ISSN 2303-4521
Periodicals of Engineering and Natural Sciences *Original Research*
Vol. 11, No. 2, March 2023, pp.62-77

Molecular analysis of 16s-rRNA and associated gene segments for identification of probiotic phenotypes

Amar Ćosić [1], Nadia Islam[1], Altijana Hromić- Jahjefendić[1]
[1] Department of Genetics and Bioengineering, International University of Sarajevo, Bosnia and Herzegovina

ABSTRACT

The 16s-rRNA consists of hypervariable regions (V1 – V9) that demonstrate considerable sequence diversity among different bacteria. Species-specific sequences within a given hypervariable region constitute useful targets for diagnostic assays and other scientific investigations. Usually, the size of the gene region is 1500 bp, which is large enough to be analyzed using bioinformatic tools and applied for detection. The need to advance the knowledge of the 16s-rRNA gene segments in bacterial strains would allow better understanding and better diagnostic possibilities when dealing with them. This could also be the basis for investigation of pathogenic microorganisms.

Keywords: Probiotics, 16s-rRNA, Hypervariable Regions, V3 and V4 regions

Corresponding Author:

Altijana Hromić- Jahjefendić
Department of Genetics and Bioengineering
International University of Sarajevo
Bosnia and Herzegovina
E-mail: ahromic@ius.edu.ba

1. Introduction

Investigation of the role of good bacteria in various diseases and illnesses, such as alcoholic liver injury, asthma, allergic rhinitis, hepatocellular carcinoma, and multidrug resistance [1]–[4], has been conducted, with a focus on five genera of bacteria in this study: *Bacillus*, *Bifidobacterium*, *Lactobacillus*, *Lactococcus*, and *Streptococcus*. These bacteria were classified according to their metabolism, specifically their heterofermentative and homofermentative abilities, and their genome. Bacteria that produce only one fermentation product were classified as homofermentative, such as *L. acidophilus* and its lactic acid fermentation product. Bacteria that produce more than one fermentation product were classified as heterofermentative, such as *L. brevis*, which produces lactic acid and ethanol.

A large number of studies have investigated the safety of probiotics, particularly those containing *Lactobacillus* species, with most showing either a positive or no influence on human health [5]–[7]. Potential negative effects of probiotics are still being explored, and while they are generally considered to be beneficial to health, they may be detrimental to individuals with compromised immunity. In such cases, the introduction of probiotics to the human gut microbiome may result in sepsis, *bacteraemia* and even death [8], [9]. Some reports have suggested an association between *lactobacillemia* in three AIDS patients and the intake of *Lactobacillus rhamnosus GG*, although further research is required to fully ascertain the impact of probiotics on human health [10], [11].

Analysis of 16s-rRNA genes remains a central method in microbiology, serving both to explore microbial diversity and as a day-to-day tool for bacterial identification. Such identification techniques are generally easier to interpret than molecular phylogenetic analyses and are often preferred when the groups are well understood. Since the seminal work of Carl R. Woese et al. in 1977 [12], research on 16s-rRNA has continued to grow in popularity, with 35577 publications appearing in PubMed in the five years preceding this project [13]. Recent studies utilizing 16s-rRNA have included the molecular identification of clinical Nocardia isolates, novel identification, sample screening, and other applications [14]–[17].

2. Research material and method

2.1. Method

Materials used in this project can be available on demand.

2.1.1. Bacterial growth and enumeration

A growth environment was established for bacteria by utilizing an incubator (Innova 42 Incubator Shaker Series, Eppendorf North America, USA) set to 37℃ for 24h. Three different media were employed for probiotic growth: the standard medium of DeMan, Rogosa and Sharpe (MRS) for *Lactobacillus*; Lysogeny broth (LB) as a control medium for verifying the sterility of the environment; and HHD agar utilized for bacteria separation and enumeration based on the fermentation process. The latter medium contains bromcresol green, which reacts with the pH changes caused by the bacterial fermentation process. A heterofermentative bacterium will not alter the colour of the medium, which remains blue, whereas a homofermentative bacterium will result in a medium colour change to green if the pH drops below 3.8 or will remain blue within the pH range of 5.4. A four-step protocol was used to carry out the growth, separation, and enumeration of bacteria. To begin, 0.5g of the bacterial mix from a capsule was dissolved in 20ml of the medium and incubated for 24h at 37°C. This yielded a mix of different probiotics, which was then transferred to agar plates (LB, MRS, HHD). The plates were incubated for 24h at 37°C, after which one colony was taken and transferred to a fresh medium for a further 24h. This enabled enumeration of a single colony. Additionally, three known bacteria (two different strains of *L. plantarum* and *B. subtilis*) were grown, and four random colonies were taken from the different plates.

2.1.2. DNA isolation, primer design and polymerase chain reaction (PCR)

Isolation of genomic bacterial DNA was conducted using a commercial bacterial kit (details available on demand). Quality control was performed with a Multiskan™ GO Microplate Spectrophotometer (Thermo Scientific™, United States), and the isolated DNA was subsequently utilized for a PCR reaction. Relevant literature was consulted, and primers were designed based on previous research [18]–[21]. These primers (details available on demand) were tested with the Silva test primer, and those with the highest coverage and specificity were chosen [19]. Coverage is a value that measures the number of sequences that are matched in one taxonomic unit by the matched or mismatched sequences, with higher values indicating more favorable primers for that taxonomic unit. Similarly, specificity indicates how accurately the primer fits the overall sequence in the database, with higher values indicating better precision. A gradient PCR was then employed to determine the optimal annealing temperature of 56 ℃, as gradient PCR allows for different temperatures in each well. The setup and thermal cycler utilized for this project are available on demand, and the StepOnePlus™ Real-Time PCR System (Germany) was employed.

2.1.3. Sequencing and sequence alignment tools

Sequencing was done in MedSankTek company in Turkey. The company utilized the Sanger method with a single read, resulting in the acquisition of sequencing data. Before the sequencing process, the samples were purified. Bioinformatic analysis was conducted to assess sequencing and nucleotide identification using the Phred score. This score is a numerical representation of the probability of a nucleotide being read incorrectly, calculated using Formula 1 based on a logarithmic relationship. Quality values were further characterized from the Phred website. Two software programs, Phred and Codon Code Aligner, were utilized to facilitate the analysis.

$$Q = -10 \log_{10} P \text{ or } P=10^{-Q/10}$$

Formula 1: Phred score calculation

After performing sequence alignment, the next step was to identify the bacterial sequence.

The efficacy of four sequence identification algorithms was tested for the purpose of this project. The algorithms, BLASTn, RDP, USEARCH, and VSEARCH, were divided into two categories: those with a graphical interface and those that do not. Upon testing, they were graded according to the ease and speed of use. BLASTn and RPD featured graphical interfaces, while USEARCH and VSEARCH needed to be programmed through batch commands or other command-based programming.

