Comparison of 16S rRNA Sequences from the Family Pasteurellaceae: Phylogenetic Relatedness by Cluster Analysis

By PAUL J. CHUBA, RENATE BOCK, GABRIELE GRAF, THOMAS ADAM AND ULF GÖBEL*

Institut für Medizinische Mikrobiologie und Hygiene der Universität Freiburg, Herman Herder Strasse 11, D-7800 Freiburg, FRG

(Received 25 January 1988; revised 28 March 1988)

The taxonomy of the family Pasteurellaceae has remained controversial despite investigations of biochemistry, serology, and nucleic acid relatedness. In an attempt to resolve some of this confusion, we have partially sequenced the 16S rRNAs of seven members of the family, representing all three genera. The sequences were aligned, similarity scores calculated, and single, average and complete linkage cluster analysis of the resulting distance matrix performed. In this way, an evolutionary branching pattern of these closely related species was reconstructed, and the approximate phylogenetic position of the family determined. Actinobacillus (Haemophilus) actinomycetemcomitans clustered with Haemophilus instead of Actinobacillus, supporting transfer of this species to the genus Haemophilus. Thus cluster analysis of phylogenetic relatedness was found to be particularly useful for studying closely related organisms, and could be performed using a microcomputer.

INTRODUCTION

Pasteurellaceae is one of three families of Gram-negative facultative anaerobes, many of which are pathogenic for humans and animals. Phylogenetically these organisms belong to the gamma subgroup of the purple bacteria, clustering closely with the Enterobacteriaceae and the Vibrionaceae in 16S rRNA cataloguing studies (Woese et al., 1985). They exhibit considerable interrelatedness as evidenced by biochemistry (Broom & Sneath, 1981; Brondz & Olsen, 1985; Sneath & Stevens, 1985), serology (Potts et al., 1985; MacInnes & Rosendal, 1987), nucleic acid hybridization (Christiansen et al., 1981; Potts & Berry, 1983; Coykendall et al., 1983), and genetic transformation (Albritton et al., 1986). Such data have indicated that the generic descriptions within this family (Pasteurella, Haemophilus and Actinobacillus) may need redefinition (Pohl et al., 1983; Mutters et al., 1984; Bisgaard et al., 1986). In particular, it has been suggested that Actinobacillus actinomycetemcomitans be transferred to the genus Haemophilus (Potts et al., 1985). As outlined below, it appears that direct sequencing of 16S rRNAs would be useful in better defining genetic relatedness within the Pasteurellaceae.

Mutations occur in macromolecules as a quasi clock-like process and the use of comparisons of primary sequences of homologous macromolecules to infer evolutionary relationships has attracted considerable attention (Zuckerkandl & Pauling, 1965; Wilson et al., 1987). The 16S rRNA molecule is particularly well suited to such studies (for reviews see Olsen et al., 1986; Woese, 1987), as relatively conserved regions of the 16S rRNA sequence can be used to study the relatedness of distantly related species, whereas variable regions can be used for the analysis of closely related species. To date, considerable progress has been made in establishing the major phylogenetic branches, based both on direct sequencing (Pace et al., 1986) and on oligonucleotide cataloguing methods (Fox et al., 1980). Recently, a method for the rapid

Abbreviations: I/L, identities per unit length; NAS, normalized alignment score.
acquisition of 16S rRNA sequence data has become available (Lane et al., 1985). The method is based on the bulk cellular RNA as template, oligonucleotides complementary to evolutionarily conserved regions as primers, and reverse transcriptase for chain elongation. We have applied this method to phylogenetic analysis of the Pasteurellaceae and report here partial sequence comparisons between Pasteurella multocida, Haemophilus influenzae, H. aphrophilus, Actinobacillus (Haemophilus) actinomyctecomitans, ‘A. hominis’, A. equuli and A. lignieresii.

**METHODS**

Bacterial strains and growth media. The following strains were obtained from the National Collection of Type Cultures (London, UK): Actinobacillus (Haemophilus) actinomyctecomitans NCTC 9710 (type strain); Actinobacillus equuli NCTC 8529 (type strain); ‘Actinobacillus hominis’ NCTC 11529 (type strain) (Frisi-Moller, 1981); Actinobacillus lignieresii NCTC 4976; Haemophilus influenzae NCTC 8143 (type strain); Haemophilus aphrophilus NCTC 5906 (type strain); Pasteurella multocida NCTC 10322 (type strain). Actinobacillus and Pasteurella strains were cultured in yeast extract (0.5%)/Difco peptone (1%) /glucose (1%) (YPD) medium, which was supplemented with NAD (0.001%) and haemin (0.001%) to support the growth of Haemophilus species. For growth on solid medium Columbia blood agar (Difco), or chocolate agar (Difco) plates (Haemophilus species) were used. A battery of biochemical tests, including sugar fermentations, urease and H2S production, and requirements for X and Y factors, were routinely performed to ensure the purity of the cultures. All cultures were incubated at 37°C with 5% carbon dioxide and without shaking.

