Chapter 3

Materials and Methods

3.1 Documentation of preparation of *napham*

The study was conducted in Kokrajhar town and its nearby areas. Kokrajhar is the capital city of Bodoland Territorial Region (BTR) which is an autonomous territory in the State of Assam. Kokrajhar town is the headquater of Kokrajhar District. It is inhibited by the people of several ethnic groups and cultures like Bodo, Rabha, Garo, Santhal, Assamese, Koch Rajbongshi, Nepali, Muslims, etc. Bodo is the largest ethnic community in Kokrajhar especially in the village areas. The documentation was mainly done in Bodo dominated area. The GPS locations where study was conducted are given as follows:

Kokrajhar Market : 26°23'50.50"N and 90°16'23.6"E

Ultapani Market : 26°41'38.76"N and 90°18'21.6"E

Jharbari : 26°33′48″N and 90°14′15″E

Dheer : 26°16′54″N and 90°23′21″E

Diplai : 26°17′31.56″N and 90°19′14.52"E

Aithugaon : 26°28'13.8"N and 90°16'18.12"E

Jornagra Pt. I : 26°17'3.55"N and 90°20'26.23"E

Dotma Market : 26°27'50.827"N and 90° 9' 11.35" E

Gambari : 26°30'6.426"N and 90°5'21.93"E

Kochugaon : 26°33'36.164"N and 90°4'20.94"E

Gosaingaon-Town Market : 26°26'50.04"N and 90°58' 52.31"E

Information on the traditional process of manufacturing *napham* was collected from different sources. Various aspects of this customary fermented fish product *napham*, viz. social significance, its place in the food culture of Bodos, and the economic benefit of villagers by selling this product were noted down during documentation. The interview was conducted with resourceful villagers, mainly the womenfolk, and *napham* vendors, from twelve different locations. Secondary information was collected from books, reports, dissertations, electronic and non-electronic sources. According to the information, the raw materials (fish and plants used) were collected, preserved, identified, and deposited in the Zoological Survey of

India and Botanical Survey of India, Shillong. The accession numbers of specimens were provided by the same authorized institutions.

3.2 Collection of samples

The samples were documented and collected from different locations for preliminary study but samples collected from Aithugaon (Debargaon) was selected for the biochemical and microbial (microbial enumeration and metagenomics) analysis based on the degree of hygiene level practiced. Freshly prepared samples were collected in bottles from the Aithugaon center and kept at ambient temperature for experimentation. The date and time of preparation of each of the samples were marked properly.

3.3 Biochemical Analysis

The biochemical analysis was done for proximate composition, mineral content, amino acid profile, and fatty acid profile. All the analysis was done in triplicate and the statistical analysis were done in excel and PAST software.

3.3.1 Proximate analysis

Proximate composition analysis was done for moisture content, total solids, ash content, crude fat, crude protein and pH.

3.3.1.1 Moisture content (AOAC, 2000)

Moisture content was determined by using standard method given in the standard method of AOAC, 2000. About 5g of sample was measured and taken in pre-weighed Petri plate. The sample was dried in hot air oven at the temperature of 105±2ºC for 12hours. After drying, the Petri plate was transferred in desiccators and allowed to cool down with lid partially open. The dried sample was weighed again. The weight loss was expressed as percentage of moisture content.

Calculation: Moisture %= (W1-W2)/W1 × 100

Where, W1= weight before drying and W2=weight after drying

3.3.1.2 Determination of ash content (AOAC, 2000)

The crucible and lid was placed in muffle furnace at 500^{o}C-550^{o}C overnight to remove any particles that is left over in the crucible. The crucible was cooled in desiccators (30 minutes). Moisture free sample was taken in pre-weighed crucible and incinerated in a muffle furnace at temperature of 500 ± 50°C for four to five hours. The value was expressed in percentage.

Calculation: Ash (%) = Weight of ash/Weight of sample ×100

3.3.1.3 Determination of pH

The pH of formulations in 1% w/v (1g: 100ml) water soluble portion was determined using standard glass electrode pH meter (Ikon). 1g of sample was suspended in 100ml distilled water in a glass beaker. After 1 hour the clear solution was measured for pH.

