An Inventory of Genes Encoding RNA Polymerase Sigma Factors in 31 Completely Sequenced Eubacterial Genomes

Gerhard Mittenhuber*
Institut für Mikrobiologie und Molekularbiologie, Ernst-Moritz-Arndt-Universität, F.-L. Jahnstr. 15, D-17487 Greifswald, Germany

Abstract
Sigma factors are important elements involved in transcriptional regulation of gene expression by conferring promoter specificity to RNA polymerase. The number of sigma factor encoding genes in 31 completely sequenced bacterial genomes were compared. Two unrelated families of sigma factors, the $\sigma^{70}$- and the $\sigma^{54}$-family were identified previously. The $\sigma^{70}$-family can be further subdivided into two distantly related groups: the $\sigma^{70}$ subfamily and the poorly characterized ECF subfamily. A total of 215 sigma factors could be attributed to these subfamilies. The construction of phylogenetic trees allows subclassifications of sigma factor encoding genes within these subfamilies. With the exception of Deinococcus radiodurans, all species possess a housekeeping primary sigma factor. Free-living species possess a higher number of both $\sigma^{70}$-type and ECF alternative sigma factors than pathogens or symbionts associated with animals. Different bacterial species exhibit large differences in the number of alternative sigma factor encoding genes and consequently huge flexibility in their transcriptional regulatory patterns. Transcriptional regulation in terms of regulons controlled by alternative sigma factors is a late evolving phenomenon. The current nomenclature for sigma factor encoding genes is confusing and should be revised.

Introduction
In eubacteria, different sigma factors exhibiting different promoter specificities provide a convenient way for differential transcriptional regulation. Sigma factors and genes under control of a specific sigma factor (the regulon) have been primarily investigated in the gram negative model organism Escherichia coli and the gram positive model organism Bacillus subtilis. The primary sigma factor (housekeeping sigma factor; $\sigma^{70}$) of E. coli was initially discovered as a control element of bacteriophage T4 (Bautz et al., 1969). In 1981, the gene encoding this sigma factor ($rpoD$) was sequenced by Burton et al. (1981). The functionally equivalent B. subtilis counterpart of $\sigma^{70}$, named $\sigma^{A}$ has a lower molecular weight (43 KDa) but is very similar to RpoD (Gitt et al., 1985). With the exception of Deinococcus radiodurans (see below), the gene encoding the major sigma factor RpoD/SigA is present in every completely sequenced eubacterial genome.

The first alternative sigma factor of E. coli discovered was the heat shock regulatory gene product HtpR (RpoH) of controlling the heat shock regulon of this organism (Grossman et al., 1984). Other minor sigma factors of E. coli include the stationary phase sigma factor RpoS (KatF) (Muley and Loewen, 1989; Lange and Hengge-Aronis, 1991; McCann et al., 1991; Tanaka et al., 1993) and a sigma factor required for flagella biosynthesis named FliA (Liu and Matsumoto, 1995).

Research on sigma factors in B. subtilis was driven by an imaginative hypothesis of Losick and Pero (1981) who postulated that minor sigma factors might play an important role in the cellular differentiation process of endospore formation. A nomenclature for sporulation sigma factors was established (Losick et al., 1986) and it could be demonstrated finally that a cascade of five sigma factors ($\sigma^{H}$, $\sigma^{F}$, $\sigma^{E}$, $\sigma^{D}$, and $\sigma^{A}$) operates in a coordinated fashion during the sporulation process (Losick and Stragier, 1992). However, the first alternative sigma factor discovered in B. subtilis, $\sigma^{37}$ (Haldenwang and Losick, 1980) which was renamed $\sigma^{B}$ (Losick et al., 1986) does not participate in the sporulation cascade. This sigma factor is involved in control of the general stress response of B. subtilis (for reviews see Hecker and Vöker, 1998, 2001). B. subtilis also possesses a FliA ortholog, which is named $\sigma^{D}$ in this organism (Marquez et al., 1990). The RpoD/SigA, RpoH, RpoS, FliA/SigD proteins and the five sporulation sigma factors constitute the $\sigma^{70}$ subfamily of sigma factors (Lonetto et al., 1992; Wösten, 1998).

Lonetto et al. (1992) defined three groups and Wösten (1998) provides an update of this classification: Group 1 constitute the primary sigma factors RpoD/SigA which are essential for cell growth. Group 2 and group 3 sigma factors are nonessential for growth and are also called alternative sigma factors. Sigma factors in Group 2 are similar in sequence to group 1 sigma factors. In group 2, RpoS-type sigma factors, sigma factors from high CG gram positive bacteria and cyanobacterial sigma factors are included. Group 3 consists of heat shock sigma factors (RpoH), flagella sigma factors and sporulation sigma factors.

