Differential subsequence conservation of interspersed repetitive Streptococcus pneumoniae BOX elements in diverse bacteria.

T Koeuth, J Versalovic and J R Lupski

Genome Res. 1995 5: 408-418
Access the most recent version at doi:10.1101/gr.5.4.408

PCR METHODS AND APPLICATIONS

Differential Subsequence Conservation of Interspersed Repetitive *Streptococcus pneumoniae* BOX Elements in Diverse Bacteria

Thearith Koeuth,[1] James Versalovic,[2] and James R. Lupski[1,3,4]

Departments of [1]Molecular and Human Genetics, [2]Medicine, and [3]Pediatrics, Baylor College of Medicine, Houston, Texas 77030

Evolutionary conservation of an interspersed repetitive DNA sequence, BOX, from *Streptococcus pneumoniae* was investigated to explore the mosaic nature of these elements. BOX elements consist of various combinations of three subunits, boxA, boxB, and boxC. Eight oligonucleotide probes were designed based on consensus DNA sequences of boxA, boxB, and boxC subunits. DNA hybridization studies and PCR using these probes/primers demonstrate that oligonucleotide sequences within the boxA subunit appear to be conserved among diverse bacterial species. The boxB and boxC subunits show only limited, if any, sequence conservation in bacteria other than *S. pneumoniae*. Intact BOX elements with boxA, boxB, and boxC subunits were only present in high copy number in pneumococcal strains. This pattern of differential conservation lends support to the modular nature of BOX repetitive elements in that boxA-like subsequences are effectively independent of boxB-like or boxC-like subunits in bacteria other than *S. pneumoniae*. Furthermore, dendrograms derived from repetitive sequence-based PCR (rep–PCR) fingerprints of *S. pneumoniae* isolates using the BOXA1R primer yielded clustering patterns that were similar to those obtained previously by other methods, suggesting that these repetitive sequence-based DNA fingerprints represent intrinsic properties of an *S. pneumoniae* strain's genome. Our results indicate widespread conservation of boxA-like subsequences in the bacterial kingdom, lend support to the mosaic nature of BOX in *S. pneumoniae*, and demonstrate the utility of boxA-based primers for rep–PCR fingerprinting of many microorganisms.

Several short interspersed repetitive DNA sequences have been identified in prokaryotic genomes.[1] Examples of well-characterized repetitive DNA sequences in bacteria include the palindromic units (PU) or repetitive extragenic palindromes (REP),[2–4] and the intergenic repeat units (IRU) or enterobacterial repetitive intergenic consensus (ERIC).[5–6] Recently, an interspersed repetitive DNA sequence, BOX, was identified from the Gram-positive bacterium *Streptococcus pneumoniae*.[7] BOX elements have dyad symmetry with the potential to form stable stem-loop structures and are located within intergenic regions. BOX elements are mosaic repetitive sequences composed of various combinations of three subunits, boxA, boxB, and boxC, which are 59, 45, and 50 nucleotides long, respectively (Fig. 1A).[7] For example, the boxB subunit was present alone as a single copy or as a variable number of direct tandem repeats flanked by boxA and boxC.[7] The DNA sequences of the BOX elements are entirely different from the prokaryotic interspersed repetitive DNA sequences REP and ERIC, although there are similarities to REP and ERIC with respect to size, copy number, and potential to form stable stem-loop structures.[2–7]

Although the exact functions of interspersed repetitive DNA elements are unknown, the presence of these repetitive DNA sequences can be utilized for rapid physical mapping procedures[8] and for DNA fingerprinting of prokaryotic genomes.[4,9–23] REP and ERIC sequences were used to design primers for PCR, in a technique known as repetitive sequence-based PCR (rep–PCR), to obtain DNA fingerprints from various microorganisms. Interspersed repetitive sequences can serve as primer binding sites that are separated by various distances in the bacterial chromosome. PCR of unique sequence located between interspersed repeats results in differently sized DNA amplification products. PCR products of different sizes constitute polymorphic DNA markers and yield DNA fingerprints that may be specific for individual bacterial strains or isolates. The apparent evolutionary conservation of these repetitive elements enables the use of a limited primer repertoire for DNA fingerprinting of a wide array of bacteria.

The aim of this study was to examine

[4]Corresponding author.
E-MAIL jlupski@bcm.tmc.edu; FAX (713) 798-5073.

5:408–418 ©1995 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/95 $5.00

the mosaic nature of BOX elements by studying the evolutionary conservation and distribution of each BOX subsequence, boxA, boxB, and boxC, in diverse bacteria. Oligonucleotide probes complementary to these individual subsequences were synthesized and utilized as probes in slot–blot and Southern hybridization[24] with total genomic DNA. The same oligonucleotide probes were used as primers to generate DNA fingerprint patterns by rep–PCR. The ability to discern relationships among bacterial isolates with different interspersed repetitive sequence primers was also examined. Our data suggest that BOX elements are present in *S. pneumoniae* and *S. agalactiae*. BOX-like elements are modular in nature, and subsequences of BOX elements are differentially conserved in bacteria. Individual sequences within the boxA subunit appear to be conserved in many bacteria, whereas the boxB-like and boxC-like subunits are predominantly observed only in *S. pneumoniae*.

MATERIALS AND METHODS

Bacterial Strains

The sources of most bacterial strains and/or genomic DNA were described previously.[9,13] Additional strains used in this study are listed in Table 1.

Isolation and Quantitation of Genomic DNA

Bacterial cells were grown in Luria–Bertani (LB) broth[25] or scraped directly from culture plates. Cell pellets from both the liquid and culture plates were suspended/washed in 1 ml of 1 M NaCl. The samples were centrifuged, and the pellets were suspended/washed in 1 ml of TE buffer (50 mM Tris, 50 mM EDTA at pH 7.8). Chromosomal DNAs were isolated and quantitated as described previously.[9] Lysostaphin (Sigma) (50 U/ml) was used instead of Mutanolysin (Sigma) for lysis of staphylococcal cells.

Oligonucleotide Synthesis and Design

Oligonucleotide primers were synthesized using a model 380B DNA synthesizer (Applied Biosystems) at the Nucleic Acids Core Facility in the Department of Molecular and Human Genetics, Baylor College of Medicine. Published sequence data[7] were used for oligonucleotide primer design. The 22-mer primers that are complementary to boxA, boxB, and boxC subunits, respectively, are listed in Figure 1.[7] The REP1R-Dt[9] and REP2-Dt[13] primer sequences were described previously.

5′ End-labeling of Oligonucleotide Probes

Primers (50 pmoles) were labeled with 5 µl of radioactive isotope [γ-^{32}P]ATP [sp. act. 4500 Ci/mmole (ICN)] and 20 units of T4 polynucleotide kinase (Pharmacia) for 45 min at 37°C. Unincorporated residues were separated from the radioactive labeled primer by dilution of ~40 µl of labeled sample in 1000 µl of distilled water and centrifugation through Centricon-3 (Amicon) filters.

