Phylogenetic, amino acid content and indel analyses of the β subunit of DNA-dependent RNA polymerase of Gram-positive and Gram-negative bacteria

Robert Morse, Karen O’Hanlon and Matthew D. Collins

Author for correspondence: Matthew D. Collins. Tel: +44 118 9357226. Fax: +44 118 9357222. e-mail: m.d.collins@reading.ac.uk

In this study, we have sequenced the rpoB gene, encoding the β subunit of DNA-dependent RNA polymerase, from a selection of Gram-positive and Gram-negative bacteria. The presence of insertions and deletions (indels) in the β subunit separate the Gram-positive and Gram-negative bacteria from each other and support the division of the Gram-positive organisms into two clades based on DNA G+C content. Phylogenetic and amino acid content analyses further separate the clostridia from bacilli, leuconostocs, listeriae and relatives, forming an early branch after the common Gram-positive ancestor. The occurrence in the β subunit of Asn–Ala at positions 471–472 in Porphyromonas cangingivalis and Asn at position 372 in Weissella paramesenteroides are postulated to be the cause of the natural rifampicin resistance of these species.

Keywords: phylogeny, Gram-positive bacteria, DNA-dependent RNA polymerase, rpoB gene, rifampicin resistance

INTRODUCTION

Due to the pioneering work of Carl Woese (Woese, 2000), 16S rRNA has increasingly occupied the centre stage in microbial systematics and has revolutionized our understanding of prokaryote evolution and diversity. Until recently, nearly all phylogenetic reconstructions amongst the bacteria were based on this single molecular chronometer. Recently, however, systematists have begun to compare and contrast phylogenetic relationships inferred from other molecules with those of 16S rRNA (Gupta, 1998; Hansmann & Martin, 2000). Examples of such alternative chronometers include GrpE (Ahmad et al., 2000), ribosomal protein (Bocchetta et al., 2000), Hsp70 (Gribaldo et al., 1999), Hsp40 (Bustard & Gupta, 1997) and EF-G (Bocchetta et al., 1995). These studies have, with a few exceptions (Hasegawa & Hashimoto, 1993), demonstrated that 16S rRNA is an excellent molecule for reconstructing phylogenetic relationships, although recent evidence of lateral transfer of rRNA gene segments suggests that caution should always be observed in interpreting such analyses (Wang & Zhang, 2000).

DNA-dependent RNA polymerase (RNAP) is a multi-subunit enzyme consisting of two α subunits (encoded by the rpoA gene), one β subunit (rpoB) and one β′ subunit (rpoC). The large sizes of the β and β′ subunits (respectively 151 and 156 kDa in Escherichia coli), together with their ancient origin in evolutionary terms, indicate that these molecules could be immensely powerful molecular chronometers for phylogenetic analysis (Rowland et al., 1992; Klenk & Zillig, 1994; Morse et al., 1996a; Mollet et al., 1997, 1998; Klenk et al., 1999; Bocchetta et al., 2000). The key role for RNAP in transcription of DNA makes this independent of the translational apparatus. Although natural intraspecific horizontal rpoB gene fragment exchange has been demonstrated (Lorenz & Sikorski, 2000), this does not necessarily preclude its use for phylogenetic analysis.

One of the earliest weighted characters employed in bacterial classification was the Gram stain, which, although now thought unsuitable for phylogenetic
Table 1. Strains used and their rpoB gene sequence accession numbers

