Revisiting bacterial phylogeny
Natural and experimental evidence for horizontal gene transfer of 16S rRNA

Kei Kitahara1 and Kentaro Miyazaki1,2,*

1Bioproduction Research Institute; National Institute of Advanced Industrial Science and Technology (AIST); Sapporo, Japan; 2Department of Medical Genome Sciences; Graduate School of Frontier Sciences; The University of Tokyo; Sapporo, Japan

Keywords: horizontal gene transfer, 16S ribosomal RNA, complexity hypothesis, phylogenetics

Abbreviations: HGT, horizontal gene transfer; rRNA, ribosomal RNA

Introduction

Conventional genetics assumes that each gene is “vertically” inherited from an ancestor to its descendants. However, recent development of genomics and bioinformatics has shown that horizontal gene transfer (HGT) frequently occurs, particularly among prokaryotes. These studies have revealed that prokaryotic genomes are mosaics of DNA fragments from different origins.1-4 HGT involves the addition or exchange of genomic fragments from one species to another5-7 through four known mechanisms:

1. transfer of DNA mediated by bacteriophage infection (transduction),
2. transfer of DNA by mating of different species (conjugation),
3. uptake of extracellular DNA fragments due to the natural competency of the cell (transformation) and
4. transfer of a DNA fragment mediated by a virus-like element (gene transfer agents).

Once incorporated into the genome, the transferred genomic fragments can be vertically inherited by descendants, generating species diversity. HGT often promotes accelerated evolution or drastic phenotypic changes. For example, 20% of the genomic DNA of the enterohemorrhagic Escherichia coli strain O157:H7 is transferred from non-E. coli species primarily by transduction to obtain pathogenic characteristics.6

How are Prokaryotic Species Defined?

In organisms that reproduce sexually, a species is often defined as a group of organisms that are capable of interbreeding and producing fertile offspring. This raises the questions regarding prokaryotic species, which reproduce asexually but often exchange genes between organisms through HGT. Historically, prokaryotic species have been defined for convenience according to their phenotypes such as morphology, metabolism and staining properties; these criteria do not fit well with the definition of a species as in sexual organisms. Prokaryotic species are currently defined relatively arbitrarily according to criteria. For example, a species of prokaryotes is defined as a group of microorganisms that share at least 70% homology according to DNA-DNA hybridization studies.7

Use of 16S rRNA Sequence for Classifying Prokaryotes

In 1970s, researchers began to challenge the taxonomic classification of prokaryotes based on the nucleotide sequence of 5S rRNA or amino acid sequences of proteins such as cytochrome c and ferredoxin. In 1977, Carl R. Woese classified prokaryotes based on the oligonucleotide catalog of 16S rRNA that make up the 30S (small) ribosomal subunit, revealing that prokaryotes are composed of 2 different domains, bacteria and archaea.8-9

Development of recombinant DNA technology and sequencing
function as essential complexes with multiple components (e.g., the ribosome). The complexity hypothesis assumes that such informational genes generally coevolve with their binding partners in the same complex. For example, let us consider a complex consisting of protein A and A'; if component A is mutated, component A' may mutate in a compensatory manner so as not to lose complete functionality of the complex. This process makes the relationship between A and A' more specific. It should be noted that genes encoding elements in a complex (i.e., A and A') are often dispersed throughout the genome (Fig. 1A), and successful HGT of A requires simultaneous HGT of A' from a different locus. As the number of components in a complex increases, the occurrence of HGT becomes less likely. Informational genes (as compared with single functional genes) in a species are thus thought to represent the species' genomic background. This makes informational genes more suitable for constructing the tree of life.