BLASTn is a widely used program for sequence comparison and has a user-friendly interface and access to a large database. It is suitable for short sequences and evaluates alignments through parameters like maximum score, total score, e-value, percentage identity, and accession number. The speed of BLAST is affected by the size of the job and time of day, with completion time ranging from seconds to days. For Europeans, the optimal time to run a BLAST job is 6-12 am.

The Ribosomal Database Project (RDP) contains 2.8 million annotated sequences from bacteria, archaea, and fungi. It provides a Hierarchical Browser, Classifier, Probe Match, FunGene, Library Compare, Sequence Match, RDPipeline, Aligner, and Tree Builder. It provides a seqmatch score to reflect the number of oligomers shared between two sequences. RDP suffers from the same problems as BLAST, with time to execution dependent on server load. USEARCH and VSEARCH are alternatives with faster execution and wider database use.

USEARCH and VSEARCH are efficient due to their combination of multiple algorithms. Both support Bash and Perl programming languages; USEARCH is open source in 32 bit version, but VSEARCH is open source with no memory limit. VSEARCH is based on the USEARCH method, which compares "words" to the query to find similar sequences. Six databases were used for comparison: HOMD, NCBI, SILVA, GreenGenes, RDP, and prokMSA, with prokMSA the largest. Both programs were run in parallel on a Linux OS with 4GB RAM.

3. Results

3.1. Cultivation of bacteria

As demonstrated in Chapter 2: Materials and Methods, the bacteria were cultivated in three separate media. These media successfully demonstrated the ability to cultivate probiotics, with the best growth being found in MRS medium.

3.2. DNA isolation and PCR

The DNA was isolated utilizing the kit provided in a period of 60 minutes. Visualization of the isolated DNA was achieved via gel electrophoresis. Table 1 depicts the purity and overall concentration of the isolated DNA yield.

Table 1. DNA quality

Samples	DNA Concentration *ng*/µl	DNA purity (260/280)
Bacillus subtilis	16.25	1.76
***Lactobacillus plantarum* strain**	17.35	1.72
Lactobacillus plantarum	16.95	1.81
Random colony	17.59	1.74
Random colony	17.30	1.76
Random colony	17.35	1.74
Random colony	17.41	1.74
Mean value (standard deviation)	**17.17 (0.416)**	**1.75 (0.026)**

PCR was performed after quality assurance in a volume of 30 µl, with an approximate total reaction time of 1.1h. To visualize the amplicons, a 1:5 ratio of 6x loading dye to PCR product was added, with 1 µl of dye on every 5 µl of PCR product.

Confirmation of the ability to amplify the target regions and the appropriateness of the primer temperature were established, thus allowing for the purification of the PCR product through the use of the gel extraction kit, as previously outlined in Table 1. Subsequently, the purified DNA was sent for sequencing.

3.3. Sequencing and bioinformatic analysis

Seven days after the samples were sent, the full sequences were acquired utilizing the Codon Code Aligner can be available on demand. Table 2 provides a synopsis of the acquired sequences.

Table 2. Length and quality of sequencing

Sample	Length	Quality	Q>20	Q>30	Q>40	GC content (%)
Sequence 1	440	393	393	378	355	53.9
Sequence 2	264	174	174	144	123	52.3
Sequence 3	262	237	237	229	218	55
Sequence 4	350	287	287	253	227	48.3
Sequence 5	215	68	68	38	25	53.5
Sequence 6	262	225	225	196	151	50.8
Sequence 7	556	456	456	424	386	52.3
Sequence 11	415	330	330	295	268	53.7
Sequence 13	262	226	226	211	195	55
Sequence 14	439	384	384	357	325	54.4
Sequence 15	261	228	228	214	202	54.4
Sequence 17	441	395	395	377	351	54.9
Sequence 18	259	233	233	232	223	55.2
Sequence 20	437	415	415	410	397	54.5
Sequence 21	236	166	168	143	130	52.9
Sequence 27	442	384	384	355	317	54.8
Sequence 28	259	234	234	232	228	55.2
Sequence 29	638	402	402	270	96	47
Sequence 30	260	204	204	160	95	51.2
Sequence 32	726	536	536	481	437	56.6
Sequence PL	48	2	2	/	/	52.1

Confirmation of sequence quality was accomplished through evaluation of read quantity and quality. It was observed that a majority of nucleotide reads possessed a satisfactory level of quality, with the exception of a sample marked by PL. Subsequently, sequence alignment was performed to further the analysis, with the results being divided into categories according to the utilized tool.

3.3.1. BLASTn and ribosomal database project

BLASTn was the first tool tested in this study. Twenty-one sequences were obtained, with twenty successfully identified; the only exception was PL, which was not identified due to its short length. As BLASTn queries the NCBI database, only a single result was returned. Table 3 provides further information, revealing six organisms as identified by BLASTn: *B. subtilis, L. herbarum, B. velezensis, L. plantarum, O. iheyensis,* and *B. coagulans*. From the list, only *L. plantarum* and *B. subtilis* were found in the mix, indicating that BLASTn made ten incorrect predictions out of twenty possible. Analysis of 21 sequences via RDP revealed a match, with the last sequence (Sequence_PL) excluded from the BLASTn analysis. Table 4 documents the results of RDP, which indicate that most of the sequences belong to the species *B. subtilis.*

Table 3. BLASTn results and quality scores

Name of Sample	BLASTn						
	Species	**Max Score**	**Total Score**	**Query Cover (%)**	**E Value***	**Per. Ident (%)**	**Accession ID**
Sequence 1	*Bacillus subtilis strain*	756	756	96.00	0	98.30	NR_112116.2
Sequence 2	*Lactobacillus herbarum*	387	387	94.00	2.00E-107	94.82	NR_145899.1
Sequence 3	*Bacillus velezensis*	448	448	95.00	1.00E-125	98.81	NR_075005.2
Sequence 4	*Lactobacillus plantarum*	595	595	97.00	4.00E-170	98.24	NR_104573.1
Sequence 5	*Bacillus subtilis subsp. Subtilis*	209	209	93.00	4.00E-54	85.71	NR_102783.2
Sequence 6	*Lactobacillus plantarum*	457	457	96.00	2.00E-128	99.21	NR_104573.1
Sequence 7	*Bacillus subtilis str*	776	921	92.00	0	99.53	NR_112116.2
Sequence 11	*Bacillus subtilis*	688	688	97.00	0	97.30	NR_112116.2
Sequence 13	*Bacillus velezensis*	446	446	95.00	3.00E-125	98.80	NR_075005.2
Sequence 14	*Bacillus subtilis*	743	743	97.00	0	97.91	NR_112116.2
Sequence 15	*Bacillus velezensis*	448	448	96.00	1.00E-125	98.44	NR_075005.2
Sequence 17	*Bacillus subtilis*	761	761	96.00	0	99.06	NR_112116.2
Sequence 18	*Bacillus velezensis*	453	453	97.00	2.00E-127	98.83	NR_075005.2
Sequence 20	*Bacillus subtilis*	773	773	97.00	0	99.53	NR_112116.2
Sequence 21	*Oceanobacillus iheyensis*	257	257	52.00	2.00E-68	100.0	NR_075027.2
Sequence 27	*Bacillus subtilis*	776	776	95.00	0	99.76	NR_112116.2
Sequence 28	*Bacillus velezensis*	453	453	96.00	2.00E-127	99.21	NR_075005.2
Sequence 29	*Lactobacillus plantarum*	737	940	93.00	0	98.11	NR_104573.1
Sequence 30	*Lactobacillus herbarum*	444	444	97.00	1.00E-124	98.05	NR_145899.1
Sequence 32	*Bacillus coagulans*	785	1180	98.00	0	100.00	NR_041523.1
Sequence PL	*X*	X	X	X	X	X	X