<table>
<thead>
<tr>
<th>Strain</th>
<th>rpoB gene accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Clostridium argentinense</em> ATCC 27322'T (toxin type G)</td>
<td>Y16472</td>
</tr>
<tr>
<td><em>Leuconostoc mesenteroides</em> NCIMB 8023'T</td>
<td>Y16467</td>
</tr>
<tr>
<td><em>Clostridium botulinum</em> NCTC 7272</td>
<td>Y16466</td>
</tr>
<tr>
<td><em>Listeria monocytogenes</em> NCTC 7973</td>
<td>Y16468</td>
</tr>
<tr>
<td><em>Listeria murrayi</em> NCTC 10812'T</td>
<td>Y16469</td>
</tr>
<tr>
<td><em>Weissella paramesenteroides</em> NCIMB 13092'T</td>
<td>Y16471</td>
</tr>
<tr>
<td><em>Porphyromonas cangengivalis</em> NCTC 4874'T</td>
<td>Y16470</td>
</tr>
</tbody>
</table>

analysis, has become an essential dichotomy for classification schemes (Shah et al., 1997). The application of 16S rRNA analyses to relationships within the Gram-positive bacteria has produced differing patterns of relationships that depend on the method of analysis used (Olsen et al., 1994; Ludwig et al., 1994) and which are sometimes in conflict with those defined either by other gene products, e.g. EF-Tu (Ludwig et al., 1994), ATPase β (Ludwig et al., 1994) and RecA (Eisen, 1995), or by whole genome analysis (Tekaia et al., 1999). The classification of Gram-positive bacteria into two groups based on DNA G+C content (high and low G+C) is, however, supported by all these chronometers, together with studies on the types of RNaseP (Haas et al., 1996) and DNA polymerase C (Huang & Ito, 1999) that they contain and the organization of their dnaK operons (Segal & Ron, 1996). Inferences within these groups are again varied, with phylogenetic studies on the HrcA (Ahmad et al., 1999) and GrpE (Ahmad et al., 2000) proteins suggesting two separate groups of low-G+C Gram-positive bacteria. The relationship of these two (three) Gram-positive groups with the Gram-negative bacteria is also unclear, with the high-G+C group being more closely linked on the basis of RNaseP RNA type (Haas et al., 1996) and the presence of shared indels in S12 protein, dihydroorotate dehydrogenase and pyruvate kinase (Gupta, 1998), whilst the low-G+C group is more closely linked on the basis of shared indels in gyrA (Gupta, 1998). Phylogenetic trees based on various chronometers have indicated close relationships of Gram-positive bacteria to various groups of other bacteria including the cyanobacteria (Ahmad et al., 1999; Eisen, 1995), halobacteria (Gribaldo et al., 1999; Haas et al., 1996) and mycoplasmas (Weisburg et al., 1989a), whilst specific links have been suggested to *Methanosarcina mazei* (Gribaldo et al., 1999) and *Borrelia burgdorferi* (Eisen, 1995). Likewise, such chronometers also give variable phylogenetic positions for the clostridia (Ahmad et al., 1999, 2000; Olsen et al., 1994; Gribaldo et al., 1999; Bustard & Gupta, 1997).

Previously, we have sequenced the rpoC gene from various leuconostocs and performed a comparative phylogenetic analysis to discern relationships and the tenet of tachyelic evolution within this group of Gram-positive bacteria (Morse et al., 1996a). In this study, we have extended this work by sequencing the rpoB gene from various Gram-positive bacteria to determine the inter- and intrarelationships of this group of bacteria.

**METHODS**

**Bacterial strains and DNA preparation.** The strains used in this study are listed in Table 1. Strains were cultured in appropriate media and at temperatures recommended in the relevant culture collection catalogues. Chromosomal DNA was extracted as described previously (Harland et al., 1993). The authenticity of each strain examined was confirmed by partial 16S rRNA gene sequence analysis (data not shown).

