Avicenna J Clin Microb Infec. 2017 November; 4(4):e13082. doi: 10.5812/ajcmi.13082.

Published online 2017 August 31. **Research Article**

Exploration of Specific DNA-Barcodes in *Shigella dysenteriae* Using In-silico Analysis

Mehdi Kamali,[1] Behnam Bakhshi,[2,*] Ali Salimi,[1] Ehsan Mohseni Fard,[3] Mohammad Hasan Darvishi,[1] and Elahe Ehghaghi[4]

[1]Nanobiotechnology Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran
[2]Young Researchers and Elite Club, Science and Research Branch, Islamic Azad University, Tehran, Iran
[3]Department of Agronomy and Plant Breeding, Faculty of Agriculture, University of Zanjan, Zanjan, Iran
[4]Department of Virology, Tarbiat Modares University, Tehran, Iran

[*]*Corresponding author*: Behnam Bakhshi, Young Researchers and Elite Club, Science and Research Branch, Islamic Azad University, Tehran, Iran. E-mail: behnam.bakhshi@gmail.com

Received 2017 March 14; **Revised** 2017 May 13; **Accepted** 2017 July 17.

Abstract

Background: *Shigella dysenteriae* are Gram-negative and non-sporulating bacteria that cause illness in epithelial tissue of the colon and rectum. According to a preliminary analysis, rare or no reports could introduce highly reliable and specific genes, primers, and probes for *S. dysenteriae* recognition. Thus, it is necessary to detect specific genome parts in *S. dysenteriae* that could be used in diagnostic laboratories to recognize *S. dysenteriae* species confidently.
Methods: Identification of specific *S. dysenteriae* genome regions as DNA-barcodes was the main objective of the current study to accrue detection of this species. To this end, *S. dysenteriae* genome was compared with other *Enterobacteriaceae* genomes.
Results: Results indicated that there is little genetic distance between *S. dysenteriae* and *E. coli*, and most of the genes are common between these 2 species. The lowest genome fluidity was observed between *S. dysenteriae* and *Escherichia coli*, and *Salmonella enterica*. Furthermore, the largest number of orthologous genes was observed between *S. dysenteriae* and *E. coli* (O157_H7). All previous markers and virulent genes were also evaluated in the current study. However, no specific DNA barcodes were identified among already identified genes. Additionally, all regions of *S. dysenteriae* genome were investigated in the current study using specific region identifier programs by comparison with other *Enterobacteriaceae* strains.
Conclusions: Finally, eight specific DNA-barcodes were identified in the current study that could be beneficial for specific recognition of *S. dysenteriae* strains.

Keywords: *Shigella*, *E. coli*, *Enterobacteriaceae*, Specific Barcodes, Comparative Genomics

1. Background

Shigellosis causes over one million fatalities with more than 160 million patients with shigellosis. Most of these patients were under 5 years (1, 2). *Shigella* infection occurs through the mouth and intestines. Accumulation of 10 to 100 of these bacteria could cause shigellosis (3). *Shigella* is categorized to 4 groups, through biochemical and O antigen characteristics, including *S. dysenteriae* (group A), *Shigellaflexneri* (group B), *Shigella boydii* (group C), and *Shigella sonnei* (group D) (4-6). *Shigella* cells include a virulent plasmid that encodes genes that are necessary for attacking Intestinal mucosal cells (7). However, there is some pathogenicity islands in *Shigella* chromosomes that could play important roles in Pathogenicity (8). All *Shigella* strains include a large virulent plasmid with 180 to 215 kb of size, which is necessary for *Shigella* pathogenicity (9, 10).

Pandemic epidemic of *S. dysenteriae* in central America led to a total of 112000 cases and 10000 deaths in Guatemala from 1969 to 1972 (11, 12). The Sd197 strain of *S. dysenteriae* includes Gram-negative and non-sporulating bacteria that cause illness in epithelial tissue of the colon and rectum (13). The Sd197 strain of *S. dysenteriae* was observed in the epidemic of Guatemala in 1968 (12, 14, 15).

Shigella species originally formed from *Escherichia coli* about 3 500 to 270 000 years ago (16). *Escherichia coli* is the most important model organism in biological and medical studies. Many of the critical approaches, including bacterial conjugation, recombination, and genetic regulation are derived from studying *E. coli* (17). Over billions of *E. coli* cells are established in a healthy human gut (18), yet there are some *E. coli* that cause illness, including diarrhea, blood

infection, pneumonia and meningitis in humans or animals (19).

Large amounts of *S. dysenteriae* genome are very much similar with the *E. coli* genome (9, 10). There is a low number of genetic assays that could distinct *S. dysenteriae* from *E. coli* or other strains of *Shigella* because most of the genes included in *S. dysenteriae* could also be found in *E. coli* or other *Shigella* strains with similar sequence structure. Additionally, their virulent plasmid showed similar characteristics (20). In some reports, shiga toxin gene has been used for *S. dysenteriae* and *E. coli* recognition (21-23) because this gene is present in both *S. dysenteria* and *Escherichia coli*. Although, this method is beneficial for recognition of toxin existence, yet it could not recognize the exact species. Because of similarities among *S. dysenteriae* and *E. coli* and also other *Shigella* strains, especially in their virulent genes, most of the designed primers or probes for *S. dysenteriae* could also recognize other related species. As an example, *stx1*, *ipaBCD*, and *ipaH* are some genes that have been introduced for specific recognition of *Shigella* species, such as *S. dysenteriae* (24). However, the researcher's in-silico analysis (but not experimental) indicated that these genes and their primer might be used to recognize other related species such as *E. coli* or other *Shigella* species instead of *S. dysenteriae*. Similar results have been found by other studies. Thus, according to our preliminary analysis, rare or no reports could introduce highly reliable and specific genes, primers, and probes for *S. dysenteriae* recognition. Therefore, laboratory specialists could be misled in diagnostic tests when they are using common genes for specific recognition. Thus, it is necessary to detect specific genome parts in *S. dysenteriae* that could be used in diagnostic laboratories to recognize *S. dysenteriae* species confidently. Although both *Shigella* and *Escherichia* species are very much similar in large amounts of their genome, an extensive study of comparative genomics between these species should be done. In the current study, comparative genomics was hired to indicate the similarity among *S. dysenteriae*, *E. coli*, and other *Shigella* strains that led to the identification of specific genome areas of *S. dysenteriae* as specific DNA-barcodes.

