JOURNAL OF BACTERIOLOGY, Mar. 1997, p. 1734–1747 Vol. 179, No. 5
0021-9193/97/$04.00+0


# Molecular Systematic Studies of Eubacteria, Using $\sigma^{70}$-Type Sigma Factors of Group 1 and Group 2

TANJA M. GRUBER AND DONALD A. BRYANT*

*Department of Biochemistry and Molecular Biology and Center for Biomolecular Structure and Function, The Pennsylvania State University, University Park, Pennsylvania 16802*

Received 22 August 1996/Accepted 26 December 1996

**Sigma factors of the $\sigma^{70}$ family were used as a phylogenetic tool to compare evolutionary relationships among eubacteria. Several new sigma factor genes were cloned and sequenced to increase the variety of available sequences. Forty-two group 1 sigma factor sequences of various species were analyzed with the help of a distance matrix method to establish a phylogenetic tree. The tree derived by using sigma factors yielded subdivisions, including low-G+C and high-G+C gram-positive bacteria, cyanobacteria, and the α, β, γ, and δ subdivisions of proteobacteria, consistent with major bacterial groups found in trees derived from analyses with other molecules. However, some groupings (e.g., the chlamydiae, mycoplasmas, and green sulfur bacteria) are found in different positions than for trees obtained by using other molecular markers. A direct comparison to the most extensively used molecule in systematic studies, small-subunit rRNA, was made by deriving trees from essentially the same species set and using similar phylogenetic methods. Differences and similarities based on the two markers are discussed. Additionally, 31 group 2 sigma factors were analyzed in combination with the group 1 proteins in order to detect functional groupings of these alternative sigma factors. The data suggest that promoters recognized by the major vegetative sigma factors of eubacteria will contain sequence motifs and spacing very similar to those for the $\sigma^{70}$ sigma factors of *Escherichia coli*.**

Genotypic information has replaced morphological or other phenotypic patterns as the primary tool to determine evolutionary relationships among microorganisms. Sequence information can be more precisely interpreted and is usually more informative of evolutionary relationships than phenotypic information (95). The primary choice for such molecular systematic studies has been the small-subunit rRNA (SS-rRNA). After molecules were first recognized as being "documents of evolutionary history" about three decades ago (98), comparisons of SS-rRNA data have revolutionized knowledge of phylogenetic relationships, especially of microorganisms (27, 62, 63). For example, sequence comparisons of rRNA sequences first established the existence of eubacteria and archaebacteria as two distinct kingdoms among the prokaryotes (1). There are a number of reasons that make SS-rRNA sequence comparisons very useful for molecular systematic studies (63, 95), some of which are the following: (i) this molecule is present in all organisms, as well as mitochondria and chloroplasts; (ii) different positions in SS-rRNA sequences change at different rates, allowing the determination of phylogenetic relationships of close and distant relatives; (iii) SS-rRNA sizes are large and secondary structure aids alignment of sequences between different species; and (iv) the SS-rRNA molecules are relatively easily cloned and sequenced.

In spite of these significant advantages, a number of potential problems are associated with SS-rRNA sequence comparisons. Rothschild et al. (70) have pointed out several sources of error if SS-rRNA is used as the sole criterion to generate a phylogeny. Some of these include the possibility of gene transfer between species as well as between cellular organelles, intraorganismal heterogeneity of the rRNA genes since they occur in multiple copies in most organisms, and the occurrence of nonrandom mutations due to the influence of adjacent bases. Misinterpretations of relationships have also been detected due to drastically different G+C contents of species (37, 90). For these and other reasons, different molecules are being used for phylogenetic studies to compare and contrast with results obtained by using SS-rRNA. A variety of protein sequences have been used for this purpose, including sequences of RecA (20, 44), transketolase and fructosebisphosphate aldolase (88), HSP70 (5, 35, 69, 88), RNA polymerase (46), elongation factor TU (16, 53), the β subunit of ATP synthase (53), genes of photosynthetic reaction centers (60), and GroEL (91). Studies that include sequence data from various enzymes across the species have also been carried out to determine the divergence times of the major kingdoms (17). It seems that many different genes should be investigated from the same organisms to be able to understand their phylogenetic relationships more completely and reliably.

In this study, group 1 $\sigma^{70}$-type sigma factors were chosen as marker molecules for eubacterial systematics. The sigma factor protein is a dissociable subunit of the eubacterial RNA polymerase holoenzyme (9). It confers on the RNA polymerase core enzyme the promoter specificity that it requires for transcription initiation. Two broad families of sigma factors have been identified: the $\sigma^{70}$-type sigma factors (for reviews, see references 38 and 51) and the $\sigma^{54}$-type sigma factor family (for reviews, see references 14 and 48), both named after the originally identified *Escherichia coli* proteins. The $\sigma^{70}$ family has been divided into three groups (51). Group 1 is composed of the primary sigma factors of the organisms, and its members are present in all known eubacteria, in which they are essential for cell viability. Group 2 and group 3 include the so-called alternative sigma factors, and these are dispensable for cell growth. Group 2 members are rather similar in sequence to primary sigma factors and include proteins such as the stationary-phase-specific sigma factor RpoS (59). Group 3 sigma factors vary more significantly in sequence from the other two groups and include functional groupings such as heat shock

* Corresponding author. Mailing address: Department of Biochemistry and Molecular Biology, S-234 Frear Bldg., Penn State University, University Park, PA 16802. Phone: (814) 865-1992. Fax: (814) 863-7024. E-mail: dab14@psu.edu.

(34) and sporulation sigma factors (for a review, see reference 36).

Group 1 sigma factors possess a variety of characteristics that make them potential candidates as phylogenetic markers. They occur across the eubacterial kingdom and are essential proteins for cell viability. Functional constraints for these proteins give rise to highly conserved structural features. Even though sigma factor proteins vary in length, four regions of very high sequence conservation have been identified (38). These regions span approximately 250 contiguous amino acids and include the subregions that are responsible for recognition of the −10 and −35 promoter elements (78, 92). Alignment of the homologous regions is rather straightforward and appears to be free of significant ambiguities. A further advantage of sigma factors is the diversity and relatively large number of sequences already available. Open reading frames encoding proteins that show high sequence similarity to the functionally characterized *E. coli* $\sigma^{70}$ proteins (80) have been found in a variety of species at a rapidly growing rate in the last few years. Considering that putative sigma factors from diverse species have been shown to carry out functions equivalent to those of the sigma factors in *E. coli*, the members of this family are now considered homologs (13, 76).

In this work, a total of 73 group 1 and group 2 $\sigma^{70}$-type sigma factors were studied to investigate if sigma factors are useful molecules for molecular systematic studies. Several additional sigma factor genes have been cloned from diverse species, including the cyanobacterium *Synechococcus* sp. strain PCC 7002, the gram-positive photosynthetic bacterium *Heliobacillus mobilis*, the green sulfur bacterium *Chlorobium tepidum*, the green gliding bacterium *Chloroflexus aurantiacus*, the photosynthetic proteobacterium *Rhodobacter sphaeroides*, and the thermophile *Thermotoga maritima*. These sequences were chosen to obtain a more balanced representation of species in the tree. A phylogenetic tree of group 1 sigma factors was established by using a distance method and was compared to a tree derived from SS-rRNA sequences that included essentially the same species as the ones included in the sigma factor analysis. Furthermore, a tree that included group 1 and group 2 sequences was generated in order to detect possible functional groupings of group 2 sigma factors. The sigma factor tree is the first protein tree which includes representatives of all phyla that include photosynthetic bacteria.

## MATERIALS AND METHODS

**Bacterial strains and plasmids.** *E. coli* DH5α (Bethesda Research Laboratories, Gaithersburg, Md.) was used for recombinant DNA manipulations, and *E. coli* MV1190 was used as the host strain for the production of single-stranded DNA template for sequencing. Plasmid pUC19 was used for most routine DNA cloning and double-stranded sequencing procedures (97). Plasmid pHSG575 (81) was chosen for the construction of the DNA library for *T. maritima*. M13 phage cloning vectors M13mp18 and M13mp19 were used for single-stranded sequencing.

**DNA preparation and DNA sequence analysis.** Chromosomal DNA from *R. sphaeroides* was kindly provided by Timothy Donohue (University of Wisconsin, Madison), DNA from *H. mobilis* was provided by Wim Vermaas (Arizona State University, Tempe), and DNA from *T. maritima* was provided by Michael Adams (University of Georgia, Athens).

Small-scale plasmid DNA preparations from *E. coli* were extracted by the alkaline lysis method (4). Large-scale DNA preparations were performed by the alkaline lysis method followed by CsCl-ethidium bromide equilibrium density gradient ultracentrifugation. Plasmid DNA used for nested deletions was purified through two successive density gradients. Single-stranded DNA templates for sequencing with M13mp vectors were purified according to the manufacturer's instructions (M13 cloning kit; Amersham, Arlington Heights, Ill.).

The DNA sequences were determined by the dideoxy-chain termination method (72), with a Sequenase version 2.0 DNA sequencing kit from United States Biochemical. Sequence data were analyzed with the MacVector sequence analysis program, version 5.0 (Eastman Kodak, Rochester, N.Y.).

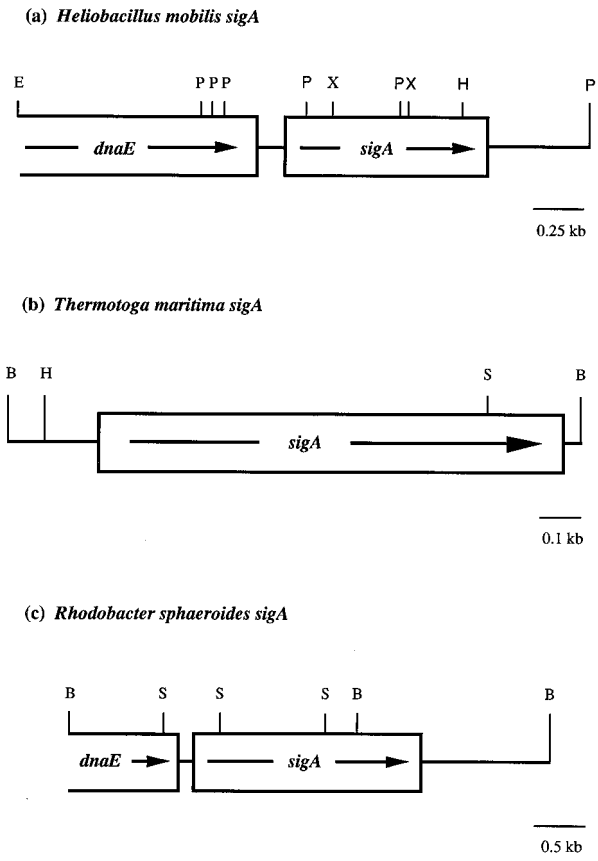


FIG. 1. Restriction maps for *sigA* genes isolated in this study. Maps are shown for *H. mobilis sigA* (a), *T. maritima sigA* (b), and *R. sphaeroides sigA* (c). For two of these (a and c), the position of the upstream *dnaE* gene is also indicated. Abbreviations for restriction endonucleases: B, *Bam*HI; H, *Hin*dIII; E, *Eco*RI; P, *Pst*I, S, *Sst*I; X, *Bst*XI.

**Southern hybridization and screening of plasmid libraries.** DNA fragments were transferred to nitrocellulose membranes (Schleicher & Schuell) by capillary transfer and hybridized overnight under conditions described previously (8). DNA fragments used as hybridization probes were purified from agarose gels (GenElute agarose spin columns; Supelco, Bellefonte, Pa.) and radiolabeled with [α-$^{32}$P]dATP (New England Nuclear, Beverly, Mass.), using a random-primed DNA labeling kit (Boehringer Mannheim, Indianapolis, Ind.).

