How conserved are the conserved 16S-rRNA regions?

Marcel Martinez-Porchas[1,*], Enrique Villalpando-Canchola[1,*], Luis Enrique Ortiz Suarez[2] and Francisco Vargas-Albores[1]

[1] Centro de Investigación en Alimentación y Desarrollo, A.C., Hermosillo, Sonora, Mexico
[2] Instituto Tecnológico de Morelia, Morelia, Michoacán, Mexico
[*] These authors contributed equally to this work.

ABSTRACT

The 16S rRNA gene has been used as master key for studying prokaryotic diversity in almost every environment. Despite the claim of several researchers to have the best universal primers, the reality is that no primer has been demonstrated to be truly universal. This suggests that conserved regions of the gene may not be as conserved as expected. The aim of this study was to evaluate the conservation degree of the so-called conserved regions flanking the hypervariable regions of the 16S rRNA gene. Data contained in SILVA database (release 123) were used for the study. Primers reported as matches of each conserved region were assembled to form contigs; sequences sizing 12 nucleotides (12-mers) were extracted from these contigs and searched into the entire set of SILVA sequences. Frequency analysis shown that extreme regions, 1 and 10, registered the lowest frequencies. 12-mer frequencies revealed segments of contigs that were not as conserved as expected ($\leq$90%). Fragments corresponding to the primer contigs 3, 4, 5b and 6a were recovered from all sequences in SILVA database. Nucleotide frequency analysis in each consensus demonstrated that only a small fraction of these so-called conserved regions is truly conserved in non-redundant sequences. It could be concluded that conserved regions of the 16S rRNA gene exhibit considerable variation that has to be considered when using this gene as biomarker.

Subjects Biodiversity, Computational Biology, Genomics
Keywords Kmers, Biodiversity, Conserved regions 16S, Primer design

INTRODUCTION

The study of microbial communities through sequencing the 16S rRNA gene by Sanger method has been abandoned by most of the scientists. Instead, high throughput sequencing technology has emerged as a masterpiece for the robust study of microbial communities, allowing laboratories to obtain millions of high quality sequences (*Martínez-Porchas & Vargas-Albores, 2015*; *Yang, Wang & Qian, 2016*). Whether this is a significant improvement for the discipline, the short size of these sequences is a limiting factor for the taxonomic classification of bacteria and archaea.

A novel technology based on single molecule real-time sequencing (SMRT), has been tested as possible solution for this problem with promising results. However, the per-base sequencing cost is still significantly higher than that of current high throughput sequencing platforms, and thus inviable for most of the laboratories (*Schloss et al., 2016*;

Singer et al., 2016). Therefore, the use of 16S rRNA fractions to study bacterial diversity seems to be the main strategy during the following years. Furthermore, there are specific studies focused on particular taxonomic groups of bacteria that would require a specific fraction of the 16S rRNA gene (*Li et al., 2009*; *Pfeiffer et al., 2014*).

In this regard, several primers targeting diverse hypervariable regions of the 16S rRNA gene have been used and reported as guarantee of wide coverage and good amplification (*Hiergeist, Reischl & Gessner, 2016*; *Takahashi et al., 2014*; *Yang, Wang & Qian, 2016*); however, none of these primers is truly universal and the coverage usually depends upon intrinsic factors such as primer design (sequence, size, position, degenerations, combinations), chemical reagents used, amplification conditions and other PCR biases, and extrinsic factors such as the kinds of samples (bacterial composition and environment) and PCR inhibitors hauled in the sampling process (*Albertsen et al., 2015*; *Tremblay et al., 2015*; *Walker et al., 2015*). Moreover, regions of the 16S rRNA gene differ in taxonomic informativeness (*Kumar et al., 2011*; *Soergel et al., 2012*); thus, some regions seem to be more useful for taxonomic classification from a general perspective and others for particular taxa.

Most of the studies regarding proposal and effectiveness of different primers are usually based on the study of biological samples. These studies have been useful to extend the panorama regarding the research of bacterial communities (*Cruaud et al., 2014*; *Logares et al., 2014*); however, the intrinsic and extrinsic factors influencing the performance of these primers do not allow concluding if they have the best possible coverage. These kinds of results allow to conclude if one pair of primers is better than others, but do not provide conclusive information regarding coverage; for example, it is possible that only a fraction of the species thriving in any environmental sample are being covered by any combination of primers, whereas the 16S rRNA fraction of others may require different amplification conditions, or do not match with the primer sequence because some fragments of the conserved regions are probably not as conserved as expected, and so on. In this case, the information provided by environmental samples regarding the coverage of any pair of primers would be incomplete.

Furthermore, many of the primers used are frequently not validated through *in silico* tests, while others are proved only with a couple of thousand sequences previously selected (*Morales & Holben, 2009*; *Peabody et al., 2015*). Additionally, the use of degenerated primers has been proposed for the amplification of DNA coding for homologous genes, covering a larger number of genes from unspecific prokaryotes. Degenerated primers were initially designed manually, inserting degenerations after multiple alignments; however sophisticated software programs are used today (*Linhart & Shamir, 2002*; *Najafabadi, Torabi & Chamankhah, 2008*; *Qu et al., 2012*). Thus, it is necessary to understand variations in conserved regions of the 16S rRNA gene and to carry out tests with all possible sequences (including degenerations) and combinations; which is impractical and unfeasible. However, this can be done virtually considering all of the information contained in robust databases, not only the sequences obtained from biological samples. Moreover, the analysis of these conserved regions could provide useful information to evaluate how much conserved are these regions and which sequence positions are truly conserved. Therefore, the aim of this

study was to evaluate the conservation degree of the so-called conserved regions flanking the hypervariable regions of the 16S rRNA gene.

MATERIALS AND METHODS

The conservation degree of the regions flanking all of the nine-hypervariable regions of the 16S rRNA gene was estimated through the analysis of data contained in the high quality ribosomal RNA database SILVA SSU Ref NR 99 (release 123) which have non-redundant bacterial sequences with at least 1,200 bases length.

Homemade PHP scripts were used for searching specific nucleotide chains, recovering fragments of bacterial sequences, making calculations and ordering information. The following process was carried out: all of the primers reported for each region were aligned to generate a continuous "primer contig" sequence in order to perform a sequence-scan analysis of regions (position by position); if primers formed separated contigs by a gap, each segment was considered as sub-contig (a, b or c). Previous tests using 9- to 15-mers, revealed that greater specificity was achieved with 12- to 15-mers, but a higher proportion of sequences were obtained with 12-mers (Fig. 1). Thereafter, sequences sizing 12 nucleotides (12-mers) were extracted from the primer contigs.

The number of 12-mers to prove was calculated by the following equation: *Number of 12-mers = consensus size − k + 1*; where, *k* was the kmer size (12 in this case). If the 12-mer contained degenerations, each isoform was considered for the analysis; for example, nucleotide ambiguity code establishes that keto or K represents T or G, therefore sequence containing this degeneration was multiplied by two possibilities. This was also considered for all kinds of degenerations 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 sequence(s) for each region was broken down into 12-mers and its respective isoforms replacing any degeneration by the corresponding nucleotides. Finally, the exact sequence of each 12-mer isoforms generated from all conservative regions was searched into the entire set of sequences contained in SILVA database.

Equivalent regions of primer contigs 2, 3, 4, 5b, 6a, 7a, 8a and 9 were recovered from all sequences of SILVA database using the 12-mer registering the highest frequency. After the position (full match) with the most frequent 12-mer was detected, nucleotides flanking both extremes of these 12-mers were added as many as necessary to achieve contig size. To avoid induced biases by repeated sequences only those non-redundant (NR) were considered. The frequency of the nucleotides occupying each position was determined in both set of sequences, all and NR. In case of degeneracy, each base received the proportional value. For example, for R (A/G), 0.5 of A and 0.5 of G was considered.