2. Methods

2.1. Genome Sequences

Enterobacteriaceae strains genome sequences were downloaded from the genome list of NCBI database (https://www.ncbi.nlm.nih.gov/genome/browse). These genome sequences were used by comparative genomics used in the current study.

2.2. Genetic Distance and Similarity Computation

Genetic distance calculation of strains could lead us to the identification of close strains. In the current study, genetic distances between *S. dysenteriae* and other bacteria strains were calculated according to oligonucleotide frequency through an online tool available at http://insilico.ehu.eus.

The researchers visualized phylogenetic trees to indicate close species. The IMG software was used for creating genome clusters between *S. dysenteriae* and other bacteria species (https://img.jgi.doe.gov/cgi-bin/m/main.cgi). The Sd197 strain of *S. dysenteriae* focused mainly on computing and comparisons with other strains of species in the current study.

Dissimilarity and genome fluidity of Sd197 strain with other strains was also evaluated in the current study. Genome fluidity is a measure of dissimilarity among genes. It is obtained through the ratio of all unique genes (not shared) to all genes present in the 2 compared genomes (25). Genome fluidity was evaluated and compared using POGO-DB (26) for 70 conserved genes described in this study.

In order to confirm genome fluidity results of studied strains, orthologous genes between *S. dysenteriae* and other *Enterobacteriaceae* species were studied using OrthoVen (http://probes.pw.usda.gov/OrthoVenn/). To this end, 2 methods, including high sensitivity and low sensitivity methods were used. In the high sensitivity method the minimum percentage of similarity was considered as 30 % for 70% of aligned sequenced, yet in low sensitivity method the minimum percentage of similarity was considered as 10 % for 50% of aligned sequences.

2.3. Comparative Genomics Analysis

The SCAN2 program was used for multiple alignments of genome sequences instead of other alignment programs because of its specific ability for analysis of multi mega byte size genome sequences that could expedite the sequence alignment. The researchers used the SCAN2 program for multiple alignments of *S. dysenteriae* and *E. coli* strains used in the current study. This program is available at http://www.softberry.com/berry.phtml?topic=scan2&group=programs&subgroup=scanh.

Synteny LinePlot analysis and was used in the current study to create a graphical overview of conserved regions. This graphical visualization was carried out using the MicroScop program (http://www.genoscope.cns.fr/agc/microscope). MicroScop is a prokaryotic annotation system widely used by the microbiologist and it has been mostly used for synteny map visualization (27).

In addition, BioEdit, Mega, and Blastall programs have been used for comparing sequences and also for some comparative genomics studies (28, 29).

2.4. Comparative Genomics Analysis of Virulent Genes to Identify Specific Virulent DNA-Barcodes

In this study, identified virulent genes were considered to detect specific and conserved virulent genes. To this end, ShiBase (http://www.mgc.ac.cn/ShiBASE/) and VFDB (http://www.mgc.ac.cn/VFs/) databases were used in the current study. ShiBase and VFDB have been introduced as important databases for identifying virulent factors of bacterial strains (30, 31). The researchers used these databases to identify specific and conserved virulent genes among *S. dysenteriae*, *Shigella*, and *Escherichia* strains.

2.5. Exploration of New DNA-Barcodes

Identification of specific genes for sd197 strain of *S. dysenteriae* has been done by removing homologous genes with other *Shigella* and also *E. coli* strains, according to MICFAM clustering algorithm through pan genome analysis using the SiLiX software (32). The MICFAM parameter was considered as 80 in this analysis.

The researchers have also used PSAT analysis (33) to identify sd197 homologues genes with other *Shigella* and also *E. coli* strains. To this end, E-value < 10, bitscore > 20 and identity percentage > 10 were considered to identify homologous genes against all *Shigella* and Escherichia strains genome. These selected strict criteria could increase the range of homologous genes and on the other hand could decrease error in identification of non-homologous genes. This could lead to an increase in confidence in identification of sd197-specific DNA-Barcodes through PSAT analysis.

Additionally, the researchers have also used the nucmer program from MUMmer3 software (34) to identify specific regions in the sd197 genome. The researchers considered 500nt as minimum lengths of specific regions.

3. Results

3.1. Genetic Distance of S. dysenteriae from Other Enterobacteriaceae

Evaluation of the phylogenetic tree of *Enterobacteriaceae* indicated that *S. dysenteriae* is genetically close to *E. coli* in addition to other *Shigella* species (Figure 1). Thus, it is very likely that many of the genome regions between *S. dysenteriae* and *E. coli* are similar. Genetic distance of *S. dysenteriae* with other *Shigella* and *E. coli* strains is presented in Table 1. As shown in Table 1 *S. dysenteriae* is genetically very close to *E. coli*. Therefore, since the aim of this study is to recognize specific regions of *S. dysenteriae*, extensive comparative genomics has been performed to identify specific *S. dysenteriae* regions not common with all *E. coli* strains.

3.2. First Comparative Results

Shigella dysenteriae are identified by 2 important strains, including sd197 and 1617. In this study sd197 was selected for further studies. Genomes of sd197 strains in comparison with other *Shigella* species and also *E. coli* are presented in ShiBASE (http://www.mgc.ac.cn/ShiBASE/). Genome characteristics of *S. dysenteriae* were compared with other *Shigella* strains and also the sakai strain of *E. coli* O157:H7 using NCBI database. Results indicated that *S. dysenteriae* chromosome is smaller than other *Shigella* and *sakai* strains. In addition, comparison results indicated that all of these strains had one large plasmid and up to three small plasmids. *Shigella dysenteriae* includes one large plasmid (pSD1_197) and one small plasmid (pSD197_spA). It is important to note that more genes exist in *Shigella* plasmids when compared with the sakai strain of *E. coli* O157:H7. Additionally, more pseudo genes exist in *Shigella* species when compared with the sakai strain. A lower number of genes and on the other hand greater number of pseudo genes in *S. dysenteriae* indicated that fewer regions of this species genome could encode proteins compared to others. The Conserved Synteny LinePlot was used to show and overview existence of homologous regions between *S. dysenteriae* and other *Shigella* species and also *S. dysenteriae* and *E. coli* (Figure 2). As shown in Figure 2, large amounts of *S. dysenteriae* regions are conserved with other *Shigella* species and *E. coli*.