Slot–blot Hybridization Conditions

A slot–blot membrane, called the "bug blot,"[9] contained 44 chromosomal DNAs from seven different bacterial phyla, as defined by Woese.[26] The bug blot was prepared by using the Sure Blot Hybridization Membrane (Oncor); each slot contains 100 ng of chromosomal DNA from specific bacterial species listed in Figure 2A. One hundred nanograms of chromosomal DNA was diluted into distilled water to a total volume of 0.3 ml. DNA samples were heated at 100°C for 5 min and applied to the membrane. After evacuating the solution, 500 µl of 0.4 N NaOH was added to each slot. Membranes were rinsed in 1 × SSC, blotted dry with Whatman paper, and baked at 80°C for 1 hr. Membrane pretreatment and slot–blot preparation were performed as described.[27]

Hybridization solution was 5 × SSC, 0.5% casein (Sigma), 0.1% Sarkosyl, and 0.02% SDS as described for use with oligonucleotide probes on the *Escherichia coli* gene mapping membrane.[28] All membranes were prehybridized for 2 hr

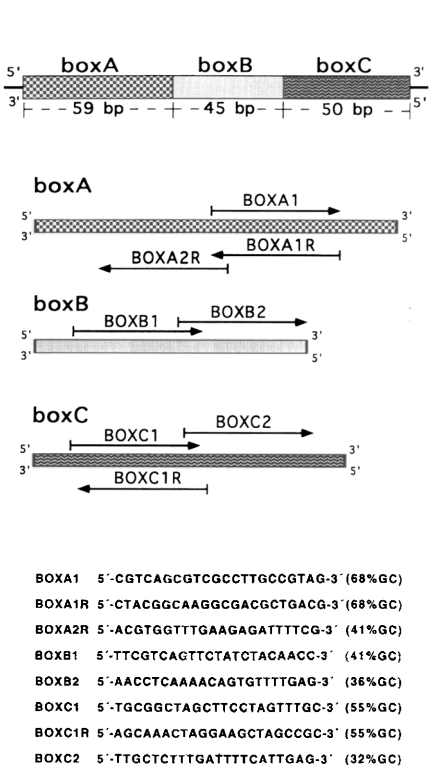

FIGURE 1 The *S. pneumoniae* interspersed repetitive BOX element structural model and primer design. (*A*) The BOX element consists of three subunits, boxA, boxB, and boxC, and may consist of various combinations of these subunits. The size of each BOX subunit is shown underneath. Positions of the BOX forward and reverse primers within the boxA, boxB, and boxC subsequences are shown by arrows with the arrowheads depicting the 3′ end of each BOX primer. (*B*) The actual sequences of primers were designed to be complementary to the consensus sequences of boxA, boxB, and boxC subunits, respectively.[7] The percent GC contents are indicated next to each primer sequence.

TABLE 1 Sources of Bacterial Strains and/or Genomic DNA

Species	Source
Chlorobium vibrioforme	D. Majumdar (Brown University, Providence, RI)
Coxiella burnetii	L.P. Mallavia (Washington State University, Pullman)
Haemophilus influenzae	J.M. Musser (Baylor College of Medicine, Houston, TX)
Leptospira interrogans	M. Fukunaga (University of Fukuyama, Hiroshima, Japan)
Pasteurella haemolytica	G. Weinstock (The University of Texas Medical School, Houston)
Streptococcus pyogenes	J.M. Musser (Baylor College of Medicine, Houston, TX)
Treponema vincentii	R. Baughn (Baylor College of Medicine, Houston, TX)
Yersinia enterocolitica	V. Miller (University of California, Los Angeles)

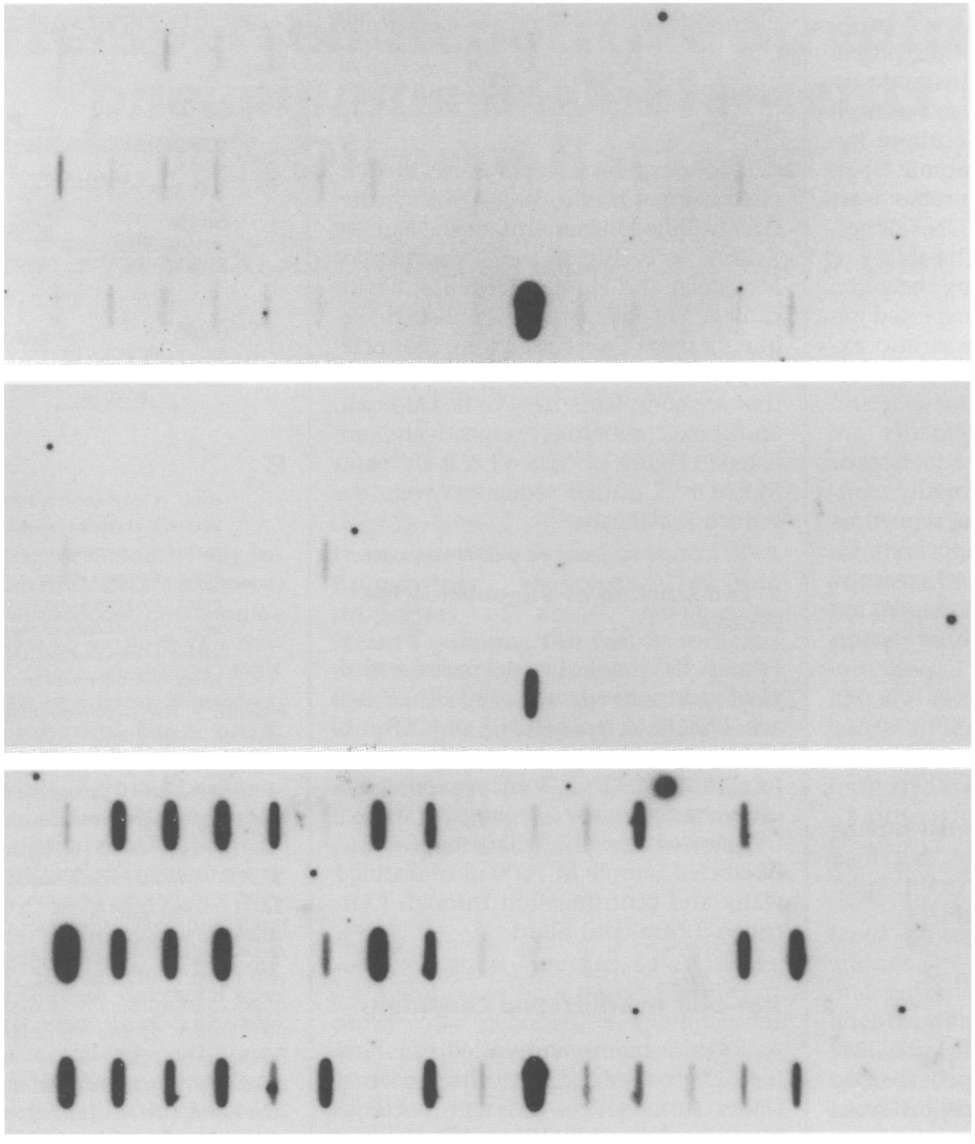

FIGURE 2 Evolutionary conservation of BOX subunit sequences by bug blot DNA hybridization. (*A*) List of bacterial species used as sources for genomic DNAs. Species names correspond to genomic DNA added to slots in hybridization panels in *B*, *C*, and *D*. Hybridizations were performed with end-labeled BOXA1 (*B*), BOXB1 (*C*), or BOXC1 (*D*) probes, respectively.

at 50°C, followed by 18 hr of hybridization with specific radioactive BOX probes (Fig. 1B) at 50°C. Probe was used at 1×10^6 cpm/ml of hybridization solution and added to the prehybridized bug blot membrane. After hybridization, the membranes were washed three times with 2× SSPE, 0.1% SDS for 5 min at room temperature. The membranes were washed one final time with 1× SSPE, 0.1% SDS for 1 min at room temperature. All membranes were blotted dry with Whatman paper and exposed on Kodak X-OMAT film with two intensifying screens at –80°C for 24 hr.

