CHAPTER 5

MOLECULAR IDENTIFICATION AND PHYLOGENY OF POTENTIAL PROBIOTIC LACTIC ACID BACTERIA ISOLATED FROM MALAYSIAN FERMENTED FOOD

5.1 Introduction

Despite having an energy-providing diet, the consumer has switched to better food categories, which include diets with adequate nutrients and health benefits such as preventing infections in the digestive system, improving lactose metabolism, reducing cholesterol, and stimulating calcium absorption (Panghal et al., 2018). Probiotic foods are an example of a food which falls under this category. Examples of probiotic groups that have mainly exhibited health advantages include *Lactobacillaceae*, *Bifidobacterium*, *Saccharomyces*, *Enterococcus*, *Streptococcus*, *Pediococcus*, and *Bacillus* (Fijan, 2014). The benefits of probiotics have been shown to be diverse and strain-specific (Halder et al., 2017; Ida Muryany et al., 2017; Lim et al., 2017). In fact, the use of probiotic strains singly or in combinations provides different reactions to the host; some produce lower impact when used in combinations (Herbel et al., 2013). Thus, each particular strain must be identified correctly to prevent the use of probiotics lacking function as desired.

The conventional biochemical and physiological methods turn out to have some constraints in distinguishing isolates displaying comparable physiological attributes, thus may produce false identification. Similar strains with different phenotypic patterns;

non-reactivity of isolates to test; and whether isolates were not explicitly identified or were misidentified are some of the issues and arguments surrounding biochemical identification (Herbel et al., 2013). Identification using a genotypic approach is more appropriate to properly identify isolates prior to any biotechnological or industrial applications (Adesulu-Dahunsi et al., 2017). For example, several bacterial strains were mis-identified as *Aeromonas sobria* using the biochemical test, but molecular identification detected the strains as *Aeromonas hydrophila* and *Aeromonas veronii* (Diyana-Nadhirah and Ina-Salwany, 2016). Other than that, two LAB were identified as *Bacillus megaterium* and *Pediococcus pentosaceus* when using a biochemical identification kit, but identification through molecular techniques demonstrated that the isolates were *Lactobacillus plantarum* and *Lactobacillus pentosus* (Ida Muryany, 2017).

Thus far, the genotypic identification of LAB isolated from belacan and budu is accessible but limited (Liasi et al., 2009; Abbasiliasi et al., 2011; Sim et al., 2012; Haitham, 2017; Khalil et al., 2018), whilst there is no study that has been documented on the molecular identification of LAB isolated from bosou. Therefore, this study aimed to identify the LAB strains found in these samples. The seventeen LAB strains isolated from belacan (BE3, BE7 and BE16), bosou (BO1, BO8, BO10, and BO16) and non-cooked budu (BUM5, BUM6, BUM7, BUM12, BUM14, BUM15, BUM18, BUM22, BUM23, and BUM24), which exhibited various probiotic properties, were subjected to molecular characterisation using two genes named the 16S ribosomal RNA (16S rRNA) gene and the internally transcribed spacer (ITS) gene. The outcomes using both genes were compared to evaluate the most precise species identification.

5.2 Materials and Methods

5.2.1 DNA Extraction

The genomic DNA of each isolate was extracted using the One-Tube Bacterial Genomic DNA Extraction Kit (Bio Basic, Canada) in accordance with the manufacturer's protocol with slight modifications. In brief, the overnight culture grown in MRS broth was adjusted to an optical density of 2.0 at 600 nm. Then, 150 µL of the bacterial suspension was transferred into a microcentrifuge tube and centrifuged at 13000 rpm for 30 seconds. The supernatant was discarded before 100 µL of Lysis-Buffer-B was added and vortexed gradually to prevent DNA smearing. The sample was further incubated at 65°C for 10 min to allow complete digestion. After incubation, 100 µL of Universal Buffer NST was added and vortexed to mix it thoroughly. The mixture was used as a DNA template directly, while the remaining was stored at -20°C.

5.2.2 Amplification of 16S rRNA and ITS Genes

The extracted DNA was used as a template to amplify a segment of the 16S rRNA and ITS genes, respectively, by the polymerase chain reaction (PCR) technique using the universal primers synthesised by Macrogen Laboratory, South Korea (Table 5.1). The PCR mixture consisted of 0.5 µl of forward and reverse primers, respectively, 1 µL template DNA and 12.5 µL MyTaq Red Mix (Bioline, UK). The mixture was added with autoclaved distilled water until the volume was 25 µL. The negative controls consisted of all PCR mixture components and autoclaved distilled water to replace the template DNA, while *Lacticaseibacillus casei* strain Shirota (Yakult, Japan) was used as a reference strain. The PCR reactions were performed using a T100 thermal cycler (Bio-Rad, US) following the standard MyTaq Red Mix protocols. For 16S rRNA gene PCR amplification, the cycle was performed as follows, a cycle of initial denaturation at 95°C (1 min), followed by 35 cycles of denaturation at 95°C (15 sec), annealing at 50°C (15 sec), and extension at 72°C (10 sec). The final extension was set at 72°C for 5 min before the reaction ended and was held at 12°C. Meanwhile, amplification conditions for the ITS gene consisted of 1.5 min of initial denaturation at 95°C, followed by 35 cycles of denaturation at 95°C for 15 sec, annealing at 60°C for 15 sec, and extension at 72°C for 10 sec, with a final elongation at 72°C for 5 min.