3.3. Similarity Evaluation of Shigella Dysenteriae with Other Enterobacteriaceae

As mentioned earlier, genome fluidity is a measure of dissimilarity among genes, which is the ratio of all unique genes (not shared) to all genes that exist in the two genomes (25). Higher genome fluidity indicates existence of more specific genes between 2 species and finally shows the difference between evaluated species (25). In this study, genome fluidity has been used to compare *S. dysenteriae* to other *Enterobacteriaceae* species. The lowest genome fluidity (less than 40%) was observed between *S. dysenteriae* and *E. coli* and also *S. dysenteriae* and *S. enterica*. This indicates high similarity among *S. dysenteriae* and the other two species. On the other hand, the highest genome fluidity (more than 80%) was observed between *S. dysenteriae* and *B. aphidicola* and also *S. dysenteriae* and *C. Moranella endobia*. Thus, these two species have the lowest common genes with *S. dysenteriae*.

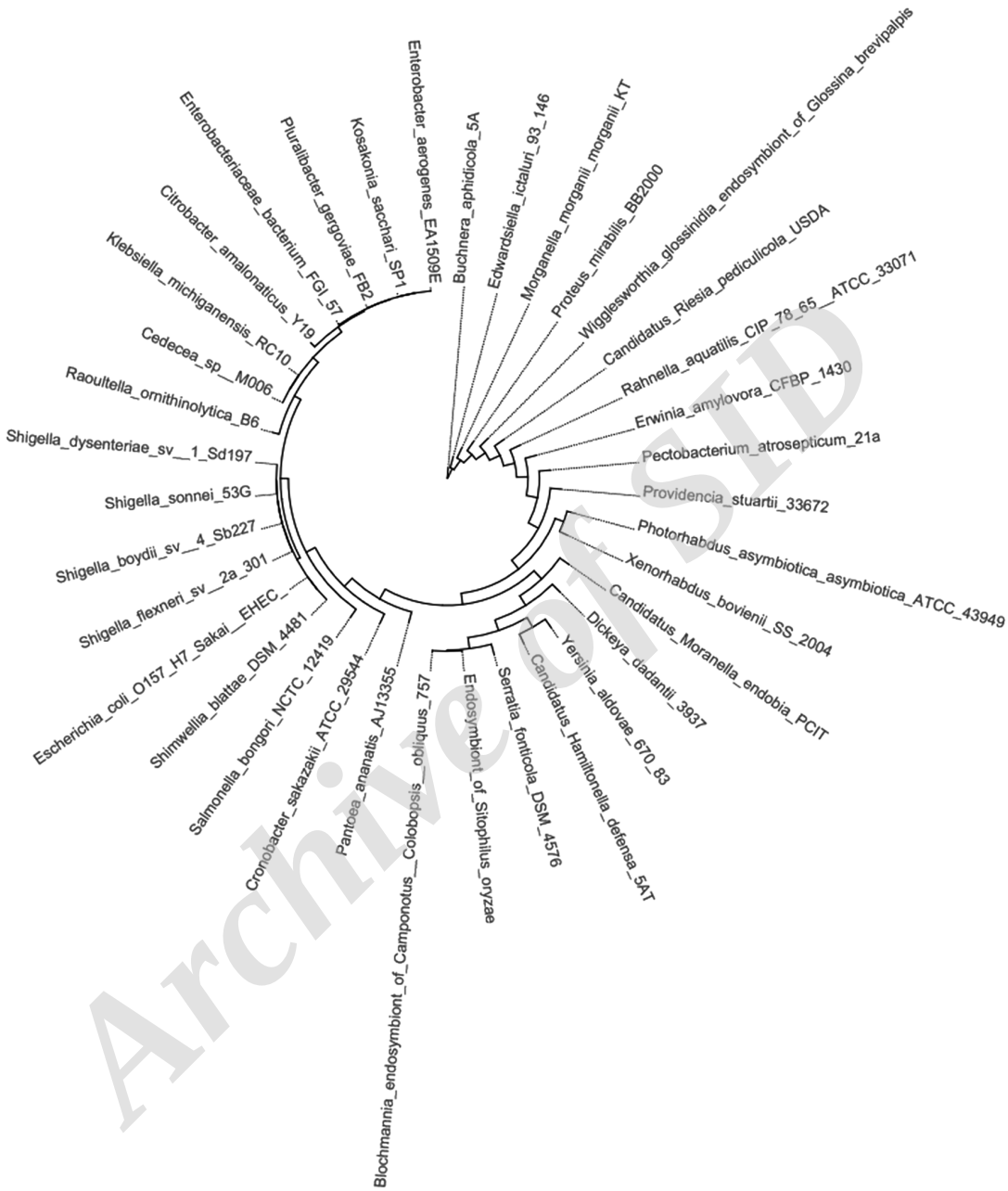


Figure 1. Phylogenetic Tree of *Enterobacteriaceae*

Furthermore, results of orthologous genes between *S. dysenteriae* and other *Enterobacteriaceae* species indicated that the lowest number of orthologous genes belonged to *S. dysenteriae* and *B. aphidicola* (high sensitivity: 321 number, low sensitivity: 351 number) and likewise *S. dysenteriae* and *C. Moranella endobia* (high sensitivity: 359, low sensitivity: 398). On the other hand, the largest number of orthologous genes belonged to *S. dysenteriae* with *E. coli* O157:H7 *en-*

Table 1. Genetic Distance of *Shigella dysenteriae* with Other *Shigella* and *Escherichia coli* Strains