To create size-directed plasmid libraries, total chromosomal DNA was first digested with restriction endonucleases, size fractionated on an agarose gel, and screened by Southern hybridization to identify fractions containing the desired fragment. Fragments from this fraction were purified by electrophoresis, isolated from the agarose gel by using GenElute agarose spin columns (Supelco), and ligated into plasmids as indicated. After transformation of ligation products into *E. coli* DH5α cells, the libraries were screened by colony hybridization to identify clones containing the desired DNA fragments. Confirmation of the correctness of these clones was made by restriction endonuclease mapping and Southern hybridization experiments with plasmids isolated from all putatively positive clones.

**Cloning and sequence analysis of *H. mobilis sigA*, *T. maritima sigA*, and *R. sphaeroides sigA*.** A Southern blot with digests of *H. mobilis* chromosomal DNA was hybridized with a 124-bp *Eco*RI-*Sma*I fragment of *Synechococcus* sp. strain PCC 7002 *sigB* (11), encoding amino acids 117 to 158 of the SigB sequence (see Fig. 2). Southern blots containing *T. maritima* and *R. sphaeroides* DNAs were hybridized with a 400-bp *Bst*XI fragment of *H. mobilis sigA* (Fig. 1); this fragment spans regions 2.1 to 2.4 of this $\sigma^{70}$-type sigma factor. Based on these and other hybridization experiments, a 2.3-kbp *Eco*RI-*Hin*dIII fragment and a 0.9-kbp *Pst*I fragment of *H. mobilis* were cloned into pUC19 (Fig. 1) to obtain the entire sequence of the putative *sigA* gene. In the case of *T. maritima*, a 1.4-kbp *Bam*HI fragment (Fig. 1) was cloned into pHSG575 (81). For *R. sphaeroides*, a 0.4-kbp *Bam*HI-*Sst*I fragment (Fig. 1) was cloned into pUC19. This cloned fragment was used as a probe to isolate overlapping cosmid pUI8315 from the pLA2917 cosmid bank of *R. sphaeroides* (from Timothy Donohue). The nucleotide sequence of a 5-kbp *Bam*HI fragment was determined (Fig. 1). Analyses of the

2.4

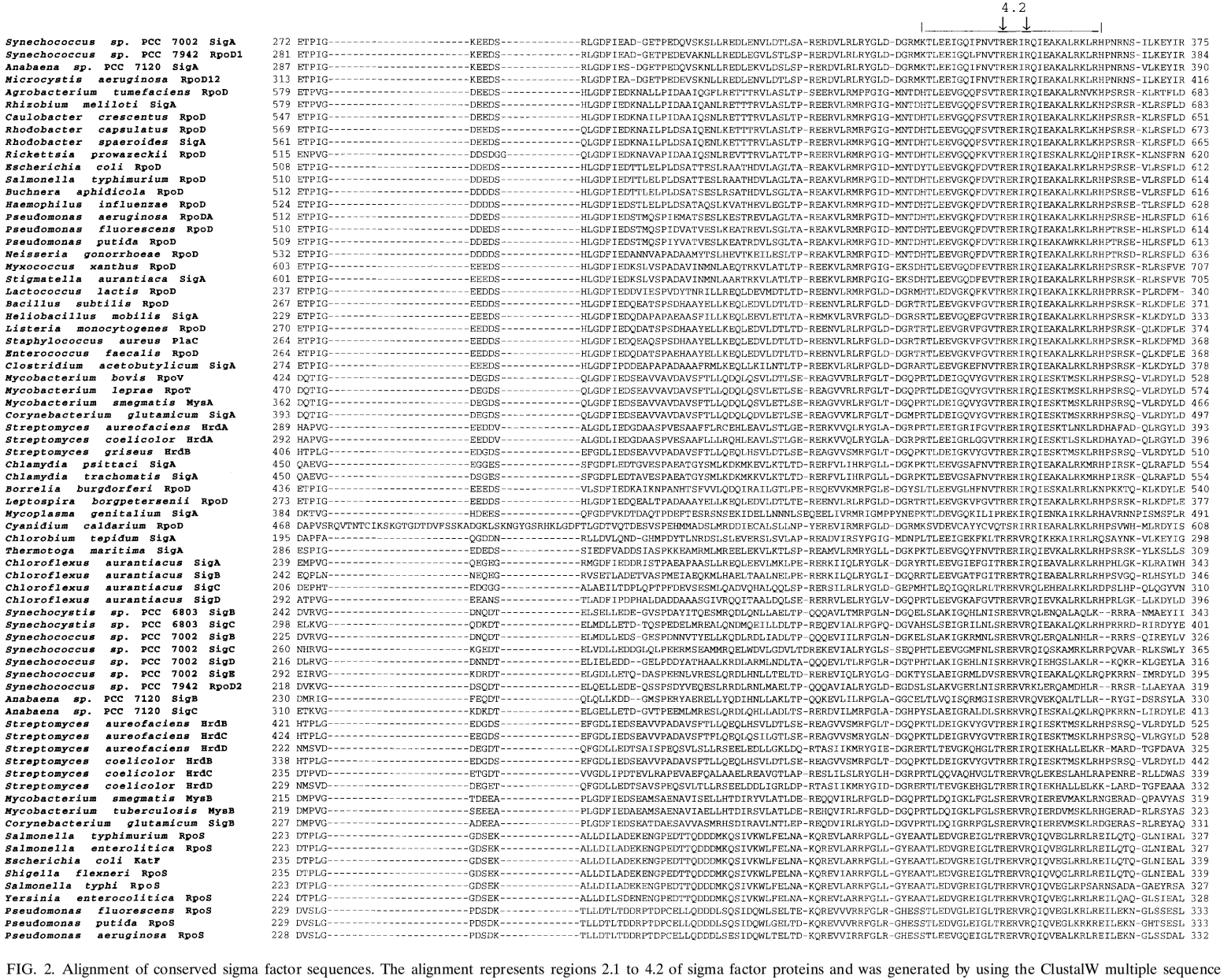


FIG. 2. Alignment of conserved sigma factor sequences. The alignment represents regions 2.1 to 4.2 of sigma factor proteins and was generated by using the ClustalW multiple sequence alignment program. Numbers at the beginning and end of each line indicate amino acid positions relative to the start of each protein sequence. Hyphens indicate gaps introduced to maximize the alignment. Conserved regions 2.4 and 4.2, which are believed to interact with promoter regions −10 and −35, respectively, are indicated by a bar across the top of the alignment. Residues that have been shown in *Bacillus subtilis* and *E. coli* to interact specifically with the −10 and −35 motifs are indicated by arrows above the alignments.

deduced SigA protein sequences of all three organisms showed very strong sequence similarity (Fig. 2) to the four conserved regions of $\sigma^{70}$-type sigma factors as defined by Helmann and Chamberlin (38), which indicated that these genes encode $\sigma^{70}$-type sigma factors.

The hybridization analyses of *H. mobilis* and *T. maritima* DNAs showed a single hybridizing DNA fragment for each restriction enzyme combination tested, suggesting that a single group 1+2 $\sigma^{70}$-type sigma factor homolog occurs in these organisms. One strongly hybridizing band and at least one weaker hybridizing band were detected by the Southern blot analyses of *R. sphaeroides*, indicating the probable existence of additional $\sigma^{70}$-type sigma factors, probably group 2 members, in this organism.

**Sequence alignments.** Amino acid sequences obtained in this work were combined with sequences of sigma factors obtained from the National Center for Biotechnology Information databases (39) and aligned with the ClustalW multiple sequence alignment program (86), using default parameters. The alignment was inspected and no further refinement was necessary. Database searches were conducted with various sigma factor sequences (*E. coli* RpoD, *Bacillus subtilis* RpoD, and *Synechococcus* sp. strain PCC 7002 SigA), and group 1+2 members that were detected by the searches were included in the alignment. Only the most conserved regions of sigma factors (regions 2.1 to 4.2 [51]) were included in the final alignment, shown in Fig. 2. Table 1 shows the accession numbers and gene designations of the sigma factor sequences that were used in creating the alignment. Table 1 also indicates to which group the various sigma factors have been assigned. For the SS-rRNA tree, sequences of species for which $\sigma^{70}$-type group 1 sigma factor sequences were also available were chosen. In most cases, a complete or nearly complete SS-rRNA sequence was available for each organism from which a group 1 sigma factor sequence was available. Table 1 lists the SS-rRNA sequences used and indicates those species for which replacement SS-rRNA sequences were used as well as those species for which no suitable replacement was available. Replacement sequences were chosen from organisms that are considered to be very close relatives of the species when no matching SS-rRNA sequence was available. The aligned SS-rRNA sequences were obtained by using the ribosomal database project (RDP) computer server (54).

**Phylogenetic trees.** The sequence alignments described above were used to construct phylogenetic trees. Regions 2.1 to 4.2 of $\sigma^{70}$ type sigma factors were chosen, as they constitute the most highly conserved regions of that protein family (51). The SS-rRNA sequences were extracted from the RDP server in aligned format. The alignment was based on both primary and secondary structure considerations (54) and included 1,683 alignment positions. This number constitutes all positions as supplied by the RDP server.

The phylogenetic trees were generated with computer algorithms supplied by the PHYLIP software package (25). Trees derived from sigma factor sequences were generated by using the method developed by Fitch and Margoliash (26), using the estimated evolutionary distances calculated by the protdist program. The PAM matrix-based distance correction (25) was used to calculate the pairwise distances between sigma factor proteins. For the $\sigma^{70}$ group 1 tree, aligned sequences of known or presumed group 1 $\sigma^{70}$-type sigma factors were included in the analysis, whereas all of the sequences indicated in Table 1 were used for the construction of the tree that also includes group 2 sigma factors. An alternative treeing method was also used to analyze the relationships among group 1 sigma factor sequences. The protpars program (24), supplied by the PHYLIP software package, was used to produce a parsimony tree. To obtain bootstrap values, 100 bootstrap analyses were performed. This tree was compared to the results obtained by using the distance method. The phylogenetic tree for the SS-rRNA was generated by a distance method. The pairwise distances between SS-rRNA sequences were calculated by using the Kimura two-parameter distance correction implemented in the dnadist program of PHYLIP. The tree was constructed by the Fitch-Margoliash method. In all cases, the trees were generated without an outgroup and can thus be considered unrooted. For both the sigma factor and SS-rRNA trees which were generated by distance methods, 500 bootstrap replications were carried out to obtain the bootstrap values (23).

**Nucleotide sequence accession numbers.** The DNA sequences for the *H. mobilis sigA* gene, the *T. maritima sigA* gene, and the *R. sphaeroides rpoD* gene have been deposited in GenBank under accession numbers U67424, U67423, and U67425. Molecular cloning, sequencing, and characterization of the *sigA* gene of *Chlorobium tepidum*, the four sigma factors of *C. aurantiacus*, and the *sigD* and *sigE* genes of *Synechococcus* sp. strain PCC 7002 will be described in detail elsewhere, although the sequences have been deposited in GenBank (34a) (Table 1).