Table 5.1: Universal Primers Used in The DNA Amplification

Gene	Universal Primers	Sequences (5' – 3')*	Fragment size (bp)	Reference
16S rRNA	27F 1492R	AGAGTTTGATCMTGGCTCAG TACGGYTACCTTGTTACGACTT	1500	Lane (1991)
ITS	ITS_F ITS_R	CGGTGAATACGTTCCCGGGYCTTG TTTCRCCTTTCCCTCACGGTA	600-700	Ina-Salwany et al. (2015)

* A:adenine; C:cytosine; G:guanine; T:thymine; M: A/C; R:A/G; Y:C/T.

5.2.3 PCR Products Analysis

Electrophoresis's amplicons obtained through PCR reactions were analysed using HyAgarose™ LE Agarose gel (HydraGene Co., Ltd., China). In the preparation of 1.2% w/v agarose gel, 1.2 g agarose powder was melted in 100 ml of 1× TAE buffer (Promega, US) until dissolved, before 3 µL ViSafe Red Gel Stain (Vivantis, Malaysia) was mixed thoroughly. The solidified gel was placed in the gel electrophoresis tank filled with 1× TAE buffer and the gel chamber wells were loaded with 6 µL of amplicons, respectively. In order to analyse the sizes of DNA samples qualitatively, 1kb DNA ladder (Promega, US) mixed with DNA gel loading dye (Thermo Scientific, US) was also loaded as a reference. The agarose gel electrophoresis was run at 85 V for 45 min, and the bands were visualised under the Gel Documentation System (Uvitec Cambridge, UK).

5.2.4 Sequence Analysis and Phylogenetic Tree Construction

The PCR products were commercially sequenced by Macrogen Laboratory (South Korea). The amplified 16S rRNA and ITS gene sequences were further used for the analysis of sequence similarity through the Basic Local Alignment Search Tool (BLAST) from the National Center for Biotechnology Information (NCBI). The nucleotide sequences were imported and aligned with MEGA X software (Kumar et al. 2018). The comparison of the isolates with closely related species available in the GenBank database *(*http://www.ncbi.nlm.nih.gov/genbank*)* and the construction of a phylogenetic tree was made using the Maximum Likelihood method (Felsenstein 1981) with 1000 replications. Then, all nucleotide sequences were deposited into GenBank.

5.3. Results

5.3.1 Amplification of Potential Probiotic LAB Isolates

PCR amplification using primers for 16S rRNA and ITS genes was successfully performed with DNA isolated from all seventeen LAB isolates. Both sets of primers used in the study provided good gene amplification for all the isolates. The band patterns of all amplicons and the reference strain for both genes are shown in Figures 5.1 and 5.2, respectively. The band size of all amplicons using 27F and 1492R primers has corresponded to the expected size of the 16S rRNA gene, which is approximately 1500 bp (Figure 5.1). In contrast, the band size of amplicons using ITS_F and ITS_R primers was roughly 750 bp, which matched the ITS gene's size (Figure 5.2). All gel chamber wells loaded with negative control did not show any bands due to the absence of amplified DNA or contamination of debris, whilst the reference strain showed bands approximately at the same level as the isolated LAB strains.

