

**Studies on Microbial Community in Some  
Naturally Fermented Soybean Foods of the  
Eastern Himalayas using Culture-dependent  
and Metagenomics Approaches**

**A Thesis Submitted**

**To  
Sikkim University**



**In Partial Fulfilment of the Requirement for the  
Degree of Doctor of Philosophy**

**By  
Pynhunlang Kharnaioir**

**Supervisor: Professor Dr. Jyoti Prakash Tamang, *FNA, FNASc, FNAAS***

**Department of Microbiology  
School of Life Sciences  
SIKKIM UNIVERSITY  
Gangtok 737102, INDIA**

**JULY 2024**



## Highlights of Thesis

- The culture-dependent method identified *Bacillus subtilis*, *Enterococcus faecalis*, *Enterococcus faecium*, *Enterococcus lactis*, and *Pediococcus acidilactici* from naturally fermented soybean (NFS) foods of the Eastern Himalaya.
- High-throughput sequencing of *kinema* in India, Nepal, and Bhutan showed that *B. subtilis* was the most common bacterial species, *Wallemia canadensis* was the prevalent mould, and *Pichia sporocuriosa* was the main yeast.
- Shotgun metagenomic analysis of *kinema* from India, Nepal, and Bhutan revealed that the bacterial population was dominated by *B. subtilis*, followed by *B. glycinifermentans*, *B. cereus*, *Brevibacillus borstelensis*, *B. licheniformis*, *B. thermoamylovorans*, *B. coagulans*, *B. circulans*, *B. paralicheniformis*, and *B. amyloliquefaciens*.
- High-throughput sequencing identified 27 *Bacillus* species in *kinema*, while shotgun metagenomic sequencing detected 280 species of *Bacillus*, showing high species diversity.
- Metagenomic analysis NFS foods from Arunachal Pradesh revealed genomes of *B. subtilis*, *B. thermoamylovorans*, *B. velezensis*, *Enterococcus faecalis*, and *Pediococcus acidilactici*, along with biofunctional genes contributing to food fermentation.
- *B. subtilis* strains Ki52 and Kn16 were identified as promising for fibrinolytic enzyme production.
- *Pediococcus acidilactici* Ki20 was recognized as a potential probiotic strain.
- Analysis of metabolic pathways showed key features essential for flavour and aroma development, secondary metabolite biosynthesis, and vitamin production in the fermentation process.
- Metabolomes of the samples detected primary and secondary metabolites linked to flavour development and potential therapeutic effects.
- *In silico* analysis indicated the microbiome's potential role in metabolite production.

## SUMMARY

Soybean fermentation, a common practice in the Eastern Himalayas, was the focus of a study investigating the microbial community in unexplored naturally fermented soybean (NFS) foods. The research utilized both culture-dependent and culture-independent methods to isolate spore-forming bacteria and lactic acid bacteria (LAB) for assessing fibrinolytic activity and potential probiotic properties, and metagenomic sequences and metabolite profiles were used for predictive functional analysis. Aseptically collected NFS food samples from various Eastern Himalayan locations were transported to the lab in pre-sterilized packaging. The pH value of all samples ranges from slightly neutral to alkaline ( $6.60 \pm 0.48$  to  $8.24 \pm 0.01$ ). In addition, the product characterization was observed based on the stickiness ( $18.1 \pm 1.33$  to  $24 \pm 1.41$  cm) and viscosity ( $10.9 \pm 0.57$  to  $17 \pm 1.41$  cP) of the samples. The *Bacillus* count was observed in the range of  $4.3 \pm 1.63 \times 10^7$  to  $7.4 \pm 1.24 \times 10^7$ , and the LAB count ranging from  $3.9 \pm 0.70 \times 10^4$  to  $5.6 \pm 0.50 \times 10^4$ , respectively. A total of 631 spore-forming bacteria and 352 LAB isolates from NFS samples were screened and examined for cell morphology and catalase assay. Isolates were subjected to preliminary screening for fibrinolytic enzyme activity using proteolytic activity and fibrin plate assay. On the other hand, the preselection of potential probiotic isolates was checked based on acid tolerance, bile tolerance, and bacterial adhesion to hydrocarbons (BATH). After preliminary screening, 9 strains producing fibrinolytic enzymes were selected and identified as *Bacillus subtilis* by 16S rRNA gene sequencing, including four isolates from *kinema*, one isolate each from *grep-chhurpi*, *peha*, and *peruñyaan*, and two from *peron namsing*. Similarly, after a preliminary screening of potential probiotic strains, 8 isolates were selected, including two isolates each from *kinema*, *peron naming* and *peruñyaan* and one each from *grep-chhurpi* and *peha*. Genotypic identification based on 16S rRNA gene sequencing revealed the identity of isolates such as *Pediococcus acidilactici*, *Enterococcus faecium*, *Enterococcus faecalis* and *Enterococcus lactis*.

