APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Mar. 1996, p. 872–879 Vol. 62, No. 3
0099-2240/96/$04.00+0

A Computer Analysis of Primer and Probe Hybridization Potential with Bacterial Small-Subunit rRNA Sequences

CLIFFORD F. BRUNK,[1,2]* ERIK AVANISS-AGHAJANI,[1] AND CLIFFORD A. BRUNK[3]

Biology Department[1] and Molecular Biology Institute,[2] University of California, Los Angeles, California 90024-1606, and Department of Computer Science, University of California, Irvine, California 92717[3]

Received 31 March 1995/Accepted 31 December 1995

Analysis of restriction fragment length polymorphism of bacterial small-subunit (SSU) rRNA sequences represents a potential means for characterizing complex bacterial populations such as those found in natural environments. In order to estimate the resolution potential of this approach, we have examined the SSU rRNA sequences in the Ribosomal Database Project bank using a computer algorithm which simulates hybridization between DNA sequences. Simulated hybridizations between a primer or probe sequence and an SSU rRNA sequence yield a value for each potential hybridization. This algorithm has been used to evaluate sites for PCR primers and hybridization probes used for classifying SSU rRNA sequences. Our analysis indicates that length variation in terminal restriction fragments of PCR products from the SSU rRNA sequences can identify a wide spectrum of bacteria. We also observe that the majority of restriction fragment length variation is the result of insertions and deletions rather than restriction site polymorphisms. This approach is also used to evaluate the relative efficiency and specificity of a number of published hybridization probes.

The survey of bacterial populations from natural samples, to date, has been a difficult and cumbersome task. Traditionally, microscopy and culturing techniques, each with severe limitations, have been used for surveying such populations. Recently, attempts have been made to characterize bacterial populations by molecular techniques (1, 5, 10, 12, 14, 16, 17, 20–23). These approaches rely on hybridization to specific DNA segments or sequence determination of conserved regions from a bacterial genome, usually after amplification of the DNA by PCR (13). Although such techniques have been demonstrated to be more sensitive and less biased than traditional approaches, they still have limited specificity or are labor-intensive. An alternative approach is to directly analyze the PCR-amplified sequences by restriction enzyme digestion (2, 3).

An ideal bacterial identification protocol should be sensitive enough to allow detection of bacteria at very low concentrations and have the ability to discriminate among a wide range of bacterial taxa, such as might be found in a sample taken from a natural environment. Sensitivity in such a protocol can be achieved by PCR amplification of virtually all of the small-subunit (SSU) rRNA genes present in the sample. Therefore, the choice of appropriate primers is of paramount importance. The primers must hybridize well to all SSU rRNA genes present so that all of the different SSU rRNA sequences are represented in the final PCR product. Simultaneously, the region between the primer sites must have sufficient sequence variability to differentiate the bacteria present in the sample.

In the protocol that we have developed, a fluorescent label is attached to one of the primers so that the PCR product is labeled at one end (2, 3). Following amplification, the PCR product is cleaved with one or more restriction enzymes and the lengths of the labeled terminal restriction fragments are determined. The length determinations are performed by an automated DNA sequencer (3). The exact lengths of the labeled restriction fragments in the range between 50 and 450 nucleotides can be readily and accurately determined.

The variation in length of labeled restriction fragments from different SSU rRNA genes is due to restriction site polymorphisms and/or the presence of insertions and deletions (indels) in the SSU rRNA genes. In our initial work it was clear that this approach could characterize a mixture with a limited number of different bacteria (2). As the number of different bacteria present in the sample increases and the mixture of taxa becomes more complex, it is uncertain whether this approach is feasible for characterization of the population.

We have examined most of the published bacterial SSU rRNA sequences to determine if this approach is capable of resolving the majority of these sequences (1, 5, 6, 9, 11, 15, 16, 19, 22). The composite of published SSU rRNA sequences provides the best available representation of the bacterial SSU rRNA genes that might be encountered in a sample from a natural source. We developed a computer algorithm that simulates hybridization between a primer or probe and an SSU rRNA sequence, yielding a value for the hybridization potential (HP) at each site. The hybridization site with the highest HP value (provided that the HP value exceeds a threshold) is the most probable site at which a given primer will hybridize. In our protocol, the PCR product is labeled at one end and cleaved with one or more restriction enzymes. The algorithm computes the length from the beginning of the labeled primer to the first occurrence of one of the designated restriction enzyme sites or to the end of the other primer if no restriction sites are present in the amplified region. For a given set of primers and a given set of restriction enzymes the algorithm yields a restriction fragment length for each taxon.

The results of our analysis indicate that this approach should readily resolve very complex bacterial populations. This approach will most probably underestimate the diversity in a complex population, but it does give an indication of that complexity. The algorithm provides an efficient means of selecting optimal primers and identifying the restriction enzyme combinations that are appropriate for characterizing a bacterial population. In addition, it is apparent that the majority of the length variation in the labeled fragments is the result of

* Corresponding author. Mailing address: Dept. of Biology, UCLA, 405 Hilgard Ave., Los Angeles, CA 90024-1606. Phone: (310) 825-3114. Fax: (310) 206-3987.

