Korean Journal of Microbiology (2013) Vol. 49, No. 4, pp. 383-390
DOI http://dx.doi.org/10.7845/kjm.2013.3091

| 보 문 |

Characterization of Phylogenetic Incongruence among Protein Coding Genes of *Vibrio* Strains Pathogenic to Humans

Young-Gun Zo

Department of Biology, Kyungsung University, Busan 608-736, Republic of Korea

인체 병원성 비브리오 균주간 유전자 계통의 불일치성 분석

조영근

경성대학교 생물학과

(Received December 13, 2013 / Accepted December 27, 2013)

Lateral gene transfer (LGT) of genes from other bacteria into *Vibrio cholerae* is expectable because of the pronounced natural competence of the bacterium. In this study, quantitative aspects of LGT among the three species of *Vibrio* pathogenic to humans were characterized. Genome sequences of *V. cholerae* N16961, *V. parahaemolyticus* RIMD2210633, *V. vulnificus* CMCP6, and *Escherichia coli* K12 substrain MG1655 were analyzed to determine orthologous quartets of protein coding genes present in all four genomes. Phylogenetic analyses on the quartets were conducted to resolve vertical versus lateral patterns of gene polymorphisms based on congruence versus incongruence of phylogenetic trees. About 70% of the quartets could be resolved as either cohesive topology (75%) or LGT tree topologies (25%). The amount of LGT genes in *Vibrio* spp. appeared to be abnormally high for a genus and comparable to those of families. Patched distributions of LGT from different donors were observed on a chromosome. In the small chromosome of *V. cholerae*, physical linkages among LGT loci spanned half the length of the chromosome. Either accumulative selection for the donor alleles in LGT or presence of large-scale LGT events was hypothesized. These findings warrant further studies on the nature of donor-specificity of LGT alleles and its influence on evolution of *Vibrio* virulence to humans.

***Keywords*:** *Vibrio cholerae*, *Vibrio parahaemolyticus*, *Vibrio vulnificus*, lateral gene transfer, quartet analysis

Vibrio cholerae has a natural competence induced by chitin (Meibom *et al.*, 2005). Recently two-step actions of the DNA-uptake machinery have been unveiled (Seitz and Blokesch, 2013). According to the findings, the DNA-uptake complex that depends on a Type IV pilus translocates foreign DNAs across outer membrane while a pilus-independent shuttling system, which is made and controlled by a 19-gene regulon, brings the DNA into cytoplasm. What is notable is that competence of *V. cholerae* in biofilms formed on chitin materials are also regulated by quorum sensing (Antonova and Hammer, 2011; Lo Scrudato and Blokesch, 2013). In nature, the natural competence will allow *V. cholerae* to uptake genetic materials from other bacteria that are present in biofilms in marine environments. As an example, weakness of *V. cholerae* species boundary for genetic resources has been demonstrated by endemic population structure of mobile genetic elements among environmental *V. cholerae* populations (Boucher *et al.*, 2011). Therefore, virulence or host-defense disarming genes of other bacteria might be laterally transferred to *V. cholerae*, generating a novel toxigenic strain. Virulence genes of *V. parahaemolyticus* and *V. vulnificus*, which are the two other well-known *Vibrio* species pathogenic to humans, can be examples of the case.

Notably, extensive lateral gene transfer (LGT) between species implies a network of co-evolving organisms, i.e., an evolutionary guild (Ochman *et al.*, 2000; Feil *et al.*, 2001). In the cases of *V. cholerae* and other *Vibrio* spp., the plasticity of their genomic makeup was pronounced, as evidenced by presence of a super-integron structure in their genome (Rowe-Magnus *et al.*, 2003; Boucher *et al.*, 2011) and abundance of island insertion sites of *V. cholerae* genomes (Chun *et al.*, 2009). The three human-pathogenic species, *V. cholerae*, *V. parahaemolyticus*, and *V. vulnificus*, thriving in

*For correspondence. E-mail: zoyful@gmail.com; Tel.: +82-51-663-4643; Fax: +82-51-627-4645

multi-species biofilms covering chitin-rich marine detritus or biomass might form such evolutionary guild, and subsequently the guild might generate a novel virulent strain via the natural competence and DNA recombination. However, most detail aspects of evolutionary-guild formation are not well known. In the case of the three pathogenic *Vibrio* spp., one might ask questions on utilization and regulation of a natural competence system by *V. parahaemolyticus* and *V. vulnificus,* the extent of gene exchange among the species, and directions of gene flow from one species to the others. In this study, the three *Vibrio* species were hypothesized to be co-evolving by mutual exchanges of protein coding genes. As a support of the hypothesis, the extent of LGT among the three *Vibrio* species was quantified.

