



# Phylogenetic Network Analysis Revealed the Occurrence of Horizontal Gene Transfer of 16S rRNA in the Genus *Enterobacter*

Mitsuharu Sato<sup>1,2</sup> and Kentaro Miyazaki<sup>1,2\*</sup>

<sup>1</sup> Bioproduction Research Institute, Department of Life Science and Biotechnology, National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan, <sup>2</sup> Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Japan

## OPEN ACCESS

### Edited by:

Zhe Zhao,  
Hohai University, China

### Reviewed by:

Alessio Mengoni,  
University of Florence, Italy  
Siomar De Castro Soares,  
Universidade Federal do Triângulo  
Mineiro, Brazil

### \*Correspondence:

Kentaro Miyazaki  
miyazaki-kentaro@aist.go.jp

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 10 August 2017

**Accepted:** 30 October 2017

**Published:** 16 November 2017

### Citation:

Sato M and Miyazaki K (2017)  
Phylogenetic Network Analysis  
Revealed the Occurrence  
of Horizontal Gene Transfer of 16S  
rRNA in the Genus *Enterobacter*.  
*Front. Microbiol.* 8:2225.  
doi: 10.3389/fmicb.2017.02225

Horizontal gene transfer (HGT) is a ubiquitous genetic event in bacterial evolution, but it seldom occurs for genes involved in highly complex supramolecules (or biosystems), which consist of many gene products. The ribosome is one such supramolecule, but several bacteria harbor dissimilar and/or chimeric 16S rRNAs in their genomes, suggesting the occurrence of HGT of this gene. However, we know little about whether the genes actually experience HGT and, if so, the frequency of such a transfer. This is primarily because the methods currently employed for phylogenetic analysis (e.g., neighbor-joining, maximum likelihood, and maximum parsimony) of 16S rRNA genes assume point mutation-driven tree-shape evolution as an evolutionary model, which is intrinsically inappropriate to decipher the evolutionary history for genes driven by recombination. To address this issue, we applied a phylogenetic network analysis, which has been used previously for detection of genetic recombination in homologous alleles, to the 16S rRNA gene. We focused on the genus *Enterobacter*, whose phylogenetic relationships inferred by multi-locus sequence alignment analysis and 16S rRNA sequences are incompatible. All 10 complete genomic sequences were retrieved from the NCBI database, in which 71 16S rRNA genes were included. Neighbor-joining analysis demonstrated that the genes residing in the same genomes clustered, indicating the occurrence of intragenomic recombination. However, as suggested by the low bootstrap values, evolutionary relationships between the clusters were uncertain. We then applied phylogenetic network analysis to representative sequences from each cluster. We found three ancestral 16S rRNA groups; the others were likely created through recursive recombination between the ancestors and chimeric descendants. Despite the large sequence changes caused by the recombination events, the RNA secondary structures were conserved. Successive intergenomic and intragenomic recombination thus shaped the evolution of 16S rRNA genes in the genus *Enterobacter*.

**Keywords:** ribosome, 16S ribosomal RNA, homogenization, horizontal gene transfer, phylogenetic network analysis, intragenomic recombination, intergenomic recombination, *Enterobacter*

**Abbreviations:** HGT, horizontal gene transfer; PIS, parsimonious informative site.

## INTRODUCTION

Based on the structural complexity of the ribosome (Brodersen et al., 2002; Klein et al., 2004; Schuwirth et al., 2005), it was believed that each ribosomal component is species-specific and the genes have seldom experienced HGT (Jain et al., 1999). In particular, the 16S rRNA gene has been recognized as the “ultimate chronometer” (Woese, 1987) and the sequence is widely used for taxonomic classification of bacteria (Lane et al., 1985; Woese, 1987; Woese et al., 1990).

One unique feature of rRNA genes is that the genomes of many bacteria contain several copies (Stoddard et al., 2015). For example, *Escherichia coli* contains 7 copies per genome and *Bacillus subtilis* contains 13 copies per genome. The 16S rRNA gene copies in the same genome are, in general, almost identical (>99.5% identity); this concerted evolution of 16S rRNA genes is due to gene conversion (or homogenization) within the genome (Hillis et al., 1991; Liao, 1999, 2000). In *E. coli*, the frequency of such intragenomic recombination between 16S rRNA genes is as high as  $5 \times 10^{-9}$  per generation (Hashimoto et al., 2003). Such a high sequence similarity among the copies of the 16S rRNA gene is also reported for bacterial genomes; bacterial species (genomes), which contain dissimilar (showing less than 98% sequence identity) 16S rRNA genes, are rare, counting for only 28 out of 2,143 genomes investigated (Tian et al., 2015).

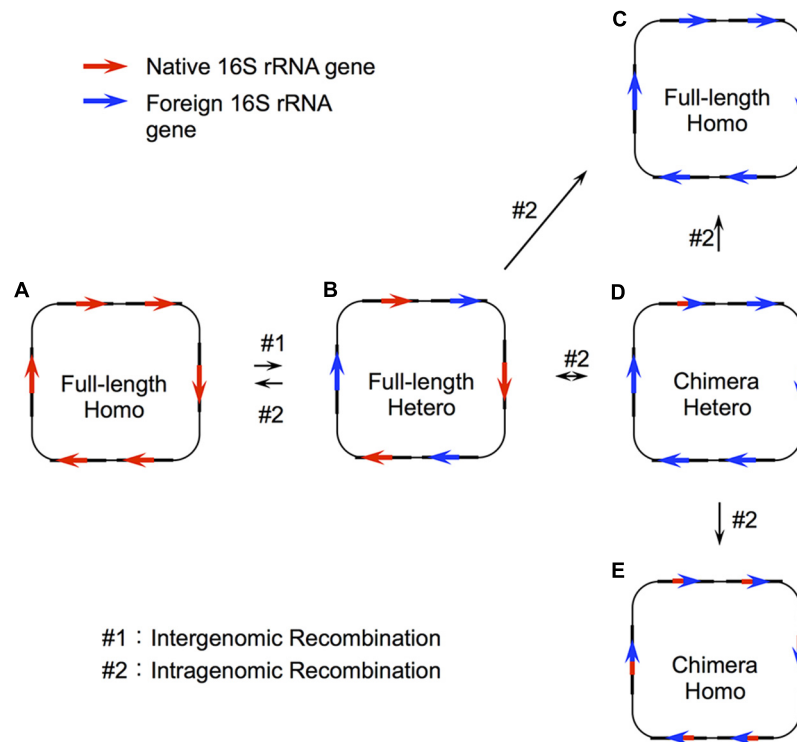
Despite this general tendency of uniformity in 16S rRNA sequences within the same genomes, the genomes of several bacteria possess heterogeneous 16S rRNA genes (Mylvaganam and Dennis, 1992; Ueda et al., 1999; Yap et al., 1999; Acinas et al., 2004; Boucher et al., 2004) or chimeric 16S rRNA genes (Mylvaganam and Dennis, 1992; Eardly et al., 1996, 2005; Lan and Reeves, 1998; Ueda et al., 1999; Yap et al., 1999; Wang and Zhang, 2000; Schouls et al., 2003). With respect to the former, these instances can arise from inefficient intragenomic recombination after interspecies recombination (or HGT) [the evolutionary model for the 16S rRNA gene is illustrated in **Figure 1** (Schouls et al., 2003)] or each allele can play functionally different and essential roles for growth (Lopez-Lopez et al., 2007). With respect to the latter situation, chimeric sequences can arise through intergenomic recombination (or HGT, #1 in **Figure 1**) followed by intragenomic recombination between the copies in the same genome (Schouls et al., 2003) (#2 in **Figure 1**).

