Nucleic Acids Research, Vol. 19, No. 24 6823–6831

Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes

James Versalovic[1], Thearith Koeuth[1] and James R.Lupski[1,2]
[1]Institute for Molecular Genetics and [2]Department of Pediatrics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Received September 13, 1991; Revised and Accepted November 18, 1991

ABSTRACT

Dispersed repetitive DNA sequences have been described recently in eubacteria. To assess the distribution and evolutionary conservation of two distinct prokaryotic repetitive elements, consensus oligonucleotides were used in polymerase chain reaction [PCR] amplification and slot blot hybridization experiments with genomic DNA from diverse eubacterial species. Oligonucleotides matching Repetitive Extragenic Palindromic [REP] elements and Enterobacterial Repetitive Intergenic Consensus [ERIC] sequences were synthesized and tested as opposing PCR primers in the amplification of eubacterial genomic DNA. REP and ERIC consensus oligonucleotides produced clearly resolvable bands by agarose gel electrophoresis following PCR amplification. These band patterns provided unambiguous DNA fingerprints of different eubacterial species and strains. Both REP and ERIC probes hybridized preferentially to genomic DNA from Gram-negative enteric bacteria and related species. Widespread distribution of these repetitive DNA elements in the genomes of various microorganisms should enable rapid identification of bacterial species and strains, and be useful for the analysis of prokaryotic genomes.

INTRODUCTION

Interspersed repetitive DNA sequence elements have been characterized extensively in eucaryotes. The best known of these elements is the *Alu* family of sequences identified in mammalian species (1–3). The conserved nature and interspersed distribution of these *Alu* repeats have been exploited to amplify unique sequences between *Alu* repeats by the polymerase chain reaction (4,5) in a technique known as *Alu*-PCR (6). Species-specific repetitive DNA elements have been used to differentiate between closely related murine species (7). The function of these repetitive DNA sequences in mammalian genomes remains largely unknown.

Prokaryotic genomes are much smaller than the genomes of mammalian species with approximately 10^6 versus 10^9 base pairs of DNA, respectively. These smaller prokaryotic genomes may have been maintained through selective pressures for rapid DNA replication and cell reproduction (8). Noncoding repetitive DNA would likely be kept to a minimum under natural selection for rapid growth, unless these sequences maintain themselves as 'selfish' DNA. Regions of the *Escherichia coli* genome where DNA sequence information is available demonstrate a high density of transcribed sequences (9,10). There are even examples of coding regions where the termination codon of one gene overlaps with the start codon of the next gene (11,12). Introns are lacking in all currently known *E. coli* chromosomal genes, although they have been described in the coliphage, T4 (13,14), and cyanobacteria (15,16). Nevertheless, families of short intergenic repeated sequences have been described in enteric bacteria (17–20).

In this study the distribution of repetitive DNA sequences in eubacteria was examined by analysis of the repetitive extragenic palindromic [REP] elements (17), otherwise known as palindromic units [PU] (18), and the Enterobacterial Repetitive Intergenic Consensus [ERIC] sequences (20). The REP elements were first described as potential regulatory sequences within untranslated regions of operons by virtue of their palindromic nature and ability to form stable stem-loop structures in transcribed RNA (21). A consensus REP sequence was formulated by multiple alignment of REP-like sequences from *E. coli* and *Salmonella typhimurium* (17,18). This 38 bp REP consensus sequence contains six totally degenerate positions, including a 5 bp variable loop between each side of the conserved stem of the palindrome (17,22). Multiple functions have been proposed for these highly conserved, dispersed REP elements including roles in transcription termination, mRNA stability, and chromosomal domain organization *in vivo* (21–26). Recently a new family of repetitive elements, enterobacterial repetitive intergenic consensus (ERIC) sequences (20), otherwise known as intergenic repeat units [IRUs] (19), has been defined using genomic sequence information obtained primarily from *E. coli* and *S. typhimurium* (19,20). Like their REP counterparts, these larger 126 bp ERIC elements contain a highly conserved central inverted repeat and are located in extragenic regions (19,20). The ERIC consensus sequence does not appear to be related to the REP consensus sequence (20).

The distribution of REP and ERIC elements in diverse prokaryotic genomes was examined by the polymerase chain reaction [PCR] (4,5) with consensus primers and by slot blot hybridization with radiolabeled, consensus probes. PCR analysis using primers to repeat sequences, with bacterial genomic DNA

as a template, reveals inter-REP [or inter-ERIC] distances and patterns specific for bacterial species and strains but is limited to adjacent repeat elements within the limitations of polymerase extension [~5 kb]. Slot blot hybridization with consensus probes allows one to measure the relative distribution of similar repetitive elements in entire genomes regardless of the distance and orientation between consecutive elements. This study demonstrates that REP- and ERIC-like sequences are present in many diverse eubacterial species, and that REP and ERIC sequences can be utilized as efficient primer binding sites in the polymerase chain reaction to produce fingerprints of different bacterial genomes.

MATERIALS AND METHODS

Bacterial strains

The sources of bacterial strains and/or genomic DNA are shown in Table 1.

Isolation and quantitation of genomic DNA

Bacterial cells were lysed by different methods depending on whether they were Gram-negatives/spirochetes or Gram-positives. Gram-negatives were grown overnight in Luria-Bertani [LB] broth (27). Spirochetes were kindly provided by Dr. Robert E. Baughn, Dept. of Immunology, VA Medical Center, Baylor College of Medicine, Houston, TX, and were collected directly from freshly thawed stocks. After cells were collected, both Gram-negative and spirochete cells were treated identically in the following steps. Cells were pelleted and washed twice in 1 ml of 1M NaCl by centrifugation in a fixed angle microfuge [Eppendorf] at 15,000 rpm for 5 min. Cells were washed twice and resuspended in TE [10mM Tris, 25mM EDTA, pH 8.0] and incubated in 0.2 mg/ml lysozyme [Sigma] and 0.3 mg/ml RNase A [Sigma] for 20 min, 37°C. If lysis by lysozyme was not visible with refractory pathogenic strains, 0.6% SDS was added. To these suspensions, 1% Sarkosyl and 0.6 mg/ml proteinase K [Boehringer Mannheim] were added, and the cells were incubated for 1 hr, 37°C. Cell lysates were extracted twice with phenol and twice with chloroform. The aqueous phase was precipitated with 0.33M NH_4 acetate and 2.5 volumes of ethanol. Precipitated threads of DNA were removed with a sterile Pasteur pipette tip, and dissolved in TE [10mM Tris, 1 mM EDTA, pH 8.0].

Gram-positive bacteria were obtained as concentrated cell pellets kindly provided by Dr. Edward Mason, Dept. of Pediatrics, Texas Children's Hospital, Houston, TX. Pellets were washed twice in 1 M NaCl and twice in TE [50 mM Tris, 50 mM EDTA, pH 7.8] and spun in a fixed-angle microfuge [Eppendorf] for 5 min. Cell pellets were resuspended in TE and incubated with 250 U/ml mutanolysin [Sigma] and 0.3 mg/ml RNAse A for 30 min at 37°C. To this reaction, 0.6% SDS and 0.6 mg/ml proteinase K were added, and the mixture was incubated for 1 hr at 37°C, followed by 65°C, 45 min. These lysates were extracted twice with phenol and twice with chloroform. Chromosomal DNA was precipitated and dissolved exactly as described above. Additional genomic DNAs were kindly provided by New England Biolabs [Table 1 (NEB)].