**PCR amplification.** In order to obtain nearly complete copies of the rpoB gene, PCR products were generated using oligonucleotide primers (Table 2) complementary to DNA sequences encoding conserved amino acid sequences in the rpoB genes of *Mycobacterium tuberculosis* (EMBL accession no. U12205), *Mycobacterium leprae* (Z14314), *Mycobacterium smegmatis* (U24494), *Bacillus subtilis* (L24376) and *Staphylococcus aureus* (X64172) and the rplL genes of *Escherichia coli* (V00339) and *Staphylococcus aureus* (X64172). PCRs were performed in 1× PCR buffer (Perkin–Elmer) containing 10 mM dNTPs and 4 ng of each oligonucleotide primer 1 µM in a final volume of 50 µl under a layer of PCR-grade mineral oil (Sigma) using a Biometra PCR thermal cycler. After an initial denaturation step of 94 °C for 5 min, AmpliTaq polymerase (Perkin–Elmer) containing 100 mM dNTPs and 4 ng of each oligonucleotide primer 1 µM in a final volume of 50 µl under a layer of PCR-grade mineral oil (Sigma) was used to amplify a final concentration of 0.02 U µl−1 and 25 cycles of 92 °C for 1 min, 48 °C for 1 min and 65 °C for 3 min were performed, ending with a final extension step of 65 °C for 10 min.

**Cloning, transformation and sequencing.** PCR products were purified from primers, nucleotides and enzyme using a QIAquick PCR purification kit (Qiagen), ligated into pCRII vector (TA cloning kit; Invitrogen) and transformed into *Escherichia coli* INV Pl cells according to the manufacturer’s instructions. DNA sequencing was performed by using the dideoxynucleotide chain termination method on both positive and negative strands of each cloned PCR product. Two independent PCR products were sequenced for each rpoB gene to check AmpliTaq polymerase fidelity, with a third PCR product being generated to resolve any nucleotide base discrepancies. On average, one nucleotide base error was detected for every 3 kb of sequence.

**Sequence analysis.** Sequences were aligned using the PILEUP program from the Wisconsin Molecular Biology software package (Devereux et al., 1984) and the multiple alignments were edited manually. Distance matrices were calculated for the DNA alignments using the programs PRETTY and
DNA-dependent RNA polymerase phylogeny

Table 2. Oligonucleotides used to generate PCR products

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>1475</td>
<td>GA(A/G)AA(A/G)ACNGA(A/G)TT(T/C)GA(T/C)GT</td>
<td>151–166</td>
</tr>
<tr>
<td>2233</td>
<td>GTNTT(T/C)ATGGGNGA(T/C)TT(T/C)CC</td>
<td>331–350</td>
</tr>
<tr>
<td>2234</td>
<td>GGG(A)/AAA(G)/ATCNCCCAT(G/A)AANAC</td>
<td>330–331</td>
</tr>
<tr>
<td>2235</td>
<td>AA(G)/A(A)/CNGNAT(C)/GA(T/C)(T/C)TNGC</td>
<td>787–803</td>
</tr>
<tr>
<td>2236</td>
<td>GCNA(G/A)(G/A)/AT(T/G)(C/T)TANC(T/G)(C/T)TT</td>
<td>803–877</td>
</tr>
<tr>
<td>2237</td>
<td>GA(T/C)GA(C/T)ATNGA(C/T)(C/A)(C/T)(C/T)T</td>
<td>1195–1211</td>
</tr>
<tr>
<td>2238</td>
<td>ATG(A/G)/ATG(G)/AJTNCAT(G/A)(T/C)(G/A)(T/C)TC</td>
<td>1211–1215</td>
</tr>
<tr>
<td>2239</td>
<td>GA(G)/AGTN(C)/AGNGNCA(T/C)TGTNCA</td>
<td>1504–1520</td>
</tr>
<tr>
<td>2240</td>
<td>TGNAC(G/A)(G/A)/NAC(T/C)(C/T)TC</td>
<td>1520–1540</td>
</tr>
<tr>
<td>2074</td>
<td>AA(C/T)ATGCAGA(G/A)(C/A)GNCA(G/A)GCNGT</td>
<td>1924–1943</td>
</tr>
<tr>
<td>2243</td>
<td>ACNGC(C/T)TGNC(G/T)(C/T)TGCAT(G/A)TT</td>
<td>1943–1924</td>
</tr>
<tr>
<td>2241</td>
<td>GA(G)/AGA(G/A)/ATNACNGNGA(T/C)AT</td>
<td>2419–2438</td>
</tr>
<tr>
<td>2242</td>
<td>AT(A/G)TNCNGNGA(T/C)(C/T)(C/T)TC</td>
<td>2438–2419</td>
</tr>
<tr>
<td>2076</td>
<td>GCNATTT(T/C)GGNGA(G/A)AA(G/A)AGC</td>
<td>2587–2606</td>
</tr>
<tr>
<td>2244</td>
<td>GC(T/C)TT(T/C)TCNCC(A/G)AAATNG(T/G)(T/C)C</td>
<td>2606–2587</td>
</tr>
</tbody>
</table>