In addition to these natural instances, we recently demonstrated experimentally that the *E. coli* ribosome is capable of using foreign 16S rRNA from diverse bacterial origins with sequence identity as low as ~78% (Kitahara et al., 2012; Miyazaki et al., 2017; Tsukuda et al., 2017). We also found that some domain- and helix-swapping were also permissible between distantly related species (Kitahara and Miyazaki, 2011; Tsukuda et al., 2017). Thus, as opposed to the “complexity hypothesis,” which claims that the major components included in complex supramolecules (or biosystems) do not transfer between species because such drastic genetic changes should disrupt the function of the whole molecule (or system)

(Jain et al., 1999), 16S rRNA, which is the center of the 30S subunit and has many contacts with surrounding proteins, may be transferrable between species. Taken together, these instances suggest the occurrence of HGT in 16S rRNA genes, just as in enzyme-encoding genes (Kitahara and Miyazaki, 2013). The resulting products caused by HGT can be the replacement of the whole gene or segmental replacement depending on the recombination points (**Figure 1**).

Although several sporadic studies have reported HGT for 16S rRNA genes, little has been done to systematically investigate the possibility (and if present, the frequency) of HGT in 16S rRNAs. Recently, Tian et al. (2015) investigated 2,143 prokaryotic genomes, and they concluded that the frequency of HGT in 16S rRNA genes is “rare.” However, there are several significant mistakes in their study, primarily because of their use of an incorrect evolutionary model (**Figure 1**). Briefly, they focused their investigation on the genomes containing the 16S rRNA genes showing >2% dissimilarity (assuming that this dissimilarity could be brought by foreign bacterial 16S rRNA genes). However, such analysis overlooks the state E in **Figure 1**, and the state B when the initial dissimilar sequence is lost through intragenomic recombination where the degree of dissimilarity falls to a level below 2%. Another problem in their paper lies in their method used to detect recombination. Specifically, they used the phylogenetic tree approach based on maximum likelihood method, which is based on tree-shape evolution (driven by point mutations) and inadequate to decipher evolutionary history of genes including recombination.

To complement all these drawbacks, we here applied a “phylogenetic network” method (Bandelt, 1994), which has previously been used to detect evolutionary events, such as recombination (Kitano et al., 2009, 2012), gene conversion (Kitano and Saitou, 1999; Ezawa et al., 2010) and gene fusion (Tomiki and Saitou, 2004), in homologous alleles. For example, Kitano et al. reported that genetic recombination created the haplogroups of the ABO blood type genes of gibbon monkeys (Kitano et al., 2009) and humans (Kitano et al., 2012). We targeted the genus *Enterobacter*, because the phylogenetic relationships inferred by multi-locus sequence alignment (MLSA) analysis and 16S rRNA-based analysis in this genus are incompatible (Brady et al., 2013). We collected all 10 complete genomic sequences of *Enterobacter* from the NCBI database and made phylogenetic analysis for 16S rRNA genes therein and MLSA analysis for the genomes. The results suggested a topological difference between the phylogenetic trees, confirming the results by Brady et al. (2013) and the occurrence of recombination-driven evolution in 16S rRNA genes. We then applied phylogenetic network analysis to the 16S rRNA genes, which suggested that there were three ancestral groups of 16S rRNA genes among the 10 groups, of which two groups appeared to create a further seven groups, through five recombination events. The present results strongly suggest that HGT is a major driving force for the evolution of 16S rRNA genes in the genus *Enterobacter*.



**FIGURE 1 |** Two step recombination model for 16S rRNA genes. Recombination of the 16S rRNA gene takes place through two steps: intergenomic recombination (#1), followed by intragenomic recombination (#2). In the first step, a genetic fragment containing a foreign 16S rRNA gene is transferred to a host organism (**A**) to produce a genome with heterogeneous copies of the gene (**B**). This initial step (#1) is the actual HGT (or interspecies exchange). From this heterogeneous status, there are four possible patterns of intragenomic recombination depending on the recombination sites. (**A**) All copies are taken over by the native 16S rRNA gene, (**C**) all copies are taken over by the full-length foreign 16S rRNA gene, and (**D**) part of the foreign 16S rRNA gene remains in some copy number. The status (**D**) would be changed to (**C**) or (**E**) through recursive intragenomic recombination.

## MATERIALS AND METHODS

### Dataset Collection

16S rRNA genes of *Enterobacter* were retrieved from the National Center for Biotechnology Information (NCBI) database on July 14, 2015. The information (name of each sequence, genome accession numbers, locations in the genome and species name) of the 71 retrieved *Enterobacter* sequences is shown in Supplementary Table S1.

### Multiple Sequence Alignment

Multiple sequence alignment of 16S rRNA genes was carried out using MUSCLE (Edgar, 2004) implemented in the MEGA 6.0 software (Tamura et al., 2013) with default parameters. All single insertion sites of *Enterobacter* sequences were eliminated after multiple sequence alignment. Singleton sites and PISs were identified using MEGA 6.0 software (Tamura et al., 2013). *E. coli* site numbering was used.

### Phylogenetic Tree Analysis

A phylogenetic tree of 16S rRNA genes was reconstructed using the neighbor-joining method (Saitou and Nei, 1987) implemented in the MEGA 6.0 software (Tamura et al.,

2013) with the Kimura two-parameter distance (Kimura, 1980). Bootstrap values were calculated with 1,000 resamples. A parsimonious phylogenetic network was constructed following the procedures set out in Bandelt (1994) and Saitou and Yamamoto (1997).

