https://doi.org/10.1051/bioconf/20235901009

Phylogeny and codon usage bias of bacterial genomes in *Bifidobacterium animalis*

Yongzong Yang[a]

Department of applied biology and biotechnology, The Hong Kong polytechnic university, 999077, Hong Kong, China

Abstract—In nature, the phenomenon of an equal probability distribution of four nucleotides did not exist. Due to the influence of selection, the nucleotides of *Bifidobacterium animalis* would not be presented with equal probability. *Bifidobacterium animalis* was frequently added to food because of its special metabolic pathway, which could catalyze fructose and lactic acid. This study analyzed industrial *Bifidobacterium animalis* and environmental *Bifidobacterium animalis* through strategies such as a phylogenetic tree, ENC, RSCU, PR2, neutral graph, and ENC top/bottom gene enrichment graph. The result was that the *Bifidobacterium animalis* as a whole is greatly affected by the environment, while the difference between the internal industrial bacteria and environmental bacteria was not apparent. This study could provide a reference for the screening of industrial strains from *Bifidobacterium* and the further development of *Bifidobacterium*.

1. Introduction

Since genetic code was decoded in the 1960s, the biological community has lasted researching codon to try to figure out the mystery of genetic code. In the wake of DNA sequencing being introduced in the late 1970s [2], it was revealed that codon usage existed in different organisms utilized with different frequencies. Whereafter, depending on the study of codon usage, some deductions were concluded [3]: In similar genomes, the aforementioned usage frequency emerged with striking consistency, but in diverse genomes, that usage frequency showed a huge difference. This phenomenon, called codon usage bias (CUB) [4], refers to the priority usage of synonymous codons or nonrandom usage. This result could be caused by mutation, gene drift, and gene recombination [5]. Various factors could also generate influence in CUB, like GC content, codon location, a function of translation product, folding of mRNA, an abundance of tRNA, and genome composition. CUB could reveal the evolution of species, the genetic relationship, and horizontal gene transfer between organisms.

Until now, several hypotheses have been developed to decipher the CUB [6,7,8]: genome hypothesis, mutation hypothesis, and selection-mutation-drift model. In those models, the selection-mutation-drift model occupied the most attention and this model proposed that [7, 8, 9] three evolution factors (mutation factor, selection factor, and genetic drift factor) coordinately affect CUB. Therefore, this study stood on this bedrock model to explore desired species codon usage patterns.

As a familiar probiotic, *Bifidobacterium animalis*, a high G+C content positive bacteria, was widely found in mammals, birds, and several ectotherms' gut-intestinal tract [10]. Since it employed a particular pathway [11,12], a number of products like foods and health care products were added *to Bifidobacterium animalis*. Besides, other functions, such as anti-tumor, anti-inflammation, and reducing blood fat, also supported that *Bifidobacterium animalis* was a significant probiotic in intestinal flora [13]. Moreover, some particular strains, such as BB-12 [11], which was employed in dairy production, may have more special codon usage patterns than other strains. However, information on the genetics of *Bifidobacterium animalis* was limited [9]. Although some of the complete genomes of *Bifidobacterium animalis* have been sequenced, the genetics of the reactions in these genomes remained vague.

As aforementioned, CUB was an effective tool to facilitate researching organisms' mutation, selection, and genetic drift. Meanwhile, the study about *Bifidobacterium animalis* in CUB field was infrequent. Hence, this study utilized relevant technologies of bioinformatics and genetics to analyze the genomic sequences of different strains of *Bifidobacterium animalis* from industry or environment and tried to obtain several results, GC, CUB, and potential evolution patterns under the selection-mutation-drift model. This result may promote the further optimization of the use of *Bifidobacterium animalis*.

[a] Email: 20079401d@connect.polyu.hk

2. Method and Materials

2.1 Bifidobacterium animalis sequences extraction

Twenty-four different strains of *Bifidobacterium animalis* are selected for analysis (see Table 1). Those *Bifidobacterium animalis* could be divided into industry, and environment strains and the below provided illustration. Their sequences were acquired from the NCBI database using the accession number of required sequences (CP001853, CP001892, NZ_CP031154, NC_012814, NC_011835, CP002915, NZ_CP009045, NC_022523, NC_012815, CP085838, CP047190, NZ_CP028460, NZ_CP035497, NZ_CP042940, NZ_CP017098, NZ_CP094969, NZ_CP007755, NZ_CP045589, NZ_CP084315, CP069248, CP069249, CP080571, NZ_CP031703, NZ_CP015407) from nucleotide database of NCBI.

Table 1. Extracted 24 Bifidobacterium animalis sequences and their accession number, purpose and specific usage if they are industry purpose strains.

