1 Short Communication

2 Predictive Metabolic Pathways of Lactic Acid Bacterial Strains Isolated from 3 Fermented Foods

4 Pynhunlang Kharnaor¹, Prakash M. Halami² and Jyoti Prakash Tamang^{1*}

5
6 ¹DAICENTRE (DBT-AIST International Centre for Translational and Environmental
7 Research) and Bioinformatics Centre, Department of Microbiology, School of Life Sciences,
8 Sikkim University, Gangtok 737102, Sikkim, India

9 ²CSIR-Central Food Technological Research Institute, Microbiology and Fermentation
10 Technology, Mysuru, 570020, Karnataka, India

11 *Corresponding author: Professor Jyoti Prakash Tamang (e-mail:jyoti_tamang@hotmail.com)

13 Abstract

14 We attempted to use PICTRUST2 software and bioinformatics tool to infer the raw sequences
15 obtained from pure strains of *Lactococcus lactis* and *Lactobacillus plantarum* isolated from
16 some fermented foods in India, which were identified by 16S rRNA gene sequencing method.
17 Predictive metabolic pathways of 16S sequences of LAB strains were predicted by PICRUST2
18 mapped against KEGG database, which showed genes associated with metabolism (36.74%),
19 environmental information processing (32.34%), genetic information processing (9.86%) and
20 the unclassified (21.06%). KEGG database also showed the dominant genes related to
21 predictive sub-pathways of metabolism at level-2 were membrane transport (31.16%) and
22 carbohydrate metabolism (12.42%).

23
24 **Keywords:** Predictive functionality, Metabolic Pathways, Fermented Foods, LAB,
25 PICTRUST,

26

27 **Introduction**

28 Ethnic fermented foods of India have socio-cultural values, rich gastronomy and also
 29 contribute health benefits to consumers (Tamang 2020), and are composed by diverse types of
 30 bacteria, fungi and yeasts (Tamang et al. 2012). Lactic acid bacteria present in fermented
 31 foods have several functional properties, therapeutic uses and health promoting benefits
 32 (Tamang et al. 2016; Goel et al. 2020). Phylogenetic investigation of communities by
 33 reconstruction of unobserved states (PICRUST2), an omics-based machine learning software
 34 in bioinformatics (Douglas et al. 2019), is commonly used to predict the genes markers of
 35 sequences generated by high-through sequences and shot-gun sequences from food samples
 36 (Tamang et al. 2020). However, there is a limited application of PICRUST2 to predict
 37 functional metabolic pathways in sequences obtained from isolated bacterial strains
 38 (Medvecky et al. 2018). In the present study, we analysed sequence of five lactic acid bacteria
 39 isolated from Indian fermented and sun-dried foods to predict the metabolic pathways by
 40 bioinformatics tools.

41

42 **Materials and Methods**

43 Three different types of Indian fermented foods were collected from Sikkim and Karnataka
 44 states in India viz. *sidra* (dried fish products of Sikkim), *kinema* (fermented soybean food of
 45 Sikkim), and *dahi* (fermented milk product of Karnataka and Sikkim). Ten gram each of
 46 samples was homogenized in 90 mL of 0.85% physiological saline using stomacher lab
 47 blender 40 (Seward, United Kingdom) for few min, and were plated MRS (Man-Rogosa-
 48 Sharpe) agar (M641, HiMedia, Mumbai, India), and incubated at 30°C for 24–48 h. About 60
 49 isolates were randomly isolated and preliminarily tested for Gram-stain, catalase test, arginine
 50 hydrolysis and other phenotypic characteristics (Pradhan and Tamang 2019). Out of 60
 51 isolates, five isolates were grouped on the basis of phenotypic characteristics viz. isolate FS2
 52 (*sidra*), C2D (*dahi*-Mysore), DHCU70 (*dahi*, Gangtok), SP2C4 (*kinema*) and KP1 (*kinema*).

53 Genomic DNA of 5 strains were extracted following the method of Cheng and Jiang (2006).
 54 Amplified 16S rDNA was obtained from each strain by polymerase chain reaction (PCR) with
 55 the universal primers in a Thermal cycler (Applied Biosystems-2720, USA). Sequencing of
 56 the amplicons was performed in an automated DNA Analyzer (ABI 3730XL Capillary
 57 Sequencers, Applied Biosystems, Foster City, CA, USA). Taxonomy of bacterial strains was
 58 assigned by comparing the DNA sequences in NCBI (National Center for Biotechnology
 59 Information) database using BLAST (basic local alignment search tool) 2.0 program and the
 60 phylogenetic tree was constructed by neighbor-joining method using MEGA7.0 software
 61 (Kumar et al. 2016).