## RESULTS

**Group 1 sigma factors.** Sigma factor sequences were aligned by using the program ClustalW as shown in Fig. 2. This alignment includes the most highly conserved regions of all members of the $\sigma^{70}$ family of sigma factors (51). Most of the gaps introduced to optimize this alignment derive from only two sequences, those of *Mycoplasma genitalium* and the nucleus-encoded, chloroplast-targeted SigA protein of *Cyanidium caldarium*. Figure 3 shows the phylogenetic tree obtained for $\sigma^{70}$ group 1 sigma factors, using the procedures described in the Materials and Methods. For all organisms except *C. aurantiacus*, the known or assumed group 1 sigma factor was included, as indicated in Table 1. In the case of *C. aurantiacus*, it was not possible to assign one of the four sigma factors as the group 1 representative (34a); consequently, all four protein sequences were included in this alignment. Figure 4 shows the tree obtained by using SS-rRNA sequences of the same species that were included in the group 1 sigma factor tree. This allows a direct comparison between the trees for the two molecular markers, since factors such as differing sample sizes within separate phyla should not influence the results.

The proteobacterial phylum has been divided into five phylogenetically distinct subdivisions (α, β, γ, δ, and ε), based on SS-rRNA sequences (62). All proteobacterial sequences are monophyletic in both the sigma factor tree and the SS-rRNA tree of this work, with four of the subgroups (α, β, γ, and δ) being represented. It is interesting that, based on the bootstrap values, the four subdivisions are more strongly supported in the sigma factor tree than in the SS-rRNA tree, considering that the subdivisions were originally based on SS-rRNA sequence comparisons. The bootstrap values for α, β plus γ, γ, and δ in the sigma factor tree are 99, 100, 72, and 100, respectively, while in the SS-rRNA tree, the corresponding values are 66, 100, 55, and 100. The species composing the four subdivisions of proteobacteria are equivalent between the two trees. The positionings of the four proteobacterial subdivisions relative to each other are also similar in the two trees; the β and γ subgroups are relatively closely related to each other, and the δ group is positioned somewhat separately from the other three subgroups. This pattern has been observed in several other studies (20, 62, 95). Another phylogenetic pattern that is consistent in both trees is the separation of *Pseudomonas* species from *E. coli*, *Buchnera aphidicola*, *Haemophilus influenzae*, and *Salmonella typhimurium* within the γ subdivision.

The gram-positive phylum consists of four subdivisions (95), two of which have been well characterized. These two subgroups, the low-GC and high-GC groups, are represented in this study, as well as a member of the third subgroup, namely, *H. mobilis* of the photosynthetic subdivision. The represented sequences do not form a monophyletic gram-positive clade that is composed of all subdivisions in either tree. The high-GC organisms form a monophyletic cluster in both the RNA tree and the protein tree. The composition of species varies in two respects between the two trees for the high-GC subgroup. In the rRNA tree, the sequence of *Corynebacterium mediolanum* was substituted for that of *Corynebacterium glutamicum*, but no suitable substitute for *Streptomyces aureofaciens* was available. Nevertheless, the relationships of the organisms forming this subgroup are virtually identical between the two trees; the *Streptomyces* and *Mycobacterium* spp. form tight clusters, and *Corynebacterium* is positioned between these two clusters. In the low-GC cluster, most organisms included are also found in equivalent positions in the two trees. The only member of the photosynthetic subdivision, *H. mobilis*, is found within this cluster (*Heliobacterium chlorum* as the substitute species in the RNA tree). In both trees, the photosynthetic representative is most closely related to *Clostridium acetobutylicum.*

In the sigma factor tree, *Leptospira borgpetersenii* is placed within the low-GC group, and this placement is supported by a very high bootstrap value. However, *L. borgpetersenii* is a spirochete, which is also apparent from its position on the SS-RNA tree. There are several possible explanations for this incongruency. First, a lateral gene transfer from another species, most likely involving a member of the low-GC gram-positive grouping, could have occurred. Alternatively, the *L.*

TABLE 1. Sigma factor and SS-rRNA sequences

| Species (by phylum) | Gene | $\sigma^{70}$ group[a] | Size (amino acids) | GenBank accession no. | SS-rRNA sequence[b] | Reference[c] |
|---|---|---|---|---|---|---|
| Cyanobacteria/chloroplasts | | | | | | |
| *Anabaena* sp. strain PCC 7120 | *sigA* | 1 | 390 | M60046 | NOST.MUSCR | 7 |
| | *sigB* | 2 | 331 | M95760 | | 6 |
| | *sigC* | 2 | 416 | M95759 | | 6 |
| *Cyanidium caldarium* | *rpoD* | NA | 609 | L42639 | | 50 |
| *Microcystis aeruginosa* | *rpoD1* | (1) | 416 | D50318 | MCS.AERUG2 | |
| *Synechococcus* sp. strain PCC 7002 | *sigA* | 1 | 374 | U15574 | NA | 12 |
| | *sigB* | 2 | 328 | U82435 | | 11 |
| | *sigC* | 2 | 365 | U82436 | | 11 |
| | *sigD* | 2 | 317 | U82484 | | 34a |
| | *sigE* | 2 | 398 | U82485 | | 34a |
| *Synechococcus* sp. strain PCC 7942 | *rpoD1* | (1) | 384 | D10973 | SYN.6301* | 82 |
| | *rpoD2* | 2 | 320 | D78583 | | 82 |
| *Synechocystis* sp. strain PCC 6803 | *sigB* | 2 | 345 | D63999 | | 43 |
| | *sigC* | 2 | 404 | D64002 | | 43 |
| Gram-positive bacteria | | | | | | |
| *Bacillus subtilis* | *rpoD* | 1 | 371 | M10089 | B.SUBTILI3 | 33 |
| *Clostridium acetobutylicum* | *sigA* | (1) | 378 | Z23080 | C.ACEBUTY3 | 74 |
| *Corynebacterium glutamicum* | *sigA* | (1) | 497 | Z49822 | COR.MEDIS1* | 61 |
| | *sigB* | 2 | 331 | Z49824 | | 61 |
| *Enterococcus faecalis* | *rpoD* | (1) | 368 | X86176 | ECO.FAECAL | |
| *Heliobacillus mobilis* | *sigA* | 1 | 333 | U67424 | HEL.CHLORM* | This work |
| *Lactococcus lactis* | *rpoD* | (1) | 340 | X71493 | LCC.LACTIS | 30 |
| *Listeria monocytogenes* | *rpoD* | (1) | 374 | U13165 | LIS.MONOC3 | 58 |
| *Mycobacterium bovis* | *rpoV* | (1) | 528 | U21130 | MYB.BOVIS | 15 |
| *Mycobacterium leprae* | *rpoT* | (1) | 574 | U15181 | MYB.LEPRA2 | |
| *Mycobacterium smegmatis* | *mysA* | (1) | 466 | U09821 | MYB.SMEGM1 | 65 |
| | *mysB* | 2 | 319 | U09863 | | 65 |
| *Mycobacterium tuberculosis* | *mysB* | 2 | 323 | U10059 | | 19 |
| *Mycoplasma genitalium* | *sigA* | 1 | 497 | U39703 | M.GENITAL2 | 28 |
| *Staphylococcus aureus* | *plaC* | (1) | 368 | M63177 | STP.AUREUS | 41 |
| *Streptomyces aureofaciens* | *hrdA* | 2 | 393 | M90410 | NA | 47 |
| | *hrdB* | (1) | 525 | M90411 | | 47 |
| | *hrdC* | 2 | 528 | M90412 | | 47 |
| | *hrdD* | 2 | 325 | M90413 | | 47 |
| *Streptomyces coelicolor* | *hrdA* | 2 | 396 | X52980 | STM.COELI3 | 83 |
| | *hrdB* | 1 | 442 | X52983 | | 83 |
| | *hrdC* | 2 | 339 | X52981 | | 83 |
| | *hrdD* | 2 | 332 | X52982 | | 83 |
| *Streptomyces griseus* | *hrdB* | (1) | 510 | X75952 | STM.GRISEUS | 57 |
| Proteobacteria | | | | | | |
| *Agrobacterium tumefaciens* | *rpoD* | (1) | 684 | X69388 | AG.TUMEFAC | 77 |
| *Buchnera aphidicola* | *rpoD* | (1) | 617 | M90644 | BUC.APHIDI | 49 |
| *Caulobacter crescentus* | *rpoD* | 1 | 652 | U35138 | CAU.CRES2 | 55 |
| *Escherichia coli* | *rpoD* | 1 | 613 | J01687 | E.COLI | 10 |
| | *katF* | 2 | 342 | Z14969 | | 59 |
| *Haemophilus influenzae* | *rpoD* | (1) | 629 | L45174 | H.INFLUENZ | 85 |
| *Myxococcus xanthus* | *rpoD* | (1) | 708 | U20669 | MYX.XANTHU | 40 |
| *Neisseria gonorrhoeae* | *rpoD* | (1) | 642 | L42289 | NIS.GONORR | |
| *Pseudomonas aeruginosa* | *rpoDA1* | 1 | 617 | D90118 | PS.AERUGIN | 84 |
| | *rpoS* | 2 | 334 | D26134 | | 29 |
| *Pseudomonas fluorescens* | *rpoD* | 1 | 615 | X84416 | PS.FLUORE3 | 75 |
| | *rpoS* | 2 | 335 | U34203 | | 73 |
| *Pseudomonas putida* | *rpoD* | (1) | 614 | D30045 | PS.PUTIDA | 29 |
| | *rpoS* | 2 | 335 | X91654 | | |
| *Rhizobium meliloti* | *sigA* | (1) | 684 | L47288 | NA | 71 |
| *Rhodobacter capsulatus* | *rpoD* | (1) | 674 | U28162 | RB.CAPSUL5 | |
| *Rhodobacter sphaeroides* | *sigA* | 1 | 666 | U67425 | RB.SPHAER2 | This work |
| *Rickettsia prowazekii* | *rpoD* | (1) | 635 | U02878 | RIC.PROWAZ | |
| *Salmonella typhi* | *rpoS* | 2 | 384 | X81641 | | |
| *Salmonella typhimurium* | *rpoD* | (1) | 615 | M14427 | S.TYMURLT2 | 22 |
| | *rpoS* | 2 | 330 | U05011 | | 66 |
| *Serratia entomophila* | *rpoS* | 2 | 332 | U35777 | | |
| *Shigella flexneri* | *rpoS* | 2 | 362 | U00119 | | |
| *Stigmatella aurantiaca* | *sigA* | (1) | 706 | M94370 | SMA.AURANT | 79 |
| *Yersinia enterocolitica* | *rpoS* | 2 | 331 | U16152 | | 42 |

*Continued on following page*

TABLE 1—*Continued.*

| Species (by phylum) | Gene | $\sigma^{70}$ group[a] | Size (amino acids) | GenBank accession no. | SS-rRNA sequence[b] | Reference[c] |
|---|---|---|---|---|---|---|
| Thermotogales | | | | | | |
| *Thermotoga maritima* | *sigA* | 1 | 397 | U67423 | TT.MARITIM | This work |
| Spirochetes | | | | | | |
| *Borrelia burgdorferi* | *rpoD* | (1) | 543 | U17591 | BOR.BURGD6 | |
| *Leptospira borgpetersenii* | *rpoD* | (1) | 331 | M96579 | LPS.BORGP2 | |
| Chlamydia | | | | | | |
| *Chlamydia psittaci* | *sigA* | 1 | 571 | U04442 | CLM.PSITTA | 18 |
| *Chlamydia trachomatis* | *sigA* | 1 | 571 | M36475 | CLM.TRACHO | 21 |
| Green gliding bacteria | | | | | | |
| *Chloroflexus aurantiacus* | *sigA* | NA | 342 | U67719 | CFX.AURANT | 34a |
| | *sigB* | NA | 346 | U67720 | | 34a |
| | *sigC* | NA | 312 | U67721 | | 34a |
| | *sigD* | NA | 398 | U67722 | | 34a |
| Green sulfur bacteria | | | | | | |
| *Chlorobium tepidum* | *sigA* | 1 | 299 | U67718 | CHL.TEPIDU | 34a |

[a] Represented sigma factor proteins are classified as either group 1 or group 2 members, except in the cases of *Cyanidium caldarium* and *C. aurantiacus*, where this information was not available (NA). Group 1 sigma factor proteins that have been confirmed as such by genetic or biochemical means, or that are the only group 1 or group 2 member of that organism, are indicated by the number 1; unconfirmed group 1 sigma factors are indicated by parentheses.