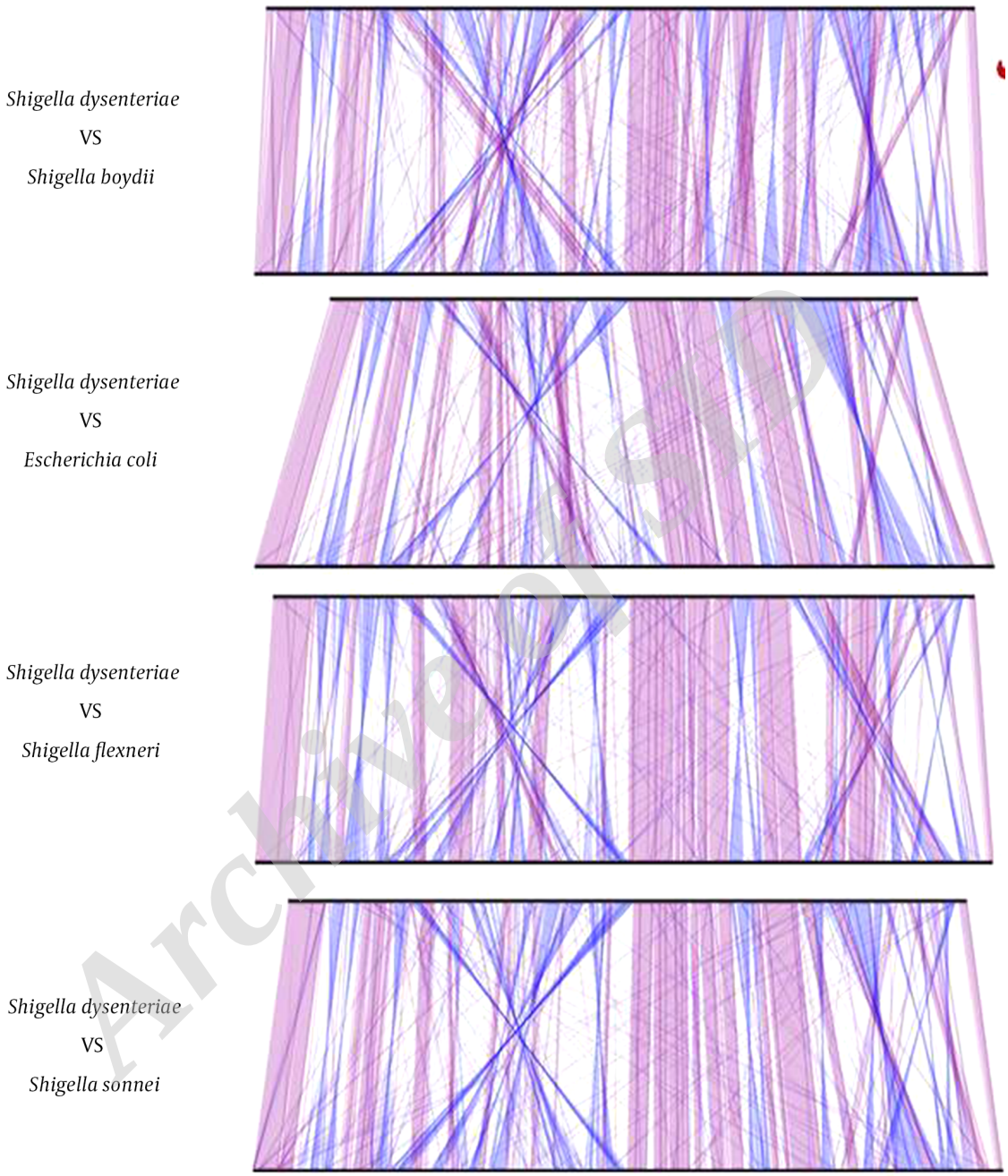
Row	Race Name	Genome Id	Genetic Distance	Row	Race Name	Genome Id	Genetic Distance	Row	Race Name	Genome Id	Genetic Distance
1	*S. dysenteriae* Sd197	NC_007606	0	28	*E. coli* O104:H4 str. 2009EL-2050	NC_018650	0.005460488	55	*E. coli* ABU 83972	NC_017631	0.006632288
2	*S. dysenteriae* 1617	NC_022912	0.000550692	29	*E. coli* str. K-12 substr. DH10B	NC_010473	0.005498053	56	*E. coli* O7:K1 str. CE10	NC_017646	0.006723061
3	*Shigella boydii* CDC 3083-94	NC_010658	0.001831203	30	*E. coli* SE11	NC_011415	0.005499051	57	*E. coli* str. clone D i14	NC_017652	0.006834802
4	*Shigella boydii* Sb227	NC_007613	0.001841915	31	*E. coli* str. K-12 substr. MDS42 DNA	NC_020518	0.005549004	58	*E. coli* str. clone D i2	NC_017651	0.006835676
5	*Shigella sonnei* Ss046	NC_007384	0.001901669	32	*E. coli* DH1	NC_017625	0.005561443	59	*E. coli* APEC O1	NC_008563	0.006878332
6	*Shigella sonnei* 53G	NC_016822	0.002198125	33	*E. coli* K-12 substr. W3110	NC_000091	0.00558117	60	*E. coli* 042	NC_017626	0.007084083
7	*Shigella flexneri* 2a str 301	NC_004337	0.002692293	34	*E. coli* str. K-12 substr. W3110	NC_007779	0.00558117	61	*E. coli* 536	NC_008253	0.007168004
8	*Shigella flexneri* 2a str. 2457T	NC_004741	0.002698727	35	*E. coli* PMV-1	NC_022370	0.005606612	62	*E. coli* O103:H2 str. 12009	NC_013353	0.007219182
9	*Shigella flexneri* 5 str. 8401	NC_008258	0.002698731	36	*E. coli* O104:H4 str. 2009EL-2071	NC_018661	0.005642387	63	*E. coli* O26:H11 str. 11368	NC_013361	0.00722457
10	*Shigella flexneri* 2002017	NC_017328	0.002920613	37	*E. coli* B str. REL606	NC_012967	0.005646166	64	*E. coli* O111:H- str. 11128	NC_013364	0.007441441
11	*E. coli* NA114	NC_017644	0.003198381	38	*E. coli* IHE3034	NC_017628	0.005654157	65	*E. coli* CFT073	NC_004431	0.007721105
12	*E. coli* P12b	NC_017663	0.00413917	39	*E. coli* S88	NC_011742	0.005662353	66	*E. coli* SMS-3-5	NC_010498	0.007789663
13	*E. coli* KO11FL	NC_017660	0.004926643	40	*E. coli* O104:H4 str. 2011C-3493	NC_018658	0.005677694	67	*E. coli* O55:H7 str. CB9615	NC_013941	0.007969988
14	*E. coli* JJ1886	NC_022648	0.004936043	41	*E. coli* str. K-12 substr. MG1655	NC_000913	0.005726779	68	*E. coli* Xuzhou21	NC_017906	0.008153557
15	*E. coli* ETEC H10407	NC_017633	0.004947672	42	*E. coli* 55989	NC_011748	0.005799974	69	*E. coli* O157:H7 str. Sakai	NC_002695	0.008279665
16	*E. coli* ATCC 8739	NC_010468	0.005055756	43	*E. coli* SE15	NC_013654	0.005824402	70	*E. coli* O55:H7 str. RM12579	NC_017656	0.008571297
17	*E. coli* HS	NC_009800	0.005101879	44	*E. coli* ED1a	NC_011745	0.005825414	71	*E. coli* O157:H7 str. TW14359	NC_013008	0.008576832
18	*E. coli* UMNK88	NC_017641	0.005168637	45	*E. coli* IAI1	NC_011741	0.005894765	72	*E. coli* O157:H7 str. EC4115	NC_011353	0.008618833
19	*E. coli* BL21-Gold	NC_012947	0.00517856	46	*E. coli* IAI39	NC_011750	0.005903618	73	*E. coli* O157:H7 EDL933	NC_002655	0.009314847
20	*E. coli* LY180	NC_022364	0.005252523	47	*E. coli* UMN026	NC_011751	0.006089112				
21	*E. coli* W	NC_017664	0.005278604	48	*E. coli* LF82	NC_011993	0.006106167				
22	*E. coli* KO11FL	NC_016902	0.005294661	49	*E. coli* E24377A	NC_009801	0.006195796				
23	*E. coli* BL21(DE3)	NC_012892	0.005295794	50	*E. coli* APEC O78	NC_020163	0.006262813				
24	*E. coli* BL21(DE3)	NC_012971	0.005296196	51	*E. coli* O83:H1 str. NRG 857C	NC_017634	0.006315047				
25	*E. coli* W	NC_017635	0.005305845	52	*E. coli* UTI89	NC_007946	0.006348379				
26	*E. coli* DH1	NC_017638	0.005435895	53	*E. coli* 0127:H6 E2348/69	NC_011601	0.006460836				
27	*E. coli* BW2952	NC_012759	0.005442963	54	*E. coli* UM146	NC_017632	0.006614888				

