Proceedings of the 6th International Conference on Biological Engineering and Medical Science
DOI: 10.54254/2753-8818/2026.38372

Current Status, Challenges, and Future Trends in Mining and Functional Characterization of Novel Protease Genes in Traditional Fermented Foods

Yuxin Jia

QAU-RAU Joint Institute for Advanced Agritechnology, Qingdao Agricultural University, Qingdao, China
yuxinjia6@gmail.com

Abstract. Traditional fermented foods are rich in microorganisms and serve as important sources of proteases that possess different biochemical and technological characteristics. During fermentation, microbial proteases help degrade proteins for the development of flavors, modification in texture, release of nutrients, and formation of bioactive peptides. New technologies in the fields of genome sequencing, metagenomics, transcriptomics, proteomics, and bioinformatics have paved the way for a shift from traditional culture-dependent screening to sequence- and function-guided protease discovery. Protease-producing Bacillus, Aspergillus, and Rhizopus are important in plant-based fermented foods, particularly fermented soybean foods, while animal-based fermented foods harbor diverse protease-producing microorganisms, including lactic acid bacteria, Staphylococcus, Tetragenococcus, Bacillus, yeasts, and molds. However, the identification of a putative protease gene does not imply its catalytic or technological role. This review summarizes the latest advances in protease gene mining, food applications, and functional characterization, identifies key methodological challenges, and outlines future prospects for next-generation food-grade proteases using integrated multi-omics, artificial intelligence, protein engineering, high-throughput screening, and synthetic biology. Special focus is placed on validation in real food matrices, salt tolerance, and substrate specificity.

Keywords: Traditional fermented foods, protease genes, genome mining, functional characterization, protease discovery

1. Introduction

Traditional fermented foods are complex systems that are influenced by raw materials, indigenous microorganisms, processing conditions, and endogenous enzymes. Fermented soybean products, vegetables, fish, meat, and dairy foods are home to a variety of bacteria, fungi, and yeasts, which play various roles in preservation, enrichment of nutritional quality, flavor development, and the production of bioactive metabolites. Proteolysis is one of the most significant biochemical reactions during fermentation as proteins are progressively degraded into peptides and free amino acids, affecting digestibility, texture, taste, aroma, and biological activity [1, 2].

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.38372

Fermentation-associated microorganisms use extracellular endopeptidases and diverse exopeptidases to hydrolyze food proteins. Adaptation to high salinity, acidity, low water activity, and fluctuating temperatures may provide proteases with useful stability or substrate specificity, making traditional fermentations valuable resources for food enzyme discovery.

Conventional discovery relies on microbial isolation and activity screening with model proteins, restricting access to cultivable organisms. Genome sequencing, shotgun metagenomics, metatranscriptomics, metaproteomics, and peptidomics broaden the search and help connect candidate genes with proteolytic activity. However, sequence annotation alone cannot establish catalytic function or technological relevance. The central challenge is therefore to build experimentally validated gene–enzyme–substrate–food phenotype relationships.

This review examines protease gene mining in plant- and animal-based traditional fermented foods, functional characterization, and food applications. It also discusses methodological challenges and future directions involving integrated multi-omics, artificial intelligence, protein engineering, high-throughput screening, and synthetic biology.

2. Current research status of protease discovery and applications in fermented foods

Proteases can be discovered from traditional fermented foods through two complementary approaches: long-term ecological selection and direct application to food processing. Their microorganisms are already adapted to particular substrates and stresses, and their proteolytic activities are often correlated with desirable fermentation outcomes. Hence, current studies increasingly combine classical activity screening with genomic and multi-omics approaches. This combination allows a shift from observing protein degradation at the food level to identifying the microorganisms, genes, and enzymes responsible. This is especially useful for identifying food enzymes that tolerate processing-related stresses and have proven technological relevance.

2.1. Mining of novel protease genes from fermented foods

Currently, protease discovery in traditional fermented foods mainly follows two complementary strategies: culture-dependent screening and culture-independent gene mining. Culture-based methods isolate microorganisms with measurable proteolytic activity and thus offer a direct link between strains and phenotype, while genome sequencing and metagenomic approaches provide access to protease genes from both cultured and uncultured members of the fermentation microbiota. These two strategies should not be seen as mutually exclusive. Sequence-based mining is more effective at broadening enzyme diversity and identifying potential genetic determinants of proteolysis, while phenotype-based screening remains useful for discovering technologically relevant strains.