3.3.1.4 Determination of crude fat (AOAC, 2000)

Crude fat content was measured by the method described in AOAC, 2000. 10g of moisture-free sample was taken in extraction thimble. The sample was then placed in the extractor with an attached receiving flask. The solvent was poured into the thimble with the help of a glass funnel. The receiver containing petroleum benzene was heated (40°C-60°C) at such a rate that the ether dropped from the condenser to the thimble at the rate of 5 to 6 drops per second. When sufficient solvent was transferred to the extraction tubes to fill the siphon arm, it was siphoned back to the receiver. This was continued until the extraction process was completed. Then the flask was removed and the volatile solvent was evaporated at 60°C-80°C on a water bath. The least weight of residue gives a load of fat within the sample. The fat content of the sample was expressed on a wet weight basis percentage.

Calculation: % of crude fat= (W1-W2)/W1×100

Where, W1= weight before extraction and W2=weight after extraction

3.3.1.5 Determination of crude protein (Kjeldahl, AOAC, 2000)

The total nitrogen was estimated by Kjeldahl method (AOAC, 2000) and as per the company guidelines. Protein value was calculated by multiplying the total nitrogen value by a factor of 6.25. Sample of 0.2 g was transferred to digestion tube and to it 10 ml of concentrated Sulphuric acid (H_2SO_4) and 2g digestion mixture ($CuSO_4$ + KSO_4) was added. The sample was digested on Kjeldahl digestion chamber till a clear mixture was obtained. After cooling, the volume was made up to 100ml with distilled water. Then 5 ml of distilled solution was taken in Kjeldahl unit for distillation with 10 ml 40% NaOH solution. The liberated ammonia was absorbed in 2% boric acid solution containing mixed indicator (1% methyl red and 1% brohmocresol green in 1:4 ratio). The pink colour of Boric acid solution turned green. Then boric acid was titrated against N/70 standard Hydrochloric acid solution until the Boric Acid solution turned pink. Total nitrogen was calculated and expressed as g/100g of sample.

3.3.2 Analysis of Minerals and trace elements

Minerals and Trace elements were estimated according to the standard protocol, AOAC (2000). 1 gram of each of the dry powdered samples were taken in porcelain crucibles and digested at 500°C in muffle furnace. The samples were allowed to cool. To this 10ml of deionized Mili Q water was added which was followed by 3-4 drops of HNO_3.The excess of $HNO_3$was allowed to evaporate and the samples were further digested at 500°C. The digested samples were then detected for minerals by Graphite Furnace Atomic Absorption Spectrometer (GF ASS) Model: Analytic Jena Vario-6 in Sophisticated Analytical Instrument Facility **(**SAIF), NEHU, Shillong, Meghalaya**.**

3.3.3 Analysis of Amino acids

Amino acid composition was analysed by HPLC (WATERS Alliance Separations module 2695). 20µL of extracted sample was loaded on to the instrument, which is quantified using a Sigma standard. Column: Luna C18 (250 x 4.6mm; 5µ), Temperature: 27°C, Flow rate: 1.0 ml/min, Detector: PDA at 254nm, Gradient Run time: 80 min were the set up used during analysis.

3.3.4 Analysis of Fatty Acids and volatile compounds (Bligh & Dyer, 1959)

For fatty acid analysis, the Gas Chromatograph GC-2010 Plus was used. Derivatization of the fatty acids was carried out with BF3-Methanol kit. The resultant pellet after derivatization was dissolved in chloroform and further taken for GCMS analysis. Column: Spinotech RB-Wax, Column Flow rate: 1 ml/min, Injection, Volume: 1µl, Run time: 60min m/z and Range: 50-600 were the set up for the run in Gas Chromatogram. The fatty acid methy esters were then identified by comparing the retention times with authentic ones.

3.4 Microbial analysis: Enumeration of the total cultivable microbes

The microbial analysis was determined by following standard methods used by Food and Drug Administration (2001), Federation, W. E., & APH Association (2005), and the method followed by Sim et al., 2015 and Romapati, 2016.

3.4.1 Sample homogenate preparation (Sim et al., 2015)

For the preparation of sample homogenate 10g of the sample was taken. It was soaked in 90ml of physiological saline which consisted of 0.1%w/v bacteriological peptone and 0.85%w/v NaCl. The pH was adjusted at 6.1. The sample was then homogenized at 250 rpm for 3 minutes at room temperature.

