Heliyon

ELSEVIER

Received:
17 May 2017
Revised:
11 July 2017
Accepted:
21 July 2017

Cite as:
Marcel Martínez-Porchas, Francisco Vargas-Albores. An efficient strategy using *k*-mers to analyse 16S rRNA sequences.
Heliyon 3 (2017) e00370.
doi: 10.1016/j.heliyon.2017.e00370

An efficient strategy using *k*-mers to analyse 16S rRNA sequences

Marcel Martínez-Porchas, Francisco Vargas-Albores *

Centro de Investigación en Alimentación y Desarrollo, A. C. Km 0.6 Carretera a La Victoria. Hermosillo, Sonora, México

* Corresponding author.
E-mail address: fvalbores@ciad.mx (F. Vargas-Albores).

Abstract

The use of *k*-mers has been a successful strategy for improving metagenomics studies, including taxonomic classifications, or *de novo* assemblies, and can be used to obtain sequences of interest from the available databases. The aim of this manuscript was to propose a simple but efficient strategy to generate *k*-mers and to use them to obtain and analyse *in silico* 16S rRNA sequence fragments. A total of 513,309 bacterial sequences contained in the SILVA database were considered for the study, and homemade PHP scripts were used to search for specific nucleotide chains, recover fragments of bacterial sequences, make calculations and organize information. Consensus sequences matching conserved regions were constructed by aligning most of the primers used in the literature. Sequences of *k* nucleotides (9- to 15-mers) were extracted from the generated primer contigs. Frequency analysis revealed that *k*-mer size was inversely proportional to the occurrence of *k*-mers in the different conserved regions, suggesting a stringency relationship; high numbers of duplicate reactions were observed with short *k*-mers, and a lower proportion of sequences were obtained with large ones, with the best results obtained using 12-mers. Using 12-mers with the proposed method to obtain and study sequences was found to be a reliable approach for the analysis of 16S rRNA sequences and this strategy may probably be extended to other biomarkers. Furthermore, additional applications such as evaluating the degree of conservation and designing primers and other calculations are proposed as examples.

http://dx.doi.org/10.1016/j.heliyon.2017.e00370

Keywords: Bioinformatics, Microbiology, Biological sciences

1. Introduction

The study of 16S ribosomal RNA (rRNA) sequences has become popular among microbiologists due to the need to study the diversity and structure of microbiomes thriving in specific ecosystems, including those as small as phycosphere [1, 2]. The number of genomic descriptions has greatly increased due to new sequencing technologies and tools for the analysis of metagenomics data [3]. This improved sequencing throughput has allowed for statistically robust quantitative comparisons between communities. The 16S rRNA is a core component of the 30S small subunit of prokaryotes that is currently used for phylogeny building because of its slow rate of evolution [4]. The molecule contains ten conserved (C) regions that are separated by variable (V) regions. The C regions have been used to design and anchor primers for polymerase chain reaction (PCR) amplification, whereas V regions have been useful for taxonomic identification [5].

Several sets of primers have been reported and used for amplifying specific V regions of 16S rRNA; specifically, the V3, V4 and V5 regions have been widely used for studies where taxonomic classification or understanding phylogenic relationships is required [6, 7, 8, 9]. However, no single region can differentiate among all bacteria and therefore the remaining regions have also been used for the same purpose but are preferentially used for studying particular communities. For example, V1 has been demonstrated to be particularly useful for differentiating among species in the genus Staphylococcus [10]. Moreover, Shakya, et al. [11] performed a comparative metagenomics microbial diversity characterization using synthetic communities and reported that all tested primers sets lead to significant taxon-specific biases; not to mention the new and rare sequences deposited daily that do not match with conventional primers. Therefore, it is clear that sometimes in silico analyses may not fully correspond to biological trails, particularly when a specimen whose genome or 16S rRNA sequences has not been uploaded in the databases, and it is considerable abundant in a given niche.

Primer sets are usually evaluated by performing PCR on well-characterized samples, often with knowledge of the size and number of the expected products. However, in many circumstances, random environmental samples are used and a positive reaction for most variants cannot be guaranteed for the primers used. In these cases, it is possible to perform virtual PCR on a set of known or reference sequences, with the advantage of avoiding inherent problems of PCR for environmental samples, such as PCR inhibitors, cation requirements or physicochemical properties of primers. This virtual approach would serve to improve coverage when using real biological samples.

http://dx.doi.org/10.1016/j.heliyon.2017.e00370

The use of k-mers has been a successful strategy for improving metagenomics studies [12, 13, 14, 15, 16], including taxonomic classifications [17] and de novo assemblies [18], and can be used to get other sequences of interest from available databases. The k-mer length should be adjusted according to the requirements, seeking an appropriate balance between short k-mers that may exhibit low specificity and long k-mers that may increase stringency and exclude a considerable proportion of sequences. The aim of this manuscript was to propose a simple but efficient strategy to generate k-mers and use them to obtain and analyse in silico 16S rRNA sequences fragments.

2. Materials and methods

The 513,309 non-redundant bacterial sequences contained in the high quality ribosomal RNA database SILVA (release 123) were used for the study. Homemade PHP scripts were used for searching specific nucleotide chains, recovering fragments of bacterial sequences, making calculations and organizing information.

2.1. Primer contigs and generation of *k*-mers

Prior to contig generation, 214 primers used for the amplification of the different fragments of the bacterial 16S rRNA gene were found in literature; however, only 101 primers contributing with an increase of the amplicon size or containing a different nucleotide were selected for the study. These primers were assembled to generate a continuous "primer contig" sequence to perform a position-by-position sequence-scan analysis of regions (Table 1). Specifically, we tested if the continuity of contigs was interrupted by a 1 nucleotide (or more) gap and if each segment was considered as a sub-contig (a, b or c). Thereafter, sequences of k nucleotides (9 to 15) were extracted from the generated primer contigs.

The number of k-mers of each size was calculated as consensus size−k + 1; where k was the k-mer size (9 to 15 in this case). If degeneracies were detected in any k-mer, each isoform was considered for the analysis; for instance, the nucleotide ambiguity code establishes that keto or K represents T or G, therefore the sequence containing this degeneracy was multiplied by two possibilities. This was applied to all types of degeneracies detected in all sequences; for instance, Y, M, S, R, W = 2 possibilities, V, H, B, D = 3, and N = 4. Thus, the primer contig sequences for each region were broken down into 9-, 10-, 11- . . . 15-mers with their respective isoforms replacing any degeneracy with the corresponding nucleotides. Finally, the exact sequence of each k-mer isoform generated from all conserved regions was queried against the entire set of sequences contained in the SILVA database (Fig. 1).

http://dx.doi.org/10.1016/j.heliyon.2017.e00370

Table 1. Primer contigs generated by the assembly of all of the primers reported for each conserved region of the 16S rRNA gene. Locations are based on *E. coli* sequence.

