Relationships between 16S-23S rRNA gene internal transcribed spacer DNA and genomic DNA similarities in the taxonomy of phototrophic bacteria

K. Okamura¹,²*, T. Hisada¹, K. Takata¹, and A. Hiraishi¹,²
¹ Department of Environmental and Life Sciences and ² Electronics-inspired Interdisciplinary Institute (EIIRIS), Toyohashi University of Technology, Toyohashi 441-8580, Japan
E-mail: okamura@recab.tut.ac.jp

Abstract. Rapid and accurate identification of microbial species is essential task in microbiology and biotechnology. In prokaryotic systematics, genomic DNA-DNA hybridization is the ultimate tool to determine genetic relationships among bacterial strains at the species level. However, a practical problem in this assay is that the experimental procedure is laborious and time-consuming. In recent years, information on the 16S-23S rRNA gene internal transcribed spacer (ITS) region has been used to classify bacterial strains at the species and intraspecies levels. It is unclear how much information on the ITS region can reflect the genome that contain it. In this study, therefore, we evaluate the quantitative relationship between ITS DNA and entire genomic DNA similarities. For this, we determined ITS sequences of several species of anoxygenic phototrophic bacteria belonging to the order Rhizobiales, and compared with DNA-DNA relatedness among these species. There was a high correlation between the two genetic markers. Based on the regression analysis of this relationship, 70% DNA-DNA relatedness corresponded to 92% ITS sequence similarity. This suggests the usefulness of the ITS sequence similarity as a criterion for determining the genospecies of the phototrophic bacteria. To avoid the effects of polymorphism bias of ITS on similarities, PCR products from all loci of ITS were used directly as genetic probes for comparison. The results of ITS DNA-DNA hybridization coincided well with those of genomic DNA-DNA relatedness. These collective data indicate that the whole ITS DNA-DNA similarity can be used as an alternative to genomic DNA-DNA similarity.

1. Introduction
There is general agreement that taxonomic information about microorganisms is incorporated in their whole genome [1, 2]. The current concept of the bacterial species – a ‘genomically coherent’ group of strains that share many common traits [3] – depends evidently on the
relatedness of bacterial genomes measurable by available molecular techniques. Next-generation sequencing technology has made it possible to relatively rapidly get whole genome sequence information on a microorganism and is giving a new insight into our understanding of the biodiversity and ecology of microorganisms [4, 5]. Nevertheless, genomic DNA-DNA hybridization still remains the ‘gold standard’ to discriminate bacterial species [1, 2]. Compared to PCR-based techniques, quantitative DNA-DNA hybridization is a laborious and time-consuming technique, because it requires relatively large-scale of cultures to prepare sufficient quantity and high quality of genomic DNA. Also, the DNA-DNA reassociation assay has limitation in quantitativity and reproducibility. This situation calls for alternatives to DNA-DNA hybridization experiments for taxonomic purpose [6-9].

An alternative approach to quantification of genome relatedness is to compare selected DNA sequences that can reliably represent whole genomes of different organisms. An ad hoc committee for the re-evaluation of the species definition in bacteria pointed out that an analysis of at least five genes of diverse chromosomal loci and wide distribution could provide sufficient information to distinguish a bacterial species from related taxa [2]. It is an open question how much information any given gene sequences can reflect the genome that includes them. Sequence differences in a given gene between closely related organisms are presumably low, because continual acquisition of random mutations has resulted from natural selection. On the other hand, such non-encoding regions as spacers on genome are functionally much less important and, therefore, detectable mutations in these regions are mostly neutral. Sequence differences between related organisms are higher in given spacers than in given genes.

Information on the 16S-23S rRNA gene internal transcribed spacer (ITS) region, in which hypervariable transcribed sequences with a number of deletion and insertion are found among species, has increasingly been used to discriminate bacterial strains at the species and intraspecies levels [10]. In general, multiple copies of the rRNA operon occur per genome, and the ITS region shows variation in length and sequences not only among species but also among different rRNA gene loci within a single organism. The ITS region also includes tRNA genes depending upon the copies and species. Since ITS DNA is easily PCR-amplified with a universal primer set, the experiment can start with single colonies or small-scale cultures of a wide variety of species. The ITS parameters used for taxonomic purpose are direct electrophoretic patterns of PCR products, their restriction fingerprints, and sequences of one or more copies. The ITS sequence analysis is an easy and reproducible tool for the discrimination of closely related strains and species. In *Bradyrhizobium* strains and related organisms, some positive relationships between ITS sequence and genomic DNA-DNA similarities were found, although there was a less correlation between the two parameters in closely related groups [11, 12].