3.4.2 Isolation of Micro-organism (Sim et al., 2015 and Romapati, 2016)

Bacterial and fungal communities were isolated by using selective and enriched media according to the methods specified by Sim et al., 2015 and Romapati, 2016. The *napham* samples were denoted as S1 for a 1-month old product, S2 for 2-months old product, S 6 for a 6-month old product, and S12 for a 12-month old product. The serial dilutions were prepared and the microbial communities were isolated by the pour plate method. The isolation was done in triplicates. After the incubation period, the CFU (colony forming unit) was calculated. The media used for the isolation of microbial communities were:

- For determination of the total number of viable aerobic heterotrophic bacteria in the sample, Plate count Agar (Himedia) was used and incubated at 37±1^{o}C for 48 hrs. (1 no., Annexure III)
- To determine the total halophilic count PCA was enriched with 10% NaCl. Halophiles are the organisms that thrive in high salt concentration. The inoculated sample was then incubated at 37±1^{o}C for 48 hrs. (2 no., Annexure III)
- Micro aerophilic Lactobacilli and other LAB of the samples were isolated in De Megan, Rogosa and Sharpe media (Himedia), a medium designed to favor luxuriant growth bacteria, supplemented with 2 % $CaCO_3$ at 30 ± 1^{o}C under anaerobic condition for 48-72 hours. Candle jar method was used to isolate the LAB from the sample. (3 no., Annexure III)
- Total proteolytic Count of the samples was done by using Nutrient Agar Media (Himedia) supplemented with 10% NaCl and 1% Casein hydrolysate at 37±1^{o}C for 48 hrs. Proteolytic ability is the ability of microbes to breakdown proteins into smaller peptides and amino acids. (4 no., Annexure III)
- Gram-negative pathogenic bacteria were determined in Eosin Methylene Blue Agar (Himedia) at 37±1^{o}C for 48 hrs. It is both selective and differential medium for gram-negative bacteria against gram-positive bacteria and is differential to coliform and fecal coliform bacteria. (7no., Annexure III)
- Yeast and Molds were isolated in PDA media (Himedia) at 25 ±1^{o}C for 72 hrs. (5 no., Annexure III)
- M17 Agar (Himedia) was used for isolation of Lactococci and Streptococcus. (6 no., Annexure III)

3.5 Whole metagenome study through Next Generation Sequencing

Understanding the microbial community structure is crucial for bio-prospection of food product. Metagenomics is analysis of totality of genomes of prokaryotic and eukaryotic microbial communities present in a sample by the application of sequencing technique (Pérez-Cobas, et al. 2020). Metagenomic studies through Next Generation Sequencing allow comprehensive and faster way of community analysis in food and environmental samples (Lozupone, 2013). Whole-genome shotgun (WGS) metagenomics is a technique that not only sequences all genomes existing in the samples, but the functional capabilities of the studied microbial community (Pérez-Cobas, et al. 2020). For the metagenomic study the samples were outsourced to AgriGenome private limited, Kochi, Kerela. The workflow for whole genome metagenomics in the present study was Metagenomic DNA extraction, DNA quality control, fragmentation, library preparation, quality check, sequencing in Illuminia HiSeq SBS (https://www.illumina.com/science/ technology/next-generation-sequencing/sequencing-technology.html)
and bioinformatics pipeline which included *denovo* metagenomic assembly. Three samples were collected from Aithugoan (26°28'13.8"N and 90°16'18.12"E) and sample codes used were S1, S2 and S3. S1 was one month old sample; S2 was 2 months old sample and S3 was the finished proted. From the local literatures available and the belief of local people these two months are regarded as crucial for the production of good *napham*. After two and half months to three months the product is ready to be consumed after cooking. So, these two months were taken for WGS metagenomics study and to analyse the biodiversity and the functional capabilities of microbial community in *napham*.