Name	Sequence	Location	References
1	AGAGTTTGATYMTGGCTCAG	8-27	[29, 30, 31, 32, 33, 34, 35, 36, 37]
2	ASYGGCGNACGGGTGAGTAA	100-119	[38, 39]
3	ACTGAGAYACGGYCCARACTCCTACGGRNGGCNGCAGTRRGGAA	320-363	[7, 10, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53]
4	GGCTAACTHCGTGNCVGCNGCYGCGGTAANAC	504-535	[27, 45, 46, 47, 49, 50, 52, 54, 55, 56, 57, 58, 59, 60, 61, 62]
5a	GTGTAGMGGTGAAATKCGTAGAT	682-704	[50, 63, 64]
5b	CAAACRGGATTAGAWACCCNNGTAGTCCACGC	778-809	[7, 36, 43, 45, 50, 55, 56, 58, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76]
6a	AAANTYAAANRAATWGRCGGGGRCCCGCACAAG	906-938	[47, 48, 50, 51, 56, 58, 60, 74, 77, 78, 79, 80, 81, 82]
6b	ATGTGGTTTAATTCGA	948-963	[50, 83]
6c	CAACGCGARGAACCTTACC	966-984	[50, 84, 85, 86]
7a	AGGTGNTGCATGGYYGYCGTCAGCTCGTGYCGTGAG	1045-1080	[50, 55, 84, 85, 87, 88, 89]
7b	TGTTGGGTTAAGTCCCRYAACGAGCGCAACCCT	1082-1114	[43, 45, 47, 50, 52, 59]
8a	GGAAGGYGGGGAYGACG	1176-1192	[50, 90]
8b	GGGCKACACACGYGCTAC	1219-1236	[55, 87]
9	GCCTTGYACWCWCCGCCCGTC	1386-1406	[45, 47, 52, 69, 74, 81, 82, 86, 87, 91, 92]
10	GGGTGAAGTCRTAACAAGGTANCC	1486-1509	[34, 36, 37, 72, 75, 93, 94, 95]

http://dx.doi.org/10.1016/j.heliyon.2017.e00370

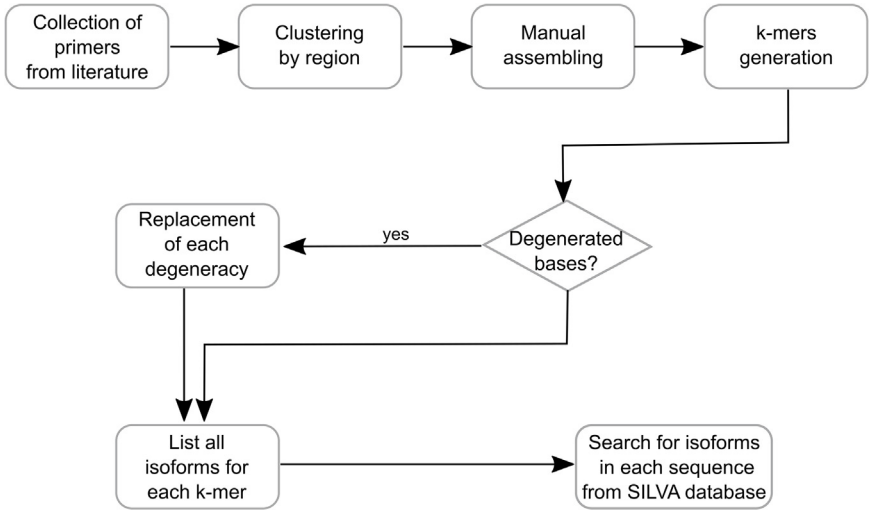

Fig. 1. Workflow established for obtaining primer contigs and the subsequent generation of *k*-mers.

2.2. Frequency and duplicate analysis

For the frequency analysis, each k-mer was considered as a set of possible isomers after degeneracies were replaced by the corresponding nucleotides, as stated above. Thus, every identical isoform of k-mer (iso-k-mer) sequence was searched in the position of the corresponding k-mer. If the iso-k-mer sequence was found, the second occurrence of the same k-mer was searched for to determine any duplicate reaction.

2.3. Search for region C1

In many sequences of the SILVA database, the C1 region located at the 5′-terminus of the molecule was not detected by region-specific primers. Therefore, the most frequent k-mer for the C3 region was used to obtain C3 sequences attached to fragments towards the 3' end containing C1 due to the high proportion of C3 sequences obtained with k-mers, and then the positions of each of the database sequences were determined. Briefly, C3-positive sequences were grouped by position, and from these, some were extended towards the 3' end containing the C1 region, whereas others were incomplete. For those containing the C1 region, the location of any k-mer within this region was examined. The percentage of positive C1 hits for each sized-group was recorded.

2.4. Estimating primers for DGGE

The utility of k-mers for retrieving sequence fragments through simulated PCR was tested by a comparison with primers commonly used to study microbiomes through denaturing gradient gel electrophoresis (DGGE). As recommended by [19], primers E341F (5'-CCTACGGGNGGCNGCA-3') and the universal reverse primer

http://dx.doi.org/10.1016/j.heliyon.2017.e00370

U926 (5'-CCGNCNATTNNTTTNAGTTT-3') were used. These corresponded to positions 340–355 and 906–925 for forward and reverse primers, respectively, in E. coli 16S rRNA. Thus, a fragment of approximately 685 nucleotides was expected. These primers were designed for annealing in the conserved regions 3 and 6. The 12-mers used for comparison were those registering the highest frequency in the corresponding positions.

3. Results and discussion

Several primers designed to match conserved regions have been proposed for the amplification and study of complete or partial sequences of the 16S rRNA gene [20, 21]. Differences among primers used for the same conserved region can be as simple as a single base substitution, but in other cases the difference may be a pair of extra nucleotides or the use or degenerate bases for particular positions, etcetera. The main challenge in the study of environmental samples is to obtain sequences from most of the microbes thriving in any niche and consequently, assess comprehensive genomic information about the diversity, structure and functions of the microorganisms forming complex communities.

The information contained in databases has exponentially increased in recent years due to the implementation of current high throughput sequencing technologies [22, 23], and therefore is a useful resource for determining the most conserved fraction of each region, which should be considered not only for designing primers but also for evaluating evolutionary or mutational patterns of this molecule.

3.1. Primer contigs and generation of *k*-mers

Contigs were successfully obtained by assembling the specific reported primers for each conserved region of the bacterial 16S rRNA; however, two or more gap-free primer contigs were obtained for regions 5, 6, 7, and 8, for a total of 15 primer contigs (Table 1). Contig sizes ranged from 16 to 44 nucleotides, and most of them (except one) exhibited ambiguities (Table 2). The only primer contig that did not require degeneracies was 6b, while primer contig 6a required the highest number of degeneracies with seven out of 33 nucleotides (21%). All primer contigs together covered 388 nucleotide positions, which corresponds to 25% of the average size of the 16S rRNA gene.