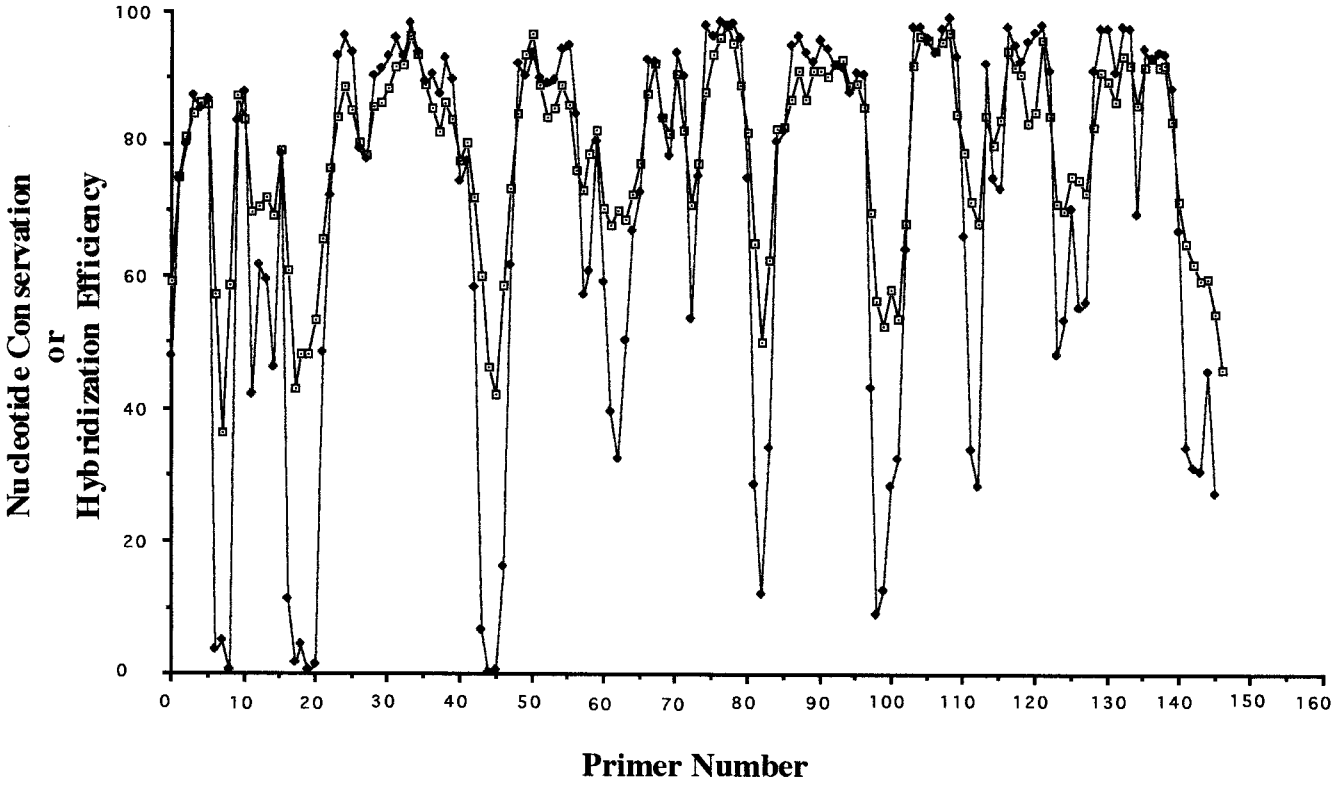

FIG. 1. Relative nucleotide conservation, *C* value (⊡), and relative efficiency of hybridization, *B* value (◆), are displayed as a function of primer number. Each primer is 20 nucleotides long.

indels rather than restriction enzyme polymorphisms. The algorithm can also evaluate the specificity of hybridization probes used to characterize organisms on the basis of SSU rRNA sequences.

MATERIALS AND METHODS

Prokaryotic and eukaryotic SSU rRNA sequences. Currently there are about 2,500 bacterial and 500 eukaryotic SSU rRNA sequences available in the Ribosomal Database Project (RDP) database (11), representing the majority of the published SSU rRNA sequences. This is a convenient source of SSU rRNA sequences because the sequences have been aligned. For our examinations we used a set of 2,250 bacterial SSU rRNA sequences, all of which were used for the probe evaluation studies; however, only 1,484 sequences were sufficiently complete to be used for the primer binding and restriction analysis study. In addition, 483 eukaryotic SSU rRNA sequences were used for probe evaluation studies. Three subsets were produced from the bacterial sequences to evaluate specific hybridization probes: a eubacterial set (1,484 sequences), an archaeal set (including crenarchaea; 83 sequences), and a crenarchaeal set (20 sequences).

Determination of a consensus sequence for the 1,484 SSU rRNA sequences. The 1,484 aligned SSU rRNA sequences contained 2,834 nucleotide positions. A computer algorithm was constructed to determine the consensus sequence that displays, for each position, the percentage of each character present (A, T, C, G, and "others"). The category "others" includes gaps used for alignment and characters used to represent nucleotides for which the sequence determination was ambiguous. The consensus sequence was constructed by choosing the most frequent nucleotide (A, T, C, or G) occurring at each position. If more than 50% of the characters at a position belonged to the category "others," that position was omitted from the consensus sequence. In cases in which the most frequent nucleotide belonged to the category "others" yet was less than 50% of the sequence, the next most frequent nucleotide was chosen to represent that position. Our final consensus sequence contained 1,485 nucleotides.

Primer selection. In selecting potential primers for PCR amplification of the bacterial SSU rRNA genes, we chose to use sequences 20 nucleotides in length. This length is long enough to provide a high degree of specificity yet short enough to be contained within a highly conserved region of the SSU rRNA sequence and to be easily synthesized.

In order to have a wide range of potential primers for the SSU rRNA gene, primers 20 nucleotides long were generated from the consensus sequence. Each consecutive primer was numbered (number 1 at the 5′ end of the SSU rRNA gene; number 146 at the 3′ end of the gene) and overlapped the previous primer by 10 nucleotides. In this manner 146 relatively independent consensus primers were generated.