The aim of the present study is to elucidate the relationship between ITS sequence and genomic DNA similarities and to determine whether ITS relatedness can serve as an alternative to genome relatedness in bacterial systematics. For this, we first compared the G+C content of all loci of ITS regions and entire genomic DNAs of prokaryotic organisms whose whole complete genome information has been available. Second, we determined ITS sequences of several species of the

---

**Fig. 1** Schematic structure of bacterial *rrn* operon.
phototrophic bacteria belonging to the order Rhizobiales, and examined the relationships between pairwise ITS sequence similarities and DNA-DNA relatedness. Finally, we performed ITS DNA-DNA hybridization and genomic DNA-DNA hybridization and evaluated the relationships between the two genetic parameters.

2. Materials and methods

**Bacterial strains and cultivation.** Thirty-five strains belonging to species of the genera *Afifella*, *Blastochloris*, *Rhodobium*, *Rhodoplanes*, and *Rhodopseudomonas* (*Rps.*) were used in this study (Table 1). The strains with ATCC and DSM numbers were obtained from the American Type Culture Collection, Manassas, U.S.A. and from the Deutsche Sammlung von Milroorganismen und Zellkulturen GmbH, Braunschweig, Germany, respectively. Cultivation was performed anaerobically in screw-cap test tubes or bottles filled with medium [13-18] under incandescent illumination (2,000 lx). Incubation was at 30°C for mesophilic strains and at 40°C for thermotolerant strains [13].

| Species                  | Source or designated as:                          | Reference |
|--------------------------|--------------------------------------------------|-----------|
| *Blastochloris viridis*  | DSM 133T, DSM 134                                 | [14]      |
| *Blastochloris sulfoviridis* | DSM 729T                                     | [14]      |
| *Blastochloris* sp.      | 5 strains from New Zealand hot spring           | [15]      |
| *Rhodobium orientis*     | MB312T, MB303                                   | [16]      |
| *Afifela marina*         | DSM 2698T and other 5 strains                   | [16]      |
| *Rhodoplanes elegans*    | AS130T                                          | [17]      |
| *Rhodoplanes roseus*     | DSM 5909T                                       | [17]      |
| *Rhodoplanes cryptolactis* | DSM 9987T                                    | [13]      |
| *Rhodoplanes sp.*        | 5 strains from New Zealand hot spring           | [15]      |
| *Rhodopseudomonas palustris* | DSM 123T                                    | [18]      |
| *Rhodopseudomonas rhenobacensis* | RbT                                      | [18]      |
| *Rhodopseudomonas faecalis* | g-c                                      | [18]      |
| *Rhodopseudomonas sp.*   | 8 strains from New Zealand hot spring and Japanese pond | [15] |

**Amplification, cloning and sequencing of the ITS region.** Crude cell lysates as a DNA source were prepared for PCR use as described [18]. The ITS region was amplified with two pair sets of PCR primers, SS1512f and LS23r or LS117r [18], which corresponded to positions 1512-1531 of *Escherichia coli* 16S rRNA and positions 21-37 or 117-137 of *E. coli* 23S rRNA, respectively. The ITS region of several *Rhodopseudomonas* strains were not amplified by using primers SS1512f and LS23r. For these strains, we used another reverse primer, LS23r-p (5’-AAGGCATCCGTCGAACA-3’), which was designed based on the conserved region of 23S rRNA gene of the strains. Amplified fragments were separated by agarose gel electrophoresis, purified with the Wizard Gel and PCR Clean-up System (Promega), and subcloned using a pT7Blue Perfectly Blunt™ Cloning kit (Novagen) according to the manufacturer’s instructions. At least 5 clones per strain were subjected to sequencing. Inserts were sequenced using plasmid primers T7 and M13 and a Dye Terminator Cycle Sequencing kit, and analyzed using a PRISM 3100 DNA sequencer (Applied Biosystems).

**Sequence analysis.** Sequence data were complied with the GENETYX-MAC ver.13 program
The multiple alignment of sequence and the calculation of the corrected evolutionary distance was performed using the CLUSTAL X program [19]. Searching for tRNA genes and predicting the secondary structure of tRNA were also performed using the GENETYX-MAC program.

