JOURNAL OF BACTERIOLOGY, Aug. 1990, p. 4420–4426
0021-9193/90/084420-07$02.00/0

Vol. 172, No. 8

# Multiple Copies of a Bile Acid-Inducible Gene in *Eubacterium* sp. Strain VPI 12708

RASHMI GOPAL-SRIVASTAVA,† DARRELL H. MALLONEE, W. BRUCE WHITE,‡ AND PHILLIP B. HYLEMON*

*Department of Microbiology and Immunology, Medical College of Virginia, Virginia Commonwealth University, Richmond, Virginia 23298-0678*

Received 20 February 1990/Accepted 13 May 1990

***Eubacterium* sp. strain VPI 12708 is an anaerobic intestinal bacterium which possesses inducible bile acid 7-dehydroxylation activity. Several new polypeptides are produced in this strain following induction with cholic acid. Genes coding for two copies of a bile acid-inducible 27,000-dalton polypeptide (*baiA1* and *baiA2*) have been previously cloned and sequenced. We now report on a gene coding for a third copy of this 27,000-dalton polypeptide (*baiA3*). The *baiA3* gene has been cloned in lambda DASH on an 11.2-kilobase DNA fragment from a partial *Sau*3A digest of the *Eubacterium* DNA. DNA sequence analysis of the *baiA3* gene revealed 100% homology with the *baiA1* gene within the coding region of the 27,000-dalton polypeptides. The *baiA2* gene shares 81% sequence identity with the other two genes at the nucleotide level. The flanking nucleotide sequences associated with the *baiA1* and *baiA3* genes are identical for 930 bases in the 5′ direction from the initiation codon and for at least 325 bases in the 3′ direction from the stop codon, including the putative promoter regions for the genes. An additional open reading frame (occupying from 621 to 648 bases, depending on the correct start codon) was found in the identical 5′ regions associated with the *baiA1* and *baiA3* clones. The 5′ sequence 930 bases upstream from the *baiA1* and *baiA3* genes was totally divergent. The *baiA2* gene, which is part of a large bile acid-inducible operon, showed no homology with the other two genes either in the 5′ or 3′ direction from the polypeptide coding region, except for a 15-base-pair presumed ribosome-binding site in the 5′ region. These studies strongly suggest that a gene duplication (*baiA1* and *baiA3*) has occurred and is stably maintained in this bacterium.**

One of the quantitatively most important bile acid biotransformations of cholic and chenodeoxycholic acids is 7-dehydroxylation, yielding deoxycholic and lithocholic acid, respectively. This biotransformation is carried out by anaerobic intestinal bacteria (27, 28). Deoxycholic and lithocholic acids differ markedly from their 7-hydroxylated precursors in physicochemical properties as well as in physiological effects (3, 6, 30, 33, 36, 37, 41). Deoxycholic acid makes up approximately 20 to 25% of the total bile acid pool in humans (45). Intestinal bacteria capable of 7-dehydroxylation of bile acids have been isolated by several laboratories (15, 28). Most intestinal bacteria possessing 7-dehydroxylation activity have been identified as members of the *Clostridium* (21, 23, 47) or *Eubacterium* (20, 23) genus. It has been demonstrated that the fecal population of 7-dehydroxylating intestinal bacteria in humans and in rats is in the range of $10^3$ to $10^5$ cultivable organisms per g (wet weight) of feces.

*Eubacterium* sp. strain VPI 12708 is an anaerobic intestinal bacterium that possesses bile acid 7α-dehydroxylation activity which is induced by culturing the bacterium in the presence of C-24 bile acids containing a 7α-hydroxyl group. At least four new polypeptides with estimated molecular weights of 77,000, 45,000, 27,000, and 23,000 have been shown to be synthesized following cholic acid induction (38). Immunoinhibition and immunoprecipitation studies have indicated involvement of the 27,000-dalton polypeptide in bile acid 7-dehydroxylation (38). The genes coding for two bile acid-inducible polypeptides, a 27,000-dalton polypeptide (27K-1) and a 45,000-dalton polypeptide (45K), have been cloned on separate DNA fragments (7, 52). Northern (RNA) blot analysis of *Eubacterium* RNA has indicated that the genes coding for the 27K-1 and 45K polypeptides reside on separate bile acid-inducible transcripts with relative sizes of 950 base and 6 to 8 kilobases, respectively (7, 52).

Nucleotide sequence analysis in the 5′ direction from the gene encoding the 45K polypeptide revealed the presence of a gene which encoded a second 27,000-dalton polypeptide (27K-2). The two genes encoding the 27,000-dalton polypeptides exhibited extensive homology (53). Southern hybridization analysis of *Eco*RI-cut chromosomal DNA probed with a 23-mer oligonucleotide consisting of a common sequence found within the coding regions for both of the 27K genes revealed the presence of three equally intense bands of 2.2, 2.9 (previously estimated to be 2.6), and 3.5 kilobases (53). Hybridization of the 2.2- and 2.9-kilobase fragments is consistent with the restriction maps of the genes encoding the 27K-1 and 27K-2 polypeptides and flanking regions. The 3.5-kilobase *Eco*RI fragment indicated the possible presence of an additional gene coding for a third 27,000-dalton polypeptide (27K-3). In the present study, a DNA fragment coding for a third 27,000-dalton polypeptide was cloned and characterized. We propose to label the members of this gene family and other related genes as *bai* (bile acid-inducible) genes and will label the genes encoding the 27,000-dalton

* Corresponding author.
† Present address: Laboratory of Molecular and Developmental Biology, National Eye Institute, National Institutes of Health, Bethesda, MD 20892.
‡ Present address: Anaerobe Laboratory, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061.