Isolation of rRNA. Bulk cellular RNA was isolated using a differential LiCl extraction method essentially as previously described for eukaryotic cells (Auffray & Rougeon, 1980), followed by phenol and chloroform extractions. Cultures (500 ml) were grown to mid-exponential phase, washed once with phosphate-buffered saline (PBS), and the pellets were resuspended in 20 ml 4M LiCl/8M urea and incubated at 0°C for 12-18 h. The mixture was centrifuged at 10000 r.p.m. for 20 min in a Solvall SS34 rotor and the pellets resuspended in 10 ml TE buffer (0.1M-Tris/HCl, 1 mM-EDTA, pH 8.0), extracted two to four times with an equal volume of TE-saturated phenol (BRL), and finally extracted with chloroform/isoamyl alcohol (24:1, v/v). RNA of sufficient quality for use in sequencing was recovered by precipitation with two volumes of ice-cold ethanol and incubation for 30 min at −70°C. The major contaminant in these preparations was bacterial lipopolysacharide, as assessed by SDS gel electrophoresis and silver staining.

Partial sequencing of 16S rRNA. Partial sequences of the 16S rRNAs were obtained using avian myeloblastoma virus reverse transcriptase (Pharmacia), bulk cellular RNA as template (1-3μg ml−1), and oligonucleotides (0-1μg ml−1) complementary to conserved regions of the 16S rRNA molecules as primers in chain-terminating sequencing reactions as described by Lane et al. (1985), using [α-32P]dCTP (800 Ci mmol−1, 29.6 TBq mmol−1; Amersham). The data shown in Fig. 1(a) were obtained using the oligonucleotide primer 5′GTAT-TACCGCGCTGCTGGCAC 3′ and the data shown in Fig. 1(b) were obtained using the oligonucleotides 5′TACCAAGGTATCTAAATCCTGT 3′ and 5′TCGGTGCAGGACTAACCACAA 3′. Oligonucleotides were synthesized using phosphoramidite chemistry on an Applied Biosystems 380A DNA synthesizer, and purified by electrophoresis on 20% (w/v) acrylamide gels followed by electroelution (Göbel et al., 1987a).

Alignment of sequences and cluster analysis. A standard algorithm (Needleman & Wunsch, 1970) was used for sequence alignment using gap penalties (Feng et al., 1985) assigned as described below. To calculate normalized alignment scores (NAS) for each pair of sequences, positions of identity and mismatch were assigned values of ten and zero, respectively, and gaps in the alignment were penalized with values from −10 to −90. The total was then divided by the length of the shorter sequence to give the NAS. The percentage of identities per unit length (I/L) was also calculated for all pairs of sequences. Positions which could not be unambiguously aligned, or where one of the bases was not determined, were not considered in the subsequent analysis. The resulting distance matrices were analysed using the Clustan 3/PC program suite (release 3.2; Wishart, 1987) using a Tandon microcomputer equipped with a math coprocessor. The Jaccard similarity coefficient was chosen for cluster analysis with single, average (unweighted pair group arithmetic average) and complete linkage conditions, as described previously for SS rRNA sequences (Nearhos & Fuerst, 1987).

