

Comparative genome analysis of three *Lactobacillus plantarum* strains isolated from the human gastrointestinal tract and fermented foods

Natalia Garcia-Gonzalez¹, Francesca Bottacini², Douwe van Sinderen², Aldo Corsetti¹, Cormac Gahan^{2,3}

¹ Faculty of Bioscience and Technology for Food, Agriculture and Environment, University of Teramo, Teramo, Italy

² APC Microbiome Ireland and School of Microbiology, National University of Ireland, Cork, Ireland

³ School of Microbiology, University College Cork, Cork, Ireland

INTRODUCTION

Lactobacillus plantarum is a versatile species commonly found in a wide variety of ecological niches including dairy products and vegetables and it is also a natural inhabitant of the human gastrointestinal tract. The capacity of adapting to different environments is associated with the high plasticity and adaptability of this species, ultimately producing a broad range of different phenotypes. Although there is evidence that specific strains of *Lb. plantarum* are beneficial to the host, the mechanisms underlying these host-microbe interactions are still being determined. In this regard, the genome-scale *in silico* analysis of putative probiotic bacteria represents a central approach for the identification of potential biomarkers for probiotics selection, in order to identify antibiotic resistance and to predict potentially novel phenotypes.

The aim of this study was to employ comparative genome analysis in order to characterize the bacterial genome sequence of three *Lb. plantarum* strains isolated from different sources and predict their probiotic capabilities.

MATERIALS AND METHODS

The DNA of three *Lactobacillus plantarum* strains (Table 1) was isolated from pure cultures and the whole-genome sequencing was performed combining Illumina short reads with Oxford Nanopore long reads (Table 2). Raw sequence data were filtered and assembled using Unicycler. Genome annotation was performed via the RAST annotation server.

Table 1. *Lactobacillus plantarum* strains and origin

<i>Lb. plantarum</i> strains	Origin
IMC 513	Synbiotec s.r.l.
C904	Table olives
LT52	Raw-milk cheese

Table 2. Raw sequencing output after Nanopore and Illumina sequencing

Strain	Number of reads	# contig (>=0 bp)	# contig (>=1000 bp)	# contig	Largest contig	Total length	N50*
IMC513	598,282	47	23	25	549,450	3,303,141	369,769
C904	481,788	53	34	37	454,255	3,352,718	211,991
LT52	443,179	56	25	28	812,920	3,282,907	411,061

*N50 is defined as the length contigs (N) for which 50% of all bases in the assembly are in a sequence of length L < N.

GENERAL FEATURES OF *Lb. plantarum* STRAINS

The general features of the three *Lb. plantarum* strains are summarized below (Table 3 and Figure 1).

Table 3. Comparison of chromosomal properties of the three *Lb. plantarum* strains

Strain	IMC 513	C904	LT52
Genome size	3,206,211	3,275,253	2,799,038
GC content (%)	45.72	45.57	45.63
Coding Sequence (CDS)	3,022	3,069	2,579
tRNA	71	65	55
rRNA	11	11	7