*For correspondence. Email Gerhard.Mittenhuber@biologie.uni-greifswald.de; Tel. +49-3834-864218; Fax. +49-3834-864202.
Further Reading

Caister Academic Press is a leading academic publisher of advanced texts in microbiology, molecular biology and medical research. Full details of all our publications at caister.com

• MALDI-TOF Mass Spectrometry in Microbiology
  Edited by: M Kostrzewa, S Schubert (2016)
  www.caister.com/malditof

• Aspergillus and Penicillium in the Post-genomic Era
  Edited by: RP Vries, IB Gelber, MR Andersen (2016)
  www.caister.com/aspergillus2

• The Bacteriocins: Current Knowledge and Future Prospects
  Edited by: RL Dorfl, SM Roy, MA Rilley (2016)
  www.caister.com/bacteriocins

• Omics in Plant Disease Resistance
  Edited by: V Bhadauria (2016)
  www.caister.com/opdr

• Acidophiles: Life in Extremely Acidic Environments
  Edited by: R Quatrini, DB Johnson (2016)
  www.caister.com/acidophiles

• Climate Change and Microbial Ecology: Current Research and Future Trends
  Edited by: J Marxsen (2016)
  www.caister.com/climate

• Biofilms in Bioremediation: Current Research and Emerging Technologies
  Edited by: G Lear (2016)
  www.caister.com/biorem

• Microalgae: Current Research and Applications
  Edited by: MN Tsaioglou (2016)
  www.caister.com/microalgae

• Gas Plasma Sterilization in Microbiology: Theory, Applications, Pitfalls and New Perspectives
  Edited by: H Shintani, A Sakudo (2016)
  www.caister.com/gasplasma

• Virus Evolution: Current Research and Future Directions
  Edited by: SC Weaver, M Denison, M Roossinck, et al. (2016)
  www.caister.com/virus

• Arboviruses: Molecular Biology, Evolution and Control
  Edited by: N Vasilakis, DJ Gubler (2016)
  www.caister.com/arbo

• Shigella: Molecular and Cellular Biology
  Edited by: WD Picking, WL Picking (2016)
  www.caister.com/shigella

• Aquatic Biofilms: Ecology, Water Quality and Wastewater Treatment
  Edited by: AM Romani, H Guasch, MD Balaguer (2016)
  www.caister.com/aquaticbiofilms

• Alphaviruses: Current Biology
  Edited by: S Mahalingam, L Herrero, B Herring (2016)
  www.caister.com/alpha

• Thermophilic Microorganisms
  Edited by: F Li (2015)
  www.caister.com/thermophile

• Flow Cytometry in Microbiology: Technology and Applications
  Edited by: MG Wilkinson (2015)
  www.caister.com/flow

• Probiotics and Prebiotics: Current Research and Future Trends
  Edited by: K Venema, AP Carmo (2015)
  www.caister.com/probiotics

• Epigenetics: Current Research and Emerging Trends
  Edited by: BP Chadwick (2015)
  www.caister.com/epigenetics

• Corynebacterium glutamicum: From Systems Biology to Biotechnological Applications
  Edited by: A Burkovski (2015)
  www.caister.com/cory

• Advanced Vaccine Research Methods for the Decade of Vaccines
  Edited by: F Bagnoli, R Rappuoli (2015)
  www.caister.com/vaccines

• Antifungals: From Genomics to Resistance and the Development of Novel Agents
  Edited by: AT Coste, P Vandeputte (2015)
  www.caister.com/antifungals

• Bacteria-Plant Interactions: Advanced Research and Future Trends
  www.caister.com/bacteria-plant

• Aeromonas
  Edited by: J Graf (2015)
  www.caister.com/aeromonas

• Antibiotics: Current Innovations and Future Trends
  Edited by: S Sánchez, AL Demain (2015)
  www.caister.com/antibiotics

• Leishmania: Current Biology and Control
  Edited by: S Adak, R Dafta (2015)
  www.caister.com/leish

• Acanthamoeba: Biology and Pathogenesis (2nd edition)
  Author: NA Khan (2015)
  www.caister.com/acanthamoeba

• Microarrays: Current Technology, Innovations and Applications
  Edited by: Z He (2014)
  www.caister.com/microarrays

• Metagenomics of the Microbial Nitrogen Cycle: Theory, Methods and Applications
  Edited by: D Marco (2014)
  www.caister.com/n2