LGT can be quantified by distinguishing genes vertically inherited from those with other origins. In this sense, gene phylogeny congruent to overall phylogeny of the genome, notably to that of 16S rRNA gene, imply cohesive inheritance of a gene while phylogenetic relationship incongruent to the genome phylogeny indicate LGT (Daubin *et al.*, 2003; Boucher *et al.*, 2011). Therefore, the number of genes showing gene phylogeny incongruent to that of 16S rRNA genes was analyzed as an estimate of abundance of LGT in *Vibrio* genomes.

Materials and Methods

Determination of orthologous proteins and orthologous quartets

Annotated complete genome sequences of *V. cholerae* N16961, *V. parahaemolyticus* RIMD2210633, *V. vulnificus* CMCP6 and *Escherichia coli* K12 substrain MG1655 were downloaded from the GenBank database at the web site of the National Center for Biotechnology Information (NCBI), Bethesda, MD (http://www.ncbi.nlm.nih.gov). Their accession numbers were AE003852-AE003853, BA000031-BA000032, AE016795-AE016796, and U00096, respectively. Protein sequences of the four genomes were made to a BLAST database (Altschul *et al.*, 1997). The best hit pairs (BeT) of a query protein and subject proteins were catalogued from the BLASTP program output, which was commanded to search all proteins of the *Vibrio* strains from their entire genomes. BeT pairs with BLAST Expect value of less than 10^{-10} were collected (Tatusov *et al.*, 2000; Clarke *et al.*, 2002), and proteins were clustered to form a single cluster by pooling any protein in the BeT pairs that had a common protein. The total of 13,204 proteins from the three genomes was pooled into 2,878 clusters, comprising one to 200 proteins per cluster. Among them, those clusters which included three proteins, each of which was from one *Vibrio* genome, were selected. Further screening was done to determine whether the three BeT were symmetric (i.e., proteins in a BeT identify each other as unique matches). Thus, clusters with three symmetrically BeT proteins, one from each species, were identified as a unique set of orthologs on *Vibrio* genomes.

To resolve the true topology of a phylogenetic tree for a set of three orthologs of *Vibrio* spp., an unequivocal outgroup is required. Proteins in the genome of *E. coli* K12 were used as the outgroup. A quartet of orthologous proteins was decided by determining the BeT of *Vibrio* orthologs on the *E. coli* K12 genome. Each of three proteins in the orthologous set of the *Vibrio* genes were queried on the *E. coli* protein database using BLASTP. When all three queries returned a unique protein from the *E. coli* genome, with the BLAST Expect value of less than 10^{-10}, the four proteins were accepted as a quartet of orthologous proteins in the four species.

Analysis on topology of quartets

The four protein sequences were aligned by CLUSTALW. The tree topology of the quartets was evaluated by the maximum likelihood (ML) method implemented in the TREEPUZZLE version 5.2 (Schmidt *et al.*, 2002). It weighs the three possible tree topologies (Fig. 1) by their posterior probabilities (P_i). The probabilities were determined as $P_i = L_i / (L_1 + L_2 + L_3)$ by Bayes' theorem, where i is 1, 2, or 3 representing one of the three possible trees and L_i is the maximum likelihood of a tree (Strimmer and von Haeseler, 1997). When the probability of one of the three possible trees was larger than 0.95, the topology is accepted as fully resolved.

Statistical analyses

To analyze uniform or differential distributions of genes with different quartet topologies on the chromosomes of *V. cholerae*, Rao's spacing test for uniformity in circular space, implemented in the S-PLUS library CIRCSTAT (Jammalamadaka and SenGupta, 2001), was used. The Fisher-Freeman-Halton exact test on contingency tables was employed using the STATXACT version 6 (Cytel Software, USA), either as an exact method or Monte Carlo approximation.

Results

Topologies of quartets

As shown in Fig. 1, three kinds of topologies can be constructed from a quartet of homologous genes. The topology of a gene tree comprising three *Vibrio* genes and a truthful outgroup can produce one of the three outcomes when phylogenetic information is significantly divergent to be detectible. If the phylogenetic signals are not divergent enough

($P \leq 0.95$), the tree is non-resolving.

By matching all 1,535 orthologous proteins of the *Vibrio* spp. to the *E. coli* K12 genomic protein database, 1,090 orthologous protein quartets were determined. The figure corresponded to 23–28% of the total ORFs in the three *Vibrio* genomes. Among the 1,090 quartets, trees of 769 quartets (71%) were resolved. A total of 578 quartets (75% of those resolved) produced topologies identical to the 16S rRNA gene tree (Topology B in Fig. 1). Topology A was found among 118 (15%) of the resolved quartets. The rest of the resolved orthologs (73 quartets or 10% of the resolved) were found to comprise Topology C, in which the *V. cholerae* proteins are more related to *V. parahaemolyticus* orthologs than to *V. vulnificus* orthologs. The distribution indicates that the majority of the genes on genomes of *Vibrio* followed the same path of evolution as with 16S rRNA gene, while some proportion was influenced by LGT with neighboring species. Therefore, three quarters of the core genes of *V. cholerae* appeared have evolved in cohesion with each other and to 16S rRNA gene, but one quarter of those genes were significantly influenced by neighboring species via LGT.