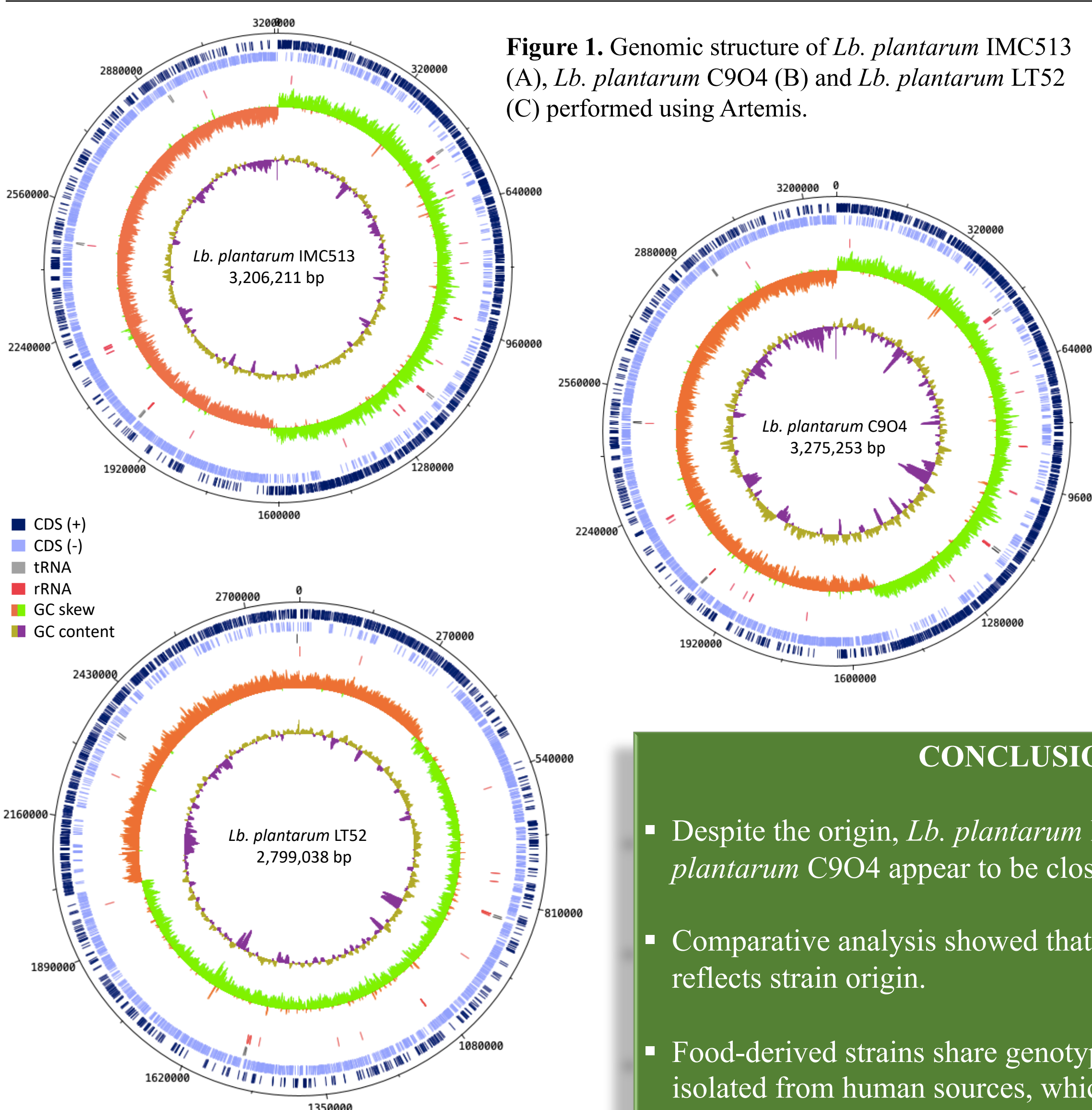


Figure 1. Genomic structure of *Lb. plantarum* IMC513 (A), *Lb. plantarum* C904 (B) and *Lb. plantarum* LT52 (C) performed using Artemis.

COMPARATIVE ANALYSIS WITHIN *Lb. plantarum* SPECIES

The draft genome of the three novel *Lb. plantarum* strains was analyzed together with 39 *Lb. plantarum* genomes obtained from NCBI database and aligned to the *Lb. plantarum* WCFS1 reference genome. A comparative pan-genome analysis has been performed in order to compare the core and the variable genome of all *Lb. plantarum* strains with our three newly sequenced strains. As result from the analysis the degree of variability (number of shared genes) among the three novel strains was low, especially between *Lb. plantarum* IMC513 and *Lb. plantarum* C904, which appeared to be closely related. Moreover, the analysis also showed that no strain grouping correlating with the origin of isolation (Figure 2).