The relative conservation (*C* value) of each primer was calculated by computing the average frequency of the nucleotides in the primer (18). The *C* values for the set of overlapping SSU rRNA primers are plotted in Fig. 1.

Primer binding algorithm. An effective PCR primer must hybridize efficiently to a unique site on the SSU rRNA sequence. A computer algorithm was used to estimate the efficiency, HP value, with which a primer or probe would be expected to hybridize with a region of SSU rRNA. The relative HP for the hybridization between two DNA sequences is calculated as $[2(A+T) + 3(G+C)]/L$ (4, 5, 8), where A+T is the number of the adenine-thymine pairs in the hybrid, G+C is the number of the guanine-cytosine pairs in the hybrid, and *L* is the length of the probe or primer (in nucleotides). These calculations provide only a guide to hybridization performance; however, the HP value should yield a reasonable estimate of the potential hybridization between two sequences. Indels within the hybridization region can dramatically reduce the HP value, which is similar to the effect of indels on physical hybridization.

The HP values can be computed for a primer with each potential site of hybridization in an SSU rRNA sequence. Figure 2 shows a plot of the HP values for a specific sequence (primer 52), 20 nucleotides long, as a function of SSU rRNA sequence position for *Escherichia coli*. In this case it is clear that the primer has high specificity (HP value = 2.6) for a single site (position 518), while the HP values for most sites are below 1.0.

For a specific primer sequence, the maximum HP value and the corresponding hybridization position can be determined for each of the 1,484 SSU rRNA sequences. This collection of maximum HP values is plotted as a function of sequence position in Fig. 3 for several different primers. It should be noted that nucleotide positions are the number of nucleotides from the beginning of the SSU rRNA sequence rather than the aligned sequence position, i.e., the alignment gaps have been removed. In Fig. 3a the sequence (primer 52; universal probe 3) clearly hybridizes to the vast majority of the SSU rRNA sequences at a homologous site, with maximum HP values greater than 1.8. The primer shown in Fig. 3b (primer 62) does not hybridize well to many of the SSU rRNA sequences. Most of the HP values are below 1.8, and their positions are highly variable, indicating that there is not a clear homologous site for hybridization. In Fig. 3c the primer is a random sequence with a base composition comparable to the SSU rRNA consensus sequence. Like primer 62 (Fig. 3b), this primer has low HP values and no preferred site for hybridization.

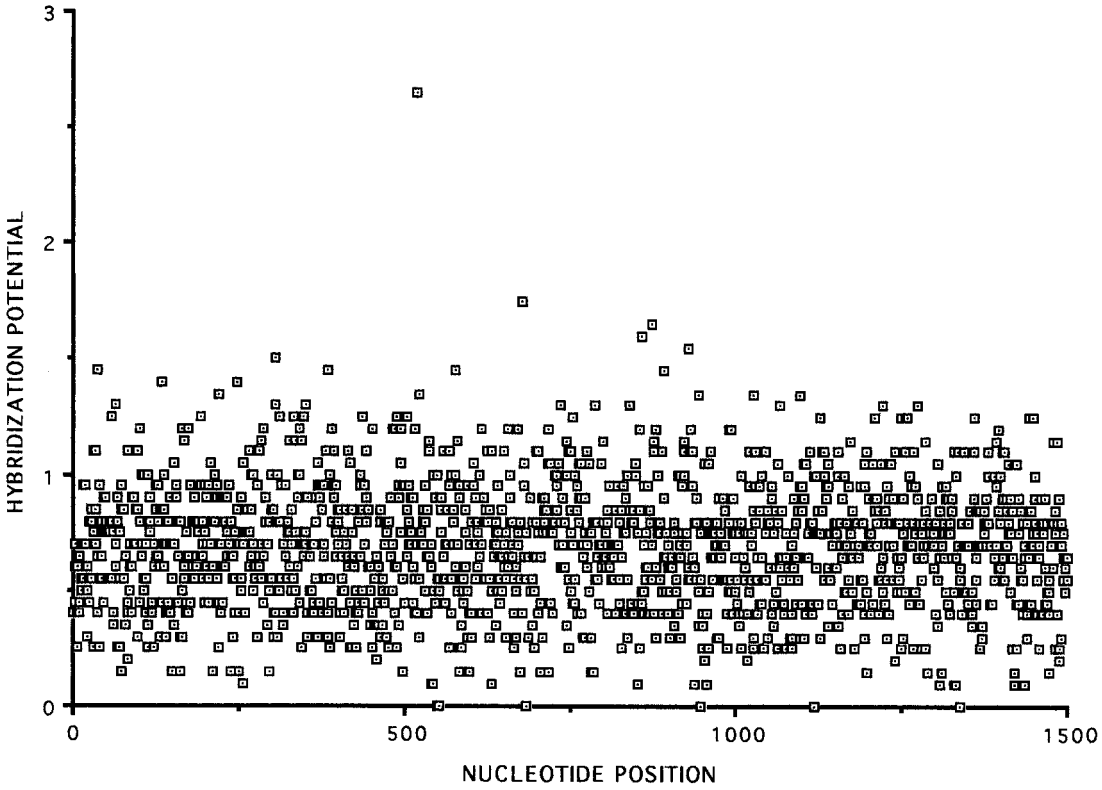

FIG. 2. HP values for primer 52 for all possible positions in the *E. coli* SSU rRNA gene.