Table S1 basic information of *Bifidobacterium animalis*

Species name	Accession number	Purpose	Patents or Products
***Bifidobacterium animalis subsp*. lactis BB-12**	CP001853	Industry	BRPI0802285A2
***Bifidobacterium animalis subsp*. lactis V9**	CP001892	Industry	CN111493261A
***Bifidobacterium animalis subsp*. lactis strain HN019**	NZ_CP031154	Industry	DuPont™ Danisco® range
***Bifidobacterium animalis subsp*. lactis Bl-04**	NC_012814	Industry	Snow Brand Milk
***Bifidobacterium animalis subsp*. lactis AD011**	NC_011835	Industry	US9453232B2
***Bifidobacterium animalis subsp.* lactis CNCM I-2494**	CP002915	Industry	US20150328266A1
***Bifidobacterium animalis subsp.* lactis strain BF052**	NZ_CP009045	Industry	Minas cheese
***Bifidobacterium animalis subsp.* lactis ATCC 27673**	NC_022523	Industry	Minas cheese
***Bifidobacterium animalis subsp*. lactis DSM 10140**	NC_012815	Industry	Snow Brand Milk
***Bifidobacterium animalis subsp*. lactis strain DSM 15954**	CP085838	Industry	Snow Brand Milk
***Bifidobacterium animalis* strain Probio-M8**	CP047190	Industry	US20150320807A1
***Bifidobacterium animalis* subsp. animalis strain CNCM I-4602 chromosome**	NZ_CP028460	Environment	
***Bifidobacterium animalis* strain 01 chromosome**	NZ_CP035497	Environment	
***Bifidobacterium animalis* strain B06 chromosome**	NZ_CP042940	Environment	
***Bifidobacterium animalis* strain BL3 chromosome**	NZ_CP017098	Environment	
***Bifidobacterium animalis* strain HY8002 chromosome**	NZ_CP094969	Environment	
***Bifidobacterium animalis* strain RH chromosome**	NZ_CP007755	Environment	
***Bifidobacterium animalis* strain TK-J6A chromosome**	NZ_CP045589	Environment	
***Bifidobacterium animalis subsp*. lactis strain 19-D-1 chromosome**	NZ_CP084315	Environment	
***Bifidobacterium animalis subsp*. lactis strain H1 chromosome**	CP069248	Environment	
***Bifidobacterium animalis subsp.* lactis strain H3 chromosome**	CP069249	Environment	
***Bifidobacterium animalis subsp*. lactis strain i797 chromosome**	CP080571	Environment	
***Bifidobacterium animalis subsp*. lactis strain IDCC4301 chromosome**	NZ_CP031703	Environment	
***Bifidobacterium animalis subsp*. animalis strain YL2 chromosome**	NZ_CP015407	Environment	

2.2 Selected sequences alignment

Using DAMBE 7.0 [14] aligned the aforementioned 24 genomes. Subsequently, using MAFFT online version [15] aligned those sequences.

2.3 Codon usage indices

To analyze codon usage pattern, the effective number of codons (ENC), relative synonymous codon usage (RSCU), the G+C content in the third locus of synonymous codon (GC3), PR2 and Neutrality. Those indices were obtained through codon W 1.4.4 [6].

The effective number of codons, an important index of CUB, could reflect the usage frequency of codons [15]. A general view stated that the higher the ENC value, the possibility of CUB would be lower, and the lower the ENC value, the possibility of CUB would be higher [6,16,17]. GC3 was utilized as the abscissa, and the ENC value was used as the ordinate. The result was a scatter plot, and the reference value of ENC was also marked in the figure. The specific calculation method was as follows: ENC = 2 + GC3s + 29/ [GC3s 2 + (1 − GC3s) 2] [6, 24, 25]. If the ENC value was proximate or on the reference curve, it proved that mutation is the main pressure for species evolution. Conversely, if the ENC value was lower than the reference curve, it meant that the evolution of the species had received the pressure of environmental selection [23].

Relative synonymous codon usage could emerge the usage bias in synonymous codon [18]. This index assumed that each synonymous codon had a usage probability of 1. Therefore, a lower ratio than 1 indicates that the codon was below the average nature expression level of the synonymous codon, and a higher ratio indicates that the codon was above the average nature expression level of the synonymous codon [19]. One detail worth noting was that values above 1.4 and below 0.6 were considered high or low usage frequency, respectively [7].

Because most of the synonymous codons differed only in the third base, the detection of GC3 content was equally essential for the analysis of CUB [20]. Parity rule 2, also known as PR2, which was introduced by Chargaff [26], was developed by subsequent studies and was responsible for the exploration of the influence of mutations and selection on genome and CUB [26]. In this study, G3/ (G3 + C3) was used as the abscissa, and A3/ (A3 + T3) was used as the ordinate. Number 3 represented the composition of the third nucleotide on the codon. The coordinate center was (0.5, 0.5), so values diverging in other directions could be regarded as biases that occur [28].