Southern Hybridization Conditions

One microgram of genomic DNA of each bacteria was digested with each of two restriction enzymes [either *Eco*RI (40 units) and *Bam*HI (40 units) or *Eco*RI (40 units) and *Hin*dIII (35 units)] and incubated at 37°C for 7 hr. Digested genomic DNA samples were added to each well of a 1.0% agarose gel, electrophoresed at 4 V/cm for 8 hr, and stained subsequently with 0.5 μg/ml of ethidium bromide. Following electrophoresis, the DNA in the gel was transferred to a Sure Blot Hybridization Membrane (Oncor) by osmotic transfer for 15 hr at room temperature. End-labeled BOX probes were hybridized separately with the membrane at 50°C for 18 hr. Membranes were washed three times with 2 × SSPE, 0.1% SDS for 5 min at room temperature and one time with 1× SSPE, 0.1% SDS for 1 min at room temperature. The washed membranes were exposed on Kodak X-OMAT film with two intensifying screens at –80°C for 24 hr.

PCR Conditions

PCR amplifications were performed as described previously,[9] except that 0.625 mM of each of four dNTPs (Pharmacia) per reaction were used. PCR reaction conditions were as follows: an initial denaturation step at 95°C for 7 min, followed by 30 cycles of denaturation at 90°C for 30 sec; variable annealing temperature (BOX, 52°C; REP, 40°C) for 1 min; and extension at 65°C for 8 min with a single final extension at 65°C for 16 min. Amplification with the BOXA2R, BOXB1, BOXB2, and BOXC2 oligonucleotide primers were also performed at a lower annealing temperature of 40°C to compensate for the lower primer GC contents. PCR amplifications with BOX primers included only a single primer except when the BOXA1 and BOXC1R primers were used together (50 pmoles each). All PCR reaction tubes were placed in the internal rows of the thermal cycler and surrounded by tubes containing only water and mineral oil at peripheral rows of the thermal cycler.[29]

Agarose Gel Preparation and Electrophoresis Conditions

Gels of 1.5% agarose (GIBCO BRL) and 1× Tris-acetate/EDTA (TAE)[27] were used in electrophoresis in 1 × TAE buffer at 5 V/cm for 8–10 hr. Eight microliters of each PCR reaction sample was loaded in each gel slot. After electrophoresis, gels were stained with 0.5 μg/ml of ethidium bromide in 1× TAE for 30–45 min. Ethidium bromide-stained PCR amplification products were visualized and photographed with a short wavelength ultraviolet source for 45–60 sec on Polaroid type 55 film.

Cluster Analysis and Dendrogram Preparation

Gel data documented on Polaroid Type 55 film were developed into 8 × 10-inch prints for measurement of gel migration distances of PCR products. The distance migrated by each DNA band was measured from the well origin. All specific distances (band positions) were then scored for presence ("1") or absence ("0") of a band at that position in the fingerprint from a given bacterial isolate. Simple Matching[30] similarity coefficients were generated based on this binary data with the SIMQUAL subprogram within NTSYS–PC version 1.70 (Applied Biostatistics, Inc., Setauket, NY). Cluster analysis and resultant dendrograms were generated by the UPGMA method using the SAHN and TREE subprograms in NTSYS-pc.

Data Base Queries with BOX Subsequence Probes

The oligonucleotide sequences, BOXA1, BOXA1R, BOXA2R, BOXB1, BOXB2, BOXC1, BOXC1R, and BOXC2, and the BOX subunit consensus sequences, boxA, boxB, and boxC,[7] were used as probes against the GenBank (version 80) and European Molecular Biology Laboratory (EMBL) (v. 35) DNA sequence data bases. Queries were performed with the FASTA[31] program in the Genetics Computer Group (GCG) software package.[32]

RESULTS

Differential Evolutionary Conservation of BOX Element Subsequences in Bacteria

Evolutionary conservation of the BOX subsequences was assessed by DNA hybridization and PCR with the oligonucleotides complementary to individual BOX subunits (Fig. 1A,B). Slot–blot hybridization experiments consisted of the BOX subunit probes and the bug blot[9] containing chromosomal DNAs from 44 different species, representing 7 of 10 different bacterial phyla.[26] Southern hybridization experiments were performed with the same oligonucleotide probes to confirm the patterns of evolutionary conservation and examine the distribution and relative copy number of these elements in different genomes. PCR-based DNA fingerprinting experiments were also performed with the different BOX subunit primers and diverse bacterial DNA templates.

Subsequences within boxA appear to be conserved in diverse bacterial species (Fig. 2B). Hybridization of the bug blot with the consensus boxA probes, BOXA1, BOXA1R, and BOXA2R, yielded discrete hybridization signals with a majority of organisms present on the filter. Both Gram-positive and Gram-negative organisms representing different phyla hybridized with the boxA subunit probes complementary to 41 of 59 nucleotides of the consensus boxA sequence. A prominent slot–blot hybridization signal with *S. pneumoniae* was obtained with each of the three boxA probes (Fig. 2B; data not shown). Surprisingly, other streptococcal species, *Streptococcus agalactiae* and *Streptococcus pyogenes*, failed to hybridize with the BOXA1R or BOXA1 probes (Fig. 2B; data not shown) and demonstrated that intact boxA subunits were not present in these related Gram-positive bacteria. In contrast, hybridization of the same bug blot with oligonucleotide probes matching the boxB subunit, BOXB1 and BOXB2, yielded prominent hybridization signals only with *S. pneumoniae* (Fig. 2C; data not shown). Therefore, the boxB subunit is not present in the bac-

teria examined other than *S. pneumoniae*. Slot–blot hybridization with the boxC probes, BOXC1, BOXC1R, and BOXC2, revealed a very limited degree of conservation, but, clearly, the most prominent signal was obtained with *S. pneumoniae* (Fig. 2D; data not shown).

Southern hybridization data support the results obtained with slot–blot hybridization and indicate the widespread genomic distribution of boxA-like subsequences. Subsequences of boxA appear to represent interspersed repetitive elements present in at least 14 different species belonging to five different phyla tested (Figs. 3A and 4C; data not shown). The boxA subsequences were present in highest copy number in the genomes of the three *S. pneumoniae* strains when each of three boxA probes were used, suggesting the presence of intact boxA subunits in *S. pneumoniae*. Interestingly, the other streptococci examined, *S. pyogenes* and *S. agalactiae*, and the Gram-positive bacterium *Staphylococcus aureus* did not yield signals with the BOXA1R or BOXA1 probes, suggesting the lack of intact boxA-like subunits (Fig. 3A). Organisms distantly related to *S. pneumoniae* may contain intact boxA-like elements, whereas several closely related Gram-positive bacteria lack this repetitive sequence. The boxB-like subunits were only present, though distributed differently, in the *S. pneumoniae* strains (Figs. 3B and 4D). Like boxB, the boxC-like sequences were present primarily in the pneumococcal strains (Fig. 3C; data not shown). As expected, hybridization signals were most intense with *S. pneumoniae* genomic DNA.

To compensate for differences in GC content between hybridization probes, BOXA2R and BOXB1, each having a GC content of 41% (Fig. 1B), were used as probes with the same conditions for slot and Southern hybridization experiments. The BOXA2R probe was highly conserved in diverse species, whereas the BOXB1 probe only matched sequences in the *S. penumoniae* genome (Fig. 4). Many of the bands that contain boxA-like sequences in *S. pneumoniae* also contain boxB-like and boxC-like sequences, but differences exist in the banding patterns with each probe. These results suggest that though these subsequences may often be found together, they also can exist independently of one another.