The profile of the maximum HP values as a function of position for each of the 1,484 SSU rRNA sequences provides a valuable measure of a primer's HP. In order to compare different profiles on a quantitative basis, a *B* (box) value for a profile is calculated. The *B* value is computed as the fraction of HP values within a box above 1.8 and positioned within 200 nucleotides of the mean position for HP values. To establish the lower limit for the HP values (HP value = 1.8), the profiles for 20 random primers, like the one shown in Fig. 3c, were generated and the means and standard deviations for these HP values were computed. The lower limit (HP value = 1.8) is 2.5 standard deviations above the mean for the random primers. The *B* value is essentially the fraction of SSU rRNA sequences that hybridize well with the primer at a homologous site. The *B* values were calculated for each of the 146 consensus primers and plotted in Fig. 1.

Computation of restriction fragment length. A computer program was developed to simulate our protocol for analyzing restriction fragments from labeled PCR-amplified SSU rRNA sequences. Given two primer sequences and a set of restriction sequences, the length from the beginning of the labeled primer to the first restriction sites is calculated. If no restriction site is encountered before the second primer binding position, the length from primer to primer is returned. The set of restriction fragment lengths calculated for the different bacterial SSU rRNA sequences simulates our protocol for analyzing bacterial populations by terminally labeled restriction fragments from PCR-amplified SSU rRNA sequences (2, 3).

Analysis of hybridization probe specificity. The SSU rRNA sequences available were separated into four sets: eukaryotic (483 sequences), eubacterial (1,484 sequences), archaebacterial (83 sequences), and crenarchaebacterial (20 sequences). Three universal SSU rRNA probes, seven eubacterium-specific SSU rRNA probes, three eukaryote-specific SSU rRNA probes, four archaebacterium-specific SSU rRNA probes, one euryarchaea-specific SSU rRNA probe, and three crenarchaea-specific SSU rRNA probes were evaluated for hybridization specificity. The sequences of the probes and their hybridization positions relative to the *E. coli* SSU rRNA sequence are given in Table 1. The fraction of SSU rRNA sequences in each category with a maximum HP value of 1.8 or greater was calculated for each of the hybridization probes, and these values are shown in Table 2.

Computer algorithms. All of the computer algorithms used in this work were written in LISP and implemented on a Macintosh Quadra 840 AV computer. The programs are available upon request from C.A.B.

RESULTS AND DISCUSSION

Optimal primers for PCR amplification derived from the consensus SSU rRNA sequence can be determined easily from Fig. 1. The plots of the *C* values and the *B* values both indicate about a dozen regions that contain primer sites that meet optimization criteria. As expected, the relative conservation (*C* values) and the relative efficiency of hybridization (*B* values) parallel each other closely. The *B* values are a bit more sensitive and are considerably more appropriate in selecting optimal PCR primers. The most apparent difference between the *C* and *B* values is in the nonconserved regions such as the region between primers 18 through 22. In such regions, the *B* values are virtually zero while the *C* values dip only to about 50%. The increased sensitivity of *B* values in poor primer regions is due to the sensitivity of this calculation to the presence of indels.

When consensus primers 18 through 22 are examined, it is clear that the hybridization efficiency is low for virtually all of the SSU rRNA sequences because there are numerous variable indels in this region. Actually, a primer sequence that accommodates the most common indels will have a *B* value considerably higher (about 15%) in this region. Thus, the consensus sequence is a poor compromise in a region with many variable indels. In practice, sensitivity to variable indels is a major advantage of the *B*-value algorithm. When identification of SSU rRNA sequences on the basis of their different terminal restriction fragment lengths is used, regions of highly variable indels located between two highly conserved regions are ideal. Thus, the *B*-value plot is preferred for identifying primers and regions for analysis.