Fig. 1 (on facing page). Alignments of 16S rRNA sequences determined for seven members of the Pasteurellaceae. The sequences are numbered as follows: 1, P. multocida; 2, H. influenzae; 3, H. aphrophilus; 4, A. (H.) actinomyctecomitans; 5, ‘A. hominis’; 6, A. equuli; 7, A. lignieresii. Positions in the sequences which are identical to the P. multocida sequence are indicated with a period. The letter ‘n’ designates positions which were not determined. (a) Base positions in the sequences are numbered consecutively, where nucleotide 1 corresponds to nucleotide 380 in the standard E. coli 16S rRNA numbering (Brosius et al., 1978). (b) Base positions in the sequences are numbered consecutively, where nucleotide number 1 corresponds to nucleotide 543 in the standard E. coli 16S rRNA numbering. The different oligonucleotide primers used to generate sequences for (a) and (b) are described in Methods.
Alignments of partial 16S rRNA sequences from seven members of the family Pasteurellaceae are shown in Fig. 1. The sequences in Fig. 1(a) correspond to *Escherichia coli* 16S rRNA positions 380-480, and those in Fig. 1(b) correspond to *E. coli* 16S rRNA positions 543-888, so that about 60 bases between the two contiguous segments are missing. These two segments contain nearly all of the variable regions V3, V4 and V5 (Huysmans & De Wachter, 1986). It was not necessary to introduce gaps into these sequences in order to produce a good alignment; however, when a very low gap penalty (−10, see Methods) was used for the pairwise alignments, gaps were introduced between *Pasteurella multocida* and all other sequences in the region around positions 450-480 (loop 17 after Huysmans & De Wachter, 1986). These gaps represent a compensatory insertion/deletion mutational event which retains the length of the loop (P. J. Chuba & U. Göbel, unpublished observations). From a total of 445 positions, 24 were uncertain (5%), and 67 of the remainder were different for the seven members of the family. Of these, 20 were different only for *P. multocida*, so that within the cluster haemophili/actinobacilli 47 of 421 positions in this region were variable. Many of these mutated positions are involved in base pairing, but we have not weighted base pairing positions differently in the construction of distance matrices as suggested by other authors (Olsen *et al.*, 1986), since an analysis of secondary structure does not indicate that a majority of these mutations have arisen in compensatory-type events.

Identities per unit length (I/L) and normalized alignment scores (NAS, gap penalty = 50) were calculated by pairwise comparison of the sequences determined here, plus seven additional sequences from the literature (compiled in Huysmans & De Wachter, 1986). These values are listed in Table 1. *E. coli* and *Proteus vulgaris* sequences served as representative of the Enterobacteriaceae (and the gamma subgroup of the purple bacteria). *Agrobacterium tumefaciens* and *Pseudomonas testosteroni* represented the alpha and beta subgroups of the purple bacteria, respectively. The Gram-positive bacterium *Bacillus subtilis*, the cyanobacterium *Anacystis nidulans*, and the archaebacterium *Methanococcus vannielii* were also included. Alignment scores were also calculated using higher and lower gap penalties (data not shown).

Within the family Pasteurellaceae, *Pasteurella multocida* was most different from all the other organisms examined (I/L score ranging from 0.887 to 0.91), approximately defining the genetic variability in this region between bacteria at the family level (see below). With the exception of *A. (H.) actinomycetemcomitans*, the *Actinobacillus* sequences were remarkably similar, indicating close genotypic relatedness. The sequences of *A. equuli* and *A. lignieresii* differed at only four positions (99% similar), approaching the limits of accuracy of the method (Romaniuk et al., 1987), and could therefore be considered as identical within experimental error. Nonetheless, rechecking the sequencing gels at the nonidentical positions indicated that the sequences were, in fact, different. *A. (H.) actinomycetemcomitans* was more closely related to *Haemophilus* spp. (I/L = 0.994 for *H. influenzae*, and 0.951 for *H. aphrophilus*) than to the other (true) actinobacilli (mean I/L = 0.917), but *H. influenzae* and *H. aphrophilus* were slightly more closely related to each other (I/L = 0.953) than to *A. (H.) actinomycetemcomitans*. Inspection of specific variable positions (Fig. 1) reinforced these observations. At certain positions (e.g. positions 51, 61 and 106 in Fig. 1b), *A. (H.) actinomycetemcomitans* had bases in common with the two *Haemophilus* sequences, which differed from the three other actinobacilli. At other positions, *A. (H.) actinomycetemcomitans* was the same as one or the other *Haemophilus* species (see positions 91 in Fig. 1a, and 107 in Fig. 1b), suggesting divergence at different times from a common ancestor.