Comparison to other taxa

To determine the significance of the levels of cohesion versus LGT among *Vibrio* spp., the results of quartet analysis were compared with other taxonomic groups. Daubin *et al.* (2003) published quartet analysis results for various taxonomic levels: four species, four genera, and two families. *Vibrio* showed 53:18:29 ratio of orthologs congruent to 16S rRNA gene, incongruent to 16S rRNA gene, and non-revolving, respectively, and those data were compared with the published data (Fig. 2). The proportion of orthologous quartets with phylogeny congruent with the 16S rRNA gene phylogeny can be interpreted as the level of cohesion of the genes to a vertical evolutionary path. Cohesion with the 16S rRNA gene phylogeny was the universally prevalent force in intra-family and intra-genus evolution (>50% of total quartets or >75% of the resolved). In comparing those cases with *Vibrio*, two interesting points characterizing the unique status of the genus *Vibrio* were noted. The first was that *Vibrio* showed the highest level of LGT, and the second was that non-resolving cases occurred at the level of family of other taxa, rather than the level of genus. The latter can be also derived from a strong LGT. Technically, repeated LGT of a gene among various species can increase ambiguity in delineating sequence phylogeny. The presence of unique mechanisms among *Vibrio* species, i.e., the natural competence, causes extensive LGT between different *Vibrio* species. One possibility, inferred from differences in the habitat of the genera of compared taxa, is the uniqueness of the *Vibrio* species habitat. Unlike other genera in Fig. 2, the natural habitat of *Vibrio* spp. is the aquatic environment, a relatively more homogenizing (open) less-patched environment, perhaps more conducive to LGT.

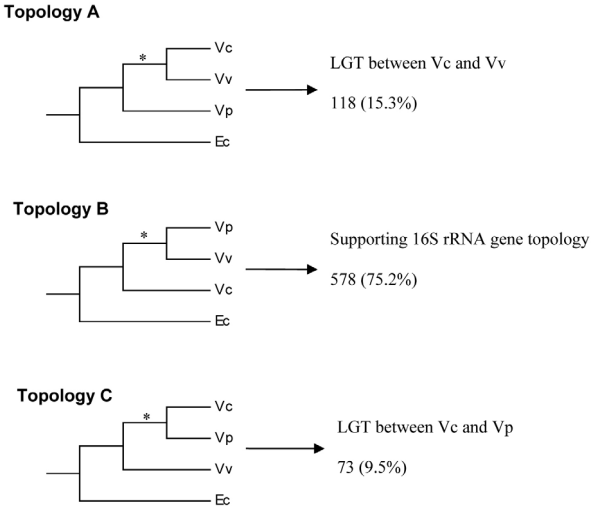

Fig. 1. Possible outcomes from a quartet comprised of *V. cholerae* (Vc), *V. vulnificus* (Vv), *V. parahaemolyticus* (Vp) and the outgroup *E. coli* (Ec). The quartet is regarded as resolved when the branch marked with * is significant by Bayesian posterior probability (P>0.95).

Distribution of LGT alleles in *V. cholerae*

A follow up question that concerns the unusually high incidence of LGT among species of the genus *Vibrio* is whether the probability of a gene being subjected to LGT is the same for all genes or if it depends on certain properties of each gene. While evenness in the frequency of sequence of LGT can indicate a genome-wide generality of LGT, distributions skewed toward a particular collection of genes support the presence of selective forces for LGT. Thus, the unevenness or the selectivity of LGT in *Vibrio* was investigated by analyzing the distribution of LGT orthologs in *Vibrio* genomes.

The distribution of orthologs by quartet topology is shown in Fig. 3. The assumption of uniform distribution was tested by Rao's spacing test for uniformity in circular space, using the middle location of a gene as its coordinate. Excluding the segment of VCR super-integrons from the small chromosome, the orthologous genes were evenly distributed (P>0.1). However, the 769 orthologs resolved from the three kinds of topologies showed significantly uneven distribution in both chromosomes (P<0.01). As noted from Fig. 3, the occurrence of segments where orthologous genes are absent or very rare, e.g., from S1 to S7 and the VCR island, were thought as the main cause of the uneven distributions. Genes in these regions are considered as species-specific or strain-specific genes