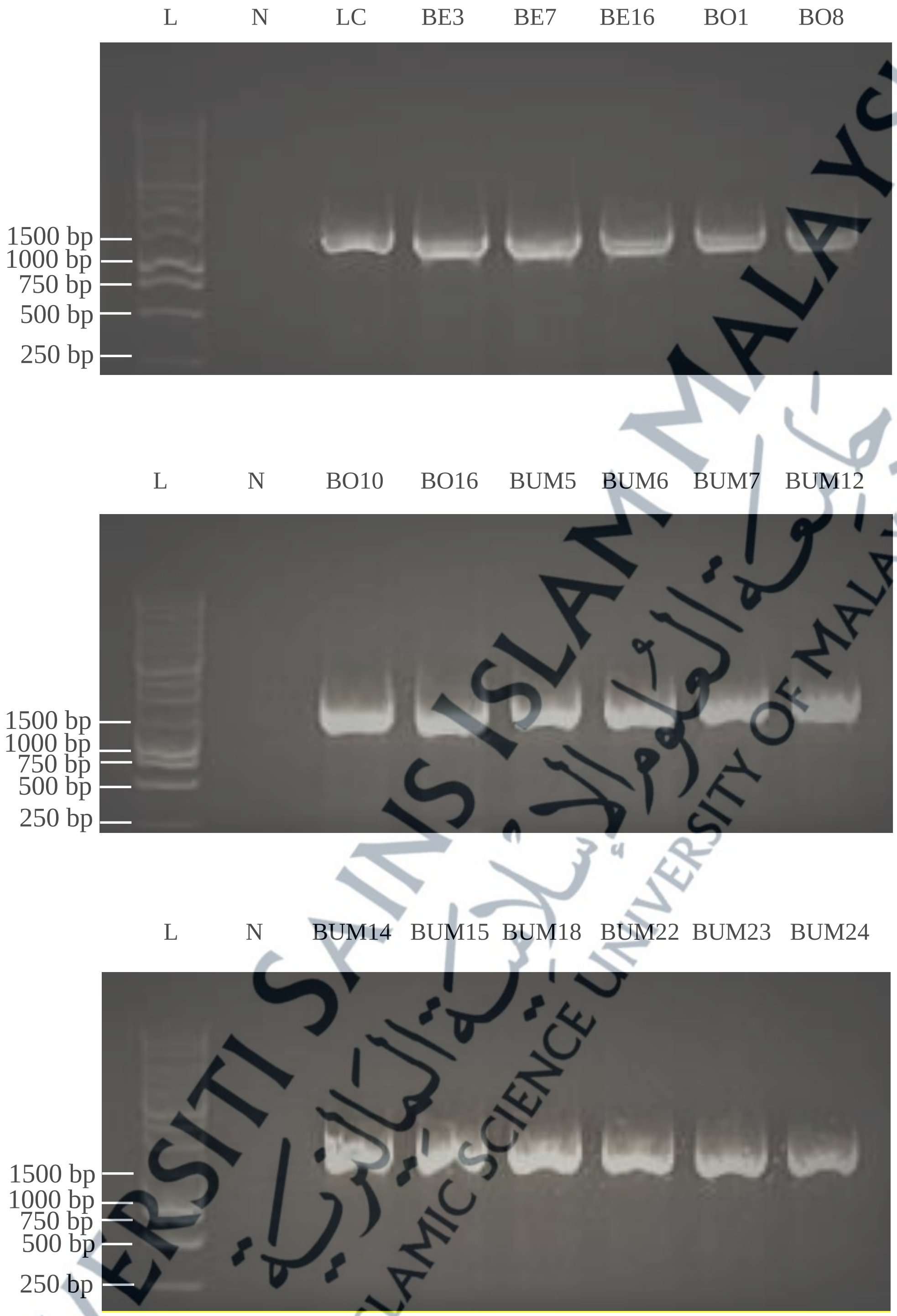

Figure 5.1: Agarose Gel Electrophoresis Analysis of PCR Amplification of the 16S rRNA Gene of BE Isolates, BO Isolates, BUM Isolates and *L. casei* Strain Shirota Using 27F and 1492R Primers

Abbreviations: L: 1kb DNA ladder; N: negative control (sterile distilled water); LC: reference strain (*L. casei* strain Shirota).