Order from caister.com/order
The existence of a distinctly related group 3 subfamily of sigma factors became evident, when a careful analysis of the sigE gene of *Streptomyces coelicolor* was performed (Lonetto et al., 1994). Common features of this subfamily include regulation of extracytoplasmic functions and functions as effector molecules responding to extracytoplasmic stimuli (Missiakas and Raina, 1998). This group was therefore named the ECF (extra cytoplasmic function) family. Members of the ECF family of *E. coli* are a second heat shock sigma factor (RpoE, Rouvière et al., 1995) and Fecl involved in transcriptional regulation of ferric citrate transport (Angerer et al., 1995). *B. subtilis* possesses seven ECF sigma factors (σ^M_, σ^V_, σ^W_, σ^X_, σ^Y_, σ^Z_, YiaC) (Kunst et al., 1997). The functions of the ECF sigma factors in *B. subtilis* are just being elucidated (Horsburgh and Moir, 1999; Huang and Helmann, 1998; Huang et al., 1998).

The σ^Z0_-family is functionally and structurally distinct from the σ^Z0_-family (for recent reviews, see Buck et al., 2000, and Studholme and Buck, 2000). No sequence similarity can be detected between the σ^Z0_- and the σ^Z4_-family. The first sigma factor belonging to the σ^Z4_-family discovered was σ^Z4_ (RpoN). It is specifically required for the transcription of nitrogen-regulated and of nitrogen-fixation promoters (Hunt and Magasanik, 1985; Reitzer and Magasanik, 1986) in *E. coli*. The RpoN equivalent of *B. subtilis* is called SigL (Débarbouillé et al., 1991).

With the advent of completely sequenced bacterial genomes it becomes possible to perform comparative genomic studies. This emerging discipline is a powerful tool for a variety of applications. For example, potential virulence genes can be identified by comparing the genomes of a virulent and an avirulent strain of the same species. In this paper, we compare the genes encoding sigma factors in completely sequenced bacterial genomes by a phylogenetic analysis. We examine whether a relationship between the number of sigma factors in the genome of a given organism and the habitat of the bacterium exists and whether a relationship between the number of sigma factors in the genome of a given organism and its genome size and/or the number of genes in the genome can be established.

**Results**

A total of 215 amino acid sequences from genes encoding sigma factors from 31 species were obtained from the NCBI server (supplementary Table 1 at http://www.jmmb.net)\(^*\). For some species (e.g. *E. coli*, *Helicobacter pylori*) the complete sequences of two strains are available. In such cases, only one genome was examined for the sigma factor content. The individual sigma factors were classified as either σ^Z0_-type, ECF, or σ^Z4_-type sigma factors (Table 2).

Table 2 also shows that a correlation between the number of ECF sigma factor encoding genes and the habitat of the species possessing these sigma factors may be deduced. Pathogenic species tend to have no or only few ECF sigma factors. The versatile species *Pseudomonas aeruginosa* possesses the highest number of ECF sigma factors (19), followed by the nitrogen fixing symbiont *Mesorhizobium loti* (18) and the stalked bacterium *Caulobacter crescentus* (14).

Information given in Table 2 also tries to determine whether a correlation exists between genome size and the number of sigma factors in the genome. For this purpose, two quotients were calculated. The first quotient is defined as “genome size (in Mb)” divided by “total number” of sigma factors. The second quotient is defined as “genome size (in Mb)” divided by “total number of genes” divided by 100. The basis for this idea is simple: Obligately pathogenic species are associated with their host and are used to their constant environment. In contrast, free-living species and those who are associated with hosts but are also able to survive in the environment (e.g. *E. coli*, *Vibrio* species) are faced with frequent changes of the environmental conditions. Since it is essential for the bacterial cell to adjust its metabolism to changing environmental conditions, it seems reasonable to speculate that a species adapted to a constant environment has a smaller repertoire of regulatory mechanisms than a species which is forced to cope with many changing environmental conditions. Quotients like “genome size/number of sigma factors” and “total gene number/number of sigma factors” might provide an estimate for transcriptional complexity: Free-living species tend to have lower values of these quotients than those associated with animals (pathogens).

The distribution pattern of the σ^Z0_-family members (RpoD/SigA, RpoS, RpoH, FliA/SigD, SigB, sporulation specific sigma factors and others with poorly characterized functions) within the 31 species is shown in Table 3. It is noteworthy that *Mycobacterium* species possess two *rpoD/sigA* genes and that the cyanobacterium *Synechocystis* sp. PCC 6803 has five paralogs. Also *M. loti* possesses two RpoN and RpoH paralogs. The occurrence of the heat shock sigma factor RpoH of *E. coli* is restricted to the proteobacteria. The FliA/SigD sigma factors are not restricted to motile bacteria; for example, *C. crescentus*, which does not possess a *fliA* gene, is a motile bacterium. Two general stress sigma factors (RpoS and SigB) show a mutually exclusive distribution pattern. The occurrence of sporulation specific sigma factors is restricted to spore-forming bacteria.

