FEMS Microbiology Letters 228 (2003) 45–49

FEMS MICROBIOLOGY Letters

www.fems-microbiology.org

Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes

Tom Coenye *, Peter Vandamme

Laboratorium voor Microbiologie, Ghent University, K.L. Ledeganckstraat 35, B-9000 Ghent, Belgium

Received 1 September 2003; received in revised form 15 September 2003; accepted 15 September 2003

First published online 3 October 2003

Abstract

The availability of a large number of completely sequenced bacterial genomes allows the rapid and reliable determination of intragenomic sequence heterogeneity of 16S rRNA genes. In the present study we assessed the intragenomic sequence heterogeneity of 16S rRNA genes in 55 bacterial genomes, representing various phylogenetic groups. The total number of rRNA operons in genomes included ranged from 2 to 13. The maximum number of nucleotides that were different between any pair of 16S rRNA genes within a genome ranged from 0 to 19. The corresponding minimal similarity ranged from 100 to 98.74%. This indicates that the intragenomic heterogeneity between multiple 16S rRNA operons in these genomes is rather limited and is unlikely to have a profound effect on the classification of taxa. Among the multiple copies of the 16S rRNA genes present in the genomes included, 199 mutations were counted with transitions being the dominant type of mutations over the total length of the 16S rRNA gene. Most heterogeneity occurred in variable regions V1, V2, and V6.

© 2003 Federation of European Microbiological Societies. Published by Elsevier B.V. All rights reserved.

Keywords: 16S rRNA microheterogeneity; Whole-genome sequence; Taxonomy

1. Introduction

The comparison of 16S ribosomal RNA gene sequences to infer phylogenetic relationships among bacteria has been widely used for several decades (see for example references [1,2]). 16S rRNA is generally accepted as the ultimate molecular chronometer because it is functionally constant, shows a mosaic structure of conserved and more variable regions, occurs in all organisms and because its length allows easy sequencing [1,3,4]. Nevertheless, it has been shown that the resolution of the 16S rRNA gene is often too low to allow the differentiation of closely related species [3,5]. It was also shown that there may be considerable intraspecific variation in 16S rRNA sequence [6]. Part of this intraspecific diversity is caused by the fact that rRNA genes are often organised as part of a multigene family, with the copy number ranging from 1 to 15 [7]. The rRNA operon copy number reflects the ecological strategy of the organism as there seems to be a correlation between the response rate of the organism to changing conditions and the rRNA operon copy number [8]. In general, members of multigene families tend to coevolve [9,10], but the ultimate degree of sequence polymorphism within the family will depend on the frequency of molecular interaction mechanisms such as gene conversion [9,11]. So far, little attention has been paid to the systematic study of variability in multiple copies of the 16S rRNA gene, although this now has been reported for organisms belonging to various major bacterial lineages, including the *Proteobacteria*, the *Firmicutes* and the *Actinobacteria* (for an overview see reference [12]). Overall, intragenomic sequence heterogeneity seems to be relatively low, although values of up to 6.5% have been reported for some actinomycetes [13,14]. Most studies regarding intragenomic sequence heterogeneity of 16S rRNA genes have relied on cloning and sequencing of the individual genes, although separation of the multiple copies by DGGE (denaturing gradient gel electrophoresis) [15] or TGGE (temperature gradient gel electrophoresis) [16] followed by sequencing has been used as well. The availability of an increasing number of completely sequenced bacterial genomes allows for the rapid determination of intragenomic sequence heterogeneity of 16S rRNA genes without the need for further experimental work. In this study we

* Corresponding author. Tel.: +32 (9) 264 5114; Fax: +32 (9) 2645092. *E-mail address:* tom.coenye@ugent.be (T. Coenye).

0378-1097/03/$22.00 © 2003 Federation of European Microbiological Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/S0378-1097(03)00717-1

Downloaded from https://academic.oup.com/femsle/article/228/1/45/571392 by guest on 19 April 2024

