Draft genome sequences of four lactic acid bacteria from fermented chicken meat unveil biosynthetic gene clusters for antimicrobial compounds

Mahabub Alam[a], Tamanna Hassan[b], Tanvir Ahmad Nizami[a], Lipi Akter[c], Tofazzal Md Rakib[a#]

a. Faculty of Veterinary Medicine, Chattogram Veterinary and Animal Sciences University, Chattogram, Bangladesh.

b. Faculty of Food Science and Technology, Chattogram Veterinary and Animal Sciences University, Chattogram, Bangladesh.

c. Transboundary Animal Diseases Research Center, Faculty of Veterinary Medicine, Kagoshima University, Kagoshima, Japan.

Running title: Biosynthetic gene clusters of lactic acid bacteria

#Address correspondence to Tofazzal Md Rakib, rakibtofazzal@cvasu.ac.bd.

Abstract

Lactic acid bacteria play a crucial role in fermented food production and serve as important sources of antimicrobial peptides. This study reports four lactic acid bacteria strains, isolated from fermented chicken meat, which harbor biosynthetic gene clusters encoding antimicrobial compounds. These strains are classified within the genera *Pediococcus* and *Lactiplantibacillus*.

Announcement

Fermented foods have been consumed for centuries, serving as carriers for probiotics, prebiotics, and postbiotics. Lactic acid bacteria (LAB) play a crucial role in fermentation by enhancing food safety, inhibiting spoilage, and offering probiotic benefits like improved immunity and gut health (1, 2). With rising antibiotic resistance, bacteriocins are gaining attention as potential alternatives (3). Genome mining and high-throughput screening of biosynthetic gene clusters present a promising pathway for discovering novel bacteriocins, addressing resistance challenges, and advancing sustainable food preservation (4, 5). Three broiler chickens from a live bird market in Chattogram, Bangladesh, were slaughtered under standard procedures. Breast meat (100 g) was stored anaerobically at 4°C for three days to propagate probiotic bacteria. One gram of meat was mixed with 5 mL MRS broth (Oxoid, Hampshire, England) and incubated anaerobically at 37°C for 48 hours, promoting LAB growth. Single colonies were isolated on MRS agar, purified, and confirmed by PCR. Genomic DNA was extracted from 1.5 mL of 24-hour cultured broth using a AllPrep Bacterial DNA/RNA/Protein Kit (Qiagen, Hilden, Germany), and purity was assessed with a Nanodrop One (Thermo Fisher Scientific, MA, USA). DNA libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA). Genome sequencing was conducted on an Illumina NextSeq2000 platform at the Poultry Research and Training Center, Chattogram Veterinary and Animal Sciences University, Bangladesh, generating 2 × 150 bp paired-end reads. Raw data quality,

35 including read quality, GC content, and adapter contamination, was assessed using FastQC v0.12.1 (6).
36 Trim Galore v0.6.5dev (7) was employed to trim low-quality bases and adapter sequences, followed by
37 normalization of read coverage with bbnorm (8) to reduce redundancy. De novo genome assembly was
38 performed using Unicycler v0.4.8 (9), with subsequent polishing using Pilon v1.24 (10) to enhance
39 accuracy. Genome assembly quality metrics were evaluated using QUAST v5.2.0 (11). Samtools v1.13 (12)
40 was used for alignment file manipulation and indexing to aid downstream analysis. The de novo genome
41 assembly, conducted through the Bacterial and Viral Bioinformatics Resource Center (13) comprising
42 above tools, yielded coverage ranging from 65.4× to 145.6×, with raw reads between 1,452,975 and
43 1,662,521. The assembled genomes were annotated using NCBI PGAP v6.9 (14) and PATRIC (15), with
44 transport proteins identified via TCDB (16), drug targets predicted using DrugBank v6.0 (17), and
45 additional analyses performed for sequence types using MLST (18, 19), CRISPR arrays and prophages using
46 CRISPRCasFinder (20), and Integrative and Conjugative Elements using ICEFinder (21). Detailed genomic
47 characteristics are provided in Table 1.

