Variations in the Multiple Copies of Bacterial 16S rRNA Genes do not Affect Species Identification

S4-1

Variations in the Multiple Copies of Bacterial 16S rRNA Genes do not Affect Species Identification

Jae-Ho Shin

School of Applied Biosciences, Kyungpook National University

Variable region analysis of 16S rRNA gene sequences is the most common tool in bacterial taxonomic studies. Although being used for distinguishing bacterial species, its use remains limited, due to the presence of variable copy numbers with sequence variation in the genomes. In this study, 16S rRNA gene sequences, obtained from completely assembled whole genome and Sanger electrophoresis sequencing of cloned PCR products from Serratia fonticola GS2, were compared. Sanger sequencing produced a combination of sequences from multiple copies of 16S rRNA genes. To determine whether the variant copies of 16S rRNA genes affected Sanger sequencing, two ratios (5:5 and 8:2) with different concentrations of cloned 16S rRNA genes were used; it was observed that the greater the number of copies with similar sequences, the higher its chance of amplification. The effect of multiple copies for taxonomic classification of 16S rRNA gene sequences was investigated using the strain GS2 as a model. The 16S rRNA copies with the maximum variation had a minimum 99.42% pairwise similarity, and this did not affect species identification. Thus, PCR products from genomes containing variable 16S rRNA gene copies can provide sufficient information needed for species identification, except for species which have a high similarity of sequences in their 16S rRNA gene copies, like the case of *Bacillus thuringiensis* and *Bacillus cereus*. *In silico* analysis of 1,616 bacterial genomes from long-read sequencing was also done. The average minimum pairwise similarity for each phylum was reported with their average genome size and average “unique copies” of 16S rRNA genes, and we found that the phyla Proteobacteria and Firmicutes showed the highest amount of variation in their copies of their 16S rRNA genes. Overall, our results sheds light on how the variations in the multiple copies of the 16S rRNA genes of bacteria can be used in appropriate species identification.