Genomic DNA was quantitated by spectrofluorimetry at excitation and emission wavelengths of 365 nm and 460 nm respectively using the DNA-specific dye, Hoechst 33258 [Hoefer], and a Model TKO-100 mini-fluorometer [Hoefer] according to manufacturer's instructions.

Oligonucleotide synthesis and design

Oligonucleotides were synthesized with dimethoxytrityl phosphoramidites using an Applied Biosystems Model 380B DNA synthesizer and DNA sequence information from published consensus sequence data (17,20,28). Dried oligonucleotide pellets were suspended in HPLC-purified water [Fisher] and quantitated by UV-VIS spectrophotometry with absorption measured at 260 nm.

REP oligonucleotide sequences are listed in Fig. 1A. Degenerate 38-mer REPALL probes were designed which encompassed the entire consensus REP sequence. REP oligonucleotides were also designed from each half of the conserved stem of the palindrome in opposite orientations such that the 3′ ends were directed outwards from each REP sequence. Since one side of the stem sequence is shorter, three inosines were added to the 5′ end of the REP1R primers so that REP1R primer lengths matched the 18-mer lengths of REP2 primers. Total degeneracy is represented either by any one of the four common bases [A, G, C, or T] at specific positions, or inosines placed at specific positions. Inosine contains the purine base, hypoxanthine, and is capable of forming Watson-Crick base pairs with A, G, C, or T (29). Inosine base-pairs are weaker than A:T base-pairs, but inosine forms the least destabilizing and discriminating mismatches overall (29,30). Positions can be partially degenerate with two of these four bases placed at specific positions as chosen from the consensus REP sequence.

ERIC oligonucleotide sequences are listed in Fig. 1B. The ERICALL oligonucleotide contains the entire conserved central core inverted repeat (20). Nondegenerate, consensus ERIC1R and ERIC2 oligonucleotides were designed from each half of this core inverted repeat (20) in opposite orientations such that the 3′ ends are directed outwards from the center of the ERIC element. The same repeat sequence oligonucleotides can be used in both hybridization and PCR experiments to allow one to effectively compare and contrast the results obtained with both techniques.

For experiments demonstrating ERIC primer specificity, oligonucleotides matching sequences within the *hsdR* gene (28) were synthesized. Oligonucleotide sequences are as follows: hsdR+2758—5′-CAGCCATGAACAACTGGTGGCG-3′ and hsdR+3235R—5′-TGCTTTGCGCAGGGAAGATTCC-3′.

5′ End labeling

End-labeling of each oligonucleotide probe was performed as described by Maniatis et al. (31). Fifty pmol of each primer were used with 20 U T4 polynucleotide kinase (Pharmacia) and 5 μl [γ-^{32}P] ATP [6000 Ci/mmol; Du Pont]. Labeled DNA was separated from unincorporated isotope by diluting the 50 μl reaction volume to 1 ml in deionized, filtered water, followed by centrifugation of this solution through Centricon-3 [Amicon] tubes. Unincorporated radioisotope was filtered through the size exclusion membrane present in the Centricon-3 tubes. The solution containing the end-labeled oligonucleotide, which was excluded by the membrane, was used as the hybridization probe.

Hybridization conditions

A single membrane containing genomic DNA from 39 different eubacterial species representing 7 of 10 different phyla as defined by Woese (32), based on rDNA sequence comparisons, was named the 'bug blot.' This 'bug blot' was made by adding 100 ng of denatured genomic DNA, from each species listed in Fig. 5A, per slot on GeneScreen Plus [Du Pont] membranes. These membranes were pretreated as described in Maniatis (31).

Genomic DNAs were denatured at 100°C, 5 min. DNA samples were then applied to the membrane, and 500 μl 0.4N NaOH were added to each slot. Membranes were rinsed in 1×SSC, and blotted dry with Whatman paper. Membranes were baked at 80°C, 1 hr, and stored in sealed plastic bags at −20°C.

The hybridization solution was prepared as recently described for use with oligonucleotide probes on a membrane containing ordered lambda phages representing the *E. coli* W3110 genome (33). For REP oligonucleotide hybridization, membranes were prehybridized at 42°C for 1.5 hrs. The probe was denatured at 100°C, 5 min. Probe was added at 1×10^6 cpm/ml hybridization solution and the membranes were incubated at 42°C for 15 hrs. Both ERIC oligonucleotide prehybridizations and hybridizations were performed at 65°C. After incubation the membranes were washed twice at room temperature for 10 min with 2×SSPE and 0.1% SDS, followed by one final wash (REP, 37°C, 15 min; ERIC, 40°, 1 min). Autoradiograms were exposed on Kodak X-Omat film with two intensifying screens at −85°C for 24 hrs.

Polymerase chain reaction (PCR) conditions

Each 25 μl PCR reaction contained 50 pmol each of 2 opposing primers, 100 ng of template [genomic] DNA, 1.25 mM of each of 4 dNTPs, 2 U AmpliTaq DNA polymerase [Perkin-Elmer/Cetus] in a buffer with 10% DMSO [v/v] (34). PCR amplifications were performed in an automated thermal cycler [Perkin-Elmer/Cetus DNA Thermal Cycler] with an initial denaturation [95°C, 7 min] followed by 30 cycles of denaturation [90°C, 30 sec], annealing [REP, 40°C, 1 min; ERIC, 52°C, 1 min], and extension [65°C, 8 min] with a single final extension [65°C, 16 min]. All PCR reaction tubes were placed in internal rows of the thermal cycler and all peripheral tubes were surrounded by 'dummy' tubes containing water and mineral oil (35). Eight μl [for Fig. 1, 5 μl] of each PCR reaction were then electrophoresed directly on 1% agarose gels containing 1×TAE [Tris acetate-EDTA; (31)], 0.5 μg/ml ethidium bromide. These gels were photographed with 20 second exposures to Polaroid Type 55 film.

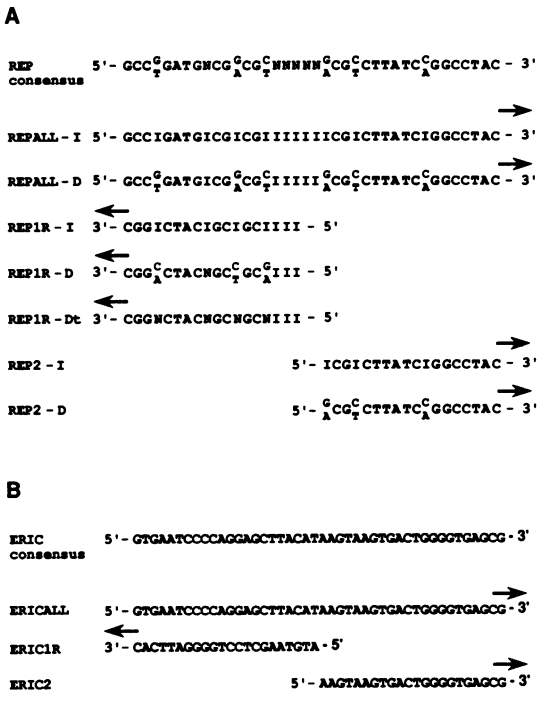

Fig. 1. Oligonucleotide design and evaluation by PCR amplification. (A) Alignment of various REP oligonucleotide primer sequences with respect to the REP consensus sequence determined previously (17). 'I' represents inosine. (B) Alignment of ERIC oligonucleotide primer sequences with respect to the central inverted repeat of the published ERIC consensus sequence (20).