Neutrality, one nonnegligible index for mutation analysis. Since most species had received different degrees of external pressure during evolution, their GC content would be selected accordingly [24]. Neutrality was an important parameter for analyzing external selection. In this study, GC1/2 (average value of GC1 and GC2) was used as the abscissa and GC3 was used as the ordinate to draw the neutrality graph. The result would be a scatter plot and the regression line would be labeled in the plot.

2.4 Phylogenetic analysis

All the genomes were annotated by Prokka [29]. The GFF3 files were obtained for further analysis. The phylogenetic tree was constructed by Roary [30]. The tree was drawn by Figtree [31].

2.5 Chart analysis

In order to present the data more intuitively, the phylogenetic tree, neutrality, ENC, and parity rule 2-bias (PR2) were graphed into figures. These figures were graphed with ggplot2 [21], which was under the R v4.2.

3. Results

3.1 24 different strains of Bifidobacterium animalis nucleotide compositions

A total of 34826 codons were extracted from the 24 types of selected strains. Among them, 17443 codons were derived from 12 strains of environment strains, while the remaining 17383 codons were derived from 12 types of industry strains.

After analyzing the 34826 codons, it was found that the base composition of *Bifidobacterium animalis* was higher than the theoretical result (equal GC and AT content). In those codons, the GC content in the completed gene sequence was as high as 61.22%, the median was 61%, and the variance was only 0.13%. This meant that the overall codon selected from *Bifidobacterium animalis* had a high fraction of GC content. For codons from environment and industry strains, their GC content was 61.20% and 61.23%, respectively, which demonstrated non-distinct variation between environment and industry strains.

What was more noteworthy was that the GC3 content reached 76.8%, the median was 77.8%, and the variance was only 0.6%. This revealed that the GC3 content of *Bifidobacterium animalis* had a high selection rate. However, this appeared to be the case only for selectivity preferences among synonymous codons. The average ENC value was determined to be 40.22, which indicated that the emerging GC bias was only exhibited in synonymous codons. Therefore, it could be concluded that there is a positive correlation between the GC content in synonymous codons and in whole codons in *Bifidobacterium animalis*. This conclusion was also raised in previous studies on other species. [32] As for the GC3 content in environment and industry strains, the value was 76.79% and 76.86%, respectively. Since either overall GC content or GC3 content in environment and industry strains had non-difference, in subsequent analysis, only the completed genome of overall strains would be analyzed.

Due to the complete gene sequence, the phylogenetic tree was also drawn by the NJ method, as Figure 1 seen. The phylogenetic tree provided potential evolution route. According to the phylogenetic tree, it was not untoward to find that the internal connection between industrial bacteria was relatively intimate. For example, BB-12 was the most common probiotic, and it is most closely related DSM 15954 was also a probiotic for industrial usage. This suggested that their evolutionary lines could have followed a consistent path. Because the industrialization time was not long, but the bifidobacteria had undergone a long evolution, distinct preferential evolution of industrial bacterium strains did not arise. But those with a close relationship were often for the same purpose. For example, the HY8002 strain and the H1 strain were both environmental bacterium strains.