Table 1
Whole-genome sequences used in this study

Species and strain designation	GenBank accession no.	rRNA copy no.	Max. difference (nucleotides)	Min. similarity (%)	Max. difference in free energy (kcal/mol)
α-Proteobacteria					
Agrobacterium tumefaciens C58	AE007869, AE007870	4	0	100	0
Brucella melitensis 16M	AE008917, AE008918	3	0	100	0
Caulobacter crescentus CB15	AE005673	2	0	100	0
Mesorhizobium loti MAFF303099	BA000012	2	0	100	0
Sinorhizobium meliloti 1021	AL591688	3	0	100	0
β-Proteobacteria					
Bordetella pertussis Tohama I	NC_002929	3	0	100	0
Burkholderia cenocepacia J2315*		6	3	99.80	1.2
Burkholderia pseudomallei K96243*		4	1	99.93	1.0
Neisseria meningitidis MC58	AE002098	4	0	100	0
Ralstonia solanacearum GMI1000	AL646052, AL646053	4	0	100	0
γ-Proteobacteria					
Escherichia coli O157:H7 Sakai	BA000007	7	11	99.29	1.1
Haemophilus influenzae Rd	L42023	6	0	100	0
Pseudomonas aeruginosa PAO1	AE004091	4	1	99.94	3.7
Pseudomonas putida KT2440	AE015451	7	3	99.81	3.3
Pseudomonas syringae DC300	NC_004578	5	0	100	0
Shewanella oneidensis MR-1	AE014299	9	4	99.74	2.5
Salmonella enterica CT18	AL513382	7	2	99.87	9.3
Vibrio parahaemolyticus RIMD 2210633	NC_004603, NC_004605	11	5	99.67	2.8
Xanthomonas axonopodis 306	AE008923	2	0	100	0
Xanthomonas campestris ATCC 33913	AE008922	2	0	100	0
Xylella fastidiosa 9a5c	AE003849	2	0	100	0
ε-Proteobacteria					
Campylobacter jejuni NCTC 11168	AL111168	3	0	100	0
Helicobacter pylori J99	AE001439	2	1	99.93	0.3
Wolinella succinogenes DSM 1740	NC_005090	2	0	100	0
Spirochaetacea					
Leptospira interrogans 56601	NC_004342	2	0	100	0
Treponema pallidum Nichols	AE000520	2	0	100	0
Firmicutes					
Bacillus anthracis Ames	NC_003997	11	5	99.67	1.6
Bacillus cereus ATCC 14579	NC_004722	13	3	99.81	8.0
Bacillus halodurans C-125	BA000004	8	5	99.66	1.9
Bacillus subtilis 168	AL009126	10	12	99.23	13.1
Clostridium acetobutylicum ATCC 824	AE001437	10	12	99.21	5.7
Clostridum perfringens 13	BA000016	10	19	98.74	6.5
Enterococcus faecalis V583	NC_004668	4	1	99.94	2.3
Lactococcus lactis IL1403	AE005176	6	1	99.94	3.4
Lactobacillus plantarum WCFS1	AL935263	5	2	99.87	0.4
Listeria innocua Clip11262	AL592022	6	4	99.74	4.5
Listeria monocytogenes EGD-e	NC_003210	6	4	99.74	2.8
Oceanobacillus iheyensis HTE831	BA000028	7	17	98.92	7.4
Staphylococcus aureus Mu50	BA000017	5	4	99.75	2.6
Staphylococcus epidermidis ATCC 12228	AE015929	5	11	99.29	4.8
Streptococcus agalactiae 2603V/R	AE009948	7	0	100	0
Streptococcus mutans UA159	AE014133	5	3	99.81	0.2
Streptococcus pneumoniae TIGR4	AE005672	4	0	100	0
Streptococcus pyogenes MGAS8232	AE009949	6	0	100	0
Actinobacteria					
Bifidobacterium longum NCC2705	AE014295	4	0	100	0
Streptomyces avermitilis MA-4680	BA000030	6	0	100	0
Streptomyces coelicolor A3(2)	AL645882	6	3	99.80	9.5

Downloaded from https://academic.oup.com/femsle/article/228/1/45/571392 by guest on 19 April 2024

Table 1 (*Continued*).

Species and strain designation	GenBank accession no.	rRNA copy no.	Max. difference (nucleotides)	Min. similarity (%)	Max. difference in free energy (kcal/mol)
Cytophaga–Flavobacterium–Bacteroides group					
Bacteroides thetaioatomicron VPI-5482	NC_004663	5	18	98.92	36.0
Porphyromonas gingivalis W83	NC_002950	4	0	100	0
Cyanobacteria					
Nostoc sp. PCC7120	NC_003272	4	1	99.93	0
Synechocystis sp. PCC6803	NC_000911	2	0	100	0
Other taxa					
Aquifex aeolicus VF5	NC_000918	2	0	100	0
Chlorobium tepidum TLS	NC_0023932	2	2	100	0
Deinococcus radiodurans R1	AE000513	3	2	99.87	1.9
Fusobacterium nucleatum ATCC 25586	NC_003454	5	2	99.86	3.3

*These sequence data were produced by the Wellcome Trust Sanger Institute and can be obtained from their website (http://www.sanger.ac.uk/).

have assessed the intragenomic sequence heterogeneity of 16S rRNA genes in 55 bacterial genomes, representing various phylogenetic groups.