RESULTS

Amplification of *E. coli* strain W3110 genomic DNA with different REP and ERIC oligonucleotides as PCR primers

Nondegenerate oligonucleotides of either random, arbitrary sequences [AP-PCR;(36)] or coding sequences of repeated genes (37) have been used as primers to generate PCR-based fingerprints of bacterial genomes. Nondegenerate, consensus REP oligonucleotides have been used in hybridization experiments with eubacterial genomic DNA (38). However, the use of degenerate, repetitive sequence oligonucleotides as primers in PCR-based DNA amplification of bacterial genomic DNA has not yet been examined. We empirically determined the optimal degenerate REP primer set for the production of species- and strain-specific genomic fingerprints of bacterial DNA. Our hypothesis was that repetitive DNA sequences were dispersed in the *E. coli* chromosome in different orientations and separated by various distances. These dispersed repetitive sequences could be used as primer binding sites and PCR amplification between them would yield distinct patterns of DNA fragments varying in size when separated by agarose gel electrophoresis. Since both REP and ERIC primers were designed to match inverted repeat sequences, we compared the effectiveness of each pair of outwardly-directed primers versus each primer alone.

The outwardly-directed primer set, REP1R-I and REP2-I [Fig. 1A], provided the most distinct genomic fingerprint of *E. coli* strain W3110 chromosomal DNA [Fig. 2]. REPALL, REP1R, and REP2 oligonucleotides were all tested as primers

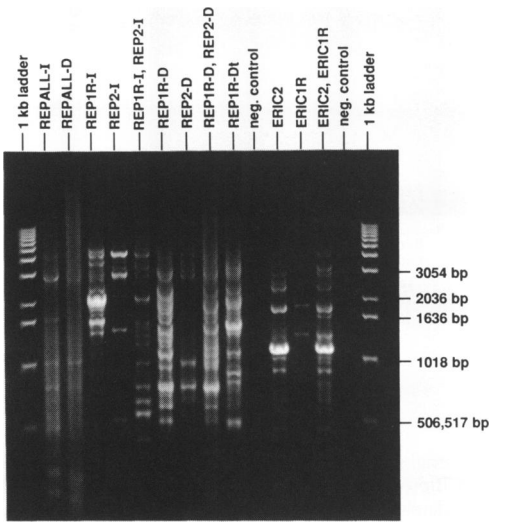

Fig. 2. PCR amplification of *E. coli* strain W3110 genomic DNA with different REP and ERIC oligonucleotide primer sets as indicated. No template DNA was added to the negative control lanes. REP1R-I and REP2-I primers were used in negative control lane 11; ERIC1R and ERIC2 primers were used in negative control lane 15. The DNA molecular weight marker is a 1-kb ladder [BRL]. The gels were 1% agarose− 1×Tris-acetate-EDTA and contained 0.5 μg of ethidium bromide per ml. to stain the DNA.

for DNA amplification because these outwardly-directed primers can amplify DNA between successive REP sequences in any orientation. The inosine-containing primers, REP1R-I antd REP2-I, provided more distinct DNA amplification band patterns and less smearing [Fig. 2], possibly because each primer is represented by a single primer sequence instead of a pool of multiple primer sequences as with REP1R-D and REP2-D. Each REP primer alone yielded visible amplification products [Fig. 2] of relatively limited complexity. This result may stem from the fact that each side of the inverted repeat has a slightly different consensus sequence. The use of both primers REP1R-I and REP2-I appears to allow optimal annealing with both sides of the conserved stem of each REP-like sequence in the genome. Inefficient amplification with REPALL-I and REPALL-D was observed [Fig. 2], presumably because a palindrome is present in the primer. Potential primer dimer formation between REPALL primers of opposite orientation precluded us from designing these primers in both orientations.

Amplification results obtained with the single consensus ERIC primer set, ERIC1R and ERIC2 [Fig. 1B], were matched in complexity by the results obtained with ERIC2 alone [Fig. 2]. In contrast PCR amplification with ERIC1R alone yielded limited amplification products [Fig. 2]. Two possible reasons for this observation are that either greater sequence conservation exists in the side of the inverted repeat complementary to ERIC2 or homologous, unrelated sequences complementary to ERIC2 exist outside ERIC elements in the genome.

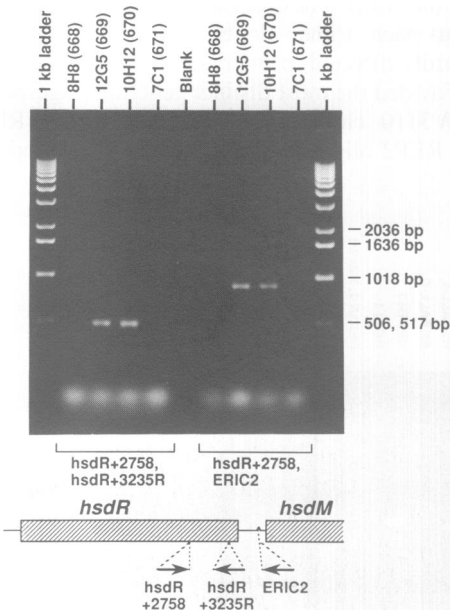

Fig. 3. Demonstration of specificity of ERIC oligonucleotide primer/template interactions. PCR products of the expected sizes were generated by amplification of Kohara lambda phages (41) containing the *E. coli hsdR* locus (9,28) and an adjacent ERIC sequence (20). The Kohara lambda phages used are listed by clone numbers (41) and miniset serial numbers are shown in parentheses. One μl of each Kohara phage lysate was used as template DNA; PCR conditions were as described in Materials and Methods. Lanes 2–5 represent PCR amplifications with primers within the *hsdR* gene, hsdR+2758 and hsdR+3235R. Lanes 7–10 represent PCR products generated by primers hsdR+2758 and ERIC2. Lane 6 is a blank lane where nothing was added to the gel. The molecular weight marker is a 1-kb ladder [BRL]. The gels were 1% agarose-1×Tris-acetate-EDTA and contained 0.5 μg of ethidium bromide per ml.

PCR reactions using primer binding sites at known distances from ERIC sequences were performed to verify the size of amplification products. Specificity of REP primer/template interactions has been demonstrated by amplification between a known REP sequence (39) [with primer REP2-I] and a Tn*5* insertion in the *glpD* gene of *E. coli* (40). The specificity of ERIC-PCR was demonstrated by PCR amplification of a defined DNA segment between a published ERIC sequence (20), or IRU sequence (19), and a sequence within the *E. coli hsdR* gene (28) using the ordered Kohara phage library (41,42) [Fig. 3]. Single PCR products of the expected size were amplified both within the *hsdR* gene and between the *hsdR* and ERIC sequences carried by Kohara phages containing the *E. coli hsdR* locus (9,41) [Fig. 3]. Amplification with only a single *hsdR* primer failed to yield any product [data not shown].