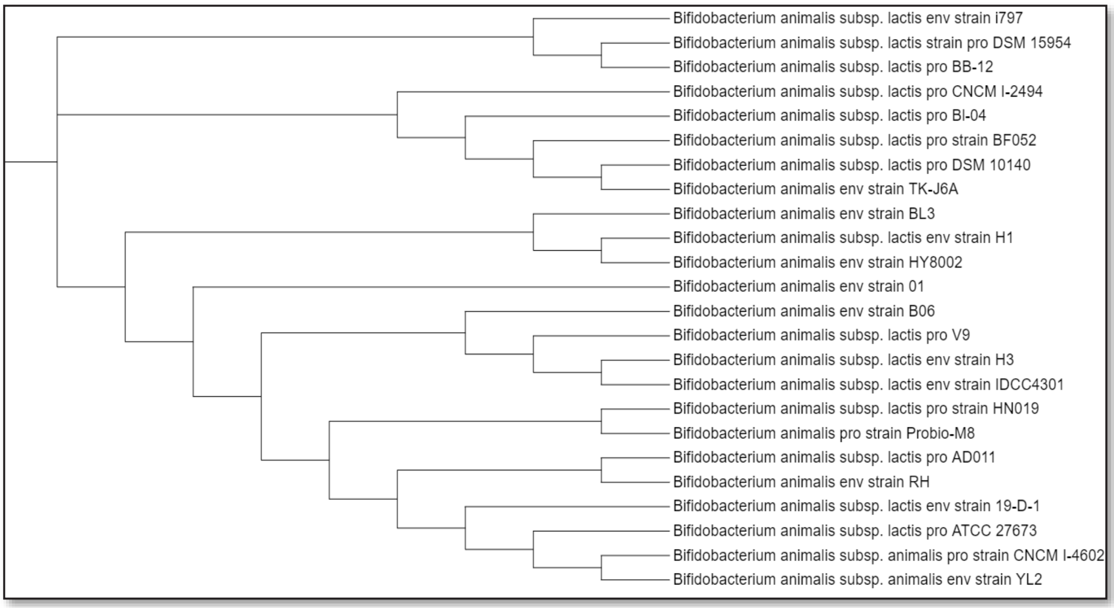
Bifidobacterium animalis subsp. lactis env strain i797
Bifidobacterium animalis subsp. lactis strain pro DSM 15954
Bifidobacterium animalis subsp. lactis pro BB-12
Bifidobacterium animalis subsp. lactis pro CNCM I-2494
Bifidobacterium animalis subsp. lactis pro BI-04
Bifidobacterium animalis subsp. lactis pro strain BF052
Bifidobacterium animalis subsp. lactis pro DSM 10140
Bifidobacterium animalis env strain TK-J6A
Bifidobacterium animalis env strain BL3
Bifidobacterium animalis subsp. lactis env strain H1
Bifidobacterium animalis env strain HY8002
Bifidobacterium animalis env strain 01
Bifidobacterium animalis env strain B06
Bifidobacterium animalis subsp. lactis pro V9
Bifidobacterium animalis subsp. lactis env strain H3
Bifidobacterium animalis subsp. lactis env strain IDCC4301
Bifidobacterium animalis subsp. lactis pro strain HN019
Bifidobacterium animalis pro strain Probio-M8
Bifidobacterium animalis subsp. lactis pro AD011
Bifidobacterium animalis env strain RH
Bifidobacterium animalis subsp. lactis env strain 19-D-1
Bifidobacterium animalis subsp. lactis pro ATCC 27673
Bifidobacterium animalis subsp. animalis pro strain CNCM I-4602
Bifidobacterium animalis subsp. animalis env strain YL2

Figure 1. Phylogenetic tree of 24 selected Bifidobacterium animalis. Values labeled in the figure corresponded to distance, and the closer distances and fewer interval branches represented closer evolutionary routes.

3.2 Neutrality graph

To explore the selection pressure and the mutation pressure influence on the genome, the neutrality figure was graphed with GC12 versus GC3, as Figure 2 shown. A positive correlation between GC3 and GC12 was observed depending on this figure, and the regression coefficient was measured as 0.001356. It could be observed from Fig. 1 that the distribution of points was not relevantly concentrated (If compared with the regression line), which could be inferred that the factors from natural selection accounted for most of the proportion. The regression coefficient represents that the proportion of mutation factors was only 1.356%, and the remaining 98.644% was contributed by natural selection factors.

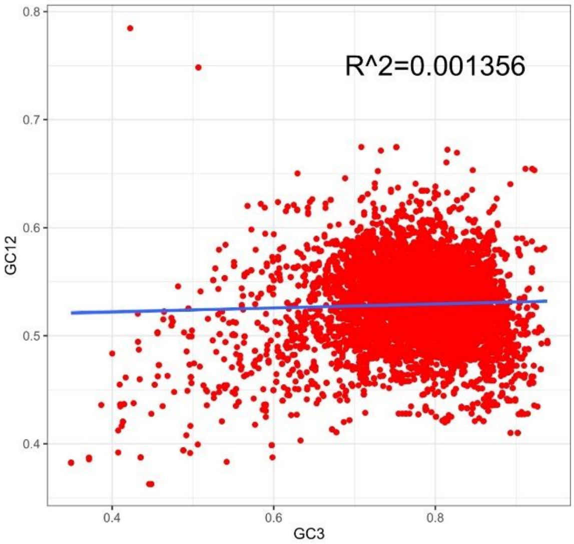

Figure 2. Neutrality graph. GC1/2 represented abscissa, GC3 represented ordinate, the red dot was the single gene ratio, the regression line was constructed on the basis of the full ratio, and R^2 represented the regression coefficient.

3.3 PR2 graph

In the subsequent analysis after the neutrality graph, the third base pair of the codon was further analyzed specifically, and a PR2 graph, as Fig. 3 seen, was created based on a single base. It could be concluded from the graph that the overall data was scattered, and there were even a few extreme cases close to thelimit value. Therefore, the influence brought by the force of natural selection was larger.