Table 4. RDP results and quality scores

Name of Sample	RDP				
	Species	**Accession ID**	**S_ab score**	**Unique common oligomers**	**short ID**
Sequence 1	*Bacillus subtilis*	GQ392049	0.939	541	S001610594
Sequence 2	Uncultured bacterium	GQ477887	0.76	1413	S002039504
Sequence 3	*Bacillus subtilis*	AY879290	0.937	1407	S000481202
Sequence 4	*Lactobacillus plantarum*	JX003595	0.907	1223	S003299715
Sequence 5	*Bacillus sp.*	LN874212	0.44	1968	S004508187
Sequence 6	*Lactobacillus paraplantarum*	AJ306297	0.941	1400	S000000066
Sequence 7	*Bacillus subtilis*	HM588154	0.869	1370	S002231960
Sequence 11	*Bacillus subtilis*	HM216569	0.885	1167	S002227183
Sequence 13	*Bacillus subtilis*	AY879290	0.913	1407	S000481202
Sequence 14	*Bacillus sp.*	DQ643081	0.94	397	S000712062
Sequence 15	*Bacillus sp.*	AY859753	0.945	586	S000478601
Sequence 17	*Bacillus subtilis*	GQ392049	0.941	541	S001610594
Sequence 18	*Bacillus sp.*	AY859753	0.984	586	S000478601
Sequence 20	*Bacillus subtilis*	AY917143	0.967	1384	S000491520
Sequence 21	*Bacillus subtilis*	AY881638	0.647	1404	S000481489
Sequence 27	*Bacillus subtilis*	DQ057582	0.942	1379	S000537305
Sequence 28	*Bacillus sp.*	AY859753	0.972	586	S000478601
Sequence 29	*Lactobacillus pentosus*	AB362657	0.703	1367	S000941694
Sequence 30	*Lactobacillus plantarum*	AM157432	0.925	1437	S000617874
Sequence 32	*Bacillus coagulans*	AF346895	0.788	1378	S000005947
Sequence PL	unidentified bacterium	X87269	0.452	1406	S000006576

If the value is lower than 5e-120 then the BLAST interprets this as a 0. Average time required was 00:04:43,37 which represents higher amount of time then other algorithms mentioned in the research. Average time required to complete every sequence on RDP in on this list was 00:00:35, 92. This indicates a very low amount of time required actually to run and to get results using online RDP compared to NCBI which was substantially slower.

3.3.2. USEARCH

USEARCH was the first tool to lack a graphical user interface. The results were stratified according to the database utilized; out of six available databases, results were obtained for GreenGenes, HOMD, NCBI, and SILVA. However, the other two databases, RDP and prokMSA, could not be processed due to the constraints of the 32-bit version of the software.

Table 5. USEARCH result for GreenGene, HOMD, and SILVA databases

*Percentage of matched sequence. **Sequence length after the Gap was introduced. ***Number of Mismatches. **** Original Sequence Length before gap

Sample	Database	Species	Database ID	NCBI ID	ID % *	Seq. Length **	MM ***	Gap	Org. Seq. Length ***	Target Length
Sequence 1	**GreenGene**	*Bacillus subtilis subsp. spizizenii*	14756	AF074970.1	96.6	441	12	3	439	1409
	HOMD	*Bacillus subtilis* str. ATCC 6633	14760	AB018486.1	96.4	441	13	3	439	1507
	NCBI	*Bacillus subtilis*	14754	Z82044.1	96.6	441	12	3	439	1551
	SILVA	*Bacillus subtilis* subsp. *spizizenii*	14756	AF074970.1	96.6	441	12	3	439	1409
Sequence 2	**GreenGene**	*Lactobacillus plantarum*	64381	AF515222.1	90.4	271	14	7	264	1528
	HOMD	*Lactobacillus casei subsp. rhamnosus* str. JCM 1136	15214	D16552.1	91.1	271	12	8	264	1521
	NCBI	*Lactobacillus plantarum*	64381	AF515222.1	90.4	271	14	7	264	1528
	SILVA	*Lactobacillus sp.* str. B5407	46388	AB070610.1	90.4	271	14	7	264	1481
Sequence 3	**GreenGene**	*Bacillus subtilis subsp. Marburg*	14731	D26185.1	94.8	268	8	2	262	1552
	HOMD	*Bacillus subtilis* str. ATCC 6633	14760	AB018486.1	94.4	268	9	2	262	1507
	NCBI	*Lactobacillus sp.* str. MR-2	64526	AF516755.1	95.7	351	14	1	350	1522
	SILVA	*Bacillus subtilis* subsp. *Marburg*	14731	D26185.1	94.8	268	8	2	262	1552
Sequence 4	**GreenGene**	*Lactobacillus plantarum*	64381	AF515222.1	95.7	351	14	1	350	1528
	HOMD	*Lactobacillus brevis* str. K9	15172	AF090328.1	92.6	351	25	1	350	1449
	NCBI	*Bacillus sp.* SGE173(2010) str. SGE173	712682	HM566699.1	98.1	262	4	1	262	1318
	SILVA	*Lactobacillus plantarum*	64381	AF515222.1	95.7	351	14	1	350	1528
Sequence 6	**GreenGene**	*Lactobacillus paraplantarum*	31888	AJ306297.1	95.9	267	6	2	262	1502
	HOMD	*Lactobacillus rhamnosus* str. F11	48136	AF243146.1	94.8	267	9	2	262	1516
	NCBI	*Lactobacillus plantarum*	64381	AF515222.1	95.9	267	6	2	262	1528
	SILVA	*Lactobacillus plantarum*	15163	M58827.1	95.9	267	6	2	262	1570
Sequence 11	**GreenGene**	*Bacillus subtilis subsp. spizizenii*	14756	AF074970.1	95.2	416	18	2	415	1409
	HOMD	*Bacillus subtilis* str. ATCC 6633	14760	AB018486.1	95	416	19	2	415	1507
	NCBI	*Bacillus subtilis subsp. spizizenii*	14756	AF074970.1	95.2	416	18	2	415	1409
	SILVA	*Bacillus subtilis* subsp. *spizizenii*	14756	AF074970.1	95.2	416	18	2	415	1409
Sequence 13	**GreenGene**	*Bacillus subtilis*	112048	AY881643.1	95.4	262	12	0	262	1485
	HOMD	*Bacillus subtilis* str. ATCC 6633	14760	AB018486.1	94.7	263	13	1	262	1507
	NCBI	*Bacillus sp.* SGE173(2010)	712682	HM566699.1	96.6	262	9	0	262	1318
	SILVA	*Bacillus subtilis*	112048	AY881643.1	95.4	262	12	0	262	1485