REP- and ERIC-PCR generates species- and strain-specific DNA fingerprints of Gram-negative enteric bacteria

Phenotypic descriptions such as biotyping, phage typing, serotyping, and antibiotic resistance patterns have been used in the past to characterize bacterial strains. Recently molecular approaches have been utilized to assess genotypic distinctions directly in prokaryotes. Examples include restriction fragment analysis of genomic DNA [RFLPs] by conventional agarose gel electrophoresis (43) or pulsed field gel electrophoresis (44,45), and generation of plasmid profiles (46). We examined bacterial species- and strain-specific fingerprint patterns directly by PCR amplification of their genomic DNA. Consensus oligonucleotides

PHYLA/Species	Strain	Source	PHYLA/Species	Strain	Source
BACTEROIDES, FLAVOBACTERIA	.....	.....	PURPLE BACTERIA	.....	.....
Bacteroides fragilis	12256	A. Salyers	*Citrobacter diversus*	4036, 1216m123	E. Mason
Flavobacterium meningosepticum	4626	E. Mason	*Enterobacter sakazakii*	4584,4585	E. Mason
Flavobacterium okeanokoites		NEB	*Escherichia coli*	E2, 2956 3049,4038	E. Mason
Fusobacterium nucleatum		NEB	*Escherichia coli*	HB101	M. Winkler
CYANOBACTERIA	.....	.....	*Escherichia coli*	W3110	Y. Kohara
Anabaena sp.	PCC 7120	S. Golden	*Klebsiella pneumoniae*	4732	E. Mason
GRAM-POSITIVE BACTERIA	.....	.....	*Klebsiella pneumoniae*		NEB
Arthrobacter luteus		NEB	*Myxococcus xanthus*	DK 1622	D. Kaiser
Bacillus subtilis	DB-2	R. Doi	*Neisseria gonorrheae*	FA 19	P.F. Sparling
Caryophanon latum		NEB	*Neisseria meningitidis*	ATCC 13077	E. Mason
Listeria monocytogenes	5032	E. Mason	*Proteus vulgaris*	ATCC 13315	NEB
Mycobacterium aurum		D. Stahl	*Pseudomonas aeruginosa*	4998,5014	E. Mason
Mycoplasma pneumoniae		D. Krause	*Rhizobium meliloti*	RM1021	F. J. deBruijn
Nocardia otitidiscaviarum	ATCC 14630	NEB	*Rhodobacter sphaeroides*	NEB 233	NEB
Staphylococcus aureus	PS96	NEB	*Salmonella sp.*	4077,4340, 4359	E. Mason
Streptococcus Group B	5023	E. Mason	*Salmonella typhi*	2304	E. Mason
Streptococcus pneumoniae	230	E. Mason	*Salmonella typhimurium*	LT-2	E. Mason
Streptomyces albus G		NEB	*Serratia marcescens*		NEB
GREEN NON-SULFUR BACTERIA	.....	.....	*Shigella flexneri*	106	E. Mason
Herpetosiphon giganteus	NEB 198	NEB	*Shigella sonnei*	104	E. Mason
RADIORESISTANT MICROCOCCI	.....	.....	*Shigella sp.*	170	E. Mason
Deinococcus radiophilus		NEB	*Sphaerotilus sp.*	ATCC 13925	NEB
Thermus aquaticus		NEB	*Vibrio vulnificus*	4029	E. Mason
Thermus thermophilus	NEB 249	NEB	*Xanthomonas manihotis*		NEB
SPIROCHETES	.....	.....	ARCHAEBACTERIA	.....	.....
Borrelia burgdorferi	G 2	L. Katz	*Halobacterium halobium*	89	N. Hackett
Treponema pallidum		R. Baughn	EUCARYOTES	.....	.....
Treponema phagedenis		R. Baughn	*Saccharomyces cerevisiae*		S. Elledge
			Schizosaccharomyce pombe	972	H. Klein
			Candida parapsilosis	4947	E. Mason
			Homo sapiens	-	J.R. Lupski

TABLE 1. Genomic DNA sources.

matching highly conserved extragenic, repetitive sequences were used as primers to amplify DNA between successive repetitive elements in different Gram-negative enterobacterial strains.

The REP-PCR genomic fingerprint of different strains/isolates from several bacterial species revealed distinct patterns [Fig. 4A]. PCR amplification of DNA from multiple strains of different enterobacterial species using primers REP1R-I and REP2-I [Fig. 4A] demonstrated subspecies or strain-specific band patterns. These DNA fingerprint patterns are reproducible and diagnostic for specific strains. Randomly chosen individual colonies on a plate or daily samples from 10-day serial cultures of a single strain yielded identical and consistent REP-PCR based DNA fingerprints (data not shown). In lanes 2 and 3 [Fig. 4A], *E. coli* K-12 strains HB101 and W3110 were distinguished clearly by an extra band of approximately 400 bp in W3110. The laboratory strains of *E. coli* K12 were related to each other and distinct from the pathogenic strains of *E. coli* [Fig. 4A]. Interestingly the *Salmonella typhimurium* laboratory strain LT-2 revealed a close similarity to *Salmonella typhi* strain 2304 [Fig. 4A]. Both of these strains showed REP-PCR patterns clearly distinct from other pathogenic *Salmonella* isolates of undetermined species [Fig. 4A]. The two *Klebsiella pneumoniae* strains shown were obtained from different sources [Table 1] and showed different banding patterns [Fig. 4A]. In lanes 14−15 and lanes 20−21 clinical isolates of pathogenic *Salmonella* and *Enterobacter sakazakii* respectively were represented by apparently identical REP-PCR patterns [Fig. 4A]. These 2 *Enterobacter sakazakii* isolates were isolated from the same hospital. In lane 13 a different isolate of pathogenic *Salmonella* contains a REP-PCR based DNA fingerprint distinct from the *Salmonella* in lanes 14−15 [Fig. 4A]. Interestingly the *Salmonella* isolates in lanes 13 and 14 were collected from different patients at different hospitals 10 months apart [Fig. 4A].

PCR amplification of different enteric bacterial species with the ERIC1R and ERIC2 primers also revealed species-specific band patterns [Fig. 4B]. Because the complexity was less than that obtained with REP-PCR, the differences between species were easier to distinguish. However, this decreased complexity of the genomic fingerprints made it more difficult to make fine distinctions between strains, e.g. *E. coli* laboratory strains HB101 and W3110. Interestingly, greater ERIC-PCR pattern differences existed when comparing laboratory strains of *E. coli* to pathogenic isolates of the same species than between laboratory *E. coli* strains and pathogenic *Shigella* species. The ERIC-PCR patterns of greatest complexity were observed with *Salmonella* [Fig. 4B] and these results are consistent with previous database searches revealing an abundance of ERIC in *Salmonella* (20). Both REP- and ERIC-PCR yielded common bands between strains of a given species [Fig. 4A, 4B] which enable one to group strains within a certain species.

The presence of REP and ERIC sequences in genomic DNA from diverse bacterial species

Extragenic, repetitive sequences, REP and ERIC, have been described in enteric bacteria (17−21). In addition, repetitive sequences have been found in *Neisseria* (47) and *Deinococcus* (48), which approximate the sizes of REP and ERIC elements respectively. Only one previous study has attempted to examine the conservation of REP sequences using nondegenerate probes in hybridization experiments (38).

A survey of the eubacterial kingdom represents an important test of evolutionary conservation for any DNA sequence since short generation times and long time spans afford the greatest opportunities for selective pressures to act. To assess evolutionary conservation of DNA sequences rapidly among eubacteria, the 'bug blot' was developed [Fig. 5A, 6A]. The presence of REP and ERIC elements throughout the eubacterial kingdom was examined by slot blot hybridization with a 'bug blot' and PCR analysis of genomic DNA from the same bacterial species.