[b] The names of the SS-rRNA sequences refer to the ribosomal database project entries (54). Those ones marked with asterisks come from a different species than the corresponding sigma factor sequence(s). These replacement species are SYN.6301 (*Synechococcus* sp. PCC 6301), COR.MEDIS1 (*Corynebacterium mediolanum*), and HEL.CHLORM (*Heliobacterium chlorum*). Species where no replacement sequences were available are indicated as not available (NA).

[c] References for sigma factor entries as found in GenBank. The absence of a reference indicates that only a GenBank entry is available for that sequence.

*borgpetersenii* sigma gene product could be a group 2 sigma factor that happens to be unusually similar to the sigma factors of the low-GC gram-positive group, although this idea is not supported by inclusion of other group 2 sigma factor sequences (see below). It is possible that the cloned sigma factor gene is not actually from *L. borgpetersenii* but from some other low-GC gram-positive bacterium. It should be noted that the position of the spirochetes, represented by *Borrelia burgdorferi*, is not highly supported in either tree. In the SS-rRNA tree, the spirochetes are most closely related to *M. genitalium*, but the relationship to other species is ambiguous. In the sigma factor tree, *B. burgdorferi* is placed between the high-GC gram-positive bacteria and the green gliding bacteria, but this placement is not highly supported.

Mycoplasmas are generally placed within the low-GC gram-positive bacteria (93, 95), however, their phylogenetic classification is not entirely clear (68). In the sigma factor tree, the only representative of the mycoplasmas, *M. genitalium*, is most closely related to *T. maritima*, a positioning that is relatively well supported. In the SS-rRNA tree, *M. genitalium* is most closely related to the spirochetes; however, this assignment is unreliable as apparent by low bootstrap values.

The cyanobacteria form a coherent clade in both the SS-rRNA and sigma factor trees. Unfortunately, the SS-rRNA tree does not include the sequence for *Synechococcus* sp. strain PCC 7002, and the sequence of *Synechococcus* sp. strain PCC 6301 was used as a replacement for *Synechocystis* sp. strain PCC 6803. The low bootstrap values for the placement of the cyanobacterial clade together with other bacterial clades does not allow confident placement of the cyanobacteria relative to other groupings within both trees. In the SS-rRNA tree, the cyanobacteria are most closely related to the chlamydiae, whereas in the sigma factor tree, the cyanobacteria are most closely related to other photosynthetic organisms, namely, the green gliding bacterium *C. aurantiacus* and the green sulfur bacterium *Chlorobium tepidum.*

A number of phyla are represented by only one or two sequences (e.g., the thermotogales, chlamydiae, spirochetes, green gliding bacteria, and green sulfur bacteria) which contributes to the inability to produce a precise placement of some of these organisms within the trees. This is reflected by relatively low bootstrap values for these species in both the SS-rRNA and sigma factor trees. A notable exception is the very strongly supported position of the chlamydiae in the sigma factor tree. According to these results, a placement of the chlamydiae within or very near the proteobacterial group would seem to be appropriate.

**Group 1 and group 2 sigma factors.** Figure 5 shows the tree that was established by using both group 1 $\sigma^{70}$-type sigma factors and alternative sigma factor sequences of group 2. The group designations are outlined in Table 1. Comparing this tree with the group 1 tree shows that the group 1 divisions largely remain unchanged in terms of their species composition and relative positions to one another. The high-GC subdivision of the gram-positive bacteria is the only group 1 cluster that obtained an additional sequence, the HrdC protein of *S. aureofaciens*. This sequence is very similar to the group 1 sigma factor of this organism, and this could be due to a gene duplication that has occurred reasonably late in the history of the high-GC subdivision; alternatively, this gene could have been obtained by lateral gene transfer from another *Streptomyces* species. In the tree composed of all sigma factor sequences, the spirochetes have changed positions slightly relative to other phyla compared to the tree based exclusively on the group 1 sigma factors. This is not surprising considering that the bootstrap values do not support these nodes highly.

The sigma factors believed to be involved in the transcription of stationary-phase-specific genes (KatF and RpoS proteins) form a tight clade. All cyanobacterial group 2 sigma factors also form a coherent clade that is not closely related to the group 1 sigma factors of these same organisms. Contrary to the grouping composed of the stationary phase sigma factors, the functions of the individual members of these cyanobacterial gene products are not yet entirely clear (see Discussion).

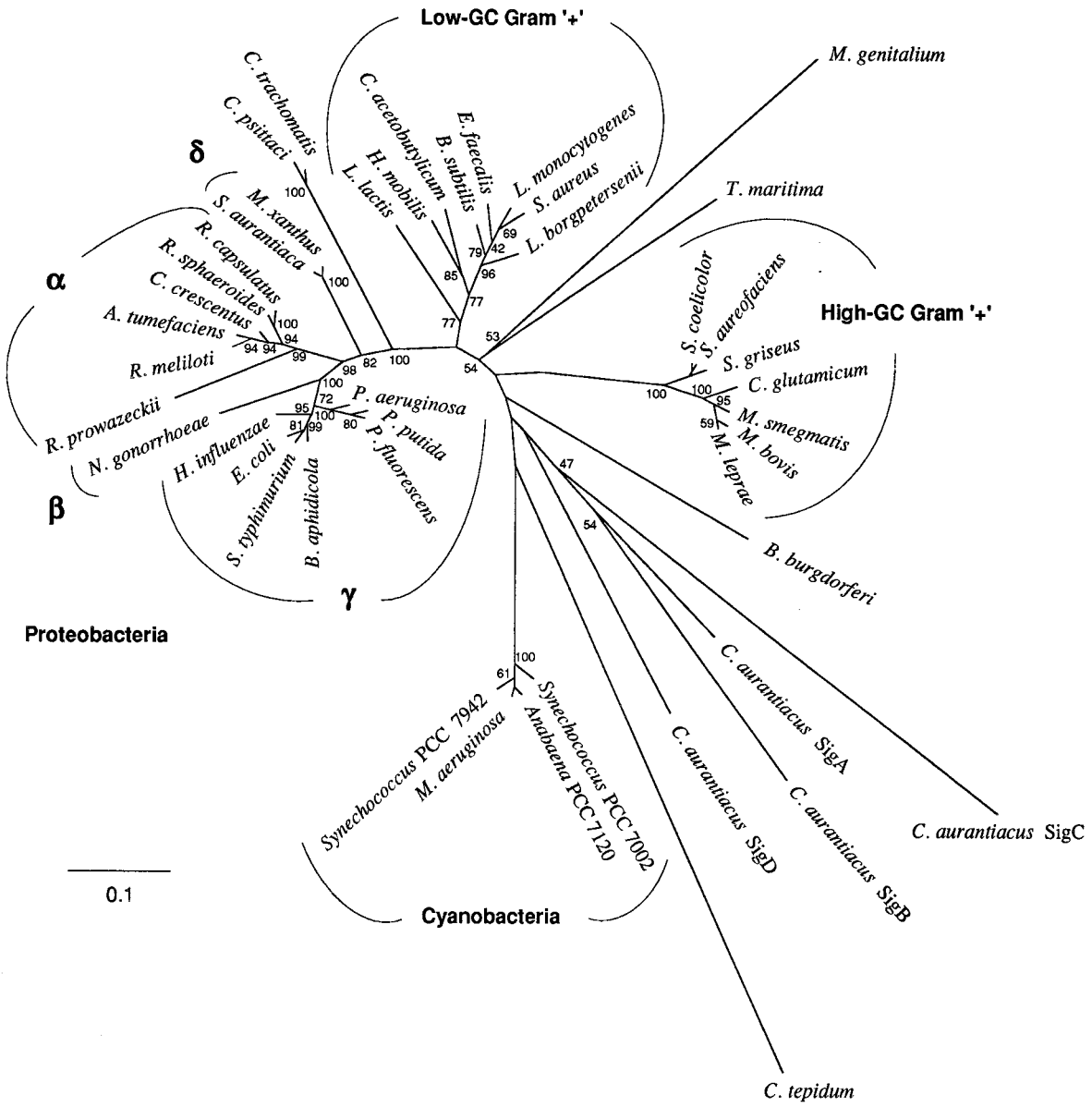


FIG. 3. Fitch-Margoliash tree for group 1 sigma factors. The tree was generated by the method of Fitch and Margoliash (26) from the multiple sequence alignments depicted in Fig. 2. The distances were calculated by using the protdist program of the PHYLIP package, using a PAM matrix-based distance correction. The scale bar represents 0.1 substitution per site. All sequences included are from group 1 sigma factors, except for the four *C. aurantiacus* sequences (CfxA, CfxB, CFxC, and CfxD), for which it was not possible to determine the identity of the group 1 member. Bootstrap values were obtained after 500 replications and are indicated when greater than 40%.

The high-GC gram-positive group 2 sigma factors do not form such a tight grouping as do the cyanobacterial group 2 members. Some are very closely related to high-GC gram-positive group 1 sigma factors (e.g., *S. aureofaciens* HrdC), whereas others (e.g., *Streptomyces coelicolor* HrdC and HrdD and *S. aureofaciens* HrdD) are only distantly related to the corresponding group 1 sequences. The four *C. aurantiacus* sigma factors are split into two groups: CfxD, CfxB, and CfxA, which are placed together near the cyanobacterial group 1 proteins, and CfxC, which is more closely related to the stationary-phase sigma factors. The group 1 sigma factor of *Chlorobium tepidum*, the only sigma factor sequence of groups 1 and 2 that could be detected in this species, is the only group 1 protein placed deep within group 2. Like CfxC, it is more closely related to the stationary-phase sigma factors. The only eukaryotic representative of this tree, the nucleus-encoded, chloroplast-targeted sigma factor RpoD of the unicellular rhodophyte *Cyanidium caldarium* (50), is found associated with the cyanobacterial group 1 sigma factors. Although the length and depth of this branch suggest that this sequence has diverged significantly from its cyanobacterial counterparts, this result is not surprising considering the established relatedness between cyanobacteria and chloroplasts (32, 94).

