Vol. 47 No. SIG 17(TBIO 1) IPSJ Transactions on Bioinformatics Nov. 2006

Original Paper

Elucidating Conservation of Genes in Multiple Genomes Based on Graphs Configured by Bidirectional Best-Hit Relationships

Hijiri Maeno,† Md. Altaf-Ul-Amin,† Yoko Shinbo,†
Ken Kurokawa,† Naotake Ogasawara† and Shigehiko Kanaya†

Gene classification based on orthologous relations is an important problem to understand species-universal or species-specific conservation of genes in genomes associated with phenotype in species. In the present study, we proposed a classification system of genes based on configuration of networks concerning bidirectional best-hit relations (called orthologous relation group), which makes it possible to compare multiple genomes. We have applied this method to five *Bacillus* species (*B. subtilis*, *B. anthracis*, *B. cereus*, *B. halodurans*, and *B. thuringiensis*). With regards to the the five species, 4,776 orthologous relation groups have been obtained, and those are classified into 113 isomorphic groups. An isomorphic group may contain only orthologs or a combination of orthologs and paralogs. Gene functions and the conservativeness are discussed in view of configuration of orthologous relation groups.

1. Introduction

With the progress in determining genome sequences, the relationships between genes from different genomes can be represented naturally as a system of homologous families such as orthologs and paralogs. The conventional approach to the identification of orthologs between two genomes is the so-called bidirectional best-hit (BBH) criterion, that is, if ith gene of genome s_t (denoted by $g_i(s_t)$) is the best-hit of the jth gene of genome s_u ($g_j(s_u)$) and vice versa, genes $g_i(s_t)$ and $g_j(s_j)$ are regarded as a pair of orthologs. Given the existence of one-to-many and many-to-many orthologous relationships, the task of identifying orthologs is redefined as the delineation of clusters of orthologous group (COG). The identification of COGs was based on consistent pattern of triangle graphs of BBH relations between genes 1).

However, orthologous grouping is not a simple task because of co-orthologous relationship of genes between species. Co-orthologs are paralogs produced by duplication of orthologs subsequent to a given speciation event that separates the given lineage from the other lineage under consideration, which is commonly observed between species 2)∼4), and should be classified logically into outparalogs and inparalogs, that is, the former corresponds to those without orthologous relationships, derived from an ancestral duplication, and latter corresponds to those with orthologous relationships, derived from a lineage-specific duplication 5). Analysis of inparalogs is important for detecting lineage-specific adaptations. In the present study, we address graph configurations based on BBH relationships of genes in multiple genomes to detect paralogous genes.

Gene classification based on BBH relations in multiple genomes is the important problem to understand conservation of genes in individual genomes, and classification of orthologs based on clustering algorithm 6),7) has been proposed in the context of multiple genomes. In the closely related species, the network configuration constructed by orthologs in multiple genomes is also important to understand gene duplication process concerning inparalogs which are bona fide orthologs by definition proposed by Remm, et al. 4)

Here we present a classification of genes using configuration of networks concerning BBH relations, which reflects diversity of orthologs in more than two genomes based on simple networks in the closely related species. We also examine relationship between the configuration and gene function. The present approach has been applied to five *Bacillus* species (*B. subtilis*, *B. anthracis*, *B. cereus*, *B. halodurans*, *B. thuringiensis*).

2. Method

In this work we determined ortholog pairs of genes using BBH relationship. If ith gene of genome s_u which is denoted by $g_i(s_u)$ is the highest hit to jth gene of genome s_v, i.e., $g_j(s_v)$