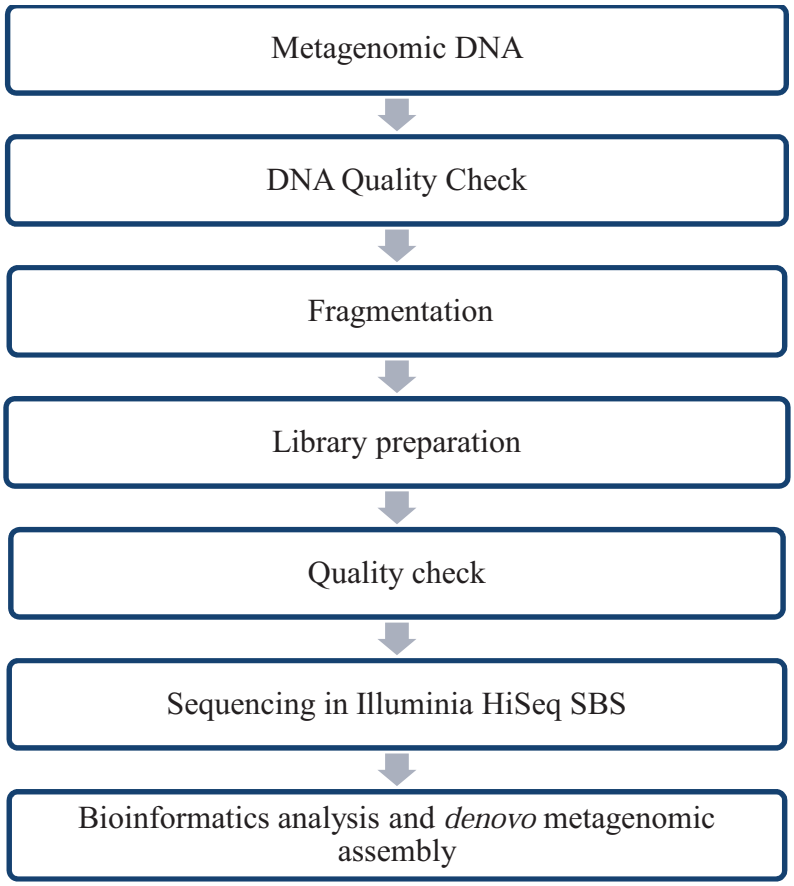

Figure 3.1 Workflow of Whole Metagenome sequencing

3.5.1 DNA extraction

DNA was extracted from the samples for whole genome metagenomic analysis using the Power Soil DNA kit according to the manufacturer's protocols. The concentration and quality of extracted DNA were determined in both Qubit 3.0 Fluorimeter and Nanodrop 8000 spectrophotometer at 260-280nm, and evaluated with a 2% agarose gel.

3.5.2 Library construction and whole metagenome sequencing

The libraries were generated with NEBNext® Ultra™ DNA Library Prep Kit for Illumina and quantification & QC was determined with TapeStation and Qubit (Qubit 3.0 Fluorimeter). High quality Genomic DNA was fragmented and Paired END Library (2x150 bp PE) was prepared according to Illumina protocol (Minoche, 2011).The DNA libraries was validated by Bioanalyzer and sequenced on Illumina HiSeq System.

3.5.3 Data Analysis and Bioinformatics

The following Bioinformatics steps were performed for *De novo* Whole Genome Metagenome assembly and analysis.

3.5.4 Fasta QC & Filter

The fastq read quality was assessed and necessary filters were applied to retain high quality reads. In addition, the low-quality sequence reads are excluded from the analysis. Following analysis was done before assembly: a) Base quality Score distribution, b) Sequence quality score distribution, c) Average base content per read, d) GC distribution in the reads and d) Adapter trimming. The adapter trimming was performed using Cutadapt (version - 1.8.1).

3.5.5 *De novo* Assembly

The de novo assembly of the adapter trimmed fastq files is carried out using MetaSPAdes (v3.10.1) (Nurk et al., 2017). MetaSPAdes construct de Bruijn graph (Compeau et al., 2011) and incorporate several modules which can be used optionally and independently for different processing and assembly steps. The contigs obtained from the assembly are used as input to MetaGene Annotator (MGA) (Noguchi et al., 2008) (http://metagene.nig.ac.jp/) for the prediction of open reading frames (ORFs). The gene-finding program MGA was used to identify the coding regions and distinguish them from noncoding DNA. The ORFs were taken for the functional annotation and taxonomic classification.