RESULTS AND DISCUSSION

16S rRNA sequence analyses which assume that conserved regions allow for the design of universal primers have been performed to elucidate the taxonomic affinities of a wide range of taxa and to robustly assessing the prokaryotic diversity of environmental samples (*Baker, Smith & Cowan, 2003*; *Martínez-Porchas & Vargas-Albores, 2015*; *Martinez-Porchas,*
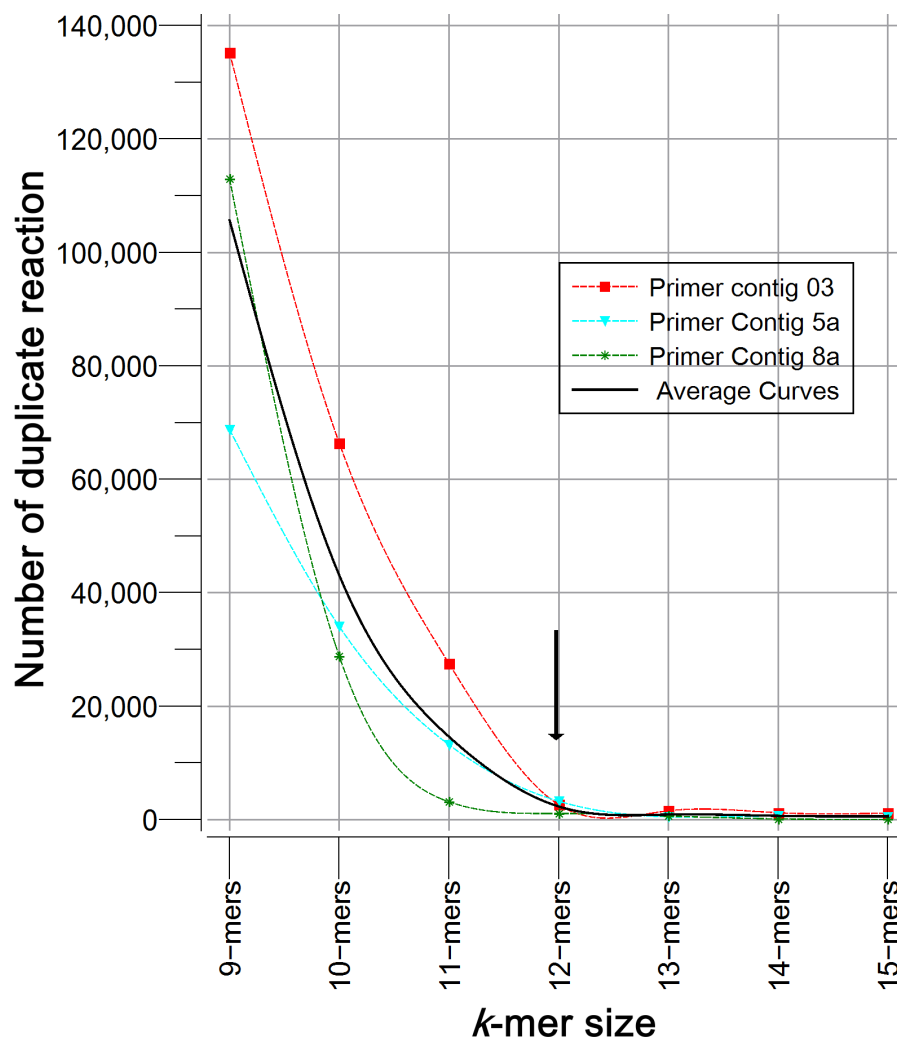
Figure 1 Specificity analysis of *k*-mers with different size (9 to 15 nucleotides). Figure describes the number of sequences of 16S rRNA gene (Silva 123) that showed duplicate reaction. *k*-mers tested corresponded to primer contigs 3, 5a and 8a. As expected, the longer the *k*-mer the greater astringency, and the probability of finding a duplicate is reduced. The optimal size was determined by the inflection point.

Villalpando-Canchola & Vargas-Albores, in press; *Wang et al., 2016*; *Wang & Qian, 2009*).
Fifteen primer contigs with lengths ranging from 16 (primer contig 6b) to 44 nucleotides (primer contig 3) were constructed by assembling all reported primers designed for matching the conservative regions (Table 1).

Except for primer contig 6b, the rest of primer contigs registered at least one degenerated base. Primer contig 3 was the most degenerated base-container sequence with eight degenerations, followed by primer contig 6a with seven and primer contig 4 with six and etcetera. The length of all primer contigs covered 388 nucleotides of the gene, which corresponded to 25% of the molecule size. Herein, 223 12-mers were generated; however this number increased to 2,886 after considering non-degenerated bases containing isomers (Table 2).

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

Name	Sequence	Location	References
1	AGAGTTTGATYMTGGCTCAG	8–27	(*Edwards et al., 1989*; *Isenbarger et al., 2008*; *Kumar et al., 2005*; *Ludwig, Mittenhuber & Friedrich, 1993*; *McInerney et al., 1995*; *Muyzer et al., 1995*; *Oguntoyinbo, 2007*; *Suzuki & Giovannoni, 1996*; *Wilson, Blitchington & Greene, 1990*)
2	ASYGGCGNACGGGTGAGTAA	100–119	(*Schmalenberger, Schwieger & Tebbe, 2001*; *Sundquist et al., 2007*)
3	ACTGAGAYACGGYCCARACTCCTACGGRNGGCNGCAGTRRGGAA	320–363	(*Amann et al., 1990*; *Baker, Smith & Cowan, 2003*; *Chakravorty et al., 2007*; *Dethlefsen et al., 2008*; *Fierer et al., 2008*; *Hang et al., 2014*; *Hansen et al., 1998*; *Lane, 1991*; *Muyzer, De Waal & Uitterlinden, 1993*; *Reysenbach & Pace, 1995*; *Rudi et al., 1997*; *Walter et al., 2000*; *Wang & Qian, 2009*; *Watanabe, Kodama & Harayama, 2001*; *Wuyts et al., 2002*)
4	GGCTAACTHCGTGNCVGCNGCYGCGGTAANAC	504–535	(*Cho et al., 1996*; *DasSarma & Fleischmann, 1995*; *Kumar et al., 2011*; *Lane, 1991*; *Liu et al., 2007*; *Makemson et al., 1997*; *Muyzer, De Waal & Uitterlinden, 1993*; *Nelson et al., 2014*; *Ovreås et al., 1997*; *Quince et al., 2011*; *Reysenbach & Pace, 1995*; *Ruff-Roberts, Kuenen & Ward, 1994*; *Walter et al., 2000*; *Wang & Qian, 2009*; *Wuyts et al., 2002*)
5a	GTGTAGMGGTGAAATKCGTAGAT	682–704	(*Klijn, Weerkamp & De Vos, 1991*; *Stults et al., 2001*; *Wang & Qian, 2009*)
5b	CAAACRGGATTAGAWACCCNNGTAGTCCACGC	778–809	(*Baker, Smith & Cowan, 2003*; *Barns et al., 1994*; *Brunk & Eis, 1998*; *Claesson et al., 2010*; *Colquhoun, 1997*; *Cruaud et al., 2014*; *DasSarma & Fleischmann, 1995*; *Engelbrektson et al., 2010*; *Flores et al., 2011*; *Lane, 1991*; *Liu et al., 2007*; *McBain et al., 2003*; *Nelson et al., 2014*; *Roesch et al., 2007*; *Sakai et al., 2004*; *Teske & Sorensen, 2007*; *Tremblay et al., 2015*; *Wang & Qian, 2009*; *Weisburg et al., 1991*; *Wilson, Blitchington & Greene, 1990*)
6a	AAANTYAAANRAATWGRCGGGGRCCCGCACAAG	906–938	(*Amann et al., 1992*; *Casamayor et al., 2002*; *Henckel, Friedrich & Conrad, 1999*; *Jurgens, Lindström & Saano, 1997*; *Lane et al., 1985*; *Liu et al., 2007*; *Mao et al., 2012*; *Nelson et al., 2014*; *Quince et al., 2011*; *Reysenbach & Pace, 1995*; *Rudi et al., 1997*; *Tremblay et al., 2015*; *Wang & Qian, 2009*; *Watanabe, Kodama & Harayama, 2001*)
6b	ATGTGGTTTAATTCGA	948–963	(*Iwamoto et al., 2000*; *Wang & Qian, 2009*)
6c	CAACGCGARGAACCTTACC	966–984	(*Jonasson, Olofsson & Monstein, 2002*; *Nübel et al., 1996*; *Sogin et al., 2006*; *Wang & Qian, 2009*)