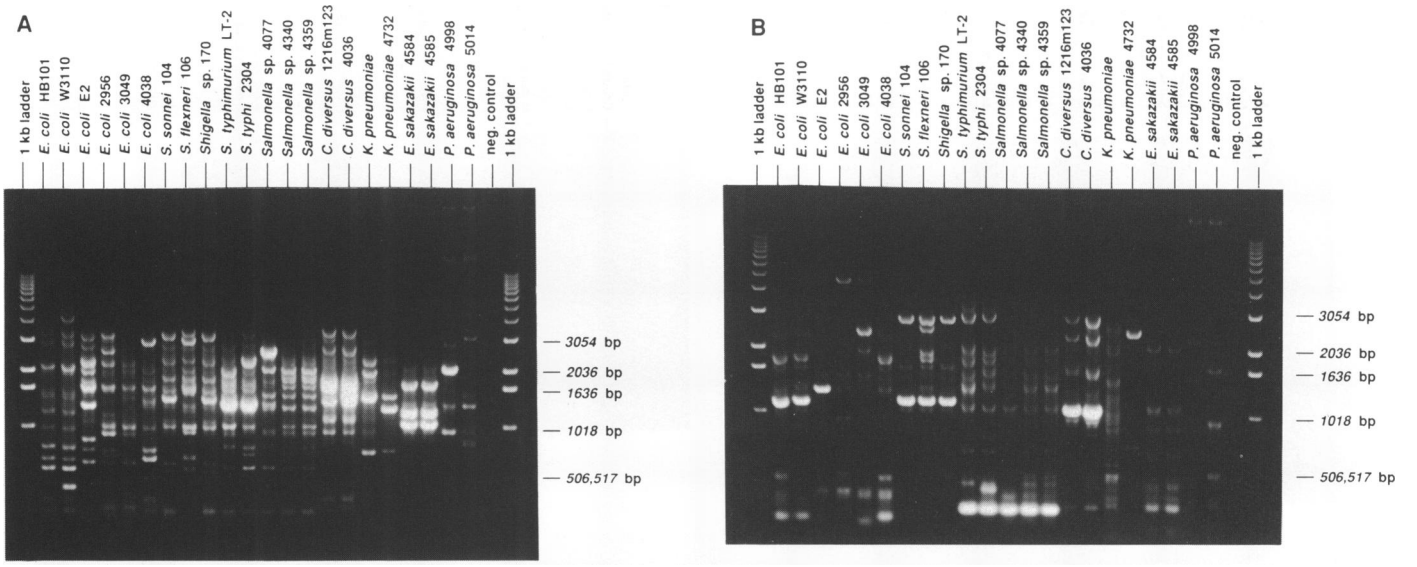

Fig. 4. REP- and ERIC-PCR distinguishes strains within Gram-negative enterobacterial species. (A) PCR products generated by amplification of enterobacterial genomic DNA with REP primers, REP1R-I and REP2-I. Complete names of species listed appear in Table 1. (B) PCR amplification of enterobacterial genomic DNA by ERIC oligonucleotide primers, ERIC1R and ERIC2. PCR reactions were performed as described in Materials and Methods. No template DNA was added to the negative control lanes. The DNA molecular weight marker is a 1-kb ladder [BRL]. The gels were 1% agarose-1×Tris-acetate-EDTA and contained 0.5 μg of ethidium bromide per ml.

Slot blot hybridization of the 'bug blot' with REP and ERIC probes in Figs. 5B and 6B, respectively, suggests that Gram-negative enterics and related species from the same phyla comprise the majority of REP- and ERIC-positive species. Hybridizations with REPALL-I and REP2-I [data not shown] yielded results similar to hybridization with REPALL-D [Fig. 5B]. The 38-mer REPALL probes were used for the hybridization because the increased length provides a longer homologous stretch and, hence, greater stability for hybridization. Indeed the REP2-I probe was effectively removed from the membrane with minimum washing. Hybridization with ERIC-ALL [data not shown] yielded results consistent with hybridization with ERIC2 [Fig. 6B]. As expected several species of Gram-positive bacteria and spirochetes in addition to the phylogenetically distant eucaryotic fungi failed to yield hybridization signals [Fig. 5B, 6B]. Surprisingly, hybridization signals were observed with the distantly related radioresistant bacterium, *Deinococcus radiophilus*, the green non-sulfur bacterium, *Herpetosiphon giganteus*, and the archaebacterium, *Halobacterium halobium* [Fig. 5B, 6B].

PCR amplification of these same bacterial species with primers REP1R-I and REP2-I yielded results consistent with the 'bug blot' hybridization described above. The species that showed the most intense hybridization signals in Fig. 5B generally demonstrated the most complex amplification patterns by REP-PCR [Fig. 5C, 5D]. PCR amplification of genomic DNA from different species clearly revealed species-specific REP patterns [Fig. 5C, 5D].

ERIC-PCR also provided results [Fig. 6C, 6D] consistent with ERIC hybridization of the 'bug blot' [Fig. 6B]. Gram-negative enteric species yielded the amplification patterns of greatest complexity [Fig. 6C,6D]. Most Gram-positive species [e.g. *Bacillus subtilis*] showed minimal ERIC-PCR amplification