RESULTS AND DISCUSSION

The PCR amplification strategy generated complete copies of the rpoB gene from the Gram-positive bacterium Listeria monocytogenes and the Gram-negative bacterium Porphyromonas gingivalis and partial copies of the gene from Clostridium argentinense (encoding a region equivalent to positions 117–1068 of the Bacillus subtilis β subunit), Clostridium botulinum (113–1059), Leuconostoc mesenteroides (111–1081), Weissella paramesenteroides (117–1047) and Listeria murrayi (121 to the end of the subunit). The fidelity of the sequence data generated from PCRs was checked rigorously and discrepancies were resolved by sequencing additional products from independent PCRs. A multiple alignment of the β subunit amino acid sequences deduced from the gene sequences together with the equivalent regions from a selection of other bacteria is available as supplementary material in IJSEM Online (http://ijis.sgmjournals.org/). As is evident from the alignment, and as we have shown previously for the β subunit (Morse et al., 1996b), the numerous regions of high primary structural conservation that are interspersed throughout the β subunit molecule greatly facilitate the alignment of less-conserved regions. Exclusion of such regions due to poor alignment can greatly affect inferred phylogenies (Hansmann & Martin, 2000). The natural rifampicin resistance of Porphyromonas gingivalis and Porphyromonas camingivalis (Klimpel & Clark, 1990; Collins et al., 1994) is possibly due to the sequence Asn→Ala found at positions 347–471 in the former (NCBI BLAST search of unpublished sequence) and 471–472 in the latter species (present work). Mutations at the equivalent highly conserved site Ser→Ala in other β subunits result in rifampicin resistance, such as Ser464Pro and Gln465Arg in Staphylococcus aureus (Dubry-Damon et al., 1998; Wichelhaus et al., 1999) and Ser→Gln (509–510) mutated to Val in Escherichia coli (Jin et al., 1988). The natural rifampicin resistance of Weissella paramesenteroides (Swenson et al., 1990) is possibly due to the presence of Asn at position 372, which is usually occupied by a highly conserved His in most other β subunits (Morse et al., 1999). Mutation of this His to Asn induces rifampicin resistance in Helicobacter pylori (Heep et al., 1999), Streptococcus pneumoniae (Enright et al., 1998), Staphylococcus aureus (Wichelhaus et al., 1999) and Mycobacterium tuberculosis (Telenti et al., 1993). Asn also occurs at the equivalent position in the β subunit of the naturally rifampicin-resistant mycoplasmas (Liberal & Boughon, 1991) and Spiroplasma citri (Gaurivaud et al., 1996). Another region of RNAP that has recently been identified as important to rifampicin binding is the F segment of the β subunit (Naryshkina et al., 2001). In particular, a stretch of four amino acids in this region (Thr→Ala–Leu–Lys at positions 786–789 in Escherichia coli) forms van der Waals interactions with amino acids in the β subunit (567–571 in Escherichia coli) to form the rifampicin-binding site. However, we have shown previously that these four amino acids are also found at the equivalent positions in the β subunits of

DNA-DIST (using the Kimura two-correction parameter) (Felsenstein, 1989) and for the amino acid alignment using the GeneBase program (Applied Maths). Phylogenetic trees were generated using the neighbour-joining method with the program NEIGHBOR and were then tested statistically by bootstrap analysis (Felsenstein, 1989). Bootstrap values were calculated for the DNA trees from 1000 replicates using the programs SEQBOOT, DNADIST, NEIGHBOR and CONSENSE (Felsenstein, 1989) and for the protein tree from 10 replicates using the GeneBase program.
Porphyromonas cangiivalis (EMBL accession no. Y16470) and Weissella paramesenteroides (Morse et al., 1996a) and, therefore, cannot be the cause of the natural rifampicin resistance in these bacteria.