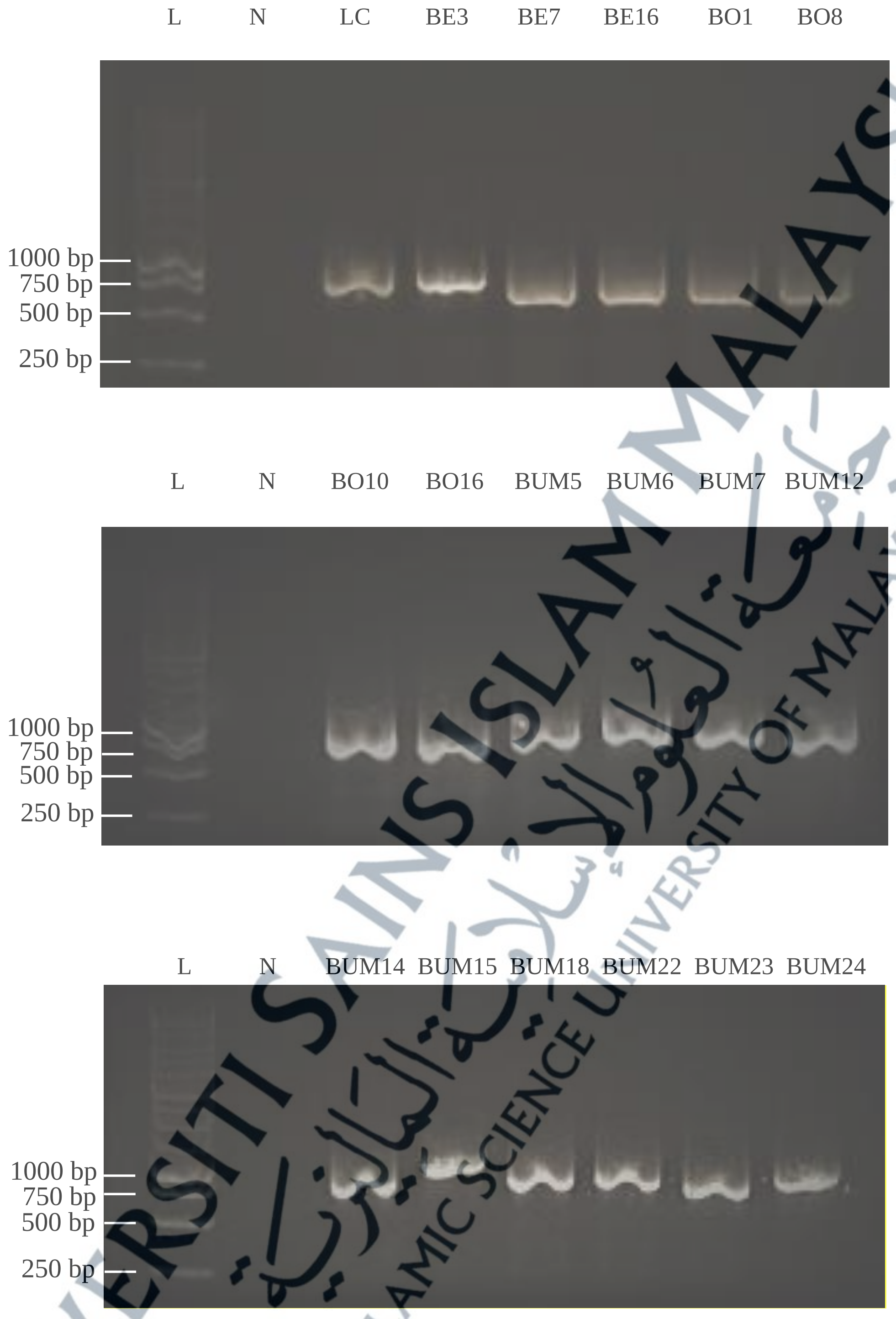

Figure 5.2: Agarose Gel Electrophoresis Analysis of PCR Amplification of the 16S rRNA Gene of BE Isolates, BO Isolates, BUM Isolates and *L. casei* Strain Shirota Using ITS_F and ITS_R Primers

Abbreviations: L: 1kb DNA ladder; N: negative control (sterile distilled water); LC: reference strain (*L. casei* strain Shirota).

5.3.2 Identification and Phylogeny Analysis of Potential Probiotic LAB Isolates Using 16S rRNA gene

The results of seventeen LAB isolates based on 16S rRNA sequences and their accession numbers are shown in Table 5.2. The BLASTn results of 16S rRNA gene sequences revealed that all BE and BO isolates were identified as *Lactiplantibacillus plantarum*, with a minimum of 98% similarity to the nucleotide sequences database available in the GenBank. The lineage reports showed that BE and BO isolates recorded a high number of hits (81.2% to 93.5%) to similar species according to the available taxonomy database. Meanwhile, the BUM strains were identified as *Lacticaseibacillus paracasei* with at least 98% similarity to the GenBank nucleotide sequences database. The lineage reports of BUM strains showed a broader number of hits (58.0% to 87.4%) in between the focus strain and the available *Lacticaseibacillus paracasei* sequences, which generated the taxonomy database. A phylogenetic tree of 16S rRNA sequences of the 17 LAB isolates and 14 *Lactobacillaceae* strains obtained from the Genbank was constructed using MEGA X and displayed in Figure 5.3. *Staphylococcus aureus* D83357.1 was used as the outgroup. The phylogenetic tree showed that all BUM isolates were in a separate branch from the BE and BO isolates. All BE and BO isolates were grouped with *L. plantarum* and *Lactiplantibacillus. pentosus* strains from the GenBank (accession no. MZ476213, D79211, and LC638737). Meanwhile, the BUM isolates were clustered together with other *L. paracasei* strains from GenBank (accession no. NR025880 and MW924095).

Table 5.2: Identified LAB Isolates by 16S rRNA Gene Sequencing with Their GenBank Accession Number