## DISCUSSION

In this study, several new group 1+2 sigma factors have been cloned and sequenced. The deduced protein sequences were

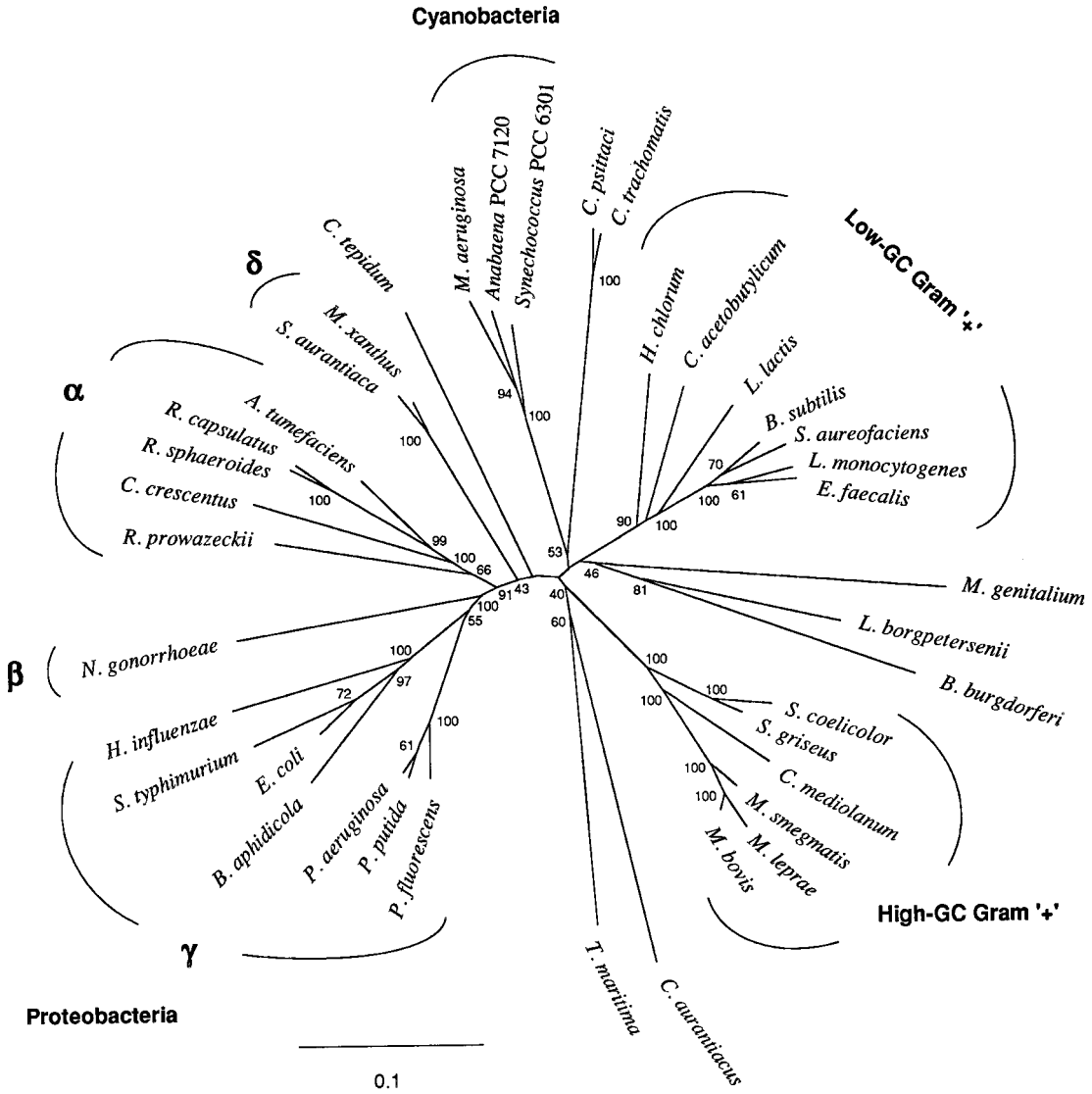


FIG. 4. Fitch-Margoliash tree for SS-rRNA sequences. The distances were calculated by using the dnadist program of the PHYLIP package, with the Kimura two-parameter as the distance correction. *Synechococcus* sp. strain PCC 6301, *Corynebacterium mediolanum*, and *Heliobacterium chlorum* were used as substitute species, as indicated in Table 1. The scale bar represents 0.1 substitution per site. Bootstrap values were obtained after 500 replications and are shown when greater than 40%.

combined with amino acid sequences of sigma factors obtained from the National Center for Biotechnology Information databases, aligned, and phylogenetically analyzed. The resulting group 1 sigma factor tree was directly compared to a tree derived from SS-rRNA sequences, which was comprised of essentially the same species set that was represented in the sigma factor tree. This direct comparison to an established phylogenetic marker was intended to test the utility of sigma factors as markers for molecular systematics.

Phylogenetic analyses of sigma factor sequences were carried out for several reasons. First, the group 1 sigma factors perform similar if not identical functions and are essential for cell survival. As shown in Fig. 2 and by others (for reviews, see references 38 and 51), this functional conservation results in a high degree of conservation throughout eubacterial species. Present evidence suggests that only a single primary sigma factor is present in an organism. It was thus considered likely that primary sigma factors would be related by phylogeny and thus could be a potentially useful addition to molecular systematics. Additionally, alternative sigma factors (groups 2 and 3) from diverse organisms are believed to be related by function (51). Thus, a study including alternative sigma factors could propose possible functional groupings composed of sigma factors of different organisms.

In this work, only group 1 and group 2 $\sigma^{70}$-type sigma factors were studied. Group 1 sigma factors were analyzed to obtain phylogenetic relationships of the respective organisms, whereas group 1 and group 2 sequences combined were investigated to obtain additional information on the relationships among the sigma factors themselves. Group 3 sigma factors and members of the $\sigma^{70}$-type subfamily involved in the regulation of extracytoplastic functions, the so-called ECF subfamily (52), have not been included. The conditions used in Southern blotting experiments in this work to screen for sigma factors allowed for the detection of only group 1 and group 2 members (see Materials and Methods). Group 3 sigma factors are quite divergent from primary sigma factors and exhibit a maximum sequence identity of 27% (51) with group 1 members. Studies

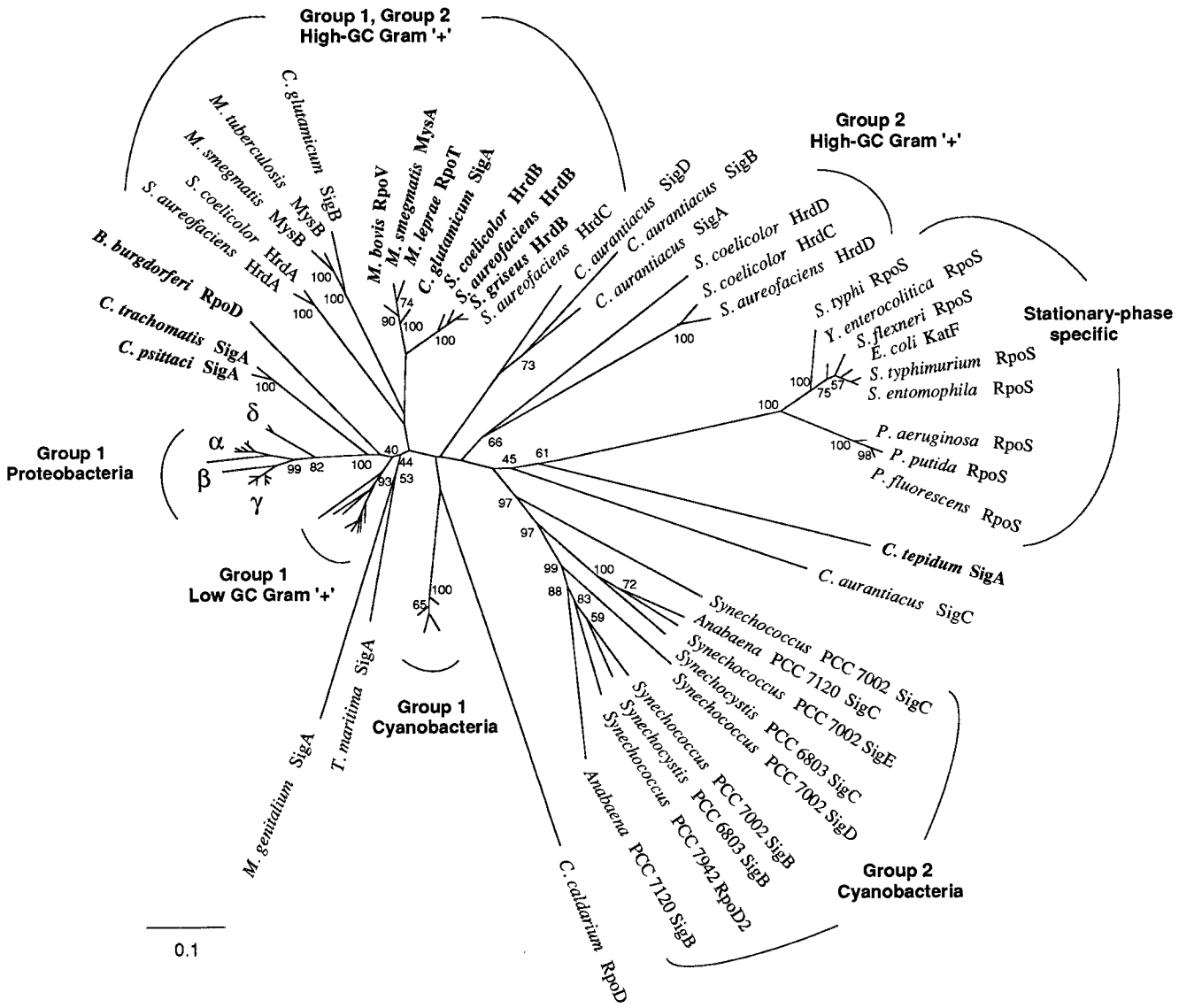


FIG. 5. Fitch-Margoliash tree for group 1 and group 2 sigma factor sequences. The tree was generated by the method of Fitch and Margoliash (26) from the total multiple sequence alignments depicted in Fig. 2. The distances were calculated by using the protdist program of the PHYLIP package, using a PAM matrix-based distance correction. The scale bar represents 0.1 substitution per site. Bootstrap values were obtained after 500 replications and are indicated when greater than 40%. For group 1 proteobacteria, cyanobacteria, and group 1 low-GC gram-positive bacteria, individual species and corresponding bootstrap values are not shown. These three clusters are virtually identical to the ones shown in Fig. 3. Group 1 sequences outside these three clusters are indicated by boldface print and different font.

of sigma factors which include members of group 3 are discussed by Lonetto et al. (51, 52).

The comparison of the phylogenetic tree obtained by using group 1 sigma factor sequences (Fig. 3) with the tree using SS-rRNA sequences (Fig. 4) shows that overall, the branching pattern and robustness, as deduced by the bootstrap values, are very similar. This finding strongly supports the idea that sigma factors can be useful markers for molecular systematics. Furthermore, the sigma factor tree generated by the parsimony method (data not shown) confirms this statement. The well-supported branches are practically identical in both branching patterns and bootstrap values to those for the tree generated by the distance method. Major differences between the two methods will be discussed below.

The proteobacterial phylum presently contains the largest number of species in the trees (Table 1). Four (α, β, γ, and δ) of the five subgroups of proteobacteria are included in the analyses performed in this study. With the sample size used, the four subgroups are supported with higher bootstrap values in the sigma factor tree than in the SS-rRNA tree. The branching order is practically identical between the two and is consistent with SS-rRNA trees which include a much larger number of sequences (62). Thus, it certainly appears that the conserved regions of the protein sequences of sigma factors can correctly portray evolutionary relationships between these closely related organisms.