Obtaining good sequence alignments is a prerequisite of molecular systematic analysis because each aligned position has to include only homologous residues from the different molecules. However, several stretches of amino acid sequence were found only in the mycobacteria and Gram-negative species. In order to minimize alignment ambiguities due to these regions and to alleviate possible error in phylogenetic reconstruction due to incorrect weighting of such differences between species (e.g. residues 941–1040 in Escherichia coli and residues 283–361 in Mycoplasma pneumoniae), these indels were omitted. The reliability of phylogenetic inferences made from comparative analyses of 16S rRNA for a range of bacterial species was assessed by constructing neighbour-joining trees from the β subunit data. Trees were generated using both amino acid similarities and the equivalent DNA sequence at the second codon position. Studies using the B subunit (the equivalent of the bacterial β subunit) of archaeal RNAP have shown the nucleotide at the second codon positions to be unbiased by G-C contents and more reliable than positions 1 and 3 for phylogenetic reconstructions (Klenk & Zillig, 1994). Since differences in species sampling influence tree generation (Lecointre et al., 1993), a 16S rRNA tree was constructed using the same set of species to assess congruence between the two chronometers. A comparison of the neighbour-joining trees derived from the β subunit and 16S rRNA analyses is shown in Fig. 1. The two β subunit trees, based on amino acid similarities and second codon position of rpoB, possess very similar topologies and the latter is therefore not shown. All statistically significant groups (bootstrap values ≥ 90%) were confirmed by maximum-parsimony analysis (data not shown). The overall topology of the β subunit tree (Fig. 1b) is in good agreement with that constructed from 16S rRNA (Fig. 1a) with the exception of the positions of Synechocystis and the mycoplasmas. Discrepancies in the phylogenetic position of these organisms have been noted using other chronometers and ascribed to either variation in data selection or long-branch effects (Klenk et al., 1999; Tekaiia et al., 1999; Bocchetta et al., 2000; Hansmann & Martin, 2000). Other minor topological differences between the trees invariably involve reference species (e.g. Brochothrix thermosphacta) where bootstrap values are low (data not shown) and are therefore of uncertain branching order. Little significance can therefore be attached to these differences. The β subunit tree shows the separation of the Gram-positive bacteria into two distinct clades representing the high-G+C bacteria (Actinobacteria) and the low-G+C bacteria. The latter group appears to be further subdivided into a clade containing clostridial species (26–28 mol % G+C) and one containing the listeriae, Bacillus subtilis, Staphylococcus aureus and the leuconostocs (37–42 mol % G+C), although additional representatives of these groups are needed to be confident of this division. The aforementioned divisions based on G+C content are reflected in the amino acid content of the β subunit proteins. It has been shown (Singer & Hickey, 2000), based on the amino acids encoded by GC- and AT-rich codons, that GC-rich coding DNA tends to produce proteins rich in the amino acids Gly, Ala, Arg and Pro, whilst AT-rich coding DNA favours the inclusion of the amino acids Phe, Tyr, Met, Ile, Asn and Lys. The ratio of the G+C content of the GC-rich-DNA-encoded amino acids to the AT-rich-DNA-encoded
Fig. 2. Amino acid sequences encoded by selected regions of rpoB gene sequences demonstrating (a) an insert specific to the high-G+C Gram-positive bacteria, (b) an insert specific to the low-G+C Gram-positive bacteria and (c) a partial and a complete deletion in the high- and low-G+C Gram-positive bacteria, respectively. Abbreviations: E.c., Escherichia coli; H.i., Haemophilus influenzae; N.m., Neisseria meningitidis; C.d., Corynebacterium diphtheriae; M.t., Mycobacterium tuberculosis; M.l., Mycobacterium leprae; M.s., Mycobacterium smegmatis; A.m., Amycolatopsis mediterranei; T.w., Tropheryma whippelii; St., Streptomyces coelicolor; T.f., Thermobifida fusca; C.a., Clostridium argentinense; C.b., Clostridium botulinum; L.me., Leuconostoc mesenteroides; W.p., Weissella paramesenteroides; L.mo., Listeria monocytogenes; L.mu., Listeria murrayi; B.s., Bacillus subtilis; S.a., Staphylococcus aureus.