An unrooted phylogenetic tree based on an alignment of whole amino acid sequences of the σ^Z0_-family is shown in Figure 1, a bootstrapped tree based on an alignment of the conserved C-terminal regions is shown in Figure 2. As observed previously (Lonetto et al., 1992; Gruber and Bryant, 1997; Wösten, 1998), members of the individual families form distinct clusters. SigI of *B. subtilis* (Zuber et al., 2001) is related to the putative sigma factor SigI of *C. acetobutylicum*, the only other species possessing this sigma factor. Similarly, the putative competence

\(^*\) Table 1 can be viewed at http://Jmmb.net/supplementary.
Table 2. Distribution pattern of sigma factors belonging to the s\(^{70}\)-family, the ECF subfamily and the s\(^{54}\)-family within 31 bacterial species. Additionally, information on the phylogenetic position, habitat and the genome size of each species is given. Two quotients “genome size divided by number of sigma factors in genome” (quotient 1) and “total number of genes divided by number of sigma factors in genome divided by 100” (quotient 2) were calculated.

<table>
<thead>
<tr>
<th>Name</th>
<th>Abbreviation</th>
<th>Phylogenetic position</th>
<th>Habitat and features</th>
<th>free living</th>
<th>Genome size; number of genes</th>
<th>Total number of sigma factors</th>
<th>Number of sigma factors in s(^{70}) family</th>
<th>Number of sigma factors in s(^{54}) family</th>
<th>Number of ECF sigma factors</th>
<th>Quotients 1; 2</th>
<th>Reference (for completely sequenced genome)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Aquifex aeolicus</em></td>
<td>Aae</td>
<td>Aquificales</td>
<td>marine, deep sea vents; chemolithoautotroph, hyperthermophilic (96°C)</td>
<td>Yes</td>
<td>1.55 Mb; 1522</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0.39; 0.38</td>
<td>Deckert et al., 1998</td>
</tr>
<tr>
<td><em>Bacillus halodurans</em></td>
<td>Bha</td>
<td>Low GC gram positive</td>
<td>alkaliphilic organism from deep sea sediment</td>
<td>Yes</td>
<td>4.2 Mb; 4066</td>
<td>19</td>
<td>8</td>
<td>1</td>
<td>10</td>
<td>0.22; 0.21</td>
<td>Takami et al., 2001</td>
</tr>
<tr>
<td><em>Bacillus subtilis</em></td>
<td>Bsu</td>
<td>Low GC gram positive</td>
<td>soil, association with plants arachnid and mammalian hosts symbiont of pea aphid involved in food poisoning, especially in contaminated poultry; causes inflammatory enterocolitis nutrient poor aquatic habitats</td>
<td>Yes</td>
<td>4.21 Mb; 4221</td>
<td>18</td>
<td>10</td>
<td>1</td>
<td>7</td>
<td>0.23; 0.23</td>
<td>Kunst et al., 1997</td>
</tr>
<tr>
<td><em>Borrelia burgdorferi</em></td>
<td>Bbu</td>
<td>Spirochaetales</td>
<td></td>
<td>No</td>
<td>0.91 Mb; 903</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0.30; 0.30</td>
<td>Fraser et al., 1997</td>
</tr>
<tr>
<td><em>Buchnera sp.</em></td>
<td>Bsp</td>
<td>a-subdivision of proteobacteria</td>
<td></td>
<td>No</td>
<td>0.64 Mb; 564</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0.32; 0.28</td>
<td>Shigenobu et al., 2000</td>
</tr>
<tr>
<td><em>Campylobacter jejuni</em></td>
<td>Cje</td>
<td>e-subdivision of proteobacteria</td>
<td></td>
<td>No</td>
<td>1.64 Mb; 1731</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0.55; 0.58</td>
<td>Parkhill et al., 2000a</td>
</tr>
<tr>
<td><em>Caulobacter crescentus</em></td>
<td>Ccr</td>
<td>a-subdivision of proteobacteria</td>
<td></td>
<td>Yes</td>
<td>4.02 Mb; 3794</td>
<td>17</td>
<td>2</td>
<td>1</td>
<td>14</td>
<td>0.24; 0.22</td>
<td>Nierman et al., 2001</td>
</tr>
<tr>
<td><em>Chlamydia trachomatis</em></td>
<td>Ctr</td>
<td>Chlamydiales</td>
<td>mouse pathogen (pneumonits)</td>
<td>No</td>
<td>1.04 Mb; 825</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0.35; 0.28</td>
<td>Read et al., 2000</td>
</tr>
<tr>
<td><em>Clostridium acetobutylicum</em></td>
<td>Cac</td>