Isolate	Species	Lineage report (%)	Accession number
BE3	*Lactiplantibacillus plantarum*	93.5	MT163337
BE7	*Lactiplantibacillus plantarum*	81.2	MT163338
BE16	*Lactiplantibacillus plantarum*	84.0	MT163339
BO1	*Lactiplantibacillus plantarum*	92.6	MT163340
BO8	*Lactiplantibacillus plantarum*	81.9	MT163358
BO10	*Lactiplantibacillus plantarum*	84.8	MT163341
BO16	*Lactiplantibacillus plantarum*	90.2	MT163342
BUM5	*Lacticaseibacillus paracasei*	65.2	MT163343
BUM6	*Lacticaseibacillus paracasei*	62.6	MT163344
BUM7	*Lacticaseibacillus paracasei*	79.2	MT163345
BUM12	*Lacticaseibacillus paracasei*	58.0	MT163346
BUM14	*Lacticaseibacillus paracasei*	79.4	MT163347
BUM15	*Lacticaseibacillus paracasei*	94.0	MT163359
BUM18	*Lacticaseibacillus paracasei*	69.0	MT163348
BUM22	*Lacticaseibacillus paracasei*	87.4	MT163360
BUM23	*Lacticaseibacillus paracasei*	82.9	MT163349
BUM24	*Lacticaseibacillus paracasei*	81.3	MT163350

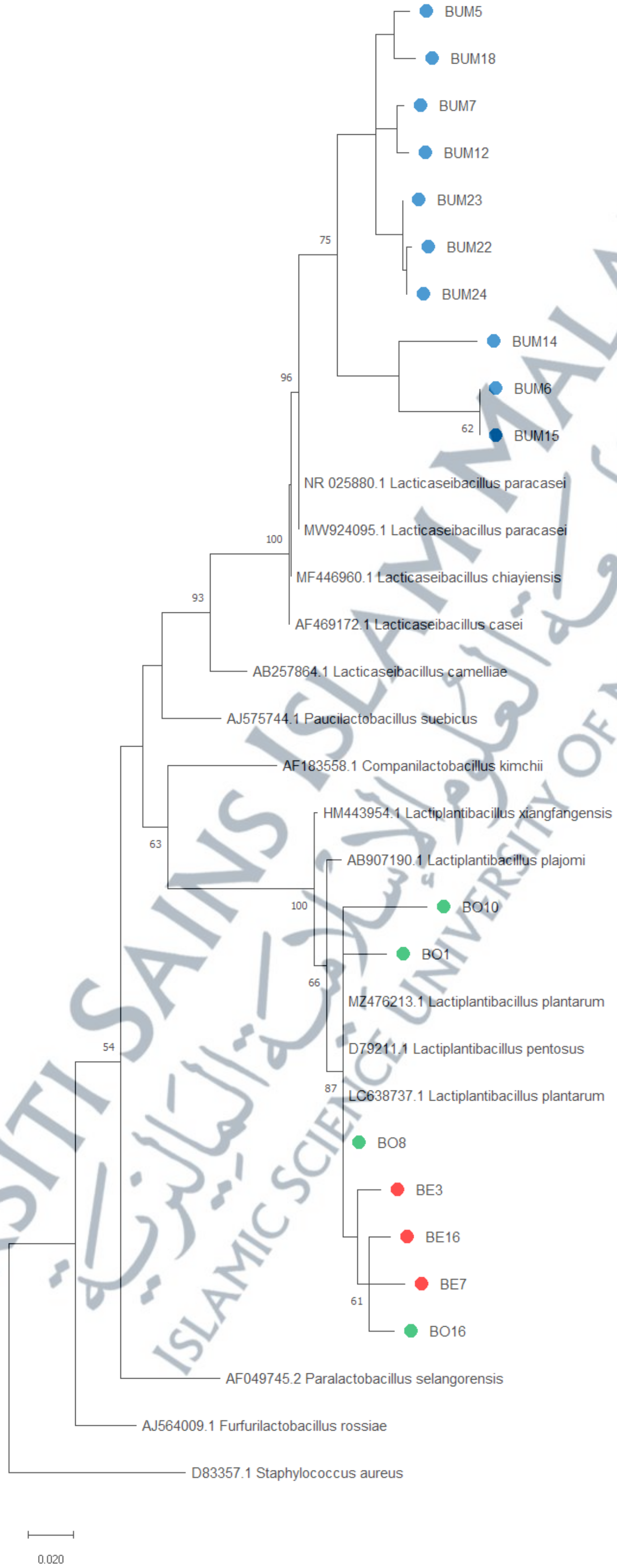

Figure 5.3: Maximum-likelihood Phylogenetic Tree of BE Isolates, BO Isolates, BUM Isolates, and Reference Strains Based on the 16S rRNA Gene Sequences

5.3.3 Identification and Phylogeny Analysis of Potential Probiotic LAB Isolates Using ITS gene

The identified LAB isolates based on ITS gene sequencing and their accession numbers are shown in Table 5.3. The BLASTn results of ITS gene sequences supported the prior identification of both BE and BO isolates using the 16S rRNA gene, which was identified as *L. plantarum*, with at least 98% similarity to the GenBank nucleotide sequence database. All BUM strains were also identical to *L. paracasei* with a minimum of 98% similarity to the GenBank nucleotide sequence database. The lineage reports revealed that all *L. plantarum* and *L. paracasei* strains showed similar percentages of hits (88.9% and 40.4%) to the taxonomy database available in the GenBank, respectively. A phylogenetic tree of ITS sequences of the seventeen LAB isolates and ten *Lactobacillaceae* strains obtained from the Genbank is shown in Figure 5.4. *S. aureus* DQ256396 was used as the outgroup. The branches formed for BUM isolates were almost identical to those constructed using the 16S rRNA gene, in which the isolates were clustered together with the reference ITS sequence of *L paracasei* (accession no. AB109028). Meanwhile, BE and BO isolates have formed a separate branch, together with the reference *L. plantarum* strain from the GenBank (accession no. AF080101 and AF429622).