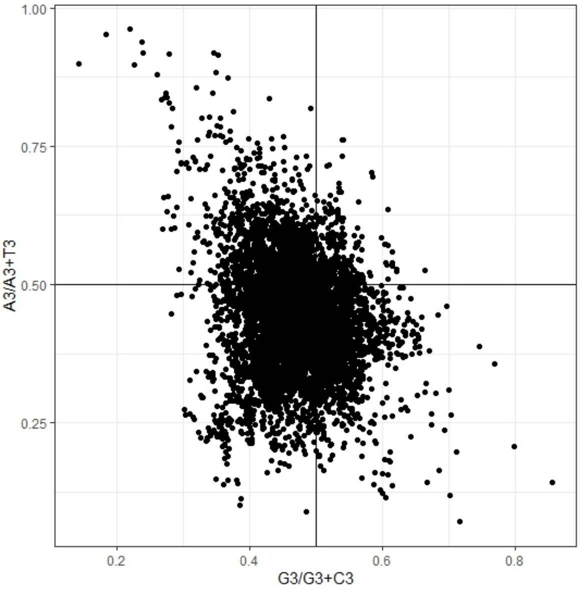

Figure 3. Parity rule2 graph. Draw with A3/(A3+T3) as the abscissa, G3/(G3+C3) as the ordinate, and (0.5,0.5) as the coordinate center.

3.4 ENC graph

In addition, a specific analysis of ENC was also carried out. In addition to the above-mentioned average ENC value of 40.22, in order to provide more accurate feedback on the relationship between ENC and GC3, an ENC graph with the coordinates of these two values was created as Fig. 4 shown. It could be observed that most of the data are in the lower right of the prediction curve (Expectation curves assume only mutation effects), which not only showed the bias of GC3 usage but also showed that natural selection factors have a great influence on codons since when only mutation factor had the effect, the value would be located on the curve.

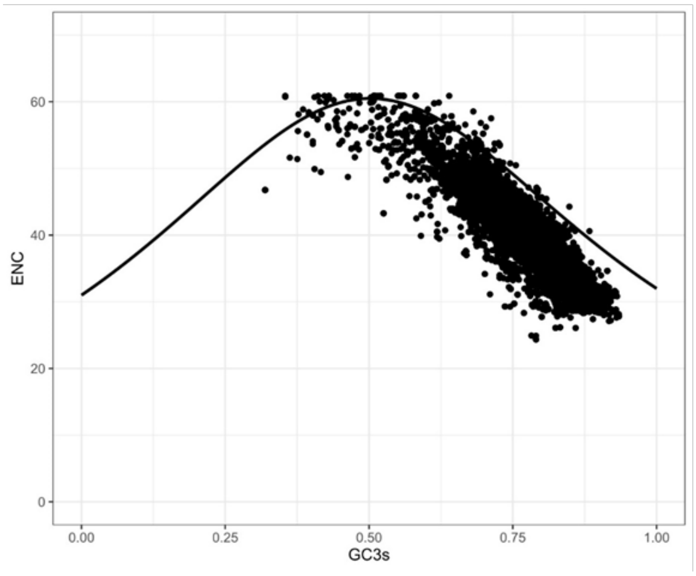

Figure 4. ENC-GC3 graph. The curve was the prediction value which assumed that only mutation factor had influence.

3.5 RSCU value

Moreover, after analyzing the ENC value, simple statistics about RSCU value were also performed on the usage of synonymous codons, as shown in Tab. 2. By observing the table, it could be found that except for the codon encoding Met (Because there was only one codon transcription), the codons for most of the remaining amino acids generated usage bias (the occurrence value exceeding 1.4 was regarded as bias). This phenomenon was uniform with the previously observed GC content bias and ENC value bias.

Among selected strains, most of the proteins had RSCUs with large gaps, and only four amino acids, Tyr, His, Asp, and Glu, had gaps lower than 0.6. This could be induced by larger influences from the environment rather than mutations. It was deservedly mentioned that Arg had a bias greater than 3, and four codons could encode this amino acid. Arginine is a positively charged essential amino acid that may be closely linked to metabolic pathways at the core of probiotics. Synonymous codon bias could also occur due to factors such as heat resistance.

Table 2. RSCU value of overall amino acids.

Table 1 Overall codon usage data of *B. animalis*

AA	Codon	RSCU	AA	Codon	RSCU
Phe	TTT	0.11	Tyr	TAT	0.74
	TTC	1.86		TAC	1.26
Leu	TTA	0.04	Stop	TAA	0.00
	TTG	0.75	Stop	TAG	0.00
	CTT	0.48	His	CAT	0.96
	CTC	2.50		CAC	0.99
	CTA	0.08	Gln	CAA	0.33
	CTG	2.16		CAG	1.64
Lie	ATT	0.66	Asn	AAT	0.49
	ATC	2.15		AAC	1.46
	ATA	0.20	Lys	AAA	0.52
Met	ATG	1.00		AAG	1.48
Val	GTT	0.24	Asp	GAT	0.71
	GTC	1.26		GAC	1.29
	GTA	0.19	Glu	GAA	0.72
	GTG	2.31		GAG	1.28
Ser	TCT	0.26	Cys	TGT	0.22
	TCC	1.65		TGC	1.53
	TCA	0.43	Stop	TGA	0.00
	TCG	1.89	Trp	TGG	0.94
Pro	CCT	0.27	Arg	CGT	1.40
	CCC	0.81		CGC	3.14
	CCA	0.54		CGA	0.37
	CCG	2.38		CGG	0.61
The	ACT	0.24	Ser	AGT	0.33
	ACC	1.96		AGC	1.44
	ACA	0.36	Arg	AGA	0.16
	ACG	1.45		AGG	0.32
ala	GCT	0.18	Gly	GGT	0.63
	GCC	1.80		GGC	2.42
	GCA	0.71		GGA	0.45
	GCG	1.31		GGG	0.50