3.5.6 Protein Homology Search

The predicted ORFs are searched against the non-redundant (NCBI-nr) data base using DIAMOND (v0.7.9.58) (Buchfink et al., 2015). DIAMOND is a new high-throughput program for aligning a file of short reads against a protein reference database such as NR, at 20,000 times the speed of BLASTX, with high sensitivity.

3.5.7 Functional Annotation

The alignment file along with the filtered ORFs are used as in put to the functional annotation using MEGAN6 (MetaGenome ANalyzer) software. MEGAN6 (Beier et al., 2017) is a fast and easy tool for metagenomics data analysis and can be used to analyze and compare metagenomics and metatranscriptomic data, both at taxonomical and functional level. To explore the metabolic potential of microbial community during fermentation gene or protein functions of all the ORF from DIAMOND BLASTX output was parsed using in-house PERL script. Further, functional annotation of all the contigs are carried out by SEED Classification. MEGAN software was used to assign the function of each contigs. The protein functions of each contig having highest alignment score from DIAMOND BLASTX results were

considered for functional assignment. The metagenome sequences were submitted to the NCBI GeneBank with sample ID SAMN17224118 and SAMN17224117.

3.6 *16S rRNA* Metagenome analysis

In *16S rRNA* Metagenome study the target organism is prokaryotes containing *16S* ribosomal RNA (rRNA) sequence. In this analysis, amplicon sequencing of the 16S rRNA as a phylogenetic marker is used. The 16S ribosomal RNA (rRNA) sequence is composed of nine hypervariable regions interspersed with conserved regions. The 16S gene contains highly conserved sequences between hypervariable regions, enabling the design of universal primers and for taxonomic classification. With the development of high-throughput sequencing platforms, sequence variation in the 16S gene is widely used to characterize diverse microbial communities (Caporaso et al., 2011; Youssef et al., 2009; Hess et al., 2011). Two finished product samples were collected from Aithugoan (26°28'13.8"N and 90°16'18.12"E) and sample codes were S3 and S4. S3 was four to five months old and S4 was five to six months old and collected from the market place of Aithugaon. This study was performed to study the bacterial diversity and composition in finished products of *napham*. The samples were outsourced for sequencing and bioinformatics pipeline in Hydrabad (Pathcare Private Limited). The flowchart of procedure is given in Figure 3.2.