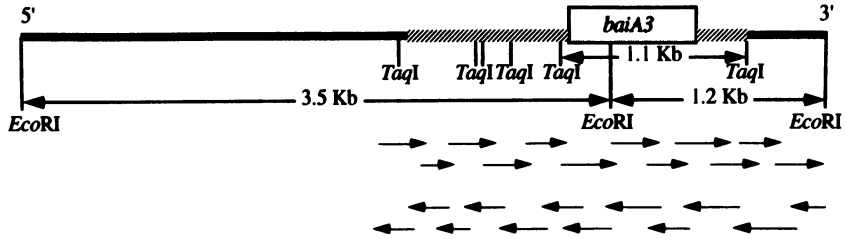



FIG. 1. Restriction map and cloning strategy for the *baiA3* gene and surrounding region. The region encoding the *baiA3* gene is boxed and the region of homology with the *baiA1* clone ( ▨ ) is indicated. Kb, Kilobases.

polypeptides as *baiA1* (encoding the 27K-1 polypeptide), *baiA2* (27K-2), and *baiA3* (27K-3).

## MATERIALS AND METHODS

**Materials.** T4 DNA ligase and T4 polynucleotide kinase were obtained from Bethesda Research Laboratories, Gaithersburg, Md. Restriction enzymes were obtained from Pharmacia LKB Biotechnology, Inc., Piscataway, N.J., and Bethesda Research Laboratories. Dideoxynucleotide sequencing reagents and modified T7 polymerase were obtained from United States Biochemical Corp., Cleveland, Ohio. Radionucleotides [$\gamma$-$^{32}$P]ATP and [$\alpha$-$^{35}$S]dATP were obtained from Du Pont, NEN Research Products, Boston, Mass., as were the GeneScreen hybridization membranes. The lambda gt11 and lambda DASH vectors were obtained from Stratagene, La Jolla, Calif.

**Bacterial strains and culture conditions.** *Eubacterium* sp. strain VPI 12708 stock cultures were maintained in chopped-meat medium, as described by Holdeman and Moore (24). Bacteria for DNA isolation were grown under anaerobic conditions, as previously described (51). *Escherichia coli* strains LE392, Y1090 (58), and JM101 (56) were used as host strains for bacteriophage lambda DASH, lambda gt11, and M13mp18, respectively.

**Nucleic acid isolation.** *Eubacterium* sp. strain VPI 12708 chromosomal DNA was isolated by the method of Marmur (34). Plasmid DNA was isolated by the method of Birnboim and Doly (5). Single-stranded M13 DNA was prepared for sequencing by the procedure described by Davis et al. (12).

**DNA hybridization procedures.** DNA restriction fragments from agarose gels were transferred to nitrocellulose as described by Southern (46). Lambda gt11 and lambda DASH plaques were transferred onto nitrocellulose filters as described by Benton and Davis (4). Baked filters were prehybridized, hybridized, and washed as described by Woods (55). The hybridized filters were placed in a cassette with Kodak X-RP film and a Du Pont Cronex Lightning-Plus intensifying screen for 24 to 48 h at −70°C before the film was developed. Oligonucleotides were synthesized with a Cyclone DNA synthesizer (Biosearch, Inc., San Rafael, Calif.) and purified as previously described (7). Purified oligonucleotides were end labeled for Southern blots with [$\gamma$-$^{32}$P]ATP (3,000 Ci/mmol). Unincorporated label was removed with Nensorb 20 cartridges (Du Pont, NEN Research Products).

**Recombinant DNA methods.** For preparation of the lambda gt11 library, *Eubacterium* sp. strain VPI 12708 genomic DNA was digested to completion with *Eco*RI and size fractionated on 0.8% agarose. Fragments of approximately 3 to 4 kilobases were isolated by electroelution onto NA-45 DEAE membranes (Schleicher & Schuell, Inc., Keene, N.H.) (32) and ligated to *Eco*RI-digested lambda gt11 arms.

For construction of the lambda DASH library, *Eubacterium* DNA was purified by the method of Marmur (34), with modifications. The sodium perchlorate step was omitted and a hexadecyltrimethyl ammonium bromide extraction (2) was performed following the RNase treatment of the DNA. The *Eubacterium* DNA was partially digested with *Sau*3A and size fractionated with a linear, 10- to-40% sucrose gradient. Fractions (200 μl) were collected from the bottom of the tube and stored at 4°C until sucrose and salt removal with a Centricon 10 microconcentrator (Amicon Corp., Danvers, Mass.). The molecular sizes of the collected DNA fractions were determined by electrophoresing samples from the gradient on a 0.6% agarose gel, and the fractions with the desired DNA size range were pooled and used as inserts for *Bam*HI-digested lambda DASH arms.