The number of k-mers is directly dependent on the size of the primer contig and is a positional marker, whereas the number of iso-k-mers is a product of the number of degeneracies in each k-mer and is a sequential marker. In this case, as the k-mers size increased from 9 to 15, the overall number of k-mers decreased from 268 to 178, but the number of iso k-mers increased from 1,762 to 4,717, respectively (Table 2).

http://dx.doi.org/10.1016/j.heliyon.2017.e00370

Table 2. Descriptive information of contigs generated after assembly of the reported primers for each conserved region of the 16S rRNA gene. The size of each contig, number of ambiguities detected and the number of iso-*k*-mers are shown. The number of generated *k*-mers is dependent on the primer contig size and is easily calculated (*k*-mers = primer contig size $- k + 1$), while the number of isomers is related to the number of degeneracies in each *k*-mer.

Primer	Contig		Iso 9-mers	Iso 10-mers	Iso 11-mers	Iso 12-mers	Iso 13-mers	Iso 14-mers	Iso 15-mers
Name	Length	Ambiguities							
1	20	2	38	39	38	36	32	28	24
2	20	3	64	63	62	61	60	56	52
3	44	8	288	333	390	488	612	734	856
4	32	6	455	574	765	970	1,175	1,476	1,792
5a	23	2	30	30	30	30	30	30	30
5b	32	4	213	244	275	306	334	350	372
6a	33	7	346	437	504	602	704	926	1,148
6b	16	0	8	7	6	5	4	3	2
6c	19	1	20	19	18	16	14	12	10
7a	36	5	110	125	140	167	194	222	246
7b	33	2	51	53	55	57	59	61	63
8a	17	2	24	24	24	22	20	16	12
8b	18	2	22	22	22	22	22	20	16
9	21	3	59	64	66	68	64	60	56
10	24	2	34	34	34	36	38	40	38
Total	**388**	**49**	**1,762**	**2,068**	**2,429**	**2,886**	**3,362**	**4,034**	**4,717**

3.2. *k*-mer search and frequency analysis

Frequency analysis revealed that k-mer size was inversely proportional to the occurrence of k-mers in the different conserved regions, suggesting an increase in stringency. Frequency curves for k-mers of different sizes of three primer contigs are shown in Fig. 2 as an example. In general, all curves exhibited similar trends in most cases; however, significant deviations were observed, particularly for shorter k-mers, suggesting that the occurrence of non-specific reactions that could be the result of a decrease in stringency (Fig. 2B). In other cases (Fig. 2C), the curves had the same shape indicating a low influence of size on specificity.

Another important feature for selecting the appropriate k-mer size is specificity, defined as the ability to react only at one site of the 16S rRNA sequence.

http://dx.doi.org/10.1016/j.heliyon.2017.e00370

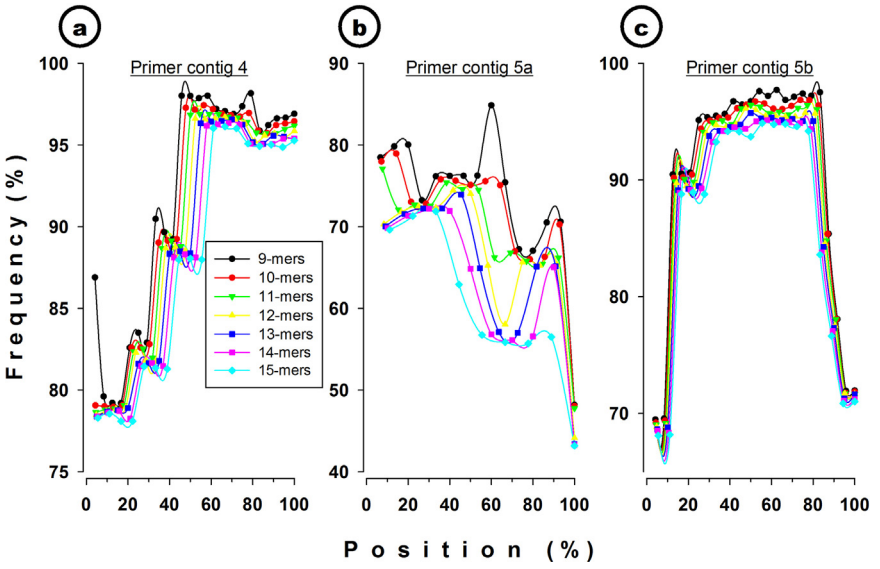

Fig. 2. Frequency of *k*-mers of 9 to 15 nucleotides detected in different conserved regions of 16S rRNA sequences contained in the SILVA database.

Therefore, the presence of duplicate reactions was investigated for each set of k-mers of different sizes. As shown in Fig. 3, the number of duplicates decreased as k-mer size increased in all cases, although with different slopes.

Duplicate reactions were substantially reduced when 12-mers were used, and the same occurred when 13-, 14- or 15-mers were considered (Fig. 3). The inflection

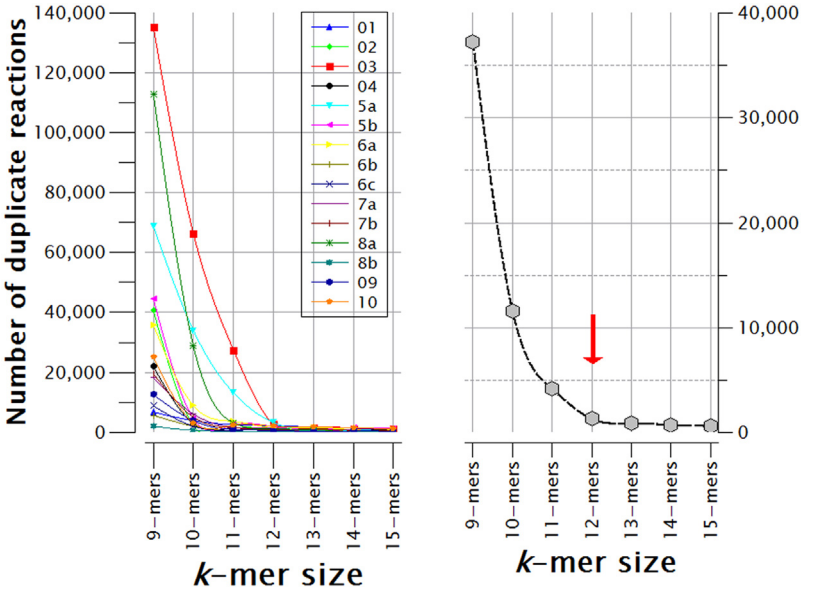

Fig. 3. Duplicate reactions detected within sequences obtained from the SILVA database when using 9- to 15-mers constructed from the primer contigs matching all conserved regions of the 16S rRNA.

http://dx.doi.org/10.1016/j.heliyon.2017.e00370

point at a size of 12 nucleotides suggests that 12-mers represent the optimal k-mer size for maintaining high coverage while avoiding non-specific reactions and with a minimal occurrence of duplicate reactions.