PCR-based DNA fingerprinting with these BOX subunit oligonucleotides as primers yielded results consistent with those obtained by DNA hybridization (Fig. 5). That is, the boxA subunit primers, BOXA1, BOXA1R, and BOXA2R, generated complex DNA fingerprints among diverse bacterial species. Again, the streptococcal species other than *S. pneumoniae*, *S. agalactiae*, and *S. pyogenes* failed to provide complex DNA fingerprints. Because of its lower GC content, lower stringency conditions were required for rep–PCR with BOXA2R. However, similar low-stringency amplification conditions with the BOXB1 and BOXB2 primers failed to yield distinct DNA fingerprints in organisms other than *S. pneumoniae* (data not shown). The boxC subunit oligonucleotide primers yielded DNA fingerprints of limited complexity with different bacterial species (data not shown). Only the oligonucleotide sequences within boxA appear to be conserved and useful for PCR-based DNA fingerprinting with a large array of bacterial species.

The above rep–PCR experiments with BOX subsequence primers utilized outwardly facing primers to amplify unique sequences located between individual BOX elements or subsequences in the genome. To examine the structural integrity of BOX elements in various microorganisms, inwardly facing boxA and boxC primers were used in the PCR. Intact BOX elements are defined as those elements that contain variable numbers of boxB subunits bounded on opposite sides by boxA and boxC subunits. PCR amplification (Fig. 6A,B) with the primer combination BOXA1 and BOXC1R followed by subsequent hybridization with the BOXB1 probe (Fig. 6C,D) indicated that *S. pneumoniae* contains intact BOX elements of uniform size. Different bacterial species revealed multiple PCR amplicons of different sizes but do not appear to contain intact BOX elements. An exception appears to be *S. agalactiae* that appears to contain intact BOX elements in low copy number that were only detectable with serial PCR amplification and DNA hybridization (Fig. 6A,C). Four differently sized PCR products (~104, 149, 194, and 239 bp) predominate in amplification reactions with many different pneumococcal strains (Fig. 6B) and presumably correspond to BOX elements containing 1, 2, 3, and 4 boxB subunits, respectively. The presence of variable numbers of boxB subunits within BOX elements lying between the

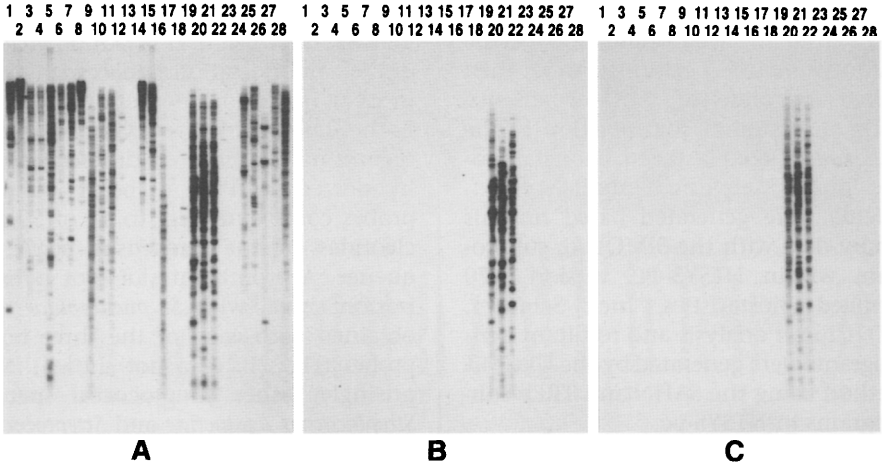

FIGURE 3 Nature of the genomic distribution and repetition of BOX subsequences in diverse bacteria. Southern hybridizations were performed with the end-labeled BOXA1R (*A*), BOXB1 (*B*), or BOXC1 (*C*) probe, respectively. The genomic DNAs (BE, DNA digested with *Bam*HI and *Eco*RI; BH, *Bam*HI and *Hin*dIII) were obtained from different bacterial species in the following order: (1) *Neisseria gonorrhoeae* BE, (2) *Sphaerotilus natans* BE, (3) *E. coli* W3110 BH, (4) *E. coli* MG1655 BH, (5) *Shigella* sp. 170 BH, (6) *Salmonella typhimurium* BH, (7) *Salmonella typhi* BH, (8) *Citrobacter diversus* BH, (9) *Klebsiella pneumoniae* BE, (10) *Pseudomonas aeruginosa* BE, (11) *Xanthomonas holcicola* BE, (12) *Haemophilus influenzae* BE, (13) *Vibrio vulnificus* BE, (14) *Myxococcus xanthus* BH, (15) *Arthrobacter luteus* BH, (16) *Nocardia otitidiscaviarum* BE, (17) *B. subtilis* BE, (18) *S. aureus* BE (19) *S. pneumoniae* 158 BE, (20) *S. pneumoniae* 242 BE, (21) *S. pneumoniae* 294 BE, (22) *S. pyogenes* BE, (23) *S. agalactiae* BE, (24) *Treponema vincentii* BE, (25) *Deinococcus radiophilus* BE, (26) *Thermus thermophilus* BE, (27) *Herpetosiphon giganteus* BE, and (28) *Halobacterium halobium* BE.

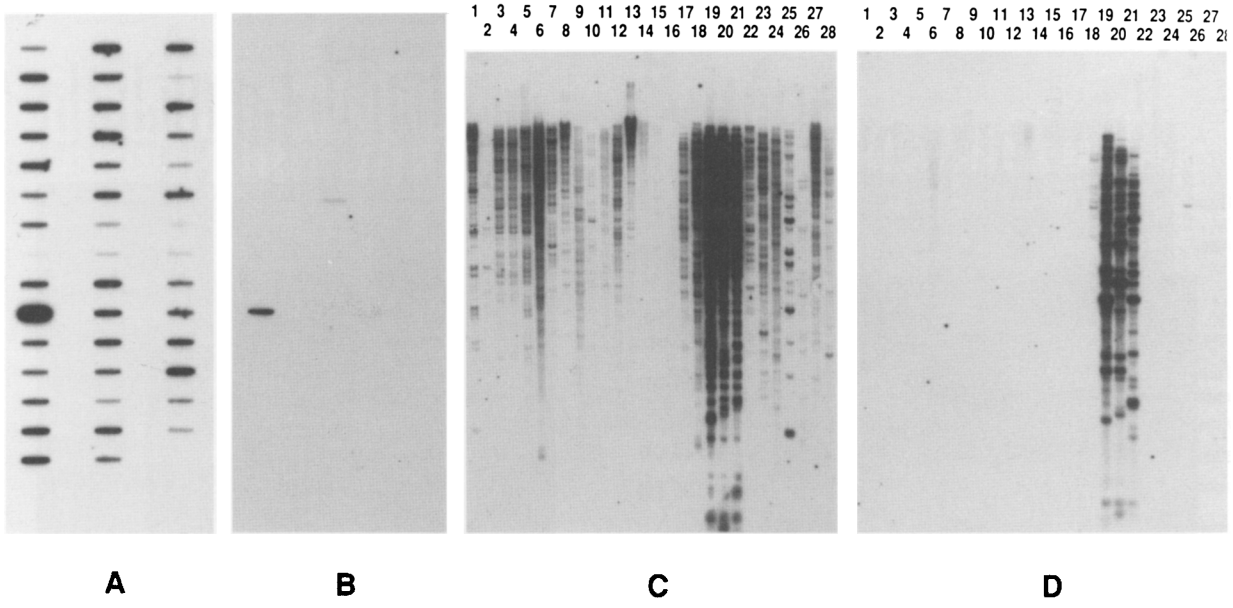

FIGURE 4 Comparative slot–blot and Southern hybridizations with the equivalent GC content (41%) BOXA2R and BOXB1 probes. Slot–blot hybridization with BOXA2R (*A*) and BOXB1 (*B*) as indicated in Fig. 2. Southern hybridization with BOXA2R (*C*) and BOXB1 (*D*) as indicated in Fig. 3.

boxA and boxC subunits has been documented previously by sequence analysis.[7] The diffuse hybridization signals (Fig. 6D) that are absent with PCR amplification only (Fig. 6B) probably represent hybridization of minor amplicons and unamplified genomic DNA.