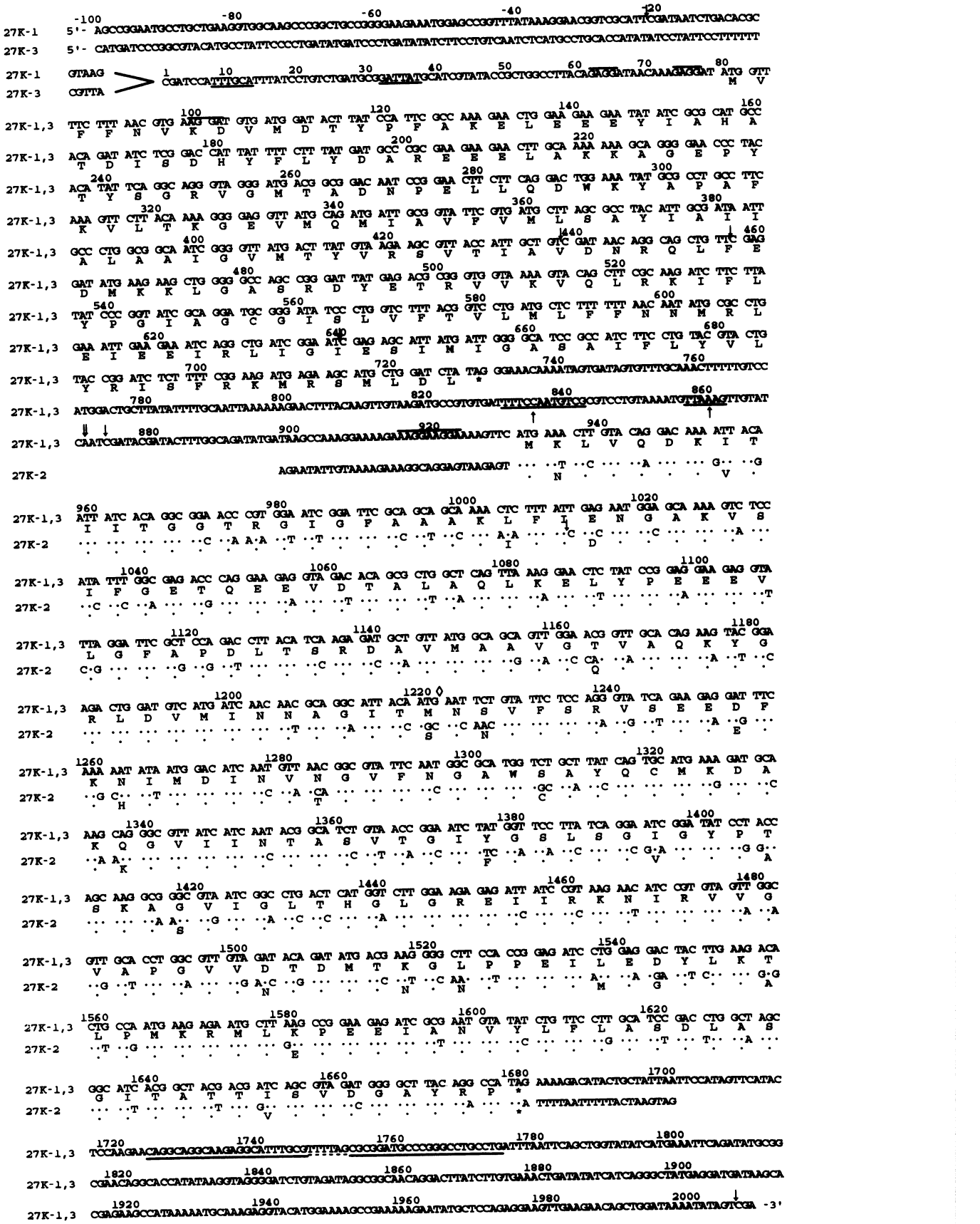
**DNA sequencing.** DNA sequences were obtained by the dideoxy sequencing method (42) and the Sequenase procedure of United States Biochemical Corp. Commercially available universal sequencing primers (17-mer) and other synthetic oligonucleotides were used in the sequencing procedures. Both DNA strands in all reported regions were sequenced. Regions of ambiguous DNA sequence were further analyzed by use of dITP in the sequencing reactions. DNA was labeled with [$\alpha$-$^{35}$S]dATP.

**Analysis of sequence data.** Analysis of nucleic acid and protein sequence data was performed with the Microgenie sequence analysis program (Beckman Instruments, Inc., Fullerton, Calif.), the IBI/Pustell DNA analysis program (International Biotechnologies, Inc., New Haven, Conn.), and the GCG program (University of Wisconsin Biotechnology Center, Madison, Wis.).

**Nucleotide sequence accession number.** The nucleotide sequence data presented in this paper have been submitted to GenBank (accession no. M34658).

## RESULTS

**Cloning of the *baiA3* gene.** Initial cloning experiments were done with *Eco*RI digests of chromosomal DNA from *Eubacterium* sp. strain VPI 12708 cloned into lambda gt11. Of approximately 3,000 plaques containing inserts (clear plaques), 20 positive plaques were identified following screening with a $^{32}$P-labeled 23-mer probe (bases 1183 to 1205 [see Fig. 2]) that hybridizes to all three copies of the *baiA* gene family (53). DNA from phage obtained from positive plaques was digested with *Eco*RI, fractionated on a 0.8% agarose gel, and Southern blotted with the 23-mer probe. Two of the positive plaques were found to contain a 2.9-kilobase insert which has been previously cloned and shown to contain the *baiA2* gene (53). The remaining recombinant phage contained a 3.5-kilobase insert which strongly hybridized to the 23-mer probe. One of the phage with a 3.5-kilobase insert was selected for further analysis. Diges-



tion of this recombinant phage DNA with *Eco*RI was followed by agarose gel electrophoresis and Southern blot hybridization with probes MCV-1 and UAB-6, corresponding to nucleotides 940 to 956 and 1574 to 1589, respectively, of the *baiA1* gene (see Fig. 2). Although strong hybridization was observed with MCV-1, there was no hybridization with probe UAB-6. This indicated that DNA coding for the N-terminal end of the putative 27K-3 polypeptide was on the 3.5-kilobase *Eco*RI fragment and that the part of the gene coding for the C-terminal end of the polypeptide was not present on this fragment.

In order to obtain the complete *baiA3* gene, a lambda DASH library containing fragments from a partial *Sau*3A digest of *Eubacterium* sp. strain VPI 12708 DNA was screened with the 23-mer probe. A positive plaque containing an 11.2-kilobase insert was identified. DNA was isolated from this clone and digested with *Eco*RI. Agarose gel electrophoresis and Southern blotting indicated that the recombinant phage contained the 3.5-kilobase *Eco*RI fragment which hybridized to the 23-mer probe and probe MCV-1 (corresponding to the N-terminal region of the 27K-3 polypeptide). Probe UAB-6 hybridized to a 1.2-kilobase *Eco*RI fragment, indicating that the coding region for the C-terminal end of the 27K-3 polypeptide was on this fragment.