Fig. 2 shows dendrograms resulting from cluster analysis of the NAS matrix (gap penalty = 50) given in Table 1. Single linkage, average linkage and complete linkage results are given. Using all three methods, the branching pattern within the facultative anaerobes (*E. coli*, *Proteus vulgaris* and the Pasteurellaceae) was the same, and in all cases the archaebacterium *Methanococcus vannielii* clustered apart from all other species tested, forming the ‘root’ of the tree. As expected, *A. (H.) actinomycetemcomitans* clustered with the members of the genus *Haemophilus*. The remaining three actinobacilli clustered together and separately from *Pasteurella multocida*. Importantly, the tree branching was consistent with previous studies of DNA/DNA hybridization between strains of Pasteurellaceae (Christiansen et al., 1981;
Coykendall et al., 1983; Potts & Berry, 1983; Escande et al., 1984). The Enterobacteriaceae diverged from the Pasteurellaceae at an average NAS of 7.8, and it would be interesting to see if the Vibrionaceae also diverge at a similar level. Interestingly, the genera \textit{Escherichia} and \textit{Proteus} diverge from one another at a level similar to that of \textit{Pasteurella} from \textit{Haemophilus} and \textit{Actinobacillus} (NAS of 8.9), but the genera \textit{Actinobacillus} and \textit{Haemophilus} clustered more closely at an NAS of about 9-3, perhaps justifying their inclusion in an intermediate group or ‘tribe’.

In contrast to the branching pattern for the facultative anaerobes, the relative branching pattern for \textit{Agrobacterium tumefaciens}, \textit{Pseudomonas testosteroni}, \textit{Bacillus subtilis} and \textit{Anacystis nidulans} did not remain constant using the three linkage methods (Fig. 2). The situation did not improve when distance matrices were calculated using higher or lower gap penalties (data not shown). This may be explained by the fact that we used a relatively small number of positions to perform the analysis, and that these sequences were relatively variable, and therefore of more limited use in examining distantly related organisms.

We expect that the continued use of direct rRNA sequencing methods will simplify and improve the taxonomic analysis of a wide variety of micro-organisms. Although accurate phylogenetic analysis of distantly related species requires the analysis of a large number of relatively constant positions, the evolutionary branching patterns of closely related organisms can be quickly established as demonstrated here. The use of these methods is increasing rapidly. For example, similar sequencing studies have recently been reported for the protozoon

Table 1. \textit{Identities per unit length and normalized alignment scores for pairwise comparisons of 16S rRNA sequences of seven Pasteurellaceae and seven other prokaryotic species}