48 The genomes of four LAB isolates were mined for biosynthetic gene clusters (BGCs) associated with
49 antimicrobial production using the software antiSMASH 7.0 (22) and BAGEL4 (23). The genomes were
50 found to contain multiple BGCs encoding ribosomally synthesized and posttranslationally modified
51 peptides (RiPP like compounds), Cyclic Lactone Autoinducer, Type III Polyketide Synthase (T3PKS),
52 Terpene Precursors, Terpene, and Class IIa (Pediocin) and IIb (Plantaricin_J) bacteriocins. The isolates or
53 their purified bacteriocins may be used in the future for food preservation and commercial probiotics to
54 manage gut health and related diseases.

55 Data availability statement

56 The accession numbers for the genomes of the four LAB strains are shown in Table 1. All genomes are
57 available under BioProject accession number PRJNA1189463.

58 Acknowledgments

59 The genome sequencing was supported by Chattogram Veterinary and Animal Sciences University WGS
60 establishment project.

61 Funding

62 This study was supported by a grant from the University Grants Commission of Bangladesh through
63 Chattogram Veterinary and Animal Sciences University, Bangladesh (Grant number: 32, Session: 2023-
64 2024).

65 **Author ORCIDs**

66 Mahabub Alam https://orcid.org/0000-0001-9837-2598
67 Tamanna Hassan https://orcid.org/0009-0009-4875-9386
68 Tanvir Ahmad Nizami https://orcid.org/0000-0002-7024-4970
69 Lipi Akter https://orcid.org/0000-0002-6317-9410
70 Tofazzal Md Rakib https://orcid.org/0000-0003-2642-9908

71 References

72 1. Zhao J, Zhao J, Zang J, Peng C, Li Z, Zhang P. 2024. Isolation, identification, and evaluation of
73 lactic acid bacteria with probiotic potential from traditional fermented sour meat. Frontiers in
74 Microbiology 15:1421285. https://doi.org/10.3389/fmicb.2024.1421285.

75 2. Fernandes A, Jobby R. 2022. Bacteriocins from lactic acid bacteria and their potential clinical
76 applications. Applied Biochemistry and Biotechnology 194:4377-4399.
77 https://doi.org/10.1007/s12010-022-03870-3.
78 3. Cleveland J, Montville TJ, Nes IF, Chikindas ML. 2001. Bacteriocins: safe, natural antimicrobials
79 for food preservation. International journal of food microbiology 71:1-20.
80 https://doi.org/10.1016/S0168-1605(01)00560-8.
81 4. Gerst M, Yousef A. 2018. Modified microassay for the isolation of antimicrobial-producing,
82 spore-forming and nonspore-forming bacteria. Journal of applied microbiology 124:1401-1410.
83 https://doi.org/10.1111/jam.13724.
84 5. Zhou H, Fang J, Tian Y, Lu XY. 2014. Mechanisms of nisin resistance in Gram-positive bacteria.
85 Annals of microbiology 64:413-420. https://doi.org/10.1007/s13213-013-0679-9.
86 6. Andrews S. 2010. FastQC: A quality control tool for high throughput sequence data. URL
87 https://wwwbioinformaticsbabrahamacuk/projects/fastqc/.
88 7. Krueger F. 2019. A wrapper tool around Cutadapt and FastQC to consistently apply quality and
89 adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type
90 (Reduced Representation Bisufite-Seq) libraries. URL
91 https://wwwbioinformaticsbabrahamacuk/projects/trim_galore/.
92 8. Bushnell B. 2017. BBNorm: Normalize coverage by down-sampling reads over high-depth areas
93 of a genome, to result in a flat coverage distribution. Department of Energy, Joint Genome
94 Institute, CA, USA. URL https://jgidoegov/data-and-tools/software-tools/bbtools/bb-tools-user-
95 guide/bbnorm-guide/.
96 9. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies
97 from short and long sequencing reads. PLoS computational biology 13:e1005595.
98 https://doi.org/10.1371/journal.pcbi.1005595.

