

論文の内容の要旨

論文題目 **Panoramic View on Genome Diversity and Evolution
of Lactic Acid Bacteria**
(乳酸菌ゲノムの多様性と進化に関する俯瞰的解析)

氏 名 谷澤 靖洋

Lactic acid bacteria (LAB) have long been associated with human culture and industrially exploited in production and preservation of food and feed for centuries. They are isolated across the world from nutrient rich environments, such as dairy products, fermented foods, plants, and animal intestines. From a taxonomic point of view, they are distributed into over 30 genera from six families under the order *Lactobacillales* (Figure 1). Among them, the genus *Lactobacillus* is the largest and highly heterogeneous group comprising nearly 200 species and subspecies. Recent advance of genome sequencing technologies has realized access to enormous genomic data. Particularly in the field of microbiology, genome sequences for a variety of organisms, not limited for model organisms or human pathogens, have become available, which gave rise to new opportunities for investigating diverse species. As of April 2016, NCBI Assembly Database stores more than 700 genomes for the genus *Lactobacillus*, marking the largest number except for model microorganisms and pathogenic bacteria. In particular, they include genomic data of 179 *Lactobacillus* spp. covering over 90% of its known species. The ecological characteristics of LAB and its wealth of genomic data make this microorganism particularly attractive for revealing the diversity of microbial world and their evolutionary background.

This work contains three research projects. The first two address case analyses of LAB that exhibit atypical characteristics: *L. hokkaidonensis* and the genus *Fructobacillus*. The last one addresses the development of a genome archive and annotation pipeline specialized for LAB.

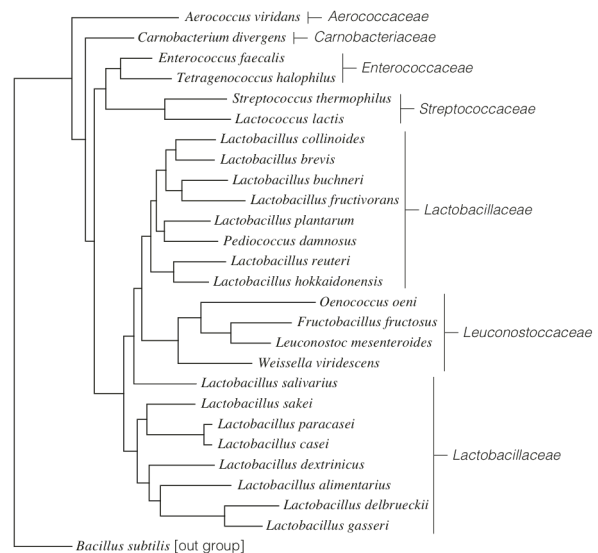


Figure 1. Phylogenetic tree of 26 representative LAB species from six families constructed from concatenated multiple alignments of 262 single-copy conserved genes.

Psychrotolerant LAB: *Lactobacillus hokkaidonensis*

Lactobacillus hokkaidonensis is an obligate heterofermentative LAB, which was isolated from Timothy grass

Compared to *Leuconostoc* spp., *Fructobacillus* species have significantly smaller numbers of protein coding sequences in their smaller genomes, especially lacking genes for carbohydrate transport and metabolism. Asymmetric distribution of conserved genes in each genus also shows that *Fructobacillus* spp. have lost more genes rather than have acquired new genes, indicating the streamlined genomes of *Fructobacillus* spp. (Figure 4). The lack of *adhE* genes in all *Fructobacillus* spp. exemplified the relevance of this gene in fructophilic characteristic, as postulated in previous studies. The general trend of reductive evolution, especially in metabolic simplification based on sugar availability, was revealed through this analysis.