Table 1 (*continued*)

Name	Sequence	Location	References
7a	AGGTGNTGCATGGYYGYCGTCAGCTCGTGYCGTGAG	1045–1080	(*DasSarma & Fleischmann, 1995*; *Ferris, Muyzer & Ward, 1996*; *Huber et al., 2007*; *Jonasson, Olofsson & Monstein, 2002*; *Sogin et al., 2006*; *Wang & Qian, 2009*; *Youssef et al., 2009*)
7b	TGTTGGGTTAAGTCCCRYAACGAGCGCAACCCT	1082–1114	(*Cruaud et al., 2014*; *Ovreås et al., 1997*; *Reysenbach & Pace, 1995*; *Stackebrandt & Goodfellow, 1991*; *Wang & Qian, 2009*; *Wuyts et al., 2002*)
8a	GGAAGGYGGGGAYGACG	1176–1192	(*Bodenhausen, Horton & Bergelson, 2013*; *Wang & Qian, 2009*)
8b	GGGCKACACACGYGCTAC	1219–1236	(*DasSarma & Fleischmann, 1995*; *Youssef et al., 2009*)
9	GCCTTGYACWCWCCGCCCGTC	1386–1406	(*Kunin et al., 2010*; *Lane, 1991*; *Lane et al., 1985*; *Mao et al., 2012*; *Marchesi et al., 1998*; *Nübel et al., 1996*; *Reysenbach & Pace, 1995*; *Tremblay et al., 2015*; *Wuyts et al., 2002*; *Youssef et al., 2009*; *Yu & Morrison, 2004*)
10	GGGTGAAGTCRTAACAAGGTANCC	1486–1509	(*Hang et al., 2014*; *Isenbarger et al., 2008*; *Lin & Stahl, 1995*; *Oguntoyinbo, 2007*; *Reysenbach et al., 1992*; *Roesch et al., 2007*; *Weisburg et al., 1991*; *Wilson, Blitchington & Greene, 1990*)

Table 2 **Characteristics of primer contigs (Table 1) obtained after assembling all of the primers reported for each conserved region of the 16S rRNA gene.** The number of possible 12-mers is size-dependent, while the number of isomers is related to the number of degenerated bases.

Name	Length	Degenerated bases	Number of 12-mers	Number of Iso 12-mers
1	20	2	9	36
2	20	3	9	61
3	44	8	33	488
4	32	6	21	970
5a	23	2	12	30
5b	32	4	21	306
6a	33	7	22	602
6b	16	0	5	5
6c	19	1	8	16
7a	36	5	25	167
7b	33	2	22	57
8a	17	2	6	22
8b	18	2	7	22
9	21	3	10	68
10	24	2	13	36
Total	388	49	223	2,886

Table 3 12-mers registering the highest frequency within each primer contig.

Primer contig	12-mer		Frequency	
	Number	Sequence	Number	Percent
01	8	ATYMTGGCTCAG	195,901	38.16%
02	1	SYGGCGNACGGG	405,570	79.01%
03	25	GGRNGGCNGCAG	500,253	97.46%
04	14	CVGCNGCYGCGG	496,412	96.71%
5a	5	GMGGTGAAATKC	382,156	74.45%
5b	10	TAGAWACCCNNG	493,348	96.11%
6a	10	RAATWGRCGGGG	501,792	97.76%
6b	2	GTGGTTTAATTC	389,530	75.89%
6c	6	GARGAACCTTAC	393,614	76.68%
7a	12	GYYGYCGTCAGC	499,976	97.40%
7b	20	CGAGCGCAACCC	489,290	95.32%
8a	3	AGGYGGGGAYGA	454,807	88.60%
8b	2	GCKACACACGYG	382,857	74.59%
09	5	GYACWCWCCGCC	388,911	75.77%
10	6	AGTCRTAACAAG	172,918	33.69%

12-mers search

When all iso12-mers, from each primer contig were searched in the sequences deposited in SILVA database release 123, highly variable coverages were revealed, which may explain in part the different results and PCR biases reported in these kinds of studies. Frequency analysis showed that extreme regions, 1 and 10, registered lower frequencies; for example, iso12-mers of these primer contigs were detected in less than 40% of the +513,000 sequences of SILVA (Table 3). These low frequencies detected at the extreme regions could be due to the interest focused on the central 16S rRNA regions. However, the most important factor is the absence of end segments (200–300 nucleotides) detected in several sequences of the database. This has been also reported in other studies (*Wang et al., 2016*; *Yarza et al., 2014*).

Regarding more commonly used regions, the first 12-mers of primer contig 3 registered low frequencies (~80%); however, this value raised up to 90% from position 16 and forward, reaching values of 97.5% (24th 12-kmer) (Fig. 2). A similar pattern was observed for primer contig 4 with a detection frequency of 78.5% and a progressive increase in forward direction, reaching values of 96.7% (13th 12-kmer). From the two primers contigs assembled in region 5, the 5a (positions 682 to 704) exhibited low frequencies for all 12-mers (60–80%); whereas higher frequencies were recorded for 5b (up to 96.1% in 9th 12-kmer) (Fig. 2). Regarding primer contigs from region 6 (6a, 6b and 6c), the highest frequency was detected for 12-mers of primer contig 6a, while none of 6b and 6c reached 80%. Two primer contigs were also detected for region 7 (7a and 7b), with the highest 12-kmer frequencies reported for 7a (95–97%) in the first segment of the molecule; however a frequency decrease was detected from 16th 12-kmer and forward; whereas the