**Comparison between GC contents of ITS and genomic DNA.** The guanine + cytosine (G+C) contents of the genomic DNA of 215 prokaryotic organisms whose whole genomes have been completely sequenced were calculated by using the ‘GC Plot’ program from Genome Information Broker of DDBJ database (http://gib.genes.nig.ac.jp/) [20]. The G+C contents of ITS regions within genomes were calculated using the GENETYX-MAC program. For this, all loci of ITS sequences within genomes were retrieved from Genome Information Broker or Sequence Retrieval System [http://srs.ddbj.nig.ac.jp] of DDBJ. Statistical analyses of all numerical data including calculation of the Pearson correlation coefficient $r$ were performed using the Microsoft Excell.

**Genomic DNA-DNA hybridization.** Genomic DNA was extracted and purified by the method of Marmur [21]. DNA-DNA hybridization studies were performed by the dot-blot hybridization method with alkaline phosphatase labeling and chemiluminescence detection using an Amersham-Pharmacia AlkalPhos kit, as described [18]. Labelling and detection were performed according to the manufacturer’s instructions. Hybridizations were performed at 50°C. Hybridization signals were detected with an Amersham-Pharmacia ECL mini-camera, and their intensity was measured by the NIH image program available at the web site [http://rsb.info.nih.gov/nih-image/].

**ITS DNA-DNA hybridization.** PCR products of ITS regions were obtained as noted above and detected by agarose gel electrophoresis. In the case that a single band was observed on the gel, the DNA fragment was purified by using the PCR clean up system (Takara). When the non-specific amplification was observed, the DNA was purified with the Wizard Gel and PCR Clean-up System (Promega). All DNA samples were prepared at a concentration of 100 ng/µl. To eliminate the effect of conserved DNA regions, *i.e.*, primer-annealing regions and tRNA genes, on hybridization signals, a PCR product of *Escherichia coli* ITS regions was used as competitor DNA. The competitor DNA was purified and digested with *Dra*I (Takara) and *Mbo*I (Takara) prior to use and blotted onto nylon membrane at a concentration of 50 times (w/w) of a total blotted DNA. The ITS DNA-DNA hybridization assay was performed according to the protocol for genomic DNA-DNA hybridization with slight modifications. The blotted volume of test DNA solutions (10 ng/µl) was 2 µl. Prehybridization was performed at 50°C for 1 h, in the presence of denatured competitor DNA and 30% deionized formamide in hybridization buffer. Then, hybridization was performed at 50°C for 2 h. The hybridized membranes were washed twice at 50°C with primary wash buffer according to the manufacturer’s instructions. Hybridization signals were detected as described above for genomic
3. Results

Relationship between the G+C content of ITS regions and whole genomes. We first studied the relationships between the G+C content of ITS regions and genomic DNAs of 215 prokaryotic organisms including mesophiles and moderately thermophiles of archaea and bacteria (Fig. 3). Although ITS DNA had a lower G+C content than that of genomic DNA in general, there was a strong correlation between the two variables (slope of the regression line =0.614, \( r = 0.75, P < 0.01 \)). An exceptional case was that, in hyperthermophilic archaea and bacteria, ITS G+C ratios were higher than genomic G+C ratios (slope of the regression line =0.913, \( r = 0.593, P < 0.01 \)). These results suggest that ITS regions can be used as genetic markers reflective of the entire genome.

ITS sequence analysis. ITS sequences were analyzed for 35 strains of the phototrophic bacteria belonging to the genera Afifella, Blastochloris, Rhodobium, Rhodoplanes, and Rhodopseudomonas. From most of the test strains, ITS regions could be amplified with a PCR primer set of SS1512f and LS23r. In the case of the New Zealand hot spring and Japanese pond isolates (see Table 1), however, ITS regions were amplified only using a primer set of SS1512f and LS117r. Since most of the test strains gave a single band of PCR amplicons upon agarose gel electrophoresis, direct sequencing of the PCR products without subcloning was possible. Also, the PCR products from these strains were subcloned, and at least five clones per strain were sequenced. As a result, all the clones from a single strain had identical sequences. Therefore, the test bacteria from which a single band of amplicon was produced may have one single type of rRNA operon, although the number of the operon copies was unknown. On the other hand, three different clones of ITS having different sequences were obtained from Rps. rhenobacensis strain RbT.