Therefore, the use of 12-mers represents a reliable alternative to study 16R rRNA sequences and to determine some characteristics of the conserved regions as well as the usefulness of the reported primers. Table 3 shows the most frequent 12-mers detected in each primer consensus. Conserved regions 3, 4, 5b, 6a, 7a, and 7b registered k-mers with a frequency of ≥95%, whereas conserved regions located at the extremes recorded the lowest frequencies (Table 3). This Table containing the highest frequencies of 12-k-mers in different regions allows for visualizing the most conserved areas and identifying the best candidates for the design and use of primers. The uses for this information are diverse and will depend on research requirements, as discussed below. In addition, the use of k-mers represents an ideal strategy for obtaining major proportions of sequences of this biomarker, which, according to recent evidence, may not be as conserved as expected across the breadth of microbial diversity [24].

Table 3. Primer contigs constructed for the different conserved regions. 12-mers registering the highest frequency in each primer contig are underlined and in bold. The number and sequence of each primer contig, as well as position (following the numbering of *E. coli* rRNA) and frequency for each 12-mer are indicated. Minor primer contigs are italicized.

Primer Contig		12-mer	
Number	**Sequence**	**Position**	**Frequency**
1	AGAGTTT**GATYMTGGCT**CAG	15	195,901 (38.2%)
2	**ASYGGCGNACGG**GTGAGTAA	100	405,570 (79.0%)
3	ACTGAGAYACGGYCCARACTCCTA**CGGRNGGCNGCA**GTRRGGAA	344	500,253 (97.5%)
4	GGCTAACTHCGTG**NCVGCNGCYGCG**GTAANAC	517	496,412 (96.7%)
5a	*GTGTAGMGGTGAAATKCGTAGAT*	*686*	*382,156 (74.4%)*
5b	CAAACRGGA**TTAGAWACCCNN**GTAGTCCACGC	787	493,348 (96.1%)
6a	AAANTYAAA**NRAATWGRCGGG**GRCCCGCACAAG	915	501,792 (97.8%)
6b	*ATGTGGTTTAATTCGA*	*949*	*389,530 (75.9%)*
6c	*CAACGCGARGAACCTTACC*	*971*	*393,614 (76.7%)*
7a	AGGTGNTGCAT**GGYYGYCGTCAG**CTCGTGYCGTGAG	1056	499,976 (97.4%)
7b	*TGTTGGGTTAAGTCCCRYAACGAGCGCAACCCT*	*1101*	*489,290 (95.3%)*
8a	**GGAAGGYGGGGA**YGACG	1176	457,537 (89.1%)
8b	*GGGCKACACACGYGCTAC*	*1220*	*382,857 (74.6%)*
9	GCCT**TGYACWCWCCGC**CCGTC	1390	388,911 (75.8%)
10	GGGTG**AAGTCRTAACAA**GGTANCC	1491	172,918 (33.7%)

http://dx.doi.org/10.1016/j.heliyon.2017.e00370

Consensus sequences of conserved regions obtained by this technique, may serve to design primers for biological samples, but only as one of many considerations that a primer design requires. This is a virtual approach and may not fully correspond to the biological reality of metagenomes; therefore these results should be taken with caution and be considered only as additional information for biological experiments.

3.3. Querying the C1 region

The most frequent 12-mer from primer contig 1 was detected in less than 40% of the +513,000 bacterial sequences from the SILVA database. This may be associated with a lack of primer specificity, but mainly stems from the absence of complete sequences corresponding to the C1 region in the database. Despite the exponential increase of information uploaded to databases over the past 5–10 years, most of these data are composed of short sequences of central 16S rRNA positions [25]. Therefore, the lower frequency of terminal regions is expected. According to the data shown in Table 3, only 198,419 out of the 513,000+ sequences contain the most common 12-mer of the C1 region, whereas the most common 12-mer of the C3 region (E. coli, position 344) was detected in 500,057 sequences.

The C3 region was selected because of the higher detection frequency along with its proximity to the C1 region (~344 nucleotides). The results revealed that C1 was detected in a low proportion (if any) of those sequences where C3 was located closer to the 5' end (260–300). This low proportion of C1 was maintained until C3 was located at position 335 or beyond (Fig. 4), whereupon the proportion of C1 was greater than 60% and even approached 100% in some cases. For example, more than 90% of the sequences whose C3 was located at position 355 registered the presence of C1. Apparently, the low frequency of C1 is associated with 5' end-truncation; something similar may occur in the C10 region, which registers 12-mers with low frequency. The use of the most frequent 12-mers determined for each region was useful to explain and verify that several of the sequences deposited in the database SILVA are truncated. This may represent a source of bias for experiments considering extreme regions of the 16S rRNA molecule for phylogenetic studies. For example, variable regions located at the extremes (V1, V9) have been used to detect particular taxonomic groups [26, 27]; however, a brief study of these sequences through the use of k-mers revealed that many studies considering such variable regions have not considered a substantial missing proportion.

3.4. Comparison with DGGE primers

Performing virtual PCR on a set of sequences is a very useful practice and allows for estimating the response of specific primers in real samples. Although inherent

http://dx.doi.org/10.1016/j.heliyon.2017.e00370

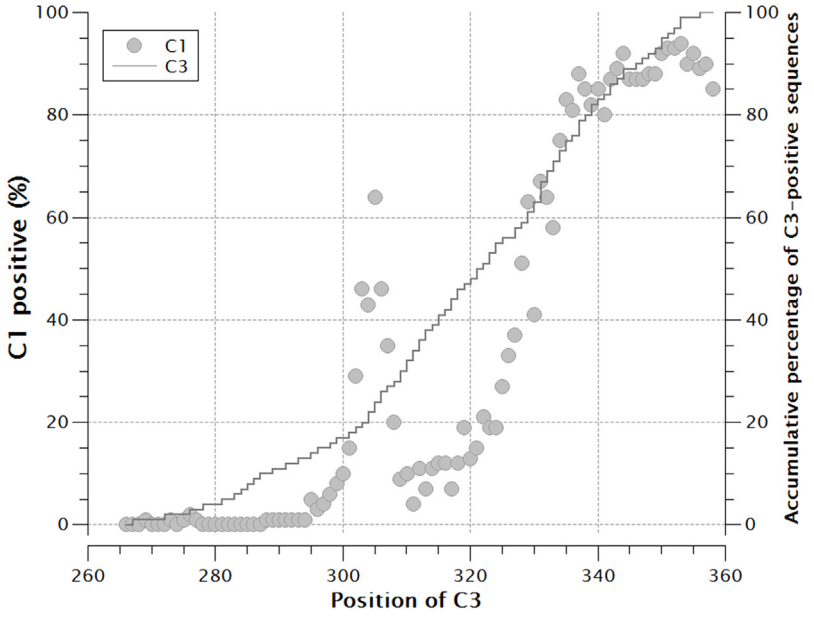

Fig. 4. Proportion of C1 sequences obtained when using 12-mers matching C3 located at different nucleotide positions. For example, C1 was not detected in sequences when C3 is located at position 295 or lower; meanwhile, when C3 is located at position 340 or higher, 80% or more of the sequences contained C1. The cumulative percentage of C3-positive sequences is indicated by the step line.

factors to the reaction and the physicochemical properties of the primers are not fully considered, the advantages are substantial, especially in terms of cost and time. From the 513,309 sequences contained in the SILVA 123 database, both DGGE primers considered for the study were found in 468,079 sequences (91.19%), while the pair of 12-mers was detected in 489,734 (95.41%). This difference (greater than 4%) was explained by only one of the primers being detected in $>$ 20,000 sequences, whereas the same pattern occurred in $<$12,100 sequences when 12-mers were used (Table 4).