Hence, the current trend is to combine phenotypic data with genomic and multi-omics information. Shotgun metagenomics can be used to recover candidate protease genes, while metatranscriptomics and metaproteomics can help determine which of these genes are transcribed and translated during fermentation [3]. This strategy is more informative than sequence annotation alone because it helps distinguish predicted protease genes from those potentially active during fermentation. Based on this, the following discussion briefly outlines the protease resources of plant- and animal-based fermented foods, both of which provide different substrate environments for enzyme discovery.

DOI: 10.54254/2753-8818/2026.38372

2.1.1. Mining of protease genes from plant-based traditional fermented foods

Fermented plant foods, particularly those made from soybeans, like soy sauce, miso, douchi, sufu, natto, tempeh, and kinema, are significant sources of proteolytic Bacillus, Aspergillus, Rhizopus, Mucor, lactic acid bacteria, and yeasts. Their protein-rich substrates and, in some products, high-salt conditions select for enzymes that have strong proteolytic capacity and environmental stability. An example of such a case is the salt-tolerant prolyl endopeptidase from Aspergillus niger, which degrades the soybean protein B3 subunit, a secondary precipitant in soy sauce, while ultrasound treatment further enhanced the degradation [4, 5]. During soybean meal fermentation, Bacillus strains isolated from kinema have also produced different peptide profiles that might exhibit antioxidant activity [6].

These studies demonstrate that plant-based fermented foods are useful not only because they contain numerous protease-producing organisms, but also because the enzymes have been selected under actual food-processing conditions. Future studies should focus less on the repeated demonstration of high protease activity and more on the identification of the genetic determinants that underlie salt tolerance, substrate selectivity, and desirable release of peptides. Thus, the use of multi-omics—including whole-genome sequencing, comparative genomics, and signal-peptide prediction, together with stage-resolved analysis—may provide a more efficient pathway from high-performing strains to the functionally important protease genes.

2.1.2. Mining of protease genes from animal-based traditional fermented foods

Animal-based fermented foods, including fermented fish, fish sauce, shrimp paste, fermented meat, dry sausages, and traditional cheeses, offer a different source of protease since they are made of muscle proteins, connective-tissue proteins, caseins, and aquatic proteins. Their microbiota commonly include lactic acid bacteria, Tetragenococcus, Staphylococcus, Bacillus, yeasts, molds, and salt-adapted microorganisms [7, 8]. Studies combining microbial profiling with peptidomics in Thai fermented fish linked bacterial succession with taste-enhancing peptides and aroma, while Bacillus subtilis fermentation of shrimp by-products has associated protease activity with changes in volatile compounds [9]. Proteolytic coagulase-negative staphylococci from fermented sausages have also been evaluated as potential starters.

Compared with plant-based systems, animal-based fermentations present an additional challenge because endogenous animal enzymes and microbial proteases act simultaneously. Therefore, phenotype-first screening followed by whole-genome sequencing and safety assessment is particularly important for this group. The priority should not be to maximize the number of detected protease genes but to identify enzymes whose substrate specificity, salt or temperature tolerance, and safety profile match defined food-processing needs. This perspective can also reduce the risk of selecting strains carrying undesirable virulence or antimicrobial-resistance traits.

2.2. Applications of proteases derived from traditional fermented foods

The proteases obtained from fermented foods are primarily applied for controlled protein hydrolysis, flavour development, and bioactive peptide production. The process of proteolysis can enhance digestibility, solubility and release precursors for the formation of aroma, kokumi and umami compounds. Characteristic volatile compounds are strongly related to protease activity in sufu [10]. Proteolytic Bacillus strains from kinema resulted in unique peptide profiles during fermentation of

DOI: 10.54254/2753-8818/2026.38372

soybean meal, and fermented miso has been shown to contain modified peptides of high bioavailability and ACE inhibitory activity [6, 11].

Proteases can also be used to modify texture, speed the maturation process, valorise by-products rich in proteins, and optimise protein functionality by targeted hydrolyses [12]. However, these benefits are conditioned on controlled, not excessive, proteolysis, as excessive hydrolysis may produce bitterness, excessive softening, or undesirable nitrogenous compounds. The practical usefulness of a food protease thus should be assessed based on its ability to produce desired changes in a specific food matrix, rather than by activity against a standard lab substrate.