dobia (high sensitivity: 3112, low sensitivity: 3147 number). Therefore, results of orthologous comparisons confirmed genome fluidity results of the current study.

3.4. Could Already-Introduced Marker Genes in Bacteria Be Introduced as Suitable DNA-Barcodes for S. dysenteriae?

Different genes have already been used as marker genes for identification of bacteria. The 16s *rRNA* is one of the frequent used genes in these studies (35-38). One of the important reasons for selection of 16s *rRNA* in these studies is that 16s *rRNA* exists in all bacteria species with no variability in its gene structure (39). Blast results of *S. dysenteriae* 16s *rRNA* with other *Enterobacteriaceae* showed that this gene could be aligned with all bacteria, especially with *E. coli*. Furthermore, 97% to 99% similarity has been observed between all *E. coli* strains and *S. dysenteriae* for 16s *rRNA* gene in this study. The lowest similarity has been observed between *S. dysenteriae* and C. Riesia with 87% similarity. Thus, 16s *rRNA* could not act as an efficient marker for recognition of closely related species like *S. dysenteriae* and *E. coli*. Therefore, marker genes with higher distinctive features are necessary for *S. dysenteriae* recognition. To this end, other introduced markers for recognition of different bacteria species were evaluated in this study including *amoA, pmoA, nirS, nirK, nosZ*, and *pufM* (35, 40, 41) and also *dnaG, frr, infC, nusA, pgk, pyrG, rplA, rplB, rplC, rplD, rplE, rplF, rplK, rplL, rplM, rplN, rplP, rplS, rplT, rpmA, rpoB, rpsB, rpsC, rpsE, rpsI, rpsJ, rpsK, rpsM, rpsS, smpB* and *tsf* (42). The current inves-

Figure 2. Synteny LinePlot Among *Shigella* and *Escherichia coli* Species



Red line indicated homologous regions and blue line indicated homologous and reversed regions, respectively.

tigation indicated that all of these genes could be aligned with the *S. dysenteriae* genome. In all cases, more than 50% similarity was observed between *S. dysenteriae* and other *Enterobacteriaceae* for these marker genes. However, more than 90% similarity was observed between *S. dysenteriae* and all *E. coli* strains for all the mentioned genes. There-

fore, it could be concluded that these already-introduced marker genes could not be useful for *S. dysenteriae* recognition.

3.5. Could Already-Introduced Virulent Genes in Bacteria Be Introduced as Suitable DNA-Barcodes for S. dysenteriae?

The use of virulent genes has been introduced as one of effective ways for recognition of bacteria. As an example, the *stx* gene has been used for *S. dysenteriae* recognition (21-23). In this context, all virulent *S. dysenteriae* genes were compared with other *Shigella* and also with *E. coli* strains to identify their conservation in this study (Figure 3). Results indicated that all virulent genes of *S. dysenteriae* could be aligned with other *Shigella* or *E. coli* strains using NCBI, VFDB, and ShiBASE. Thus, these virulent genes could not be introduced as suitable DNA-barcodes for *S. dysenteriae*. Therefore, these virulent genes are beneficial for recognition of multiple bacteria recognition and also virulent factors instead of specific diagnosis of bacteria like *S. dysenteriae*. As shown in Figure 3, the *stx* gene is conserved in both *S. dysenteriae* and *E. coli* strains and plays a role in shiga toxin production. Thus, these virulent genes are beneficial for recognition of shiga toxin existence. Conservation of other virulent genes is presented in Figure 3.

3.6. Investigation of Other Genome Areas of S. dysenteriae to Identify DNA-Barcodes

Using the MUMmer3 program, six specific regions were identified in the current study that could be used as specific DNA-barcodes for *S. dysenteriae* recognition, including NC_009344.1 (2791.7017 in plasmid pSD197_spA), NC_007606.1 (3886090.3886726 in complete genome), NC_007606.1 (3769230.3770547 in complete genome), NC_007606.1 (3859088.3859910 in complete genome), NC_007606.1 (2329346.2331791 in complete genome), and NC_007606.1 (1082613.1083293 in complete genome) regions. In addition, using PSAT, specific DNA-barcodes have been found for large and small plasmids, including NC_009344.1 (5327.6079 in plasmid pSD197_spA near rfp genes that was also detected by MUMmer3 but in a larger space) and NC_007606.1 (162252.162452 exist in plasmid pSD1_197 between *virA* and *spa* genes). Designed forward and reverse primers in the current study are presented in Table 2. These primers could be beneficial for further studies.