The chlamydiae are grouped closely to the proteobacteria in the sigma factor tree, and this association is very highly supported (Fig. 3). This close relationship is also apparent in the tree generated by the parsimony method (data not shown). Based on both their sizes and sequence properties, the sigma factors of chlamydiae support their inclusion as a distinctive

subgroup of the proteobacteria. It will be interesting to examine the position of the chlamydiae once sigma factor sequences of the ε subgroup of proteobacteria become available. In studies including the ε subgroup, their members are placed as the deepest of the five proteobacterial subgroups (20). Chlamydiae are obligate intracellular parasites of eukaryotic cells that cause a number of diseases in mammals and birds (2). Chlamydiae possess some unique phenotypic characteristics within the eubacteria, such as a biphasic life cycle and an outer cell membrane that differs from the membranes of other gram-negative bacteria in that it is highly cross-linked by S-S bridges (3). Examination of studies of a number of different molecular markers reveals that the position of the chlamydiae within the eubacterial kingdom is not well defined. For example, trees employing elongation factor Tu as the phylogenetic marker suggest that the chlamydia are most closely related to the spirochetes (91, 16). The use of GroEL as the marker positions the chlamydiae between the bacteroides and the spirochetes (91). Studies with HSP70s place the chlamydiae adjacent to the gram-positive bacteria and cyanobacteria (5, 69), whereas a comparison of RecA sequences places the chlamydiae between the deinococcus-thermus group and the low-GC gram-positive cluster (20). A major problem in positioning the chlamydiae within the eubacteria appears to be the small number of organisms isolated and studied from this group. However, it should be noted that of all molecular markers used so far, sigma factor comparisons position the chlamydiae with the highest confidence.

The gram-positive phylum has been divided into four phylogenetically distinct groups, two of which, the low-GC and high-GC groups, are well characterized (95). There has been much debate over whether the gram-positive bacteria are monophyletic. For example, studies on some rRNA genes (95) and GroEL genes (91) indicated that the gram-positive bacteria are of monophyletic origin, whereas different studies on rRNA genes (89), studies of elongation factor TU and genes encoding the β subunit of ATP synthase (53), and comparisons of RecA sequences (20) suggested that they are not monophyletic. With the species represented in this work, the sigma factor tree shows that gram-positive bacteria are probably not of monophyletic origin, although the bootstrap values of the nodes defining the low-GC and high-GC clusters are low. The tree generated by the parsimony method (data not shown) indicates that the gram-positive bacteria could be of monophyletic origin, even though, as is the case in the tree based on the distance method, this association is only weakly supported.

*M. genitalium*, and *T. maritima* as its closest relative, are positioned between the two gram-positive clusters in the sigma factor tree. The closest relative of *M. genitalium* in the SS-rRNA tree are the spirochetes, but the position of this mycoplasma relative to other organisms is unsupported. The rapidly evolving nature of the mycoplasmas, as evidenced by long branch lengths, is also apparent in the sigma factor tree. Other studies using SS-rRNA data place the mycoplasmas within the low-GC gram-positive group (93, 95), even though their properties are very unusual. They have no cell walls, their genomes are far smaller than other bacterial genomes (28), and they have a number of other cytological and biochemical peculiarities (67). Some microbiologists argue that mycoplasmas constitute a phylogenetically distinct group that is distantly related to other eubacteria (67). It is hard to draw any definitive conclusions on that issue from the data of this work, considering that only one mycoplasma-derived sequence is included. However, it is interesting that the SigA protein of *M. genitalium* has a unique insertion of five residues which no other sigma factor proteins possess (Fig. 2). Both in the protein tree and in SS-rRNA tree, *M. genitalium* is placed outside the low-GC cluster. This is also the case in the protein tree generated by the parsimony method (data not shown). However, it is apparent from various studies that if a large number of SS-rRNA sequences are analyzed, the mycoplasmas are placed within the gram-positive group (20, 93). In a similar study comparing phylogenetic results of the RecA protein with SS-rRNA data (20), it was found that the SS-rRNA data placed the mycoplasmas within the gram-positive group whereas their position in the protein tree was not as obvious. Further sigma factor sequences will have to be included to position the mycoplasmas more accurately and to determine whether phylogenetic analyses of sigma factor protein sequences support the notion that mycoplasmas form a phylogenetically distinct group within the eubacteria.

Another group whose relationship to the gram-positive bacteria is unresolved is the cyanobacterial phylum. Analysis of a number of genes has suggested that cyanobacteria and gram-positive bacteria might be sister groups. Both protein sequences (20, 91) and SS-rRNA sequences (89, 95) have been used to support this interpretation. For example, there is one strong signature position in SS-rRNA (C residue at position 1207) which occurs only in gram-positive bacteria and cyanobacteria, not in any other eubacteria (96). Although cyanobacteria and gram-positive bacteria appear to be related to one another in a number of analyses, the difference is that SS-rRNA data show the cyanobacteria more closely related to the low-GC gram-positive bacteria, whereas certain protein-based analyses show a closer relationship between cyanobacteria and high-GC gram-positive bacteria (20). In this study, the bootstrap values of the gram-positive and cyanobacterial nodes are too low to allow us to draw strongly supported conclusions. Nevertheless, according to the sigma factor tree, cyanobacteria are more closely related to the high-GC than to the low-GC gram-positive bacteria. This result is similar to that derived from analyses of RecA sequences in which the cyanobacteria were also more closely associated with the high-GC gram positive bacteria (20).

It does not disturb the species composition and relative positioning of group 1 divisions significantly to include group 2 sequences in the phylogenetic analysis of group 1 sequences (Fig. 5). One exception to this statement is the position of the *B. burgdorferi* RpoD sequence, which is placed closer to the proteobacterial grouping when the group 2 sequences are included. Although this positioning is not strongly supported statistically, it is interesting to note that the large size of the sigma factor of this spirochete is more typical of sigma factors of proteobacteria, chlamydia, and some high-GC gram-positive bacteria. Moreover, with the exception of some group 1 sigma factors of the high-GC gram-positive bacteria and of *Chlorobium tepidum* (see below), no known group 1 proteins clustered tightly with group 2 proteins. These results suggest that all sigma factors designated as group 1 members have probably been correctly identified. No exact function has yet been assigned to any of the group 2 proteins. Cyanobacterial group 2 sigma factors form a tight cluster, but roles of these proteins have remained uncertain. Since multiple group 2 sigma factors are found in nonheterocystous cyanobacteria (82), these sigma factors are probably not involved in controlling this differentiation process. This suggestion was confirmed by mutational studies of Brahamsha and Haselkorn (6), who found that *sigB* and *sigC* mutants of *Anabaena* sp. strain PCC 7120 could still differentiate heterocysts and fix nitrogen. More recent studies by Tsinoremas et al. (87) have suggested that the SigB (RpoD2) protein of *Synechococcus* sp. strain PCC 7942 plays a role in controlling gene expression during circadian responses

of this organism. The SigE protein of *Synechococcus* sp. strain PCC 7002 may play a role similar to that of RpoS of proteobacteria in transcribing genes in post-exponential-phase cells (34a).

Some group 2 sigma factors of high-GC gram-positive organisms infiltrate the group 1 cluster for the high-GC gram-positive organisms, suggestive of recent gene duplication events, while some others are placed near other group 2 sequences. The significance of these differences is not presently known. The so-called stationary-phase sigma factors, which are members of the group 2 family, form a very well defined cluster. This is not surprising when one considers that all known sequences within this cluster are derived from proteobacteria. The heat shock cluster, composed of proteobacterial group 3 heat shock sigma factor sequences, branches off from this part of the tree (data not shown), in accordance with previous results (51). Other studies have shown that additional group 3 sequences radiate from the region of the heat shock sequences to form other functional clusters, including the flagellar clusters, sporulation clusters, and the ECF subfamily (51, 52).

The position of *Chlorobium tepidum* in all three trees is somewhat mystifying. It is placed differently in the group 1 sigma factor tree than in the SS-rRNA tree, although the bootstrap values obtained in either case are low. The branch length for this sigma factor in the group 1 tree is extremely long, a result that could indicate that this gene is a product of very rapid evolution or that the organism is at best only very distantly related to other eubacteria. The only confident placement of this sequence occurs when both group 1 and group 2 sigma factors sequences are included. The *Chlorobium tepidum* SigA sequence is the only group 1 sigma factor that is placed deep within the group 2 sequences, and the *Chlorobium tepidum* sigma factor is in fact most closely related to the stationary-phase sigma factors of proteobacteria. However, it is highly unlikely that the *Chlorobium tepidum* sequence is anything but the primary sigma factor of this organism, considering that no other sigma factor gene was identified in spite of considerable effort to do so (34a). This result could indicate that a lateral gene transfer event has taken place between members of these groups or could indicate that a dramatic change in function has occurred for this sigma factor in the green sulfur bacteria. The green sulfur bacteria are not a particularly well established phylogenetic group; for example, only a few SS-rRNA sequences have been cataloged (31). It will be interesting to see where *Chlorobium tepidum* is positioned within other trees, using different molecules as the phylogenetic markers. In hopes of further resolving the position of *Chlorobium tepidum*, the *recA* gene of this organism has recently been cloned and sequencing is in progress (19a).

Cyanobacteria, *C. aurantiacus*, and the high-GC gram-positive bacteria are the only organisms for which multiple group 2 sigma factors have been found (Fig. 5). This could serve as a further indicator of the relatedness of these groups. Although not very strongly supported from the analysis of group 1 sigma factors, it seems that the green sulfur bacteria and green gliding bacteria (perhaps better referred to as the anoxygenic flexibacteria or the filamentous anoxygenic phototrophic bacteria) (64) may be more closely related to the cyanobacteria than to any other grouping in the sigma factor tree. It is impossible to predict from the placement of the four *C. aurantiacus* sequences which represents the primary sigma factor of this organism. Based on the conservation of sequence in the vicinity of the critically important regions 2.4 and 4.2 (Fig. 2), the SigA protein might be the most likely candidate to be the group 1 sigma factor. However, it is also possible that the SigC sequence of *C. aurantiacus* is another potential group 1 sigma factor that is placed within the group 2 sequences (Fig. 5). Attempts to isolate the *recA* gene of this organism are in progress, and it will be interesting to see results obtained with this marker. It will also be important to obtain molecular data from other members of this rather poorly characterized bacterial group.

Conserved regions 2.4 and 4.2 (Fig. 2) have been shown by mutational analysis to contain residues that interact with the −10 and −35 motifs, respectively, of the promoters of both *E. coli* (78) and *Bacillus subtilis* (45). Recently, the determination of the crystal structure of a fragment of the $\sigma^{70}$ protein of *E. coli* has provided the structural framework within which to interpret these data further (56). For the group 1 sigma factors, and for many of the group 2 sigma factors as well, the residues known to interact with these motifs are absolutely conserved. The sole exception among the eubacteria sequenced to date is *Chlorobium tepidum*, which has a leucine rather than the conserved threonine residue in the −10 recognition region. Moreover, it is obvious from the alignment shown in Fig. 2 that the structural context (i.e., the likely distance separating these domains) between those regions of sigma factors that interact with the −10 and −35 motifs of the promoter are also conserved. The principal exception to this statement is the nucleus-encoded, chloroplast-targeted sigma factor of *Cyanidium caldarium*, which could possibly accommodate a slightly different spacing of these elements due to insertion of amino acids between regions 3 and 4. Nevertheless, it is clear that the data in Fig. 2 strongly support the notion that typical promoters in all eubacteria should contain sequence motifs and spacing similar to those of the consensus promoter sequence for the $\sigma^{70}$ sigma factor of *E. coli*. Since 5′ endpoints of mRNAs from diverse bacteria frequently do not map adjacent to sequences that resemble *E. coli* consensus promoters, it could be that endonucleolytic processing of primary transcripts is more prevalent in eubacteria than has yet been demonstrated experimentally.

In conclusion, the studies described here support the idea that group 1 $\sigma^{70}$ sigma factors are a useful marker for bacterial systematics. A direct comparison to SS-rRNA sequences, using essentially the same set of species, showed that the overall branching pattern and resolution of the two molecular markers are very similar. More confident placement of a number of phylogenetic groups should be possible as more sequences of sigma factors become available. This is a very probable scenario if the recent explosive increase in the number of available sigma factors sequences continues. Because group 1 sigma factor proteins contain regions that are practically invariant in all eubacterial species, and other regions that are variable even between close relatives, relationships between close and distant relatives can be resolved.