2. Materials and methods

2.1. Whole-genome sequence data

The complete genome sequences used in this study are shown in Table 1. They were downloaded from the GenBank database (http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/micr.html) or were obtained from the Wellcome Trust Sanger Institute website (http://www.sanger.ac.uk/).

2.2. Extraction of 16S rRNA gene sequences and phylogenetic analysis

Whole-genome sequences were downloaded, imported in the Kodon 2.0 (Applied Maths) software package and 16S rDNA sequences were extracted. If 16S rRNA genes were not annotated they were located in the genome sequence using BLAST [17]. Multiple 16S rRNA genes extracted from the same genome sequence were aligned and a similarity matrix was constructed using Kodon 2.0.

2.3. Determination of secondary structure and minimal free energy

Secondary structures of all 16S rRNA genes were obtained through the *mfold* webserver [18] using the free energy data from Mathews et al. [19]. The conditions for folding were the standard conditions (37°C, 1 M NaCl, no divalent ions), which are equivalent to physiological conditions [18]. *mfold* was also used to calculate the mimimal free energy, ΔG^0, which is a measurement of the stability of the secondary structure.

2.4. Statistical analysis

Statistical analyses were performed using the SPSS 11.0.1 software package.

3. Results

The rRNA copy number, the maximum pairwise difference in nucleotides between any pair of 16S rRNA operons, the corresponding minimal pairwise similarity and the maximal difference in free energy between any pair of 16S rRNA operons for each genome are given in Table 1. The total number of rRNA operons in genomes included in the present study ranged from 2 to 13 (mean ± standard deviation 5.07 ± 2.73). The maximum number of nucleotides that were different between any pair of 16S rRNA genes within a genome ranged from 0 to 19 (mean ± standard deviation 2.91 ± 4.78). The corresponding minimal similarity ranged from 100 to 98.74% (mean ± standard deviation 99.81 ± 0.31). The maximal difference in free energy between two 16S rRNA operons from the same genome was between 0 and 36.0 kcal/mol (mean ± standard deviation 2.57 ± 5.47). The distribution of different types of mutations (transitions [A ↔ G and C ↔ T], transversions [purine ↔ pyrimidine] and insertions/deletions) in multiple copies of the 16S rRNA gene was mapped. Among the multiple copies of the 16S rRNA genes present in the 55 bacterial genomes included in this study, 199 mutations were counted (insertion/deletions at the 5′ or 3′ end of the 16S rRNA gene were not included as they may result from errors in predicting the correct borders of the 16S rRNA gene, especially in not-annotated genomes for which BLAST was used to identify the gene). These mutations included 125 transitions (62.8%), 52 transversions (26.1%) and 22 insertions/deletions (11.1%). Transitions are the dominant type of mutations over the total length

Downloaded from https://academic.oup.com/femsle/article/228/1/45/571392 by guest on 19 April 2024

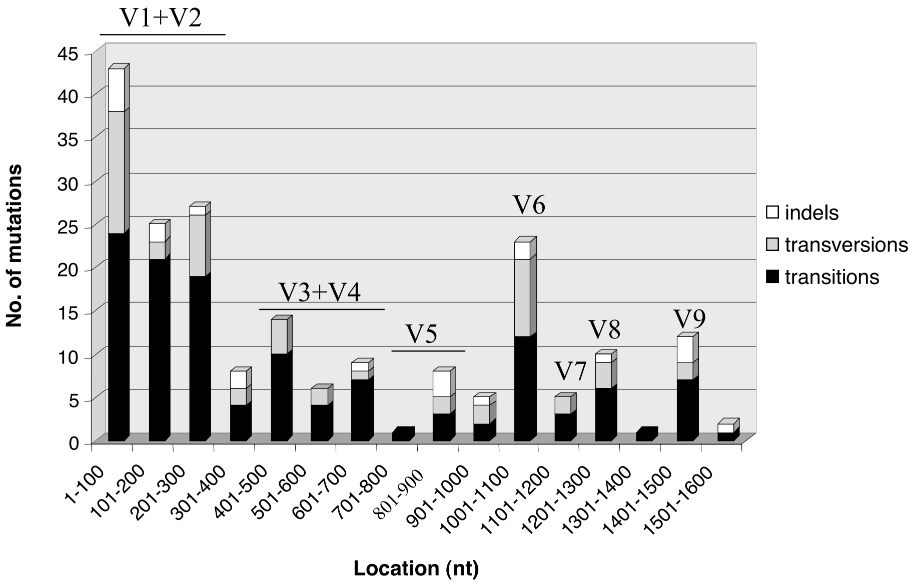

Fig. 1. Distribution of different types of mutations in multiple copies of the 16S rRNA genes of sequenced bacterial genomes. The location of the nine variable regions (V1–V9) is indicated above the bars.