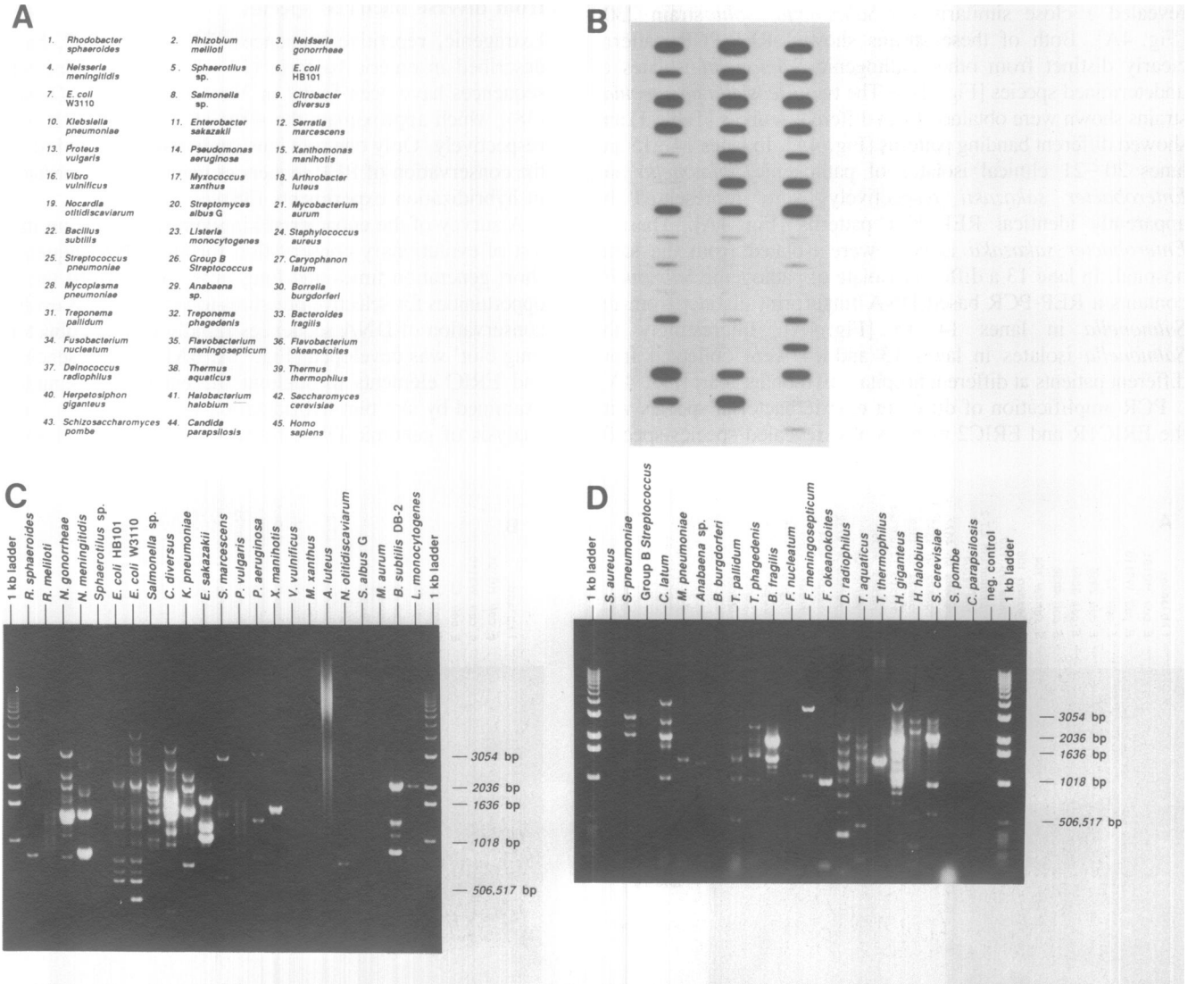

Fig. 5. Evolutionary conservation of REP sequences. (A) A listing of bacterial and non-bacterial species which matches the genomic DNA in each slot of the 'bug blot' hybridization presented in Fig. 5B. (B) The 'bug blot' representing a slot blot DNA:DNA hybridization of genomic DNA probed with ^{32}P-end-labeled REPALL-D. Filters were prepared and hybridizations were performed as described in Materials and Methods. (C) PCR amplification of bacterial genomic DNAs used in the 'bug blot' hybridization with REP primers, REP1R-I and REP2-I. These PCR reactions are presented in exactly the same order as the slots of the 'bug blot.' (D) A continuation of PCR amplifications of bacterial genomic DNAs used in the 'bug blot' hybridization. All PCR reactions were performed as described in Materials and Methods. No template DNA was added to the negative control lane. The DNA molecular weight marker is a 1-kb ladder [BRL]. The gels were 1% agarose-1×Tris-acetate-EDTA and contained 0.5 μg of ethidium bromide per ml.

[Fig. 6C,6D]. This result is consistent with a previous computer search of ERIC in the DNA sequence databases (20) and known phylogenetic distances between Gram-positive bacteria and Gram-negative enteric bacteria (32).

DISCUSSION

REP and ERIC primer sets were empirically determined which allowed the direct generation of unambiguous species- and strain-specific genomic fingerprints by PCR amplification of genomic DNA. Degenerate REP oligonucleotides were useful in detecting many REP-like elements in the eubacterial kingdom. A previous study examined the evolutionary conservation of REP sequences using nondegenerate, consensus oligonucleotides in dot blot and Southern hybridization experiments with different bacteria, primarily Gram-negative enterics (38). Significant REP hybridizations with only four enterobacterial genera, *Escherichia*, *Salmonella*, *Shigella*, and *Citrobacter* were reported. However, recent DNA sequence analysis has illustrated that *Enterobacter aerogenes* has highly conserved REP sequences located between the same genes as *E. coli* (49). The use of single sequence oligonucleotides may have prevented the detection of related sequences in a wide variety of organisms.

In this work we demonstrate that REP- and ERIC-like sequences are found primarily in Gram-negative enteric bacteria and its close relatives in the same phyla. This paper represents

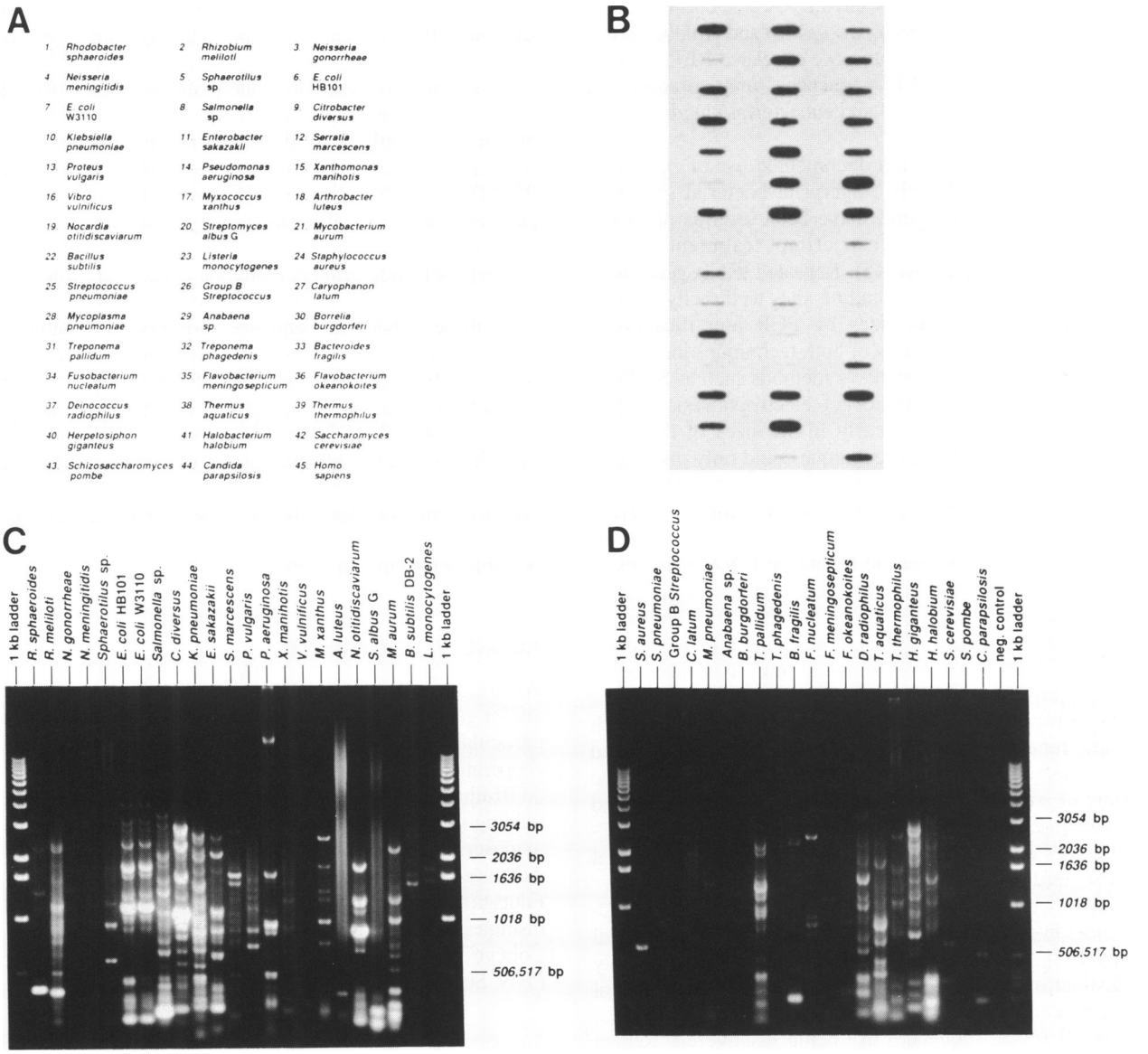

Fig. 6. Evolutionary conservation of ERIC sequences. (A) A listing of bacterial and non-bacterial species which matches the genomic DNA in each slot of the 'bug blot' hybridization presented in Fig. 6B. (B) The 'bug blot' representing a slot blot DNA:DNA hybridization of genomic DNA probed with ^{32}P-end-labeled ERIC2. Filters were prepared and hybridizations were performed as described in Materials and Methods. (C) PCR amplification of bacterial genomic DNAs used in the 'bug blot' hybridization with ERIC primers, ERIC1R and ERIC2. These PCR reactions are presented in exactly the same order as the slots of the 'bug blot.' (D) A continuation of PCR amplifications of bacterial genomic DNAs used in the 'bug blot' hybridization with ERIC primers, ERIC1R and ERIC2. All PCR reactions were performed as described in Materials and Methods. No template DNA was added to the negative control lane. The DNA molecular weight marker is a 1-kb ladder [BRL]. The gels were 1% agarose-1×Tris-acetate-EDTA and contained 0.5 μg of ethidium bromide per ml.