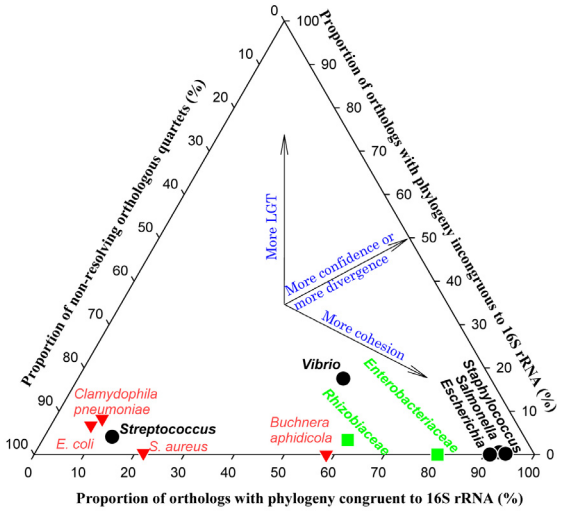

Fig. 2. Distribution of *Vibrio* and other bacterial taxonomic groups by proportion of different quartet topologies derived from protein orthologs (Texts: bold, genus or family; plain, species / Symbols: triangle, species; circle, genus; square, family).

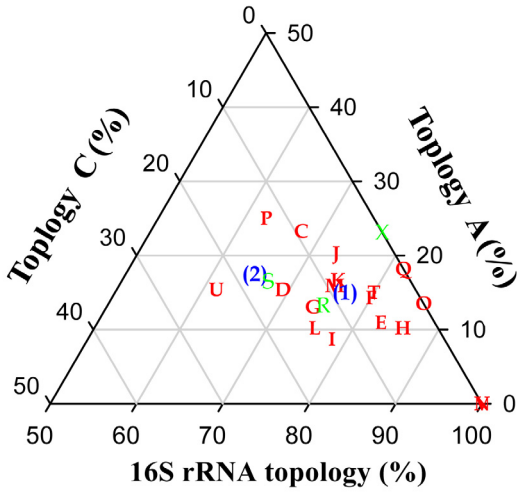

Fig. 4. Distribution of COG functional categories (A to X) and chromosomes (1 and 2) of *V. cholerae* by proportions of the three quartet-tree topologies (Symbols: 1-2, chromosomes; A, RNA processing and modification; C, energy production and conversion; D, cell cycle control, mitosis and meiosis; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; J, translation; K, transcription; L, replication, recombination and repair; M, cell wall/membrane biogenesis; N, cell motility; O, posttranslational modification, protein turnover, chaperones; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport and catabolism; R, general function prediction only; S, function unknown; T, signal transduction mechanisms; U, intracellular trafficking and secretion; V, defense mechanisms; X, not assigned).

(Chun *et al.*, 2009). To overcome this problem from the native physical coordinates of the chromosomes, circular "logical" coordinates of orthologs were created for each chromosome by serially listing orthologs in their order on the physical coordinates. There was no gap in the logical coordinates between orthologs and all orthologs were of the same size. Rao's test using the logical coordinates was performed on the ortholog groups differentiated by their quartet topologies, and the result was that none of the two LGT topologies deviated significantly from uniform distribution of LGT among common orthologs of *Vibrio* (P>0.1). Therefore, it is concluded that LGT is spatially a generalized phenomenon in *V. cholerae*, i.e.,

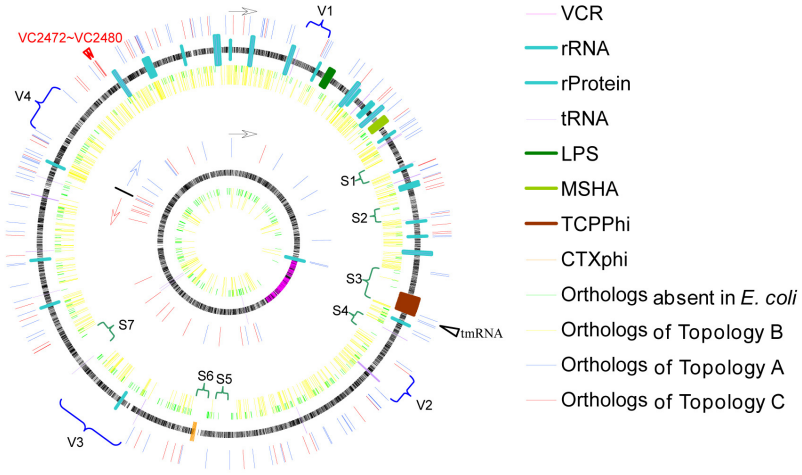

Fig. 3. Distribution of orthologous quartet genes on the two chromosomes of *V. cholerae* N16961. The two circles comprising short black spines are the ORFs of all proteins (Symbols: spines, location of ORFs listed in the inset; middle rings, ORFs and ORFs with specific function marked in the inset; inner ring, quartet orthologs with Topology B and *Vibrio* orthologs absent in *E. coli*; outer ring, quartet orthologs with Topology B and C; arrow, replication origin and direction of base numbering; S1-7, gaps of *Vibrio* orthologs; V1-4, gaps of LGT orthologs).