Table 5.3: Identified LAB Isolates by ITS Gene Sequencing with Their GenBank Accession Number

Isolate	Species	Lineage report (%)	Accession number
BE3	*Lactiplantibacillus plantarum*	88.9	MT212063
BE7	*Lactiplantibacillus plantarum*	88.9	MT212064
BE16	*Lactiplantibacillus plantarum*	88.9	MT212065
BO1	*Lactiplantibacillus plantarum*	88.9	MT212080
BO8	*Lactiplantibacillus plantarum*	88.9	MT212081
BO10	*Lactiplantibacillus plantarum*	88.9	MT212082
BO16	*Lactiplantibacillus plantarum*	88.9	MT212083
BUM5	*Lacticaseibacillus paracasei*	40.4	MT212055
BUM6	*Lacticaseibacillus paracasei*	40.4	MT995194
BUM7	*Lacticaseibacillus paracasei*	40.4	MT212056
BUM12	*Lacticaseibacillus paracasei*	40.4	MT212057
BUM14	*Lacticaseibacillus paracasei*	40.4	MT212058
BUM15	*Lacticaseibacillus paracasei*	40.4	MT212059
BUM18	*Lacticaseibacillus paracasei*	40.4	MT212060
BUM22	*Lacticaseibacillus paracasei*	40.4	MT212061
BUM23	*Lacticaseibacillus paracasei*	40.4	MT995195
BUM24	*Lacticaseibacillus paracasei*	40.4	MT212062