### Determination of Non-randomness of the Order of PISs

A two-sample runs test (Takahata, 1994; Ezawa et al., 2006) was employed to investigate the random/non-random distribution of PISs in the 16S rRNA sequences. Detailed calculation procedures are described in the Supplementary Information. The results were benchmarked following the procedure by Takahata (1994).

### MLSA Analysis

Multi-locus sequence alignment analysis was carried out for the 10 representative *Enterobacter* genomes using the following procedure. (1) Retrieved 609 essential genes, which are listed in the Database of Essential Genes (Zhang and Lin, 2009), from the *E. coli* MG1655 genome (NC\_000913.3). (2) Identified 194 essential genes (in the 609 genes as described above), which are shared by selected *Enterobacter* genomes. (3) Each gene was translated into their amino acid sequence, which

were multiply aligned using MUSCLE 3.8.31 (Edgar, 2004). (4) Removed poorly aligned regions and divergent regions using Gblocks 0.91 (Castresana, 2000). When more than 20 continuous amino acid sites were found to be different, these sites were also manually removed. (5) Concatenated all the trimmed amino acid sequences. (6) Constructed a neighbor-joining tree using MEGA. 6.0 software with the JTT matrix-based method (Jones et al., 1992) with 1,000 bootstrap resampling.

## RESULTS AND DISCUSSION

### Data Collection

We retrieved 10 complete genomic sequences of *Enterobacter* from the National Center for Biotechnology Information (NCBI) database, which were designated G1 to G10 (Supplementary Table S1). Among them, all but one genome (G5) contained multiple 16S rRNA gene copies. G1, G3, and G5 contained seven copies; G2, G4, G6, G7, and G8 contained eight copies, and G10 contained nine copies (Supplementary Table S1). Overall, a total of 71 16S rRNA genes was identified in the 10 genomes.

### Intragenomic Uniformity of 16S rRNA Gene Copies

**Figure 2A** shows a neighbor-joining tree of the 71 *Enterobacter* 16S rRNA gene sequences with *E. coli* MG1655 *rrsB* as an outgroup. All of the 16S rRNA genes that originated from the same genome shared high sequence identities ( $\geq 99\%$ ) and clustered with high bootstrap values ( $\geq 93$ ). This indicates the occurrence of intragenomic recombination (homogenization) between the copies (Hillis et al., 1991; Liao, 1999, 2000). In contrast, bootstrap values for each node were, in general, low, suggesting uncertain topology between the genomes.

To investigate the phylogenetic relationships of 16S rRNA gene sequences between clusters (genomes), we selected representative 16S rRNA genes from each cluster. As a rule, we selected one ancestral sequence (closest to the outgroup [*E. coli*]) and one sequence from each cluster to cover the variety. For example, for G1, E7 branches from the ancestral node and E1, E3, E4, E5, and E6 clustered (Supplementary Figure S1A). Thus, E1 and E7 were selected as the representatives of G1. Similarly, from each genome, we selected E8 and E9 from G2 (Supplementary Figure S1B), E19, E20, and E22 from G3 (Supplementary Figure S1C), E23 and E26 from G4 (Supplementary Figure S1D), E32 and E33 from G5 (Supplementary Figure S1E), E54 and E57 from G8 (Supplementary Figure S1F) and E63, E64, and E66 from G10 (Supplementary Figure S1G). For G6 and G7, we selected one sequence because all copies in the genomes were identical in these strains (E46 and E38, respectively). For G9, only a single 16S rRNA copy (E62) was used in the analysis. In total, we selected 19 representative sequences from 10 *Enterobacter* genomes (indicated by dots in Supplementary Figure S1).