Table 1. Frequencies of orthologous protein quartets by topologies and chromosomes of *V. cholerae* N16961

Chromosome	Topology of 16S rRNA gene quartet (Topology B)	Topology A	Topology C	Sum
Chromosome 1	515 (505)[a]	101 (103)	56 (64)	672
Chromosome 2	63 (73)	17 (15)	17 (9)	97
Sum	578	118	73	769

[a] Expected frequencies, assuming independence between topology and chromosome

all genes of the genome being subject to LGT rather than LGT being preferential to a specific location on a chromosome.

Distribution of quartet topologies by functional groups

LGT from a given donor may be preferential and fixed by a recipient genome via natural selection. Events of LGT, followed by homologous recombination, will introduce alleles which are novel to the recipient. However, not all alleles will be fixed by a recipient, but stochastic chance and positive/negative selection to the recipient will also play a role. Stochastically fixed alleles will show a random or uniform distribution, but those fixed by selectivity will show aggregated pattern in terms of locations or functional categories. Therefore, an observation of high incidence of LGT skewed toward particular functional classes may indicate the occurrence of such adaptive events in the recipient organism.

This possibility was tested by classifying the *Vibrio* orthologs according to functional categories of clusters of orthologous groups (COG) (Tatusov *et al.*, 2000). The table of frequencies of orthologs (i.e., a table of 3 columns and 19 rows comprising the three topologies of quartets in Fig. 1 and COG categories in Fig. 4, except for X, R, and S, which are not functional but arbitrary groups) were tested using the Fisher-Freeman-Halton exact test with Monte Carlo approximation. Significance (P) was estimated as 0.54–0.57, indicating absence of dependence between COG categories and quartet topologies. Therefore, function-specific preference to a specific LGT topologies was not significant. However, this result could also arise from insufficient power of the test statistic because frequencies of LGT-prone genes for each category were low (maximum of 20 and mainly less than 12 orthologs). For the reason, distributions of COG clusters in terms of the three quartet topologies were examined (Fig. 4). The scatter graph showed a pattern in agreement with the hypothesis that genes related to environmental adaptation experienced more LGT. In the case of the COG categories of central dogma (J, K, and L), about 75% of the proteins were commonly coherent with the 16S rRNA gene phylogeny, but with a moderate level of variation (10%–20%) in frequency of the two LGT topologies. In contrast, the C, P, and U categories, related to environmental adaptation via energy conversion, ion transport and material secretion, receptively, showed higher LGT. The most extreme deviation was Category P, with high LGT from *V. vulnificus* to *V. cholerae*. It is interesting to note that genes regulating ion content of cells showed greater deviation, since salinity is a prominent factor in confining the habitat of an aquatic organism. The optimal range of salinity for different species has been reported for *Vibrio* species (Baumann *et al.*, 1984). NaCl requirements for optimal growth of *V. cholerae*, *V. vulnificus*, and *V. parahaemolyticus* were 5 mM, 130 mM, and 160 mM, respectively, indicating that LGT topology distribution of the category P might reflect habitat resemblance.

Linkage of quartet topologies

To further exam uneven distribution of LGT alleles in *Vibrio* species, linkage patterns of LGT locations in *V. cholerae* genome was examined. Because a set of genes coding for a given function often occurs in a cluster, i.e., an operon, a gene cassette or island, the recipient genome in the LGT may carry a set of adjacently-linked genes that share the same LGT phylogeny, e.g., Topology A or Topology C. The presence of such a LGT gene cluster in the *V. cholerae* genome would indicate a concerted adaptive incorporation of the foreign alleles by the recipient. Such LGT clusters were sought by examining the genome of *V. cholerae* for chromosome segments containing more than two orthologous genes of the same functional category with the same LGT topology within a 10 kb length. The segments of chromosome that met these criteria included VC2472-VC2480 (Fig. 3), the largest, containing Topology C. However, functional relatedness among the genes in the segment could not be measured because functions of those genes in the particular segment were unknown. The next large segment detected was located within a large operon coding for 24 ribosomal proteins. Among those, 5 ribosomal proteins were Topology A. Because ribosomal proteins are highly conserved and interacting directly with rRNA, this segment offers a clear example of selective pressure that lead to incorporation of the *V. vulnificus*-related alleles into *V. cholerae* genome via LGT.