---

† Nara Institute of Science and Technology, Graduate School of Information Science

and vice versa, genes $g_i(s_u)$ and $g_j(s_v)$ are regarded as a pair of orthologs [1]. The orthologous pair is represented as $[g_j(s_v),\ g_i(s_u)]$. In the case of multiple genomes (>2), we can represent all the ortholog pairs of genes as a graph G(N, E), where N and E are a set of nodes corresponding to genes and a set of edges corresponding to the orthologous relations, respectively. Now, the connected components of this graph can be considered as the clusters of genes. Such a cluster, i.e., a group of closely related genes represented by a non-separable subgraph is referred to as orthologous relation group (ORGroup). The ORGroup of **Fig. 1** (a) consists of $g_h(s_t)$, $g_i(s_u)$, and $g_j(s_v)$ connected by BBH relations between all pairs among them. Grouping of orthologous genes based on the ORGroup is utilized and referred to as clusters of orthologous group by Tatsusov, et al. [1] Triangle relation of orthologous genes in three genomes is fundamentally important property of ORGroups. There are three cases for configuration of orthologous relation of genes in three genomes (Fig. 1 (a), (b), and (c)); (i) when genes for three genomes have the bidirectional best-hit relation to each other, those are very highly conserved with respect to each other (Fig. 1 (a)); (ii) in the case that one gene of a genome, for example $g_h(s_t)$ in Fig. 1 (b), has the bidirectional best hit relations to genes of two other genomes, $g_j(s_v)$ and $g_i(s_u)$, and gene $g_j(s_v)$ does not have bidirectional best hit relation to gene $g_i(s_u)$, these two genes can be regarded as not so highly conserved with respect to each other in view of gene comparison of different genomes; and (iii) when two genes ($g_h(s_t)$ and $g_i(s_u)$) from two genomes have the bidirectional best hit to each other, and each of them has the bidirectional best hit to two different genes ($g_j(s_v)$ and $g_k(s_v)$) in a third genome, we can detect two duplicated genes via orthologous relation (Fig. 1 (c)). In the present study, multiplicated genes detected via orthologous relations are referred to as 'paralogous genes' or simply 'paralogs'.

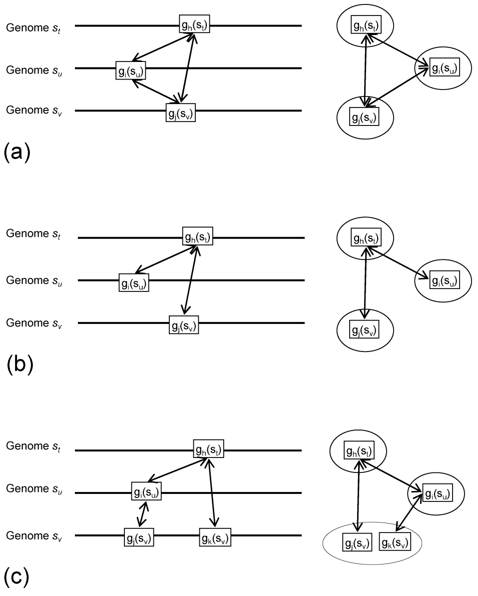

Fig. 1 BBH relations among three species.

The tightness of genes in an ORGroup can be measured by using the concept of diameter of a graph. The diameter of a graph is the longest path among all of the pairs of vertices in the graph. In the case of small diameter, genes in ORGroups are closely connected to each other, for example, in a complete graph, that is, in the case where all genes in an ORGroup have orthologous relations to each other, the diameter is always 1. Larger diameter of an ORGroup indicates increased indirect relations among its genes.

3. Results and Discussion

3.1 ORGroup without Paralogous Genes

In the present study we examined orthologous relations among genes of five *Bacillus* species (*B. subtilis*, *B. anthracis*, *B. cereus*, *B. halodurans*, *B. thuringiensis*). Among the members of Gram-positive bacteria, *B. subtilis* is the best-characterized aerobic, endospore-forming, rod-shaped bacterium commonly found in soil, water source and in association with plants [8]. *B. halodurans* is an alkaliphilic bacterium that grows optimally above pH 9.5 [9]. *B. cereus*, *B. thuringiensis* and *B. anthracis* belong to the *B. cereus* sensu lato group of rod-shaped, Gram-positive, spore-forming bacteria [10]. *B. anthracis* is the etiological agent of anthrax, an acute fatal animal and human disease that was employed as a bioterror agent in the autumn of 2001 [11]. In the present study, *B. subtilis*, *B. anthracis*, *B. cereus*, *B. halodurans*, and *B. thuringiensis* are abbreviated as BS, BA, BC, BH, and BT, respectively. To identify orthologs we selected BBH relations with a significant BLASTP E-value smaller than 10^{-6}.