the first demonstration of ERIC-like sequences in eubacteria by methods other than computer-aided analysis of sequence data. Both REP (17,18) and ERIC (19,20) elements have been highly conserved within *E. coli* and *S. typhimurium*, and from this study appear to be conserved in the eubacterial kingdom for at least hundreds of millions of years. Their evolutionary conservation suggests that their existence precedes the formation of the Gram-negative enteric bacterial lineage. Surprisingly, some distant species such as *Deinococcus radiophilus*, *Herpetosiphon giganteus*, and even the archaebacterium *Halobacterium halobium*, revealed hybridization signals when probed with REP and ERIC oligonucleotides. However, the intensity of the signal observed with *Halobacterium halobium* is not reflected by a corresponding complexity of REP-PCR in the same species. This may be due to a greater spacing of REP-like elements in this archaebacterial chromosome. Repetitive DNA elements have been identified previously in both *Deinococcus* (48) and *Halobacterium* (50). If these results indicate the presence of related REP elements in archaebacteria, then REP sequences may predate the divergence of the archaebacterial and eubacterial kingdoms over 2 billion years ago (51).

This report represents the first documented use of extragenic repetitive sequences to directly fingerprint bacterial genomes. Previously repeated rRNA genes have been used as probes in Southern blots to detect restriction fragment length polymorphisms between strains (45). Repeated tRNA genes have been used as consensus primer binding sites to directly amplify DNA fragments of different sizes by PCR amplification of different strains (37). Limitations of both techniques include the use of radioisotope and time-intensive methods such as Southern blotting (45) and polyacrylamide gel electrophoresis (37) to distinguish subtle differences clearly in the sizes of the DNA fragments generated. The latter technique could only distinguish organisms at the species and genus level. The tDNA-PCR fingerprints are generally invariant between strains of a given species and between related species (37). Other previous studies include the use of species-specific repetitive DNA elements as primer-binding sites for PCR-based bacterial species identification (52,53). Though such methods allow species identification by PCR with picogram amounts of DNA, only single PCR products are generated which precludes the generation of strain-specific genomic fingerprints. REP- and ERIC-PCR allow clear distinctions between different bacterial species and strains which contain these repetitive elements. Direct PCR amplification and agarose gel electrophoresis of PCR products provide genomic fingerprints of sufficient complexity to distinguish species and strains. Combined with rapid cell lysis methods REP- and ERIC-PCR techniques enable one to characterize different bacterial strains in several hours.

In this paper we demonstrated the presence of REP and ERIC-like sequences in many different bacterial species. Some bacterial species which lack REP- and ERIC-like sequences probably contain distinct extragenic, repetitive sequences. As additional DNA sequence information is assembled in different microbial systems, novel repeat sequences are being discovered. Recent reports of novel repeated sequences in the eubacterial genera, *Deinococcus* (48), *Calothrix* (54), *Neisseria* (47), and the fungi, *Candida albicans* (55) and *Pneumocystis carinii* (56), illustrate the presence of dispersed extragenic repetitive sequences in many organisms.

Two mechanisms can account for the high degree of evolutionary conservation of these repetitive elements. First, natural selection may constrain variation in these sequences because they represent sites of essential protein:DNA interactions. In support of this hypothesis, DNA replication proteins, *E. coli* DNA gyrase (22,23) and polymerase I (26), specifically bind to REP sequences. Secondly, these sequences may propagate themselves as 'selfish' DNA by gene conversion (57). Retroposons are thought to represent a significant source of genetic diversity in evolution by their multiplication and creation of retropseudogenes (58). REP and ERIC sequences may be transcribed into RNA and dispersed through RNA intermediates. With the recent discovery of reverse transcriptase activity in eubacteria (59), gene conversion through RNA intermediates remains a possible mechanism accounting for the widespread presence and conservation of these prokaryotic repetitive elements (20,57). The occurrence of only internal changes in repetitive elements rather than movement to new genomic locations in the *Candida albicans* genome supports the hypothesis of dispersion by gene conversion, not random transposition (55).

In addition to genomic fingerprinting, PCR methodologies employing these REP and ERIC sequences as PCR primer binding sites could be used to study the distribution of repetitive sequences in different genomes. Using REP primers, different REP-PCR amplification patterns were observed with different purified cosmids [data not shown] from an ordered *E. coli* genomic library (60).

Oligonucleotide primers matching conserved protein sequence motifs or transposon insertion ends could be used in conjunction with these repetitive sequence primers to rapidly amplify unknown DNA sequences from many different bacterial species. An example of this strategy, REP-PCR with primers from the ends of Tn*5*, was used to map Tn*5* insertions in the *glpD* gene of *E. coli* (40). The characterization of new repetitive elements in different species will permit the use of this technique in the molecular genetic analysis of any microorganism. These amplified products could be used directly in cloning and direct sequencing methods to extract genomic information more rapidly than is possible with current methods.

ACKNOWLEDGEMENTS

We are grateful to the individuals in Table 1 who provided strains and/or genomic DNA. Thanks to the New England Biolabs and, particularly to Marta Meda and Joe Bonventre, for their cooperation in obtaining samples. We thank Jurgen Brosius, Frans de Bruijn, and George M. Weinstock for discussions and critical reading of the manuscript, and Linda Haway for her aid in the manuscript preparation. This research was supported by an Advanced Technology Program Grant from the Texas Higher Education Department to J.R.L. J.V. is supported by the National Institutes of Health Medical Scientist Training Program at Baylor College of Medicine. J.R.L. acknowledges support from the PEW Scholars Program in Biomedical Sciences.

REFERENCES

1. Schmid, C.W. and Jelinek, W.R. (1982) *Science*, **216**, 1065–1070.
2. Jelinek, W.R., Toomey, T.P., Leinwand, L., et al. (1980) *Proc.Natl.Acad.Sci.USA*, **77**, 1398–1402.
3. Houck, C.M., Rinehart, F.P and Schmid, C.W. (1979) *J.Molec.Biol.*, **132**, 289–306.
4. Saiki, R.K., Scharf, S., Faloona, F., et al. (1985) *Science*, **230**, 1350–1354.
5. Mullis, K.B. and Faloona, F.A. (1987) *Methods Enzymol.*, **155**, 335–350.