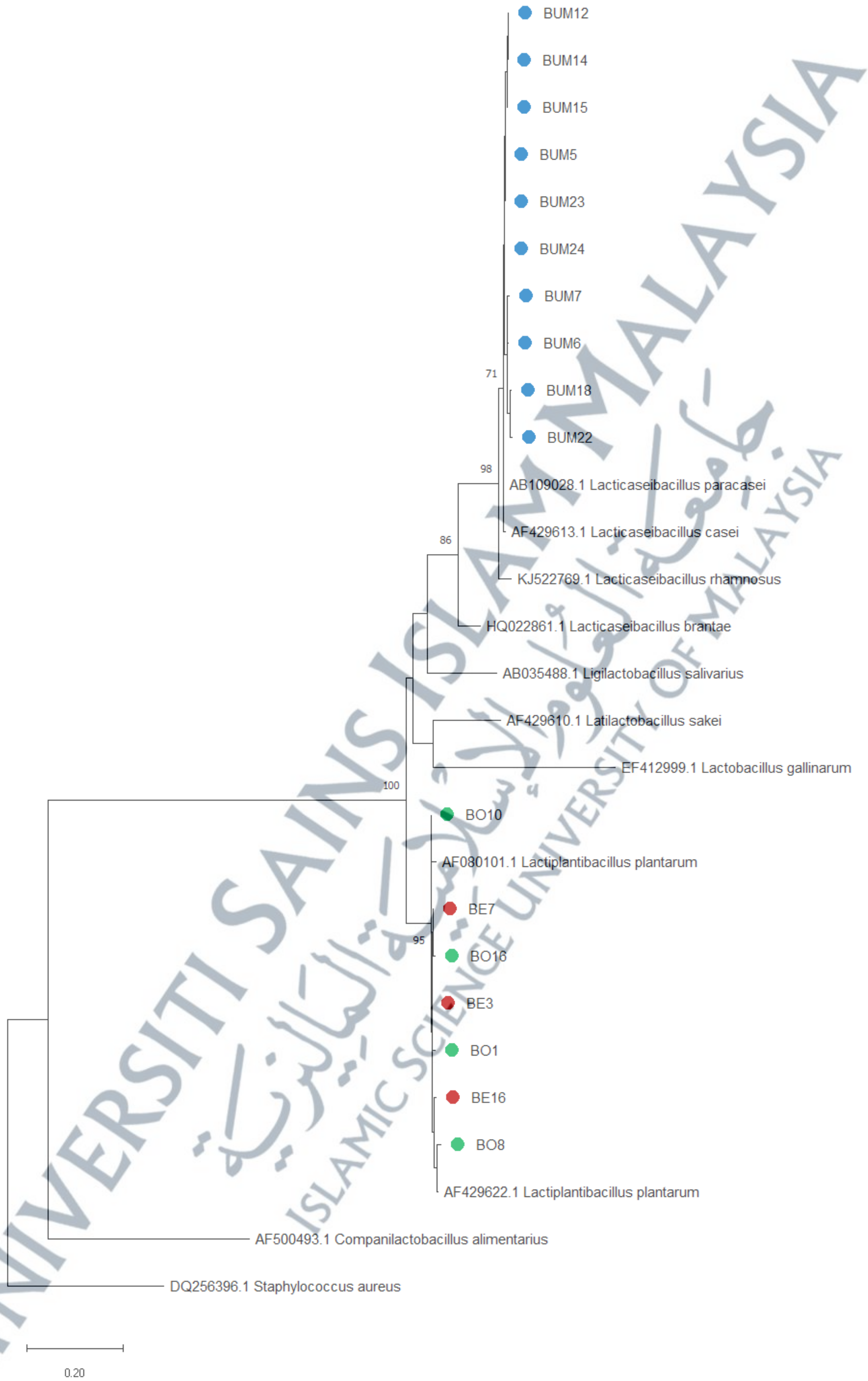

Figure 5.4. Maximum-likelihood Phylogenetic Tree of BE Isolates, BO Isolates, BUM Isolates and Reference Strains Based on the ITS Gene Sequences

5.4 Discussion

Research on bacteria with probiotic properties is widely done globally as this microbiota has demonstrated promising benefits (Han et al. 2017; Todorov et al. 2017; Yasiri et al. 2018). In this study, the identification of strains isolated from the fermented food was done using molecular techniques. The identification using the 16s rRNA gene could confirm the identity of all BE and BO isolates as different strains of *L. plantarum* species with at least 98% similarity to the GenBank nucleotide sequences database. All BE and BO isolates were further identified using the ITS gene to support the molecular identification using the 16s rRNA gene. Meanwhile, the correct species for BUM isolates was also confirmed by the use of both genes and was identified as a different strain of *L. paracasei*. The identification of these isolates is based on the reclassification of the genus *Lactobacillus* into 25 genera (Zheng et al., 2020). The reclassification of the genus was based on the extreme difference of phenotypic, ecological and genotypic levels of all 261 species as of March 2020. It is vital to ensure that identifying each particular strain is made precisely as the specific health advantages and their safety profiles have been shown to be strain-specific (Fijan 2014). At the moment, molecular approaches are considered fundamental and crucial methods to identify microorganisms. This method is vital to accurately identifying the genus and species of unknown isolates (Shehata et al. 2016; Ida Muryany et al. 2017). An earlier study by Diyana-Nadhirah and Ina-Salwany (2016) concluded that the identification based on 16S rRNA gene sequencing could classify strains up to genus level, whereas ITS gene sequencing can discriminate strains up to species level.

The absence of other LAB species in belacan, bosou, and non-cooked budu might be due to their inability to grow in the food samples or MRS broth during prior isolation of LAB. Moreover, the *L. plantarum* found in belacan and bosou, as well as *L. paracasei* isolated from non-cooked budu, might be dominant compared to the other species and might have entirely inhabited the samples (Haitham 2017). This view has supported the study by Kopermsub and Yunchalard (2010), who reported that the early stages of Thai fermented fish making were dominated by *Lactococcus garvieae*, *Streptococcus bovis*, and *Weissella cibaria*. However, *L. plantarum* started to become prevalent after a certain period before dominating the samples until the completion of fermentation. Lactobacilli are non-spore-forming gram-positive and rod-shaped bacteria, with facultatively anaerobic or microaerophilic, acid-tolerant, and catalase-negative characteristics (Huang et al. 2018).

Organisms that fall under the *Lactobacillaceae* group are widely used as probiotics, such as *L. casei* and *Lactobacillus acidophilus*, found in commercial probiotic drinks Yakult and Vitagen, respectively. The isolation of LAB under the *Lactobacillaceae* group from fermented shrimp and fish is not surprising. Previously, *Lactobacillus fermentum*, which was a strain that could produce exopolysaccharides, was isolated from budu (Khalil et al., 2018). Meanwhile, probiotic LAB identified as *L. plantarum, L. acidophilus*, and two other species from different genera named *Enterococcus faecium* and *Pediococcus acidilactici* have been isolated from belacan (Haitham, 2017). Similarly, Le and Yang (2018) concluded that one of the dominant isolates of their salted-fermented shrimp was *L. plantarum,* besides the other three isolates, which were close to *L. sakei.* However, our study was able to isolate multiple strains of *L. plantarum* from the fermented shrimp product..

By contrast, an earlier study by Kobayashi et al. (2003) using 16S rRNA molecular identification found two different species in their terasi shrimp of genus *Tetragenococcus*, which were *T. muriaticus* and *T. halophilus*, while Lee et al. (2014) was able to isolate three *P. pentosaceus* strains from jeotgals, salted and fermented

Korean seafood. Another study indicated that the microbial flora of Thai traditional fermented shrimp paste (kapi) was varied but was shown to exclude the presence of LAB (Daroonpunt et al., 2016). A more recent study of Indonesian traditional fermented whole fish by Putra et al. (2018) has identified the *Aerococcus* genus through morphology and physiology tests.