**DNA sequence analysis.** The 1.2- and 3.5-kilobase *Eco*RI fragments were subcloned into M13mp18 in both orientations for sequencing. DNA sequence analysis of these clones was performed with both universal primers and other synthetic oligonucleotide primers directed against sequences contained within the 1.2- and 3.5-kilobase *Eco*RI fragments. The sequencing strategy and a restriction map for these fragments are shown in Fig. 1.

The nucleotide sequences of the 1.2- and 3.5-kilobase *Eco*RI fragments indicated that there was an open reading frame which appeared to encode a polypeptide with a calculated molecular weight of 26,745. This open reading frame was designated the *baiA3* gene. The complete DNA sequence is shown in Fig. 2 along with the deduced amino acid sequence. The coding region for the *baiA3* gene includes nucleotides 931 through 1677.

Alignment of the sequence for the *baiA3* gene with the previously reported sequence for the *baiA1* gene (8) indicated that there was 100% nucleotide sequence identity within the 747-base coding regions of these two genes (Fig. 2). Sequences in the 5′ and 3′ directions from the *baiA3* gene, including the putative promoter region and a large 3′ element with dyad symmetry, are also identical to the *baiA1* clone. The *baiA1* and *baiA3* clones were both sequenced in the 5′ direction until a point of divergence was discovered. The homology between the two clones remains for 930 bases in the 5′ direction from the *baiA* genes and for at least 325 bases in the 3′ direction. The nucleotide sequences of the *baiA1* and *baiA3* clones are divergent in the 5′ direction after the point indicated by nucleotide 1 in Fig. 2. An open reading frame of from 621 to 648 bases, depending on the correct initiation codon, was found in the identical 5′ region upstream from the *baiA1* and *baiA3* genes. A putative promoter region for this open reading frame is located immediately to the 3′ side of the homology junction for the two cloned fragments.

The *baiA2* gene shows differences at 143 nucleotides within the coding region, for an overall 81% homology with the *baiA1* and *baiA3* genes. Significant homology between the *baiA2* and *baiA3* genes was not observed either in the 5′ direction from the initiation codon or in the 3′ direction from the stop codon, except for a 15-base-pair presumed ribosome binding site in the 5′ region. Comparative restriction maps for the three *baiA* genes and surrounding regions are shown in Fig. 3.

Data bank searches for sequences similar to the open reading frame 5′ from the *baiA1* and *baiA3* genes revealed no significantly homologous sequences. We have previously reported that the 27K polypeptides share significant sequence homology with several alcohol-polyol dehydrogenases (8, 53).

**Southern blot hybridization analysis.** In order to determine if bile acids were involved in regulating the duplication of the *baiA1* and *baiA3* genes, the following experiment was performed. Cultures of *Eubacterium* sp. strain VPI 12708 were repeatedly cultured in the presence or absence of cholic acid. Then, DNA was isolated from the cells and Southern blot analysis was carried out by using a DNA probe specific for the *baiA* genes. The results in Fig. 4 show three strongly hybridizing bands from *Eco*RI-cut DNA isolated from bacteria cultured in the presence or absence of cholic acid. Finally, Southern blot analysis of DNA digested with various restriction endonucleases was inconclusive as to whether the *baiA1* and *baiA3* genes were located in a tandem or nontandem orientation in the chromosome (data not shown).

## DISCUSSION

We have previously reported the cloning and sequencing of genes coding for two similar copies of a 27,000-dalton bile acid-inducible polypeptide (*baiA1* and *baiA2*) involved in 7-dehydroxylation of bile acids in *Eubacterium* sp. strain VPI 12708 (7, 8, 53). In this study, we have identified an additional gene (*baiA3*) in this multigene family.

The three members of the *baiA* gene family are highly conserved at the nucleotide and amino acid levels. Nucleotide sequence identities within the open reading frames are 100% between the *baiA1* and *baiA3* genes and 81% between the *baiA1* and *baiA2* genes. The amino acid sequences for the 27K-1 and 27K-3 polypeptides are identical, whereas the 27K-1 and 27K-2 polypeptides show 92% sequence identity at the amino acid level. The *baiA1* and *baiA3* genes share a block of homology that spans at least 2 kilobases of DNA, including 930 bases in the 5′ direction and at least 325 bases in the 3′ direction from the open reading frames for the *baiA* genes. A second open reading frame, which could code for at

FIG. 2. DNA sequences for the *baiA1* (27K-1), *baiA2* (27K-2), and *baiA3* (27K-3) genes and surrounding regions. Nucleotide 1 represents the start of the region of homology between the *baiA1* and *baiA3* clones. Deduced amino acid sequences are provided for the open reading frame in the 5′ region of homology for the *baiA1* and *baiA3* clones (nucleotides 80 to 730), and for the *baiA* genes (nucleotides 931 to 1677). Possible ribosome-binding sites are overlined (three possible sites for the first open reading frame). The putative −10 and −35 promoter regions are underlined. A 12-nucleotide region containing several potential −35 sites is underlined for the *baiA1* and *baiA3* genes (nucleotides 831 to 842 [8]). The mRNA start site for the *baiA1* and *baiA3* genes, as determined by primer extension (8) ( ⇓ ), and the nucleotides present at 10 and 35 bases upstream from the mRNA start site ( ↑ ) are indicated. Restriction sites for *Taq*I ( ↓ ) are indicated (after nucleotides −22, 438, 456, 639, 1010, and 2005). The *Eco*RI restriction site in the *baiA1* and *baiA3* genes (◇) is indicated after nucleotide 1221. A large potential stem-loop structure is underlined from nucleotides 1724 to 1774.