Distribution of quartet topologies by chromosome

Another way to explore physical linkage among genes is to

check at distributions of genetic loci by chromosome, which is the unit of physical linkage. Table 1 provides a contingency table constructed from results of the quartet analysis. Independence between the row and column variables was tested. Because Fisher's exact test yielded significance $P<0.05$, the source of the interaction between quartet topology and chromosome was determined by comparing the observed frequencies to expected frequencies. While chromosome 1 carried orthologs consistent with expected frequency, chromosome 2 showed an increased proportion of Topology C at the expense of reduced 16S rRNA gene topology (Table 1 and Fig. 3). Therefore, chromosome 2 of *V. cholerae* can be concluded to have been influenced by *V. parahaemolyticus* more than chromosome 1. When the physical distribution of LGT orthologs on chromosome 2 was examined (Fig. 3), a very interesting, uneven distribution was observed. The distribution of Topology A LGT and Topology C LGT were highly skewed to opposite sides of the chromosome across the split (split line between blue and red arrows at 880,000 bp location, as shown in Fig. 3) and VCR islands. In the segment from the split line to the VCR island, moving in a clockwise direction, 13 orthologs were Topology A and four 4 orthologs Topology C. In the other segment, spanning clockwise from the split line to the VCR island, four orthologs were Topology A and 13 Topology C. When these frequencies were tested by Fisher's exact test, the hypothesis of independence between topologies and the segment was rejected ($P<0.01$), with the conclusion that physical linkages among LGT loci of the same topology spanned nearly half of chromosome 2 (*ca.* 500 kb). The distribution is possible when the chromosome contains homologous recombination with the long fragment (~500 kb), half the size of the small chromosome and originating from chromosomes of *V. vulnificus* or *V. parahaemolyticus*. This phenomenon merits further study since it implies a megabase transformation mechanism, such as conjugation, and selective fixation, in the small chromosome of *V. cholerae* that is regarded as a mega-plasmid.

Discussion

Scales of LGT

Phylogenetic analysis in this study could resolve vertical or horizontal gene-phylogeny of ORFs coding quartets of orthologous proteins among three *Vibrio* spp. and *E. coli*, and the number of resolved cases corresponded to 28% of the total genomic ORFs of *V. cholerae*. Three quarters (75%) of the orthologous quartets showed cohesive phylogeny while the rest (25%) showed LGT phylogeny. The amount of LGT carried by *Vibrio* spp. appeared to be abnormally high for a genus and comparable to those of families (Fig. 2), which appeared to reflect the genome-homogenizing function of the natural competence of *V. cholerae*. While both chromosomes of *Vibrio* spp. carried LGT quartets rather evenly along the locations on each chromosome, uneven distributions were also observed when the donors of LGT alleles were taken into considerations based on difference of LGT topologies, i.e., Topology A versus Topology C. Weak but detectable level of physical or genetic linkages was detected among LGT genes. On a scale of a 10 kb-sized polycistronic operon fragment, the operon of ribosomal proteins carried only one type of LGT topology. In the small chromosome of *V. cholerae*, the physical linkage was on a scale of 500 kb, half the size of the small chromosome of *V. cholerae* excluding its VCR island. The former provides an example of selective pressure in a particular direction of incorporation of the genes received via lateral gene transfer, whereas the latter implies presence of a large scale LGT. The latter case is an intriguing finding in that the small chromosome of *V. cholerae* might went on a large scale recombination twice; once with *V. parahaemolyticus* clade and once with *V. vulnificus* clade.

Influence of LGT on speciation of *Vibrionaceae*

How bacterial lineages, including species, emerge without sexual reproduction has been an intriguing question (Doolittle and Zhaxybayeva, 2009; Doolittle, 2012; Papke and Gogarten, 2012). Cohesion of genes by vertical inheritance and genetic linkage seemed to be the primal natural process on which mutations and periodic sweeps cause lineage diversity. Therefore, formation of clonal lineage under ecological selection was regarded as the mechanism of bacterial speciation. However, discovery of frequent LGT among bacterial cells added complexity in answering the question, and it still remains as a more perplexing and intriguing question. Because the distinct genome structure, i.e., harboring two chromosomes, is shared by all known species of *Vibrionaceae* family, the bacterial family is regarded as a monophylectic taxon, with an exceptionally high confidence. The monophyly of the family, together with high LGT estimates found among *Vibrio* strains, has led to enthusiastic genome sequencing and genome-wide phylogeny studies as a model case for bacterial evolution and speciation (Kahlke *et al.*, 2012; Shapiro *et al.*, 2012; Dikow and Smith, 2013). Major findings in genome-wide phylogeny included estimation of the level of cohesion of polycistronic gene clusters in the both chromosomes of diverse species (Dikow and Smith, 2013), and identification of shared core genes unique in a taxon or in an ecological niche (Kahlke *et al.*, 2012). It was shown that the levels of cohesion were >97% of total gene clusters in a given