Sequences not reacting with the forward primer (22,376) were recovered using the most frequent 12-mer with the upstream addition of four nucleotides, as primer sequences should start at this position (Fig. 5). Using this strategy, the recovery of 10,802 additional sequences was possible, and the analysis of these four nucleotides revealed a great diversity of these positions, which explained the lack of in silico detection using the primer. For example, the most frequent tetranucleotides were TCTA (28.7%) and CCTG (27.9%), followed by 121 different tetranucleotide combinations representing an aggregate proportion of 43% (data not shown). Moreover, this small fraction of tetranucleotides is so variable that degeneracies (YHHV) were required to represent 97.5% of sequences.

Another factor affecting primer detection (CCTACGGGNGGCNGCA) was nucleotide G8, which corresponded to the ambiguity R4 of the 12-mer

http://dx.doi.org/10.1016/j.heliyon.2017.e00370

Table 4. Reactions of primers used for DGGE and for the most frequent 12-mers of regions C3 and C6. The vast majority of the SILVA database sequences reacted at both ends, indicating a possible amplification. The reaction occurred only at one end in ~8% and 4% of sequences when primers and 12-mers were used, respectively. Less than 0.5% did not react with any primer or 12-mer.

Reaction	DGGE Primers	12-mers	Primers & 12-mers
Both ends	468,079 (91.19%)	489,734 (95.41%)	466,499 (90.88%)
Forward	20,472 (3.99%)	10,517 (2.05%)	8,493 (1.65%)
Reverse	22,376 (4.36%)	12,058 (2.35%)	11,386 (2.22%)
None	2,382 (0.46%)	1,000 (0.19%)	888 (0.17%)
Total	513,309	513,309	487,266 (94.93%)

(CGGRNGGCNGCA). From the 10,802 retrieved sequences, only 70.72% (7,640) registered G for this position, whereas A was detected in the remaining 29.27% (3,162). Therefore, the ambiguity R seems to be a better choice for this position, but will depend on an overall evaluation of the primer and the microbial population to be studied or the expected taxonomic groups. Also noteworthy was the 30% increase in detection of the undetected sequences that was achieved when using the ambiguity R instead of the original nucleotide; this represented approximately 0.6% of the total sequences. The analysis of these types of details is not accessible with biological samples because the PCR reaction may still be successful considering that mismatches occur upstream of the zone containing the critical nucleotides for annealing and amplification [28].

Similar results were obtained for the analysis of the 3′ end. A portion of sequences from the SILVA database (20,472 or 3.99%) reacted only with the forward primer. The undetected sequences were recovered using the most frequent 12-mer method, and the fragment corresponding to the primer was then analysed. Several degeneracies were required to cover most of the sequences (RVVHHHVR RKRAATTGACGG). Although the origin causing the lack of reaction was found, whether to increase the number of degeneracies in the primer or to accept the loss

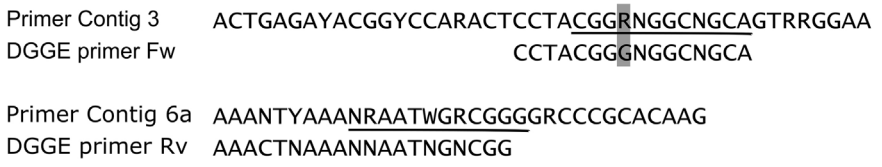

Fig. 5. Alignment of primers for DGGE and the primer contig. The most frequent 12-mer is underlined, while the difference G8 of the primer, which corresponds to R4 of the 12-mer, is shaded.

http://dx.doi.org/10.1016/j.heliyon.2017.e00370

of a small percentage (3%) of sequences will depend on the research intent or objective.

Finally, using the k-mer strategy, and particularly 12-mers, to obtain and analyse 16S rRNA sequences was found to be reliable. This may not only serve to evaluate the presence of a molecular motif but also to evaluate and design primers, study mutational or evolutionary patterns, and detect rare sequences, along with many other possible applications. Moreover, this approach can consider all of the novel and rare 16S rRNA sequences obtained through shotgun sequencing and deposited in database; adapting in real time to databases actualizations.

Declarations

Author contribution statement

Francisco Vargas-Albores, Marcel Martínez-Porchas: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This work was supported by the National Council for Science and Technology (CONACyT), Mexico, grant 84398 (to FVA).

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

We express our gratitude to E. Villalpando-Canchola for his assistance in the listing of primers and ordering the bibliography.

References

[1] S. Pandey, S. Singh, A.N. Yadav, L. Nain, A.K. Saxena, Phylogenetic diversity and characterization of novel and efficient cellulase producing bacterial isolates from various extreme environments, Biosci. Biotechnol. Biochem. 77 (2013) 1474–1480.

[2] A.-M. Lakaniemi, C.J. Hulatt, K.D. Wakeman, D.N. Thomas, J.A. Puhakka, Eukaryotic and prokaryotic microbial communities during microalgal biomass production, Bioresour. Technol. 124 (2012) 387–393.

[3] V. Lazarevic, K. Whiteson, S. Huse, D. Hernandez, L. Farinelli, M. Østerås, J. Schrenzel, P. François, Metagenomic study of the oral microbiota by Illumina high-throughput sequencing, J. Microbiol. Methods 79 (2009) 266–271.

[4] E. Stackebrandt, B. Goebel, Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology, Int. J. Syst. Evol. Microbiol. 44 (1994) 846–849.

[5] A. Klindworth, E. Pruesse, T. Schweer, J. Peplies, C. Quast, M. Horn, F.O. Glöckner, Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies, Nucleic Acids Res. 41 (2013) e1.

[6] N. Boon, W. De Windt, W. Verstraete, E.M. Top, Evaluation of nested PCR–DGGE (denaturing gradient gel electrophoresis) with group-specific 16S rRNA primers for the analysis of bacterial communities from different wastewater treatment plants, FEMS Microbiol. Ecol. 39 (2002) 101–112.

[7] G.C. Baker, J.J. Smith, D.A. Cowan, Review and re-analysis of domain-specific 16S primers, J. Microbiol. Methods 55 (2003) 541–555.

[8] Y. Wang, R.M. Tian, Z.M. Gao, S. Bougouffa, P.-Y. Qian, Optimal eukaryotic 18S and universal 16S/18S ribosomal RNA primers and their application in a study of symbiosis, PLoS One 9 (2014) e90053.

[9] X. Zhang, X. Feng, F. Wang, Diversity and Metabolic Potentials of Subsurface Crustal Microorganisms from the Western Flank of the Mid-Atlantic Ridge, Front. Microbiol. 7 (2016) 363.

[10] S. Chakravorty, D. Helb, M. Burday, N. Connell, D. Alland, A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria, J. Microbiol. Methods 69 (2007) 330–339.