With regard to the five *Bacillus* species, all ORGroups with more than one gene were classified into 113 isomorphic groups as shown in

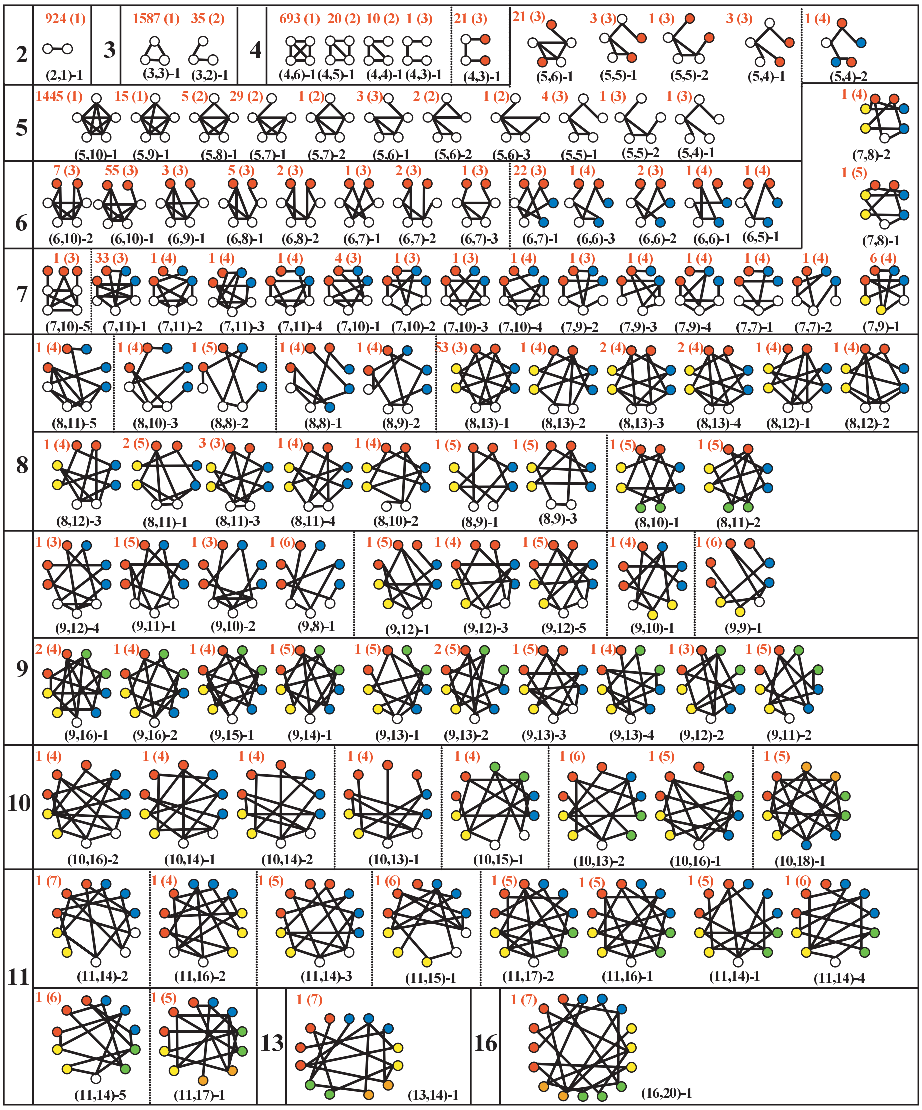

Fig. 2 All isomorphic groups in the five *Bacillus* genes. Genes and BBH relations are represented by nodes and edges. Paralogous genes are represented by identical colors other than white. Graph ID is written in black color under each graph. The two integers in parenthesis are the number of genes and the number of BBH relations, respectively. The number following the parenthesis represents ID-number for each type of isomorphic graphs. Total number of corresponding ORGroups and the diameter in parenthesis is written above each graph using red color.