The next question is how a candidate protease can be shown to have the stability, specificity, catalytic activity, and product profile needed for actual food processing. This requires functional characterization that moves beyond sequence prediction and conventional activity assays toward substrate-relevant and food-matrix-relevant validation.

3. Current status of functional characterization of novel proteases

Functional characterization is important because even if a gene is predicted to encode a protease, it cannot be used as a food enzyme until its catalytic and processing properties are confirmed. Whole-genome sequencing, or metagenomics, can be used to identify candidate genes, which are then cloned or synthesized and expressed in appropriate host cells. Rapid recombinant expression can be achieved in Escherichia coli, while Bacillus subtilis is attractive for secreted bacterial proteases and for food-related applications. For eukaryotic enzymes that need to be secreted or post-translationally processed, yeast systems like Komagataella phaffii might be more suitable.

The recombinant enzyme is purified and analysed by SDS-PAGE, together with activity assays or zymography, to verify proteolytic activity after expression. Optimum temperature and pH, thermal and pH stability, salt tolerance, effects of metal ions, inhibitor sensitivity, and substrate specificity should be included in biochemical characterization. Inhibitors can help in classification, but catalytic residues and conserved domains should also be scrutinized.

The catalytic performance of a protease should first be evaluated using conditions and substrates relevant to the target food product. Synthetic substrates are suitable for mechanistic comparison, but they lack the structural complexity, accessibility of cleavage sites, and matrix interactions of the native proteins. Thus, the potential proteases should be evaluated using suitable substrates like soy glycinin, soy β-conglycinin, casein, whey protein, collagen, myofibrillar proteins, and fish proteins. In this substrate-specific context, the kinetic parameters Km, Vmax, kcat, and kcat/Km can provide quantitative information on substrate affinity and catalytic efficiency. An example is the A. niger prolyl endopeptidase employed in soy sauce production, where, in addition to biochemical assays, the enzyme was tested directly against the actual B3 protein subunit of soybeans [4]. This progression from real-substrate selection to quantitative kinetic analysis may provide a more informative basis for predicting industrial performance than measurements based only on model substrates.

Understanding enzyme function by means of structural and computational tools is growing increasingly important. Residues related to activity, salt tolerance, thermostability, or substrate recognition can be identified by structure prediction, molecular docking, molecular dynamics, and site-directed mutagenesis. Peptidomics provides an additional level of functional information by revealing actual cleavage products and showing how an enzyme releases desirable taste-active or bioactive peptides.

Finally, the purified enzymes need to be tested in actual food matrices. Enzyme behaviour can be significantly affected by a combination of high salt, low water activity, lipids, fermentation

DOI: 10.54254/2753-8818/2026.38372

metabolites, and heterogeneous protein structures. Therefore, food-matrix validation should assess protein degradation, peptide profiles, flavor, texture, nutritive effects, and undesirable products formed. A rigorous pipeline from discovery through recombinant expression, biochemical and structural characterization, hydrolysis of specific substrates, peptidomics, and food-matrix validation, followed by appropriate safety assessment.

4. Existing challenges and future development

Despite rapid technological progress, the field remains constrained by a gap between large-scale sequence discovery and experimentally confirmed food functionality. Improving this conversion rate is essential if traditional fermented foods are to become practical enzyme-discovery platforms rather than only sources of descriptive microbiome data. The following challenges concern both the identification of gene candidates and the subsequent validation and optimization of their enzymatic functions.

4.1. Existing challenges and development trends in novel protease gene mining

Several challenges still limit efficient protease gene mining. First, sequence annotation does not equal functional identification. Proteins from the same peptidase family can vary significantly in catalytic activity, localization, and substrate preference, and metagenomic studies can result in many more candidates than can be practically tested.

Secondly, precise gene-host assignment is challenging in complex fermentation microbiomes. The fragmented assembly and ambiguous taxonomy determination of short-read sequencing data can make this assignment difficult. Long-read or hybrid sequencing approaches can improve genome reconstruction and gene-host assignment.

Third, the functional potential of uncultured microorganisms remains underexplored. Their genes can be detected by metagenomics, but heterologous expression can be unsuccessful due to codon usage, folding requirements, secretion systems, propeptide processing, or unknown cofactors. Overcoming these barriers may be possible with functional metagenomic libraries, optimized expression hosts, and cell-free systems.