Martinez-Porchas et al. (2017), *PeerJ*, DOI 10.7717/peerj.3036

7/20
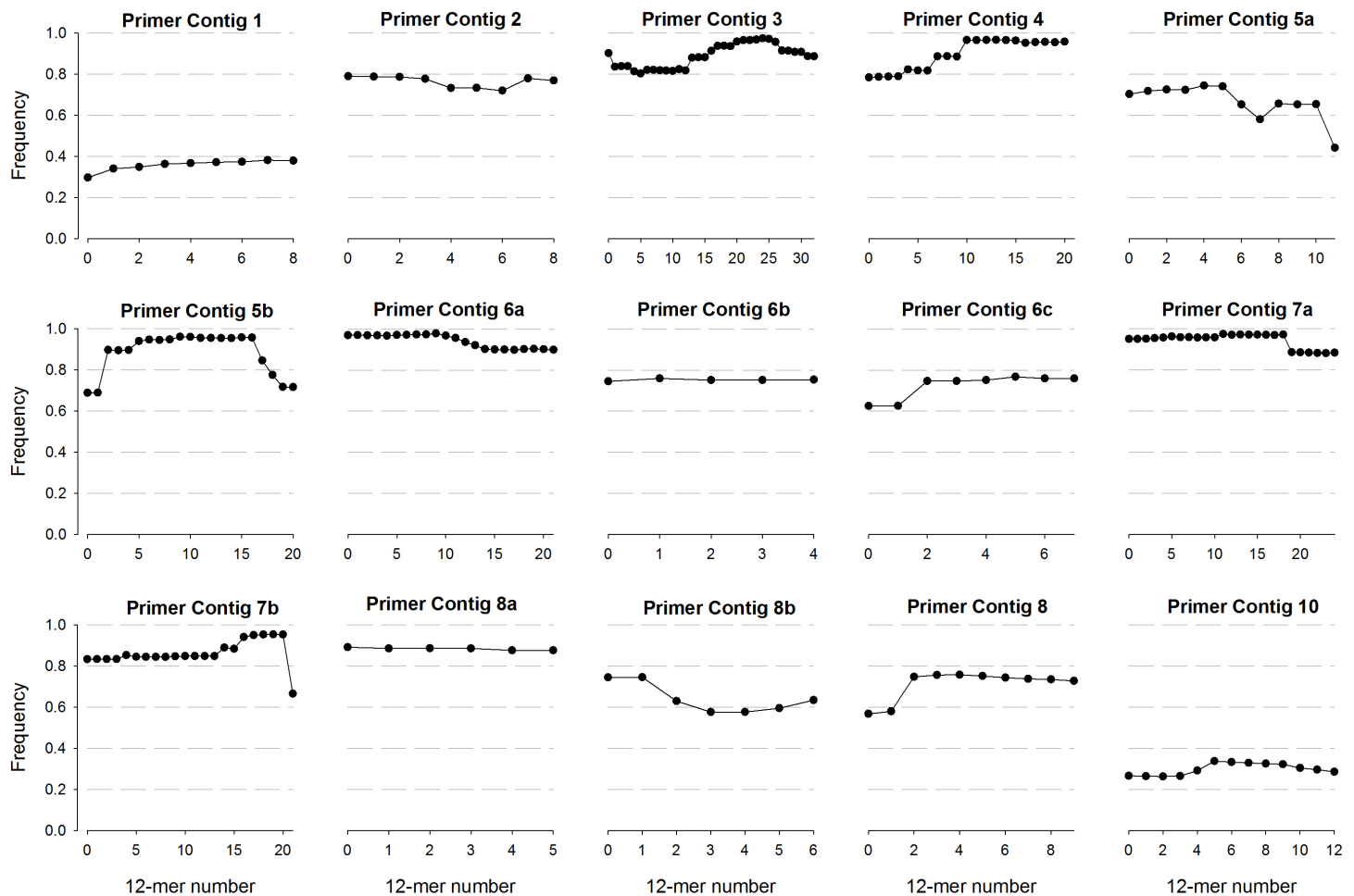
Figure 2 Frequency of 12-mers detected in contigs recovered from the different conserved regions of the 16S rRNA gene. More than one contig were recovered from a single conserved region.

highest frequency detected for 7b was 95.3% (Fig. 2). Finally, primer contigs 2, 8a, 8b and 9 registered low frequency values, ranging from 60 to 85%.

Thus, considerable coverage variability was observed in conserved regions located at the extremes of the 16S rRNA gene (1 and 2, as well as 8, 9 and 10) where none of the 12-mers reached a frequency higher than 85%. These results could call into question the suitability of some of the primers that have been used for long time. In spite of these variabilities, 12-mers frequency analysis revealed that there are yet particular segments within each region with acceptable conservation degree to be considered for the study of prokaryotic diversity (Fig. 2). In this regard, 12-mers covering segments of regions 3, 4, 5b and 6a registered the highest frequencies, which could be useful information to design primers that are more suitable to profiling and comparing microbial communities; however, additional considerations have to be taken into account to design primers (*Wang & Qian, 2009*).

**Table 4 Non-redundant (NR) sequences detected in consensuses corresponding to primer contigs.** The last two columns indicate how many of NR sequences required to reach a coverage of 95% of corresponding fragment recovered from SILVA database.

Consensus	NR sequences		NR sequences needed to cover 95% of all fragments recovered	
	Number	Percent	Number	Percent
Consensus 2	1,844	0.45%	24	1.30%
Consensus 3	11,586	2.32%	498	4.30%
Consensus 4	4,323	0.87%	30	0.69%
Consensus 5b	6,694	1.36%	89	1.33%
Consensus 6a	5,301	1.06%	42	0.79%
Consensus 7a	4,312	0.86%	26	0.60%
Consensus 8a	746	0.16%	4	0.54%
Consensus 9	1,078	0.28%	9	0.83%

CONSENSUS FRAGMENTS ANALYSIS

In order to define the conservative sequences, each region corresponding to primer contig was recovered from more than 513,000 sequences in the SILVA database. The analysis was done for contigs 3, 4, 5b and 6a, because these regions constitute the most used target for primer design in 16S rRNA gene considered for the study of prokaryotic diversity. Around 500,000 fragments were recovered for each primer contig region and, after being manually aligned, the consensuses sized equal to the corresponding primer contig (32 to 44 nucleotides) detecting several degenerated bases. The frequency of each position for each aligned sequence set was determined (Figs. 3A, 3C, 4A and 4C). However, to avoid or reduce bias, subsequent analyzes were done with non-redundant sequences. The reduction in the number of sequences was significant; for example, the largest reduction (99.84%) was obtained with consensus 8a, where the 454,807 fragments obtained were pooled into 746 NR sequences. The lowest reduction in the number of sequence, from 500,253 to 11,586 (97.68%), was obtained by grouping consensus 3 fragments (Table 4). Each NR sequence represents a different number of fragments, and in some cases a small proportion is sufficient to represent the majority of the fragments. For consensus 3, more than 11,500 NR sequences were obtained, but only 498 are required to cover 95% of all retrieved fragments; in contrast, only 4 and 9 NR sequences are required to have a coverage of 95% of the consensuses 8a and 9, respectively (Table 4).

Nucleotide frequency analysis of consensuses revealed that small segments or single nucleotide positions were far to be constant within conserved regions. For instance, from the 44-nucleotide fragment conforming consensus 3, only five nucleotides registered a frequency below 95% (Fig. 3A), which would suggest a high conservation degree. However, when redundant sequences were eliminated and only NR sequences were considered for the analysis, 35 of the 44 nucleotides resulted to have a frequency below 95% (Fig. 3B). Only a 12-nucleotide segment containing nine nucleotides with frequency ≥95% was found to be the most conserved fraction of this region. A similar decrease of conserved nucleotides was observed in the other studied fragments when non-redundant sequences
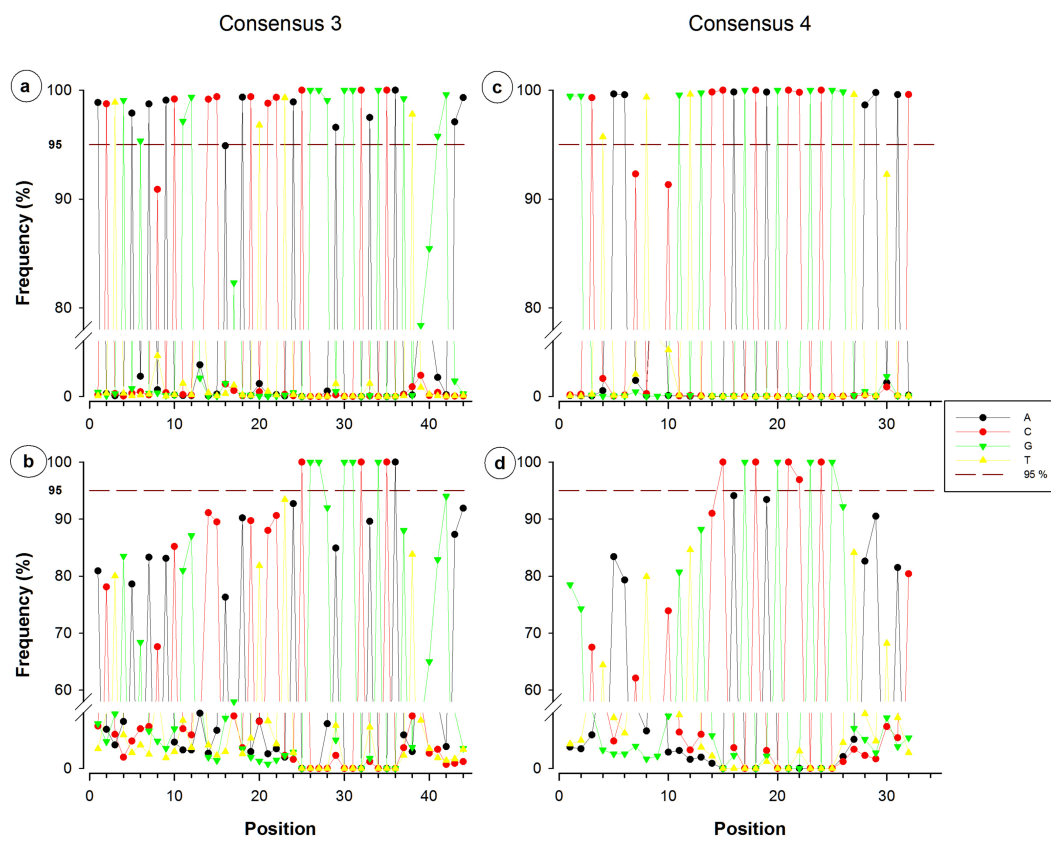
Figure 3 **Single nucleotide frequencies for the consensus corresponding to primer contigs 3 and 4.** Consensuses recovered from SILVA database are located on the upper side, whereas non-redundant sequences can be observed on the lower side. The red dotted line indicates the limit of 95%.

