Increased level of intragenomic 16S rRNA gene heterogeneity in commensal strains closely related to *Haemophilus influenzae*

Niels Nørskov-Lauritsen

Department of Clinical Microbiology, Aarhus University Hospital Skejby, DK-8200 Aarhus N, Denmark

The 16S rRNA gene sequence of strains closely related to, but excluded from, *Haemophilus influenzae* was investigated and a conspicuously high number of polymorphic nucleotide positions due to intragenomic 16S rRNA gene heterogeneity was observed. The average frequency of 16S rRNA gene polymorphic nucleotide positions in 31 variant strains was $7.0 \times 10^{-3}$, which is approximately ten times the level observed in validated strains of *H. influenzae*. Sixty-seven polymorphic nucleotide positions in seven strains most likely originated from the simultaneous presence of two distinct types of helix 18 as a consequence of prior recombinatorial events. The increased level of 16S rRNA gene polymorphism in commensal taxa excluded from the pathogenic species *H. influenzae* is unexplained. The heterogeneity imposes difficulties on rRNA gene-based classification and systematics.

**INTRODUCTION**

*Haemophilus influenzae* is an important human pathogen commonly involved in infections of the respiratory tract and sometimes in invasive infections (Murphy, 2010). The bacterium is part of the normal bacterial flora of the mucous membranes of the human upper respiratory tract, particularly in children, in whom it can be cultured from mucous membranes of the human upper respiratory tract, but is not associated with exacerbations in chronic obstructive pulmonary disease (Murphy et al., 2007).

rRNA sequences, particularly the 16S rRNA gene, are widely used as the basis for classification and identification of prokaryotes (Tindall et al., 2010; Woo et al., 2008). Indeed, the rRNA genes combine several properties which make them uniquely suited for this purpose (Acinas et al., 2004): they are universally distributed, allowing the comparison of phylogenetic relationships between all extant organisms and thus the construction of a ‘tree of life’ (Pruesse et al., 2007); their conserved nature is generally presumed to provide a solid framework for assessment of evolutionary lineages (Woese, 1987; Ludwig & Klenk, 2001); and their highly conserved regions enable the design of almost universal PCR primers within the bacterial domain (Wang & Qian, 2009).

In prokaryotes, the genes encoding the three rRNAs (16S, 23S and 5S) are usually linked into an operon, *rrn*, that may be present in multiple copies in the bacterial genome. The number of copies is typically constant within a species, although exceptions occur (Lee et al., 2009). Microbes with few *rrn* operons tend to be oligotrophic (i.e. capable of growth in environments that are poor in nutrients) and slow-growing, while those with many copies grow more rapidly in response to favourable growth conditions (Klappenbach et al., 2000; Lee et al., 2009). Mutants of *Bacillus subtilis* in which nine of ten *rrn* operons were deleted exhibited increased doubling times and extremely reduced sporulation rates (Nanamiya et al., 2010).

Intragenomic heterogeneity between multiple rRNA genes has been confirmed by genome sequencing and other methods (Klappenbach et al., 2001; Coenye & Vandamme, 2003; Acinas et al., 2004). Exceptional genomes show extreme divergence (>5%) among 16S rRNA genes, possibly indicating horizontal rRNA gene transfer between divergent genomes. The large majority of 16S rRNA sequences from the same genome are identical or display high similarities (1–5 polymorphic nucleotide positions per 16S rRNA gene), while a minority display a nucleotide divergence of ~1% (10–20 polymorphisms per 16S rRNA gene). Intragenomic diversity is not associated with
particular phyligenetic groups (Klappenbach et al., 2001; Coenye & Vandamme, 2003; Acinas et al., 2004).

In a previous study an attempt was made to delineate the borders of *H. influenzae* by comparing reference strains of this species to closely related bacteria, including strains assigned to *H. haemolyticus*, ‘cryptic genospecies biotype IV’ and the unvalidated ‘*Haemophilus intermedius*’ (Norskov-Lauritsen et al., 2009). Multilocus sequence analysis successfully encircled *H. influenzae* while 16S rRNA gene sequence comparison was of low resolution, partly due to a conspicuously high number of polymorphic sites in many of the strains that did not belong to *H. influenzae*. In the present study the increased intragenomic 16S heterogeneity in variant strains excluded from *H. influenzae* has been characterized.