As further examination of BOX subsequence conservation in the genomes of various microorganisms, BOX subsequence probes were used to search the GenBank and EMBL data bases. These analyses are of limited utility because the data bases are biased toward coding sequences and, apart from *E. coli* and *Bacillus subtilis*, there are limited entries for most microorganisms. Nevertheless, DNA sequence data base queries with the same oligonucleotide probes listed in Figure 1 and BOX subunit consensus sequence probes[7] provided results (data not shown) consistent with those obtained empirically by slot–blot and Southern hybridization. That is, the BOXA1 and BOXA1R probes extracted the greatest number of high-stringency matches with prokaryotic sequences in GenBank and EMBL, relative to BOXB1, BOXB2, BOXC1, BOXC1R, and BOXC2. Data base searches with probes containing the entire boxA, boxB, and boxC consensus sequences[7] yielded the greatest number of prokaryotic sequence matches with the boxA consensus probe.

Comparative Analysis of PCR-based DNA Fingerprints and Dendrogram Construction

A previous rep–PCR DNA fingerprinting study with primers matching REP sequences in penicillin-resistant *S. pneumoniae* revealed 17 different PCR-based DNA fingerprint patterns among 53 isolates.[13] These *S. pneumoniae* isolates were reexamined by PCR-based fingerprinting with primers matching the REP sequences and the boxA subunit. A representative sample of 13 isolates with 13 different DNA fingerprint patterns using both primer sets is shown in Figure 7. All 13 isolates were shown previously to be different from each other by multilocus enzyme electrophoresis (MLEE) and rep–PCR using REP primers.[13] Differences between these isolates were confirmed by rep–PCR with the BOXA1R primer (Fig. 7). All *S. pneumoniae* isolates yielded complex DNA fingerprint patterns with the BOXA1R primer, suggesting that the boxA-like subsequences are conserved, dispersed, and present in high copy number in all of these strains.

Twenty-three similar and distinct *S. pneumoniae* isolates were selected from a previous study[13] and subjected to PCR-based DNA fingerprinting with primers matching the REP sequences and the boxA subunit. Similarity coefficients were computed, and dendrograms were generated depicting the cluster analysis of these isolates (Fig. 8). Interestingly, the dendrograms are similar regardless of the primer set, REP1R-Dt plus REP2-Dt or BOXA1R alone, used to generate the DNA fingerprints. The largest cluster at the 1.0 identity level in both dendrograms contains isolates that are either identical or very closely related by multilocus enzyme electrophoresis [electrophoretotypes (ETs) 1 and 2].[13] Other pairs belonging to the same ET (ETs 8 and 21) were clustered together in at least one dendrogram. Finally, the only isolate, isolate 168, that was serologically nontypeable and very distantly related from other isolates by MLEE[13] was also the most genotypically dissimilar isolate by rep–PCR-based DNA fingerprinting with either primer set (Fig. 8).

DISCUSSION

DNA probes matching different subunits of *S. pneumoniae* interspersed repetitive BOX elements revealed differential conservation of these subsequences in bac-

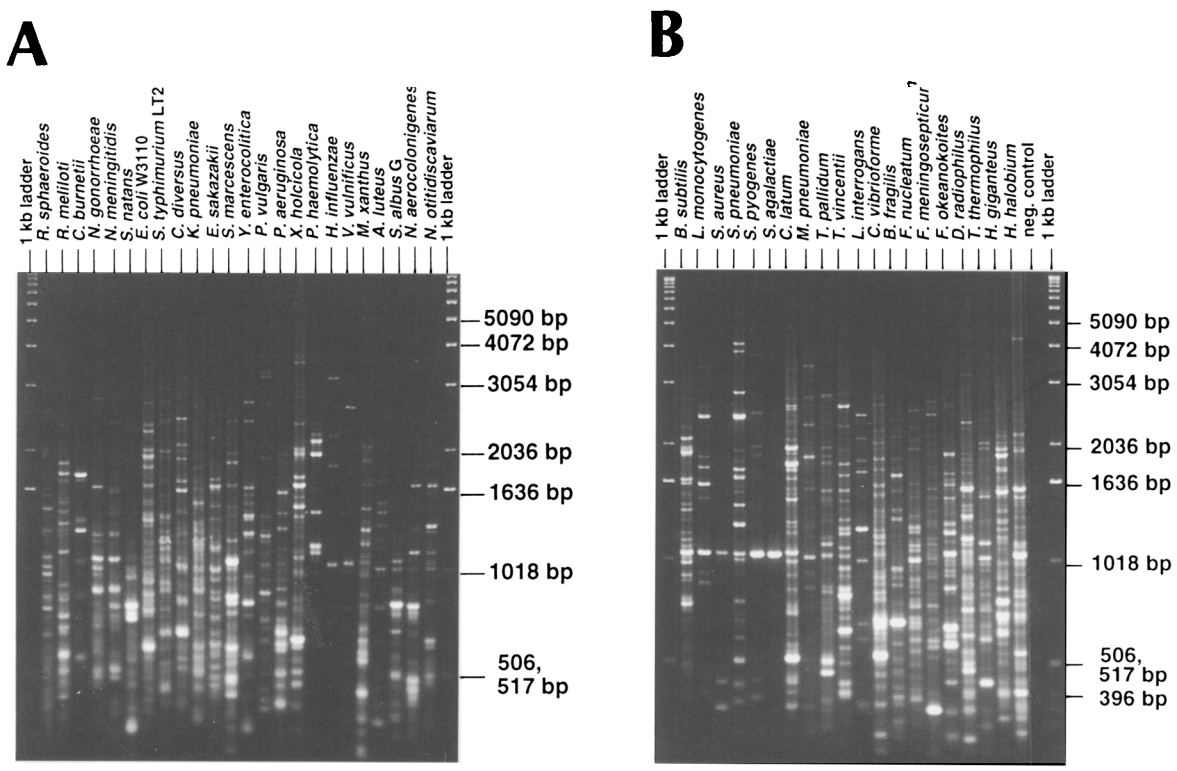

FIGURE 5 PCR-based DNA fingerprinting with the BOXA1R primer. (*A,B*) PCR-based DNA fingerprinting with the BOXA1R primer and template genomic DNA from the same bacteria listed in Fig. 2A. The negative control (*neg. control*) lane represents the same PCR reaction without template genomic DNA. The DNA molecular weight marker was a 1-kb ladder (GIBCO BRL). Gels were 1.5% agarose and 1 × TAE and were stained with 0.5 μg/ml of ethidium bromide after electrophoresis.

teria and support the proposal that BOX elements are mosaic repetitive sequences. Intact BOX elements containing the boxA, boxB, and boxC subunits are present almost exclusively in *S. pneumoniae*. Individual subsequences within boxA appear to be conserved in diverse bacteria, whereas the boxB and boxC subunits appear to be present in high copy numbers only in the genome of *S. pneumoniae*. DNA fingerprints generated by genomic Southern hybridization with each of the BOX subunit probes are different from one another and suggest the structural independence of the boxA, boxB, and boxC subsequences. Therefore, BOX elements appear to be mosaic elements containing different combinations of the boxA, boxB, or boxC repetitive sequences. Oligonucleotide primers complementary to the boxA subunit, in contrast to the boxB and boxC subunit primers, yielded complex and distinct PCR-based DNA fingerprints in diverse bacterial species.