of the 16S rRNA gene. Insertions and deletions are rare and are even totally absent from regions 401–600, 701–800, 1101–1200 and 1301–1400. The majority of all mutations were located in the first 300 nucleotides of the 16S rRNA gene ($n = 95$, 47.7%) and between positions 1001 and 1100 ($n = 23$, 11.6%) (Fig. 1).

4. Discussion

The results of the present study confirm that the number of rRNA operons is indeed not strictly correlated with phylogeny [7,8], although the mean number is slightly higher in Gram-positive organisms than in Gram-negative organisms (6.64 vs. 3.90, $P = 0.001$). However, it should be noted that the number of taxa investigated is relatively low and that this may influence the statistical analysis. Our data clearly indicate that the intragenomic heterogeneity between multiple 16S rRNA operons in a genome is rather limited. This is in agreement with a preliminary study in which it was shown that the maximum intragenomic 16S rRNA operon diversity within 14 bacterial and archeal genomes was between 0 and 1.23% [12] and with several other studies in which individual 16S rRNA operons were cloned and sequenced (see references [20–22] for recent examples). There are however examples in which extensive sequence diversity has been reported between multiple 16S rRNA operons within a genome, especially in the actinomycetes; this heterogeneity is most likely caused by recombination and/or horizontal transfer [13,14]. There is no significant difference in intragenomic diversity between the different phylogenetic groups, although the homogeneity of the 16S rRNA operons in the α-*Proteobacteria* is remarkable. We included several organisms of which the genomes consist of multiple rRNA-containing replicons (*Agrobacterium tumefaciens*, *Brucella melitensis*, *Burkholderia cenocepacia*, *Burkholderia pseudomallei*, *Ralstonia solanacearum* and *Vibrio parahaemolyticus*). We found that the diversity between 16S rRNA operons located on different replicons was not higher than the diversity between 16S rRNA operons located on the same replicon; the16S rRNA operons of *A. tumefaciens* and *B. melitensis* located on different replicons are even identical. To evaluate the impact of the sequence heterogeneity on the stability of the 16S rRNA we calculated the difference in free energy between all pairs of 16S rRNA operons within a genome. As could be expected these differences were also rather limited, indicating that, at least for the genomes included in the present study, there are no significant differences in secondary structure.

When we mapped the distribution of all mutations, it was obvious that this distribution was unequal ($P < 0.01$). Most heterogeneity occurred in variable regions V1, V2, V6, and, to a lesser extent, V3 and V4 (Fig. 1). When we compared the distribution with known distribution of substitution rates and with secondary and tertiary structure models of rRNA and ribosomes [23,24], it became obvious that most variability occurred in (i) regions which are known to have high intrataxon diversity, (ii) regions that are located further away from the centre in an assembled ribosome, and (iii) regions that are not (or only to a lesser extent) involved in tertiary structure interactions. The location of the differences in the most variable part of the 16S rRNA corroborates that the differences are true differences, and not mere sequencing errors.

Our data indicate that, although there is heterogeneity among multiple 16S rRNA operons in bacterial genomes, in general this heterogeneity is rather limited and is unlikely to have a profound effect on the classification of taxa. Although there are several mechanisms that can

Downloaded from https://academic.oup.com/femsle/article/228/1/45/571392 by guest on 19 April 2024

cause heterogeneity within a multigene family, it seems that there is a strong pressure to maintain a high level of sequence conservation among multiple copies of 16S rRNA genes, probably because of functional and structural constraints.

Acknowledgements

T.C. and P.V. are indebted to the Fund for Scientific Research – Flanders (Belgium) for a position as postdoctoral fellow and research grants, respectively. T.C. also acknowledges the support from the Belgian Federal Government (Federal Office for Scientific, Technical and Cultural Affairs).

References

[1] Woese, C.R. (1987) Bacterial evolution. Microbiol. Rev. 51, 221–271.

[2] Garrity, G.M. and Holt, J.G. (2001) The road map to the manual. In: Bergey's Manual of Systematic Bacteriology, 2nd edn. (Boone, D.R. and Castenholz, R.W., Eds.), Vol. 1., pp. 119–141. Springer, New York.