Fourthly, many studies focus on the microbiota, enzyme activity, metabolites, and sensory traits separately. In the future, genome-resolved metagenomics, metatranscriptomics, metaproteomics, peptidomics, metabolomics, and fermentation phenotypes should be combined. Priority for validation should be given to genes that have been corroborated by several approaches such as DNA abundance, transcript expression, protein detection, substrate degradation, and relevant sensory changes.

AI can also enhance the candidate selection process. Candidates can be ranked on the basis of protein language models, structure prediction, or machine-learning classifiers for catalytic motifs, secretion potential, predicted stability, or substrate-binding features. But evidence-based prioritization should use AI as a support tool rather than as a replacement for thorough biochemical verification. The overall direction of the field is to shift from general sequence mining to function-driven precision mining guided by well-defined food-processing needs.

4.2. Existing problems and future trends in functional studies

There are significant limitations to functional studies as well. One is the inefficiency of heterologous expression. Novel proteases may be toxic to the host or may suffer from inclusion-body formation,

DOI: 10.54254/2753-8818/2026.38372

autolysis, incorrect propeptide processing, or inefficient post-translational mutation. Thus, expression systems must be chosen based on the enzyme source and maturation process, not necessarily on the use of E. coli by default.

Another limitation is the excessive reliance on traditional assays. Caseinolytic activity, optimum pH, and optimum temperature are important as preliminary parameters but do not always indicate industrial performance. Food proteases should be tested under realistic salt concentrations, water activities, temperatures, metabolite compositions, and authentic protein substrates.

The correlation between enzyme structure and food function should also be investigated further. The combination of 3-D structure prediction, docking, mutagenesis, and quantitative peptidomics can help to gain insight into factors governing thermostability, acid tolerance, cold activity, and/or halotolerance. These findings can then be used to enhance performance through protein engineering. For the development of enzymes for use in food, directed evolution, semi-rational design, and high-throughput screening are becoming more significant [13].

Future studies should also take note that the most active enzyme is not necessarily the best enzyme. Over-proteolysis may create bitter hydrophobic peptides, cause excessive softening of tissue, or produce excessive amounts of undesirable nitrogenous compounds. Appropriate specificity and controllable activity can be more useful than maximum activity.

Optimization is likely to be faster with the help of AI-assisted protein engineering, microfluidic screening, and synthetic biology. Food-grade microorganisms or defined microbial consortia may eventually be constructed to express the desired proteolytic systems at the desired time during the fermentation. Prior to industrial application, however, candidate strains and enzymes should be evaluated for their safety and sensory characteristics, such as virulence determinants, acquired antimicrobial resistance genes, bitterness, allergenicity, and biogenic amine formation.

5. Conclusion

Protease-producing microorganisms and novel protease genes are still under-exploited in traditional fermented food. Plant-based systems, especially soybean fermentations, supply enzymes adapted to protein-rich and often high-salt environments, while animal-based fermented foods have other proteases that show preferences for muscle proteins, connective tissue proteins, milk proteins, and aquatic proteins. Collectively, these resources indicate that traditional fermentations, besides serving as a source of microbial diversity, can be considered as natural selection systems in which microorganisms and enzymes are exposed to food-relevant stresses over long periods.

With the advent of high-throughput sequencing, the number of candidate protease genes identified has grown significantly but has also brought to light a major challenge: the number of candidate sequences has increased faster than functional validation. The principal challenge for future research is, therefore, not to identify as many protease-like sequences as possible but to convert more of the potentially useful sequences into experimentally verified enzymes, which have defined food functions. Integrated metagenomics, metatranscriptomics, metaproteomics, and peptidomics should be combined with recombinant expression and biochemical verification for establishing reliable gene–enzyme–substrate–food phenotype relationships.

Food-protease evaluation should therefore move from activity-based screening towards application-based characterization. An enzyme that has moderate activity but a desirable substrate specificity, process stability and peptide profile might be more valuable than one that shows the highest activity on a model substrate. Therefore, enzyme optimization should consider not only activity but also application-relevant properties, supported by approaches such as AI-assisted protein engineering, structure prediction, semi-rational design, directed evolution, and high-throughput

DOI: 10.54254/2753-8818/2026.38372

screening. Authentic food-matrix testing, safety assessment, and sensory evaluation should also be incorporated into the development pipeline. Finally, the most important advances will be those that link traditional fermentation biodiversity with precision enzyme discovery to select the appropriate protease for a particular food processing application, to enhance the control of food flavor, develop functional food ingredients, valorize by-products and support more sustainable food processing.