99 10. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman
100 J, Young SK. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and
101 genome assembly improvement. PloS one 9:e112963.
102 https://doi.org/10.1371/journal.pone.0112963.
103 11. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome
104 assemblies. Bioinformatics 29:1072-1075. https://doi.org/10.1093/bioinformatics/btt086.
105 12. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T,
106 McCarthy SA, Davies RM. 2021. Twelve years of SAMtools and BCFtools. Gigascience 10:giab008.
107 https://doi.org/10.1093/gigascience/giab008.
108 13. Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, Davis JJ, Dempsey DM, Dickerman A, Dietrich
109 EM, Kenyon RW. 2023. Introducing the bacterial and viral bioinformatics resource center (BV-
110 BRC): a resource combining PATRIC, IRD and ViPR. Nucleic acids research 51:D678-D689.
111 https://doi.org/10.1093/nar/gkac1003.
112 14. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt
113 KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic acids
114 research 44:6614-6624. https://doi.org/10.1093/nar/gkw569.
115 15. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard
116 JL. 2017. Improvements to PATRIC, the all-bacterial bioinformatics database and analysis
117 resource center. Nucleic acids research 45:D535-D542. https://doi.org/10.1093/nar/gkw1017.
118 16. Saier Jr MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, Iddamsetty V, Lam KJK, Tian
119 N, Russum S, Wang J. 2021. The transporter classification database (TCDB): 2021 update. Nucleic
120 acids research 49:D461-D467. https://doi.org/10.1093/nar/gkaa1004.

121 17. Knox C, Wilson M, Klinger CM, Franklin M, Oler E, Wilson A, Pon A, Cox J, Chin NE, Strawbridge
122 SA. 2024. DrugBank 6.0: the DrugBank knowledgebase for 2024. Nucleic acids research
123 52:D1265-D1275. https://doi.org/10.1093/nar/gkad976.
124 18. Maiden MC, Van Rensburg MJJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND. 2013. MLST
125 revisited: the gene-by-gene approach to bacterial genomics. Nature Reviews Microbiology
126 11:728-736. https://doi.org/10.1038/nrmicro3093.
127 19. Carattoli A, Hasman H. 2020. PlasmidFinder and in silico pMLST: identification and typing of
128 plasmid replicons in whole-genome sequencing (WGS). Horizontal gene transfer: methods and
129 protocols doi:https://doi.org/10.1007/978-1-4939-9877-7_20:285-294.
130 https://doi.org/10.1007/978-1-4939-9877-7_20.
131 20. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, Rocha EP, Vergnaud
132 G, Gautheret D, Pourcel C. 2018. CRISPRCasFinder, an update of CRISRFinder, includes a portable
133 version, enhanced performance and integrates search for Cas proteins. Nucleic acids research
134 46:W246-W251. https://doi.org/10.1093/nar/gky425.
135 21. Gonçalves OS, de Assis JCS, Santana MF. 2022. Breaking the ICE: an easy workflow for identifying
136 and analyzing integrative and conjugative elements in bacterial genomes. Functional &
137 Integrative Genomics 22:1139-1145. https://doi.org/10.1007/s10142-022-00903-2.
138 22. Blin K, Shaw S, Medema MH, Weber T. 2024. The antiSMASH database version 4: additional
139 genomes and BGCs, new sequence-based searches and more. Nucleic acids research 52:D586-
140 D589. https://doi.org/10.1093/nar/gkad984.
141 23. van Heel AJ, de Jong A, Song C, Viel JH, Kok J, Kuipers OP. 2018. BAGEL4: a user-friendly web
142 server to thoroughly mine RiPPs and bacteriocins. Nucleic acids research 46:W278-W281.
143 https://doi.org/10.1093/nar/gky383.