Fig. 2. An ORGroup may contain only orthologs or a combination of orthologs and paralogs. Each of the identical ORGroups containing paralogous genes corresponding to a particular isomorphic group contains a set of paralogs of the same size and shows identical BBH relations between paralogous genes, between paralogous and other genes, and between the other genes. We observed 4,776 ORGroups that are without paralogous genes and can be classified into 18 isomorphic groups. The largest four isomorphic groups are (3,3)-1, (5,10)-1, (4,6)-1, and (2,1)-1 which corresponds to 1,587, 1,445, 693, and 924 ORGroups, respectively. The ID numbers of isomorphic groups, for example, (3,3)-1, (5,10)-1 and so on are same as those

Table 1 Combination of species and its occurrence in complete-type ORGroups.

Graph ID	**BA** (5311)	**BT** (5117)	**BC** (5234)	**BH** (4066)	**BS** (4105)	**The number of combinations**
(2,1)-1	BA			BH		8
		BT			BS	14
		BT		BH		15
	BA				BS	16
			BC		BS	24
			BC	BH		26
		BT	BC			146
	BA		BC			151
	BA	BT				222
				BH	BS	302
(3,3)-1		BT		BH	BS	4
		BT	BC	BH		5
			BC	BH	BS	5
	BA		BC	BH		7
		BT	BC	BH		8
	BA			BH	BS	9
	BA		BC		BS	12
	BA	BT			BS	17
		BT	BC		BS	23
	BA			BH	BS	33
	BA	BT	BC			1,464
(4,6)-1	BA		BC	BH	BS	5
		BT	BC	BH	BS	27
	BA	BT		BH	BS	34
	BA	BT	BC	BH		209
	BA	BT	BC		BS	418
(5,10)-1	BA	BT	BC	BH	BS	1,445

Graph ID corresponds to that in Fig. 2. The parentheses under the species abbreviation represents the number of genes for each species.

in Fig. 2. The numbers of ORGroups belonging to the other isomorphic groups are smaller than 30. So in the five *Bacillus* species, 78% of the ORgroups that are without paralogous genes are complete graphs based on BBH relations, that is, graphs with all possible BBH relations. The isomorphic group (5,10)-1 is the most important complete graph because in this case all pairs of genes in the five genomes have BBH relations to each other, and therefore such ORGroups consist of highly conserved genes. In *B. subtilis*, 272 genes have been determined as essential genes by gene disruption experiments [12], that is, a gene is assigned the tag 'essential', if its disruption in the genome appears to be lethal to the cell. All of the essential genes belong to the isomorphic groups of the type (5,10)-1. The ORGroups that are the complete graphs of maximum possible size, i.e., comprising of the number of genes equal to the number of species analyzed tend to contain very highly conserved genes concerning house keeping. Thus, it can be suggested that genes included in the ORGroups comprizing of complete graphs are highly conserved and uniquely exist in multiple genomes.

In this study a complete ORGroup of size five contains five genes, one from each of the five species. But the complete ORGroups of size three or four may not be identical to each other in consideration of the related species. The number of occurrences of different combinations of species in case of isomorphic groups (2,1)-1, (3,3)-1, (4,6)-1 and (5,10)-1 are summarized in **Table 1**. The highest occurrence among those combinations is 1,464 for the combination of (BA, BT, BC) in isomorphic group (3,3)-1. This implies that very large number of orthologous genes are highly conserved among these three species BA, BT, and BC. This conservation property of genomes is associated with the fact that BA, BC, and BT are classified into the same group and are very similar bacteria [10]. In case of isomorphic group (4,6)-1, two combinations involving four species occurred relatively higher number of times, one of them is BA, BT, BC, BS (418 ORGroups) and the other is BA, BT, BC, BH (209 ORGroups).

3.2 ORGroups Containing Paralogous Genes

In case of ORGroups containing paralogous genes, the occurrences of six isomorphic groups, (6,10)-1, (8,13)-1, (7,11)-1, (4,3)-1, (5,6)-1 and (6,7)-1 are much higher than the others. Their occurrences are 55, 53, 33, 21, 21 and, 20, respectively. In the present study paralogous genes are defined as a pair of genes for same species in an ORGroup that are connected by a path of indirect BBH relations. Obviously at least 3 BBH relations are needed to obtain a pair of paralogous genes. The diameter of all the above mentioned six isomorphic groups is 3. Therefore the paralogous gene pairs obtained from ORGroups belonging to these six isomorphic groups are expected to be highly conserved. In the present study 95 isomorphic groups that contain paralogous genes were obtained (Fig. 2), and only six isomorphic groups are dominant. 172 out of 328 ORGroups, i.e., 52% of the ORGroups that contain paralogous genes corresponds to these six dominant groups.