Vibrio species while it ranged 63–71% for the large chromosomes and 49–65% for the small chromosomes among genomes of a given clade of *Vibrio* spp. In the genus *Vibrio*, two clades of species were identified as the clade C, made with *V. cholerae*, *V. furnissii*, and *V. anguillarum*, and the clade V, made with *V. vulnificus*, *V. parahaemolyticus*, and *V. campbellii*. The genomes of the two clades were distinguished by 11–19 genes unique in one of the clades while the genus *Vibrio* carried 37 unique genes. Thus, the clade segregation is ecologically supported, too. Because the three *Vibrio* strains analyzed in this study were distributed across the two clades, the level of cohesion found in this study (75%) was an estimate of inter-clade cohesion and comparable to those estimates from the previous studies, although the operational units of genes and the outgroup taxa were different, i.e., individual ORFs versus clusters of ORFs and *E. coli* versus *Shewanella oneidensis*. Overall, genes in a *Vibrio* species were highly homogeneous with very strong cohesion (>97%) while cohesion is weaker within an ecological clade of *Vibrio* spp. (<75%). Therefore, makeup of more than 22% of a genome of *Vibrio* sp. appears to be ecologically determined rather than phylogenetically.

On the other hand, there is an interesting report on estimation of cohesion and LGT within a *Vibrio* species. By comparing single nucleotide polymorphisms in 20 recently diverging *Vibrio cyclitrophicus* strains, Shapiro *et al*. (2012) estimated the level of cohesion among their core genomes as less than 1%. This startling finding implies that >99% of genes in the core genomes of a *Vibrio* species undergo LGT (Papke and Gogarten, 2012). Therefore, LGT could be the homogenizing mechanism resulting in the high level of cohesion for a species. In this interpretation of bacterial genome makeup, genes sweep through a population without much restrictions, and ecological selection is the driving force fixing a set of genes within a bacterial lineage. Therefore, the evolutionary hypothesis depicts that inter-species LGT events create ecological diversity among species while highly frequent small scale LGT among highly related strains construct species-level lineages. The hypothesis warrants future studies on ecological functions of LGT traits in the pathogenic *Vibrio* spp.

Conclusion

While the genome of *V. cholerae* is pronounced to undergo LGT due to the natural competence of the bacterium, LGT with other *Vibrio* spp. pathogenic to humans is of interest in the sense of evolution of virulence of the marine pathogens. This study demonstrated that evolution of *V. cholerae* genome is occurring in an open compartment that is open to LGT events from other related species sharing habitats with *V. cholerae*. While all genes in *V. cholerae* appear to be susceptible to LGT, selective incorporation of genes from a particular species is also expected. In the case of *V. cholerae*, greater incorporation of *V. vulnificus* than *V. parahaemolyticus* alleles was observed on the large chromosome, but both were found to contribute equally in the makeup of the composition of the small chromosome. However, uneven physical linkages of LGT in the small chromosome of *V. cholerae* indicated that the small chromosome also experienced selective incorporation of LGT alleles from specific donors. These findings warrant further study on the nature of donor-specificity of LGT alleles and its influence on evolution of *Vibrio* virulence to humans.

적 요

*Vibrio cholerae*균은 자연적으로 외부 유전자를 받아들이는 능력이 있으므로, 종간 수평적 유전자 전달 작용(LGT)을 받을 것으로 예상된다. 본 연구는 인체에 질병을 일으키는 3종의 비브리오균 사이에서 일어나는 LGT 현상의 정량적 측면들을 분석하였다. *V. cholerae* N16961, *V. parahaemolyticus* RIMD2210633, *V. vulnificus* CMCP6, *Escherichia coli* K12 substrain MG1655의 유전체 염기서열을 분석하여 4개의 유전체에 모두 존재하는 단백질 발현 유전자들의 4개 일조를 결정하였다. 각 조의 4개 유전자의 계통수를 작성하는 분석을 통하여, 다른 조들 간의 계통성의 일치성과 불일치성을 결정하고, 수직적 계통성과 수평적 계통성을 구분하였다. 약 70%의 조에서 계통수가 확정될 수 있었으며, 그 중 75%는 서로 일치하는 계통성을 보였고, 25%는 LGT 계통수를 보였다. 이 결과에 따르면, 비브리오균의 LGT는 다른 세균 분류군의 속보다는 과단위에서 발생하는 빈도의 LGT계통수를 보였다. 염색체별로 관찰하였을 때, 유전자 제공자별로 LGT가 집중되는 현상이 일부 관찰되었고, *V. cholerae* 균주의 작은 염색체에서는 염색체의 약 절반 길이에 해당하는 부분에서 제공자별 LGT 위치들이 집중되는 현상을 보였다. 이런 결과는 유전자 제공자에 따라 선택성이 반복적으로 작용하거나, 대규모의 LGT가 있다는 가설을 수립하게 하였으며, 유전자 제공자별로 LGT 유전형질이 선택성을 띄게 되는 원인과 그 현상이 비브리오균의 진화에 미치는 영향에 대한 연구의 필요성을 제시하였다.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2013028666).