[11] M. Shakya, C. Quince, J.H. Campbell, Z.K. Yang, C.W. Schadt, M. Podar, Comparative metagenomic and rRNA microbial diversity characterization using archaeal and bacterial synthetic communities, Environ. Microbiol. 15 (2013) 1882–1899.

[12] E.S. Allman, J.A. Rhodes, S. Sullivant, Statistically consistent *k*-mer methods for phylogenetic tree reconstruction, J. Comput. Biol. 24 (2017) 153–171.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00370

[13] A. Gupta, I.K. Jordan, L. Rishishwar, stringMLST: a fast *k*-mer based tool for multilocus sequence typing, Bioinformatics 33 (2017) 119–121.

[14] D. Pellow, D. Filippova, C. Kingsford, Improving bloom filter performance on sequence data using k-mer bloom filters, J. Comput. Biol. (2016).

[15] A. Shajii, D. Yorukoglu, Y. William Yu, B. Berger, Fast genotyping of known SNPs through approximate *k*-mer matching, Bioinformatics 32 (2016) i538–i544.

[16] R. Wang, Y. Xu, B. Liu, Recombination spot identification based on gapped k-mers, Sci. Rep. 6 (2016) 23934.

[17] D. Wood, S. Salzberg, Kraken ultrafast metagenomic sequence classification using exact alignments, Genome Biol. 15 (2014) R46.

[18] D.R. Zerbino, E. Birney, Velvet Algorithms for *de novo* short read assembly using de Bruijn graphs, Genome Res. 18 (2008) 821–829.

[19] N.N. Perreault, D.T. Andersen, W.H. Pollard, C.W. Greer, L.G. Whyte, Characterization of the Prokaryotic diversity in cold saline perennial springs of the canadian high arctic, Appl. Environ. Microbiol. 73 (2007) 1532–1543.

[20] J. Ghyselinck, S. Pfeiffer, K. Heylen, A. Sessitsch, P. De Vos, The effect of primer choice and short read sequences on the outcome of 16S rRNA gene based diversity studies, PLoS One 8 (2013) e71360.

[21] T. Matsuki, K. Watanabe, J. Fujimoto, T. Takada, R. Tanaka, Use of 16S rRNA gene-targeted group-specific primers for real-time PCR analysis of predominant bacteria in human feces, Appl. Environ. Microbiol. 70 (2004) 7220–7228.

[22] M. Kim, K.H. Lee, S.W. Yoon, B.S. Kim, J. Chun, H. Yi, Analytical tools and databases for metagenomics in the next-generation sequencing era, Genomics Informat. 11 (2013) 102–113.

[23] S.C. Schuster, Next-generation sequencing transforms today's biology, Nat. Methods 5 (2008) 16–18.

[24] M. Martinez-Porchas, E. Villalpando-Canchola, L.E. Ortiz Suarez, F. Vargas-Albores, How conserved are the conserved 16S-rRNA regions? PeerJ. 5 (2017) e3036.

[25] P. Jeraldo, N. Chia, N. Goldenfeld, On the suitability of short reads of 16S rRNA for phylogeny-based analyses in environmental surveys, Environ. Microbiol. 13 (2011) 3000–3009.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00370

[26] M. Kim, M. Morrison, Z. Yu, Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes, J. Microbiol. Methods 84 (2011) 81–87.

[27] P.S. Kumar, M.R. Brooker, S.E. Dowd, T. Camerlengo, Target region selection is a critical determinant of community fingerprints generated by 16S pyrosequencing, PLoS One 6 (2011) e20956.

[28] C.W. Dieffenbach, T.M. Lowe, G.S. Dveksler, General concepts for PCR primer design, Genome Res. 3 (1993) S30–S37.

[29] U. Edwards, T. Rogall, H. Blöcker, M. Emde, E.C. Böttger, Isolation and direct complete nucleotide determination of entire genes Characterization of a gene coding for 16S ribosomal RNA, Nucleic Acids Res. 17 (1989) 7843–7853.

[30] P.S. Kumar, A.L. Griffen, M.L. Moeschberger, E.J. Leys, Identification of candidate periodontal pathogens and beneficial species by quantitative 16S clonal analysis, J. Clin. Microbiol. 43 (2005) 3944–3955.

[31] W. Ludwig, G. Mittenhuber, C.G. Friedrich, Transfer of *Thiosphaera pantotropha* to *Paracoccus denitrificans*, Int. J. Syst. Bacteriol. 43 (1993) 363–367.

[32] J.O. McInerney, M. Wilkinson, J.W. Patching, T.M. Embley, R. Powell, Recovery and phylogenetic analysis of novel archaeal rRNA sequences from a deep-sea deposit feeder, Appl. Environ. Microbiol. 61 (1995) 1646–1648.

[33] G. Muyzer, A. Teske, C.O. Wirsen, H.W. Jannasch, Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments, Arch. Microbiol. 164 (1995) 165–172.

[34] F.A. Oguntoyinbo, Monitoring of marine *Bacillus* diversity among the bacteria community of sea water, Afr. J. Biotechnol. 6 (2007) 163–166.

[35] M.T. Suzuki, S.J. Giovannoni, Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR, Appl. Environ. Microbiol. 62 (1996) 625–630.

[36] K.H. Wilson, R.B. Blitchington, R.C. Greene, Amplification of bacterial 16S ribosomal DNA with polymerase chain reaction, J. Clin. Microbiol. 28 (1990) 1942–1946.

[37] T.A. Isenbarger, M. Finney, C. Ríos-Velázquez, J. Handelsman, G. Ruvkun, Miniprimer PCR, a new lens for viewing the microbial world, Appl. Environ. Microbiol. 74 (2008) 840–849.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00370

[38] A. Schmalenberger, F. Schwieger, C.C. Tebbe, Effect of primers hybridizing to different evolutionarily conserved regions of the small-subunit rRNA gene in PCR-based microbial community analyses and genetic profiling, Appl. Environ. Microbiol. 67 (2001) 3557–3563.

[39] A. Sundquist, S. Bigdeli, R. Jalili, M.L. Druzin, S. Waller, K.M. Pullen, Y.Y. El-Sayed, M.M. Taslimi, S. Batzoglou, M. Ronaghi, Bacterial flora-typing with targeted, chip-based Pyrosequencing, BMC Microbiol. 7 (2007) 1–11.

[40] R.I. Amann, B.J. Binder, R.J. Olson, S.W. Chisholm, R. Devereux, D.A. Stahl, Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations, Appl. Environ. Microbiol. 56 (1990) 1919–1925.

[41] L. Dethlefsen, S. Huse, M.L. Sogin, D.A. Relman, The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing, PLoS Biol. 6 (2008) e280.

[42] N. Fierer, M. Hamady, C.L. Lauber, R. Knight, The influence of sex, handedness, and washing on the diversity of hand surface bacteria, Proc. Natl. Acad. Sci. USA 105 (2008) 17994–17999.