**METHODS**

**Strains.** The 42 study strains are described in detail by Norskov-Lauritsen et al. (2009). The type strain (NCTC 8143) and eight other reference strains of *H. influenzae* were chosen to represent different evolutionary lineages of this species. The type strain of *H. haemolyticus* (NCTC 10659) and 32 other strains were chosen to challenge the borders of *H. influenzae*. The latter group included seven haemolytic strains and eleven strains capable of porphyrin synthesis, phenotypic characters that are considered incompatible with a classification within *H. influenzae* (Kilian, 2005). Multilocus housekeeping gene sequence analysis, phenotypic tests and presence of marker genes indicated that two of the challenge strains should be classified with *H. influenzae*, while 31 strains should be excluded from the species (Norskov-Lauritsen et al., 2009).

**16S rRNA gene sequencing and identification of polymorphic nucleotide positions.** 16S rRNA genes were amplified by PCR using HotStarTaq DNA polymerase (Qiagen) and sequenced using multiple separate primers as described by Kuhnert et al. (2002). All electropherograms were carefully inspected for the presence of polymorphic positions (‘double peaks’). Polymorphic positions were only accepted if they were documented by a minimum of two sequencing reactions using separate sequencing primers. For phylogenetic analysis, a 1362 nt fragment (1361 nt in HK 2067 and HK 2122) corresponding to nucleotides 27–1388 of the 16S rRNA gene in strain Rd (Fleischmann et al., 1995) was aligned and compared. The previously deposited sequences for *Haemophilus aegyptius* strain CCUG 25716 (nucleotide sequence accession number AY362905), the cryptic genospecies biotype IV strain S32F2 (AY962106) and strain Rd were used. *H. parainfluenzae* T3T1 (http://www.sanger.ac.uk/resources/downloads/bacteria/haemophilus.html) was used as outgroup.

**Separation of intragenomic 16S rRNA genes.** To reveal sequences attributable to a single RNA operon, genomic DNA was digested with the homing endonuclease I-Ceu, which recognizes a 26 bp site in the 23S rRNA genes and fragments the DNA according to the *rrn* copy number. Fragments were separated by pulsed-field gel electrophoresis, excised from the gel, purified on silica-gel columns and subjected to 16S rRNA gene PCR and sequencing.

**Sequence analysis and rRNA helix numbering.** DNA sequences from both strands were edited, assembled and aligned using software from DNASTAR. Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 4 (Tamura et al., 2007) by the neighbour-joining method. Bootstrap tests were performed with 500 replicates. The *Escherichia coli* 16S rRNA helix numbering system of Wuyts et al. (2001) was used; the synonymy with the Gutell Lab numbering system (http://www.rna.ccb.iupui.edu/), where each helix is numbered with the 5’ nucleotide of the initial base pair, is as follows: helix 10, H1198; helix 18, H455; helix 29, H829; helix 37, H996; and helix 37b1, H11006.

DNA sequences of this study are deposited with GenBank under accession numbers EU909638–EU909680.

**RESULTS**

**Phylogeny of the study strains as revealed by 16S rRNA gene comparison**

In Fig. 1 is shown a neighbour-joining dendrogram comparing near-full-length (1362 nt) 16S rRNA gene sequences obtained from the 42 study strains, using *H. parainfluenzae* strain T3T1 as outgroup. The final classification of strains within or outside the border of the species *H. influenzae* was based on both phenotypic and genotypic evidence, with particular emphasis on multilocus housekeeping gene phylogeny (see Norskov-Lauritsen et al., 2009, for details). Six of our strains, including the type strain of *H. influenzae* and the full genome-sequenced reference strain Rd, belonged to phylogenetic group I (blue in Fig. 1; phylogenetic grouping according to Meats et al., 2003). Three *H. influenzae* reference strains represented the more rarely encountered phylogenetic group II (yellow in Fig. 1), which encompasses strains of serotypes e and f, as well as other distantly related evolutionary lineages. A single strain (CCUG 30048, grey in Fig. 1) was located on the border of the species but should probably be included in *H. influenzae* (Norskov-Lauritsen et al., 2009; Ridderberg et al., 2010). Finally, 31 study strains were excluded from *H. influenzae* (red in Fig. 1). The excluded strains were phenotypically and genotypically diverse and only some of them could be allocated to *H. haemolyticus* by current definitions (Kilian, 2005). They will collectively be referred to as ‘variant strains’ but may in fact constitute a number of unnamed species, exemplified by the four closely related representatives of the so-called ‘cryptic genospecies biotype IV’ (encircled in Fig. 1).