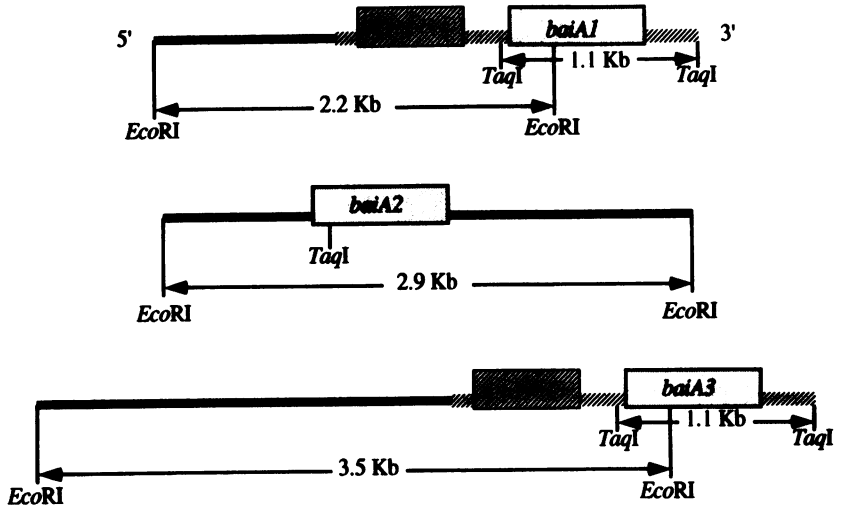



FIG. 3. Relative restriction maps for the *baiA1*, *baiA2*, and *baiA3* genes and surrounding regions. Not all *Taq*I sites are identified. The *baiA* genes are boxed, and the region of homology between the *baiA1* and *baiA3* clones (▨) is highlighted. The open reading frames in the homologous regions upstream from the *baiA1* and *baiA3* clones (▨) are also boxed. Kb, Kilobases.

least 207 amino acids, was found in the identical 5′ flanking sequences of the *baiA1* and *baiA3* clones. No discernable open reading frames were detected in the identical 3′ flanking regions of the *baiA1* and *baiA3* clones. The nucleotide sequences for the *baiA1* and *baiA3* clones were totally divergent after the 930-base 5′ block of homology. A comparison of the *baiA1* and *baiA3* clones after the 325-base block of homology in the 3′ direction could not be made because this region could not be obtained for the *baiA1* clone. Repeated attempts to obtain such a clone were unsuccessful.

Although the exact function of the 27K polypeptides in 7-dehydroxylation of bile acids is not known, protein sequence surveys indicate that these polypeptides share strong homology with the nonzinc alcohol-polyol dehydrogenase class of enzymes, especially in the proposed pyridine nucleotide-binding domains (8, 29, 53). It has therefore been suggested that the 27K polypeptides are involved in one of the pyridine nucleotide-dependent steroid oxidation-reduction reactions in the 7-dehydroxylation pathway in *Eubacterium* sp. strain VPI 12708 (8, 53).

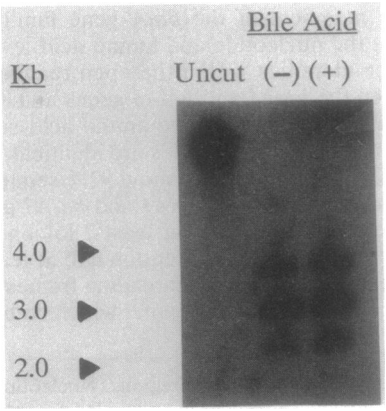



FIG. 4. Southern blot analysis of *Eubacterium* sp. strain VPI 12708 chromosomal DNA isolated from bacteria grown in the presence (+) and absence (−) of sodium cholate. DNA was digested with *Eco*RI endonuclease and probed with a $^{32}$P-labeled oligonucleotide (23-mer) consisting of a common sequence found in the three *baiA* genes (nucleotides 1183 to 1205 [Fig. 2]). Abbreviation: Kb, kilobases.

Although unusual, there are several examples of stabilized gene duplication in bacteria. In the cyanobacteria, several genes have been reported to occur as multiple copies, including the *psbA* (photosystem II) gene from *Anacystis nidulans* (three copies) (19) and a gas-vesicle protein gene and a phycobilisome protein gene from a *Calothrix* strain (two copies each) (11, 25). Other examples include highly repetitive tRNA genes in *Photobacterium phosphoreum* (18), homologous acid-soluble spore proteins from *Bacillus subtilis* and *Bacillus megaterium* (9, 16), duplicated *tuf* genes in numerous bacterial species (14, 44), and homologous *nif* genes (40, 49); there is also the possibility of numerous reiterated sequences in *Rhizobium* and *Agrobacterium* spp. (17) and the possibility of multiple chromosomes in *Azotobacter vinelandii* (39). The presence of isozymes which exhibit extensive homology has also been reported in such cases as pectic enzymes in *Erwinia chrysanthemi* (43), alkaline phosphatase enzymes in *Bacillus licheniformis* (26) and acetohydroxy acid synthase enzymes from *E. coli* (50).