## ACKNOWLEDGMENTS

We thank Michael Adams (University of Georgia) and Wim Vermaas (Arizona State University) for providing chromosomal DNAs, Timothy Donohue (University of Wisconsin) for chromosomal DNA and identifying and supplying a cosmid encoding the *rpoD* gene of *R. sphaeroides*, Blair Hedges and Tracy Nixon (Penn State University) for their help with phylogenetic analyses, Vicki Stirewalt (Penn State University) for technical assistance, and Jeff Palmer (Indiana University) for helpful discussions.

This work was supported by grant GM-31625 from the National Institutes of Health and grant DE-FG02-94ER20137 from the Department of Energy to D.A.B.

## REFERENCES

1. **Balch, W. E., L. J. Magrum, G. E. Fox, R. S. Wolfe, and C. R. Woese.** 1977. An ancient divergence among the bacteria. J. Mol. Evol. **9:**305–311.

2. **Beatty, W. L., R. P. Morrison, and G. I. Byrne.** 1994. Persistent chlamydiae: from cell culture to a paradigm for chlamydial pathogenesis. Microbiol. Rev. **58:**686–699.
3. **Birkelund, S.** 1992. The molecular biology and diagnostics of *Chlamydia trachomatis*. Dan. Med. Bull. **39:**304–320.
4. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7:**1513–1523.
5. **Boorstein, W. R., T. Ziegelhoffer, and E. A. Craig.** 1994. Molecular evolution of the HSP70 multigene family. J. Mol. Evol. **38:**1–17.
6. **Brahamsha, B., and R. Haselkorn.** 1992. Identification of multiple RNA polymerase sigma factor homologs in the cyanobacterium *Anabaena* sp. strain PCC 7120: cloning, expression, and inactivation of the *sigB* and *sigC* genes. J. Bacteriol. **174:**7273–7282.
7. **Brahamsha, B., and R. Haselkorn.** 1991. Isolation and characterization of the gene encoding the principal sigma factor of the vegetative cell RNA polymerase from the cyanobacterium *Anabaena* sp. strain PCC 7120. J. Bacteriol. **173:**2442–2450.
8. **Bryant, D. A., and N. Tandeau de Marsac.** 1988. Isolation of genes encoding components of the photosynthetic apparatus. Methods Enzymol. **167:**755–765.
9. **Burgess, R. R., A. A. Travers, J. J. Dunn, and E. K. F. Bautz.** 1969. Factor stimulating transcription by RNA polymerase. Nature (London) **221:**43–46.
10. **Burton, Z., R. R. Burgess, J. Lin, D. Moore, S. Holder, and C. A. Gross.** 1981. The nucleotide sequence of the cloned *rpoD* gene for the RNA polymerase sigma subunit from *E. coli* K12. Nucleic Acids Res. **9:**2889–2903.
11. **Caslake, L. F.** 1994. Ph.D. thesis. The Pennsylvania State University, University Park, Pa.
12. **Caslake, L. F., and D. A. Bryant.** 1996. The *sigA* gene encoding the major σ factor of RNA polymerase from the marine cyanobacterium *Synechococcus* sp. PCC 7002: cloning and characterization. Microbiology **142:**347–357.
13. **Chang, B. Y., and R. H. Doi.** 1990. Overproduction, purification, and characterization of *Bacillus subtilis* RNA polymerase sigma A factor. J. Bacteriol. **172:**3257–3263.
14. **Collado-Vides, J., B. Magasanik, and J. Gralla.** 1991. Control site location and transcriptional regulation in *Escherichia coli*. Microbiol. Rev. **55:**371–394.
15. **Collins, D. M., R. P. Kawakami, G. W. de Lisle, L. Pascopella, B. R. Bloom, and W. R. Jacobs.** 1995. Mutation of the principal sigma factor causes loss of virulence of the *Mycobacterium tuberculosis* complex. Proc. Natl. Acad. Sci. USA **92:**8036–8040.
16. **Delwiche, C. F., M. Kuhsel, and J. D. Palmer.** 1995. Phylogenetic analysis of *tufA* sequences indicates a cyanobacterial origin of all plastids. Mol. Phylogenet. Evol. **4:**110–128.
17. **Doolittle, R. F., D.-F. Feng, S. Tsang, G. Cho, and E. Little.** 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science **271:**470–477.
18. **Douglas, A. L., N. K. Saxena, and T. P. Hatch.** 1994. Enhancement of in vitro transcription by addition of cloned, overexpressed major sigma factor of *Chlamydia psittaci* 6BC. J. Bacteriol. **176:**3033–3039.
19. **Doukhan, L., M. Predich, G. Nair, O. Dussurget, I. Mandic-Mulec, S. T. Cole, D. R. Smith, and I. Smith.** 1995. Genomic characterization of the mycobacterial sigma gene cluster. Gene **165:**67–70.

19a. **Eisen, J., T. Gruber, and D. A. Bryant.** Unpublished data.

20. **Eisen, J. A.** 1995. The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecA's and 16S rRNAs from the same species. J. Mol. Evol. **41:**1105–1123.
21. **Engel, J. N., and D. Ganem.** 1990. A polymerase chain reaction-based approach to cloning sigma factors from eubacteria and its application to the isolation of a sigma-70 homolog from *Chlamydia trachomatis*. J. Bacteriol. **172:**2447–2455.
22. **Erickson, B. D., Z. F. Burton, K. K. Watanabe, and R. R. Burgess.** 1985. Nucleotide sequence of the *rpsU-dnaG-rpoD* operon from *Salmonella typhimurium* and a comparison of this sequence with the homologous operon of *Escherichia coli*. Gene **40:**67–78.
23. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:**783–791.
24. **Felsenstein, J.** 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17:**368–376.
25. **Felsenstein, J.** 1989. PHYLIP—phylogeny inference package (version 3.2). Cladistics **5:**164–166.
26. **Fitch, W. M., and E. Margoliash.** 1967. Construction of phylogenetic trees. Science **155:**279–284.
27. **Fox, G. E., E. Stackebrandt, R. B. Hespell, J. Gibson, J. Maniloff, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. S. Tanner, L. J. Magrum, L. B. Zablen, R. Blakemore, R. Gupta, L. Bonen, B. J. Lewis, D. A. Stahl, K. H. Leuhrsen, K. N. Chen, and C. R. Woese.** 1980. The phylogeny of prokaryotes. Science **209:**457–463.
28. **Fraser, C. M., J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. Sutton, J. M. Kelley, J. L. Fritchman, J. F. Weidman, K. V. Small, M. Sandusky, J. L. Fuhrmann, D. T. Nguyen, T. R. Utterback, D. M. Saudek, C. A. Phillips, J. M. Merrick, J.-F. Tomb, B. A. Dougherty, K. F. Bott, P.-C. Hu, T. S. Lucier, S. N. Peterson, H. O. Smith, C. A. Hutchison, and J. C. Venter.** 1995. The minimal gene complement of *Mycoplasma genitalium*. Science **270:**397–403.
29. **Fujita, M., Y. Hanaura, and A. Amemura.** 1995. Analysis of the *rpoD* gene encoding the principal sigma factor of *Pseudomonas putida*. Gene **167:**93–98.
30. **Gansel, X., A. Hartke, P. Boutibonnes, and Y. Auffray.** 1993. Nucleotide sequence of the *Lactococcus lactis* NCDO 763 (ML3) *rpoD* gene. Biochim. Biophys. Acta **1216:**115–118.
31. **Gibson, J., W. Ludwig, E. Stackebrandt, and C. R. Woese.** 1985. The phylogeny of green photosynthetic bacteria: absence of a close relationship between *Chlorobium* and *Chloroflexus*. Syst. Appl. Microbiol. **6:**152–156.
32. **Giovannoni, S. G., S. Turner, G. J. Olsen, S. Barns, D. J. Lane, and N. R. Pace.** 1988. Evolutionary relationships among cyanobacteria and green chloroplasts. J. Bacteriol. **170:**3584–3592.
33. **Gitt, M. A., L.-F. Wang, and R. H. Doi.** 1985. A strong sequence homology exists between the major RNA polymerase sigma factors of *Bacillus subtilis* and *Escherichia coli*. J. Biol. Chem. **260:**7178–7185.
34. **Grossman, A. D., J. W. Erickson, and C. A. Gross.** 1984. The *htpR* gene product of *E. coli* is a sigma factor for heat-shock promoters. Cell **38:**383–390.

34a. **Gruber, T., and D. A. Bryant.** Unpublished data.

35. **Gupta, R. S., and G. B. Golding.** 1993. Evolution of HSP70 gene and its implications regarding relationship between archaebacteria, eubacteria and eukaryotes. J. Mol. Evol. **37:**573–582.
36. **Haldenwang, W. G.** 1995. The sigma factors of *Bacillus subtilis*. Microbiol. Rev. **59:**1–30.
37. **Hasegawa, M., and T. Hashimoto.** 1993. Ribosomal RNA trees misleading? Nature (London) **361:**23.
38. **Helmann, J. D., and M. J. Chamberlin.** 1988. Structure and function of bacterial sigma factors. Annu. Rev. Biochem. **57:**839–872.
39. **Henikoff, S.** 1993. Sequence analysis by electronic mail server. Trends Biochem. Sci. **18:**267–268.
40. **Inouye, S.** 1990. Cloning and DNA sequence of the gene coding for the major sigma factor from *Myxococcus xanthus*. J. Bacteriol. **172:**80–85.
41. **Iordanescu, S. M.** 1989. *Staphylococcus aureus* chromosomal mutation plaC1 amplifies plasmid pT181 by depressing synthesis of its negative effector. J. Bacteriol. **171:**4831–4835.
42. **Iriarte, M., I. Stainier, and G. R. Cornelis.** 1995. The *rpoS* gene from *Yersinia enterocolitica* and its influence on expression of virulence factors. Infect. Immun. **63:**1840–1847.
43. **Kaneko, T., A. Tanaka, S. Sato, H. Kotani, T. Sazuka, N. Miyajima, M. Sugiura, and S. Tabata.** 1995. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome. DNA Res. **2:**153–166.
44. **Karlin, S., G. M. Weinstock, and V. Brendel.** 1995. Bacterial classifications derived from RecA protein sequence comparisons. J. Bacteriol. **177:**6881–6893.
45. **Kenney, T. J., and C. P. J. Moran.** 1991. Genetic evidence for interaction of $\sigma^a$ with two promoters in *Bacillus subtilis*. J. Bacteriol. **173:**3282–3290.
46. **Klenk, H.-P., and W. Zillig.** 1994. DNA-dependent RNA polymerase subunit b as a tool for phylogenetic reconstructions: branching topology of the archaeal domain. J. Mol. Evol. **38:**420–432.
47. **Kormanec, J., M. Farkasovsky, and L. Potuckova.** 1992. Four genes in *Streptomyces aureofaciens* containing a domain characteristic of principal sigma factors. Gene **122:**63–70.
48. **Kustu, S., E. Santero, J. Keener, D. Popham, and D. Weiss.** 1989. Expression of $\sigma^{54}$ (*ntrA*)-dependent genes is probably united by a common mechanism. Microbiol. Rev. **53:**367–376.
49. **Lai, C.-Y., and P. Baumann.** 1992. Sequence analysis of a DNA fragment from *Buchnera aphidicola* (an endosymbiont of aphids) containing genes homologous to *dnaG*, *rpoD*, *cysE*, and *secB*. Gene **119:**113–118.
50. **Liu, B., and R. F. Troxler.** 1996. Molecular characterization of a positively photoregulated nuclear gene for a chloroplast RNA polymerase sigma factor in *Cyanidium caldarium*. Proc. Natl. Acad. Sci. USA **93:**3313–3320.
51. **Lonetto, M., M. Gribskov, and C. A. Gross.** 1992. The $\sigma^{70}$ family: sequence conservation and evolutionary relationships. J. Bacteriol. **174:**3843–3849.
52. **Lonetto, M. A., K. L. Brown, K. E. Rudd, and M. J. Buttner.** 1994. Analysis of the *Streptomyces coelicolor sigE* gene reveals the existence of a subfamily of eubacterial RNA polymerase σ factors involved in the regulation of extracytoplasmic functions. Proc. Natl. Acad. Sci. USA **91:**7573–7577.
53. **Ludwig, W., J. Neumaier, N. Klugbauer, E. Brockmann, C. Roller, S. Jilg, K. Reetz, I. Schachtner, A. Ludvigsen, M. Bachleitner, U. Fischer, and K. H. Schleifer.** 1994. Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor TU and ATP-synthase beta-subunit genes. Antonie Leeuwenhoek. **64:**285–305.
54. **Maidak, B. L., N. Larsen, M. J. McCaughey, R. Overbeek, G. J. Olsen, K. Fogel, J. Blandy, and C. R. Woese.** 1994. The ribosomal database project. Nucleic Acids Res. **22:**3485–3487.
55. **Malakooti, J., S. P. Wang, and B. Ely.** 1995. A consensus promoter sequence for *Caulobacter crescentus* genes involved in biosynthetic and housekeeping functions. J. Bacteriol. **177:**4372–4376.
56. **Malhotra, A., E. Severinova, and S. Darst.** 1996. Crystal structure of a $\sigma^{70}$