Only a modest bootstrap value of 70% supported the *H. influenzae* cluster (including strain CCUG 30048) when 16S rRNA gene sequences were compared (Fig. 1). This is in contrast to a bootstrap value of 98% in support of the same cluster when six individual housekeeping gene sequences are concatenated and compared (Norskov-Lauritsen et al., 2009). The inferior resolution of 16S rRNA is caused by less variability, shorter length (1362 vs 2712 nt) and the presence of polymorphic positions in the 16S rRNA genes.

**Interoperonic heterogeneity of 16S rRNA genes in the study strains**

Sequencing of 16S rRNA genes from bacterial species with multiple RNA operons will occasionally display unresolved...
nucleotides when genomic DNA is used for PCR. In the present study, an unexpectedly high number of polymorphic nucleotide positions were observed with several of the variant strains. To exclude the possibility of contaminated cultures, the sequencing was repeated after three times repeated subculturing of single colonies from strains with 16S rRNA gene heterogeneity. Of 31 variant strains, no polymorphic positions were found in eleven strains, ten strains harboured 1–9 polymorphisms (0.07–0.6 %), and ten strains harboured 14–36 polymorphic nucleotide positions in the 16S rRNA genes (1.0–2.6 %). A total of 296 polymorphic nucleotide positions were observed in the 31 variant strains, corresponding to an average frequency of $7.0 \times 10^{-3}$ in the 16S rRNA gene. Four of 11 strains of H. influenzae had 16S rRNA gene polymorphism: three strains with polymorphism at one nucleotide position, and one strain (CCUG 30048) with polymorphism at eight positions. This corresponds to an average frequency of $0.73 \times 10^{-3}$ polymorphic nucleotide positions in the 16S rRNA gene of the 11 strains of H. influenzae. This frequency is similar to the results of Sacchi et al. (2005), who investigated 65 reference strains of H. influenzae and found 83 polymorphic positions within the entire 16S rRNA gene (1538 nt), or 78 sites within the 1362 nt analysed in this study (calculated from nucleotide sequence accession numbers AY613445 to AY613775). This corresponds to a frequency of $0.87 \times 10^{-2}$, which is significantly lower than the value observed with the variant strains ($P=0.0003$, Student’s unpaired t-test).

With respect to the 20 variant strains with intragenomic 16S rRNA gene heterogeneity, the polymorphism occurred at 79 nucleotide positions (5.8 % of 1362). At 14 positions, only one strain exhibited polymorphism, while 65 positions were polymorphic in 2–10 strains. When positions were polymorphic in more than one strain, the same polymorphism was usually encountered (50 of 65 positions). For example, eight strains exhibited polymorphism at nucleotide T$_{182}$ (strain Rd numbering), and in all cases the simultaneous presence of both thymidine and adenine was observed. At 11 positions, two separate polymorphisms were seen, while three different polymorphisms were found at four positions; for example, at nucleotide A$_{760}$, the dichotomies C/A, T/A and C/T were represented by two, three and two strains, respectively. At only a single position in a single strain, the ambiguity could not be reduced to a dichotomy despite repeated sequencing; this strain probably harbours 16S rRNA genes with more than two different nucleotides at this position.