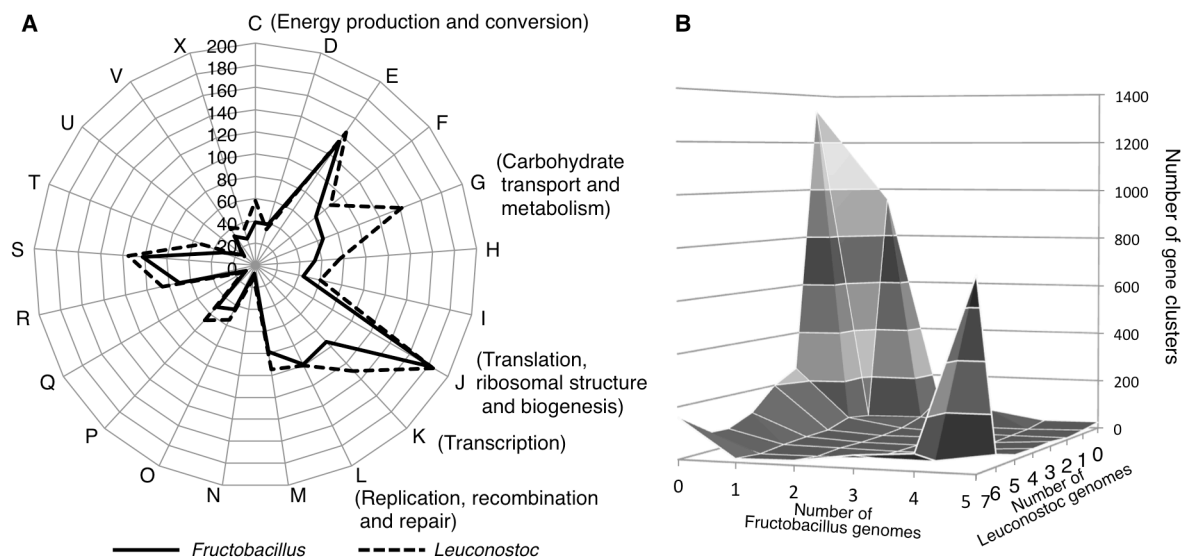


Figure 4. A) Comparison of gene content profiles for the genera *Fructobacillus* and *Leuconostoc*. B) Distribution of numbers of conserved gene clusters present in *Fructobacillus* and *Leuconostoc* [3].

LAB genome archive and annotation pipeline

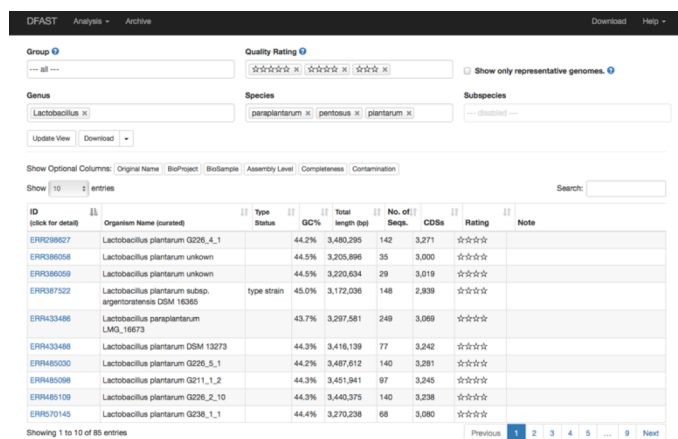
The number of LAB genomes available is drastically increasing, together with the spectrum of data quality and taxonomically mislabeled entries. They may lead to incorrect assumption and erroneous conclusions when dealt without careful consideration. In particular, some LAB species are difficult to distinguish only by the 16S rRNA gene-based identification, and a significant number of LAB genomes were deposited with incorrect taxonomic names in public databases [4]. To resolve these issues, I developed a curated genome repository DAGA (DFAST Archive of Genome Annotation) to provide reliable genome data resources for LAB [5].

DAGA currently provides 1,421 LAB genomes covering 191 (sub) species of two genera *Lactobacillus* and *Pediococcus* in the family *Lactobacillaceae* obtained from both DDBJ/ENA/GenBank and Sequence Read Archive (Figure 5). All genomes deposited in DAGA were re-annotated consistently using the identical pipeline. I used the average nucleotide identity (ANI), which showed high discriminative power to determine whether two genomes belong to the same species, to confirm the taxonomic affiliation of genomes. As a result, 155 mislabeled or unidentified genomes were assigned their correct taxonomic names and 38 genomes were marked

as ‘poor quality’. In particular, genomes for six type strains were disqualified due to possible misidentification or contamination. DAGA will improve both accessibility and reusability of genomic data for LAB.

To provide consistent annotation to genomes stored in DAGA, I developed an annotation pipeline called DFAST (DDBJ Fast Annotation and Submission Tool; <https://dfast.nig.ac.jp>), with curated reference protein databases tailored for LAB as well as quality and taxonomy assessment methods. DFAST was developed so that all the procedure required for data submission can be performed seamlessly on-line, and it can generate ‘ready-to-submit’ level annotation files to DDBJ without computational knowledge.