6. Nelson, D.L., Ledbetter, S.A., Corbo, L., et al. (1989) *Proc.Natl.Acad.Sci.USA*, **86**, 6686–6690.
7. Rikke, B.A., Garvin, L.D. and Hardies, S.C. (1991) *J.Molec.Biol.*, **219**, 635–643.
8. Doolittle, W.F. (1978) *Nature*, **272**, 581–582.
9. Bachmann, B.J. (1990) *Microbiol.Rev.*, **54**, 130–197.
10. Nesin, M., Lupski, J.R., Svec, P. and Godson, G.N. (1987) *Gene*, **51**, 149–161.
11. Yanofsky, C., Platt, T., Crawford, I.P., et al. (1981) *Nucleic Acids Res.*, **9**, 6647–6668.
12. Lawther, R.P., Wek, R.C., Lopes, J.M., Pereira, R., Taillon, B.E. and Hatfield, G.W. (1987) *Nucleic Acids Res.*, **15**, 2137–2155.
13. Chu, F.K., Maley, G.F., Maley, F. and Belfort, M. (1984) *Proc.Natl.Acad.Sci.USA*, **81**, 3049–3053.
14. Gott, J.M., Shub, D.A. and Belfort, M. (1986) *Cell*, **47**, 81–87.
15. Kuhsel, M.G., Strickland, R. and Palmer, J.D. (1990) *Science*, **250**, 1570–1573.
16. Xu, M-Q., Kathe, S.D., Goodrich-Blair, H., Nierzwicki-Bauer, S.A. and Shub, D.A. (1990) *Science*, **250**, 1566–1570.
17. Stern, M.J., Ames, G.F.L., Smith, N.H., Robinson, E.C. and Higgins, C.F. (1984) *Cell*, **37**, 1015–1026.
18. Gilson, E., Clement, J.M., Brutlag, D. and Hofnung, M. (1984) *EMBO J.*, **3**, 1417–1421.
19. Sharples, G.J. and Lloyd, R.G. (1990) *Nucleic Acids Res.*, **18**, 6503–6508.
20. Hulton, C.S.J., Higgins, C.F. and Sharp, P.M. (1991) *Molecular Microbiol.*, **5**, 825–834.
21. Higgins, C.F., Ames, G.F.L., Barnes, W.M., Clement, J.M. and Hofnung, M. (1982) *Nature*, **298**, 760–762.
22. Yang, Y. and Ames, G.F.L. (1990) In Drlica, K. (ed.), The Bacterial Chromosome. American Society for Microbiology, Washington, D.C. pp. 211–225.
23. Yang, Y. and Ames, G.F.L. (1988) *Proc.Natl.Acad.Sci.USA*, **85**, 8850–8854.
24. Gilson, E., Rousset, J.P., Clement, J.M. and Hofnung, M. (1986) *Ann.Inst.Pasteur*, **137**, 259–270.
25. Newbury, S.F., Smith, N.H., Robinson, E.C., Hiles, I.D. and Higgins, C.F. (1987) *Cell*, **48**, 297–310.
26. Gilson, E., Perrin, D. and Hofnung, M. (1990) *Nucleic Acids Res.*, **18**, 3941–3952.
27. Miller, J.H. (1972) Experiments in Molecular Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
28. Loenen, W.A.M., Daniel, A.S., Braymer, H.D. and Murray, N.E. (1987) *J.Molec.Biol.*, **198**, 159–170.
29. Martin, F.H., Castro, M.M., Aboul-ela, F. and Tinoco Jr, I. (1985) *Nucleic Acids Res.*, **13**, 8927–8938.
30. Kawase, Y., Iwai, S., Inoue, H., Miura, K. and Ohtsuka, E. (1986) *Nucleic Acids Res.*, **14**, 7727–7736.
31. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
32. Woese, C.R. (1987) *Microbiol.Rev.*, **51**, 221–271.
33. Noda, A., Courtwright, J.B., Denor, P.F., Webb, G., Kohara, Y. and Ishihama, A. (1991) *Biotechniques*, **10**, 474–477.
34. Kogan, S., Doherty, M. and Gitschier, J. (1987) *New Engl.J.Med.*, **317**, 985–990.
35. Leuven, F.V. (1991) *Trends Genet.*, **7**, 142.
36. Welsh, J. and McClelland, M. (1990) *Nucleic Acids Res.*, **18**, 7213–7218.
37. Welsh, J. and McClelland, M. (1991) *Nucleic Acids Res.*, **19**, 861–866.
38. Gilson, E., Bachellier, S., Perrin, S., et al. (1990) *Ann.Inst.Pasteur*, **141**, 1103–1116.
39. Austin, D.A. and Larson, T.J. (1991) *J.Bacteriol.*, **173**, 101–107.
40. Subramanian, P.S., Versalovic, J., McCabe, E.R.B. and Lupski, J.R. *PCR Methods Applic.* Rapid mapping of *Escherichia coli* insertion mutations by REP-Tn*5* PCR. In press.
41. Kohara, Y., Akiyama, K. and Isono, K. (1987) *Cell*, **50**, 495–508.
42. Versalovic, J., Koeuth, T., McCabe, E.R.B. and Lupski, J.R. (1991) *J.Bacteriol.*, **173**, 5253–5255.
43. Owen, R.J. (1989) *J.Med.Microbiol.*, **30**, 89–99.
44. Murray, B.E., Singh, K.V., Heath, J.D., Sharma, B.R. and Weinstock, G.M. (1990) *J.Clin.Microbiol.*, **28**, 2059–2063.
45. Carriere, C., Allardet-Servent, A., Bourg, G., Audurier, A. and Ramuz, M. (1991) *J.Clin.Microbiol.*, **29**, 1351–1355.
46. Litwin, C.M., Storm, A.L., Chipowsky, S. and Ryan, K.J. (1991) *J.Clin.Microbiol.*, **29**, 104–108.
47. Correia, F.F., Inouye, S. and Inouye, M. (1986) *J.Bacteriol.*, **167**, 1009–1015.
48. Lennon, E., Gutman, P.D., Yao, H. and Minton, K.W. (1991) *J.Bacteriol.*, **173**, 2137–2140.
49. Dahl, M.K., Francoz, E., Saurin, W., Boos, W., Manson, M.D. and Hofnung, M. (1989) *Mol.Gen.Genet.*, **218**, 199–207.
50. Sapienza, C. and Doolittle, W.F. (1982) *Nature*, **295**, 384–389.
51. Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) *Proc.Natl.Acad.Sci.USA*, **87**, 4576–4579.
52. Glare, E.M., Paton, J.C., Premier, R.R., Lawrence, A.J. and Nisbet, I.T. (1990) *J.Clin.Microbiol.*, **28**, 1982–1987.
53. Eisenach, K.D., Cave, M.D., Bates, J.H. and Crawford, J.T. (1990) *J.Infect.Dis.*, **161**, 977–981.
54. Mazel, D., Houmard, J., Castets, A.M. and Tandeau de Marsac, N. (1990) *J.Bacteriol.*, **172**, 2755–2761.
55. Scherer, S. and Stevens, D.A. (1988) *Proc.Natl.Acad.Sci.USA*, **85**, 1452–1456.
56. Stringer, S.L., Hong, S.-T., Giuntoli, D. and Stringer, J.R. (1991) *J.Clin.Microbiol.*, **29**, 1194–1201.
57. Higgins, C.F., McLaren, R.S. and Newbury, S.F. (1988) *Gene*, **72**, 3–14.
58. Brosius, J. (1991) *Science*, **251**, 753.
59. Inouye, M. and Inouye, S. (1991) *Trends Biochem.Sci.*, **16**, 18–21.
60. Tabata, S., Higashitani, A., Takanami, M., et al. (1989) *J.Bacteriol.*, **171**, 1214–1218.