3.6 ENC gene enrichment

The difference between the top and the bottom 10% gene in the enrichment graph was apparent. Both of bottom and top had several gene sequences related to metabolism, the largest gene at the bottom was related to the ribosome, and the gene at the top was related to ABC transporters and homologous recombination.

4. Discussion

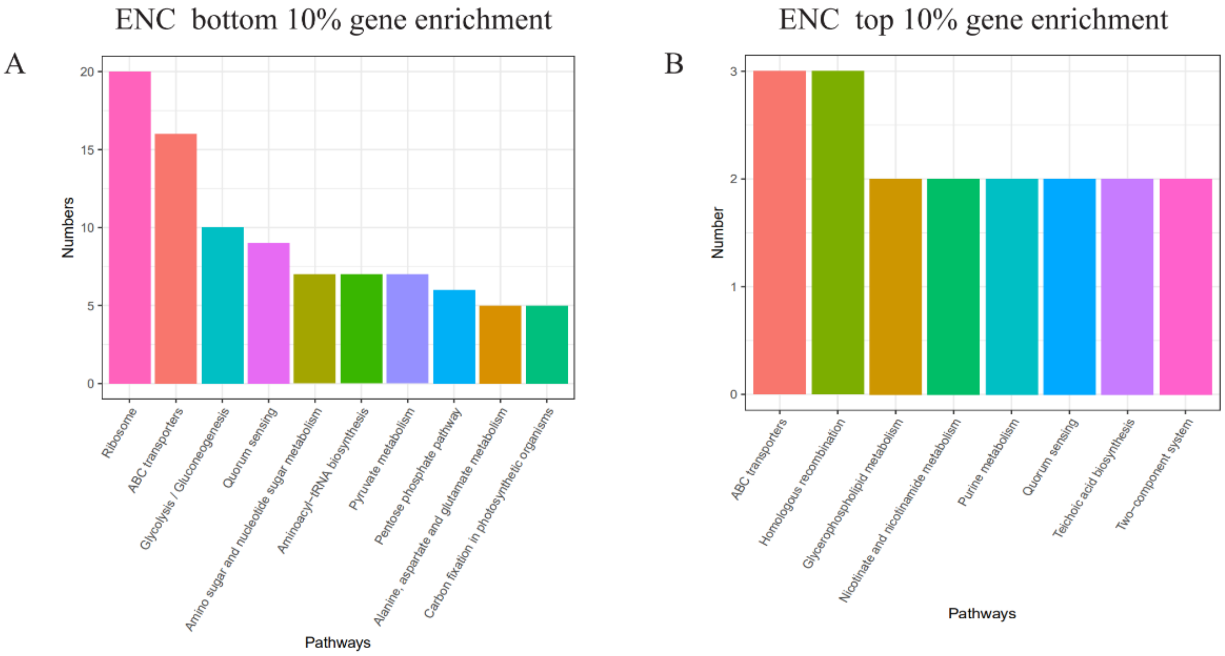

Figure 5. ENC relative gene enrichment analysis.

With the development of science and technology, some species were used for industrial purposes due to their unique properties; most of the properties were particular pathways. It was of reference significance to explore the variations between species for industrial usage purposes and other species (environment species) for the further development of industry and the selection of more excellent species. *Bifidobacterium animalis*, as well-known probiotics, were frequently utilized in the industry because of their unique metabolic pathways. Therefore, this study selected twelve common industrial *Bifidobacterium animalis* and twelve environmental *Bifidobacterium animalis* which were relative to each other for analysis, expecting to provide a reference for the screening of strains.

In general, the results shown above showed not much difference between industrial and environmental bacteria at the genetic level. The following factors may cause the reason for the insignificant difference. First, since the industrialization process of Bifidobacterium was still relatively short and the selection factor was not enough to play a big role, the differences within Bifidobacterium were not obvious. Secondly, the selection of samples could be biased. Because not many strains had been sequenced, the strains selected may not truly represent genomic differences between industrial and environmental bacteria. Finally, the difference between industrial strains and environmental strains may be in the genes that were not expressed, such as operon, promoter, and so on. These factors may lead to the presence of no significant differences appear.