The microbial community of the *kinema* samples was further analyzed using amplicon-based high-throughput sequencing. In the bacterial community, Bacillota was the predominant phylum, Bacillaceae among the family, *Bacillus* among the genus, and *B. subtilis* among the species in *kinema* samples. *B. subtilis* was the most abundant species in the *kinema* of India and Nepal, while, interestingly, *B. cereus* was dominant in the *kinema* of Bhutan. On the other hand, the fungal community showed that Ascomycota predominated in the phylum and Wallemiaceae in the family. At the genus level, *Wallemia* and *Pichia* were the most abundant mould and yeast genera, respectively, with *Wallemia canadensis* and *Pichia sporocuriosa* among the species in *kinema* samples. *Wallemia canadensis* was the most abundant species in the *kinema* of India and Nepal, while *Pichia sporocuriosa* was the most abundant species in the *kinema* of Bhutan. Furthermore, *kinema* samples (India, Nepal and Bhutan) showed the dominance of bacterial population compared to other microbial domains after being examined using shotgun metagenomic sequencing. *Bacillus* was dominant in all *kinema* samples, with *B. subtilis* being the most abundant species overall. However, based on geographical locations, the *kinema* of India and Nepal was dominated by *B. subtilis*, while in the *kinema* of Bhutan, *B. cereus* was most abundant. Similarly, the microbial community of NFS foods (*grep-chhurpi*, *peha*, *peron namsing* and *peruñyaan*) of Arunachal Pradesh also showed the dominance of bacterial population with the abundance of genera *Bacillus*, *Staphylococcus*, and *Enterococcus*. At the species level, we observed the abundance of *Bacillus* species, mainly *B. subtilis*, some lactic acid bacteria and coagulase-negative staphylococci (CNS). The presence of both beneficial and undesirable microbes was observed in the microbial community structure, which may be due to the uncontrolled fermentation process traditionally carried out using indigenous methods. The differences in microbial composition between products can be influenced by various factors such as intrinsic and extrinsic factors, preparation method, geographical location, seasonal changes, etc.

In addition to microbial diversity, shotgun metagenomic data were subjected for metagenome-assembled genomes (MAGs) analysis to extract the potential single genome from metagenomic sequences. Seven high-quality MAGs were obtained and identified with reference to the genome taxonomy database (GTDB): such as *B. subtilis* and *B. thermoamylovorans* (*kinema*), *B. subtilis*, and *E. faecalis* (*grep-chhurpi*), *P. acidilactici* (*peha*), *B. subtilis* (*peron naming*) and *B. velezensis* (*peruñyaan*). Annotation of the genome showed that several useful genes were present in each genome. The results highlight the ability of putative genes linked to the vitamin B complex to produce and increase vitamin B12 levels. It was also found that the genomes of *B. subtilis* and *B. velezensis* contain putative genes encoding serine proteases (*apr*, *aprX*) associated with the activity of fibrinolytic enzymes, genes associated for the  $\gamma$ - polyglutamic acid production, surfactin (antibacterial and antifungal) and L-asparaginase (anticancer). In addition, *E. faecalis* (*grep-chhurpi*) and *P. acidilactici* (*peha*) showed the presence of genes associated with probiotic properties. Furthermore, the functional properties of the recovered microbial genomes showed evidence of potential areas of interest (AOI) associated with secondary metabolite gene clusters, the presence of CRISPR-Cas systems and multiple classes of carbohydrate-active enzyme families (CAZymes) including auxiliary activity, carbohydrate esterase, carbohydrate binding module, glycoside hydrolase, glycosyltransferase and polysaccharide lyase. The resulting MAGs together with biofunctional genetic markers highlighted the novelty and potential contribution of bacterial genomes in the fermentation process, which could improve the health-promoting properties of the product. The genomic information of MAG is useful for selecting bacterial genomes to understand the key features required for optimizing cultivation conditions and cultivating bacterial species in appropriate environments, which can be further improved and standardized for fermentation of fermented soybean foods.

In the other part of the study, the quantitative fibrinolytic enzyme activity was found to be more than 1000 U/ml in all strains and ranged from 1024.07 U/ml to 1230.61 U/ml. Among all isolates, two strains isolated from *kinema*, *B. subtilis* Ki52 (50.47%) and *B. subtilis* Kn16 (51.37%), recorded the highest whole blood clot-degrading properties. Furthermore, fibrin clot digestion showed that *B. subtilis* Ki52 and Kn16 had a concentration of 1.94 mg/mL and 1.02 mg/mL, respectively. It was also found that all strains producing fibrinolytic enzymes showed caseinolytic activity. Since *B. subtilis* is the main species in NFS foods, its ability to exhibit fibrinolytic activity suggests the health-promoting effect of fermented soybean foods in preventing cardiovascular diseases. The ability to exhibit caseinolytic activity may also contribute to several beneficial properties such as antibiofilm, analgesic and anti-inflammatory effects.