[3] Stackebrandt, E. and Goebel, B.M. (1994) Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. 44, 846–849.

[4] Vandamme, P., Pot, B., Gillis, M., De Vos, P., Kersters, K. and Swings, J. (1996) Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiol. Rev. 60, 407–438.

[5] Fox, G.E., Wisotzkey, J.D. and Jurtshuk, P. (1992) How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int. J. Syst. Bacteriol. 42, 166–170.

[6] Clayton, R.A., Sutton, G., Hinkle, P.S., Bult, C. and Fields, C. (1995) Intraspecific variation in small-subunit rRNA sequences in GenBank: why single sequences may not adequatly represent prokaryotic taxa. Int. J. Syst. Bacteriol. 45, 595–599.

[7] Schmidt, T.M. (1997) Multiplicity of ribosomal RNA operons in prokaryotes. In: Bacterial Genomes: Physical Structure and Analysis (De Bruijn, J.F., Lupiski, J.R. and Weinstock, G., Eds.), pp. 221–229. Chapman and Hall.

[8] Klappenbach, J.A., Dunbar, J.M. and Schmidt, T.M. (2000) rRNA operon copy number reflects ecological strategies of bacteria. Appl. Environ. Microbiol. 66, 1328–1333.

[9] Ohta, T. (1991) Multigene families and the evolution of complexity. J. Mol. Evol. 33, 34–41.

[10] Hillis, D.M., Moritz, C., Porter, C.A. and Baker, R.J. (1991) Evidence for biased gene conversion in concerted evolution of ribosomal DNA. Science 251, 308–310.

[11] Cilia, V., Lafay, B. and Christen, R. (1996) Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level. Mol. Biol. Evol. 13, 451–461.

[12] Klappenbach, J.A., Saxman, P.R., Cole, J.R. and Schmidt, T.M. (2001) rrndb: the ribosomal RNA copy number database. Nucleic Acids Res. 29, 181–184.

[13] Wang, Y., Zhang, Z. and Ramanan, N. (1997) The actinomycte *Thermobispora bispora* contains two distinct types of transcriptionally active 16S rRNA genes. J. Bacteriol. 179, 3270–3276.

[14] Yap, W.H., Zhang, Z. and Wang, Y. (1999) Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. J. Bacteriol. 181, 5201–5209.

[15] Satokari, R.M., Vaughan, E.E., Akkermans, A.D., Saarela, M. and de Vos, W.M. (2001) Bifidobacterial diversity in human feces detected by genus-specific PCR and denaturing gradient gel electrophoresis. Appl. Environ. Microbiol. 67, 504–513.

[16] Nüble, U., Engelen, B., Felske, A., Snaidr, J., Wieshuber, A., Amann, R.I., Ludwig, W. and Backhaus, H. (1996) Sequence heterogeneities of genes encoding 16S rRNAs in *Paenibacillus polymyxa* detected by temperature gradient gel electrophoresis. J. Bacteriol. 178, 5336–5643.

[17] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

[18] Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.

[19] Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. 288, 911–940.

[20] Shimizu, T., Oshima, S., Ohtani, K., Hoshino, K., Honjo, K., Hayashi, H. and Shimizu, T. (2001) Sequence heterogeneity of the ten rRNA operons in *Clostridium perfringens* Syst. Appl. Microbiol. 24, 149–156.

[21] Moreno, C., Romero, J. and Espejo, R.T. (2002) Polymorphism in repeated 16S rRNA genes is a common property of type strains and environmnetal isolates of the genus *Vibrio*. Microbiology 148, 1233–1239.

[22] Marchanin, H., Teyssier, C., de Buochberg, M.S., Jean-Pierre, H., Carriere, C. and Jumas-Bilak, E. (2003) Intra-chromosomal heterogeneity between the four 16S rRNA gene copies in the genus *Veilonella*: implications for phylogeny and taxonomy. Microbiology 149, 1493–1501.

[23] Van de Peer, Y., Chapelle, S. and De Wachter, R. (1996) A quantitative map of nucleotide substitution rates in bacterial rRNA. Nucleic Acids Res. 17, 3381–3391.

[24] Wuyts, J., Van de Peer, Y. and De Wachter, R. (2001) Distribution of substitution rates and location of insertion sites in the tertiary structure of ribosomal RNA. Nucleic Acids res. 29, 5017–5028.

Downloaded from https://academic.oup.com/femsle/article/228/1/45/571392 by guest on 19 April 2024