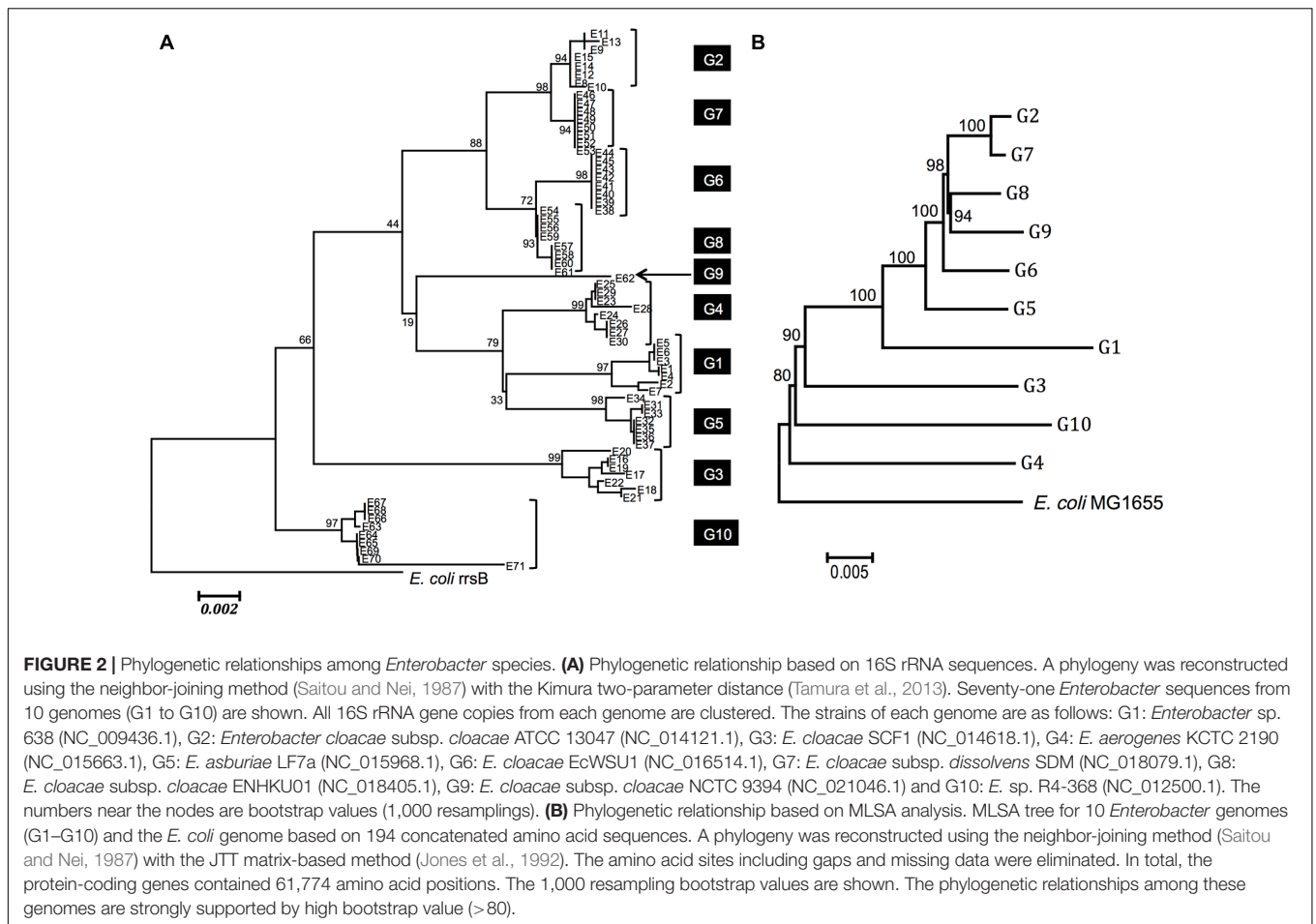
### MLSA Analysis

Brady et al. (2013) have reported that the phylogenetic relationships inferred by MLSA analysis using four protein-coding essential genes (*rpoB*, *gyrB*, *infB*, and *atpD*) and 16S rRNA-based analysis were incompatible. We performed a similar analysis for the 10 selected genomes (G1–G10) using a more comprehensive set of essential genes. We first retrieved 609 genes that were known to be essential in *E. coli*, from which 194 genes were selected. These genes were present in all 10 genomes. The 194 protein-coding genes (61,774 amino acids total) were as follows; *accA*, *accB*, *accC*, *aceF*, *ackA*, *acpP*, *adk*, *alaS*, *argS*, *aroK*, *asnS*, *aspS*, *atpC*, *atpF*, *can*, *csrA*, *cvpA*, *cydA*, *cydB*, *cysS*, *dadA*, *dapF*, *der*, *dnaA*, *dnaB*, *dnaE*, *dnaG*, *dnaK*, *dxs*, *eno*, *era*, *fabA*, *fabB*, *fabG*, *fabI*, *fbaA*, *ffh*, *fis*, *folA*, *folD*, *frf*, *ftsA*, *ftsH*, *ftsI*, *ftsQ*, *ftsY*, *ftsZ*, *fusA*, *gapA*, *glms*, *glnS*, *glyA*, *glyQ*, *glyS*, *gmk*, *gpsA*, *groL*, *grpE*, *gshB*, *gyrA*, *gyrB*, *hemE*, *hfq*, *hisS*, *hns*, *holC*, *ileS*, *infA*, *infB*, *infC*, *iscS*, *ispB*, *ispE*, *ispH*, *ispU*, *lepB*, *leuS*, *lolC*, *lpxA*, *lpxC*, *lpxD*, *metG*, *metK*, *miaA*, *minD*, *mraY*, *mreB*, *mukB*, *mukF*, *murC*, *murE*, *murI*, *nrdA*, *nrdB*, *nusG*, *obgE*, *parE*, *pgk*, *pheS*, *pheT*, *plsB*, *pnp*, *ppiB*, *priB*, *proS*, *purB*, *pyrG*, *ribA*, *ribB*, *ribE*, *ribF*, *rnc*, *rne*, *rnpA*, *rplB*, *rplC*, *rplD*, *rplE*, *rplF*, *rplJ*, *rplL*, *rplM*, *rplN*, *rplO*, *rplP*, *rplQ*, *rplR*, *rplS*, *rplT*, *rplU*, *rplV*, *rplW*, *rplX*, *rplY*, *rpmA*, *rpmB*, *rpmC*, *rpmD*, *rpmH*, *rpoA*, *rpoB*, *rpoC*, *rpoD*, *rpoE*, *rpoH*, *rpsA*, *rpsB*, *rpsC*, *rpsD*, *rpsE*, *rpsF*, *rpsG*, *rpsH*, *rpsI*, *rpsJ*, *rpsK*, *rpsL*, *rpsM*, *rpsN*, *rpsO*, *rpsP*, *rpsQ*, *rpsR*, *rpsS*, *rpsT*, *secA*, *secD*, *secE*, *secY*, *serS*, *slyD*, *spoT*, *ssb*, *sucA*, *sucB*, *suhB*, *thrS*, *tktA*, *topA*, *trmD*, *trpS*, *tsf*, *tufA*, *tufB*, *tyrS*, *valS*, *ybeY*, *ybgC*, *yceQ*, *ychF*, *yidC*, *yqgE*, *yqiB*, *zipA*. The deduced amino acid sequences for these genes were trimmed, concatenated, and used for phylogenetic analysis. **Figure 2B** shows the phylogenetic relationship between these genomes, whose topology was different from that constructed based on the 16S rRNA (**Figure 2A**), confirming the notion by Brady et al. (2013).

### Multiple Occurrences of HGT in 16S rRNA Genes in *Enterobacter*

In the multiple alignments of 20 16S rRNA gene sequences (19 representative sequences from *Enterobacter* plus *E. coli*), there were a total of 82 variable sites. Of these sites, 15 were singletons and 67 were PISs. Among the 67 PISs, those carrying more than three nucleotides (11 PISs) were removed from the analysis. Two PISs that contained gaps were also removed from the analysis. Overall, 15 singletons and 54 PISs were used for phylogenetic network analysis. A multiple alignment of the 54 PISs of the 19 representatives is shown in **Figure 3**. Based on this PIS alignment and 15 singleton sites, a parsimonious phylogenetic network was reconstructed, following the procedure as described previously (Bandelt, 1994; Saitou and Yamamoto, 1997) (**Figure 4**). The numbers along each side indicate the corresponding PISs that separate one group from another. For example, *E. coli*, G2/G7 (E8, E9, E46), G3 (E19, E20, E22) and G10 (E63, E64, E66) have the nucleotide 'G' at PIS 987 (red-colored numbers in **Figures 3, 4**), whereas others (G1 [E1, E7], G4 [E23, E26], G5 [E32, E33], G6/G8 [E38, E54, E57] and G9 [E62]) have 'A' at the same





position. Thus, the genomic groups are separated by PIS 987. Similarly, in *E. coli*, G2/G7, G3 and G10 are separated from G1, G4, G5, G6/G8, and G9 by PISs 988, 1001, 1039, 1217, and 1218. These PISs are located on the same vertical side with PIS 987.