Likewise, LAB isolates were tested for probiotic properties based on *in vitro* properties. *E. faecium* Kn19 recorded the highest survival rate against low pH ( $73.67 \pm 1.05\%$ ), while *P. acidilactici* Ph32 recorded the highest survival rate against bile ( $79.71 \pm 0.13\%$ ) and lysozyme ( $77.76 \pm 0.25\%$ ). Similarly, *E. faecalis* Gc21 showed the highest adhesion capacity towards hydrocarbons ( $90.50 \pm 10.14\%$ ) and *P. acidilactici* Ki20 recorded  $55.73 \pm 0.96\%$  and  $75.50 \pm 2.85\%$  for the autoaggregation and coaggregation properties. Of all strains, *P. acidilactici* Ki20, *E. faecium* Kn19, and *P. acidilactici* Ph32 were able to hydrolyze both sodium taurodeoxycholate and sodium taurocholate salt. In addition, *E. faecium* Kn19, *E. faecalis* Gc21, *P. acidilactici* Ph32 and *E. faecalis* Pn37 were able to show antibacterial properties against all the tested pathogens. Genetic screening of probiotic traits by gene detection showed that LAB strains contain the genes associated with low pH survival, bile salt tolerance and adhesion. *Pediococcus* spp. showed the presence of pediocin genes (*pedA*, *pedB*) and *Enterococcus* spp. showed the presence of enterocin genes (*entA*, *entB*). *In silico* analyzes performed based on overall probiotic properties suggested that *P. acidilactici* Ki20 isolated from *kinema* could be the potential probiotic strain. Although the

contribution of LAB in fermented soybean foods is unclear, their potential probiotic properties suggest that they help minimize the population of undesirable microorganisms and can be used as co-culture(s) for food fermentation.

In the other part of the work, a predictive functional analysis of high-throughput amplicon-based sequences was performed using PICRUSt2, which revealed the dominance of metabolism, followed by genetic information processing, cellular processes, environmental information processing, human diseases and organismal systems within the bacterial community (16S rRNA); and the abundance of biosynthesis followed by degradation, utilization and assimilation as well as the production of precursor metabolites and energy in the fungal community (ITS gene). In addition, the predictive pathways were also performed for shotgun metagenomic sequences using SqueezeMeta, which demonstrated the same pattern for *kinema* (India, Nepal and Bhutan), *grep-chhurpi*, *peha*, *peron namsing* and *peruñyaan* metagenomes with the abundance of metabolism followed by environmental information processing, genetic information processing, cellular processes, human diseases, and organismal systems. In addition, enzyme classification analysis revealed the presence of genes associated with biosynthesis of  $\gamma$ -PGA, serine protease, secondary metabolites and *umami* flavour. The predominance of metabolic pathways related to amino acid, carbohydrate, lipid metabolism and biosynthesis has been reported to be related to taste development. Based on the metagenomic data, the predictive metabolic functions suggested the influence of microorganisms that may have acted as precursors for metabolic activity and metabolite production. Furthermore, a metabolomics study of NFS foods was conducted using LC-MS analysis and revealed the presence of several metabolites, including both primary and secondary metabolites such as amino acids and derivatives, carbohydrates and conjugates, fatty acids and conjugates, organic acids and derivatives, vitamins and derivatives, triterpene saponins, phenylpropanoids and polyketides, aldehydes, ketones, isoflavones, alkaloids, and other bioactive compounds. Metabolites associated with taste



development have been reported, and the presence of glutamic acid and aspartic acid is the main component for the *umami* taste of NFS foods. In addition, the presence of secondary metabolites and bioactive compounds suggests that they contribute to improving immunomodulatory and therapeutic properties (anticancer, antimicrobial, anti-inflammatory, antidiabetic, antioxidant, etc.) of NFS foods of the Eastern Himalayas. Correlation analysis examined the possible contribution of microbes to the production of metabolites. The analysis also revealed the difference between the microbial composition of NFS food samples and other fermented soybean products of Asia through comparative analysis, which may be due to multiple factors such as the substrate for preparation, geographical locations, climatic conditions, and traditional preparation methods.

In Conclusion, the examination of the microbial community in Naturally fermented soybean foods like *kinema*, *grop-chhurpi*, *peha*, *peron namsing*, and *peruñyaan*, revealed that *B. subtilis* was the predominant species, exhibiting fibrinolytic properties. Lactic acid bacteria such as *Enterococcus* and *Pediococcus* were also present in the soybean foods, showing probiotic characteristics. This research illuminates the microbial composition of Himalayan soybean foods and the retrieval of microbial genomes containing genes associated with various biofunctions, underscoring their potential as a starter culture. Metabolomic analysis revealed a range of compounds, including those responsible for flavour development and others linked to health-promoting effects such as antimicrobial, anticancer, anti-inflammatory, and immunomodulatory properties. Furthermore, the presence of multiple beneficial compounds in these foods underscores their potential therapeutic effects for consumers. The genomic information of MAGs enables the potential selection of a starter culture by isolating bacteria under supportive environmental conditions in food fermentation.