<td>Low GC gram positive</td>
<td>soil, anaerobic spore former isolated from environments rich in organic nutrients; as well as from nutrient poor habitats colonizes lower gut of animals, survives release into environment</td>
<td>Yes</td>
<td>3.94 Mb; 4023</td>
<td>14</td>
<td>10</td>
<td>1</td>
<td>3</td>
<td>0.28; 0.29</td>
<td>Nölling et al., 2001</td>
</tr>
<tr>
<td><em>Deinococcus radiodurans</em></td>
<td>Dra</td>
<td>Thermus/ Deinococcus group</td>
<td></td>
<td>Yes</td>
<td>2.65 Mb; 3174</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1.33; 1.59</td>
<td>White et al., 1999</td>
</tr>
<tr>
<td><em>Escherichia coli</em> K12</td>
<td>Eco</td>
<td>a-subdivision of proteobacteria</td>
<td></td>
<td>Yes</td>
<td>4.64 Mb; 4290</td>
<td>7</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>0.66; 0.61</td>
<td>Blattner et al., 1997</td>
</tr>
<tr>
<td><strong>Haemophilus influenzae</strong></td>
<td>Hin</td>
<td>1.83 Mb; 1707</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0.46; 0.43</td>
<td>Fleischmann et al., 1995</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----</td>
<td>--------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---------</td>
<td>-------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Helicobacter pylori J99</strong></td>
<td>Hpy</td>
<td>1.64 Mb; 1491</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0.55; 0.5</td>
<td>Tomb et al., 1997</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lactococcus lactis</strong></td>
<td>Lla</td>
<td>2.37 Mb; 2310</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0.79; 0.77</td>
<td>Bolotin et al., 2001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mesorhizobium loti</strong></td>
<td>Mlo</td>
<td>7.04 Mb; 6804</td>
<td>23</td>
<td>3</td>
<td>2</td>
<td>18</td>
<td>0.31; 0.30</td>
<td>Kaneko et al., 2000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mycobacterium leprae</strong></td>
<td>Mle</td>
<td>3.27 Mb; 2770</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0.82; 0.69</td>
<td>Cole et al., 2001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mycobacterium tuberculosis</strong></td>
<td>Mtu</td>
<td>4.41 Mb; 3923</td>
<td>13</td>
<td>3</td>
<td>0</td>
<td>10</td>
<td>0.34; 0.30</td>
<td>Cole et al., 1998</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mycoplasma genitalium</strong></td>
<td>Mge</td>
<td>0.58 Mb; 468</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.58; 0.47</td>
<td>Fraser et al., 1995</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| **Mycoplasma pneumoniae**  
| M129 | Mpn | Low GC gram positive | human pathogenic bacterium causing tracheobronchitis and primary atypical pneumonia | No | 0.82 Mb; 677 | 1 | 1 | 0 | 0 | 0.58; 0.68 | Himmelreich *et al.*, 1996 |
| **Mycoplasma pulmonis** | Mpu | low GC gram positive | causal agent of murine respiratory mycoplasmosis | No | 0.96 Mb; 814 | 1 | 1 | 0 | 0 | 0.96; 0.81 | Chambaud *et al.*, 2001 |
| **Neisseria meningitidis**  
| Z2491 | Nme | β-subdivision of proteobacteria | opportunistic pathogen causing bacterial meningitis colonizing nasopharynges and oropharynges | No | 2.18 Mb; 2230 | 4 | 2 | 1 | 1 | 0.55; 0.56 | Parkhill *et al.*, 2000b |
| **Pasteurella multocida**  
| Pmu | γ-subdivision of proteobacteria | multispecies pathogen; causes serious diseases in food animals and humans | No | 2.26 Mb; 2077 | 4 | 2 | 0 | 2 | 0.56; 0.52 | May *et al.*, 2001 |
| **Pseudomonas aeruginosa**  
| Pae | γ-subdivision of proteobacteria | versatile bacterium; found in soil, marshes, coastal marine habitats and on plant and animal tissues; forms biofilms on wet surfaces (rocks, soil); opportunistic pathogen (urinary tract infections, burn wounds, pneumonia; cystic fibrosis) | Yes | 6.26 Mb; 5640 | 24 | 4 | 1 | 19 | 0.26; 0.24 | Stover *et al.*, 2000 |
Table 2. Continued