The ITS sequences determined varied in length among the test strains as follows: 787 to 839 nucleotides in the Blastochloris strains, 724 to 875 nucleotides in the Rhodobium strains, 937 to 1021 nucleotides in the Rhodoplanes strains, and 761 to 1085 nucleotides in the Rhodopseudomonas strains. All of the detected ITS regions contained the tRNA\textsuperscript{Ile} and tRNA\textsuperscript{Ala} genes. The primary and secondary structures of the tRNA\textsuperscript{Ile} and a tRNA\textsuperscript{Ala} predicted were the same in all test strains within a single genus. Therefore, the ITS-tRNA structures may be used as a genus-specific genetic maker.

An alignment analysis of ITS sequences showed that there were great sequence variations in non-encoding regions with insertions and deletions among the test species. Therefore, it was much more difficult to make correct alignments of ITS sequences than of rRNA gene sequences.
ITS sequence similarities were deduced from these alignments with gaps.

**Relationship between pairwise ITS sequence similarities and genomic DNA-DNA relatedness.** The relationship between pairwise ITS sequences similarities and DNA-DNA relatedness in the phototrophic species is shown in Fig. 4. Pairwise similarities to the corresponding ITS sequence of *E. coli* and DNA-DNA hybridization levels against *E. coli* were used as the control to make a starting point of the regression line. There is a strong correlation between the two genetic markers (*r* = 0.8448, *P* < 0.01). Based on this relationship, 70% DNA-DNA relatedness was found to correspond to a 92% level of the ITS sequence similarity. Also, 60-80% ITS sequence similarities corresponded to 20-65% DNA-DNA similarities in the phototrophic bacteria, although these relationships were different to some extent from species to species. The aforementioned results suggest the usefulness of the ITS sequence similarity with gaps as a criterion for determining genospecies of the phototrophic bacteria.

**Relationship between ITS-ITS hybridization and DNA-DNA hybridization values.** Finally, we applied the ITS DNA-DNA hybridization method using PCR-amplified DNA from all loci of ITS per genome as well as genomic DNA-DNA hybridization for genetically closely related strains of the phototrophic bacteria within each genus. Figure 5 shows the correlation between pairwise ITS DNA-DNA similarity and genomic DNA-DNA relatedness. A liner positive relationship between the two parameters was observed with a significant correlation coefficient (*r* = 0.900, *P* < 0.01). Based on this relationship, 70% DNA-DNA relatedness was estimated to correspond to an 82% level of ITS-ITS similarity.

**4. Discussion**

The differences in genomic G+C content may have been caused by directional mutation pressure,
and such mutation pressure seems to have been exerted uniformly on the entire genome during evolution [22]. However, direction and magnitude of this pressure vary among the phylogenetic lines, as the G+C content of bacterial DNA actually ranges from 25 to 75% [23] with low intraspecific heterogeneity [24-27]. Thus, the G+C content of entire genomic DNA is one of the most important characteristics for the description of the bacterial genotype [28]. In this study, the G+C content analysis of the ITS and genomic DNA of 215 prokaryotic species showed that there is a high correlation between the two, this suggesting that the ITS DNA similarities reflects entire genome similarities for taxonomic purpose. Muto and Osawa [29] demonstrated that the G+C content of various parts of the bacterial genome has a positive linear correlation with that of the whole genomic DNA. The G+C ratios of both spacers and protein-encoding genes had a strong linear correlation with genomic G+C contents.

This study has also shown that ITS sequence similarities significantly correlate with DNA-DNA relatedness in species of the four phototrophic genera belonging to the *Rhizobia* (Fig. 2). We have shown that the 92.0% level of ITS sequence similarities corresponds to 70% of DNA-DNA relatedness. Willems *et al.* [11, 12] found that, in *Bradyrhizobium* strains and related organisms, a 95.5% level of ITS sequence similarities corresponded to 60% DNA-DNA relatedness. The G+C contents of entire genome and whole ITS regions in *Affifella*, *Bradyrhizobium*, *Blastochloris*, *Rhodobium*, *Rhodoplanes*, and *Rhodopseudomonas* species, all of which belong to the order *Rhizobiales*, are similar to each other (genomic DNA, 64-66 mol%; ITS, 53.2-56.6 mol%). The mutation pressure for ITS and genomic DNA of these organisms during evolution may be in the same direction, resulting in a high correlation between the two genetic markers.