were considered, as shown in Figs. 3 and 4. Regarding consensus 4 sizing 32 nucleotides, an 11-nucleotide segment was detected to be the most conserved fragment with nine nucleotides registering frequencies above 95%; whereas fractions flanking this segment exhibited frequencies below 90% (Fig. 3D). A 10-nucleotide segment with ≥95% frequency was detected for consensus 5b constituted of 32 nucleotides (Fig. 4B); fractions flanking this segment registered extreme low frequency values, ranging from 10% to 90%. Similarly, a 10-nucleotide segment was found to be the most conserved fraction of consensus 6b composed by 33 nucleotides (Fig. 4D); from these 10 nucleotides, only eight resulted to be detected with a frequency above 95%, whereas nucleotides of fractions flanking the segment registered considerable variations (10–94%).

Despite all- and NR-sequences were analyzed in this trial, analyses based on NR sequences provided more realistic data, because all differences had same value; these kinds of approaches should be central to estimate how much conserved could be any region. Herein, primers complementing the conserved regions of the 16S rRNA gene of environmental prokaryotes are not necessarily complementary to all those that exist in the actual databases (*Baker, Smith & Cowan, 2003*). Degenerations have been used to include all these new sequences; however this replacement may difficult the design of
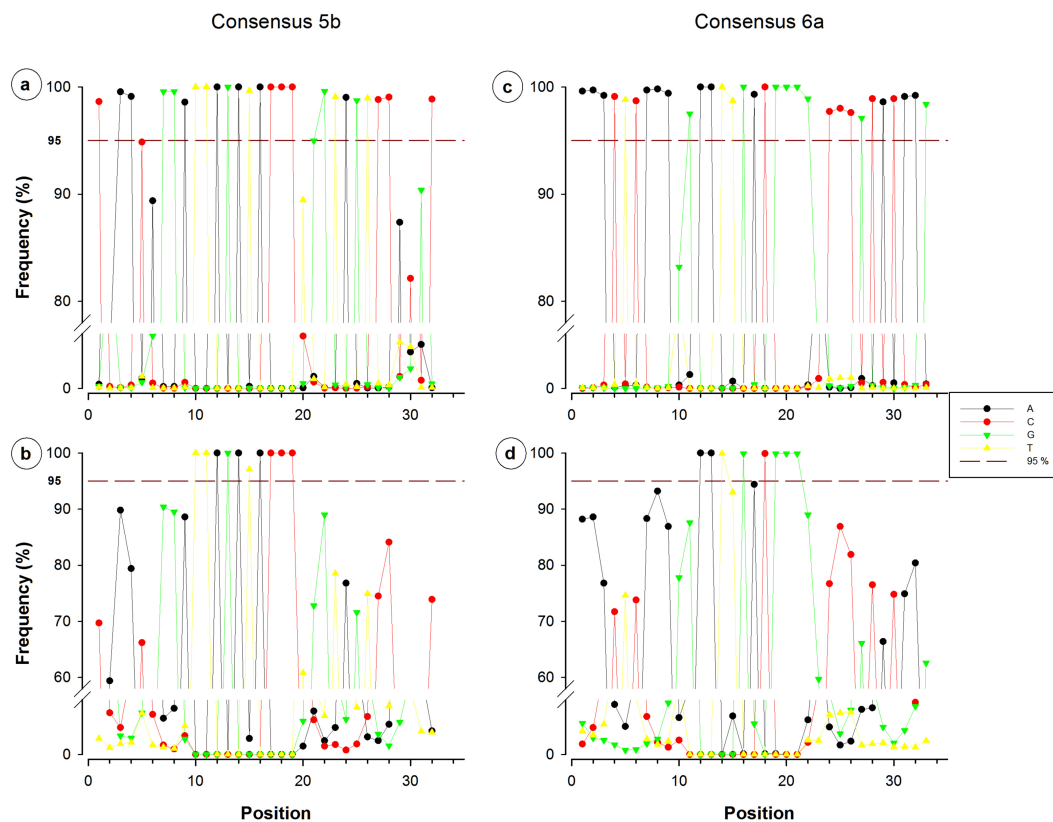
Figure 4 **Single nucleotide frequencies for the consensus corresponding to primer contig 5b and 6a, recovered from all database SILVA (upper) and NR sequences (lower).** Red dotted line indicates a limit of 95%.

adequate primers and reduces the confidence of conserved regions. Furthermore, the low iso 12-mers frequencies in SILVA sequences could be associated to a greater proportion of degenerations than those actually reported, or inclusively to the potential presence of sequencing errors; for example, any insertion or deletion of a single nucleotide cause shifting of the entire sequence and reports biased diversity.

The single nucleotide-frequency analysis provided additional information that revealed the most conserved nucleotide positions within each consensus. These results revealed that most of the positions of the conservative regions were not as conserved as expected. Herein, the sum of the consensuses covered 25% of the molecule (388 nucleotides); however, only 75 nucleotides showed frequencies of at least 95%, representing 5% of the molecule size. Such information revealed that a very small fraction 16S rRNA gene is truly conserved (≥95%); therefore, primer design must necessarily be anchored to these very short, but highly conserved segments. Furthermore, these short segments corresponded to the 12-mers that registered the highest frequencies. Table 5 shows nucleotide pattern of each of the studied regions (2, 3, 4, 5b, 6a, 7a, 8a and 9) considering the sequences of SILVA database, the sets of NR sequences, as well as primer contig for comparison.

These analyses could be also used as a methodological pathway focused to particular environments, and selecting only the inhabitants reported in specific databases, like the

Table 5 The primer contig and the corresponding consensuses from both all and non-redundant sequences are aligned. Conserved nucleotides in the consensus of all sequences are in blue, while those preserved in non-redundant sequences are in red.