Table 2. Designed Forward and Reverse Primers for Identified DNA-Barcodes

DNA-Barcodes	Forward Primer	Reverse Primer
NC_007606.1 (162252.162452)	ATTAAACCGGGGTGCCTCA	GCCTCTCGAGACGTGAAATC
NC_007606.1 (3886090.38867	GCGTAACCACCAATCCAGTT	TGCAATATTTCCAGCAGGTG
NC_007606.1 (3769230.3770547)	GGGGACACCAGCAGTACCTA	CGGTGGAGAAATCGTCATCT
NC_007606.1 (3859088.385991	CTTCTGCCAGAGCATCTTCC	CTGATTAGCGTGATACCGCA
NC_007606.1 (2329346.2331791)	TTGACCAGCAACTTCCAGTG	CTTGCTGGCTGGCTTATTTC
NC_007606.1 (1082613.108329	TGGTTTCAGCCAATGTTTCA	TGCGATTGCATTTGCTAAGA
NC_009344.1 (2791.7017)	CCATGTGGCTGCTCTGTAAA	GCGCCATTCCTGTTGATTAT
NC_009344.1 (5327.6079)	TGCCAACAACCTTAGCTGTG	CAAGTGACCCAAATGTGTTAGC

4. Discussion

The current results indicated that specific regions of *S. dysenteriae* that might have evolved recently are appropriate for DNA-barcodes detection rather than slowly evolved genes, such as 16s *rRNA* or *stx*. Comparative genomic studies have helped with identification of specific regions in the *S. dysenteriae* genome from recently evolved genes. However, slowly evolved genes could be helpful for multiple bacteria recognition or identification of virulent factors when certain strains or species are not considered. Lack of attention to these notes could lead to mistakes in *S. dysenteriae* recognition. To overcome this event, selection of correct genes form this species is an essential step. Finally, in this study, 8 specific DNA-barcodes for recognition of *S. dysenteriae* were identified. These DNA-barcodes could be useful for designing primers and probes to identify *S. dysenteriae*.

References

1. Kotloff KL, Winickoff JP, Ivanoff B, Clemens JD, Swerdlow DL, Sansonetti PJ, et al. Global burden of Shigella infections: implications for vaccine development and implementation of control strategies. *Bull World Health Organ.* 1999;**77**(8):651–66. [PubMed: 10516787].
2. Pothoulakis C, Lamont JT. Microbes and microbial toxins: paradigms for microbial-mucosal interactions II. The integrated response of the intestine to Clostridium difficile toxins. *Am J Physiol Gastrointest Liver Physiol.* 2001;**280**(2):G178–83. [PubMed: 11208538].
3. DuPont HL, Levine MM, Hornick RB, Formal SB. Inoculum size in shigellosis and implications for expected mode of transmission. *J Infect Dis.* 1989;**159**(6):1126–8. [PubMed: 2656880].
4. Li Y, Cao B, Liu B, Liu D, Gao Q, Peng X, et al. Molecular detection of all 34 distinct O-antigen forms of Shigella. *J Med Microbiol.* 2009;**58**(Pt 1):69–81. doi: 10.1099/jmm.0.000794-0. [PubMed: 19074655].
5. Martinez-Becerra FJ, Kissmann JM, Diaz-McNair J, Choudhari SP, Quick AM, Mellado-Sanchez G, et al. Broadly protective Shigella vaccine based on type III secretion apparatus proteins. *Infect Immun.* 2012;**80**(3):1222–31. doi: 10.1128/IAI.06174-11. [PubMed: 22202122].

Groups	Virulence factors	Related genes	S. dysenteriae	S. boydii	S. flexneri	S. sonnei	E. coli
Secretion system	Mxi-Spa TTSA (type III secretion apparatus)	spa32	+	+	+	+	-
		spa33	+	+	+	+	-
		spa24	+	+	+	+	-
		spa9	+	+	+	+	-
		spa29	+	+	+	+	-
		spa40	+	+	+	+	-
		ipgC	+	+	+	+	-
		ipgA	+	+	+	+	-
		ipgE	+	+	+	+	-
		ipgF	+	+	+	+	-
	Mxi-Spa TTSS effectors controlled by MxiE	ospD3	+	+	+	+	-
		ospE1	+	+	+	+	-
		ospE2	-	+	+	+	-
		ospG	+	+	+	+	-
		ipaH1.4	+	+	+	+	-
		ipaH2.5	+	+	+	-	-
		ipaH4.5	+	+	+	+	-
		ipaH7.8	+	+	+	+	-
		ipaH9.8	+	+	+	+	-
		ipaH	+	+	+	+	-
	Mxi-Spa TTSS effectors controlled by VirB	ipaA	+	+	+	+	-
		ipaD	+	+	+	+	-
		ipaC	+	+	+	+	-
		ipaB	+	+	+	+	-
		ipgB1	+	+	+	+	-
		ipgB2	+	+	+	+	-
		ipgD	+	+	+	+	-
		icsB	+	+	+	+	-
		ospC2	+	+	+	+	-
		ospC3	+	+	+	+	-
		ospC4	+	+	+	+	-
		ospD1	+	+	+	+	-
		ospD2	+	+	+	+	-
	Mxi-Spa TTSS effectors controlled by	virA	+	+	+	+	-
		ospB	+	+	+	+	-
		ospC1	+	+	+	+	-
		ospF	+	+	+	+	-
	T2SS (Type II secretion system)	gspC	+	+	-	-	-
		gspD	+	+	-	-	-
		gspE	+	+	-	-	-
		gspF	+	+	-	-	-
		gspG	+	+	-	-	-
		gspH	+	+	-	-	-
		gspI	+	+	-	-	-
		gspJ	+	+	-	-	-
		gspK	+	+	-	-	-
		gspL	+	+	-	-	-
		gspM	+	+	-	-	-
Toxin	Enterotoxin	set1A	-	-	+	-	+
	ShET-1	set1B	-	-	+	-	+
	Shiga toxin	stxA	+	-	-	-	+
		stxB	+	-	-	-	+
Others	IcsA (VirG)	icsA/virG	+	+	+	+	-
	MsbB2	msbB2	+	+	+	+	-
	VirF	virF	-	+	+	+	-
	VirK	virK	+	+	+	+	-