<table>
<thead>
<tr>
<th>Organism</th>
<th>Type</th>
<th>Genetic Material</th>
<th>Genome Size</th>
<th>GC Content</th>
<th>Pathogenicity</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rickettsia prowazekii</td>
<td>Rpr</td>
<td>γ-subdivision of proteobacteria</td>
<td>obligate intracellular parasite</td>
<td>No 1.11 Mb; 834</td>
<td>2 2 0 0</td>
<td>0.56; 0.42</td>
</tr>
<tr>
<td>Staphylococcus aureus Mu50</td>
<td>Sau</td>
<td>Low GC gram positive</td>
<td>nasal membrane and skin of warm-blooded animals</td>
<td>No 2.81 Mb; 2673</td>
<td>3 3 0 0</td>
<td>0.94; 0.89</td>
</tr>
<tr>
<td>Streptococcus pyogenes</td>
<td>Spy</td>
<td>Low GC gram positive</td>
<td>strict human pathogen associated with a wide variety of diseases</td>
<td>No 1.85 Mb; 1769</td>
<td>3 3 0 0</td>
<td>0.62; 0.59</td>
</tr>
<tr>
<td>Synechocystis sp. PCC6803</td>
<td>Ssp</td>
<td>Cyanobacteria</td>
<td>phototrophic growth, facultative glucose-heterotrophic growth</td>
<td>Yes 3.54 Mb; 3168</td>
<td>8 6 0 2</td>
<td>0.44; 0.40</td>
</tr>
<tr>
<td>Thermotoga maritima</td>
<td>Tma</td>
<td>Thermotogales</td>
<td>optimum growth temperature 80°C, isolated from geothermally heated marine sediment; metabolizes simple and complex carbohydrates</td>
<td>Yes 1.86 Mb; 1877</td>
<td>4 3 0 1</td>
<td>0.47; 0.47</td>
</tr>
<tr>
<td>Treponema pallidium</td>
<td>Tpa</td>
<td>Spirochaetales</td>
<td>causative agent of syphilis</td>
<td>No 1.14 Mb; 1031</td>
<td>5 3 1 1</td>
<td>0.23; 0.21</td>
</tr>
<tr>
<td>Vibrio cholera</td>
<td>Vch</td>
<td>γ-subdivision of proteobacteria</td>
<td>aetiological agent of cholera; free-living species in marine environments; pathogen via horizontal gene transfer</td>
<td>Yes 4.03 Mb; 3828</td>
<td>8 4 1 3</td>
<td>0.5; 0.48</td>
</tr>
<tr>
<td>Ureaplasma urealyticum</td>
<td>Uur</td>
<td>Low GC gram positive</td>
<td>opportunistic pathogen of the human urogenital tract</td>
<td>No 0.75 Mb; 611</td>
<td>1 1 0 0</td>
<td>0.75; 0.61</td>
</tr>
<tr>
<td>Xylella fastidiosa</td>
<td>Xfa</td>
<td>γ-subdivision of proteobacteria</td>
<td>Plant pathogen transmitted by leafhoppers; embedded in extracellular translucent matrix in the plant</td>
<td>No 2.68 Mb; 2830</td>
<td>4 2 1 1</td>
<td>0.67; 0.71</td>
</tr>
</tbody>
</table>
sigma factor ComX is present only in \textit{Lactococcus lactis} and \textit{Streptococcus pyogenes} (2 copies). The putative sigma factor SigH of \textit{Staphylococcus aureus} is only distantly related to the SigH cluster. An alignment of the SigH, SigI proteins and of Xpf\textsubscript{Bsu}, CAC2052\_Cac and DR0804\_Dra was performed. A similar condensed tree as in Figure 2 was obtained (data not shown). The putative sigma factor DR0804 of \textit{Deinococcus radiodurans} is not related to any other sigma factor and serves as outgroup for rooting the tree.

An unrooted phylogenetic tree based on an alignment of whole amino acid sequences of the ECF subfamily is shown in Figure 3, a bootstrapped tree based on an alignment of the conserved regions is shown in Figure 4. The sequence of DR0804 of \textit{D. radiodurans} was used again as outgroup for rooting the bootstrapped tree. It seems evident that many of the genes encoding ECF sigma factors in the genomes of \textit{C. crescentus}, \textit{M. loti} and \textit{P. aeruginosa} evolved by gene duplication. It will be a challenge for experimental studies to determine the stimuli resulting in activation of the sigma factors and to identify the corresponding regulons.

An unrooted phylogenetic tree based on an alignment of whole amino acid sequences of the $\sigma^{54}$ family is shown in Figure 5, a bootstrapped tree based on an alignment of the conserved regions is shown in Figure 6. Interestingly, the putative RpoN sigma factor NMA0049 of \textit{Neisseria meningitidis} is not present in the cluster of the other proteobacterial species.