Computer-aided analysis of PCR-based DNA fingerprints of different pneumococcal isolates generated with REP and boxA oligonucleotides yielded very similar clustering patterns. These clusters were also consistent with previous data obtained by MLEE. DNA fingerprints obtained using primers complementary to interspersed repetitive sequences, such as BOX and REP, therefore appear to represent intrinsic properties of a specific strain's genome because similar clustering patterns are observed among isolates regardless of the primer set used.

Our evolutionary conservation and DNA fingerprinting studies support the modular nature of BOX elements that was first suggested by the original DNA sequencing data obtained from *S. pneumoniae*.[7] The boxA, boxB, and boxC subunits are differentially conserved in bacteria. The boxA subsequences appear to be conserved in a wide variety of bacteria from different phyla when moderate-stringency conditions are used. These data suggest that boxA-like subunits may be conserved and maintained independently of boxB-like and boxC-like subunits. Limited DNA sequence information from *S. pneumoniae* previously suggested the modular nature of the BOX elements because the boxB subunit was found by itself, apart from the other subunits.[7] Interestingly, the boxB subunit was the most highly conserved BOX subunit within *S. pneumoniae*[7] and repesents a subunit uniquely found in high copy number in the pneumococcal genome (Figs. 2–4 and 6). Complex bacterial interspersed mosaic elements (BIMEs) are composed of blocks of different repetitive elements clustered together.[33–36] Such complex mosaic repetitive elements are composed of different combinations of smaller, modular interspersed repetitive sequences. Although BIME have been found in different enterobacterial genera,[33,35] the individual sequence motifs are generally species-, or at least, genus-specific.

Diverse bacterial species differ in genomic GC content, and this may affect interpretation of our results. However, the inclusion of Gram-positive bacteria closely related to *S. pneumoniae* that possess similar GC contents represented an important control. Significant differ-

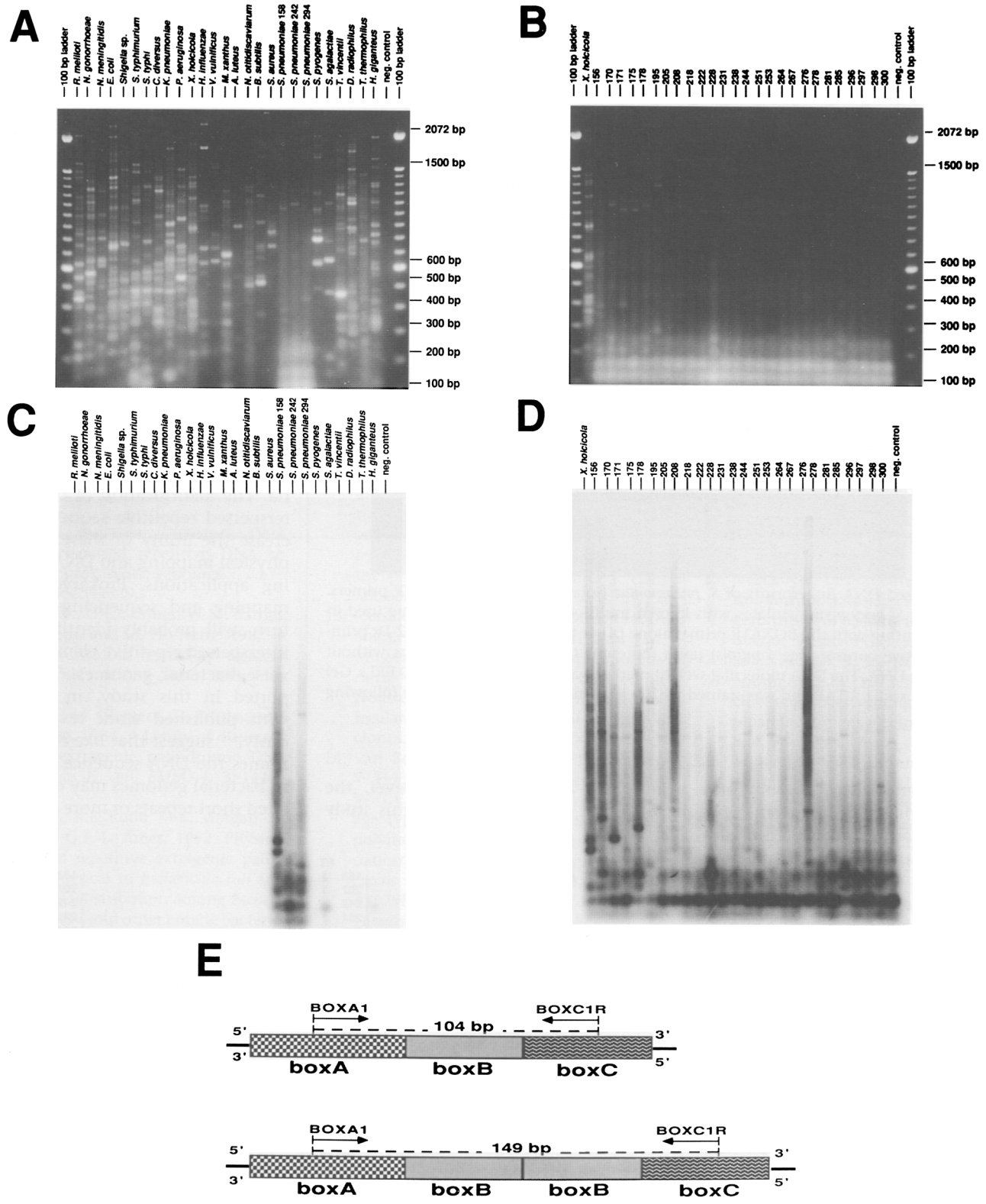

FIGURE 6 Presence of intact BOX elements in *S. pneumoniae* and *S. agalactiae*. Genomic DNAs from diverse bacteria (*A*) and pneumococcal isolates[13] (*B*) were used in PCR amplification experiments with the inwardly facing primers BOXA1 and BOXC1R. Amplicons were separated in 1.2% agarose and 1 × TAE and were stained with 0.5 μg/ml ethidium bromide after electrophoresis. PCR amplicons from *A* and *B* were transferred and subsequently hybridized with end-labeled BOXB1 as depicted in *C* and *D*, respectively. A schematic diagram in *E* depicts BOX elements with one or two boxB subunits and the expected sizes of PCR amplicons obtained with the BOXA1 and BOXC1R primers.