A single bacterium has usually multilocus rRNA operons including respective different sequences of ITS regions. ITS-based phylogenetic relationships among the organisms may be affected by which ITS copies analyzed are used. Therefore, the organisms under investigation should not be classified by using only an ITS sequence of one locus. Also, it is actually difficult to align all sequences of different ITS copies from different species for phylogenetic analysis, because ITS sequences are hypervariable. An approach to overcome this problem is to estimate the sequence similarity of whole ITS regions directly by the hybridization technique using their PCR products. As reported herein, the results of ITS-ITS hybridization assays were well agreement with those of the conventional DNA-DNA hybridization; 70% of DNA-DNA relatedness as the threshold of species discrimination roughly corresponded to an 82% level of ITS-ITS relatedness. These results indicate that ITS-ITS relatedness can be used as an alternative to genomic DNA-DNA relatedness to discriminate bacterial strains at the species level. An advantage of the ITS-ITS hybridization method is that it can be applied to multi-copies of ITS regions having different sequences.

In conclusion, the ITS-ITS hybridization method using PCR products is more rapid and simpler than the conventional DNA-DNA hybridization technique. This method is applicable to the DNA microarray analysis not only for rapid identification of bacteria but also for characterization of microbial communities. Information about the ITS region has already been applied to microbial community analyses [30-33]. To more definitely evaluate ITS-ITS similarity as a genetic marker, however, further study is necessary with more strains of wide variety of species.
Acknowledgements
This work was supported by a Grant-in-Aid for Young Scientists ((B) 23710091), the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References
[1] Stackebrandt E et al. 2002 Int. J. Syst. Evol. Microbiol. 52 1043–1047
[2] Wayne L G et al. 1987 Int. J. Syst. Bacteriol. 37 463–464
[3] Rosselló-Mora R and Amann R 2001 FEMS Microbiol. Rev. 25 39–67
[4] Trevors J T and Masson L 2010 Antonie van Leeuwenhoek 98 249–262.
[5] Dini-Andreote F et al. 2012 Microb. Ecol. 64 1–7
[6] Cho J C and Tiedje J M 2001 Appl. Environ. Microbiol. 67 3677–3682
[7] Coenye T et al. 2005 FEMS Microbiol. Rev. 29 147–167
[8] Gevers D et al. 2005 Nat. Rev. Microbiol. 3 733–739
[9] Goris J et al. 2007 Int. J. Syst. Evol. Microbiol. 57 81–91
[10] Gürtler V and Stanisich V A 1996 Microbiology 142 3–16
[11] Willems A et al. 2001 Int. J. Syst. Evol. Microbiol. 51 623–632
[12] Willems A et al. 2003 Syst. Appl. Microbiol. 26 203–210
[13] Okamura K et al. 2007 J. Gen. Appl. Microbiol. 53 357–361
[14] Hiraishi A 1997 Int. J. Syst. Bacteriol. 47 217–219
[15] Hisada T et al. 2007 Microbes Environ. 22 405–411
[16] Hiraishi A et al. 1995 Int. J. Syst. Bacteriol. 45 226–234
[17] Hiraishi A and Ueda Y 1994 Int. J. Syst. Bacteriol. 44 665–673
[18] Okamura K et al. 2009 J. Gen. Appl. Microbiol. 55 469–478
[19] Thompson J D et al. 1997 Nucleic Acids Res. 24 4876–4882
[20] Fumoto M et al. 2002 Nucleic Acids Res. 30 66–68.
[21] Mamur J 1961 J. Mol. Biol. 3 208–218
[22] Sueoka N 1962 Proc. Natl. Acad. Sci. USA 48 582–592
[23] Lee K Y et al. 1956 Ann. Inst. Pasteur 91 212–224
[24] Sueoka N et al. 1959 Nature 183 1429–1431
[25] Rolfe R and Messelson M 1959 Proc. Natl. Acad. Sci. USA 44 1039–1043
[26] Sueoka N 1961 Proc. Natl. Acad. Sci. USA 47 1141–1149
[27] Schildkraut C L et al. 1962 J. Mol. Biol. 4 430–43
[28] Barbu E et al. 1956 Ann. Inst. Pasteur. 91 212–224
[29] Muto A and Osawa S 1987 Proc. Natl. Acad. Sci. USA 84 166–169
[30] Gonzalez N et al. 2003 J. Microbiol. Methods. 55 91–97
[31] Cook K L et al. 2004 J. Microbial. Methods. 57 79–93
[32] Nübel U et al. 2004 FEMS Microbiol. Lett. 15 215–223
[33] Gunther S et al. 2006 J. Microbiol. Methods. 65 226–236