Sample	Database	Species	Database ID	NCBI ID	ID % *	Seq. Length **	MM ***	Gap	Org. Seq. Length ***	Target Length
Sequence 14	**GreenGene**	*Bacillus subtilis subsp. spizizenii*	14756	AF074970.1	96.1	440	14	2	439	1409
	HOMD	*Bacillus subtilis* str. ATCC 6633	14760	AB018486.1	95.9	440	15	2	439	1507
	NCBI	*Bacillus subtilis subsp. spizizenii*	14756	AF074970.1	96.1	440	14	2	439	1409
	SILVA	*Bacillus subtilis* subsp. *spizizenii*	14756	AF074970.1	96.1	440	14	2	439	1409
Sequence 15	**GreenGene**	*Bacillus subtilis s*	277872	EU627167.1	95.8	262	10	1	261	1514
	HOMD	*Bacillus subtilis* str. ATCC 6633	14760	AB018486.1	95	262	12	1	261	1507
	NCBI	*Bacillus sp.* SGE173(2010)	712682	HM566699.1	96.9	261	8	0	261	1318
	SILVA	*Bacillus subtilis*	277872	EU627167.1	95.8	262	10	1	261	1514
Sequence 17	**GreenGene**	*Bacillus subtilis*	14759	AB018484.1	96.6	442	12	3	440	1506
	HOMD	*Bacillus subtilis* str. ATCC 6633	14760	AB018486.1	96.4	442	13	3	440	1507
	NCBI	*Bacillus subtilis str.* PY79	14758	AF142577.1	96.6	442	12	3	440	1407
	SILVA	*Bacillus subtilis*	14759	AB018484.1	96.6	442	12	3	440	1506
Sequence 18	**GreenGene**	*Bacillus subtilis subsp. Marburg*	14731	D26185.1	95.8	265	5	2	259	1552
	HOMD	*Bacillus subtilis* str. ATCC 6633	14760	AB018486.1	95.5	265	6	2	259	1507
	NCBI	*Bacillus sp.* SGE173(2010) str. SGE173	712682	HM566699.1	99.2	259	1	1	259	1318
	SILVA	*Bacillus subtilis* subsp. *Marburg*	14731	D26185.1	95.8	265	5	2	259	1552
Sequence 20	**GreenGene**	*Bacillus subtilis subsp. spizizenii*	14756	AF074970.1	97.5	438	9	2	437	1409
	HOMD	*Bacillus subtilis* str. ATCC 6633	14760	AB018486.1	97.3	438	10	2	437	1507
	NCBI	*Bacillus subtilis* subsp. *spizizenii*	14756	AF074970.1	97.5	438	9	2	437	1409
	SILVA	Bacillus subtilis subsp. spizizenii	14756	AF074970.1	97.5	438	9	2	437	1409
Sequence 27	**GreenGene**	*Bacillus subtilis subsp. spizizenii*	14756	AF074970.1	97.5	442	10	1	442	1409
	HOMD	*Bacillus subtilis* str. ATCC 6633	14760	AB018486.1	97.3	442	11	1	442	1507
	NCBI	*Bacillus subtilis* subsp. *spizizenii*	14756	AF074970.1	97.5	442	10	1	442	1409
	SILVA	*Bacillus subtilis* subsp. *spizizenii*	14756	AF074970.1	97.5	442	10	1	442	1409
Sequence 28	**GreenGene**	*Bacillus sp.*	14734	AB017587.1	95.8	265	5	2	259	1516
	HOMD	*Bacillus subtilis* str. ATCC 6633	14760	AB018486.1	95.5	265	6	2	259	1507
	NCBI	*Bacillus sp.* str. PM3	14741	AB017588.1	95.8	265	5	2	259	1516
	SILVA	*Bacillus sp.* str. SSA3	14734	AB017587.1	95.8	265	5	2	259	1516
Sequence 30	**GreenGene**	*Lactobacillus plantarum*	141422	AM157432.1	95.8	261	10	1	260	1534
	HOMD	*Lactobacillus rhamnosus* str. F11	48136	AF243146.1	93.2	266	11	3	260	1516
	NCBI	*Lactobacillus plantarum*	141422	AM157432.1	95.8	261	10	1	260	1534
	SILVA	*Lactobacillus plantarum*	141422	AM157432.1	95.8	261	10	1	260	1534

The time required to carry out the test was found by running it three times and taking the average value. The times required for the four databases used (GreenGene, HOMD, NCBI, and SILVA) were recorded in the order they were presented as 00:01:90.23, 00:00:01.45, 00:00:45.56, and 00:01:86.66, respectively. The average time of USEARCH as a software was 00:01:05.83. It is worth noting that this was a 32-bit version of the software, and that only four out of six databases were used.

3.3.3. VSEARCH

VSEARCH was the second tool without the graphical interface. The results were segmented based on the source database, and access to the 64-bit version allowed for analysis of all six databases.

Table 6. VSEARCH results, GreenGenes, HOMD, NCBI, SILVA, RDP and prokMSA databases