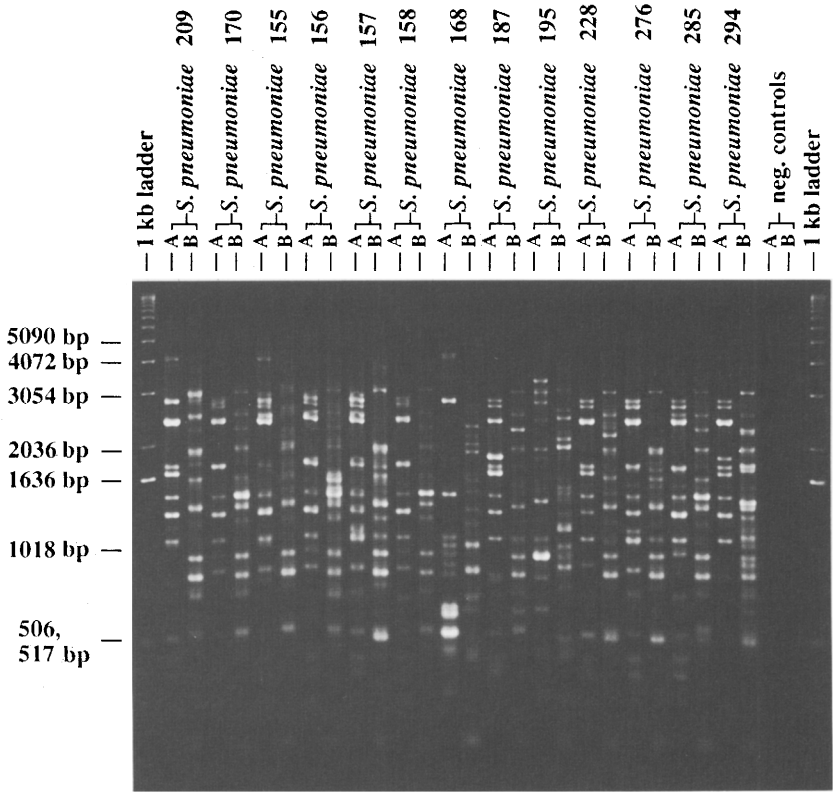

FIGURE 7 PCR-based DNA fingerprints of *S. pneumoniae* isolates with boxA and REP primers. Thirteen different *S. pneumoniae* isolates with known multilocus enzyme ETs[13] were used in rep–PCR fingerprinting with the BOXA1R primer alone (*A*) or the REP1R-Dt plus REP2-Dt primers (*B*). The negative control (*neg. controls*) lanes represent the same PCR reactions without template genomic DNA. The DNA molecular weight marker was a 1-kb ladder (GIBCO BRL). Gel was 1.5% agarose and 1 × TAE and was stained with 0.5 μg/ml of ethidium bromide following electrophoresis.

ences were observed with slot–blot and Southern hybridization patterns between *S. pneumoniae* and related organisms (Figs. 2–4 and 6). For example, the BOXB1 probe is 41% GC and hybridizes with *S. pneumoniae* (GC content = 38–39) but fails to hybridize with *S. aureus* (GC content = 32–36) and *S. pyogenes* (GC content = 34–39). Also, differences in GC content of the various probes represented another potentially confounding issue. For this reason, multiple oligonucleotide probes were designed that were complementary to each subunit and varied with respect to GC content. Figure 4 illustrated differences in evolutionary conservation obtained with probes with equivalent GC contents that are complementary to different BOX subunits.

The BOX element includes the boxA, boxB, and boxC subunits within a region of imperfect dyad symmetry.[7] This dyad structure may affect results obtained by DNA/DNA hybridization and PCR amplification. However, the consistent findings found in this study with respect to boxA subunit conservation and the lack of conservation of the boxB subunit with four different methodologies, slot–blot and Southern hybridizations and rep–PCR and data base queries, support our conclusions and probably represent deviations in the palindromic structure of BOX. The fact that alternative configurations of BOX with respect to subunit composition exist in the *S. pneumoniae* genome[7] confirm that the BOX palindrome can have significant structural alterations in the genome.

Interspersed repetitive sequences have been characterized from the genomes of different bacterial species.[1] Additionally, this study and Versalovic et al. [9] have shown that low to moderate stringency hybridization experiments reveal that sequences within these interspersed repetitive elements appear to be conserved in diverse bacteria. The identification of conserved interspersed repetitive sequences will increase the utility of these elements in physical mapping and DNA fingerprinting applications. Prokaryotic genome mapping and sequencing projects, in turn, will probably identify additional interspersed repetitive elements from diverse bacterial genomes. The data reported in this study, in conjunction with published work reviewed previously,[1] suggest that like eukaryotic genomes the DNA sequence arrangement in bacterial genomes may consist of isolated short repeats or more complex mo-

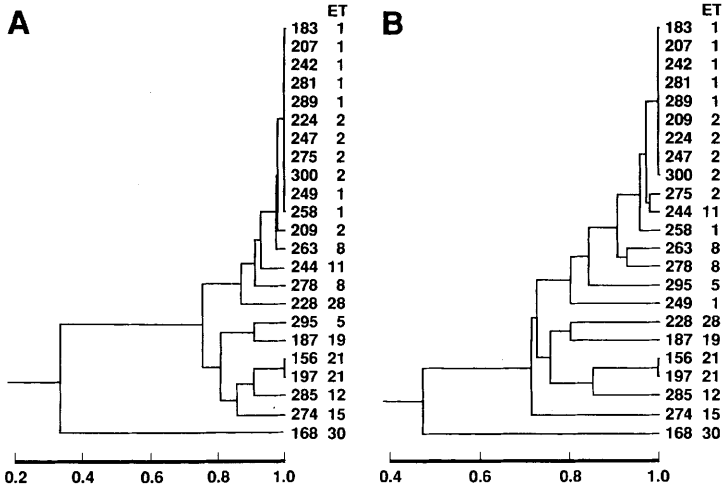

FIGURE 8 Dendrograms of bacterial isolates derived from fingerprints obtained with boxA and REP primers. Dendrograms were generated based on rep–PCR DNA fingerprints with the BOXA1R primer alone (*A*) or the REP1R-Dt plus REP2-Dt primers (*B*). Similarity levels are indicated by the axes below the dendrograms. Similarity of "1.0" indicates 100% identity between isolates. ET indicates the multilocus enzyme electrophoretotype obtained previously.[13]

saic repeats interspersed with longer single-copy sequences.

ACKNOWLEDGMENTS

We are grateful to the individuals who provided strains and genomic DNA. J.V. is supported by the National Institutes of Health Medical Scientist Training Program at Baylor College of Medicine and the NIH Minority Predoctoral Fellowship Program (MPFP) (1F31GM14601-01 SRC-7). J.R.L. acknowledges support from the Pew Scholars Program in Biomedical Sciences.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