Groups	Virulence factors	Related genes	S. dysenteriae	S. boydii	S. flexneri	S. sonnei	E. coli
Host immune evasion	LPS glucosylation	gtrA	-	-	+	-	-
		gtrB	-	-	+	-	-
		gtr	-	-	+	-	-
Iron uptake	Aerobactin synthesis	iucA	-	+	+	+	+
		iucB	-	+	+	+	+
		iucC	-	+	+	+	+
		iucD	-	+	+	+	+
	Aerobactin	iutA	-	+	+	+	+
	Enterobactin synthesis	entA	+	+	+	+	-
		entB	+	+	+	+	-
		entE	+	+	+	+	-
		entC	+	+	+	+	-
		entF	+	+	+	+	-
		entD	+	+	+	+	-
	Enterobactin transport	fepA	+	+	+	+	-
		fepB	+	+	+	+	-
		fepC	+	+	+	+	-
		fepD	+	+	+	+	-
		fepG	+	+	+	+	-
	Ferrous iron transport	sitA	+	+	+	+	+
		sitB	+	+	+	+	+
		sitC	+	+	+	+	+
		sitD	+	+	+	+	+
	Heme transport	shuS	+	-	-	-	-
		shuA	+	-	-	-	-
		shuT	+	-	-	-	-
		shuW	+	-	-	-	-
		shuX	+	-	-	-	-
		shuY	+	-	-	-	-
		shuU	+	-	-	-	-
		shuV	+	-	-	-	-
	Salmochelin synthesis and transport	iroN	+	-	-	-	+
		iroE	+	-	-	-	+
		iroD	+	-	-	-	+
		iroB	+	-	-	-	+
		iroC	+	-	-	-	+
Protease	IcsP (SopA)	icsP/sopA	+	+	+	+	-
	Pic	pic	-	-	+	-	+
	Serine	sepA	-	+	+	-	-
	SigA	sigA	-	+	+	+	-
Secretion system	Mxi-Spa TTSA (type III secretion apparatus)	virB	+	+	+	+	-
		mxiG	+	+	+	+	-
		mxiH	+	+	+	+	-
		mxiI	+	+	+	+	-
		mxiJ	+	+	+	+	-
		mxiK	+	+	+	+	-
		mxiN	+	+	+	+	-
		mxiL	+	+	+	+	-
		mxiM	+	+	+	+	-
		mxiE	+	+	+	+	-
		mxiD	+	+	+	+	-
		mxiC	+	+	+	+	-
		mxiA	+	+	+	+	-
		spa15	+	+	+	+	-
		spa47	+	+	+	+	-
		spa13	+	+	+	+	-

Figure 3. Conservation of *Shigella dysenteriae* Virulent Genes in Comparison with Other *Shigella* Species and also *Escherichia coli*

6. Torres AG. Current aspects of Shigella pathogenesis. *Rev Latinoam Microbiol.* 2004;**46**(3-4):89–97. [PubMed: 17061529].
7. Sansonetti PJ, Kopecko DJ, Formal SB. Involvement of a plasmid in the invasive ability of Shigella flexneri. *Infect Immun.* 1982;**35**(3):852–60. [PubMed: 6279518].
8. Wei J, Goldberg MB, Burland V, Venkatesan MM, Deng W, Fournier G, et al. Complete genome sequence and comparative genomics of Shigella flexneri serotype 2a strain 2457T. *Infect Immun.* 2003;**71**(5):2775–86. [PubMed: 12704152].
9. Jin Q, Yuan Z, Xu J, Wang Y, Shen Y, Lu W, et al. Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157. *Nucleic Acids Res.* 2002;**30**(20):4432–41. [PubMed: 12384590].