Sample	Database	Species	Database ID	NCBI ID	Matched Nucleotides (%)	Gap %	Target Length
Sequence 1	**GreenGene**	Halotolerant aerobic waters and shal groundwater along Rouge southeastern Michigan river water clone 9-sw-su5-2 *(Bacillus)*	195322	DQ981833.1	426 (96.6)	3 (0.7)	1399
	HOMD	*Bacillus subtilis*	14760	AB018486.1	425 (96.4)	3 (0.7)	1507
	NCBI	*Bacillus subtilis*	790492	HM117721.1	426 (96.6)	3 (0.7)	1298
	SILVA	Halotolerant aerobic waters and shal groundwater along *(Bacillus)*	195322	DQ981833.1	426 (96.6)	3 (0.7)	1399
	RDP	*Bacillus subtilis* str. SK09	790492	HM117721.1	426 (96.6)	3 (0.7)	1298
	prokMSA	*Bacillus sp.* str. QLPB08	1105502	JF346899.1	426 (96.6)	3 (0.7)	1511
Sequence 2	**GreenGene**	Metagenomic gut microbiome *(Lactobacillus)*	147397	DQ327193.1	245 (90.4)	12 (4.4)	1387
	HOMD	*Lactobacillus rhamnosus*	48136	AF243146.1	246 (90.8)	12 (4.4)	1516
	NCBI	*Lactobacillus plantarum*	161185	AM279764.2	245 (90.4)	12 (4.4)	1488
	SILVA	Metagenomic gut microbiome healthy *(Lactobacillus)*	147397	DQ327193.1	245 (90.4)	12 (4.4)	1387
	RDP	Mastoidis- O. var. koroneiki-generated wastewaters: influence cultivation and harvesting practice on structure Olea europaea cv. *(Lactobacillus)*	704038	GQ477887.1	246 (90.8)	12 (4.4)	1510
	prokMSA	Mastoidis- O. var. koroneiki-generated wastewaters: influence cultivation and harvesting practice on structure Olea europaea cv. Mastoidis (*Lactobacillus)*	704038	GQ477887.1	246 (90.8)	12(4.4)	1510
Sequence 3	**GreenGene**	structure receiving long-term augmentations chromium contaminated wastes landfill sediments *(Bacillus)*	237766	DQ899879.1	254 (95.1)	5 (1.9)	1386
	HOMD	*Bacillus subtilis*	14760	AB018486.1	253 (94.4)	6 (2.2	1507
	NCBI	*Bacillus sp*. SGE173(2010) str. SGE173	712682	HM566699.1	257 (98.1)	1 (0.4)	1318
	SILVA	Structure receiving long-term augmentations chromium contaminated wastes landfill sediments *(Bacillus)*	237766	DQ899879.1	254 (95.1)	5 (1.9)	1386
	RDP	*Bacillus sp.* SGE173(2010) str. SGE173	712682	HM566699.1	257 (98.1)	1 (0.4)	1318
	prokMSA	Mastoidis- O. var. koroneiki-generated wastewaters: influence cultivation and harvesting practice on structure Olea europaea cv. Mastoidis *(Lactobacillus)*	712682	GQ477898.1	257 (98.1)	1 (0.4)	1318
Sequence 4	**GreenGene**	*Lactobacillus plantarum*	257487	EU081013.1	337 (95.7)	2 (0.6)	1392
	HOMD	*Lactobacillus brevis*	15172	AF090328.1	325 (92.6)	1 (0.3)	1449
	NCBI	*Lactobacillus plantarum*	257487	EU081013.1	337 (95.7)	2 (0.6)	1392
	SILVA	*Lactobacillus plantarum*	257487	EU081013.1	337 (95.7)	2 (0.6)	1392
	RDP	Mastoidis- O. var. koroneiki-generated wastewaters: influence cultivation and harvesting practice on structure Olea europaea cv.	762182	GQ477898.1	337 (96.0)	1 (0.3)	1519

Sample	Database	Species	Database ID	NCBI ID	Matched Nucleotides (%)	Gap %	Target Length
		(Lactobacillus)					
	prokMSA	Mastoidis- O. var. koroneiki-generated wastewaters: influence cultivation and harvesting practice on structure Olea europaea cv. Mastoidis *(Lactobacillus)*	762182	GQ477898.1	337 (96.0)	1 (0.3)	1519
Sequence 6	**GreenGene**	*Lactobacillus plantarum*	557831	GU138564.1	256 (95.9)	5 (1.9)	1481
	HOMD	*Lactobacillus rhamnosus*	48136	AF243146.1	253 (94.8)	5 (1.9)	1516
	NCBI	*Lactobacillus plantarum*	713476	FJ861328.1	256 (95.9)	5 (1.9)	1295
	SILVA	*Lactobacillus plantarum*	557831	GU138564.1	256 (95.9)	5 (1.9)	1481
	RDP	Temporal succession biological degreasing systems clone CapF3B.11 *(Bacillus)*	557831	HM152578.1	256 (95.9)	5 (1.9)	1481
	prokMSA	Temporal succession biological degreasing systems clone CapF3B.11 *(Bacillus)*	557831	HM152578.1	256 (95.9)	5 (1.9)	1481
Sequence 11	**GreenGene**	*Bacillus subtilis*	105773	AY296804.1	397 (95.4)	2 (0.5)	1373
	HOMD	*Bacillus subtilis*	14760	AB018486.1	395 (95.0)	2 (0.5)	1507
	NCBI	*Bacillus subtilis*	105773	AY296804.1	397 (95.4)	2 (0.5)	1373
	SILVA	*Bacillus subtilis*	105773	AY296804.1	397 (95.4)	2 (0.5)	1373
	RDP	*Bacillus subtilis* str. IBT012	105773	AY296804.1	397 (95.4)	2 (0.5)	1373
	prokMSA	*Bacillus subtilis*	105773	AY296804.1	397 (95.4)	2 (0.5)	1373
Sequence 13	**GreenGene**	*Bacillus amyloliquefaciens*	248942	EU164542.1	251 (95.4)	1(0.4)	1406
	HOMD	*Bacillus subtilis*	14760	AB018486.1	249 (94.7)	1 (0.4)	1507
	NCBI	*Bacillus sp*. SGE173(2010) str. SGE173	712682	HM566699.1	253 (96.6)	0 (0)	1318
	SILVA	*Bacillus amyloliquefaciens*	248942	EU164542.1	251 (95.4)	1 (0.4)	1406
	RDP	*Bacillus sp*. SGE173(2010) str. SGE173	712682	HM566699.1	253 (96.6)	0 (0.0)	1318
	prokMSA	*Bacillus sp.* SGE173(2010)	712682	HM566699.1	253 (96.6),	0 (0.0)	1318
Sequence 14	**GreenGene**	*Bacillus sp.*	251574	AB330409.1	423 (96.1)	3 (0.7)	1475
	HOMD	*Bacillus subtilis*	14760	AB018486.1	422 (95.9)	3 (0.7)	1507
	NCBI	*Bacillus sp*. str. BAM522	251574	AB330409.1	423 (96.1)	3 (0.7)	1475
	SILVA	*Bacillus sp*. str. BAM522	251574	AB330409.1	423 (96.1)	3 (0.7	1475
	RDP	Temporal succession biological degreasing systems clone CapF3B.11 *(Bacillus)*	754955	HM152578.1	424 (96.4)	3 (0.7)	1514
	prokMSA	Temporal succession biological degreasing systems clone CapF3B.11 *(Bacillus)*	754955	HM152578.1	424 (96.4)	3 (0.7)	1514
Sequence 15	**GreenGene**	*Bacillus amyloliquefaciens*	248942	EU164542.1	249 (95.0)	1 (0.4)	1406
	HOMD	*Bacillus subtilis*	14760	AB018486.1	249 (95.0)	1 (0.4)	1507
	NCBI	*Bacillus sp*. SGE173(2010) str. SGE173	712682	HM566699.1	253 (96.9)	0 (0)	1318
	SILVA	*Bacillus amyloliquefaciens*	248942	EU164542.1	249 (95.0)	1 (0.4)	1406
	RDP	*Bacillus sp*. SGE173(2010) str. SGE173	712682	HM566699.1	253 (96.9)	0 (0.0)	1318
	prokMSA	*Bacillus sp.* SGE173(2010)	712682	HM566699.1	253 (96.9)	0 (0.0)	1318
Sequence 17	**GreenGene**	Halotolerant aerobic waters and shal groundwater 2 *(Bacillus)*	195322	DQ981833.1	427 (96.6)	3 (0.7)	1399
	HOMD	*Bacillus subtilis*	14760	AB018486.1	426 (96.4)	3 (0.7)	1507
	NCBI	*Bacillus subtilis*	613361	FJ772085.1	427 (96.6)	3 (0.7)	1254
	SILVA	Halotolerant aerobic waters and shal groundwater along	195322	DQ981833.1	427 (96.6)	3 (0.7)	1399