2	Primer contig	ASYGGCGNACGGGTGAGTAA
	All sequences	ASYGGCGVACGGGTGMGTAA
	NR sequences	ASYGGCGVACGGBNNNNNNN
3	Primer contig	ACTGAGAYACGGYCCARACTCCTACGGRNGGCNGCAGTRRGGAA
	All sequences	ACTGAGAYACGGHCCRRACTCCTACGGGAGGCAGCAGTVRGGAA
	NR sequences	VHBDVVVHVVBBNYMVNVHHYHYRCGGRDGGCWGCAVBNDVRRR
4	Primer contig	GGCTAACTHCGTGNCVGCNGCYGCGGTAANAC
	All sequences	GGCTAAYTHYGTGCCAGCAGCCGCGGTAAKAC
	NR sequences	BBNHHHHHHBBKBSCMGCMGCCGCGKDDWNHV
5b	Primer contig	CAAACRGGATTAGAWACCCNNGTAGTCCACGC
	All sequences	CRAAYRGGATTAGATACCCYGGTAGTCCWHRC
	NR sequences	VVVVNVRRHTTAGATACCCBNKDDDBBHNNVV
6a	Primer contig	AAANTYAAANRAATWGRCGGGGRCCCGCACAAG
	All sequences	AAACTCAAAKGAATTGACGGGGRCCCGCACAAG
	NR sequences	DHHHHHMRRDRAATWGRCGGGRVNBBVVVMVVV
7a	Primer contig	AGGTGNTGCATGGYYGYCGTCAGCTCGTGYCGTGAG
	All sequences	AGGTGSTGCATGGYTGTCGTCAGCTCGTGTCGTGAG
	NR sequences	VDDBBNBSMWYGGYYGTCGTCAGYYBBBDBBBBBDDR
8a	Primer contig	GGAAGGYGGGGAYGACG
	All sequences	GGAAGGTGGGGATGACG
	NR sequences	GGAAGGYGGGGANNNNN
9	Primer contig	GCCTTGYACWCWCCGCCCGTC
	All sequences	GYCTTGYACWCACCGCCCGTC
	NR sequences	NNNBTGYACWCWCCGCHNNNN

marine waters with their particular microorganisms, or exclusive microbial pathogens that require more specific discrimination. It is important to consider that only fractions of the entire bacteria species could be detected in different environments, and that from such diversity some kinds of bacteria would have a greater representation than others; for this reason the use of NR sequences improves coverage of bacteria thriving into random environments. Finally, it could be concluded that conserved regions of the 16S rRNA gene exhibit considerable variations; however, it was demonstrated that it is possible to achieve more reliable primers designs.

ADDITIONAL INFORMATION AND DECLARATIONS

Author Contributions

- Marcel Martinez-Porchas conceived and designed the experiments, wrote the paper, reviewed drafts of the paper.
- Enrique Villalpando-Canchola conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, reviewed drafts of the paper.
- Luis Enrique Ortiz Suarez performed the experiments, contributed reagents/materials/-analysis tools, prepared figures and/or tables, reviewed drafts of the paper.
- Francisco Vargas-Albores conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.

Data Availability

The following information was supplied regarding data availability:
The raw data has been supplied as a Supplementary File.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.3036#supplemental-information.

REFERENCES

Albertsen M, Karst SM, Ziegler AS, Kirkegaard RH, Nielsen PH. 2015. Back to basics—the influence of DNA extraction and primer choice on phylogenetic analysis of activated sludge communities. *PLOS ONE* **10(7)**:e0132783 DOI 10.1371/journal.pone.0132783.

Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA. 1990. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Applied and Environmental Microbiology* **56**:1919–1925.

Amann RI, Stromley J, Devereux R, Key R, Stahl DA. 1992. Molecular and microscopic identification of sulfate-reducing bacteria in multispecies biofilms. *Applied and Environmental Microbiology* **58**:614–623.

Baker GC, Smith JJ, Cowan DA. 2003. Review and re-analysis of domain-specific 16S primers. *Journal of Microbiological Methods* **55(3)**:541–555 DOI 10.1016/j.mimet.2003.08.009.

Barns SM, Fundyga RE, Jeffries MW, Pace NR. 1994. Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proceedings of the National Academy of Sciences of the United States of America* **91**:1609–1613 DOI 10.1073/pnas.91.5.1609.

Bodenhausen N, Horton MW, Bergelson J. 2013. Bacterial communities associated with the leaves and the roots of *Arabidopsis thaliana*. *PLOS ONE* **8(2)**:e56329 DOI 10.1371/journal.pone.0056329.

Brunk CF, Eis N. 1998. Quantitative measure of small-subunit rRNA gene sequences of the kingdom Korarchaeota. *Applied and Environmental Microbiology* **64**:5064–5066.

Casamayor EO, Massana R, Benlloch S, Ovreas L, Diez B, Goddard VJ, Gasol JM, Joint I, Rodriguez-Valera F, Pedros-Alio C. 2002. Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern. *Environmental Microbiology* **4(6)**:338–348 DOI 10.1046/j.1462-2920.2002.00297.x.

Chakravorty S, Helb D, Burday M, Connell N, Alland D. 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiological Methods* **69**(2):330–339 DOI 10.1016/j.mimet.2007.02.005.

Cho JC, Lee DH, Cho YC, Cho JC, Kim SJ. 1996. Direct extraction of dna from soil for amplification of 16s rrna gene sequences by polymerase chain reaction. *Journal of Microbiology* **34**:229–235.

Claesson MJ, Wang Q, O'Sullivan O, Greene-Diniz R, Cole JR, Ross RP, O'Toole PW. 2010. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Research* **38(22)**:e200 DOI 10.1093/nar/gkq873.

Colquhoun JA. 1997. Discovery of deep-sea actinomycetes. PhD Dissertation, School of Biosciences, University of Kent, Canterbury UK.

Cruaud P, Vigneron A, Lucchetti-Miganeh C, Ciron PE, Godfroy A, Cambon-Bonavita M-A. 2014. Influence of DNA extraction method, 16s rrna targeted hypervariable regions, and sample origin on microbial diversity detected by 454 pyrosequencing in marine chemosynthetic ecosystems. *Applied and Environmental Microbiology* **80(15)**:4626–4639 DOI 10.1128/AEM.00592-14.

DasSarma S, Fleischmann EF. 1995. *Archaea: a laboratory manual—halophiles.* New York: Cold Spring Harbour Laboratory Press.

Dethlefsen L, Huse S, Sogin ML, Relman DA. 2008. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLOS Biology* **6(11)**:e280 DOI 10.1371/journal.pbio.0060280.

Edwards U, Rogall T, Blöcker H, Emde M, Böttger EC. 1989. Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Research* **17(19)**:7843–7853 DOI 10.1093/nar/17.19.7843.

Engelbrektson A, Kunin V, Wrighton KC, Zvenigorodsky N, Chen F, Ochman H, Hugenholtz P. 2010. Experimental factors affecting PCR-based estimates of

microbial species richness and evenness. *ISME Journal* **4**:642–647
DOI 10.1038/ismej.2009.153.

Ferris MJ, Muyzer G, Ward DM. 1996. Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community. *Applied and Environmental Microbiology* **62**:340–346.

Fierer N, Hamady M, Lauber CL, Knight R. 2008. The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **105**:17994–17999
DOI 10.1073/pnas.0807920105.

Flores GE, Campbell JH, Kirshtein JD, Meneghin J, Podar M, Steinberg JI, Seewald JS, Tivey MK, Voytek MA, Yang ZK, Reysenbach AL. 2011. Microbial community structure of hydrothermal deposits from geochemically different vent fields along the Mid-Atlantic Ridge. *Environmental Microbiology* **13**:2158–2171
DOI 10.1111/j.1462-2920.2011.02463.x.

Hang J, Desai V, Zavaljevski N, Yang Y, Lin X, Satya RV, Martinez LJ, Blaylock JM, Jarman RG, Thomas SJ, Kuschner RA. 2014. 16S rRNA gene pyrosequencing of reference and clinical samples and investigation of the temperature stability of microbiome profiles. *Microbiome* **2**: Article 31 DOI 10.1186/2049-2618-2-31.

Hansen MC, Tolker-Nielsen T, Givskov M, Molin S. 1998. Biased 16S rDNA PCR amplification caused by interference from DNA flanking the template region. *FEMS Microbiology Ecology* **26(2)**:141–149 DOI 10.1111/j.1574-6941.1998.tb00500.x.

Henckel T, Friedrich M, Conrad R. 1999. 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. *Applied and Environmental Microbiology* **65**:1980–1990.

Hiergeist A, Reischl U, Gessner A. 2016. Multicenter quality assessment of 16S ribosomal DNA-sequencing for microbiome analyses reveals high inter-center variability. *International Journal of Medical Microbiology* **306(5)**:334–342
DOI 10.1016/j.ijmm.2016.03.005.