In a parsimonious phylogenetic network, recombination results in the formation of reticular structures. In the present case, there are many reticules (or vertical and horizontal sides) in the parsimonious phylogenetic network, suggesting multiple occurrences of recombination (Kitano et al., 2009, 2012). From the reticule, the following evolutionary information can be obtained (Kitano et al., 2012; Saitou and Kitano, 2013):

- (1) The recombinant sequence is located at the corner opposite the outgroup.
- (2) The parental sequences are located at the corner of another diagonal of the same reticule.
- (3) The two sides (vertical and horizontal) of the reticule are dominated by the upstream or downstream PISs in question.

To deduce each recombination event in **Figure 4**, we dissected the whole network into several groups and attempted to identify

hypothetical recombinants and their parents in each group. Let us focus on the reticule consisting of G2/G7, G6/G8, G9, and *E. coli* as an example (Supplementary Figure S2A). In this case, G6/G8 are the hypothetical recombinants because they are located in the corner opposite the outgroup (*E. coli*), and G2/G7 and G9 are the hypothetical parents, because they are located in the two corners of another diagonal (Supplementary Figure S2A). The PISs of one side (side 1, colored in red) of this reticule are 75, 76, 77, 93, 95, 226, 459, 460, 473, 474, 1,355, and 1,367, and the PISs of another side (side 2, colored in blue) are 987, 988, 1,001, 1,039, 1,217, and 1,218 (Supplementary Figure S2A). The arrangement of PISs on sides 1 and 2 is as follows (the PISs from side 2 are shown in bold typeface):

75, 76, 77, 93, 95, 226, 459, 460, 473, 474, **987, 988, 1,001, 1,039, 1,217, 1,218**, 1,355, 1,367.

This consecutive distribution of the PISs strongly suggests that the G2/G7 group and G9 recombinantly created the G6/G8 group, and the recombination points are located between PISs 474 and 987, and 1,218 and 1,355, suggesting the occurrence of two recombinations.

To confirm the non-random distribution of these PISs statistically, we employed the two-sample runs test (Takahata, 1994; Ezawa et al., 2006). A “run” is defined as a sequence of letters of the same kind bounded by letters of another kind,



with the exception of the first and last positions (Saitou and Yamamoto, 1997). In this case, the total number of PISs is 20 ( $n = 20$ ), of which 14 are derived from side 1 ( $\alpha = 12$ ) and 6 from side 2 ( $\beta = 6$ ). The sequence is separated into three blocks (PISs 72–474, 987–1,218, and 1,355–1,367), so that the number of runs ( $Z$ ) is three. The probability of the number of runs becoming fewer than three is significantly low ( $P = 0.0005$ ,  $< 0.05$ ), indicating that the PISs are distributed non-randomly. The details of the calculation for each side are described in the Supporting Information.

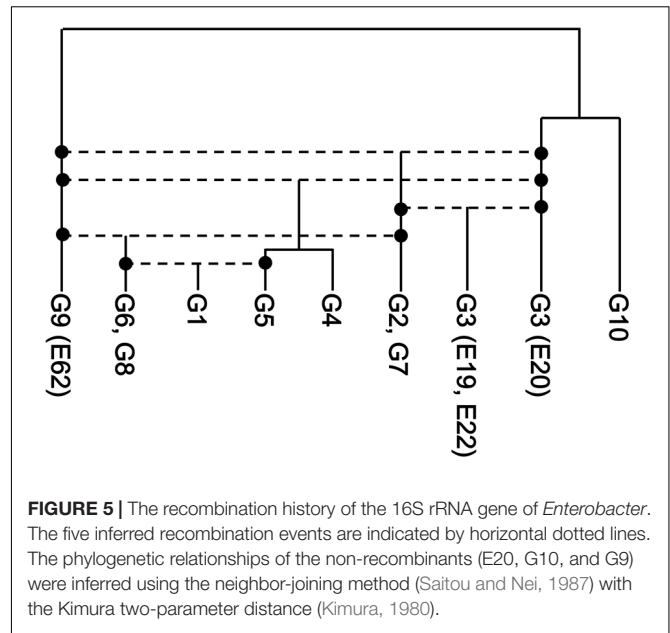
## Recombination Is a Major Driving Force for the Evolution of *Enterobacter* 16S rRNA Genes

Similar analysis was performed for the other groups in the network. We found that G1, G2/G7, G4/G5 and two of G3 (E19 and E22) were the recombinants, whose hypothetical parents are G5 and G6/G8, G3 (E20) and G9, G3 (E20) and G9 and G3 (E20) and G2/G7, respectively (Supplementary Figures S2B–E). The results of the analysis are summarized in Supplementary Table S2 (see also Supporting Information) and the evolutionary history of representative *Enterobacter* 16S rRNA genes is illustrated in **Figure 5**, in which horizontal dotted lines indicate recombination events. All 16S rRNA gene copies of G1, G2, G3, G4, G6, G7, and G8, and six of seven 16S rRNA gene copies of G9 were created by intergenomic recombination between two parents, E20 (G9) and E20 (G3). G10 shares a lineage with E20, but did not serve as a parent. Collectively, a total of five intergenomic recombination events took place in *Enterobacter*.

As illustrated in **Figure 2A**, the majority of 16S rRNA gene copies in the same genome are highly homologous, which suggests that the existence of non-homogeneous 16S rRNA genes in the same genome is not favorable and intragenomic recombination (homogenization) is a “fast” process. In fact, the frequency of intragenomic recombination in *E. coli* 16S rRNA gene copies was estimated to be  $5 \times 10^{-9}$  per generation (Hashimoto et al., 2003), whereas the spontaneous point mutation rate is estimated to be  $2.2 \times 10^{-10}$  per generation per site in the same organism (Lee et al., 2012). Thus, the “transient states” would not be frequently found in the 16S rRNA genes in *Enterobacter*, but as a rare exception we identified a possible “on-going” homogenizing process in G9. In this genome, one copy (E20) was a parent and the others (E19 and E22) were recombinants between E20 and the G2/G7 group (**Figure 4**). Thus, E20 likely pre-existed in the genome and the second parent G2/G7 integrated into the cell (either in-genome or ex-genome), followed by recombination between the alleles. Many types of chimerization might have occurred, but subsequent intragenomic recombination should have selected one of those (i.e., E19/E22 type). We speculate that the relic E20 would eventually be replaced by the E19/E22 type.