\textbf{Discussion}

The number of sigma factor encoding genes in completely sequenced bacterial genomes exhibits a broad variation. With the exception of \textit{D. radiodurans}, all eubacteria investigated in this study possess a housekeeping sigma factor. This may be surprising; however very little is known about transcriptional regulation in this species (Meima \textit{et al.}, 2001). The two RpoD paralogs in \textit{M. tuberculosis} fulfill different functions: SigA performs housekeeping functions related to the SigH cluster. A heat shock inducible sigma factor in \textit{B. subtilis} (Zuber \textit{et al.}, 1994), a putative competence sigma factor from \textit{L. lactis} (Bolotin \textit{et al.}, 2001), a putative competence sigma factor similar to SigH of \textit{B. subtilis} from \textit{Staphylococcus aureus} (accession numbers AAK15307, BAB41722), and two putative competence sigma factors from \textit{S. pyogenes} (ComX1, ComX2; Lee and Morrison, 1999). Two putative sigma factors from \textit{C. acetobutylicum} (CAC1226, CAC2052) have unknown functions and are not related to any other sigma factor investigated in this study.
functions, whereas SigB seems to be involved in stress responses, similar to RpoS (Hu and Coates, 1999). This result may indicate that M. tuberculosis (and perhaps also Mycobacterium leprae) possess both the RpoS mediated stress response typical for E. coli and the SigB mediated general stress response typical for B. subtilis (this sigma factor is named SigF in M. tuberculosis (Chen et al., 2000)). The gram-positive anaerobic bacterium Clostridium acetobutylicum as well as Lactococcus lactis and Streptococcus pyogenes which tolerate low oxygen doses do not possess a homologous gene encoding the general stress response sigma factor \( \sigma^B \) of B. subtilis (Mittenhuber, 2002). This observation supports a conclusion that the presence of oxygen might have contributed to the evolution of the \( \sigma^B \) dependent general stress response in B. subtilis and aerobic gram positive bacteria (Hecker and Völker, 2001).

In E. coli, the sigma factor RpoS controls a regulon which is functionally similar to the \( \sigma^B \) regulon of B. subtilis (Hengge-Aronis, 2000). Several studies have been also performed on the RpoS regulons in other bacteria besides E. coli: A connection between an intact rpoS gene in Borrelia burgdorferi and stationary phase gene expression was established recently (Elias et al., 2000). In P. aeruginosa, RpoS plays a role in regulating stress responses and virulence (Suh et al., 1999), quorum sensing (Whiteley et al., 2000) and biofilm formation (Xu et al., 2001). In Vibrio cholerae, RpoS is required for stress resistance (Yildiz and Schoolnik, 1998) and colonization of the...
intestine (Merrell et al., 2000) but not for in vivo survival (Yildiz and Schoolnik, 1998). Also, the absence of an rpoS gene in the genome of Campylobacter jejuni has been proposed as explanation for the absence of a stationary phase associated stress response (Kelly et al., 2001).

A few interesting observations supporting some aspects of this study have been reported previously: In cyanobacteria, multiple paralogous genes encoding RpoD variants have been described previously (Tanaka et al., 1992). Similar to M. loti, the nitrogen fixing bacterium Bradyrhizobium japonicum carries two RpoN paralogs (Kullik et al., 1991) but three RpoH paralogs (Narberhaus et al., 1997). C. acetobutylicum possesses a complete set of sporulation specific sigma factors supporting previous work which indicated that clostridia may use the same pathway as bacilli for spore formation (Sauer et al., 1995).

However, the presence of a sigma factor in a genome does not indicate that the same set of genes as in model organisms forms the regulon controlled by the individual sigma factor in different species. For example, in Caulobacter crescentus RpoN controls a regulon involved in biosynthesis of polar structures (flagellum, stalk) and disruption of rpoN results in cell division defects (Brun and Shapiro, 1992). A combination of the RpoN and FliA regulons is involved in flagellar biosynthesis in Vibrio cholerae (Prouty et al., 2001) and Campylobacter jejuni (Jagannathan et al., 2001). In B. subtilis, SigL is required for certain catabolic reactions (use of branched amino acids and arginine and ornithine as nitrogen source; (Débarbouillé et al., 1991). As reviewed by Studholme and Buck (2000), individual RpoN regulons are quite diverse in the microbial world.