Sample	Database	Species	Database ID	NCBI ID	Matched Nucleotides (%)	Gap %	Target Length
		(Bacillus)					
	RDP	Halotolerant aerobic waters and shal groundwater along *(Bacillus)*	195322	DQ981833.1	427 (96.6)	3 (0.7)	1399
	prokMSA	*Bacillus sp.*	1105502	JF346899.1	427 (96.6)	3 (0.7)	1511
Sequence 18	**GreenGene**	Structure receiving long-term augmentations chromium contaminated wastes landfill sediments *(Bacillus)*	237766	DQ899879.1	254 (96.2)	5 (1.9)	1386
	HOMD	*Bacillus subtilis*	14760	AB018486.1	253 (95.5)	6 (2.3)	1507
	NCBI	*Bacillus sp.* SGE173(2010) str. SGE173	712682	HM566699.1	257 (99.2)	1 (0.4)	1318
	SILVA	Structure receiving long-term augmentations chromium contaminated wastes landfill sediments *(Bacillus)*	237766	DQ899879.1	254 (96.2)	5 (1.9)	1386
	RDP	*Bacillus sp.* SGE173(2010) str. SGE173	712682	HM566699.1	257 (99.2)	1 (0.4)	1318
	prokMSA	*Bacillus sp.* SGE173(2010) str. SGE173	712682	HM566699.1	257 (99.2)	1 (0.4)	1318
Sequence 20	**GreenGene**	*Bacillus subtilis*	112861	AY917143.1	427 (97.5)	2 (0.5)	1469
	HOMD	*Bacillus subtilis*	14760	AB018486.1	426 (97.3)	2 (0.5)	1507
	NCBI	*Bacillus subtilis*	112861	AY917143.1	427 (97.5)	2 (0.5)	1469
	SILVA	*Bacillus subtilis*	112861	AY917143.1	427 (97.5)	2 (0.5)	1469
	RDP	*Bacillus subtilis*	112861	AY917143.1	427 (97.5)	2 (0.5)	1469
	prokMSA	*Bacillus subtilis*	112861	AY917143.1	427 (97.5)	2 (0.5)	1469
Sequence 27	**GreenGene**	*Bacillus subtilis*	589582	GQ303255.1	431 (97.5)	1 (0.2)	1502
	HOMD	*Bacillus subtilis*	14760	AB018486.1	430 (97.3)	1 (0.2)	1507
	NCBI	*Bacillus subtilis*	613361	FJ772085.1	431 (97.5)	1 (0.2%)	1254
	SILVA	*Bacillus subtilis*	589582	GQ303255.1	431 (97.5)	1 (0.2)	1502
	RDP	*Bacillus subtilis*	589582	GQ303255.1	431 (97.5)	1 (0.2)	1502
	prokMSA	*Bacillus subtilis str.* ME-N11 Bacteria	589582	GQ303255.1	431 (97.5)	1 (0.2)	1502
Sequence 28	**GreenGene**	Structure receiving long-term augmentations chromium contaminated wastes landfill sediments *(Bacillus)*	237766	DQ899879.1	254 (96.2)	5 (1.9)	1386
	HOMD	*Bacillus subtilis*	14760	AB018486.1	253 (95.5)	6 (2.3)	1507
	NCBI	*Bacillus sp.* SGE173(2010) str. SGE173	712682	HM566699.1	256 (98.8)	1 (0.4)	1318
	SILVA	Structure receiving long-term augmentations chromium contaminated wastes landfill sediments *(Bacillus)*	237766	DQ899879.1	254 (96.2)	5 (1.9	1386
	RDP	*Bacillus sp.* SGE173(2010) str. SGE173	712682	HM566699.1	256 (98.8)	1 (0.4)	1318
	prokMSA	*Bacillus sp.* SGE173(2010) str. SGE173	712682	HM566699.1	256 (98.8)	1 (0.4)	1318
Sequence 30	**GreenGene**	*Lactobacillus plantarum*	141422	AM157432.1	250 (95.8)	1(0.4)	1534
	HOMD	*Lactobacillus rhamnosus*	48136	AF243146.1	248 (93.2)	7 (2.6)	1516
	NCBI	*Lactobacillus*	141422	AM157432.1	250 (95.8)	1 (0.4)	1534
	SILVA	*Lactobacillus plantarum*	141422	AM157432.1	250 (95.8)	1 (0.4)	1534
	RDP	*Lactobacillus plantarum*	141422	AM157432.1	250 (95.8)	1 (0.4)	1534
	prokMSA	*Lactobacillus plantarum*	141422	AM157432.1	250 (95.8)	1 (0.4)	1534

The average time required to use the VSEARCH software with six different databases (GreenGene, HOMD, NCBI, SILVA, RDP, and prokMSA) was 01:95.23, 00:01.19, 00:51.56, 01:95.66, 03:55.41, and 06:43.12, respectively. USEARCH, the software used for comparison, took an average of 00:14:41.36 to process the same six databases. USEARCH took longer, however, it was able to use larger databases such as RDP and prokMSA. Using a threshold of 50%, the species of each sequence were identified by determining which organism had the highest overall hits. This score was divided by the maximum score (12) and multiplied by 100%. If the score was higher than 50%, it was considered to be a positive confirmation. Based on the 21 results obtained, the most common results are provided in Table 6.