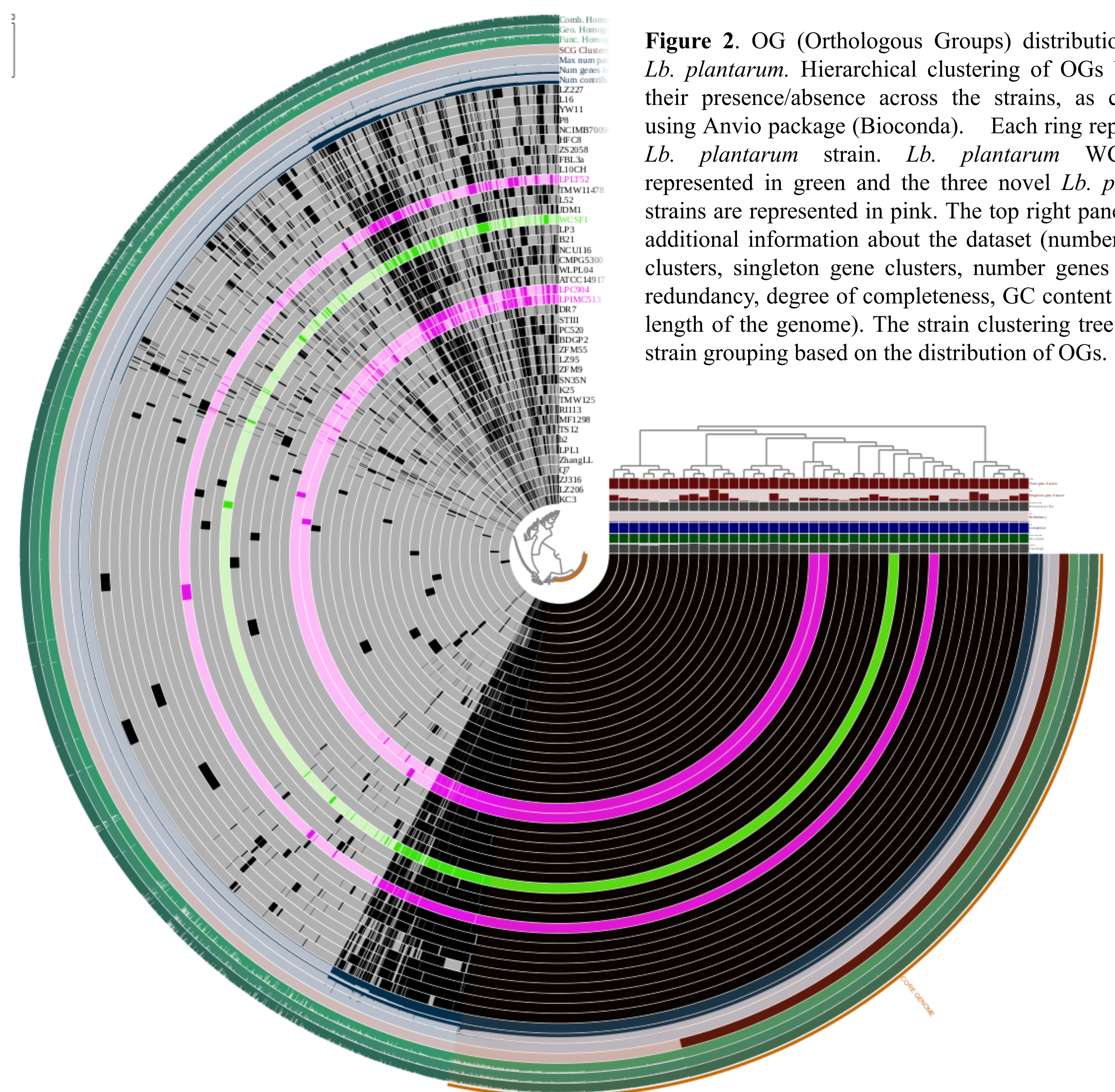


Figure 2. OG (Orthologous Groups) distribution across *Lb. plantarum*. Hierarchical clustering of OGs based on their presence/absence across the strains, as computed using Anvio package (Bioconda). Each ring represents a *Lb. plantarum* strain. *Lb. plantarum* WCFS1 is represented in green and the three novel *Lb. plantarum* strains are represented in pink. The top right panel reports additional information about the dataset (number of gene clusters, singleton gene clusters, number genes per kbp, redundancy, degree of completeness, GC content and total length of the genome). The strain clustering tree refers to strain grouping based on the distribution of OGs.

GENES POTENTIALLY INVOLVED IN PROBIOTIC EFFECTS

Comparative analysis showed that around 53% of *Lb. plantarum* WCFS1 genes are shared by the three *Lb. plantarum* strains investigated. Among these genes, we focused our research on those conferring putative possible probiotic properties. A probiotic bacteria should have the ability to survive, and transiently persist in the gastrointestinal tract, where has to be able to exert a beneficial effect. Ability to resist host stressful conditions or hydrolyse bile salts are key genes to consider for potential candidate probiotic bacteria.

Genes involved in	WCFS1	IMC513	C904	LT52
Bile resistance	lp_3536			
	lp_0067			
	lp_3362			
	lp_2572			
	lp_3324			
Osmotic stress	lp_0031			
	lp_0727			
	lp_0728			
	lp_1129			
	lp_1128			
Temperature stress	lp_1097			
	lp_1096			
	lp_0927			
	lp_3122			
	lp_0331			
Oxidative stress	lp_1385			
	lp_2370			
	lp_2366			
	lp_2365			
	lp_2369			
pH stress	cns2			
	cas2			
	cas9			
	cas2			
	cas1			
CRISPR-Cas system				

Figure 3. Genes potentially involved in probiotic effects. Presence and absence of genes in each genome are indicated in pink and grey, respectively

Gene present
Gene absent

CONCLUSIONS

- Despite the origin, *Lb. plantarum* IMC513 and *Lb. plantarum* C904 appear to be closely related.
- Comparative analysis showed that gene distribution no reflects strain origin.
- Food-derived strains share genotypic properties to those isolated from human sources, which make them suitable as potential probiotic candidates.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Synbiotec (Camerino, Italy) to kindly provide us *Lb. plantarum* IMC513.