The aforementioned CUB was constituted of several indexes. ENC as an important indicator was critical to weighing CUB, and the 24 selected strains in this study exhibited ENCs exceeding 35. Moreover, ENCs exceeding 35 were generally considered to have no codon bias in the screened species case. The advantage of ENC was that it could reflect the overall codon usage preference, but it could not especially reflect the usage preference among synonymous codons. Therefore, RSCUs could complement this concept. By processing the data of Bifidobacterium RSCUs, several codons with serious preferences could be sorted out, and it was also of reference significance to analyze the metabolic gene locations of these codons. The neutral chart, PR2, and other charts could be more specific to reflect which occupies the greater influence of selection pressure and abrupt pressure. It was not hard to see from the results that selection pressure played a bigger role. Codon usage biases can be seen in both eukaryotic as well as prokaryotic genomes, and highly expressed genes have a tendency to use preferred codons more frequently than other gene types. Prior to this discovery, researchers believed that the impacts of codon use on the expression of genes were mostly mediated by the effects it had on translation. In this regard, Zhou et al. (2016) discovered an unexpectedly important function of codon usage in ORF sequences in influencing transcription levels. Furthermore, they argue that codon biases represent an adaptation of protein nucleotide regions to the transcriptional and translational machinery. Thus, the employment of codons not only helps determine protein sequences and the dynamics of translation but also helps determine the amounts of gene expression.

By analyzing the codon bias, several results from the study could reflect that the evolutionary pressure of *Bifidobacterium animalis* came from the selection. For example, prominent deviation in the composition of nucleotides, the value of RSCU having no preference for codons corresponding to only four amino acids, the values of the neutral graph not on the standard line, the valuesof the PR2 graph not in the coordinate neutral and ENC values all below the reference line. All these indicated that *Bifidobacterium animalis* had undergone more selection so far in evolution.

5. Conclusion

In summary, this study analyzed 24 genomes of *Bifidobacterium animalis* utilizing bioinformatics in order to find out the impacts on their evolution under different conditions. This study obtained nucleotide composition, phylogenetic tree, ENC, neutral graph, RSCUs, and ENC enrichment data. Although the difference between industrial and environmental evolutionary pressures was not yet available from these data, it could be predicted that they collectively evolved under greater environmental pressures. The reason why internal distinctions were not available may be that the industrialization of *Bifidobacterium animalis* was still relatively recent. These results could provide a reference for future screening of better industrial bacteria.

Reference

[1] Sharp PM, Emery LR and Zeng, K. (2010). Forces that influence the evolution of codon bias. Philosophical transactions of the Royal Society of London. *Series B, Biological sciences, 365*(1544), 1203–1212. https://doi.org/10.1098/rstb.2009.0305

[2] Grantham, R., Gautier, C., Gouy, M., Mercier, R., & Pavé, A. (1980). Codon catalog usage and the genome hypothesis. *Nucleic acids research, 8*(1), r49–r62. https://doi.org/10.1093/nar/8.1.197-c

[3] Parvathy, S. T., Udayasuriyan, V., & Bhadana, V. (2022). Codon usage bias. *Molecular biology reports, 49*(1), 539–565. https://doi.org/10.1007/s11033-021-06749-4

[4] Iriarte, A., Lamolle, G., & Musto, H. (2021). Codon Usage Bias: An Endless Tale. *Journal of molecular evolution, 89*(9-10), 589–593. https://doi.org/10.1007/s00239-021-10027-z

[5] Yang, J., Ding, H., & Kan, X. (2021). Codon usage patterns and evolution of HSP60 in birds. *International journal of biological macromolecules, 183*, 1002–1012. https://doi.org/10.1016/j.ijbiomac.2021.05.017

[6] Bulmer M. (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics, 129*(3), 897–907. https://doi.org/10.1093/genetics/129.3.897

[7] Frank, A. C., & Lobry, J. R. (1999). Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. *Gene, 238*(1), 65–77. https://doi.org/10.1016/s0378-1119(99)00297-8

[8] Carbone, A., Zinovyev, A., & Képès, F. (2003). Codon adaptation index as a measure of dominating codon bias. *Bioinformatics (Oxford, England), 19*(16), 2005–2015. https://doi.org/10.1093/bioinformatics/btg272

[9] Turroni, F., van Sinderen, D., & Ventura, M. (2011). Genomics and ecological overview of the genus Bifidobacterium. *International journal of food microbiology, 149*(1), 37–44. https://doi.org/10.1016/j.ijfoodmicro.2010.12.010

[10] Dantas, A., Verruck, S., Canella, M. H. M., Hernandez, E., & Prudencio, E. S. (2021). Encapsulated Bifidobacterium BB-12 addition in a concentrated lactose-free yogurt: Its survival during storage and effects on the product's properties. *Food research international (Ottawa, Ont.), 150*(Pt A), 110742. https://doi.org/10.1016/j.foodres.2021.110742