[43] P. Cruaud, A. Vigneron, C. Lucchetti-Miganeh, P.E. Ciron, A. Godfroy, M.-A. Cambon-Bonavita, Influence of DNA extraction method, 16s rRNA targeted hypervariable regions, and sample origin on microbial diversity detected by 454 pyrosequencing in marine chemosynthetic ecosystems, Appl. Environ. Microbiol. 80 (2014) 4626–4639.

[44] M.C. Hansen, T. Tolker-Nielsen, M. Givskov, S. Molin, Biased 16S rDNA PCR amplification caused by interference from DNA flanking the template region, FEMS Microbiol. Ecol. 26 (1998) 141–149.

[45] D.J. Lane, 16S/23S rRNA sequencing, In: E. Stackebrandt, M. Goodfellow (Eds.), Nucleic acid techniques in bacterial systematics, John Wiley & Sons, Chichester, United Kingdom, 1991, pp. 115–175.

[46] G. Muyzer, E.C. de Waal, A.G. Uitterlinden, Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA, Appl. Environ. Microbiol. 59 (1993) 695–700.

[47] A.L. Reysenbach, B. Pace, Archaea: A Laboratory Manual—Thermophiles, In: F.T. Robb, A.R. Place (Eds.), Cold Spring Harbour Laboratory Press, New York, 1995, pp. 101–107.

[48] K. Rudi, O.M. Skulberg, F. Larsen, K.S. Jakobsen, Strain characterization and classification of oxyphotobacteria in clone cultures on the basis of 16S rRNA

sequences from the variable regions V6, V7, and V8, Appl. Environ. Microbiol. 63 (1997) 2593–2599.

[49] J. Walter, G.W. Tannock, A. Tilsala-Timisjarvi, S. Rodtong, D.M. Loach, K. Munro, T. Alatossava, Detection and identification of gastrointestinal *Lactobacillus* species by using denaturing gradient gel electrophoresis and species-specific PCR primers, Appl. Environ. Microbiol. 66 (2000) 297–303.

[50] Y. Wang, P.-Y. Qian, Conservative fragments in bacterial 16s rRNA genes and primer design for 16s ribosomal DNA amplicons in metagenomic studies, PLoS One 4 (2009) e7401.

[51] K. Watanabe, Y. Kodama, S. Harayama, Design and evaluation of PCR primers to amplify bacterial 16S ribosomal DNA fragments used for community fingerprinting, J. Microbiol. Methods 44 (2001) 253–262.

[52] J. Wuyts, Y. Van de Peer, T. Winkelmans, R. De Wachter, The European database on small subunit ribosomal RNA, Nucleic Acids Res. 30 (2002) 183–185.

[53] V.P. Natarajan, X. Zhang, Y. Morono, F. Inagaki, F. Wang, A Modified SDS-Based DNA Extraction Method for High Quality Environmental DNA from Seafloor Environments, Front. Microbiol. 7 (2016) 986.

[54] J.C. Cho, D.H. Lee, Y.C. Cho, J.C. Cho, S.J. Kim, Direct extraction of DNA from soil for amplification of 16s rrna gene sequences by polymerase chain reaction, J. Microbiol. 34 (1996) 229–235.

[55] S. DasSarma, E.F. Fleischmann, Archaea A Laboratory Manual—Halophiles, Cold Spring Harbour Laboratory Press, New York USA, 1995.

[56] Z. Liu, C. Lozupone, M. Hamady, F.D. Bushman, R. Knight, Short pyrosequencing reads suffice for accurate microbial community analysis, Nucleic Acids Res. 35 (2007) e120.

[57] J.C. Makemson, N.R. Fulayfil, W. Landry, L.M. Van Ert, C.F. Wimpee, E.A. Widder, J.F. Case, *Shewanella woodyi* sp nov., an exclusively respiratory luminous bacterium isolated from the Alboran Sea, Int. J. Syst. Bacteriol. 47 (1997) 1034–1039.

[58] M.C. Nelson, H.G. Morrison, J. Benjamino, S.L. Grim, J. Graf, Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys, PLoS One 9 (2014) e94249.

[59] L. Ovreås, L. Forney, F.L. Daae, V. Torsvik, Distribution of bacterioplankton in meromictic Lake Saelenvannet, as determined by denaturing gradient gel

 http://dx.doi.org/10.1016/j.heliyon.2017.e00370

electrophoresis of PCR-amplified gene fragments coding for 16S rRNA, Appl. Environ. Microbiol. 63 (1997) 3367–3373.

[60] C. Quince, A. Lanzen, R.J. Davenport, P.J. Turnbaugh, Removing noise from pyrosequenced amplicons, BMC Bioinformatics 12 (2011) 38.

[61] A.L. Ruff-Roberts, J.G. Kuenen, D.M. Ward, Distribution of cultivated and uncultivated cyanobacteria and *Chloroflexus*-like bacteria in hot spring microbial mats, Appl. Environ. Microbiol. 60 (1994) 697–704.

[62] X. Zhang, J. Fang, W. Bach, K.J. Edwards, B.N. Orcutt, F. Wang, Nitrogen Stimulates the Growth of Subsurface Basalt-associated Microorganisms at the Western Flank of the Mid-Atlantic Ridge, Front. Microbiol. 7 (2016) 633.

[63] N. Klijn, A.H. Weerkamp, W.M. De Vos, Identification of mesophilic lactic acid bacteria by using polymerase chain reaction-amplified variable regions of 16S rRNA and specific DNA probes, Appl. Environ. Microbiol. 57 (1991) 3390–3393.

[64] J.R. Stults, O. Snoeyenbos-West, B. Methe, D.R. Lovley, D.P. Chandler, Application of the 5′ fluorogenic exonuclease assay (TaqMan) for quantitative ribosomal DNA and rRNA analysis in sediments, Appl. Environ. Microbiol. 67 (2001) 2781–2789.

[65] S.M. Barns, R.E. Fundyga, M.W. Jeffries, N.R. Pace, Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment, Proc. Natl. Acad. Sci. USA 91 (1994) 1609–1613.

[66] C.F. Brunk, N. Eis, Quantitative measure of small-subunit rRNA gene sequences of the kingdom Korarchaeota, Appl. Environ. Microbiol. 64 (1998) 5064–5066.

[67] M.J. Claesson, Q. Wang, O. O'Sullivan, R. Greene-Diniz, J.R. Cole, R.P. Ross, P.W. O'Toole, Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions, Nucleic Acids Res. 38 (2010) e200.

[68] J.A. Colquhoun, Discovery of deep-sea actinomycetes, Research School of Biosciences, University of Kent, Canterbury UK, 1997.

[69] A. Engelbrektson, V. Kunin, K.C. Wrighton, N. Zvenigorodsky, F. Chen, H. Ochman, P. Hugenholtz, Experimental factors affecting PCR-based estimates of microbial species richness and evenness, ISME J. 4 (2010) 642–647.