144

Table 1: Genomic features of the sequenced four LAB strains isolated from fermented chicken meat in Chattogram, Bangladesh

Strain Name / Information Title	CVASU1	CVASU3	CVASU2	CVASU4
LAB Species	*Pediococcus pentosaceus*	*Pediococcus pentosaceus*	*Lactiplantibacillus plantarum*	*Lactiplantibacillus argentoratensis*
Genome Submission ID	SUB14898186	SUB14898186	SUB14898218	SUB14898234
Bioproject ID	PRJNA1189463	PRJNA1189463	PRJNA1189463	PRJNA1189463
BioSample accession number	SAMN44971012	SAMN44971014	SAMN44971013	SAMN44971015
SRA accession number	SRR31442436	SRR31442434	SRR31442435	SRR31442433
GenBank accession number	JBJPFK000000000	JBJPFJ000000000	JBJPFL000000000	JBJPFM000000000
FastQC Phred score	36	35	37	35
Raw Reads	1452975	1925165	1600310	1662521
Coverage	113.9×	145.6×	65.4×	70.9×
Contig count	24	26	62	99
Coarse consistency (%)	99.1	99.1	97.9	97.9
Fine consistency (%)	98.2	98.2	96	96.3
Completeness (%)	100	100	100	100

Contamination (%)	1.5	1.5	0	0.6
Genome size (bp)	1775093	1775155	3346512	3162126
Contigs N50 (bp)	190398	174289	197047	118879
Contigs L50	4	4	6	9
Largest contig (bp)	297937	297937	471909	305211
GC Content (%)	37.153152	37.153095	44.325077	45.074894
Genes (Total), PGAP	1800	1,799	3247	3008
CDS (PGAP; PATRIC)	1749; 1783	1748; 1783	3175; 3353	2947; 3108
tRNA	45	45	66	56
rRNA	3	3	3	2
CDS Ratio	1.0044544	1.0044193	1.0019387	0.98288304
Hypothetical CDS	350	350	1507	1333
Hypothetical CDS Ratio	0.29220414	0.29220414	0.50611395	0.49131274
PLFAM CDS	1700	1701	0	0
PLFAM CDS Ratio	0.95344925	0.95401007	-	-
Hypothetical proteins	351	351	1508	1334
Proteins with functional assignments	1432	1432	1845	1774

Proteins with EC number assignments	491	491	687	669
Proteins with GO assignments	407	407	584	561
Proteins with Pathway assignments	336	336	488	470
Proteins with Subsystem assignments	570	570	716	716
Proteins with PATRIC genus-specific family (PLfam) assignments	1700	1701	0	0
Proteins with PATRIC cross-genus family (PGfam) assignments	1727	1728	3233	2990
Proteins with FIGfam assignments	0	0	0	0
Transporter (TCDB)	2	2	15	8
Drug Target (DrugBank)	1	1	1	0
Antibiotic Resistance Genes (PATRIC)	21	21	25	27
Integrative and Conjugative Elements (ICEs)	2 (819699..838639, 18941 bp; 1645942..1683899, 37958 bp)	3 (819699..838639, 18941 bp; 1174724..1193031, 18308 bp; 1662631..1699235, 36605 bp)	2 (2983152..3034384, 51233 bp; 3313868..3325625, 11758 bp)	1 (3096716..3139121, 42406 bp)
No. of Clustered Interspaced Short Repeats (CRISPR) (number of Cas clusters) (name of cas genes)	6 (4) (cas9_TypeII, cas1_TypeII, cas2_TypeI-II-III, csn2_TypeIIA)	6 (4) (cas9_TypeII, cas1_TypeII, cas2_TypeI-II-III, csn2_TypeIIA)	2 (1) (cas3_TypeI)	22 (8) (cas6_TypeIE, cas5_TypeIE, cas7_TypeIE, cse2_TypeIE, cse1_TypeIE,

				cas3_TypeI, cas2_TypeIE, cas1_TypeIE)
Bacteriocins and biosynthetic gene clusters (BGCs)	Pediocin, terpene-precursor, T3PKS	Pediocin, T3PKS	Plantaricin_J, cyclic-lactone-autoinducer, RiPP like, terpene, T3PKS	terpene-precursor (3 loci), T3PKS, terpene