Huber JA, Mark Welch DB, Morrison HG, Huse SM, Neal PR, Butterfield DA, Sogin ML. 2007. Microbial population structures in the deep marine biosphere. *Science* **318(5847)**:97–100 DOI 10.1126/science.1146689.

Isenbarger TA, Finney M, Ríos-Velázquez C, Handelsman J, Ruvkun G. 2008. Miniprimer PCR, a new lens for viewing the microbial world. *Applied and Environmental Microbiology* **74**:840–849 DOI 10.1128/AEM.01933-07.

Iwamoto T, Tani K, Nakamura K, Suzuki Y, Kitagawa M, Eguchi M, Nasu M. 2000. Monitoring impact of *in situ* biostimulation treatment on groundwater bacterial community by DGGE. *FEMS Microbiology Ecology* **32**:129–141
DOI 10.1111/j.1574-6941.2000.tb00707.x.

Jonasson J, Olofsson M, Monstein H-J. 2002. Classification, identification and subtyping of bacteria based on pyrosequencing and signature matching of 16s rDNA fragments. *APMIS* **110(3)**:263–272 DOI 10.1034/j.1600-0463.2002.100309.x.

Jurgens G, Lindström K, Saano A. 1997. Novel group within the kingdom Crenarchaeota from boreal forest soil. *Applied and Environmental Microbiology* **63**:803–805.

Klijn N, Weerkamp AH, De Vos WM. 1991. Identification of mesophilic lactic acid bacteria by using polymerase chain reaction-amplified variable regions of 16S rRNA and specific DNA probes. *Applied and Environmental Microbiology* **57**:3390–3393.

Kumar PS, Brooker MR, Dowd SE, Camerlengo T. 2011. Target region selection is a critical determinant of community fingerprints generated by 16S pyrosequencing. *PLOS ONE* **6(6)**:e20956 DOI 10.1371/journal.pone.0020956.

Kumar PS, Griffen AL, Moeschberger ML, Leys EJ. 2005. Identification of candidate periodontal pathogens and beneficial species by quantitative 16S clonal analysis. *Journal of Clinical Microbiology* **43**:3944–3955 DOI 10.1128/JCM.43.8.3944-3955.2005.

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology* **12(1)**:118–123 DOI 10.1111/j.1462-2920.2009.02051.x.

Lane DJ. 1991. 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M, eds. *Nucleic acid techniques in bacterial systematics.* Chichester: John Wiley & Sons, 115–175.

Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences of the United States of America* **82(20)**:6955–6959 DOI 10.1073/pnas.82.20.6955.

Li H, Zhang Y, Li D-S, Xu H, Chen G-X, Zhang C-G. 2009. Comparisons of different hypervariable regions of rrs genes for fingerprinting of microbial communities in paddy soils. *Soil Biology and Biochemistry* **41(5)**:954–968 DOI 10.1016/j.soilbio.2008.10.030.

Lin C, Stahl DA. 1995. Taxon-specific probes for the cellulolytic genus *Fibrobacter* reveal abundant and novel equine-associated populations. *Applied and Environmental Microbiology* **61**:1348–1351.

Linhart C, Shamir R. 2002. The degenerate primer design problem. *Bioinformatics* **18**:S172–S180.

Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R. 2007. Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Research* **35(18)**:e120 DOI 10.1093/nar/gkm541.

Logares R, Sunagawa S, Salazar G, Cornejo-Castillo FM, Ferrera I, Sarmento H, Hingamp P, Ogata H, De Vargas C, Lima-Mendez G, Raes J, Poulain J, Jaillon O, Wincker P, Kandels-Lewis S, Karsenti E, Bork P, Acinas SG. 2014. Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environmental Microbiology* **16(9)**:2659–2671 DOI 10.1111/1462-2920.12250.

Ludwig W, Mittenhuber G, Friedrich CG. 1993. Transfer of *Thiosphaera pantotropha* to *Paracoccus denitrificans*. *International Journal of Systematic Bacteriology* **43**:363–367 DOI 10.1099/00207713-43-2-363.

Makemson JC, Fulayfil NR, Landry W, Van Ert LM, Wimpee CF, Widder EA, Case JF. 1997. *Shewanella woodyi* sp. nov., an exclusively respiratory luminous bacterium

isolated from the Alboran Sea. *International Journal of Systematic Bacteriology* **47**:1034–1039 DOI 10.1099/00207713-47-4-1034.

Mao D-P, Zhou Q, Chen C-Y, Quan Z-X. 2012. Coverage evaluation of universal bacterial primers using the metagenomic datasets. *BMC Microbiology* **12**:66 DOI 10.1186/1471-2180-12-66.

Marchesi JR, Sato T, Weightman AJ, Martin TA, Fry JC, Hiom SJ, Dymock D, Wade WG. 1998. Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. *Applied and Environmental Microbiology* **64**:795–799.

Martínez-Porchas M, Vargas-Albores F. 2015. Microbial metagenomics in aquaculture: a potential tool for a deeper insight into the activity. *Reviews in Aquaculture* Epub ahead of print July 6 2015 DOI 10.1111/raq.12102.

Martinez-Porchas M, Villalpando-Canchola E, Vargas-Albores F. 2016. Significant loss of sensitivity and specificity in the taxonomic classification occurs when short 16S rRNA gene sequences are used. *Heliyon* In Press.

McBain AJ, Bartolo RG, Catrenich CE, Charbonneau D, Ledder RG, Rickard AH, Symmons SA, Gilbert P. 2003. Microbial characterization of biofilms in domestic drains and the establishment of stable biofilm microcosms. *Applied and Environmental Microbiology* **69**:177–185 DOI 10.1128/AEM.69.1.177-185.2003.

McInerney JO, Wilkinson M, Patching JW, Embley TM, Powell R. 1995. Recovery and phylogenetic analysis of novel archaeal rRNA sequences from a deep-sea deposit feeder. *Applied and Environmental Microbiology* **61**:1646–1648.

Morales SE, Holben WE. 2009. Empirical testing of 16S rRNA gene PCR primer pairs reveals variance in target specificity and efficacy not suggested by *in silico* analysis. *Applied and Environmental Microbiology* **75**:2677–2683 DOI 10.1128/AEM.02166-08.

Muyzer G, De Waal EC, Uitterlinden AG. 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology* **59**:695–700.

Muyzer G, Teske A, Wirsen CO, Jannasch HW. 1995. Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Archives of Microbiology* **164**(3):165–172 DOI 10.1007/BF02529967.

Najafabadi HS, Torabi N, Chamankhah M. 2008. Designing multiple degenerate primers via consecutive pairwise alignments. *BMC Bioinformatics* **9**:55 DOI 10.1186/1471-2105-9-55.

Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J. 2014. Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. *PLOS ONE* **9**(4):e94249 DOI 10.1371/journal.pone.0094249.

Nübel U, Engelen B, Felske A, Snaidr J, Wieshuber A, Amann RI, Ludwig W, Backhaus H. 1996. Sequence heterogeneities of genes encoding 16S rRNAs in *Paenibacillus polymyxa* detected by temperature gradient gel electrophoresis. *Journal of Bacteriology* **178**:5636–5643 DOI 10.1128/jb.178.19.5636-5643.1996.

Oguntoyinbo FA. 2007. Monitoring of marine *Bacillus* diversity among the bacteria community of sea water. *African Journal of Biotechnology* **6**:163–166.

Ovreås L, Forney L, Daae FL, Torsvik V. 1997. Distribution of bacterioplankton in meromictic Lake Saelenvannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. *Applied and Environmental Microbiology* **63**:3367–3373.

Peabody MA, Van Rossum T, Lo R, Brinkman FSL. 2015. Evaluation of shotgun metagenomics sequence classification methods using *in silico* and *in vitro* simulated communities. *BMC Bioinformatics* **16(8)**:363 DOI 10.1186/s12859-015-0788-5.

Pfeiffer S, Pastar M, Mitter B, Lippert K, Hackl E, Lojan P, Oswald A, Sessitsch A. 2014. Improved group-specific primers based on the full SILVA 16S rRNA gene reference database. *Environmental Microbiology* **16**:2389–2407 DOI 10.1111/1462-2920.12350.