subunit fragment from *E. coli* RNA polymerase. Cell **87:**127–136.
57. **Marcos, A. T., S. Gutierrez, B. Diez, F. J. Fernandez, J. A. Oguiza, and J. F. Martin.** 1995. Three genes *hrdB*, *hrdD* and *hrdT* of *Streptomyces griseus* IMRU 3570, encoding sigma factor-like proteins, are differentially expressed under specific nutritional conditions. Gene **153:**41–48.
58. **Metzger, R., D. P. Brown, P. Grealish, M. J. Staver, J. Versalovic, J. R. Lupski, and L. Katz.** 1994. Characterization of the macromolecular synthesis (MMS) operon from *Listeria monocytogenes*. Gene **151:**161–166.
59. **Mulvey, M. R., and P. C. Loewen.** 1989. Nucleotide sequence of *katF* of *Escherichia coli* suggests KatF protein is a novel sigma transcription factor. Nucleic Acids Res. **17:**9979–9991.
60. **Nagashima, K. V. P., S. Hanada, A. Hiraishi, K. Shimada, and K. Matsuura.** 1995. Phylogenetic analysis of photosynthetic reaction centers of purple bacteria and green filamentous bacteria, p. 975–978. *In* P. Mathis (ed.), Photosynthesis: from light to biosphere, vol. 1. Kluwer Academic Publishers, Dordrecht, The Netherlands.
61. **Oguiza, J. A., A. T. Marcos, M. Malumbres, and J. F. Martin.** 1996. Multiple sigma factor genes in *Brevibacterium lactofermentum*: characterization of *sigA* and *sigB*. J. Bacteriol. **178:**550–553.
62. **Olsen, G. J., C. R. Woese, and R. Overbeek.** 1994. The winds of (evolutionary) change: breathing new life into microbiology. J. Bacteriol. **176:**1–6.
63. **Pace, N. R., G. J. Olsen, and C. R. Woese.** 1986. Ribosomal RNA phylogeny and the primary lines of evolutionary descent. Cell **45:**325–326.
64. **Pierson, B. K., and R. W. Castenholz.** 1995. Taxonomy and physiology of filamentous anoxygenic phototrophs, p. 32–47. *In* R. E. Blankenship, M. T. Madigan, and C. E. Bauer (ed.), Anoxygenic photosynthetic bacteria, vol. 2. Kluwer Academic Publishers, Dordrecht, The Netherlands.
65. **Predich, M., L. Doukhan, G. Nair, and I. Smith.** 1995. Characterization of RNA polymerase and two sigma-factor genes from *Mycobacterium smegmatis*. Mol. Microbiol. **15:**355–366.
66. **Prince, R. W., Y. Xu, S. J. Libby, and F. C. Fang.** 1994. Cloning and sequencing of the gene encoding the RpoS (KatF) sigma factor from *Salmonella typhimurium* 14028s. Biochim. Biophys. Acta **1219:**198–200.
67. **Razin, S.** 1985. Molecular biology and genetics of mycoplasmas (*Mollicutes*). Microbiol. Rev. **49:**419–455.
68. **Razin, S.** 1978. The mycoplasmas. Microbiol. Rev. **42:**414–470.
69. **Rensing, S. A., and U. G. Maier.** 1994. Phylogenetic analysis of the stress-70 protein family. J. Mol. Evol. **39:**80–86.
70. **Rothschild, L. J., M. A. Ragan, A. W. Coleman, P. Heywood, and S. A. Gerbi.** 1986. Are rRNA sequence comparisons the Rosetta stone of phylogenetics? Cell **47:**640.
71. **Rushing, B. G., and S. R. Long.** 1995. Cloning and characterization of the *sigA* gene encoding the major sigma subunit of *Rhizobium meliloti*. J. Bacteriol. **177:**6952–6957.
72. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
73. **Sarniguet, A., J. Kraus, M. D. Henkels, A. M. Muehlchen, and J. E. Loper.** 1995. The sigma factor affects antibiotic production and biological control activity of *Pseudomonas fluorescens* Pf-5. Proc. Natl. Acad. Sci. USA **92:** 12255–12259.
74. **Sauer, U., A. Treuner, M. Buchholz, J. D. Santangelo, and P. Durre.** 1994. Sporulation and primary sigma factor homologous genes in *Clostridium acetobutylicum*. J. Bacteriol. **176:**6572–6582.
75. **Schnider, U., C. Keel, C. Blumer, J. Troxler, G. Defago, and D. Haas.** 1995. Amplification of the housekeeping sigma factor in *Pseudomonas fluorescens* CHA0 enhances antibiotic production and improves biocontrol abilities. J. Bacteriol. **177:**5387–5392.
76. **Schyns, G., A. Sobczyk, N. Tandeau de Marsac, and J. Houmard.** 1994. Specific initiation at a cyanobacterial promoter with RNA polymerase purified from *Calothrix* sp. PCC 7601. Mol. Microbiol. **13:**887–896.
77. **Segal, G., and E. Z. Ron.** 1993. Cloning, sequencing, and transcriptional analysis of the gene coding for the vegetative sigma factor of *Agrobacterium tumefaciens*. J. Bacteriol. **175:**3026–3030.
78. **Siegele, D. A., J. C. Hu, W. A. Walter, and C. A. Gross.** 1989. Altered promoter recognition by mutant forms of the $\sigma^{70}$ subunit of *Escherichia coli* RNA polymerase. J. Mol. Biol. **206:**591–603.
79. **Skladny, H., M. Heidelbach, and H.-U. Schairer.** 1994. Cloning and characterization of the gene coding for the major sigma factor of *Stigmatella aurantiaca*. Gene **143:**123–127.
80. **Strickland, M. S., N. E. Thompson, and R. R. Burgess.** 1988. Structure and function of the sigma-70 subunit of *Escherichia coli* RNA polymerase. Monoclonal antibodies: localization of epitopes by peptide mapping and effects on transcription. Biochemistry **27:**5755–5762.
81. **Takeshita, S., M. Sato, M. Toba, W. Masahashi, and T. Hashimoto-Gotoh.** 1987. High-copy-number and low-copy-number plasmid vectors for lacZ α-complementation and chloramphenicol- or kanamycin-resistance selection. Gene **61:**63–74.
82. **Tanaka, K., S. Masuda, and H. Takahashi.** 1992. Multiple *rpoD*-related genes of cyanobacteria. Biochim. Biophys. Acta **1132:**94–96.
83. **Tanaka, K., T. Shiina, and H. Takahashi.** 1991. Nucleotide sequence of genes *hrdA*, *hrdC*, and *hrdD* from *Streptomyces coelicolor* A3(2) having similarity to *rpoD* genes. Mol. Gen. Genet. **229:**334–340.
84. **Tanaka, K., and H. Takahashi.** 1991. Cloning and analysis of the gene (*rpoDA*) for the principal sigma factor of *Pseudomonas aeruginosa*. Biochim. Biophys. Acta **1089:**113–119.
85. **Tatusov, R. L., A. R. Mushegian, P. Bork, N. P. Brown, W. S. Hayes, M. Borodovsky, K. E. Rudd, and E. V. Koonin.** 1996. Metabolism and evolution of *Haemophilus influenzae* deduced from a whole genome comparison with *Escherichia coli*. Curr. Biol. **6:**279–291.
86. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
87. **Tsinoremas, N. F., M. Ishiura, T. Kondo, C. R. Andersson, K. Tanaka, H. Takahashi, C. H. Johnson, and S. S. Golden.** 1996. A sigma factor that modifies the circadian expression of a subset of genes in cyanobacteria. EMBO J. **15:**2488–2495.
88. **van den Bergh, E. R. E., C. Baker, R. J. Raggers, P. Terpstra, E. C. Woudstra, L. Dijkhuizen, and W. G. Meijer.** 1996. Primary structure and phylogeny of the Calvin cycle enzymes transketolase and fructosebisphosphate aldolase of *Xanthobacter flavus*. J. Bacteriol. **178:**888–893.
89. **van der Peer, Y., J. M. Neefs, P. De Rijk, P. De Vos, and R. De Wachter.** 1994. About the order of divergence of the major bacterial taxa during evolution. Syst. Appl. Microbiol. **17:**32–38.
90. **Vawter, L., and W. M. Brown.** 1993. Rates and patterns of base change in the small subunit ribosomal RNA gene. Genetics **134:**597–608.
91. **Viale, A. M., A. K. Arakaki, F. C. Soncini, and R. G. Ferreyra.** 1994. Evolutionary relationships among eubacterial groups as inferred from GroEL (chaperonin) sequence comparisons. Int. J. Syst. Bacteriol. **44:**527–533.
92. **Waldburger, C., T. Gardella, R. Wong, and M. M. Susskind.** 1990. Changes involved in conserved region 2 of *Escherichia coli* $\sigma^{70}$ affecting promoter recognition. J. Mol. Biol. **215:**267–276.
93. **Weisburg, W. G., J. G. Tully, D. L. Rose, J. P. Petzel, H. Oyaizu, D. Yang, L. Mandelco, J. Sechrest, T. G. Lawrence, J. van Etten, J. Maniloff, and C. R. Woese.** 1989. A phylogenetic analysis of the mycoplasmas: basis for their classification. J. Bacteriol. **171:**6455–6467.
94. **Wilmotte, A.** 1994. Molecular evolution and taxonomy of the cyanobacteria, p. 1–25. *In* D. A. Bryant (ed.), The molecular biology of cyanobacteria, vol. 1. Kluwer Academic Publishers, Dordrecht, The Netherlands.
95. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51:**221–271.
96. **Woese, C. R., E. Stackebrandt, T. J. Macke, and G. E. Fox.** 1985. A phylogenetic definition of the major eubacterial taxa. Syst. Appl. Microbiol. **6:**143–151.
97. **Yanish-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.
98. **Zuckerkandl, E., and L. Pauling.** 1965. Molecules as documents of evolutionary history. J. Theor. Biol. **8:**357–366.