References

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.

Antonova, E.S. and Hammer, B.K. 2011. Quorum-sensing autoinducer molecules produced by members of a multispecies biofilm promote horizontal gene transfer to *Vibrio cholerae*. *FEMS Microbiol. Lett.* **322**, 68−76.

Baumann, P., Furniss, A.L., and Lee, J.V. 1984. Genus I. *Vibrio*. *In* Krieg, N.R. and Holt, J.G. (eds.), *Bergey's Manual of Systematic Bacteriology,* Vol. 1, pp. 518−538. Williams & Wilkins, Baltimore, M.D., USA.

Boucher, Y., Cordero, O.X., Takemura, A., Hunt, D.E., Schliep, K., Bapteste, E., Lopez, P., Tarr, C.L., and Polz, M.F. 2011. Local mobile gene pools rapidly cross species boundaries to create endemicity within global *Vibrio cholerae* populations. *MBio* **2**, e00335-10.

Chun, J., Grim, C.J., Hasan, N.A., Lee, J.H., Choi, S.Y., Haley, B.J., Taviani, E., Jeon, Y.S., Kim, D.W., Brettin, T.S., and *et al.* 2009. Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proc. Natl. Acad. Sci. USA* **106**, 15442−15447.

Clarke, G.D., Beiko, R.G., Ragan, M.A., and Charlebois, R.L. 2002. Inferring genome trees by using a filter to eliminate phylogenetically discordant sequences and a distance matrix based on mean normalized BLASTP scores. *J. Bacteriol.* **184**, 2072−2080.

Daubin, V., Moran, N.A., and Ochman, H. 2003. Phylogenetics and the cohesion of bacterial genomes. *Science* **301**, 829−832.

Dikow, R.B. and Smith, W.L. 2013. Genome-level homology and phylogeny of *Vibrionaceae* (Gammaproteobacteria: *Vibrionales*) with three new complete genome sequences. *BMC Microbiol.* **13**, 80.

Doolittle, W.F. 2012. Population genomics: how bacterial species form and why they don't exist. *Curr. Biol.* **22**, R451−453.

Doolittle, W.F. and Zhaxybayeva, O. 2009. On the origin of prokaryotic species. *Genome Res.* **19**, 744−756.

Feil, E.J., Holmes, E.C., Bessen, D.E., Chan, M.S., Day, N.P., Enright, M.C., Goldstein, R., Hood, D.W., Kalia, A., Moore, C.E., and *et al.* 2001. Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc. Natl. Acad. Sci. USA* **98**, 182−187.

Jammalamadaka, S.R. and SenGupta, A. 2001. Topics in Circular Statistics. World Scientific Publishing Co., Singapore.

Kahlke, T., Goesmann, A., Hjerde, E., Willassen, N.P., and Haugen, P. 2012. Unique core genomes of the bacterial family *Vibrionaceae*: insights into niche adaptation and speciation. *BMC Genomics* **13**, 179.

Lo Scrudato, M. and Blokesch, M. 2013. A transcriptional regulator linking quorum sensing and chitin induction to render *Vibrio cholerae* naturally transformable. *Nucleic Acids Res.* **41**, 3644−3658.

Meibom, K.L., Blokesch, M., Dolganov, N.A., Wu, C.Y., and Schoolnik, G.K. 2005. Chitin induces natural competence in *Vibrio cholerae*. *Science* **310**, 1824−1827.

Ochman, H., Lawrence, J.G., and Groisman, E.A. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299−304.

Papke, R.T. and Gogarten, J.P. 2012. Ecology. How bacterial lineages emerge. *Science* **336**, 45−46.

Rowe-Magnus, D.A., Guerout, A.M., Biskri, L., Bouige, P., and Mazel, D. 2003. Comparative analysis of superintegrons: engineering extensive genetic diversity in the *Vibrionaceae*. *Genome Res.* **13**, 428−442.

Schmidt, H.A., Strimmer, K., Vingron, M., and von Haeseler, A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502−504.

Seitz, P. and Blokesch, M. 2013. DNA-uptake machinery of naturally competent *Vibrio cholerae*. *Proc. Natl. Acad. Sci. USA* **110**, 17987−17992.

Shapiro, B.J., Friedman, J., Cordero, O.X., Preheim, S.P., Timberlake, S.C., Szabo, G., Polz, M.F., and Alm, E.J. 2012. Population genomics of early events in the ecological differentiation of bacteria. *Science* **336**, 48−51.

Strimmer, K. and von Haeseler, A. 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl. Acad. Sci. USA* **94**, 6815−6819.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33−36.