1. Lupski, J.R. and G.M. Weinstock. 1992. Short, interspersed repetitive DNA sequences in prokaryotic genomes. *J. Bacteriol.* **174:** 4525–4529.
2. Higgins, C.F., G.F.L. Ames, W.M. Barnes, J.M. Clement, and M. Hofnung. 1982. A novel intercistronic regulatory element of prokaryotic operons. *Nature* **298:** 760–762.
3. Gilson, E., J.M. Clement, D. Brutlag, and M. Hofnung. 1984. A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*. *EMBO J.* **3:** 1417–1421.
4. Dimri, G.P., K.E. Rudd, M.K. Morgan, H. Bayat, and G.F.-L. Ames. 1992. Physical mapping of repetitive extragenic palindromic sequences in *Escherichia coli* and phylogenetic distribution among *Escherichia coli* strains and other enteric bacteria. *J. Bacteriol.* **174:** 4583–4593.
5. Sharples, G.J. and R.G. Lloyd. 1990. A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. *Nucleic Acids Res.* **18:** 6503–6508.
6. Hulton, C.S.J., C.F. Higgins, and P.M. Sharp. 1991. ERIC sequences: A novel family of repetitive elements in the genomes of *Escherichia coli, Salmonella typhimurium* and other enterobacteria. *Mol. Microbiol.* **5:** 825–834.
7. Martin, B., O. Humbert, M. Camara, E. Guenzi, J. Walker, T. Mitchell, P. Andrew, M. Prudhomme, G. Alloing, R. Hakenbeck, D.A. Morrison, G.J. Boulnois, and J.-P. Claverys. 1992. A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. *Nucleic Acids Res.* **20:** 3479–3483.
8. Subramanian, P.S., J. Versalovic, E.R.B. McCabe, and J.R. Lupski. 1992. Rapid mapping of *Escherichia coli*::Tn5 insertion mutations by REP-Tn5 PCR. *PCR Methods Applic.* **1:** 187–194.
9. Versalovic, J., T. Koeuth, and J.R. Lupski. 1991. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Res.* **19:** 6823–6831.
10. de Bruijn, F.J. 1992. Use of repetitive (repetitive extragenic element and enterobacterial repetitive intergenic consensus) sequences and the polymerase chain reaction to fingerprint the genomes of *Rhizobium meliloti* isolates and other soil bacteria. *Appl. Environ. Microbiol.* **58:** 2180–2187.
11. Versalovic, J., T. Koeuth, Y.-H. Zhang, E.R.B. McCabe, and J.R. Lupski. 1992. Quality control for bacterial inhibition assays: DNA fingerprinting of microorganisms by rep-PCR. *Screening* **1:** 175–183.
12. Woods, C.R., J. Versalovic, T. Koeuth, and J.R. Lupski. 1992. Analysis of relationships among isolates of *Citrobacter diversus* using DNA fingerprints generated by repetitive sequence-based primers in the polymerase chain reaction. *J. Clin. Microbiol.* **30:** 2921–2929.
13. Versalovic, J., V. Kapur, E.O. Mason Jr., U. Shah, T. Koeuth, J.R. Lupski, and J.M. Musser. 1993. Penicillin resistant *Streptococcus pneumoniae* strains recovered in Houston, Texas: Identification and molecular characterization of multiple clones. *J. Infect. Dis.* **167:** 850–856.
14. van Belkum, A., R. Bax, P. Peerbooms, W.H.F. Goessens, N. van Leeuwen, and W.G.V. Quint. 1993. Comparison of phage typing and DNA fingerprinting by polymerase chain reaction for discrimination of methicillin-resistant *Staphylococcus aureus* strains. *J. Clin. Microbiol.* **31:** 798–803.
15. Giesendorf, B.A.J., A. van Belkum, A. Koeken, H. Stegeman, M.H.C. Henkens, J. Van der Plas, H. Goossens, H.G.M. Niesters, and W.G.V. Quint. 1993. Development of species-specific DNA probes for *Campylobacter jejuni, Campylobacter coli*, and *Campylobacter lari* by polymerase chain reaction fingerprinting. *J. Clin. Microbiol.* **31:** 1541–1546.
16. Judd, A.K., M. Schneider, M.J. Sadowsky, and F.J. de Bruijn. 1993. Use of repetitive sequences and the polymerase chain reaction technique to classify genetically related *Bradyrhizobium japonicum* serocluster 123 strains. *Appl. Environ. Microbiol.* **59:** 1702–1708.
17. Struelens, M.J., R. Bax, A. Deplano, W.G.V. Quint, and A. van Belkum. 1993. Concordant clonal delineation of methicillin-resistant *Staphylococcus aureus* by macrorestriction analysis and polymerase chain reaction genome fingerprinting. *J. Clin. Microbiol.* **31:** 1964–1970.
18. van Belkum, A., M. Struelens, and W.G.V. Quint. 1993. Typing of *Legionella pneumophila* strains by polymerase chain reaction-mediated DNA fingerprinting. *J. Clin. Microbiol.* **31:** 2198–2200.
19. Woods, C., J. Versalovic, T. Koeuth, and J.R. Lupski. 1993. Whole cell rep–PCR allows rapid assessment of clonal relationships of bacterial isolates. *J. Clin. Microbiol.* **31:** 1927–1931.
20. Georghiou, P.R., A.M. Doggett, M.A. Kielhofner, J.E. Stout, D.A. Watson, J.R. Lupski, and R.J. Hamill. 1994. Molecular fingerprinting of *Legionella* species by repetitive element PCR. *J. Clin. Microbiol.* **32:** 2989–2994.
21. Georghiou, P., R.J. Hamill, C.E. Wright, J. Versalovic, T. Koeuth, D.A. Watson, and J.R. Lupski. 1995. Molecular epidemiology of infections due to *Enterobacter aerogenes*: Identification of hospital outbreak strains by molecular techniques. *Clin. Infect. Dis.* **20:** 84–94.
22. van Belkum, A. 1994. DNA fingerprinting of medically important microorganisms by use of PCR. *Clin. Microbiol. Rev.* **7:** 174–184.
23. Versalovic, J., M. Schnieder, F.J. de Bruijn, and J.R. Lupski. 1994. Genomic fingerprinting of bacteria using repetitive sequence based PCR (rep-PCR). *Methods Mol. Cell Biol.* **5:** 25–40.
24. Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98:** 503–517.
25. Miller, J.H. 1972. *Experiments in molecular genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
26. Woese, C.R. 1987. Bacterial evolution. *Microbiol. Rev.* **51:** 221–271.
27. Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
28. Noda, A., J.B. Courtwright, P.F. Denor, G. Webb, Y. Kohara, and A. Ishihama. 1991. Rapid identification of specific genes in *E. coli* by hybridization to membranes containing the ordered set of phage clones. *BioTechniques* **10:** 474–477.
29. van Leuven, F. 1991. The trouble with PCR machines: Fill up the empty spaces. *Trends Genet.* **7:** 142.
30. Sokal, R.R. and P.H.A. Sneath. 1963. *Principles of numerical taxonomy*. W.H. Freeman and Company, San Francisco, CA.
31. Pearson, W.R. and D.J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci.* **85:** 2444–2448.
32. Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12:** 387–395.

33. Gilson, E., W. Saurin, D. Perrin, S. Bachellier, and M. Hofnung. 1991. Palindromic units are part of a new bacterial interspersed mosaic element (BIME). *Nucleic Acids Res.* **19:** 1375–1383.
34. Gilson, E., W. Saurin, D. Perrin, S. Bachellier, and M. Hofnung. 1991. The BIME family of bacterial highly repetitive sequences. *Res. Microbiol.* **142:** 217–222.
35. Bachellier, S., D. Perrin, M. Hofnung, and E. Gilson. 1993. Bacterial interspersed mosaic elements (BIMEs) are present in the genome of *Klebsiella*. *Mol. Microbiol.* **7:** 537–544.
36. Bachellier, S., W. Saurin, D. Perrin, M. Hofnung, and E. Gilson. 1994. Structural and functional diversity among bacterial interspersed mosaic elements (BIMEs). *Mol. Microbiol.* **12:** 61–70.

Received September 14, 1995; accepted in revised form October 13, 1995.