References

[1] An, F., Wu, J., Feng, Y., et al. (2023). A systematic review on the flavor of soy-based fermented foods: Core fermentation microbiome, multisensory flavor substances, key enzymes, and metabolic pathways. *Comprehensive Reviews in Food Science and Food Safety, 22*, 2773–2801. https://doi.org/10.1111/1541-4337.13162

[2] Liu, L., Chen, X., Hao, L., et al. (2022). Traditional fermented soybean products: Processing, flavor formation, nutritional and biological activities. *Critical Reviews in Food Science and Nutrition, 62*, 1971–1989. https://doi.org/10.1080/10408398.2020.1848792

[3] Xie, M., Wu, J., An, F., et al. (2019). An integrated metagenomic/metaproteomic investigation of microbiota in dajiang-meju, a traditional fermented soybean product in Northeast China. *Food Research International, 115*, 414–424. https://doi.org/10.1016/j.foodres.2018.10.076

[4] Shan, P., Ho, C. T., Zhang, L., et al. (2022). Degradation mechanism of soybean protein B3 subunit catalyzed by prolyl endopeptidase from Aspergillus niger during soy sauce fermentation. *Journal of Agricultural and Food Chemistry, 70*, 5869–5878. https://doi.org/10.1021/acs.jafc.2c01796

[5] Zhang, Z., Shan, P., Zhang, Z. H., et al. (2023). Efficient degradation of soybean protein B3 subunit in soy sauce by ultrasound-assisted prolyl endopeptidase and its primary mechanism. *Food Chemistry, 429*, 136972. https://doi.org/10.1016/j.foodchem.2023.136972

[6] Kumari, R., Sharma, N., Sharma, S., et al. (2023). Production and characterization of bioactive peptides in fermented soybean meal produced using proteolytic Bacillus species isolated from kinema. *Food Chemistry, 421*: 136130. https://doi.org/10.1016/j.foodchem.2023.136130

[7] Belleggia, L., & Osimani, A. (2023). Fermented fish and fermented fish-based products, an ever-growing source of microbial diversity: A literature review. *Food Research International, 172*: 113112. https://doi.org/10.1016/j.foodres.2023.113112

[8] Wang, Y., Chen, Q., Li, L., et al. (2023). Transforming the fermented fish landscape: Microbiota enable novel, safe, flavorful, and healthy products for modern consumers. *Comprehensive Reviews in Food Science and Food Safety, 22*: 3560–3601. https://doi.org/10.1111/1541-4337.13208

[9] Phuwapraisirisan, P., Phewpan, A., Lopetcharat, K., et al. (2024). Exploring the relationships between bacterial community, taste-enhancing peptides and aroma in Thai fermented fish (Pla-ra). *Journal of Agricultural and Food Chemistry, 72*: 10558–10569. https://doi.org/10.1021/acs.jafc.3c09003

[10] He, W., Chen, Z., & Chung, H. Y. (2022). Dynamic correlations between major enzymatic activities, physicochemical properties and targeted volatile compounds in naturally fermented plain sufu during production. *Food Chemistry, 378*: 131988. https://doi.org/10.1016/j.foodchem.2021.131988

[11] Nagao, A., Nakamoto, Y., Miyauchi, S., & Sato, K. (2024). Presence of modified peptides with high bioavailability and angiotensin-converting enzyme inhibitory activity in Japanese fermented soybean paste (miso). *Journal of Agricultural and Food Chemistry, 72*, 18942–18956. https://doi.org/10.1021/acs.jafc.4c02603

[12] Gregersen Echers, S., Jafarpour, A., Yesiltas, B., et al. (2023). Targeted hydrolysis of native potato protein: A novel workflow for obtaining hydrolysates with improved interfacial properties. *Food Hydrocolloids, 137*, 108299. https://doi.org/10.1016/j.foodhyd.2022.108299

[13] Wu, J., Wang, Z., Zeng, M., et al. (2024). Comprehensive understanding of laboratory evolution for food enzymes: From design to screening innovations. *Journal of Agricultural and Food Chemistry, 72*, 24928–24943. https://doi.org/10.1021/acs.jafc.4c08453