The reason for having three copies of the *baiA* gene is unclear. The need for a large amount of an enzyme involved in 7-dehydroxylation would appear to be unlikely but cannot be discounted. Duplication of the regions either upstream or downstream from the *baiA1* and *baiA3* genes could also be potentially useful to the *Eubacterium* strain. Culturing bacteria in the presence or absence of cholic acid does not appear to alter or select for the number of copies of the *baiA* genes under laboratory conditions. Another possibility is that the 27K-2 polypeptide may have a slightly different activity than the 27K-1 and 27K-3 polypeptides. Recent evidence has indicated the formation of isomeric forms of bile acids (allo bile acids) during 7α-dehydroxylation of cholic acid in *Eubacterium* sp. strain VPI 12708 (unpublished data). Perhaps the 27K polypeptides, acting as steroid oxidoreductases, recognize and/or catalyze the formation of different isomers in this pathway. A third possibility could involve differences in transcriptional or translational control of these bile acid-induced polypeptides. In this regard, both the *baiA1* and *baiA3* genes (8) and the *baiA2* gene (52) have

been shown by Northern blot analysis to be induced by cholic acid. The *B. licheniformis* alkaline phosphatase genes were reported to be transcribed from different RNA polymerases (26), and the *E. coli* acetohydroxy acid synthases have been shown to be under different transcriptional and translational control mechanisms (10, 50). Similar differences could exist between the *baiA* genes. The putative promoter region for the operon associated with the *baiA2* gene is currently being studied and compared with the two identical promoter regions associated with the *baiA1* and *baiA3* genes.

The presence of two identical stretches of DNA, encompassing the *baiA1* and *baiA3* genes and spanning at least 2 kilobases of DNA, poses additional questions. Why and how would the *Eubacterium* strain sustain these two identical copies? Possibilities include the presence of insertion sequence elements on the ends of both copies, which could allow the segments to integrate into multiple sites on the chromosome. Another possibility is that the *baiA1* and *baiA3* genes could be present on an integrated bacteriophage. An examination of both ends of the homologous blocks of DNA associated with the *baiA1* and *baiA3* clones could help in testing these possibilities. It is unclear how large these two identical stretches of DNA actually are, and the 3′ junctions of these two copies have not been determined. It is also unclear whether the two identical stretches of DNA are present as tandem repeats or are positioned on different parts of the chromosome. Southern blot analysis of chromosomal digests from *Eubacterium* sp. strain VPI 12708 have been inconclusive in determining the sizes and relative locations of these copies. However, the *baiA3* gene has been detected in approximately 19 lambda DASH clones, some containing 12- to 15-kilobase inserts. None of these phages contained the *baiA1* gene, suggesting either a very large duplicated region or a nontandem arrangement (data not presented). Further studies are necessary to investigate these possibilities.

A third possible explanation for the two identical copies could be random chromosomal crossover events, which could produce a duplicated region. Random genetic duplication is not uncommon in bacteria. One report suggests that 3% of *Salmonella typhimurium* isolates have duplicated regions of DNA (1). Many genetic duplications are thought to be associated with unequal recombination events involving highly repetitive chromosomal elements such as rRNA genes (1, 22) or *rhs* genes (13, 31). However, such duplications are generally not stably maintained unless they provide the microorganism with a selective advantage. There have been numerous reports on the integration of duplicated DNA segments into the *B. subtilis* chromosome and on the apparent stability of these duplicated regions (35, 48, 54, 57). However, the possibility of a random duplication maintaining a completely identical nucleotide sequence over a long period of time would be remote, and the relative intensities of hybridizing bands from Southern blots performed on *Eubacterium* sp. strain VPI 12708 DNA suggest that the entire population of cells possesses the duplicated regions.

## ACKNOWLEDGMENTS

We thank Mark Lijewski for his help in screening the lambda DASH library and Pam Melone for help with Fig. 4.

This work was supported by Public Health Service research grants DK40986 and P01DK38030 from the National Institutes of Health. D.H.M. was supported by Public Health Service postdoctoral training grant T32-CA-0956404 from the National Institutes of Health.