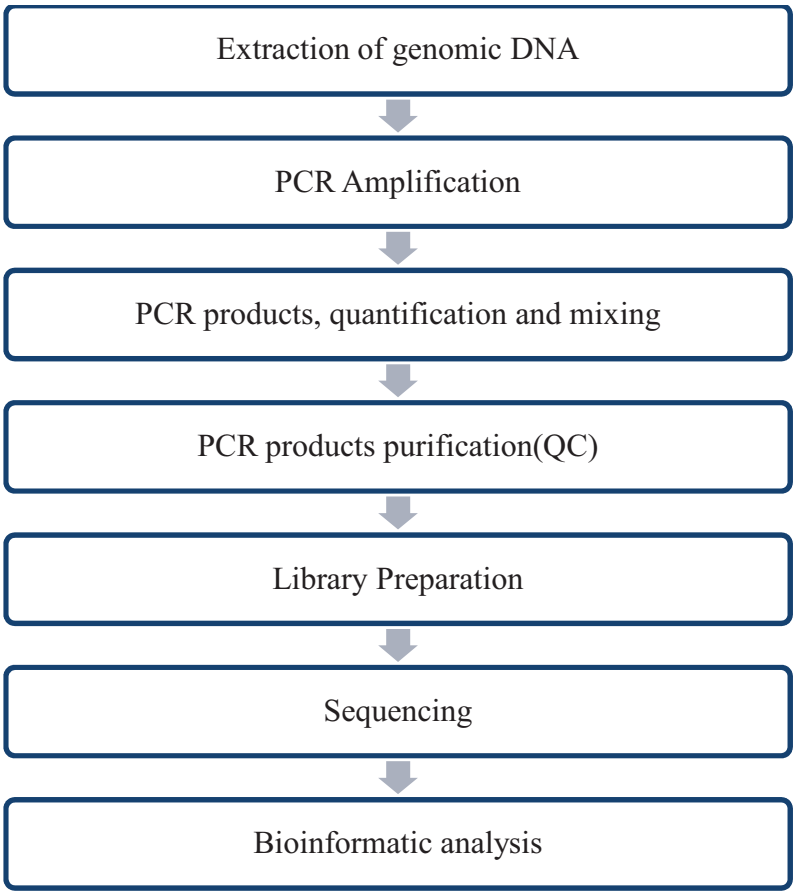

Figure 3.2 The workflow of 16S rRNA metagenome sequencing

3.6.1 Extraction of genomic DNA and Amplicon Generation

Total genome DNA from samples was extracted using CTAB/SDS method (Verma, 2017). According to the concentration, DNA was diluted to 1ng/µl using sterile distilled water. DNA purity and was monitored on 1% Agarose gel and concentration with Qubit 3.0 Fluorimeter at A_{260}/A_{280}. 16S rRNA genes of distinct regions (16SV3-V4) were amplified using specific barcoded primers CCTAYGGGRBGCASCAG, GGACTACNNGGGTATCTAAT. All Polymerase chain reaction (PCR) reactions were carried out with Phusion® High-Fidelity PCR Master Mix (New England Biolabs) and procedure was carried out according to the manual guide.

3.6.2 PCR Products quantification and qualification

To verify the amplification the same volume of 1X loading buffer (contained SYB green) was mixed with PCR products and operated in electrophoresis on 2% agarose gel. Samples with bright main strip between 400-450bp were chosen for further experiments.

3.6.3 PCR Products Mixing and Purification

PCR product was mixed in equidensity ratios. Then, mixture PCR product was

purified with Qiagen Gel Extraction Kit (Qiagen, Germany). The libraries generated with NEBNext® Ultra™ DNA Library Prep Kit for Illumina and quantified via Qubit and Q-PCR, was analysed by Illumina platform.

3.6.4 Sequencing and data processing

Paired-end reads were assigned to samples based on their unique barcode and truncated by cutting off the barcode and primer sequence. Paired-end reads were merged using FLASH (V1.2.7) (http://ccb.jhu.edu/software/FLASH/) (Magoč & Salzberg, 2011). Quality filtering on the raw tags was performed under specific filtering conditions to obtain the high-quality clean tags (Bokulich et al., 2013) according to the QIIME (V1.7.0) (Caporaso et al., 2010) quality controlled process. These were compared with the reference database (Gold database) using UCHIME algorithm (UCHIME Algorithm) (Edgar et al., 2011) to detect and remove chimera sequences. The chimera sequences were removed (Haas et al., 2011) and then the Effective Tags were finally obtained.

Software used:

- http://qiime.org/scripts/split_libraries_fastq.html
- http://drive5.com/uchime/uchime_download.html
- http://www.drive5.com/usearch/manual/uchime_algo.html

3.6.5 OTU cluster and Species annotation

Sequences analysis was performed by UPARSE software (UPARSE v7.0.1001) (http://drive5.com/uparse/) (Edgar, 2013) using all the effective tags. Sequences with ≥97% similarity were assigned to the same OTUs. Representative sequence for each OTU was screened for further annotation. For each representative sequence, Mothur software was performed against the SSU rRNA database of SILVA Database (Wang et al., 2007) for species annotation at each taxonomic rank (Threshold: 0.8~1) (kingdom, phylum, class, order, family, genus, species). To get the phylogenetic relationship of all OTUs representative sequences, MUSCLE (Edgar, 2004) (Version 3.8.31) was utilised to compare multiple sequences rapidly. OTUs abundance information was normalized using a standard of sequence number corresponding to the sample with the least sequences.

3.6.6 Alpha Diversity

Alpha diversity is measured to study the diversity within a particular ecosystem. To study the Alpha diversity, 6 indices were used in analyzing complexity of species

diversity for a sample which included Observed-species, PD_whole_tree-PD_whole_tree index, Chao1, Shannon, Simpson, ACE, Good's-coverage. All these indices for the samples were calculated with QIIME (Version 1.7.0) and displayed with R software (Version 2.15.3). The 16S Metagenome sequences were then submitted to GeneBank with sample ID SAMN20088403.

Software used:

- http://qiime.org/scripts/split_libraries_fastq.html
- http://scikitbio.org/docs/latest/generated/generated/skbio.diversity.alpha.chao1.html#skbio.diversity.alpha.chao1