amino acids for the β subunits is 1–33 for Mycobacterium tuberculosis, 1–04–13 for listeriae and leuconostocs and 0–86–089 for the clostridial species examined.

Although the bifurcation separating the high- and low-G+C Gram-positive clades was not supported by high bootstrap values, this division was, however, demonstrated unequivocally by the presence of a 12–13 amino acid insert found only in the mycobacteria, Tropheryma whippelii (EMBL accession no. AF243072), Mycolactobacillus leprae; M.s., Mycobacterium smegmatis; A.m., Amycolatopsis mediterranei; T.w., Tropheryma whippelii; St., Streptomyces coelicolor; T.f., Thermobifida fusca; C.a., Clostridium argentinense; C.b., Clostridium botulinum; L.me., Leuconostoc mesenteroides; W.p., Weissella paramesenteroides; L.mo., Listeria monocytogenes; L.mu., Listeria murrayi; B.s., Bacillus subtilis; S.a., Staphylococcus aureus.

flanking sequence homology was present to ensure positional and size equivalence of the inserts (Philippe et al., 1999). The β subunits of Gram-negative bacteria are distinguished from those of Gram-positive bacteria by two rather less well-defined indels (Figs 2c and 3a). One of these (Fig. 2c) represents a sequence that has been partially deleted in mycobacteria, T. whipplei and Amycolatopsis mediterranei (Escherichia coli positions 1150–1177) and completely deleted in the low-G+C Gram-positive bacteria (Escherichia coli positions 1124–1177). The conserved sequence flanking the right-hand side of the deletion suggests that the partial deletion occurred first, before the separation of the high- and low-G+C bacteria, and that the second occurred only in the low-G+C bacteria. This supports an affinity between the mycobacteria and the Gram-negative bacteria indicated previously by work on the classification of RNaseP RNA (Haas et al., 1996) and the presence of indels in S12 protein, dihydroorotate dehydrogenase and pyruvate kinase (Gupta, 1998).
Fig. 3. Amino acid sequences encoded by selected regions of rpoB gene sequences demonstrating (a) an insert specific to the Gram-negative bacteria and (b) an alignment showing how a region of the latter insert has been triplicated in Reclinomonas americana. Abbreviations (in addition to those listed in the legend to Fig. 2): T.a., Thermus aquaticus; T.m., Thermotoga maritima; P.c., Porphyromonas cangingivalis; A.p., Aquifex pyrophilus; A.a., Aquifex aeolicus; R.a., Reclinomonas americana; R.p., Rickettsia prowazekii.