We performed a computer simulation of a terminal restric-

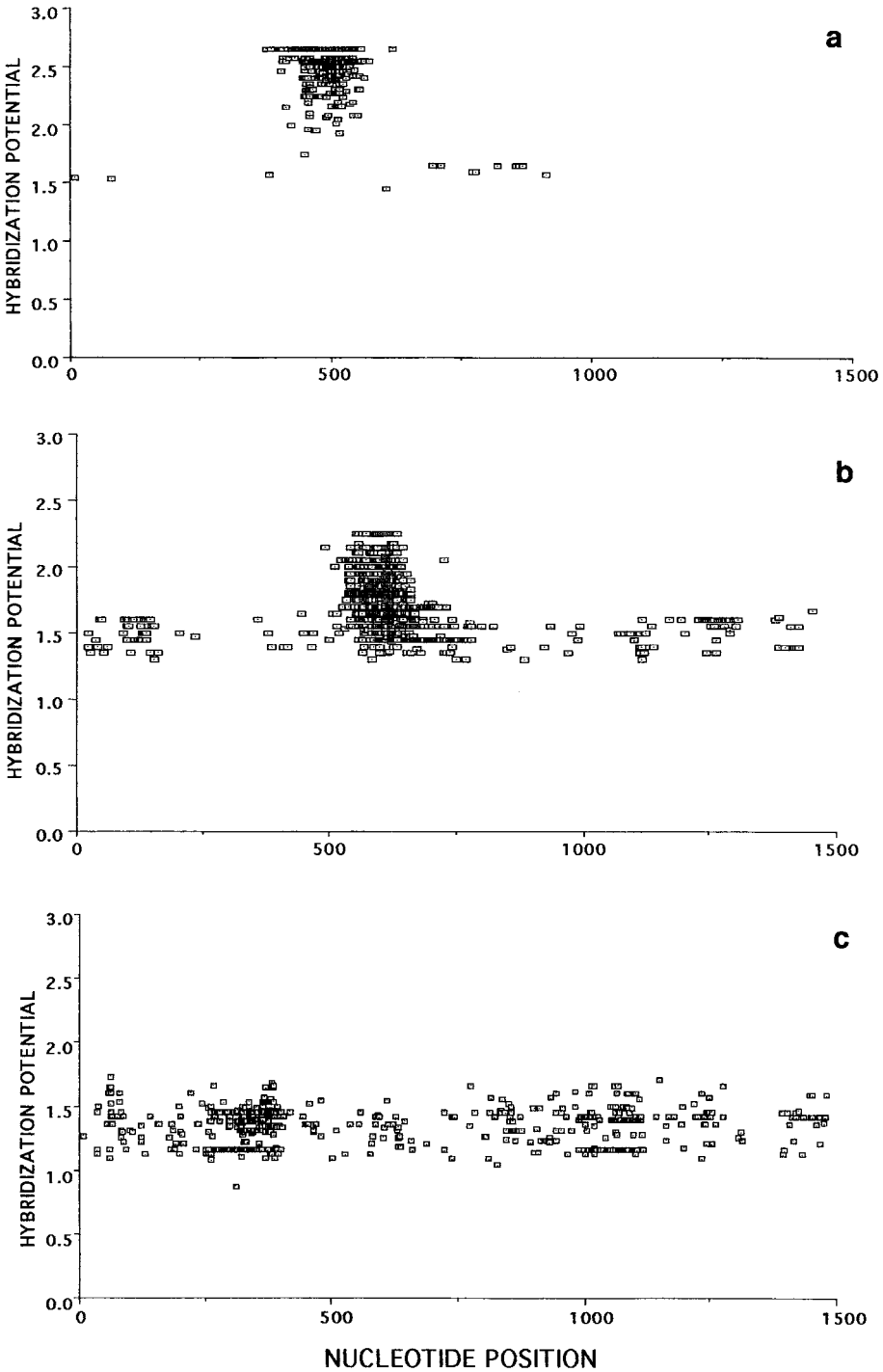

FIG. 3. Maximum HP values for various primers against 1,484 bacterial SSU rRNA sequences, shown as a function of nucleotide position. (a) A good primer (primer 52); (b) a poor primer (primer 62); (c) a random primer.

tion fragment length analysis of the SSU rRNA sequences in the RDP database to estimate the potential resolution of this approach with a complex set of SSU rRNA sequences. From the consensus PCR primer sets we selected primers 6 and 52 as optimal. They have HP values considerably above 1.8 for 99 and 90% of the SSU rRNA sequences, respectively. These PCR primers also flank three regions that are highly variable, particularly with regard to indels. The predicted terminal restriction fragments from SSU rRNA sequences for most available restriction enzymes were analyzed.

As a practical matter we can easily resolve labeled DNA fragments ranging from 50 to 450 nucleotides in length using an automated DNA sequencer. Our experiments have shown that this range allows accurate fragment size determination to the base pair (3). In comparing the characterization of PCR-amplified SSU rRNA regions by different restriction enzymes,

TABLE 1. Sequences, positions, and the original references for all of the probes used in this study[a]

Probe	Sequence	RDP aligned positions	Reference(s)
Universal			
1	CCAGCAGCCGCGGTAATACG	916–936	Primer 52
2	CAGCMGCCGCGGTAATWC	936–954	1, 10
3	GYACACACCGCCCGT	2584–2600	10
Eukaryotic			
1	CAGGTCTGTGATGCCC	3462–3478	1, 10
2	GGTTCGATTCCGGAGARGGAGC	867–891	10
3	GGAGGGCAAGTCTGGT	1362–1378	1
Bacterial			
1	GGGCCCGCACAAGCGGT	1463–1487	1
2	GCYTAACACATGCAAGTCGA	89–102	Primer 6
3	AGAGTTTGATCCTGGCTCAG	41–61	6
4	ACTCCTACGGGAGGCAGC	671–696	1
5	GGATTAGATACCCTGGTAG	1265–1285	1
6	ATGGCTGTCGTCAGCTCGTG	1781–1803	1
7	GGTTAAGTCCCGCAACGAGC	1820–1842	1
Archaeal			
1	GGCCCTACGGGGSGCASCAGGCGC	671–705	10
2	AATTGGAKTCAACGCCGGR	1503–1526	6
3	TTCCGGTTGATCCYGCCGGA	40–61	1, 6
4	AGGAATTGGCGGGGGAGCAC	1449–1473	6
Euryarchaeal probe 1	AAGGGCYGGGCAAG	911–926	1
Crenarchaeal			
1	CGGGAGCCCCGAGATGGGC	608–648	This work
2	AGCGGGGGGCAAGYCTGG	912–932	1
3	GTAGACGGTACTCGG	1093–1108	7

[a] The position of each probe was determined by the aligned sequence positions in the RDP bank. These sequences may be published as reverse complements.

we used as a criterion the number of different lengths between 50 and 450 nucleotides that are produced. The maximum number of different lengths is 400; thus, this sets the upper limit. In addition to production of a large number of different restriction fragments, a relatively uniform distribution of lengths is preferred. We compared the number of terminal restriction fragments produced by each of a series of restriction enzyme and then tested combinations of enzymes which had the highest number of fragments as individual enzymes. Figure 4 shows the profile of lengths for several restriction enzymes (*Hha*I, *Msp*I, *Rsa*I, and *Hha*I plus *Rsa*I). The total numbers of restriction fragments between 50 and 450 nucleotides in length for each restriction enzyme (or combination) are as follows: *Hha*I, 252; *Msp*I, 245; *Rsa*I, 231; and *Hha*I plus *Rsa*I, 282. The combination of restriction enzymes *Hha*I plus *Rsa*I produces the most attractive profile. It has the largest number of fragments with different lengths (282) and the fewest lengths with large numbers of representatives. In the case of the *Hha*I plus *Rsa*I profile, only one length has over 20 representatives.