<table>
<thead>
<tr>
<th>Species</th>
<th>TUME</th>
<th>TEST</th>
<th>SUBT</th>
<th>NIDU</th>
<th>VANN</th>
<th>COLI</th>
<th>VULG</th>
<th>MULT</th>
<th>INFL</th>
<th>APER</th>
<th>ACTI</th>
<th>HOMI</th>
<th>EQUU</th>
<th>LIGN</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{Agrobacterium}</td>
<td>7.51</td>
<td>7.48</td>
<td>6.96</td>
<td>5.76</td>
<td>5.71</td>
<td>7.34</td>
<td>7.13</td>
<td>7.03</td>
<td>7.16</td>
<td>7.03</td>
<td>7.03</td>
<td>6.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{tumefaciens}</td>
<td>0.763</td>
<td>0.98</td>
<td>0.84</td>
<td>0.67</td>
<td>0.71</td>
<td>0.78</td>
<td>0.70</td>
<td>0.69</td>
<td>0.64</td>
<td>0.63</td>
<td>0.62</td>
<td>0.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{Pseudomonas}</td>
<td>0.786</td>
<td>0.721</td>
<td>0.76</td>
<td>0.60</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.70</td>
<td>0.69</td>
<td>0.69</td>
<td>0.69</td>
<td>0.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{testosteroni}</td>
<td>0.721</td>
<td>0.709</td>
<td>0.75</td>
<td>0.59</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td>0.66</td>
<td>0.65</td>
<td>0.65</td>
<td>0.65</td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{Bacillus subtilis}</td>
<td>0.602</td>
<td>0.633</td>
<td>0.641</td>
<td>0.56</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.56</td>
<td>0.55</td>
<td>0.55</td>
<td>0.55</td>
<td>0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{Anacystis nidulans}</td>
<td>0.793</td>
<td>0.789</td>
<td>0.745</td>
<td>0.719</td>
<td>0.720</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{Methanococcus}</td>
<td>0.798</td>
<td>0.797</td>
<td>0.773</td>
<td>0.704</td>
<td>0.610</td>
<td>0.932</td>
<td>0.932</td>
<td>0.932</td>
<td>0.932</td>
<td>0.932</td>
<td>0.932</td>
<td>0.932</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{vannielli}</td>
<td>0.761</td>
<td>0.754</td>
<td>0.718</td>
<td>0.682</td>
<td>0.597</td>
<td>0.857</td>
<td>0.836</td>
<td>0.847</td>
<td>0.847</td>
<td>0.847</td>
<td>0.847</td>
<td>0.847</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{Escherichia coli}</td>
<td>0.750</td>
<td>0.751</td>
<td>0.718</td>
<td>0.686</td>
<td>0.584</td>
<td>0.826</td>
<td>0.847</td>
<td>0.887</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{Proteus vulgaris}</td>
<td>0.740</td>
<td>0.761</td>
<td>0.688</td>
<td>0.659</td>
<td>0.594</td>
<td>0.831</td>
<td>0.838</td>
<td>0.901</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{Pasteurella}</td>
<td>0.753</td>
<td>0.742</td>
<td>0.695</td>
<td>0.667</td>
<td>0.607</td>
<td>0.824</td>
<td>0.836</td>
<td>0.892</td>
<td>0.944</td>
<td>0.951</td>
<td>0.951</td>
<td>0.951</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{multocida}</td>
<td>0.741</td>
<td>0.756</td>
<td>0.716</td>
<td>0.686</td>
<td>0.602</td>
<td>0.833</td>
<td>0.861</td>
<td>0.876</td>
<td>0.939</td>
<td>0.939</td>
<td>0.939</td>
<td>0.939</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{Haemophilus}</td>
<td>0.741</td>
<td>0.749</td>
<td>0.721</td>
<td>0.684</td>
<td>0.584</td>
<td>0.840</td>
<td>0.859</td>
<td>0.842</td>
<td>0.932</td>
<td>0.920</td>
<td>0.916</td>
<td>0.972</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{influenzae}</td>
<td>0.736</td>
<td>0.744</td>
<td>0.723</td>
<td>0.684</td>
<td>0.581</td>
<td>0.838</td>
<td>0.854</td>
<td>0.842</td>
<td>0.930</td>
<td>0.923</td>
<td>0.913</td>
<td>0.974</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{Haemophilus}</td>
<td>0.736</td>
<td>0.744</td>
<td>0.723</td>
<td>0.684</td>
<td>0.581</td>
<td>0.838</td>
<td>0.854</td>
<td>0.842</td>
<td>0.930</td>
<td>0.923</td>
<td>0.913</td>
<td>0.974</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{aprhophilus}</td>
<td>0.736</td>
<td>0.744</td>
<td>0.723</td>
<td>0.684</td>
<td>0.581</td>
<td>0.838</td>
<td>0.854</td>
<td>0.842</td>
<td>0.930</td>
<td>0.923</td>
<td>0.913</td>
<td>0.974</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{Actinobacillus}</td>
<td>0.736</td>
<td>0.744</td>
<td>0.723</td>
<td>0.684</td>
<td>0.581</td>
<td>0.838</td>
<td>0.854</td>
<td>0.842</td>
<td>0.930</td>
<td>0.923</td>
<td>0.913</td>
<td>0.974</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{actinomyctetemcomitans}</td>
<td>0.736</td>
<td>0.744</td>
<td>0.723</td>
<td>0.684</td>
<td>0.581</td>
<td>0.838</td>
<td>0.854</td>
<td>0.842</td>
<td>0.930</td>
<td>0.923</td>
<td>0.913</td>
<td>0.974</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
'Actinobacillus hominis'
P. J. CHUBA AND OTHERS

Average linkage

Single linkage

Complete linkage

Fig. 2. Dendrograms illustrating the results of cluster analysis of the normalized alignment score matrix from Table 1. The lengths of the branches are equivalent to the normalized alignment scores given on the scale at the bottom.

Toxoplasma gondii (Johnson et al., 1987) and the prokaryotic genus Campylobacter (Lau et al., 1987; Romanuk et al., 1987). The PC version of the Clustan program allows cluster analysis of sequence data on a microcomputer and this will make the rRNA sequencing approach less expensive and more widely accessible to taxonomists. Aside from taxonomic and evolutionary applications, the rapid accumulation of rRNA sequence data is also important for the
production of specific oligonucleotide probes complementary to variable regions of rRNA (Göbel et al., 1987b), which may prove to have wide applications in diagnostic medicine.

Our thanks go to Professor W. Mannheim and to Bodo Rak for their assistance and encouragement. Financial support for this project was provided by the Bundesministerium für Forschung und Technologie.

REFERENCES