## Recombination Does Not Perturb RNA Secondary Structures

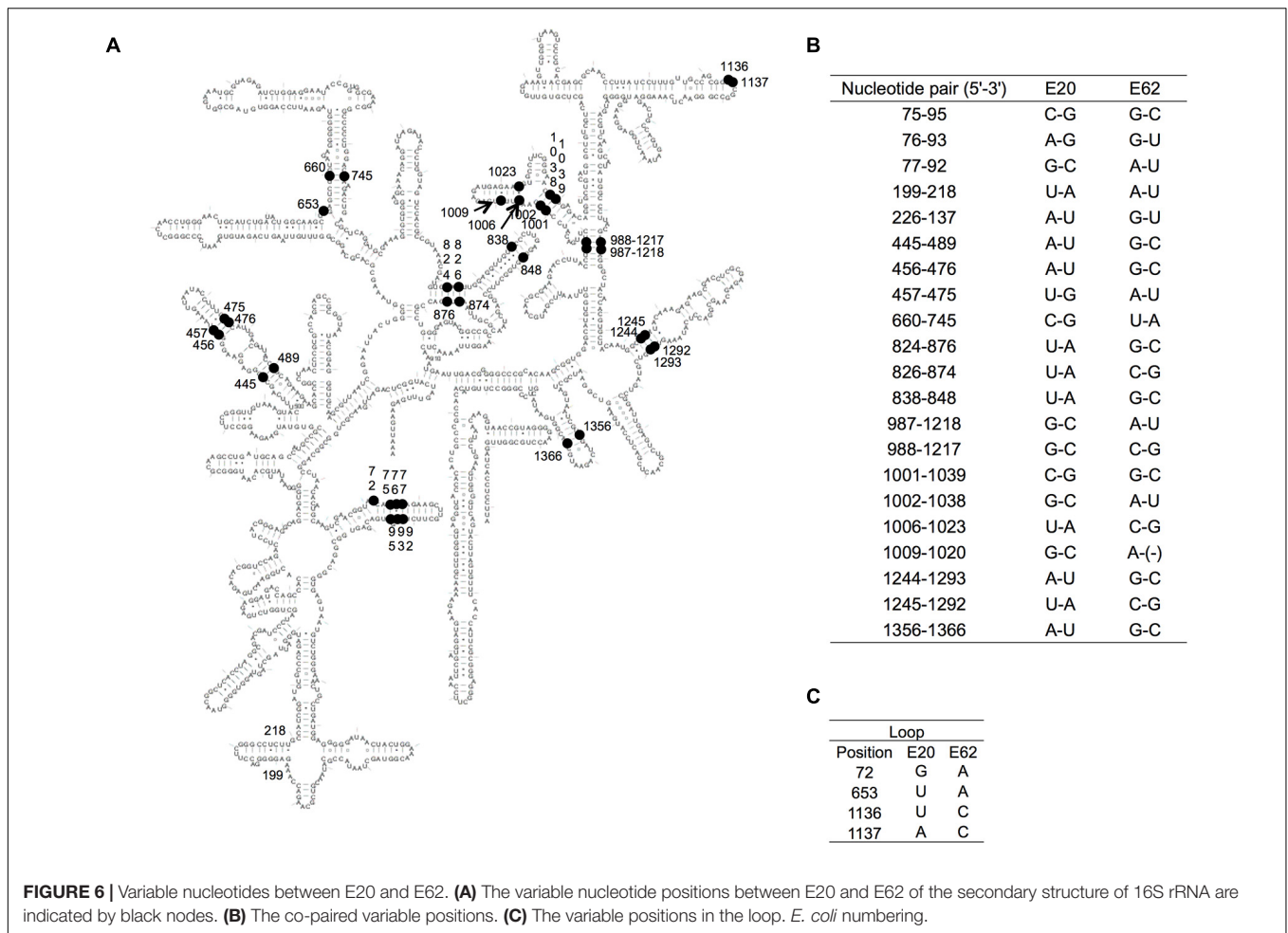
In general, secondary structures are key for the functionality of 16S rRNAs and must be conserved during evolution (Laios



et al., 2004). From this point of view, we mapped the different sites between two parental *Enterobacter* 16S rRNAs, E20 (of G3) and E62 (of G9) (**Figure 6A**). There were 46 different nucleotides, of which 42 were positioned in stem regions, which co-varied to maintain the base pairs in each recombinant (**Figure 6B**). With a sole exception, the 1,020th nucleotide (in h33) was C in E20 and missing (gap) in E62, which was found only in E20 in 19 representatives. Four nucleotides were located in loop regions (72, 653, 1,136, and 1,137) (**Figure 6C**). Our previous study suggests that all of these nucleotides are variable (72R, 653N, 1136N, 1137H) in *E. coli* 16S rRNA (Kitahara et al., 2012), suggesting them to be functionally neutral.

## Notes on Previous Studies

With the rapid increase in the number of bacterial genomic sequences, attempts have been made to investigate the occurrence of HGT in 16S rRNA genes. Tian et al. (2015) focused on the complete prokaryotic genomes and concluded that HGT is a rare event (Tian et al., 2015). More specifically, they first retrieved 2,143 complete prokaryotic genomes. They then selected 28 genomes that showed the pairwise identities between 16S rRNA genes that were less than 98%. They assumed that such “dissimilar” 16S rRNA genes could be brought by HGT from different bacterial species. Next, to investigate the origin of the dissimilarity, they used each 16S rRNA gene sequence in the 28 genomes as a query and BLAST searched for the top 50 hits in the NCBI database. They then reconstructed a maximum likelihood tree of these hits and queries. When the query sequence positioned in a cluster different from the host lineage, they considered that the query gene was brought by HGT. For example, they found one of the two copies of the 16S rRNA gene (copy A) of strain *Chlamydia trachomatis* clustered with



*Chlamydia trachomatis* sequences, and the other (copy B, 97.1% identical to copy A) clustered with *C. suis* sequences with only one mismatch over 1,520 bases. Based on this, they concluded that the copy B was horizontally transferred from *C. suis* to the *C. trachomatis* genome (Tian et al., 2015).

However, their analysis (and thus the deduced results) suffers from several fatal errors. First, their analysis focused only on some specific states of the whole HGT process (states B and D in **Figure 1**). A high rate of intragenomic recombination is already known (Hashimoto et al., 2003) and their analysis simply approved the results. Additionally, some argue that all copies of the 16S rRNA sequences are chimeric (Schouls et al., 2003), and that they originated by HGT followed by intragenomic recombination (state E in **Figure 1**). However, Tian's criteria (i.e., 2% dissimilarity) could overlook such chimeric sequences. In this study, we also found that, out of the 10 genomes, 8 *Enterobacter* genomes contained chimeric 16S rRNA gene copies that were nearly identical (<0.4% dissimilarity). The 28 genomes that Tian et al. (2015) identified may be the transition state of intragenomic recombination, which is certainly rare as suggested by previous studies (Hashimoto et al., 2003).