Table 7: Comparison of all software results

Sample	Species	#Hit algorithm/#Total algorithms (%)*
Sequence 1	*Bacillus subtilis*	(10/12, 83.33%)
Sequence 2	*Lactobacillus plantarum*	(3/12, 25%)
Sequence 3	*Bacillus subtilis*	(7/12, 58,33%)
Sequence 4	*Lactobacillus plantarum*	(5/12, 41,67%)
Sequence 5**	/	/
Sequence 6	*Lactobacillus plantarum*	(6/12, 60%)
Sequence 7	/	/
Sequence 11	*Bacillus subtilis*	(11/12, 91,67%)
Sequence 13	*Bacillus subtilis*	(6/12, 50%)
Sequence 14	*Bacillus subtilis*	(6/12, 50%)
Sequence 15	*Bacillus sp.*	(5/12, 41.67%)
Sequence 17	*Bacillus subtilis*	(8/12, 66,67%)
Sequence 18	*Bacillus sp.*	(5/12, 41.67%)
Sequence 20	*Bacillus subtilis*	(12/12, 100%)
Sequence 21	/	/
Sequence 27	*Bacillus subtilis*	(11/12, 91.67%)
Sequence 28	*Bacillus sp.*	(5/12, 41.67%)
Sequence 29	/	/
Sequence 30	*Lactobacillus plantarum*	(9/12, 75%)
Sequence 32	/	/
Sequence PL	/	/

*This score is calculated based on how many of the same Specie was found on all different tools and algorithms.
**Sequences that have "/" indicate they do not have enough data to be predicted.

Out of the 21 sequences, 11 (52.4%) were predicted with a confidence greater than the threshold. Sequences 6, 11, 13, 29 and 30 correctly predicted the bacteria, and we also assume that sequences 1, 3, 14, 17, 20 and 27 are correctly identified. Sequences 4, 15, 18 and 28 have a low percentage of identification, so we cannot confidently predict the species. The remaining sequences (5, 7, 21, 29, 32 and PL) could not be analysed by VSEARCH or USEARCH, and therefore are marked as none identified due to insufficient information.

4. Discussion

The aim of this project was to demonstrate the use of 16s-rRNA sequence identification for probiotics using the example of the supplement pills. Despite the presence of multiple bacterial strains, only a few of them could be isolated and characterized due to the absence of a biological safety level 2 laboratory, which is essential for the identification of pathogenic bacteria. Additionally, the anaerobic conditions required for the cultivation of Bifidobacterium could not be provided by the incubator used, thus further limiting the scope of the project.

The primary challenge in this project was the accurate separation of bacterial colonies. While colonies were obtained, they were randomly selected and used for further characterization, making it difficult to ascertain whether the chosen colonies were from different or the same bacteria, even when utilizing selective characterization media. This difficulty is shared by laboratories worldwide. After obtaining the DNA, the same was sent for sequencing to Turkey. Transportation of the DNA, which must be kept on ice, posed a challenge due to temperature fluctuations and transport vibrations which could potentially cause degradation of the sample. Fortunately, this did not occur and 21 sequences of either 350 or 500 bp amplicons were obtained. Initially, the whole 1500 bp gene was isolated and amplified, but due to a lack of resources, it was not sequenced. Consequently, the decision was made to focus only on the V3 – V4 region.

The obtained sequences were utilized for a comprehensive bioinformatic analysis, as detailed in Chapter 3: Results. This analysis revealed that the V3 and V4 hyper-variable regions of the 16s-rRNA can be employed for prediction of bacterial identification; however, the extent of accuracy is highly dependent on the sequence of the region, which is variable among different bacteria. For instance, *B. subtilis* has a high chance of being identified by the V3 – V4 region, yet bacteria such as *L. rhamnosus* often become confused with *L. plantarum* and *L. paraplantarum* due to the fact that the V3 and V4 regions of all three bacteria are more than 90% identical in sequence, rendering the software unable to recognize the difference. When this project was initially designed, the research conducted at the time suggested that the V3 – V3 regions were better for comparison than V1 – V2 regions [22]–[24]. The results obtained from this analysis indicate that the use of these regions is not recommended for the identification of bacteria. Additionally, the precision of these algorithms and the speed at which they can be run is worth noting. Algorithms with graphical interfaces are generally easier to use and are employed by biologists with limited programming knowledge. In such cases, users may find that the amount of time required to complete the task is contingent on the number of other users utilizing the software at the same time. However, local software can be used if there are any issues with the online software, as they do not require an internet connection so long as the relevant databases and sequences are locally stored. The disadvantage of these programs is that they are confined to the capabilities of the computer they are running on and will run more quickly on better machines.

Based on the results obtained and discussed, several recommendations can be made for future studies. Firstly, it is suggested that the overall size of the regions should be increased, either by taking three or more hyper-variable regions of 16s-rRNA, or even the entire gene, which was the original plan for this project. However, due to the issues encountered when performing PCR and the need to send samples for sequencing abroad, multiple repetitions of the experiment were not possible. For this type of study, it is necessary to have a sequencing device in the institution or country, thus enabling overnight sequencing and enabling mistakes to be avoided or corrected. Secondly, it is recommended to use more software and algorithms with a higher percentage of sequence available, combined with an increased working memory. The minimal amount of RAM should ideally be 32 GB to start this type of analysis, while 64 GB is preferable for smooth functioning. Additionally, it is suggested to use software such as FASTCAR, GASSS.T. and Genoogle.

5. Conclusion

This study assessed the growth and DNA extraction from *Lactobacillus, Bacillus, Lactococcus, Bifidobacterium,* and *Streptococcus* colonies in MRS medium. The mean value of DNA concentration was 17.17 ng/µl (standard deviation of 0.416) with a DNA purity of 1.75 (standard deviation of 0.026). Furthermore, BLASTn, RDP, USEARCH, and VSEARCH software were tested to analyze DNA sequences. BLASTn yielded a single result, with six organisms identified and ten incorrect predictions. RDP provided a match with an average time of 00:00:35,92. USEARCH was a 32-bit version with an average time of 00:01:05.83 per sequence. VSEARCH was the only tool to access all six databases, however, the average time for each database varied from 00:51.56 to 06:43.12. The findings demonstrated that VSEARCH successfully identified *B. subtilis* with the highest accuracy rate of 91.67%, followed by *L. plantarum* with an accuracy rate of 75%. The lowest accuracy rate was observed for *Bacillus sp*. at 41.67%. Subsequently, the potential of the V3 and V4 hyper-variable regions of the 16s-rRNA gene for bacterial identification was also investigated. Results suggested that the performance of the V3 – V4 regions in bacterial identification is highly dependent on the sequence of the region, which is variable among different bacteria. Moreover, the analysis revealed that while the V3 – V4 region of *B. subtilis* is highly likely to be successfully identified, other bacteria such as *L.*

rhamnosus may be confused with *L. plantarum* and *L. paraplantarum* due to the fact that the V3 and V4 regions of all three bacteria are more than 90% identical in sequence. As a consequence, the software was unable to make a distinction between them. Consequently, the results of this analysis provide evidence that the use of the V3 – V4 regions is not recommended for bacterial identification.

Declaration of competing interest

The authors declare that they have no / any known financial or non-financial competing interests in any material discussed in this paper.

Funding information

This research was funded by International University of Sarajevo, Bosnia and Herzegovina.

References and citations

[1] Z. Gu, Y. Liu, S. Hu, Y. You, W. Li, and Y. Wang, "Probiotics for Alleviating Alcoholic Liver Injury," *Gastroenterol. Res. Pract.*, vol. 2019, pp. 1–8, May 2019, doi: 10.1155/2019/9097276.