TABLE 2. Specificity of different hybridization probes when hybridized with sequences from the different SSU rRNA sets

Primer	% of sequences with HP value of ≥1.8			
	Eukaryotic	Bacterial	Archaeal	Crenarchaeal
Universal				
1	96	99	82	70
2	97	99	83	70
3	94	97	83	70
Eukaryotic				
1	99	0	0	0
2	98	0	0	0
3	97	5	76	70
Bacterial				
1	1	92	1	10
2	0	90	8	10
3	0	70	10	0
4	86	99	85	55
5	74	100	91	75
6	12	100	91	80
7	71	100	89	60
Archaeal				
1	1	5	94	95
2	56	5	79	60
3	70	4	75	40
4	66	72	90	75
Euryarchaeal primer 1	2	4	86	30
Crenarchaeal				
1	0	4	34	95
2	94	9	17	70
3	0	0	1	5

The resolution capacity of this approach is set by the length resolution of the analytic equipment (in this case an automated DNA sequencer). Our optimal PCR primer set (primers 6 and 52) in combination with restriction enzymes *Hha*I plus RsaI utilizes over half of the available resolution capacity. This is quite remarkable and bodes well for the ability of analyzing bacterial populations from natural sources.

Initially, we expected the majority of terminal restriction fragment length variation to result from restriction site polymorphisms. An estimate of the amount of length variation due to restriction site polymorphisms can be made by using aligned sequence positions for the calculation of fragment length (aligned lengths are independent of indels). We recomputed the number of different terminal restriction fragment lengths that would result if aligned positions were used instead of actual nucleotide length (without indels). Comparing the different number of actual lengths (245) with the number of different aligned fragment lengths (65) suggests that about 73% of the length variation is due to variability in indels and about 27% is the result of restriction site polymorphism.

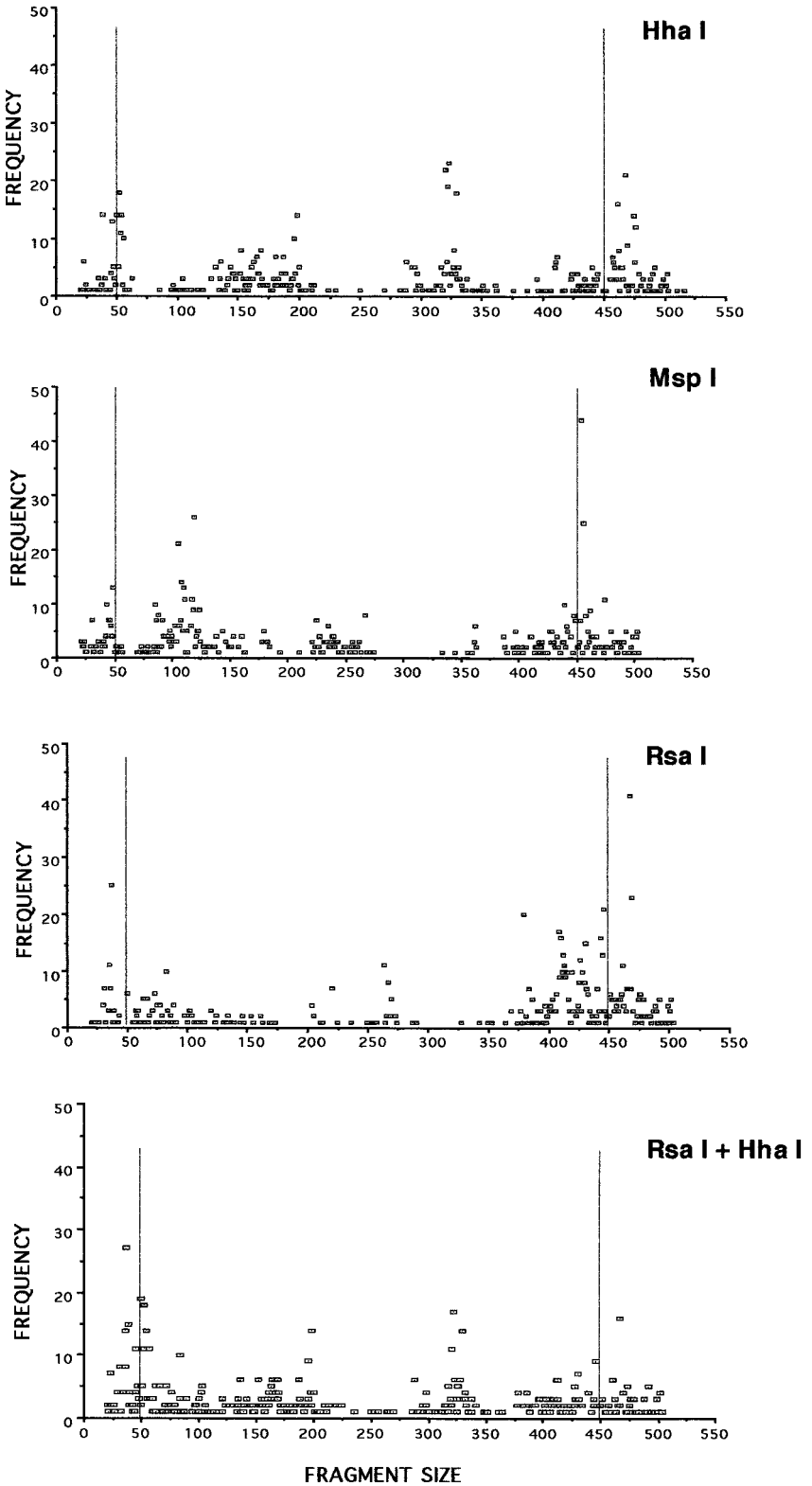

FIG. 4. Frequencies of different size restriction fragments that are generated when the PCR products yielded from the SSU rRNAs of 856 different bacteria are digested with various restriction enzymes: *Msp*I, *Hha*I, *Rsa*I, and a combination of *Hha*I and *Rsa*I. The vertical lines at 50 and 450 show the region in which fragment length can be easily analyzed.