The analytical procedure used after the selection of the 28 genomes is also problematic, in that they used Blast search (using individual 16S rRNA gene sequences as a query) to identify genomes. However, in the case of segmental HGT (which usually results in the creation of chimeric genes), the segmental sequence will not be recovered from the database due its short length.

Furthermore, the phylogenetic tree method is inadequate for the detection of recombination as it is beyond the application limit as described above. Overall, their analysis will underestimate the frequency of HGT.

## CONCLUSION

We identified at least five intergenomic recombination events within the *Enterobacter* genome. Taking the high susceptibility of 16S rRNAs for point mutations, which can disrupt secondary structure (Laios et al., 2004), into account, it is reasonable to think that the evolution of 16S rRNA genes was facilitated by HGT. Recombination-based evolution (intergenomic, followed by intragenomic) effectively generated a higher order of sequence diversity, shaping a specific phylogenetic group, the genus *Enterobacter*.



In this study, we investigated the possibility (and frequency) of HGT in 16S rRNA genes using a bioinformatic approach. Very recently, through an experimental approach, we found an unexpectedly high functional compatibility between the 16S rRNA of *E. coli* and Acidobacteria, which are different at the phylum level (Tsukuda et al., 2017). Based on the high compatibility of 16S rRNAs, we proposed a new evolutionary model for 16S rRNA (and the ribosome). The model, which we named the “Cradle Model,” suggests that bacterial 16S rRNAs are significantly connected by a neutral network, and thus, functional interactions can be maintained before and after interspecies exchange of 16S rRNA. Taken together, our results strongly suggest that 16S rRNA genes are amenable to HGT.

## AUTHOR CONTRIBUTIONS

MS and KM designed the study, MS conducted the data analysis and MS and KM wrote the manuscript.

## REFERENCES

- Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V., and Polz, M. F. (2004). Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. *J. Bacteriol.* 186, 2629–2635. doi: 10.1128/JB.186.9.2629-2635.2004
- Bandelt, H. (1994). Phylogenetic networks. *Verh. Natwiss Ver. Hamburg* 34, 51–71.
- Boucher, Y., Douady, C. J., Sharma, A. K., Kamekura, M., and Doolittle, W. F. (2004). Intrageneric heterogeneity and intergeneric recombination among haloarchaeal rRNA genes. *J. Bacteriol.* 186, 3980–3990. doi: 10.1128/jb.186.12.3980-3990.2004
- Brady, C., Cleenwerck, I., Venter, S., Coutinho, T., and De Vos, P. (2013). Taxonomic evaluation of the genus *Enterobacter* based on multilocus sequence analysis (MLSA): proposal to reclassify *E. nimipressuralis* and *E. amnigenus* into *Lelliottia* gen. nov. as *Lelliottia nimipressuralis* comb. nov. and *Lelliottia amnigena* comb. nov., respectively, *E. gergoviae* and *E. pyrinus* into *Pluralibacter* gen. nov. as *Pluralibacter gergoviae* comb. nov. and *Pluralibacter pyrinus* comb. nov., respectively, *E. cowanii*, *E. radincitans*, *E. oryzae* and *E. arachidis* into *Kosakonia* gen. nov. as *Kosakonia cowanii* comb. nov., *Kosakonia radincitans* comb. nov., *Kosakonia oryzae* comb. nov. and *Kosakonia arachidis* comb. nov., respectively, and *E. turicensis*, *E. helveticus* and *E. pulveris* into *Cronobacter* as *Cronobacter zurichensis* nom. nov., *Cronobacter helveticus* comb. nov. and *Cronobacter pulveris* comb. nov., respectively, and emended description of the genera *Enterobacter* and *Cronobacter*. *Syst Appl Microbiol.* 36, 309–319. doi: 10.1016/j.syapm.2013.03.005
- Brodersen, D. E., Clemons, W. M., Carter, A. P., Wimberly, B. T., and Ramakrishnan, V. (2002). Crystal structure of the 30 S ribosomal subunit from *Thermus thermophilus*: structure of the proteins and their interactions with 16 S RNA. *J. Mol. Biol.* 316, 725–768. doi: 10.1006/jmbi.2001.5359
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334
- Eardly, B., Wang, F., and Berkum, P. V. (1996). Corresponding 16S rRNA gene segments in Rhizobiaceae and *Aeromonas* yield discordant phylogenies. *Plant Soil* 186, 69–74. doi: 10.1007/978-94-011-5700-1\_9
- Eardly, B. D., Nour, S. M., van Berkum, P., and Selander, R. K. (2005). Rhizobial 16S rRNA and *dnaK* genes: mosaicism and the uncertain phylogenetic placement of *Rhizobium galegae*. *Appl. Environ. Microbiol.* 71, 1328–1335. doi: 10.1128/aem.71.3.1328-1335.2005
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Ezawa, K., Ikeo, K., Gojobori, T., and Saitou, N. (2010). Evolutionary pattern of gene homogenization between primate-specific paralogs after human and

## FUNDING

This work was supported in part by a JSPS Grant-in-Aid for Scientific Research (B) (Grant 26292048 to KM), Grant-in-Aid for Scientific Research on Innovative Areas (Grant 26670219 to KM), and Grant-in-Aid for Challenging Exploratory Research (Grant 15H01072 to KM).

## ACKNOWLEDGMENT

We thank Professor Naruya Saitou for discussion of our findings.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02225/full#supplementary-material>