has no homologue in the subunit from Thermotoga maritima. This is consistent with the inference from indels found in Hsp70 and glutamine synthase that this bacterium exhibits a resemblance to Gram-positive bacteria, although this is discordant with work based on signature sequences in pyruvate kinase and gyrase A, which indicates that it is Gram-negative (Gupta, 1998). However, a homologous sequence is found in the equivalent subunit of the bacterial-type RNAP present in the mitochondrion of the heterotrophic flagellate Reclinomonas americana (EMBL accession no. AF007261), lending support to the endosymbiotic origin of mitochondria (Sogin, 1997). The contention that the ancestor was from the α-proteobacteria (Lang et al., 1997) is supported by the close homology between the Reclinomonas americana insert and the corresponding insert in the β subunit of RNAP from Rickettsia prowazekii (Fig. 3a). A close evolutionary relationship between Rickettsia prowazekii and the mitochondria has previously been demonstrated by phylogenetic analysis of ribosomal proteins and NADH dehydrogenase (Andersson et al., 1998). Interestingly, and not previously reported, part of this insert appears to have been triplicated in Reclinomonas americana (Fig. 3b). As noted previously (Klenk et al., 1999), both subunits from Aquifex aeolicus and Aquifex pyrophilus also possess this insert which, together with their position in RNAP phylogenetic trees, has led to the assertion that these species are related to the proteobacteria and do not form the deepest branch of the bacteria, as indicated by analysis based on 16S rRNA (Pitulle et al., 1994), Hsp70 (Gribaldo et al., 1999) and EF-G (Bocchetta et al., 1995). It has been suggested that this erroneous placement of Aquifex species in RNAP trees is due to a distortion effect generated by the inclusion of Mycoplasma RNAP sequences (Bocchetta et al., 2000). Even whole genome analysis of Aquifex species in RNAV trees is due to a distortion effect generated by the inclusion of Mycoplasma RNAP sequences (Bocchetta et al., 2000). Even whole genome analysis of Aquifex species in RNAV trees is due to a distortion effect generated by the inclusion of Mycoplasma RNAP sequences (Bocchetta et al., 2000).
of *Deinococcus radiodurans* (EMBL accession no. AE001944, positions 984–1027) and *Thermus aquaticus* (Y19223, positions 925–967) and containing the consensus sequence DKRE(Q/K)EV(L/D/A)RA(G/E)KLG confirms the common genealogical origin of these two species proposed previously based on chemotaxonomic (*Hensel et al.*, 1986) and 16S rRNA (*Weisburg et al.*, 1989b) data.

Analysis of the RNAP β subunits divides the high- and low-G+C bacteria on the basis of amino acid content, phylogenetic analysis (Fig. 1b) and the presence of indels (Fig. 2a and b). The phylogenetic and amino acid content analyses of the species examined further divides the low-G+C bacteria into a clostridial clade and a *Bacillus, Listeria, Staphylococcus* and leucostoc clade, with the former forming an early branch after the common Gram-positive ancestor (Fig. 1b), as described previously for other chronometers (*Ahmad et al.*, 2000; *Olsen et al.*, 1994; *Gribaldo et al.*, 1999).

Although a number of inserts can be identified in one group of the low-G+C bacteria but not in the others [e.g. VHKD(M/I) in *Leuconostoc mesenteroides* (115–119) and *Weissella paramesenteroides* (110–114)], they also have equivalents in the Gram-negative bacteria, making their use in differentiating the low-G+C bacteria problematic. Although no indels support the previously proposed specific relationships between the Gram-positive bacteria and either mycoplasmas (*Weisburg et al.*, 1989a) or *Borrelia burgdorferi* (Eisen, 1995), there is an insert, (S/G)(S/C/A)(T/V)(L/K)/FPNRD, found in the mycobacteria (*Mycobacterium smegmatis*), *B. burgdorferi*, *Staphylococcus* and *L. monocytogenes*, which may be homologous to an insert in a unique heat-shock regulatory gene.

**ACKNOWLEDGEMENTS**

This work was supported by grants from the BBSRC and the European Commission (B102-CT94-3098).

**REFERENCES**


*Helicobacter pylori*. Development of a DNA probe for the detection of the *helicobacter pylori* DNA. *J Gen Microbiol* 139, 756–769.


*Jin, D. J., Cashel, M., Friedman, D. I., Nakamura, Y., Walter, W. A. & Gross, C. A. (1988).* Effects of rifampicin resistant *rpoB* mutations on...