Qu W, Zhou Y, Zhang Y, Lu Y, Wang X, Zhao D, Yang Y, Zhang C. 2012. MFEprimer-2.0: a fast thermodynamics-based program for checking PCR primer specificity. *Nucleic Acids Research* **40(W1)**:W205–W208 DOI 10.1093/nar/gks552.

Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ. 2011. Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* **12**:38 DOI 10.1186/1471-2105-12-38.

Reysenbach AL, Giver LJ, Wickham GS, Pace NR. 1992. Differential amplification of rRNA genes by polymerase chain reaction. *Applied and Environmental Microbiology* **58**:3417–3418.

Reysenbach AL, Pace B. 1995. Robb FT, Place AR, eds. *Archaea: a laboratory manual–thermophiles.* New York: Cold Spring Harbour Laboratory Press, 101–107.

Roesch LFW, Fulthorpe RR, Riva A, Casella G, Hadwin AKM, Kent AD, Daroub SH, Camargo FAO, Farmerie WG, Triplett EW. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME Journal* **1**:283–290 DOI 10.1038/ismej.2007.53.

Rudi K, Skulberg OM, Larsen F, Jakobsen KS. 1997. Strain characterization and classification of oxyphotobacteria in clone cultures on the basis of 16S rRNA sequences from the variable regions V6, V7, and V8. *Applied and Environmental Microbiology* **63**:2593–2599.

Ruff-Roberts AL, Kuenen JG, Ward DM. 1994. Distribution of cultivated and uncultivated cyanobacteria and *Chloroflexus*-like bacteria in hot spring microbial mats. *Applied and Environmental Microbiology* **60**:697–704.

Sakai M, Matsuka A, Komura T, Kanazawa S. 2004. Application of a new PCR primer for terminal restriction fragment length polymorphism analysis of the bacterial communities in plant roots. *Journal of Microbiological Methods* **59(1)**:81–89 DOI 10.1016/j.mimet.2004.06.005.

Schloss PD, Westcott SL, Jenior ML, Highlander SK. 2016. Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. *PeerJ* **4**:e1869 DOI 10.7287/peerj.preprints.778v1.

Schmalenberger A, Schwieger F, Tebbe CC. 2001. Effect of primers hybridizing to different evolutionarily conserved regions of the small-subunit rRNA gene in

PCR-based microbial community analyses and genetic profiling. *Applied and Environmental Microbiology* **67**:3557–3563 DOI 10.1128/AEM.67.8.3557-3563.2001.

Singer E, Bushnell B, Coleman-Derr D, Bowman B, Bowers RM, Levy A, Gies EA, Cheng J-F, Copeland A, Klenk H-P, Hallam SJ, Hugenholtz P, Tringe SG, Woyke T. 2016. High-resolution phylogenetic microbial community profiling. *ISME Journal* **10**:2020–2032 DOI 10.1038/ismej.2015.249.

Soergel DAW, Dey N, Knight R, Brenner SE. 2012. Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *ISME Journal* **6**:1440–1444 DOI 10.1038/ismej.2011.208.

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences of the United States of America* **103**(32):12115–12120 DOI 10.1073/pnas.0605127103.

Stackebrandt E, Goodfellow M. 1991. *Nucleic acid techniques in bacterial systematics.* Chichester: John Wiley & Sons.

Stults JR, Snoeyenbos-West O, Methe B, Lovley DR, Chandler DP. 2001. Application of the 5′ fluorogenic exonuclease assay (TaqMan) for quantitative ribosomal DNA and rRNA analysis in sediments. *Applied and Environmental Microbiology* **67**:2781–2789 DOI 10.1128/AEM.67.6.2781-2789.2001.

Sundquist A, Bigdeli S, Jalili R, Druzin ML, Waller S, Pullen KM, El-Sayed YY, Taslimi MM, Batzoglou S, Ronaghi M. 2007. Bacterial flora-typing with targeted, chip-based Pyrosequencing. *BMC Microbiology* **7**:1–11 DOI 10.1186/1471-2180-7-108.

Suzuki MT, Giovannoni SJ. 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Applied and Environmental Microbiology* **62**:625–630.

Takahashi S, Tomita J, Nishioka K, Hisada T, Nishijima M. 2014. Development of a Prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using next-generation sequencing. *PLOS ONE* **9**(8):e105592 DOI 10.1371/journal.pone.0105592.

Teske A, Sorensen KB. 2007. Uncultured archaea in deep marine subsurface sediments: have we caught them all? *ISME Journal* **2**:3–18 DOI 10.1038/ismej.2007.90.

Tremblay J, Singh K, Fern A, Kirton ES, He S, Woyke T, Lee J, Chen F, Dangl JL, Tringe SG. 2015. Primer and platform effects on 16S rRNA tag sequencing. *Frontiers in Microbiology* **6**: Article 771 DOI 10.3389/fmicb.2015.00771.

Walker AW, Martin JC, Scott P, Parkhill J, Flint HJ, Scott KP. 2015. 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. *Microbiome* **3**: Article 26 DOI 10.1186/s40168-015-0087-4.

Walter J, Tannock GW, Tilsala-Timisjarvi A, Rodtong S, Loach DM, Munro K, Alatossava T. 2000. Detection and identification of gastrointestinal *Lactobacillus* species by using denaturing gradient gel electrophoresis and species-specific PCR primers. *Applied and Environmental Microbiology* **66**:297–303 DOI 10.1128/AEM.66.1.297-303.2000.

Wang Y, Qian P-Y. 2009. Conservative fragments in bacterial 16s rRNA genes and primer design for 16s ribosomal DNA amplicons in metagenomic studies. *PLOS ONE* **4(10)**:e7401 DOI 10.1371/journal.pone.0007401.

Wang S, Sun B, Tu J, Lu Z. 2016. Improving the microbial community reconstruction at the genus level by multiple 16S rRNA regions. *Journal of Theoretical Biology* **398**:1–8 DOI 10.1016/j.jtbi.2016.03.016.

Watanabe K, Kodama Y, Harayama S. 2001. Design and evaluation of PCR primers to amplify bacterial 16S ribosomal DNA fragments used for community fingerprinting. *Journal of Microbiological Methods* **44(3)**:253–262 DOI 10.1016/S0167-7012(01)00220-2.

Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 1991. 16S ribosomal DNA amplification for phylogenetic study. *Journal of Bacteriology* **173**:697–703 DOI 10.1128/jb.173.2.697-703.1991.

Wilson KH, Blitchington RB, Greene RC. 1990. Amplification of bacterial 16S ribosomal DNA with polymerase chain reaction. *Journal of Clinical Microbiology* **28**:1942–1946.

Wuyts J, Van de Peer Y, Winkelmans T, De Wachter R. 2002. The European database on small subunit ribosomal RNA. *Nucleic Acids Research* **30(1)**:183–185 DOI 10.1093/nar/30.1.183.

Yang B, Wang Y, Qian PY. 2016. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics* **17**:135 DOI 10.1186/s12859-016-0992-y.

Yarza P, Yilmaz P, Pruesse E, Glockner FO, Ludwig W, Schleifer K-H, Whitman WB, Euzeby J, Amann R, Rossello-Mora R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology* **12**:635–645 DOI 10.1038/nrmicro3330.

Youssef N, Sheik CS, Krumholz LR, Najar FZ, Roe BA, Elshahed MS. 2009. Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rrna gene-based environmental surveys. *Applied and Environmental Microbiology* **75**:5227–5236 DOI 10.1128/AEM.00592-09.

Yu Z, Morrison M. 2004. Comparisons of different hypervariable regions of rrs genes for use in fingerprinting of microbial communities by PCR-Denaturing Gradient Gel Electrophoresis. *Applied and Environmental Microbiology* **70**:4800–4806 DOI 10.1128/AEM.70.8.4800-4806.2004.