To further verify the intergenic 16S rRNA gene heterogeneity of selected strains, individual genes were obtained after cleavage of genomic DNA with the homing endonuclease I- CeuI. Two discrete 16S rRNA genes, termed rrsA and rrsB, from strains CCUG 17210 and CCUG 18082 (deposited in GenBank with accession numbers EU909652–3 and EU909656–7, respectively), were included in the phylogenetic comparison (Fig. 1). A total of 24 16S rRNA gene polymorphisms were revealed in strain CCUG 17210 when genomic DNA was used for PCR, and 21 of these were resolved by the two individually sequenced rrs genes. A total of 25 16S rRNA gene polymorphisms were revealed in strain CCUG 18082 when genomic DNA was used for PCR, and 23 of these were resolved by the two individually sequenced rrs genes.

The polymorphic positions were not evenly distributed in the 16S rRNA gene. In Fig. 2 is shown the distribution of polymorphic positions in the 20 variant strains of H. influenzae and in 32 strains of H. influenzae (described by Sacchi et al., 2005). The majority of the polymorphic sites in the variant strains, 220 of 296, occurred in helices 10, 18, 29, 37 and 37/b1 (16S rRNA helix numbering system according to Wuyts et al., 2001). A clustering of polymorphic positions between nucleotides 182–278, and to a lesser extent in helix 18, was observed for both validated and variant strains, while polymorphisms in helices 29, 37 and 37/b1 were almost exclusively observed in variant strains (Fig. 2).
16S rRNA helix 18

Three distinct types of 16S rRNA helix 18 were present in the material, represented by Rd (the full genome-sequenced reference strain of *H. influenzae*), 16N (a reference strain of ‘cryptic genospecies biotype IV’) and NCTC 10659\(^T\) (the type strain of *H. haemolyticus*) (Fig. 3). Within the region of 23 nt that encodes the outer stem of helix 18, seven variant strains had 8, 9 or 11 polymorphic positions in the 16S rRNA genes. When genomic DNA of CCUG 17210 was used for PCR and sequencing, eight polymorphic positions in the helix 18 region were revealed. However, the helix 18 gene sequences of *rrsA* and *rrsB* of strain CCUG 17210 were identical to those of strains 16N and NCTC 10659\(^T\), respectively, which exactly accounts for strain CCUG 17210 was used for PCR and sequencing, eight polymorphic positions in the 16S rRNA genes. When genomic DNA of Rd and 16N are represented in the genome; and strain CCUG 36040 exhibited the eleven polymorphic positions that will result from the simultaneous presence of helix 18 alleles from both Rd and 16N are represented in the genome; and strain CCUG 36040 exhibited the eleven polymorphic positions that will result from the simultaneous presence of helix 18 alleles from both NCTC 10659\(^T\) and Rd (Fig. 3). Thus, 67 polymorphisms (23% of the total) were most likely explained by the simultaneous presence of two separate types of helix 18 sequences in seven variant strains.

**DISCUSSION**

In a characterization of *H. influenzae* and closely related taxa, 16S rRNA genes were investigated by subjecting genomic DNA to PCR and sequencing. A conspicuously high number of 16S rRNA gene polymorphic nucleotide positions was found when electropherograms were inspected for presence of ‘double peaks’, or ambiguously determined nucleotides. This method has previously been used for documentation of 16S rRNA gene polymorphism in *H. influenzae* (Sacchi *et al.*, 2005), but it can be argued that the method is subject to inter-observer variation. Several lines of evidence do, however, confirm that our data are reproducible. First, an array of 16S rRNA sequencing primers are available for the family *Pasteurellaceae* (Kuhnert *et al.*, 2002), and polymorphic positions were validated by a minimum of two sequencing reactions using separate sequencing primers. Second, the position of the polymorphism, and the combination of nucleotides involved, was usually encountered in more than one strain. Third, a significant number of polymorphic positions could be inferred by the presumed compounding of distinct helix 18 alleles present in other strains. Finally, we were able to resolve the majority of polymorphic positions in two strains (CCUG 17210 and CCUG 18082) by isolation and sequencing of two discrete 16S rRNA genes from each strain (termed *rrsA* and *rrsB*, shown in Fig. 1). Fragmentation of genomic DNA according to *rrn* copy number by cleavage with *Ceu*I, and subsequent isolation of individual 16S rRNA genes on non-gradient gels, has been successfully accomplished with bacteria with low rRNA gene copy number (Marchandin *et al.*, 2003; Kilian *et al.*, 2008). The location and orientation of the six RNA operons in strain Rd preclude the isolation of all 16S rRNA genes by *Ceu*I cleavage, but the method resolved the majority of polymorphic positions in selected strains. Separation of PCR amplicons by density gradient gel electrophoresis (DGGE) or temporal temperature gel electrophoresis (TTGE) are other methods used to