[2] X. Du *et al.*, "Efficacy of probiotic supplementary therapy for asthma, allergic rhinitis, and wheeze: a meta-analysis of randomized controlled trials," *Allergy Asthma Proc.*, vol. 40, no. 4, pp. 250–260, Jul. 2019, doi: 10.2500/aap.2019.40.4227.

[3] M. Hagbø *et al.*, "Experimental support for multidrug resistance transfer potential in the preterm infant gut microbiota," *Pediatr. Res.*, vol. 88, no. 1, pp. 57–65, Jul. 2020, doi: 10.1038/s41390-019-0491-8.

[4] A. Nicoletti, M. Pompili, A. Gasbarrini, and F. R. Ponziani, "Going with the gut: probiotics as a novel therapy for hepatocellular carcinoma," *HepatoBiliary Surg. Nutr.*, vol. 8, no. 3, pp. 295–297, Jun. 2019, doi: 10.21037/hbsn.2019.01.16.

[5] L. K. Sarao and M. Arora, "Probiotics, prebiotics, and microencapsulation: A review," *Crit. Rev. Food Sci. Nutr.*, vol. 57, no. 2, pp. 344–371, Jan. 2017, doi: 10.1080/10408398.2014.887055.

[6] V. Mezzasalma *et al.*, "A Randomized, Double-Blind, Placebo-Controlled Trial: The Efficacy of Multispecies Probiotic Supplementation in Alleviating Symptoms of Irritable Bowel Syndrome Associated with Constipation," *BioMed Res. Int.*, vol. 2016, pp. 1–10, 2016, doi: 10.1155/2016/4740907.

[7] Saavedra J.M.* and A. Tschernia, "Human studies with probiotics and prebiotics: clinical implications," *Br. J. Nutr.*, vol. 87, no. 6, pp. 241–246, May 2002, doi: 10.1079/BJNBJN/2002543.

[8] S. Doron and D. R. Snydman, "Risk and Safety of Probiotics," *Clin. Infect. Dis.*, vol. 60, no. suppl_2, pp. S129–S134, May 2015, doi: 10.1093/cid/civ085.

[9] P. Markowiak and K. Śliżewska, "Effects of Probiotics, Prebiotics, and Synbiotics on Human Health," *Nutrients*, vol. 9, no. 9, p. 1021, Sep. 2017, doi: 10.3390/nu9091021.

[10] C. A. Horwitch, H. A. Furseth, A. M. Larson, T. L. Jones, J. F. Olliffe, and D. H. Spach, "Lactobacillemia in Three Patients with AIDS," *Clin. Infect. Dis.*, vol. 21, no. 6, pp. 1460–1462, Dec. 1995, doi: 10.1093/clinids/21.6.1460.

[11] M. K. Salminen *et al.*, "*Lactobacillus* Bacteremia during a Rapid Increase in Probiotic Use of *Lactobacillus rhamnosus* GG in Finland," *Clin. Infect. Dis.*, vol. 35, no. 10, pp. 1155–1160, Nov. 2002, doi: 10.1086/342912.

[12] C. R. Woese, O. Kandler, and M. L. Wheelis, "Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.," *Proc. Natl. Acad. Sci.*, vol. 87, no. 12, pp. 4576–4579, Jun. 1990, doi: 10.1073/pnas.87.12.4576.

[13] "About PubMed by Year." https://esperr.github.io/pubmed-by-year/about.html (accessed Feb. 05, 2023).

[14] M.-T. Kuo, J.-L. Chen, S.-L. Hsu, A. Chen, and H.-L. You, "An Omics Approach to Diagnosing or Investigating Fungal Keratitis," *Int. J. Mol. Sci.*, vol. 20, no. 15, p. 3631, Jul. 2019, doi: 10.3390/ijms20153631.

[15] L. N. Ten *et al.*, "Mucilaginibacter terrigena sp. nov. sp., A Novel Member of the Family Sphingobacteriaceae," *Curr. Microbiol.*, vol. 76, no. 10, pp. 1152–1160, Oct. 2019, doi: 10.1007/s00284-019-01748-y.

[16] M. D. Rodriguez, K. K. Yu, Z. S. Paul, M. Keller-Wood, C. E. Wood, and E. W. Triplett, "Contamination Is Not Linked to the Gestational Microbiome," *Appl. Environ. Microbiol.*, vol. 85, no. 19, pp. e01127-19, Oct. 2019, doi: 10.1128/AEM.01127-19.

[17] M. Wei, P. Wang, C. Yang, and L. Gu, "Molecular identification and phylogenetic relationships of clinical Nocardia isolates," *Antonie Van Leeuwenhoek*, vol. 112, no. 12, pp. 1755–1766, Dec. 2019, doi: 10.1007/s10482-019-01296-2.

[18] S. A. Barghouthi, "A Universal Method for the Identification of Bacteria Based on General PCR Primers," *Indian J. Microbiol.*, vol. 51, no. 4, pp. 430–444, Oct. 2011, doi: 10.1007/s12088-011-0122-5.

[19] A. Klindworth *et al.*, "Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies," *Nucleic Acids Res.*, vol. 41, no. 1, pp. e1–e1, Jan. 2013, doi: 10.1093/nar/gks808.

[20] C. M. Liu *et al.*, "BactQuant: An enhanced broad-coverage bacterial quantitative real-time PCR assay," *BMC Microbiol.*, vol. 12, no. 1, p. 56, Dec. 2012, doi: 10.1186/1471-2180-12-56.

[21] A. U. Triwijayani, I. D. Puspita, Murwantoko, and Ustadi, "Identification of chitinolytic bacteria isolated from shrimp pond sediment and characterization of their chitinase encoding gene," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 139, p. 012051, Mar. 2018, doi: 10.1088/1755-1315/139/1/012051.

[22] S. Graspeuntner, N. Loeper, S. Künzel, J. F. Baines, and J. Rupp, "Selection of validated hypervariable regions is crucial in 16S-based microbiota studies of the female genital tract," *Sci. Rep.*, vol. 8, no. 1, p. 9678, Jun. 2018, doi: 10.1038/s41598-018-27757-8.

[23] Yu. S. Bukin, Yu. P. Galachyants, I. V. Morozov, S. V. Bukin, A. S. Zakharenko, and T. I. Zemskaya, "The effect of 16S rRNA region choice on bacterial community metabarcoding results," *Sci. Data*, vol. 6, no. 1, p. 190007, Feb. 2019, doi: 10.1038/sdata.2019.7.

[24] Z. Chen *et al.*, "Impact of Preservation Method and 16S rRNA Hypervariable Region on Gut Microbiota Profiling," *mSystems*, vol. 4, no. 1, pp. e00271-18, Feb. 2019, doi: 10.1128/mSystems.00271-18.