62

63 **Results and Discussion**

64 The 16S rRNA gene sequences of five strains were analysed for the functionality using
 65 PICRUST2-algorithm (Douglas et al. 2019). The sequences were aligned for placement of the
 66 study sequences into a reference phylogeny, and the gene copy number of test sequences were
 67 effectuated at default parameters (Barbera et al. 2019). Annotation of gene function and
 68 pathways inference with high level function (Ye and Doak 2009) were mapped against KEGG
 69 (Kyoto Encyclopaedia of Genes and Genomes) database (Kanehisa et al. 2017). Raw reads
 70 were normalized into relative abundances and data visualization using MS-Excel 365.

71 Based on results of 16S rRNA gene sequences, a phylogenetic tree was constructed using
 72 neighbour-joining for identification of bacterial strains (Fig. 1). FS2 strain isolated from
 73 *sidra*, C2D strain isolated from *dahi* and SP2C4 isolated from *kinema* were identified as
 74 *Lactococcus lactis*, and DHCU70 strain isolated from *dahi* and KP1 strain isolated from
 75 *kinema* were identified as *Lactobacillus plantarum* (Fig. 1). The functional features of
 76 sequences of 5 strains of LAB obtained from 16S rRNA data were predicted by using
 77 PICRUST2 mapped against KEGG database. The overall functional features of LAB strains
 78 showed genes associated with metabolism (36.74%), environmental information processing

(32.34%), genetic information processing (9.86%) and the unclassified (21.06%) at Level-1 (Fig. 2a). At the Level-2, Dominant genes related to predictive sub-pathways of metabolism at level-2 were membrane transport (31.16%) and carbohydrate metabolism (12.42%) (Fig. 2b). KGGE database showed tyrosine metabolism was dominant in *Lactococcus lactis* (Fig. 3a), indicating its role in flavour development in the product since tyrosine is an aromatic amino acids (Parthasarathy et al., 2018). Predictive genes in *Lb. plantarum* showed the folate biosynthesis (Fig. 3b), that can possibly play an important key in the development of drug against infectious disease (Bertacine Dias et al. 2018). Phosphotransferases system (PTS), the source of transport and phosphorylation of various sugar and other sugar derivatives in bacteria (Deutscher et al. 2006), was detected only in *Lb. plantarum*. We did not detect any predictive genes for human diseases in *Lb. plantarum* and *Lc. lactis* strains. The ATP-binding cassette transporters (ABC transporters) genes were dominant in all five strains, which facilitate the selective uptake of di- and tripeptides (Doeven et al. 2005). We can conclude that predictive metabolic pathways (Baranwal et al. 2020) can be inferred more in details of even isolated strains by KGGE database which may help to understand functionality of the LAB strains in foods.

95

96 **Acknowledgement**

97 This work was supported by Department of Biotechnology, Government of India for
98 Twinning Research, project no. BT/PR16706/NER/95/259/2015.

99

100 **COMPLIANCE WITH ETHICAL STANDARDS**

101 *Conflict of interests.* The authors declare that they have no conflict of interest.

102 *Statement on the welfare of animals.* This article does not contain any studies involving
103 animals or human subjects performed by any of the authors.

104

105 DATA AVAILABILITY

106 The sequences retrieved from the 16S rRNA sequencing were deposited at GenBank-NCBI
107 under the nucleotide accession number: MT503218, MT503219, MT503220, MT503221
108 and MT503222.