![Fig. 3. The three 16S rRNA helix 18 types present in the study strains (outer stem of helix 18, nucleotides 453–477 (strain Rd numbering). Conserved nucleotides are in bold.](image)

**Fig. 2.** Distribution of polymorphic nucleotide positions in the 16S rRNA gene. Columns represent the number of strains with 16S rRNA gene heterogeneity at individual positions between nucleotides 27–1388 (strain Rd numbering). Top, 296 polymorphic positions in 20 variant strains (this study). Bottom, 78 polymorphic positions in 32 strains of *H. influenzae* (described by Sacchi *et al.*, 2005). The locations of helices 10, 18, 29 and 37+37b1 are marked. The resolution of the figure does not allow the discrimination of columns at neighbouring nucleotides.
document multiple-copy gene heterogeneity, but gradient gel-based methods also have drawbacks (Hestekin & Barron, 2006; Michon et al., 2010). Although single polymorphic positions in our material may have escaped detection, the disclosure of polymorphisms in the present report is robustly verified by the use of several experimental designs.

rRNA sequences continue to be an indispensable part of the validation of new species (Tindall et al., 2010). In this study, the comparison of 16S rRNA gene sequences was able to differentiate strains of H. influenzae from neighbouring taxa, although the bootstrap support of the H. influenzae cluster was modest. Of the taxa excluded from H. influenzae, only H. haemolyticus has been validly described, but the delineation of this species is not clear-cut (Nørskov-Lauritsen et al., 2009). It is likely that the diverse group of ‘variant strains’ used to examine the borders of H. influenzae is composed of several distinct taxa that could merit species recognition. A resolution of ‘variant strains’ into new species could probably not be accomplished by 16S rRNA gene comparison. This is indicated by the discrete locations taken up by CCUG 17210 rrsA and rrsB, intermingled with the cluster of ‘cryptic genospecies biotype IV’ and the type strain of H. haemolyticus (Fig. 1). Similar problems with classifications based on 16S rRNA gene sequences have been documented for the genera Nocardia (Conville & Witebsky, 2007), Veillonella (Marchandin et al., 2003; Michon et al., 2010) and Streptococcus (Kilian et al., 2008).

There was a significant difference in the level of 16S rRNA gene polymorphism between H. influenzae and neighbouring taxa. This difference may be obscured by the mixing of presumed distinct taxa into the group of variant strains. We included four representatives of the so-called ‘cryptic genospecies biotype IV’, and no member of this group exhibited 16S rRNA gene heterogeneity. These strains have been isolated from neonatal and post-partum infections (Quentin et al., 1996), in contrast to Haemophilus commensal species of the upper respiratory tract.

The biological consequences of rRNA heterogeneity are largely unknown. It has been hypothesized that many prokaryotes evolved multiple RNA operons with functional differentiation to cope with a variety of environmental conditions (Condon et al., 1995). One observation along this line is the recent demonstration that B. subtilis mutants with only one functional rrr operon expressed variable and operon-specific sporulation rates (Nanamiya et al., 2010). Strains of H. influenzae are commonly isolated from infections, in contrast to Haemophilus commensal species (Murphy, 2010). The observed difference in rRNA gene composition could reflect separate, adaptive strategies, where the rRNA homogeneity of H. influenzae may render this species fit for a narrow niche in the host, whereas the rRNA heterogeneity of variant strains could be advantageous in responding to a wider variety of growth conditions. Different levels of rRNA polymorphism have also been documented within the genus Streptococcus, where 16S rRNA gene heterogeneity has been documented for Streptococcus oralis but not for Streptococcus mitis/ pneumonaias (Kilian et al., 2008). However, the hypothesis needs substantiation from further comparisons of pathogenic species with related, commensal taxa, or directly from experimental assessment of virulence expressed by organisms with manipulated rrn operons.

REFERENCES


Edited by: W. Liesack