## LITERATURE CITED

1. **Anderson, R. P., and J. R. Roth.** 1981. Spontaneous tandem genetic duplications in *Salmonella typhimurium* arise by unequal recombination between rRNA (*rrn*) cistrons. Proc. Natl. Acad. Sci. USA **78**:3113–3117.
2. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1987. Current protocols in molecular biology. John Wiley & Sons, Inc., New York.
3. **Bagheri, S. A., M. G. Bolt, J. L. Boyers, and R. H. Palmer.** 1978. Stimulation of thymidine incorporation in mouse liver and biliary tract epithelium by lithocholate and deoxycholate. Gastroenterology **74**:188–192.
4. **Benton, W. D., and R. W. Davis.** 1977. Screening lambda gt recombinant clones by hybridization to single plaque in situ. Science **196**:180–182.
5. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
6. **Carey, M. C.** 1985. Physical-chemical properties of bile acids and their salts, p. 345–405. *In* H. Danielsson and J. Sjovall (ed.), Sterols and bile acids: new comprehensive biochemistry, vol. 12. Elsevier Science Publishing, Inc., New York.
7. **Coleman, J. P., W. B. White, and P. B. Hylemon.** 1987. Molecular cloning of bile acid 7-dehydroxylase from *Eubacterium* sp. strain VPI 12708. J. Bacteriol. **169**:1516–1521.
8. **Coleman, J. P., W. B. White, M. Lijewski, and P. B. Hylemon.** 1988. Nucleotide sequence and regulation of a gene involved in bile acid 7-dehydroxylation in *Eubacterium* sp. strain VPI 12708. J. Bacteriol. **170**:2070–2077.
9. **Conners, M. J., S. Howard, J. Hoch, and P. Setlow.** 1986. Determination of the chromosomal locations of four *Bacillus subtilis* genes which code for a family of small, acid-soluble spore proteins. J. Bacteriol. **166**:412–416.
10. **Dailey, F. E., and J. E. Cronan, Jr.** 1986. Acetohydroxy acid synthase I, a required enzyme for isoleucine and valine biosynthesis in *Escherichia coli* K-12 during growth on acetate as the sole carbon source. J. Bacteriol. **165**:453–460.
11. **Damerval, T., J. Houmard, G. Guglielmi, K. Csiszar, and N. Tandeau de Marsac.** 1987. Developmentally regulated *gvpABC* operon is involved in the formation of gas vesicles in the cyanobacterium *Calothrix* 7601. Gene **54**:83–92.
12. **Davis, L. G., M. D. Dibner, and J. F. Battey.** 1986. Basic methods in molecular biology. Elsevier Science Publishing, Inc., New York.
13. **Feulner, G., J. A. Gray, J. A. Kirschman, A. F. Lehner, A. B. Sadosky, D. A. Vlazny, J. Zhang, S. Zhao, and C. W. Hill.** 1990. Structure of the *rhsA* locus from *Escherichia coli* K-12 and comparison of *rhsA* with other members of the *rhs* multigene family. J. Bacteriol. **172**:446–456.
14. **Filer, D., R. Dhar, and A. V. Furano.** 1981. The conservation of DNA sequences over very long periods of evolutionary time. Evidence against intergeneric chromosomal transfer as an explanation for the presence of *Escherichia coli tuf* gene sequences in taxonomically-unrelated prokaryotes. Eur. J. Biochem. **120**:69–77.
15. **Finegold, S. M., D. J. Flora, H. R. Atteberry, and V. L. Sutter.** 1975. Fecal bacteriology of colonic polyp patients and control patients. Cancer Res. **35**:3407–3417.
16. **Fliss, E. R., C. A. Loshon, and P. Setlow.** 1986. Genes for *Bacillus megaterium* small, acid-soluble spore proteins: cloning and nucleotide sequence of three additional genes from this multigene family. J. Bacteriol. **165**:467–473.
17. **Flores, M., V. Gonzalez, S. Brom, E. Martinez, D. Pinero, D. Romero, G. Davila, and R. Palacios.** 1987. Reiterated DNA sequences in *Rhizobium* and *Agrobacterium* spp. J. Bacteriol. **169**:5782–5788.
18. **Giroux, S., J. Beaudet, and R. Cedergren.** 1988. Highly repetitive $tRNA^{Pro}$-$tRNA^{His}$ gene cluster from *Photobacterium phosphoreum*. J. Bacteriol. **170**:5601–5606.
19. **Golden, S. S., J. Brusslan, and R. Haselkorn.** 1986. Expression of a family of *psbA* genes encoding a photosystem II polypeptide in cyanobacterium *Anacystis nidulans* R2. EMBO J. **5**:

2789–2798.
20. **Gustafsson, B. E., T. Midtvedt, and A. Norman.** 1966. Isolated fecal microorganisms capable of 7α-dehydroxylating bile acids. J. Exp. Med. **123**:413–432.
21. **Hayakawa, S.** 1973. Microbial transformation of bile acids. Adv. Lipid Res. **11**:143–192.
22. **Hill, C. W., R. H. Grafstrom, B. W. Harnish, and B. S. Hillman.** 1977. Tandem duplications resulting from recombination between ribosomal RNA genes in *Escherichia coli*. J. Mol. Biol. **116**:407–428.
23. **Hirano, S., R. Nakama, M. Tamaki, N. Matsuda, and H. Oda.** 1981. Isolation and characterization of thirteen intestinal microorganisms capable of 7α-dehydroxylating bile acids. Appl. Environ. Microbiol. **41**:737–745.
24. **Holdeman, L. V., and W. E. C. Moore (ed.).** 1973. Anaerobic laboratory manual, 2nd ed. Virginia Polytechnic Institute and State University, Blacksburg.
25. **Houmard, J., V. Capuano, T. Coursin, and N. Tandeau de Marsac.** 1988. Genes encoding core components of the phycobilisome in the cyanobacterium *Calothrix* sp. strain PCC 7601: occurrence of a multigene family. J. Bacteriol. **170**:5512–5521.
26. **Hulett, F. M., P.-Z. Wang, M. Sussman, and K. L. Jung-Wan.** 1985. Two alkaline phosphatase genes positioned in tandem in *Bacillus licheniformis* MC14 require different RNA polymerase holoenzymes for transcription. Proc. Natl. Acad. Sci. USA **82**:1035–1039.
27. **Hylemon, P. B.** 1985. Metabolism of bile acids in intestinal microflora, p. 331–343. *In* H. Danielsson and J. Sjovall (ed.), Sterols and bile acids: new comprehensive biochemistry, vol. 12. Elsevier Science Publishing, Inc., New York.
28. **Hylemon, P. B., and T. L. Glass.** 1983. Biotransformation of bile acids and cholesterol by the intestinal microflora, p. 189–213. *In* D. J. Henteg (ed.), Human intestinal microflora in health and disease. Academic Press, Inc., New York.
29. **Jornvall, H., H. von Bahr-Lindstrom, K. D. Jany, W. Ulmer, and M. Froschle.** 1984. Extended super family of short alcohol-polyol-sugar-dehydrogenase: structural similarities between glucose and ribitol dehydrogenases. FEBS Lett. **165**:190–196.
30. **Kulkarni, M. S., P. M. Heideprien, and K. L. Yielding.** 1980. Production by lithocholic acids of DNA strand breaks in L1210 cells. Cancer Res. **40**:2666–2669.
31. **Lin, R.-J., M. Capage, and C. W. Hill.** 1984. A repetitive DNA sequence, *rhs*, responsible for duplications with the *Escherichia coli* K-12 chromosome. J. Mol. Biol. **177**:1–18.
32. **Lizardi, P. M.** 1981. Binding and recovery of DNA and RNA using S&S NA-45 DEAE membrane. Schleicher & Schuell sequences application update no. 364. Schleicher & Schuell, Inc., Keene, N.H.
33. **Low-Beer, T. S., and S. Nutter.** 1978. Colonic activity, biliary cholesterol saturation, and the pathogenesis of gallstones. Lancet **ii**:1063–1065.
34. **Marmur, J.** 1961. A procedure for the isolation of deoxyribonucleic acid from microorganisms. J. Mol. Biol. **3**:208–218.
35. **Mori, M., A. Tanimoto, K. Yoda, S. Harada, N. Koyama, K. Hashiguchi, M. Obinata, M. Yamasaki, and G. Tamura.** 1986. Essential structure in the cloned transforming DNA that induces gene amplification of the *Bacillus subtilis amyE-tmrB* region. J. Bacteriol. **166**:787–794.
36. **Narisawa, T., N. E. Magadia, J. H. Weisberger, and E. L. Wynder.** 1974. Promoting effect of bile acids on colon carcinogenesis after intrarectal instillation of N-methyl-N′-nitrosoguanidine in rats. J. Natl. Cancer Inst. **53**:1093–1097.
37. **Palmer, R. H.** 1976. Toxic effects of lithocholate on the liver and biliary tree, p. 227–240. *In* W. Taylor (ed.), The hepatobiliary system: fundamental and pathological mechanism. Plenum Publishing Corp., New York.
38. **Paone, D. A. M., and P. B. Hylemon.** 1984. HPLC purification and preparation of antibodies to cholic acid-inducible polypeptides from *Eubacterium* sp. VPI 12708. J. Lipid. Res. **25**: 1343–1349.
39. **Punita, S. Jafri, M. A. Reddy, and H. K. Das.** 1989. Multiple chromosomes in *Azotobacter vinelandii*. J. Bacteriol. **171**:3133–3138.
40. **Quinto, C., H. de la Vega, M. Flores, J. Leemans, M. A. Cevallos, M. A. Pardo, R. Azpiroz, M. D. L. Girard, E. Calva, and R. Palacois.** 1985. Nitrogenase reductase: a functional multigene family in *Rhizobium phaseoli*. Proc. Natl. Acad. Sci. USA **82**:1170–1174.
41. **Reddy, B. A., J. W. Weisburger, and E. L. Wynder.** 1978. Colon cancer: bile salts as tumor promoters, p. 454–464. *In* T. J. Slaga, A. Savik, and R. J. Boutwell (ed.), Carcinogenesis, vol. 2. Mechanism of tumor promotion and carcinogenesis. Raven Press, New York.
42. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
43. **Schoedel, C., and A. Collmer.** 1986. Evidence of homology between the pectate lyase-encoding *pelB* and *pelC* genes in *Erwinia chrysanthemi*. J. Bacteriol. **167**:117–123.
44. **Sela, S., D. Yogev, S. Razin, and H. Bercovier.** 1989. Duplication of the *tuf* gene: a new insight into the phylogeny of eubacteria. J. Bacteriol. **171**:581–584.
45. **Sjövall, J.** 1960. Bile acids in man under normal and pathological conditions: bile acids and steroids 73. Clin. Chim. Acta **5**:33–41.
46. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
47. **Stellwag, E. J., and P. B. Hylemon.** 1979. 7α-dehydroxylation of cholic acid and chenodeoxycholic acid by *Clostridium leptum*. J. Lipid Res. **20**:325–333.
48. **Vagner, V., and S. D. Ehrlich.** 1988. Efficiency of homologous DNA recombination varies along the *Bacillus subtilis* chromosome. J. Bacteriol. **170**:3978–3982.
49. **Wang, S.-Z., J.-S. Chen, and J. L. Johnson.** 1988. The presence of five *nifH*-like sequences in *Clostridium pasteurianum*: sequence divergence and transcription properties. Nucleic Acids Res. **16**:439–454.
50. **Wek, R. C., C. A. Hauser, and G. W. Hatfield.** 1985. The nucleotide sequence of the *ilvBN* operon of *Escherichia coli*: sequence homologies of the acetohydroxy acid synthase isozymes. Nucleic Acids Res. **13**:3995–4010.
51. **White, B. A., R. L. Lipsky, R. J. Fricke, and P. B. Hylemon.** 1980. Bile acid induction specificity of 7α-dehydroxylase activity in an intestinal *Eubacterium* species. Steroids **35**:103–109.
52. **White, W. B., J. P. Coleman, and P. B. Hylemon.** 1988. Molecular cloning of a gene encoding a 45,000-dalton polypeptide associated with bile acid 7-dehydroxylation in *Eubacterium* sp. strain VPI 12708. J. Bacteriol. **170**:611–616.
53. **White, W. B., C. V. Franklund, J. P. Coleman, and P. B. Hylemon.** 1988. Evidence for a multigene family involved in bile acid 7-dehydroxylation in *Eubacterium* sp. strain VPI 12708. J. Bacteriol. **170**:4555–4561.
54. **Wilson, C. R., and A. E. Morgan.** 1985. Chromosomal-DNA amplification in *Bacillus subtilis*. J. Bacteriol. **163**:445–453.
55. **Woods, D.** 1984. Oligonucleotide screening of cDNA libraries. Focus **6**:1–2.
56. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
57. **Young, M., and S. D. Ehrlich.** 1989. Stability of reiterated sequences in the *Bacillus subtilis* chromosome. J. Bacteriol. **171**:2653–2656.
58. **Young, R. A., and R. W. Davis.** 1983. Efficient isolation of genes by using antibody probe. Proc. Natl. Acad. Sci. USA **80**: 1194–1198.