[70] G.E. Flores, J.H. Campbell, J.D. Kirshtein, J. Meneghin, M. Podar, J.I. Steinberg, J.S. Seewald, M.K. Tivey, M.A. Voytek, Z.K. Yang, A.L. Reysenbach, Microbial community structure of hydrothermal deposits from

 http://dx.doi.org/10.1016/j.heliyon.2017.e00370

geochemically different vent fields along the Mid-Atlantic Ridge, Environ. Microbiol. 13 (2011) 2158–2171.

[71] A.J. McBain, R.G. Bartolo, C.E. Catrenich, D. Charbonneau, R.G. Ledder, A. H. Rickard, S.A. Symmons, P. Gilbert, Microbial characterization of biofilms in domestic drains and the establishment of stable biofilm microcosms, Appl. Environ. Microbiol. 69 (2003) 177–185.

[72] L.F.W. Roesch, R.R. Fulthorpe, A. Riva, G. Casella, A.K.M. Hadwin, A.D. Kent, S.H. Daroub, F.A.O. Camargo, W.G. Farmerie, E.W. Triplett, Pyrosequencing enumerates and contrasts soil microbial diversity, ISME J. 1 (2007) 283–290.

[73] A. Teske, K.B. Sorensen, Uncultured archaea in deep marine subsurface sediments: have we caught them all? ISME J. 2 (2007) 3–18.

[74] J. Tremblay, K. Singh, A. Fern, E.S. Kirton, S. He, T. Woyke, J. Lee, F. Chen, J.L. Dangl, S.G. Tringe, Primer and platform effects on 16S rRNA tag sequencing, Front. Microbiol. 6 (2015) 771.

[75] W.G. Weisburg, S.M. Barns, D.A. Pelletier, D.J. Lane, 16S ribosomal DNA amplification for phylogenetic study, J. Bacteriol. 173 (1991) 697–703.

[76] M. Sakai, A. Matsuka, T. Komura, S. Kanazawa, Application of a new PCR primer for terminal restriction fragment length polymorphism analysis of the bacterial communities in plant roots, J. Microbiol. Methods 59 (2004) 81–89.

[77] R.I. Amann, J. Stromley, R. Devereux, R. Key, D.A. Stahl, Molecular and microscopic identification of sulfate-reducing bacteria in multispecies biofilms, Appl. Environ. Microbiol. 58 (1992) 614–623.

[78] E.O. Casamayor, R. Massana, S. Benlloch, L. Ovreas, B. Diez, V.J. Goddard, J.M. Gasol, I. Joint, F. Rodriguez-Valera, C. Pedros-Alio, Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern, Environ. Microbiol. 4 (2002) 338–348.

[79] T. Henckel, M. Friedrich, R. Conrad, Molecular analyses of the methane-oxidizing microbial community in rice field soil by targeting the genes of the 16s rRNA, particulate methane monooxygenase, and methanol dehydrogenase, Appl. Environ. Microbiol. 65 (1999) 1980–1990.

[80] G. Jurgens, K. Lindström, A. Saano, Novel group within the kingdom *Crenarchaeota* from boreal forest soil, Appl. Environ. Microbiol. 63 (1997) 803–805.

 http://dx.doi.org/10.1016/j.heliyon.2017.e00370

[81] D.J. Lane, B. Pace, G.J. Olsen, D.A. Stahl, M.L. Sogin, N.R. Pace, Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses, Proc. Natl. Acad. Sci. USA 82 (1985) 6955–6959.

[82] D.-P. Mao, Q. Zhou, C.-Y. Chen, Z.-X. Quan, Coverage evaluation of universal bacterial primers using the metagenomic datasets, BMC Microbiol. 12 (2012) 66.

[83] T. Iwamoto, K. Tani, K. Nakamura, Y. Suzuki, M. Kitagawa, M. Eguchi, M. Nasu, Monitoring impact of in situ biostimulation treatment on groundwater bacterial community by DGGE, FEMS Microbiol. Ecol. 32 (2000) 129–141.

[84] M.L. Sogin, H.G. Morrison, J.A. Huber, D.M. Welch, S.M. Huse, P.R. Neal, J.M. Arrieta, G.J. Herndl, Microbial diversity in the deep sea and the underexplored rare biosphere, Proc. Natl. Acad. Sci. USA 103 (2006) 12115–12120.

[85] J. Jonasson, M. Olofsson, H.-J. Monstein, Classification, identification and subtyping of bacteria based on pyrosequencing and signature matching of 16s rDNA fragments, APMIS 110 (2002) 263–272.

[86] U. Nübel, B. Engelen, A. Felske, J. Snaidr, A. Wieshuber, R.I. Amann, W. Ludwig, H. Backhaus, Sequence heterogeneities of genes encoding 16S rRNAs in *Paenibacillus polymyxa* detected by temperature gradient gel electrophoresis, J. Bacteriol. 178 (1996) 5636–5643.

[87] N. Youssef, C.S. Sheik, L.R. Krumholz, F.Z. Najar, B.A. Roe, M.S. Elshahed, Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys, Appl. Environ. Microbiol. 75 (2009) 5227–5236.

[88] J.A. Huber, D.B. Mark Welch, H.G. Morrison, S.M. Huse, P.R. Neal, D.A. Butterfield, M.L. Sogin, Microbial population structures in the deep marine biosphere, Science 318 (2007) 97–100.

[89] M.J. Ferris, G. Muyzer, D.M. Ward, Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community, Appl. Environ. Microbiol. 62 (1996) 340–346.

[90] N. Bodenhausen, M.W. Horton, J. Bergelson, Bacterial communities associated with the leaves and the roots of *Arabidopsis thaliana*, PLoS One 8 (2013) e56329.

[91] J.R. Marchesi, T. Sato, A.J. Weightman, T.A. Martin, J.C. Fry, S.J. Hiom, D. Dymock, W.G. Wade, Design and evaluation of useful bacterium-specific

 http://dx.doi.org/10.1016/j.heliyon.2017.e00370

PCR primers that amplify genes coding for bacterial 16S rRNA, Appl. Environ. Microbiol. 64 (1998) 795–799.

[92] Z. Yu, M. Morrison, Comparisons of different hypervariable regions of rrs genes for use in fingerprinting of microbial communities by PCR-Denaturing Gradient Gel Electrophoresis, Appl. Environ. Microbiol. 70 (2004) 4800-–4806.

[93] J. Hang, V. Desai, N. Zavaljevski, Y. Yang, X. Lin, R.V. Satya, L.J. Martinez, J.M. Blaylock, R.G. Jarman, S.J. Thomas, R.A. Kuschner, 16S rRNA gene pyrosequencing of reference and clinical samples and investigation of the temperature stability of microbiome profiles, Microbiome 2 (2014) 31.

[94] C. Lin, D.A. Stahl, Taxon-specific probes for the cellulolytic genus *Fibrobacter* reveal abundant and novel equine-associated populations, Appl. Environ. Microbiol. 61 (1995) 1348–1351.

[95] A.L. Reysenbach, L.J. Giver, G.S. Wickham, N.R. Pace, Differential amplification of rRNA genes by polymerase chain reaction, Appl. Environ. Microbiol. 58 (1992) 3417–3418.

http://dx.doi.org/10.1016/j.heliyon.2017.e00370