Complexity of 16S rRNA in the Ribosome Structure

The ribosome is a universal translational apparatus and its 30S subunit, in which the 16S rRNA is located, plays a role in monitoring recognition of codons on mRNA through the anti-codons of tRNAs. 16S rRNA interacts with as many as 21 different ribosomal proteins (Fig. 1A) to construct the 30S subunit. In vitro reconstitution experiments have shown that ribosomal proteins hierarchically interact with 16S rRNA and are incorporated into the intermediate particle of the subunit as shown in the assembly map (Fig. 1B), although this hierarchy, known as the “assembly gradient” is not essential in vivo. The crystal structure model of the 30S subunit clearly shows that 16S rRNA

Figure 1. (A) Genome-wide distribution of rRNA operon and ribosomal protein-encoding genes. Location of ribosomal protein genes involved in 30S subunit (S1–S21) and 7 rRNA operons on the genomic map of *E. coli* (0–100 min). (B) Assembly map of the 30S subunit. Assembly order of ribosomal protein to 16S rRNA is shown by arrows. Figure was constructed as described by Culver et al. 
E. coli 30S ribosome

Foreign (metagenomic) 16S rRNAs

Deletion of E. coli 16S rRNA gene

Functional hybrid ribosome

A

B

16S rRNA gene from metagenome

Transformation

Selection on Zeocin plate

E. coli 16S rRNA gene

pRB103

Zeo"

16S rRNA gene from metagenome

Counter-selection of pRB101

KT101

pRB101

Amp'

KT103

pRB103

Zeo"

Figure 2. (A) Schematic illustration of laboratory HGT experiment of 16S rRNA. Escherichia coli ribosomal 30S subunits (green, 16S rRNA; white, ribosomal proteins) were genetically modified to replace 16S rRNA with foreign 16S rRNA (orange, blue and cyan). Through functional screening, 16S rRNA genes that are active in E. coli can be selected (orange). (B) Escherichia coli genetic system for laboratory HGT experiment of 16S rRNA. Escherichia coli Δ7 strain KT101 was used to screen for foreign (metagenomically retrieved) 16S rRNA genes compatible with an E. coli genetic background. The rescue plasmid contained the rmb operon of E. coli and the counterselectable marker sacB. Transformants containing pRB103 were referred to as KT103.

Evidence of HGT of 16S rRNA in Nature: Counterexamples to the Complexity Hypothesis

rRNAs are transcribed as an operon (rrn operon), with the transcript processed by various kind of endo- or exoribonucleases to generate 16S rRNA, 23S rRNA, 5S rRNA and some tRNAs, although some exceptions exist. Diversity exists in the copy numbers of rrn operons in prokaryotes; e.g., Mycoplasma pneumonia has only a single copy, whereas E. coli has 7 and Clostridium paradoxum has as many as 15 copies. In E. coli, differences in

forms the subunit structural core. The conserved region of 16S rRNA is sensitive to mutation; even subtle mutations (i.e., single point mutation) often render the ribosome completely non-functional. It is thus reasonable for evolutionary biologists to hypothesize that HGT of 16S rRNA does not occur because it requires hundreds of concurrent nucleotide mutations. In contrast, genetic studies have shown that 16S or 23S rRNA is somewhat flexible when mutations occur within variable regions (regions not conserved between species), including binding sites for some ribosomal proteins.
16S rRNA sequences among seven rrn operons (rrnA–H) are less than 1%. Interestingly, several isolated species of prokaryotes contains rrn operons that show significant diversity. For example, the thermophilic bacterium Thermoaerobacter tengcongensis contains four rrn operons, and the 16S rRNA sequence of rrnC shows only 88.4% homology with other rrn operons (rrnA, rrnB and rrnD). The 16S rRNA gene of rrnC is not a pseudogene because it correctly forms the conserved secondary structure of 16S rRNA. The sequence is more closely related (95%) to that of other Thermoaerobacter species (T. subterraneus SL9 and T. keratinophilus 2KXI), suggesting that this gene was transferred into the ancestor of T. tengcongensis and became fixed as a functional gene in this organism over evolution. Similar interoperonic diversity is observed in Desulfothermus kuznetsovi (91.7% homology), Thermodisuibispora bispora R51 (93.6%), Thermostreptus chromogena (94%) and Halobacterium marismortui (95%). These examples can be regarded as snapshots of ongoing HGT of 16S rRNA in nature and are thought to be clear counterexamples of the complexity hypothesis. Chimeric 16S rRNA genes have also been reported between two distantly related species or closely related lineages, suggesting that partial fragments of the 16S rRNA genes are also amenable to HGT.