The aligned length profile identifies the prevalence of restriction site polymorphisms. As an example, the *Msp*I restriction site polymorphisms were examined in detail. There are three locations in the aligned SSU rRNA sequences at which *Msp*I sites are frequently found (these sites give rise to aligned fragment lengths of 245, 438, and 731 nucleotides); the rest of the *Msp*I sites arise by infrequent mutations. About 24% of the sequences have an *Msp*I restriction site that gives rise to aligned fragment lengths of 245 nucleotides; however, this is the first *Msp*I site in only 14% of the sequences. Interestingly, this *Msp*I restriction site does not occur in the consensus sequence. Similarly, the restriction site that gives rise to aligned fragment lengths of 438 and 731 nucleotides occurs in 25 and 66% of all of the sequences, respectively, but the former site occurs as the first *Msp*I site in only 9% of the sequences and the latter site occurs first in 21% of the sequences. About 15% of the sequences have no *Msp*I sites and yield a 746-nucleotide aligned fragment length. Although there are substantial numbers of restriction site polymorphisms, with primitive sites being lost and new sites being created, the majority of restriction fragment length variation in this region is due to multiple indel events.

Clearly, the protocol that we have described using conserved PCR primers flanking a highly variable DNA region has the potential for distinguishing a wide spectrum of bacterial SSU rRNA sequences. This will be a powerful tool for analyzing bacterial populations from natural sources. The algorithm that we have developed for evaluation of potential hybridization efficiency of primers with long DNA sequences has potential for comparing the merits of different PCR primers and hybridization probe sequences. Finally, it is apparent that the indels contribute significantly to the sequence variability in conserved genes.

Group-specific probes are being widely used in qualitative and quantitative analysis of microbial populations (1, 6, 7, 10, 16). Most of these probes have been in use for some time and are believed to perform adequately. A direct evaluation of the merit of these probes has been difficult. Our algorithm can stimulate the performance of these probes. We tested some of the most widely used group-specific SSU rRNA probes with sets of bacterial, eukaryotic, archaeal, and crenarchaeal SSU rRNA sequences. The results of these analyses are shown in Table 2.

The universal probes all hybridize well with all of the sets of SSU rRNA sequences except the crenarchaea, which is a very small sequence set. The universal probe that we identified (universal probe 1; primer 52) is very similar to the published universal probe 2 (Table 1); our probe is 2 nucleotides longer and does not have the degenerate nucleotides (M and W) found in the published version. The universal probe 2 sequence corresponding to universal probe 1, with M = A and W = A, has a perfect match with 1,052 (71%) of the SSU rRNA sequences. The other three iterations of this probe sequence, collectively, have only 32 additional perfect matches. Clearly, using a degenerate probe at either or both positions adds only marginally to the HP. Alternatively, bacterial probe 2, which we have identified as primer 6, has a Y in its sequence. The sequence corresponding to bacterial probe 2 with Y = T has a perfect match with 434 of the SSU rRNA sequences, while the probe with Y = C has 299 perfect matches. In this case, using a degenerate probe substantially increases the effectiveness of the probe. Our hybridization simulation algorithm allows us to evaluate the HP of each individual probe sequence and determine if a degenerate probe will be more effective.

Two of the three published eukaryotic probe sequences have excellent specificity, but eukaryotic probe 3 shows a significant affinity for the archaea in addition to eukaryotes. Of the eight bacterial probes, probes 1, 2, and 3 have high specificity for bacterial sequences with only minor affinity for archaeal sequences. Probes 1 and 2 have high HP values, while probe 3 has a slightly lower HP value. Bacterial probes 4, 5, 6, and 7 also show very high affinity for bacteria, but they also have relatively high HP values with archaeal SSU rRNA sequences and eukaryotic SSU rRNA sequences. These probes would be good candidates for universal probes except that their HP with eukaryotes is significantly less than that of the universal probes. Archaeal probe 1 has high specificity for both archaeal and crenarchaeal sequences but does not distinguish among these sequences.

Although archaeal probe 2 has a moderately high HP for archaeal SSU rRNA sequences, this probe also has a high HP for eukaryotic and bacterial sequences. Archaeal probe 3 has only a nominal HP for archaeal SSU rRNA sequences and an equally high HP for eukaryotic sequences. Euryarchaeal probe 1 has a preferential HP for archaeal SSU rRNA sequences but also has significant HP for crenarchaeal SSU rRNA sequences. We have identified crenarchaeal probe 1, which has high HP and specificity for crenarchaea (crenarchaeal sequences make up 24% of the archaeal class). Crenarchaeal probe 2 has a high HP for eukaryotic SSU rRNA sequences as well as crenarchaeal sequences. Crenarchaeal probe 3 has little HP for any class of SSU rRNA sequence.