[11] Engevik, M. A., Danhof, H. A., Hall, A., Engevik, K. A., Horvath, T. D., Haidacher, S. J., Hoch, K. M., Endres, B. T., Bajaj, M., Garey, K. W., Britton, R. A., Spinler, J. K., Haag, A. M., & Versalovic, J. (2021). The metabolic profile of Bifidobacterium dentium reflects its status as a human gut commensal. *BMC microbiology, 21*(1), 154. https://doi.org/10.1186/s12866-021-02166-6

[12] Luo, J., Li, Y., Xie, J., Gao, L., Liu, L., Ou, S., Chen, L., & Peng, X. (2018). The primary biological network of Bifidobacterium in the gut. *FEMS microbiology letters, 365*(8), 10.1093/femsle/fny057. https://doi.org/10.1093/femsle/fny057

[13] Xia X. (2018). DAMBE7: New and Improved Tools for Data Analysis in Molecular Biology and Evolution. *Molecular biology and evolution, 35*(6), 1550–1552. https://doi.org/10.1093/molbev/msy073

[14] Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics, 20*(4), 1160–1166. https://doi.org/10.1093/bib/bbx108

[15] Sun, X., Yang, Q., & Xia, X. (2013). An improved implementation of effective number of codons (nc). *Molecular biology and evolution, 30*(1), 191–196. https://doi.org/10.1093/molbev/mss201

[16] Wright F. (1990). The 'effective number of codons' used in a gene. *Gene, 87*(1), 23–29. https://doi.org/10.1016/0378-1119(90)90491-9

[17] Komar A. A. (2019). *Molekuliarnaia biologiia, 53*(6), 883–898. https://doi.org/10.1134/S0026898419060090

[18] Shen, W., Wang, D., Ye, B., Shi, M., Ma, L., Zhang, Y., & Zhao, Z. (2015). GC3-biased gene domains in mammalian genomes. *Bioinformatics (Oxford, England), 31*(19), 3081–3084. https://doi.org/10.1093/bioinformatics/btv329

[19] Tatarinova, T., Elhaik, E., & Pellegrini, M. (2013). Cross-species analysis of genic GC3 content and DNA methylation patterns. *Genome biology and evolution, 5*(8), 1443–1456. https://doi.org/10.1093/gbe/evt103

[20] Ito, K., & Murphy, D. (2013). Application of ggplot2 to Pharmacometric Graphics. *CPT: pharmacometrics & systems pharmacology, 2*(10), e79. https://doi.org/10.1038/psp.2013.56

[21] Morrison D. A. (1996). Phylogenetic tree-building. *International journal for parasitology, 26*(6), 589–617. https://doi.org/10.1016/0020-7519(96)00044-6

[22] Yu G. (2020). Using ggtree to Visualize Data on Tree-Like Structures. *Current protocols in bioinformatics, 69*(1), e96. https://doi.org/10.1002/cpbi.96

[23] Sueoka N. (1988). Directional mutation pressure and neutral molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America, 85*(8), 2653–2657. https://doi.org/10.1073/pnas.85.8.2653

[24] Sun, X., Yang, Q., & Xia, X. (2013). *An improved implementation of effective number of codons (nc). Molecular biology and evolution, 30*(1), 191–196. https://doi.org/10.1093/molbev/mss201

[25] Forsdyke, D. R., & Mortimer, J. R. (2000). *Chargaff's legacy. Gene, 261*(1), 127–137. https://doi.org/10.1016/s0378-1119(00)00472-8

[26] Forsdyke D. R. (2021). *Neutralism versus selectionism: Chargaff's second parity rule, revisited. Genetica, 149*(2), 81–88. https://doi.org/10.1007/s10709-021-00119-5

[27] Sueoka N. (1999). Translation-coupled violation of Parity Rule 2 in human genes is not the cause of heterogeneity of the DNA G+C content of the third codon position. *Gene, 238*(1), 53–58. https://doi.org/10.1016/s0378-1119(99)00320-0

[28] Seemann T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics (Oxford, England), 30(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

[29] Sanger-pathogens.github.io. (n.d.). *Roary: The pan genome pipeline*. https://sanger-pathogens.github.io/Roary/.

[30] Rambaut. (n.d.). *Release figtree V1.4.4 · Rambaut/Figtree*. https://github.com/rambaut/figtree/releases/tag/v1.4.4.

[31] Hildebrand, F., Meyer, A., & Eyre-Walker, A. (2010). *Evidence of selection upon genomic GC-content in bacteria. PLoS genetics, 6*(9), e1001107. https://doi.org/10.1371/journal.pgen.1001107

[32] Zhou, Z., Dang, Y., Zhou, M., Li, L., Yu, C. H., Fu, J., Chen, S. & Liu, Y. (2016). Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proceedings of the National Academy of Sciences*, *113*(41), E6117-E6125. https://doi.org/10.1073/pnas.1606724113.