Genotypic characterisation of bacteria based on 16S rRNA gene sequence has been used broadly due to the fact that it exists in nearly all bacteria, it has a sluggish evolutionary rate and constant role, as well as the great size of a gene which is roughly 1500bp, that was adequate for bioinformatic purposes (Janda and Abbott 2007). The use of this gene is considered a common and fundamental method for the molecular identification and classification of bacteria. However, the gene is highly conserved, thus a high similarity index of the 16S rRNA gene sequence could affect the identification of bacteria at the genus and/or species level (Tokajian et al. 2016). The impact can be seen more obviously if the studied bacteria of interest are phylogenetically very closely related to one another.

Due to this difficulty, another phylogenetic marker is needed to correctly identify bacterial species. ITS, or "Internal transcribed spacer", is the spacer region positioned in the middle of the small-subunit rRNA and large-subunit rRNA genes (Felsberg et al. 2015). There is only a single ITS presence in bacteria, located between the 16S rRNA and the 23S rRNA. Thus, it is also called the 16S-23S intergenic spacer region (Liguori et al. 2011). When compared to 16S and 23S rRNA genes, The ITS region was the shortest yet the most discriminatory region for identification, differentiation, and systematic analysis of bacterial species and strains (Song et al. 2000; Man et al. 2010). The ITS region’s variability might be due to differences in the base length and sequences, thus offering better information for identifying bacteria (Tokajian et al.

2016). Variations found in ITS sequences are sufficient to surmount the 16S rRNA gene sequence's inadequacy in resolving closely related isolates. In fact, as most bacteria retain multiple alleles of the ribosomal operon in their complete genome, a significant sequence variation in the ITS region may exist in a different strain of the same species (Felsberg et al. 2015). This study's results are consistent with these arguments; the ITS gene's use seemed to be suitable for all BE, BO, and BUM isolate strains' identification and differentiation.

The ITS gene has been reported as a marker to identify bacterial species other than the 16S rRNA gene. Earlier, Song et al. (2000) were able to sequence the ITS region of several *Lactobacillus* species usually found in the human intestine, with amplicons sized between 600 and 700 bp. In another study by Diyana-Nadhirah and Ina-Salwany (2016), the ITS region's amplification on all *Aeromonas* spp. strains isolated from freshwater fishes produced an amplicon of between 1000 and 1200 bp. The different amplicon sizes may be due to the copy number of ribosomal units and tRNA-encoding genes within ITS (Singh et al. 2012). Nevertheless, prior molecular identification using the 16S rRNA gene is still needed as the sequences obtained must be matched to the available sequence from the GenBank database for correct identification. The fact that should be emphasised is that the absence of comparable sequences in the GenBank database will cause lower disability to identify the microorganism, as depicted by the lower number of hits in the lineage reports for BUM strains using the ITS gene, compared to the percentages using the 16s rRNA gene. The lineage report describes the relationships between the organisms that generates the database and the focus organism that provided the strongest BLAST hit (NCBI, 2023).

As variation in lengths and sequences of ITS regions occurs in different bacteria species, this sequence can also be used to design species-specific PCR primers. For example, the PCR primers were designed from the ITS region to be highly sensitive and specific for *Salmonella* spp. but generate negative results for non-*Salmonella* strains (Chiu et al., 2005). Other than that, primers for detection and discrimination of *L. garvieae* strains from other closely related species isolated from fish and aquaculture environments have also been designed (Dang et al., 2012). However, the primer pairs used in this study are general and non-species specific, which could detect various species from multiple genera such as *Paenibacillus* spp., *Paenibacillus pabuli, Aeromonas hydrophila, Aeromonas veronii,* and *Bacillus amyloliquefacien* (Ina-Salwany et al. 2015; Diyana-Nadhirah and Ina-Salwany 2016; Azrin et al. 2017), as well as the *Lactobacillaceae* strains which have been identified from the BE, BO, and BUM isolates.

5.5 Conclusion

In conclusion, the study confirmed the presence of LAB strains found in Malaysian fermented foods named belacan, bosou, and non-cooked budu. The BE and BO isolates were classified as different strains of *L. plantarum*, and the ten BUM strains were identical to *L. paracasei*, all with at least 98% similarity to the GenBank nucleotide sequence database. There is an association of results obtained by 16S rRNA and ITS genes for the identification of BE, BO and BUM isolates. The ITS gene, which is not commonly used in the molecular identification of bacterial strains, could effectively identify the *Lactobacillaceae* strains up to the species level. More studies are required to unravel the potential probiotic characteristics of each strain and their potential to be applied in therapeutics.