The results for most of the group-specific probes were encouraging, as they demonstrated a high degree of specificity for the sequences in their specific sets of SSU rRNAs. The specificity and utility of a few published hybridization probes appear to be questionable. To demonstrate specificity, a probe must have high affinity for a class of SSU rRNA sequences and must not hybridize with other classes of SSU rRNAs. In general, our analyses of the published probes yield results similar to those of an analysis that can be performed using the Check_Probe algorithm in the RDP database (11). Check_Probe returns the number of matches without consideration for the relative number GC or AT pairs. It is much more difficult to determine the specificity of a probe with Check_Probe, as slightly mismatched sequences may hybridize very well. Analysis using our hybridization simulation algorithm also allows us to rank the specificity of the probes.

ACKNOWLEDGMENTS

We thank Chris Collins, Gary Fogel, Kenneth Jones, Jinliang Li, and Charles Marshal for reviewing this paper.

This work was funded in part by a grant from the National Sea Grant College Program, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, under grant NA89AA-D-SG138 and project R/CZ-104 through the Sea Grant College.

REFERENCES

1. **Amann, R. I., W. Ludwig, and K. H. Schleifer.** 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. **51:**143–169.
2. **Avaniss-Aghajani, E., K. Jones, D. Chapman, and C. Brunk.** 1994. A molecular technique for identification of bacteria using small subunit ribosomal RNA sequences. BioTechniques **17:**144–149.
3. **Avaniss-Aghajani, E., K. Jones, A. Holtzman, T. Aronson, N. Glover, M. Boian, S. Froman, and C. F. Brunk.** 1996. Molecular technique for rapid identification of mycobacteria. J. Clin. Microbiol. **34:**98–102.
4. **Baldino, F., Jr., M.-F. Chesselet, and M. E. Lewis.** 1989. High resolution in-situ hybridization histochemistry. Methods Enzymol. **168:**761–777.
5. **Breslauer, K. J., R. Frank, H. Blocker, and L. A. Markey.** 1986. Predicting DNA duplex stability from the base sequence. Proc. Natl. Acad. Sci. USA **83:**3746–3750.
6. **Delong, F. E.** 1992. Archaea in coastal marine environments. Proc. Natl. Acad. Sci. USA **89:**5685–5689.
7. **Delong, E. F., K. Y. Wu, B. B. Prezelin, and R. V. M. Jovine.** 1994. High

abundance of Archaea in Antarctic marine picoplankton. Nature (London) **371:**695–697.
8. **Frier, S. M., R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, and D. H. Turner.** 1986. Improved free energy parameters for predictions of RNA duplex stability. Proc. Natl. Acad. Sci. USA **83:**9373–9377.
9. **Fuhrman, J. A., K. McCallum, and A. A. Davis.** 1993. Phylogenetic diversity of subsurface marine microbial communities from Atlantic and Pacific oceans. Appl. Environ. Microbiol. **59:**1294–1302.
10. **Giovannoni, S. J., T. B. Britschgi, C. L. Moyer, and K. G. Field.** 1990. Genetic diversity in Sargasso Sea bacterioplankton. Nature (London) **345:** 60–63.
11. **Larsen, N., J. G. Olsen, L. B. Maidak, J. M. McCaughey, R. Overbeek, J. T. Mack, L. T. Marsh, and C. R. Woese.** 1993. The ribosomal database project. Nucleic Acids Res. **21**(Suppl.)**:**3021–3023.
12. **Moyer, C. L., F. C. Dobbs, and D. M. Karl.** 1994. Estimation of diversity and community structure through restriction fragment length polymorphism distribution analysis of bacterial 16S rRNA genes from a microbial mat at an active, hydrothermal vent system, Loihi Seamont, Hawaii. Appl. Environ. Microbiol. **60:**871–879.
13. **Mullis, K. B., and F. A. Faloona.** 1987. Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. Methods Enzymol. **155:**335–350.
14. **Pace, N. R., D. A. Stahl, D. J. Lane, and G. J. Olsen.** 1986. The analysis of natural microbial populations by ribosomal RNA sequences. Adv. Microbiol. Ecol. **9:**1–55.
15. **Picard, C., C. Ponsonnet, E. Paget, X. Nesme, and P. Simonex.** 1992. Detection and enumeration of bacteria in soil by direct DNA extraction and polymerase chain reaction. Appl. Environ. Microbiol. **58:**2717–2722.
16. **Raskin, L., J. M. Stromley, B. E. Rittman, and D. Stahl.** 1994. Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. Appl. Environ. Microbiol. **60:**1232–1240.
17. **Schmidt, T. M., E. F. Delong, and N. R. Pace.** 1991. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J. Bacteriol. **173:**4371–4378.
18. **Sogin, M., and H. Gunderson.** 1987. Structural diversity of eukaryotic small subunit ribosomal RNAs. J. Endocytobiol. **503:**125–139.
19. **Spaepen, M., A. F. Angulo, P. Marynen, and J. J. Cassiman.** 1991. Detection of bacterial and mycoplasma contamination in cell cultures by polymerase chain reaction. FEMS Microbiol. Lett. **78:**89–94.
20. **Stahl, D. A., D. J. Lane, G. J. Olson, and N. R. Pace.** 1985. Characterization of a Yellowstone hot spring community by 5S rRNA sequences. Appl. Environ. Microbiol. **49:**1379–1384.
21. **Steffan, R. J., and R. M. Atlas.** 1991. Polymerase chain reaction: applications in environmental microbiology. Annu. Rev. Microbiol. **45:**137–161.
22. **Wang, R. F., W. W. Cao, and M. G. Johnson.** 1992. A 16S rRNA-based DNA probe and PCR method specific for *Listeria ivanovii*. Appl. Environ. Microbiol. **58:**2827–2831.
23. **Ward, D. M., R. Weller, and M. M. Bateson.** 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature (London) **345:**63–65.