109

110 References

- 111 Baranwal M, Magner A, Elvati P, Saldinger J, Violi A, Hero AO. 2020. A deep learning
112 architecture for metabolic pathway prediction. *Bioinform* 36(8): 2547–2553.
- 113 Barbera P, Kozlov AM, Czech L, Morel B, Darriba D, Flouri T, Stamatakis A. 2019. EPA-ng:
114 massively parallel evolutionary placement of genetic sequences. *Sys Biol* 68(2): 365-369.
- 115 Bertacine Dias MV, Santos JC, Libreros-Zúñiga GA, Ribeiro JA, Chavez-Pacheco SM. 2018.
116 Folate biosynthesis pathway: mechanisms and insights into drug design for infectious
117 diseases. *Future Med Chem* 10(8): 935-959. Deutscher J, Francke C, Postma PW. 2006.
118 How phosphotransferase system-related protein phosphorylation regulates carbohydrate
119 metabolism in bacteria. *Microbiol Mol Biol Rev* 70: 939–1031.
- 120 Doeven MK, Kok J, Poolman B. 2005. Specificity and selectivity determinants of peptide
121 transport in *Lactococcus lactis* and other microorganisms. *Mol Microbiol* 57: 640–649.
- 122 Douglas GM, Maffei VJ, Zaneveld J, Yurgel S., Brown JR, Taylor CM, Huttenhower C,
123 Langille MG. 2019. PICRUSt2: An improved and extensible approach for metagenome
124 inference. *BioRxiv* 672295. doi.org/10.1101/672295.
- 125 Goel A, Halami PM, Tamang JP. 2020. Genome analysis of *Lactobacillus plantarum* isolated
126 from some Indian fermented foods for bacteriocin production and probiotic marker
127 genes. *Front Microbiol* 11: 40.doi.org/10.3389/fmicb.2020.00040.
- 128 Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives
129 on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 45(D1): D353-D361.

130 Kumar S, Stecher G, Tamura K. 2016. MEGA 7: molecular evolutionary genetics analysis
131 version 7.0 for bigger datasets. *Mol Biol Evol* 33: 1870–1874.

132 Medvecky M, Cejkova D, Polansky O, Karasova D, Kubasova T, Cizek A, Rychlik I. 2018.
133 Whole genome sequencing and function prediction of 133 gut anaerobes isolated from
134 chicken caecum in pure cultures. *BMC Genom* 19: 561.doi.org/10.1186/s12864-018-
135 4959-4.

136 Parthasarathy A, Cross PJ, Dobson RCJ, Adams LE, Savka M, Hudson AO. 2018. A three-
137 ring circus: metabolism of the three proteogenic aromatic amino acids and their role in
138 the health of plants and animals. *Front Mol Biosci* 5:29. doi: 10.3389/fmolb.2018.00029

139 Pradhan P, Tamang JP. 2019. Phenotypic and genotypic identification of bacteria isolated
140 from traditionally prepared dry starters of the Eastern Himalayas. *Front Microbiol*
141 10:2526. doi:10.3389/fmicb.2019.02526.

142 Tamang JP. 2020. History and culture of Indian ethnic fermented foods and beverages. *In*
143 *Ethnic Fermented Foods and Alcoholic Beverages of India: Science History and Culture*,
144 Tamang JP (ed), Springer Nature, Singapore, pp. 1-40.

145 Tamang JP, Cotter P, Endo A, Han NS, Kort R, Liu SQ, Mayo B, Westerik N, Hutkins R.
146 2020. Fermented foods in a global age: east meets west. *Compreh Rev Food Sci Food*
147 *Safe* 19: 184-217.

148 Tamang JP, Shin DH, Jung SJ, Chae SW. 2016. Functional properties of microorganisms in
149 fermented foods. *Front Microbiol* 7: 578. doi: 10.3389/fmicb.2016.00578.

150 Tamang JP, Tamang N, Thapa S, Dewan S, Tamang B, Yonzan H, Rai AK, Chettri R,
151 Chakrabarty J, Kharel N. 2012. Microorganisms and nutritional value of ethnic
152 fermented foods and alcoholic beverages of North East India. *Indian J Trad Know* 11: 7-
153 25.

154 Ye Y, Doak TG. 2009. A parsimony approach to biological pathway reconstruction/inference
155 for genomes and metagenomes. PLoS Com Biol 5:
156 8.doi.org/10.1371/journal.pcbi.1000465.

157

158

159 **Figure legend:**

160 **Fig. 1:** Neighbor-joining phylogenetic analysis using MEGA7 based on 16S rRNA gene
161 sequences of lactic acid bacterial strains (FS2, C2D, SP2C4, KP1 and DHCU70) isolated
162 from fermented foods with bootstrap values (expressed as percentage of 1000 replicates).
163 *Bacillus subtilis* ATCC 6051T (Genbank accession number JF749278.1) was used as an out-
164 group.

165 **Fig. 2:** Overall predictive functional features of LAB strains via PICRUSt2 mapped against
166 KEGG database. A bar-plot for (a) Level-1 and (b) Level-2, and a heatmap of predominant
167 pathways (c) Level-3 were plotted.

168 **Fig. 3:** Doughnut representation were plotted in comparison between the two species
169 identified (a) *Lactococcus lactis* and (b) *Lactobacillus plantarum*.

170

171