The use of group 1 sigma factors as marker molecules for molecular systematics has been reported (Gruber and Bryant, 1997). To obtain additional information on the relationships between the sigma factors themselves, group 2 sigma factors were also used by Gruber and Bryant for the construction of phylogenetic trees. Group 1 sigma factors are the only suitable marker molecules for inferring phylogenetic relationships simply because the content of alternative sigma factors varies between individual species. This variation in sigma factor content might indicate immense flexibility of transcriptional regulation in individual species. Taken together, all the observations discussed above strongly indicate that regulation of gene expression in terms of regulons controlled by specialized alternative sigma factors was a late evolving phenomenon. Support for this point of view was recently obtained in a phylogenomic study of the stress response sigma factor σB of B. subtilis and its regulatory proteins. Occurrence of σB is restricted to a

![Figure 2. Phylogenetic tree of the σ70-subfamily of sigma factors. The tree was rooted using DR0804_Dra.](Inventory of Sigma Factors 85)
small group of grampositive bacteria, whereas genes encoding the regulatory proteins are also found in species which do not possess a gene related to $\sigma^B$ (Mittenhuber, 2002). In an interesting contrast, regulation of another stress response, namely the stringent response conferred by the regulatory molecule (p)pGpp is common to all eubacteria (Mittenhuber, 2001).

Different bacterial species use a wide variety of mechanisms for transcriptional regulation which is of course not restricted to alternative sigma factors (Vicente et al., 1999). This transcriptional versality conferred by alternative sigma factors can be even observed in different laboratory stocks of the same species: In different E. coli stocks, heterogeneity was discovered with respect to the two sigma subunits FliA and RpoS (Jishage and Ishihama, 1997). Sequence variability in the rpoS gene of E. coli (called katF in this paper) was also observed by Ivanova et al. (1992).

An attempt to relate genome size and/or total gene number to the number of sigma factors provided an unconvincing result. No threshold value can be

Figure 3. Unrooted phylogenetic tree of the ECF subfamily of sigma factors.
determined which definitely distinguishes free-living from pathogenic bacteria. However, free-living bacteria tend to have lower values than pathogenic species. Eisen (2000) stated that the number of currently available completely sequenced genomes does not reflect biodiversity. As more complete genome sequences from a variety of closer and more distantly related bacteria become available, time will tell whether these quotients might be a useful measure. For example, these quotients might be only useful for a unique phylum or for a group of closely related bacteria.

The present nomenclature for bacterial sigma factors is quite confusing. For example, orthologous sigma factors carry different names in different organisms: The general stress sigma factor SigB of the Bacillus group is called SigF in Mycobacterium tuberculosis and in Synechocystis sp PCC6803. In Bacillus species, however, SigF is a sporulation sigma factor. Another example: RpoE is a heat shock ECF sigma factor in E. coli (Rouvière et al., 1995). However, this designation has been also used for the delta subunit of RNA polymerase in gram positive bacteria (Lampe et al., 1988). This has resulted in an error in the annotation of the S. pyogenes genome (Ferretti et al., 2001): It is stated that S. pyogenes possesses a sigma factor encoding gene similar to rpoE of E. coli. The rpoE gene of S. pyogenes (accession number AAK34603), however encodes the delta subunit of RNA polymerase and not a sigma factor. In order to avoid similar errors in future genome projects, development of a new nomenclature system for bacterial sigma factors would be very desirable. However, finding consensus gene names is not an easy task since changes in gene names are commonly only accepted by the scientific community with some reluctance (Pearson, 2001).

Experimental Procedures

Individual amino acid sequences of sigma factors were obtained at the NCBI server (http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?db=Protein). The complete genome sequence of C. acetobutylicum was obtained from Genome Therapeutics, Inc. (http://www.genomecorp.com/programs/sequence_data_clostr.shtml). Alignments were generated using CLUSTALW (Thompson et al., 1994) at the EBI server (http://www.ebi.ac.uk/clustalw/). The construction of rooted phylogenetic trees and statistical tests of tree topologies (bootstrapping) was performed using MEGA 2 (Kumar et al., 2001; http://www.megasoftware.net). The neighbor-joining method (Saitou and Nei, 1987) was used for tree construction. Bootstrapping (Felsenstein, 1985) was performed on the basis of the p-distance (proportion of different amino acids)
with 1000 replicates. Unrooted phylogenetic trees were constructed using the data from the alignment with the help of the program TreeView (Page, 1996) (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html) and edited using the program Metafile Companion (http://www.companionsoftware.com/).

**Acknowledgements**

This work was performed in the laboratory of Michael Hecker, whose support is gratefully acknowledged. I thank Doug Smith (Genome Therapeutics, Inc.) for the permission to perform analyses using the unpublished *C. acetobutylicum* genomic sequence. I thank Michael Hecker and Ulrich Zuber for comments on the manuscript and Ulrich Zuber and Knut Büttner for helpful discussions.

**References**