It remains unclear how HGT of 16S rRNA is driven in the environment. 16S rRNA genes may be propagated by transduction because, according to a previous report, this gene can be incorporated into a broad-host-range, generalized transducing phage (SN-T) at a frequency of 1 × 10^9 transductants/PFU. The existence of several universally conserved sequences in the rrn operon may allow and facilitate recombination of a foreign rrn operon with that of the host after incorporation into the genome. An experimental study showed that 16S rRNA can also undergo HGT via transformation by showing that Helicobacter pylori acquires a tetracycline-resistant phenotype when the 16S rRNA gene containing multiple point mutations, which render this gene resistant to tetracycline, is added to the medium.

HGT Experiment Revealed Unexpected Robustness of the 16S rRNA Gene

The evidence described above that HGT of 16S rRNA has occurred in nature raises doubt on the basic assumption that 16S rRNA is unlikely to undergo HGT between species. To address this issue, we performed systematic HGT experiments for 16S rRNA using E. coli by introducing homologs of this gene that had been amplified from metagenomes extracted from various environments (i.e., soils, fermented foods, sea water and hot springs) (Fig. 2). Metagenomes are mixtures of uncultured microbial genomes (and plasmids) in environmental samples rich in microbial diversity. Because both ends (5' and 3') of 16S rRNA genes show high homology among prokaryotes, we amplified 16S rRNA genes from various (phylogenetically diverse) species by PCR using a universal primer pair. These DNA fragments were cloned into the rrn operon-expression vector pRB103 and introduced into an E. coli strain KT101 (“Δ7 prrn strain”), in which all chromosomal rrn operons were inactivated and complemented by a counter-selectable plasmid pRB101 encoding the wild-type rrnB operon. Using this system, we performed genetic selection of 16S rRNA genes that are functional in E. coli. From the 15,000 transformants, we could obtain 33 different clones (KT103) carrying the metagenomic 16S rRNA genes rather than the E. coli 16S rRNA gene. Through DNA sequencing, we found that most 16S rRNA genes were derived from Gammaproteobacteria, but some were derived from Betaproteobacteria with 80.9% identity to the E. coli sequence (note that the phylogenetic class of E. coli is Gammaproteobacteria; Fig. 3). Doubling times (DT) of KT103 strains carrying foreign 16S rRNA genes (30–50 min) were comparable to those of KT101 carrying E. coli 16S rRNA (30 min). DTs of each KT103 strain increased only modestly with decreasing sequence identity; the average increase in DT was 4.6 sec per point mutation. Because severe effects caused by mutation within the rRNA gene are known to be suppressed relatively easily through a spontaneous compensatory mutation (e.g., after continuous cultivation for two weeks), smaller defects caused by HGT of 16S rRNA may also be circumvented by similar adaptation mechanisms. Our results indicate that the ribosome is an unexpectedly plastic complex and sufficiently robust to accommodate foreign 16S rRNA.

Conclusions and Perspectives

Recent progress in genomics and bioinformatics has shown that natural 16S rRNA genes are likely to experience HGT during the course of evolution. In our HGT experiment, we showed that E. coli can accommodate a foreign 16S rRNA gene with only 80.9% identity that belonged to different phylogenetic class. This evidence strongly suggests that the 16S rRNA gene is amenable to HGT, thereby questioning the theoretical background of microbial phylogeny (e.g., the complexity hypothesis); the 16S rRNA gene may well be represented in a phylogenetic network. Results of the current method for rapid identification and classification of prokaryotes based on the 16S rRNA gene should thus be carefully analyzed and interpreted. Unexpectedly high horizontal transferability of the 16S rRNA gene suggests structural flexibility or plasticity of the ribosome because HGT introduces hundreds of mutations into the 16S rRNA at once. This finding accelerates the study of ribosome engineering mediated by 16S rRNA replacement. Because it has been known that subtle alterations in ribosomal components can induce large phenotypic changes in bacteria, the 16S rRNA gene may be a promising target for molecular engineering to manipulate bacterial phenotypes.

Disclosure of Potential Conflicts of Interest
No potential conflict of interest was disclosed.
Figure 3. Neighbor-joining phylogenetic tree of 16S rRNA genes. Environmental 16S rRNA genes that were functional in E. coli are shown as the clone ID (A01–H03) with their closest relatives (nomenclature). Several other relevant strains are also shown.