By taking advantage of the genomic data deposited in DAGA, I found previously unreported intraspecific diversity that exceeds the commonly accepted species-level differentiation within *Lactobacillus gasseri*, *L. jensenii*, and *L. vaginalis*. In addition, through the analysis of gene transfer among LAB strains, the niche-specific dissemination of genes related to anti-stress system was identified (Figure 6).



The screenshot shows the DFAST web interface. At the top, there are tabs for 'DFAST', 'Analysis', and 'Archive'. Below these are filters for 'Group' (set to 'all'), 'Quality Rating' (stars), and 'Show only representative genomes'. There are also filters for 'Genus' (set to 'Lactobacillus') and 'Species' (set to 'paraplantarum', 'pentosus', and 'plantarum'). A 'Download' button is visible. Below the filters is a table with columns: ID, Organism Name (curated), Type, Status, GC%, Total length (bp), No. of Seqs., CDSs, Rating, and Note. The table lists several *Lactobacillus* strains with their respective metrics and quality ratings (stars). A search bar is located on the right side of the table. At the bottom, it says 'Showing 1 to 10 of 85 entries' and has pagination controls.

ID	Organism Name (curated)	Type	Status	GC%	Total length (bp)	No. of Seqs.	CDSs	Rating	Note
ERR208627	Lactobacillus plantarum G226_4_1			44.2%	3,480,295	142	3,271	☆☆☆	
ERR208658	Lactobacillus plantarum unknown			44.5%	3,205,896	35	3,000	☆☆☆	
ERR208659	Lactobacillus plantarum unknown			44.5%	3,220,634	29	3,019	☆☆☆	
ERR2087322	Lactobacillus plantarum subsp. argenterotens DSM 16385	type strain		45.0%	3,172,036	148	2,839	☆☆☆	
ERR433486	Lactobacillus paraplantarum LMG_16673			43.7%	3,297,581	249	3,069	☆☆☆	
ERR433488	Lactobacillus plantarum DSM 13273			44.3%	3,416,139	77	3,242	☆☆☆	
ERR485030	Lactobacillus plantarum G226_5_1			44.2%	3,487,812	140	3,281	☆☆☆	
ERR485098	Lactobacillus plantarum G211_1_2			44.3%	3,451,941	97	3,245	☆☆☆	
ERR485109	Lactobacillus plantarum G226_2_10			44.3%	3,440,375	140	3,238	☆☆☆	
ERR570145	Lactobacillus plantarum G238_1_1			44.4%	3,270,238	68	3,080	☆☆☆	

Figure 5. Representative screenshot of DFAST and DAGA [5]. Users can query genomes of interest, browse their annotation, and download sequences.

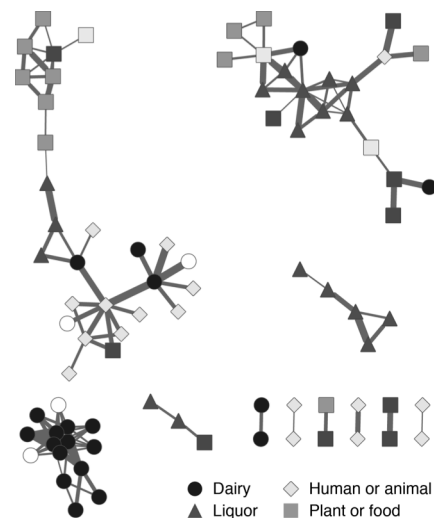


Figure 6. Gene transfer network among LAB strains. Each node represents a strain and its isolation source, and edges represent possible horizontal gene transfer.

Summary

Along this study, evolutionary trends associated with ecological habitats were presented. In addition, an integrated research platform that makes accurate and rapid genome analysis was established, which can be extended to other LAB genera as well as other microorganisms in the future plan.

References

- [1] Tanizawa Y, et al. *BMC Genomics*, 2015, 16:240.
- [2] Tanizawa Y, et al. *Jpn J Lactic Acid Bacteria*, 2016, 27(1):41-52.
- [3] Endo A, Tanizawa Y, et al. *BMC Genomics*, 2015, 16:1117.
- [4] Tanizawa Y, *Jpn J Lactic Acid Bacteria*, 2015, 26(3):206.
- [5] Tanizawa Y, et al. *Bioscience of Microbiota, Food and Health*, 2016, (accepted with minor revision).