10. Maurelli AT, Fernandez RE, Bloch CA, Rode CK, Fasano A. "Black holes" and bacterial pathogenicity: a large genomic deletion that enhances the virulence of Shigella spp. and enteroinvasive Escherichia coli. *Proc Natl Acad Sci U S A.* 1998;**95**(7):3943–8. [PubMed: 9520472].
11. Gangarosa EJ, Perera DR, Mata LJ, Mendizabal-Morris C, Guzman G, Reller LB. Epidemic Shiga bacillus dysentery in Central America. II. Epidemiologic studies in 1969. *J Infect Dis.* 1970;**122**(3):181–90. [PubMed: 4915511].
12. Mendizabal-Morris CA, Mata LJ, Gangarosa EJ, Guzman G. Epidemic Shiga-bacillus dysentery in Central America. Derivation of the epidemic and its progression in Guatemala, 1968-69. *Am J Trop Med Hyg.* 1971;**20**(6):927–33. [PubMed: 4943477].
13. Nie H, Yang F, Zhang X, Yang J, Chen L, Wang J, et al. Complete genome sequence of Shigella flexneri 5b and comparison with Shigella flexneri 2a. *BMC Genomics.* 2006;**7**:173. doi: 10.1186/1471-2164-7-173. [PubMed: 16822325].
14. Mata LJ, Gangarosa EJ, Caceres A, Perera DR, Mejicanos ML. Epidemic Shiga bacillus dysentery in Central America. I. Etiologic investigations in Guatemala, 1969. *J Infect Dis.* 1970;**122**(3):170–80. [PubMed: 4915510].
15. Venkatesan MM, Hartman AB, Newland JW, Ivanova VS, Hale TL, McDonough M, et al. Construction, characterization, and animal testing of WRSd1, a Shigella dysenteriae 1 vaccine. *Infect Immun.* 2002;**70**(6):2950–8. [PubMed: 12010984].
16. Pupo GM, Lan R, Reeves PR. Multiple independent origins of Shigella clones of Escherichia coli and convergent evolution of many of their characteristics. *Proc Natl Acad Sci U S A.* 2000;**97**(19):10567–72. doi: 10.1073/pnas.180094797. [PubMed: 10954745].
17. Zhang Y, Lin K. A phylogenomic analysis of Escherichia coli / Shigella group: implications of genomic features associated with pathogenicity and ecological adaptation. *BMC Evol Biol.* 2012;**12**:174. doi: 10.1186/1471-2148-12-174. [PubMed: 22958895].
18. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, et al. Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. *PLoS Genet.* 2009;**5**(1):e1000344. doi: 10.1371/journal.pgen.1000344. [PubMed: 19165319].
19. Donnenberg MS. Escherichia coli: virulence mechanisms of a versatile pathogen. New York: Academic Press; 2002.
20. Lan R, Lumb B, Ryan D, Reeves PR. Molecular evolution of large virulence plasmid in Shigella clones and enteroinvasive Escherichia coli. *Infect Immun.* 2001;**69**(10):6303–9. doi: 10.1128/IAI.69.10.6303-6309.2001. [PubMed: 11553574].
21. Amani J, Ahmadpour A, Imani Fooladi AA, Nazarian S. Detection of E. coli O157:H7 and Shigella dysenteriae toxins in clinical samples by PCR-ELISA. *Braz J Infect Dis.* 2015;**19**(3):278–84. doi: 10.1016/j.bjid.2015.02.008. [PubMed: 25911087].
22. Binet R, Deer DM, Uhlfelder SJ. Rapid detection of Shigella and enteroinvasive Escherichia coli in produce enrichments by a conventional multiplex PCR assay. *Food Microbiol.* 2014;**40**:48–54. doi: 10.1016/j.fm.2013.12.001. [PubMed: 24549197].
23. Newland JW, Neill RJ. DNA probes for Shiga-like toxins I and II and for toxin-converting bacteriophages. *J Clin Microbiol.* 1988;**26**(7):1292–7. [PubMed: 2842369].
24. Faruque SM, Khan R, Kamruzzaman M, Yamasaki S, Ahmad QS, Azim T, et al. Isolation of Shigella dysenteriae type 1 and S. flexneri strains from surface waters in Bangladesh: comparative molecular analysis of environmental Shigella isolates versus clinical strains. *Appl Environ Microbiol.* 2002;**68**(8):3908–13. [PubMed: 12147489].
25. Kislyuk AO, Haegeman B, Bergman NH, Weitz JS. Genomic fluidity: an integrative view of gene diversity within microbial populations. *BMC Genomics.* 2011;**12**:32. doi: 10.1186/1471-2164-12-32. [PubMed: 21232151].
26. Lan Y, Morrison JC, Hershberg R, Rosen GL. POGO-DB–a database of pairwise-comparisons of genomes and conserved orthologous genes. *Nucleic Acids Res.* 2014;**42**(Database issue):D625–32. doi: 10.1093/nar/gkt1094. [PubMed: 24198250].
27. Vallenet D, Belda E, Calteau A, Cruveiller S, Engelen S, Lajus A, et al. MicroScope–an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Res.* 2013;**41**(Database issue):D636–47. doi: 10.1093/nar/gks1194. [PubMed: 23193269].
28. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series. *Nucleic Acids Symp Ser.* 1999;**41**(2):95–8.
29. Tao T. Program Parameters for blastall. *View Article PubMed/NCBI Google Scholar.* 2006.
30. Yang J, Chen L, Yu J, Sun L, Jin Q. ShiBASE: an integrated database for comparative genomics of Shigella. *Nucleic Acids Res.* 2006;**34**(Database issue):D398–401. doi: 10.1093/nar/gkj033. [PubMed: 16381896].
31. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* 2005;**33**(Database issue):D325–8. doi: 10.1093/nar/gki008. [PubMed: 15608208].
32. Miele V, Penel S, Duret L. Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC Bioinformatics.* 2011;**12**:116. doi: 10.1186/1471-2105-12-116. [PubMed: 21513511].
33. Fong C, Rohmer L, Radey M, Wasnick M, Brittnacher MJ. PSAT: a web tool to compare genomic neighborhoods of multiple prokaryotic genomes. *BMC Bioinformatics.* 2008;**9**:170. doi: 10.1186/1471-2105-9-170. [PubMed: 18366802].
34. Delcher AL, Phillippy A, Carlton J, Salzberg SL. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* 2002;**30**(11):2478–83. [PubMed: 12034836].
35. Case RJ, Boucher Y, Dahllof I, Holmstrom C, Doolittle WF, Kjelleberg S. Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. *Appl Environ Microbiol.* 2007;**73**(1):278–88. doi: 10.1128/AEM.01177-06. [PubMed: 17071787].
36. Gevers D, Knight R, Petrosino JF, Huang K, McGuire AL, Birren BW, et al. The Human Microbiome Project: a community resource for the healthy human microbiome. *PLoS Biol.* 2012;**10**(8):e1001377. doi: 10.1371/journal.pbio.1001377. [PubMed: 22904687].
37. Gilbert JA, Meyer F, Jansson J, Gordon J, Pace N, Tiedje J, et al. The Earth Microbiome Project: Meeting report of the "1 EMP meeting on sample selection and acquisition" at Argonne National Laboratory October 6 2010. *Stand Genomic Sci.* 2010;**3**(3):249–53. doi: 10.4056/aigs.1443528. [PubMed: 21304728].
38. Shade A, Caporaso JG, Handelsman J, Knight R, Fierer N. A meta-analysis of changes in bacterial and archaeal communities with time. *ISME J.* 2013;**7**(8):1493–506. doi: 10.1038/ismej.2013.54. [PubMed: 23575374].
39. Pei AY, Oberdorf WE, Nossa CW, Agarwal A, Chokshi P, Gerz EA, et al. Diversity of 16S rRNA genes within individual prokaryotic genomes. *Appl Environ Microbiol.* 2010;**76**(12):3886–97. doi: 10.1128/AEM.02953-09. [PubMed: 20418441].
40. Achenbach LA, Carey J, Madigan MT. Photosynthetic and phylogenetic primers for detection of anoxygenic phototrophs in natural environments. *Appl Environ Microbiol.* 2001;**67**(7):2922–6. doi: 10.1128/AEM.67.7.2922-2926.2001. [PubMed: 11425703].
41. Walsh DA, Bapteste E, Kamekura M, Doolittle WF. Evolution of the RNA polymerase B' subunit gene (rpoB') in Halobacteriales: a complementary molecular marker to the SSU rRNA gene. *Mol Biol Evol.* 2004;**21**(12):2340–51. doi: 10.1093/molbev/msh248. [PubMed: 15356285].
42. Wu M, Eisen JA. A simple, fast, and accurate method of phylogenomic inference. *Genome Biol.* 2008;**9**(10):R151. doi: 10.1186/gb-2008-9-10-r151. [PubMed: 18851752].