- macaque speciation using the 4-2-4 method. *Mol. Biol. Evol.* 27, 2152–2171. doi: 10.1093/molbev/msq109
- Ezawa, K., Oota, S., and Saitou, N. (2006). Genome-wide search of gene conversions in duplicated genes of mouse and rat. *Mol. Biol. Evol.* 23, 927–940. doi: 10.1093/molbev/msj093
- Hashimoto, J. G., Stevenson, B. S., and Schmidt, T. M. (2003). Rates and consequences of recombination between rRNA operons. *J. Bacteriol.* 185, 966–972. doi: 10.1128/JB.185.3.966-972.2003
- 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. doi: 10.1126/science.1987647
- Jain, R., Rivera, M. C., and Lake, J. A. (1999). Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 96, 3801–3806. doi: 10.1073/pnas.96.7.3801
- Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275–282. doi: 10.1093/bioinformatics/8.3.275
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120. doi: 10.1007/BF01731581
- Kitahara, K., and Miyazaki, K. (2011). Specific inhibition of bacterial RNase T2 by helix 41 of 16S ribosomal RNA. *Nat. Commun.* 2, 549. doi: 10.1038/ncomms1553
- Kitahara, K., and Miyazaki, K. (2013). Revisiting bacterial phylogeny: Natural and experimental evidence for horizontal gene transfer of 16S rRNA. *Mob. Genet. Elements* 3:e24210. doi: 10.4161/mge.24210
- Kitahara, K., Yasutake, Y., and Miyazaki, K. (2012). Mutational robustness of 16S ribosomal RNA, shown by experimental horizontal gene transfer in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19220–19225. doi: 10.1073/pnas.1213609109
- Kitano, T., Blancher, A., and Saitou, N. (2012). The functional A allele was resurrected via recombination in the human ABO blood group gene. *Mol. Biol. Evol.* 29, 1791–1796. doi: 10.1093/molbev/mss021
- Kitano, T., Noda, R., Takenaka, O., and Saitou, N. (2009). Relic of ancient recombinations in gibbon ABO blood group genes deciphered through phylogenetic network analysis. *Mol. Phylogenet. Evol.* 51, 465–471. doi: 10.1016/j.ympev.2009.02.023
- Kitano, T., and Saitou, N. (1999). Evolution of Rh blood group genes have experienced gene conversions and positive selection. *J. Mol. Evol.* 49, 615–626. doi: 10.1007/PL00006583

- Klein, D. J., Moore, P. B., and Steitz, T. A. (2004). The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J. Mol. Biol.* 340, 141–177. doi: 10.1016/j.jmb.2004.03.076
- Laios, E., Waddington, M., Saraiya, A. A., Baker, K. A., O'Connor, E., Pamarathy, D., et al. (2004). Combinatorial genetic technology for the development of new anti-infectives. *Arch. Pathol. Lab. Med.* 128, 1351–1359. doi: 10.1043/1543-2165(2004)128<1351:CGTFTD>2.0.CO;2
- Lan, R., and Reeves, P. R. (1998). Recombination between rRNA operons created most of the ribotype variation observed in the seventh pandemic clone of *Vibrio cholerae*. *Microbiology* 144, 1213–1221. doi: 10.1099/00221287-144-5-1213
- Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L., and Pace, N. R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl. Acad. Sci. U.S.A.* 82, 6955–6959. doi: 10.1073/pnas.82.20.6955
- Lee, H., Popodi, E., Tang, H., and Foster, P. L. (2012). Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2774–E2783. doi: 10.1073/pnas.1210309109
- Liao, D. (1999). Concerted evolution: molecular mechanism and biological implications. *Am. J. Hum. Genet.* 64, 24–30. doi: 10.1086/302221
- Liao, D. (2000). Gene conversion drives within genic sequences: concerted evolution of ribosomal RNA genes in bacteria and archaea. *J. Mol. Evol.* 51, 305–317. doi: 10.1007/s002390010093
- Lopez-Lopez, A., Benlloch, S., Bonfa, M., Rodriguez-Valera, F., and Mira, A. (2007). Intragenomic 16S rDNA divergence in *Haloarcula marismortui* is an adaptation to different temperatures. *J. Mol. Evol.* 65, 687–696. doi: 10.1007/s00239-007-9047-3
- Miyazaki, K., Sato, M., and Tsukuda, M. (2017). PCR primer design for 16S rRNAs for experimental horizontal gene transfer test in *Escherichia coli*. *Front. Bioeng. Biotechnol.* 5:14. doi: 10.3389/fbioe.2017.00014
- Mylvaganam, S., and Dennis, P. P. (1992). Sequence heterogeneity between the two genes encoding 16S rRNA from the halophilic archaeobacterium *Haloarcula marismortui*. *Genetics* 130, 399–410.
- Saitou, N., and Kitano, T. (2013). The PNarec method for detection of ancient recombinations through phylogenetic network analysis. *Mol. Phylogenet. Evol.* 66, 507–514. doi: 10.1016/j.ympev.2012.09.015
- Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425. doi: 10.1093/oxfordjournals.molbev.a040454
- Saitou, N., and Yamamoto, F. (1997). Evolution of primate ABO blood group genes and their homologous genes. *Mol. Biol. Evol.* 14, 399–411. doi: 10.1093/oxfordjournals.molbev.a025776
- Schouls, L. M., Schot, C. S., and Jacobs, J. A. (2003). Horizontal transfer of segments of the 16S rRNA genes between species of the *Streptococcus anginosus* group. *J. Bacteriol.* 185, 7241–7246. doi: 10.1128/jb.185.24.7241-7246.2003
- Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., et al. (2005). Structures of the bacterial ribosome at 3.5 Å resolution. *Science* 310, 827–834. doi: 10.1126/science.1117230
- Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R., and Schmidt, T. M. (2015). rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res.* 43, D593–D598. doi: 10.1093/nar/gku1201
- Takahata, N. (1994). Comments on the detection of reciprocal recombination or gene conversion. *Immunogenetics* 39, 146–149. doi: 10.1007/BF00188618
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Tian, R. M., Cai, L., Zhang, W. P., Cao, H. L., and Qian, P. Y. (2015). Rare events of intragenus and intraspecies horizontal transfer of the 16S rRNA gene. *Genome Biol. Evol.* 7, 2310–2320. doi: 10.1093/gbe/evv143
- Tomiki, T., and Saitou, N. (2004). Phylogenetic analysis of proteins associated in the four major energy metabolism systems: photosynthesis, aerobic respiration, denitrification, and sulfur respiration. *J. Mol. Evol.* 59, 158–176. doi: 10.1007/s00239-004-2610-2
- Tsukuda, M., Kitahara, K., and Miyazaki, K. (2017). Comparative RNA function analysis reveals high functional similarity between distantly related bacterial 16S rRNAs. *Sci. Rep.* 7:9993. doi: 10.1038/s41598-017-10214-3
- Ueda, K., Seki, T., Kudo, T., Yoshida, T., and Kataoka, M. (1999). Two distinct mechanisms cause heterogeneity of 16S rRNA. *J. Bacteriol.* 181, 78–82.
- Wang, Y., and Zhang, Z. (2000). Comparative sequence analyses reveal frequent occurrence of short segments containing an abnormally high number of non-random base variations in bacterial rRNA genes. *Microbiology* 146, 2845–2854. doi: 10.1099/00221287-146-11-2845
- Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* 51, 221–271.
- Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U.S.A.* 87, 4576–4579. doi: 10.1073/pnas.87.12.4576
- 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.
- Zhang, R., and Lin, Y. (2009). DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. *Nucleic Acids Res.* 37, D455–D458. doi: 10.1093/nar/gkn858

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Sato and Miyazaki. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.