The major six isomorphic groups that contain paralogous genes can be derived from three fundamental isomorphic groups that do not contain paralogous genes denoted by (3,3)-1, (4,6)-1, and (5,10)-1 as shown in **Fig. 3**. The isomorphic groups (4,3)-1, (5,6)-1, and (6,10)-1 (at the middle part of Fig. 3) can be derived by adding just a single gene to isomorphic groups (3,3)-1, (4,6)-1, and (5,10)-1 (at the left side), respectively. On the basis of strong structural similarity between the corresponding pairs of these isomorphic groups, we can detect strongly conserved pairs of paralogous genes. We guess that the number of duplicated genes increases with the increment of the number of genes in complete ORGroups. Therefore, we estimate that the ratio of the occurrences of a derived group to that of the corresponding complete group should have a relation with the total number of genes in the complete ORGroups. Actually, we found the following linear relation represented in Eq. (1).

$$R = 0.012NG - 0.023 \qquad (1)$$

Here, R represents the ratio of the occurrences of a derived group (shown in middle part of Fig. 3) to that of the corresponding complete group (shown in left part of Fig. 3) and NG is the number of genes in a complete graph. High linear correlation coefficient (0.997, n=3) was obtained in the context of Eq. (1). With regard to the five *Bacillus* genomes, the ratio R is 0.013 for (4,3)-1 to the complete graph (3,3)-1, 0.030 for (5,6)-1 to (4,6)-1, and 0.038 for (6,10)-1 to (5,10)-1. Other isomorphic groups containing paralogous genes whose occurrences are relatively higher are (6,7)-1, (8,13)-1, and (7,11)-1. One characteristic of these groups is that they contain complete subgraphs.

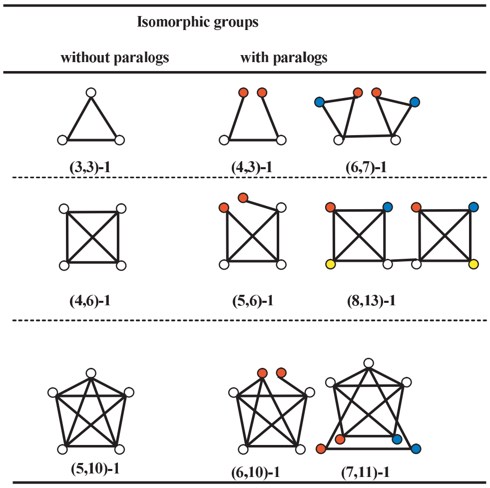

Fig. 3 Relation between major isomorphic groups without and with paralogous genes. Details are described in the text.

3.3 Comparative Genomics Based on ORGroups

Examination of the relation of conservation of genes obtained in the present study to cell function is important to validate and explain conservativeness of BBH relations and to understand the essentiality of genes in genomes. Genes of the ORGroups corresponding to the isomorphic group (5,10)-1, that is, where all pairs of genes in the five genomes have BBH relations to each other, are expected to be very highly conserved as mentioned in Section 3.1. So we examined the function categories of genes related to the isomorphic group (5,10)-1. The COG (clusters of orthologous groups of proteins) functional categories for several species have been developed by Tatsusov, et al.[1] This classification is useful for comparison of genes in case of distantly related organisms. However, the function class defined in COG functional categories is not always suitable for the purpose of comparison of genomes of closely related organisms, because of lack of species-specific function classes. For example, it is fundamentally important to examine diversity or conservation of genes concerning endospore-formation process in *Bacillus* species. Unfortunately, this type of category has not been considered in COG functional categories. So, we used the categories based on functional classification of the *Bacillus subtilis* protein-coding genes[13], which com-

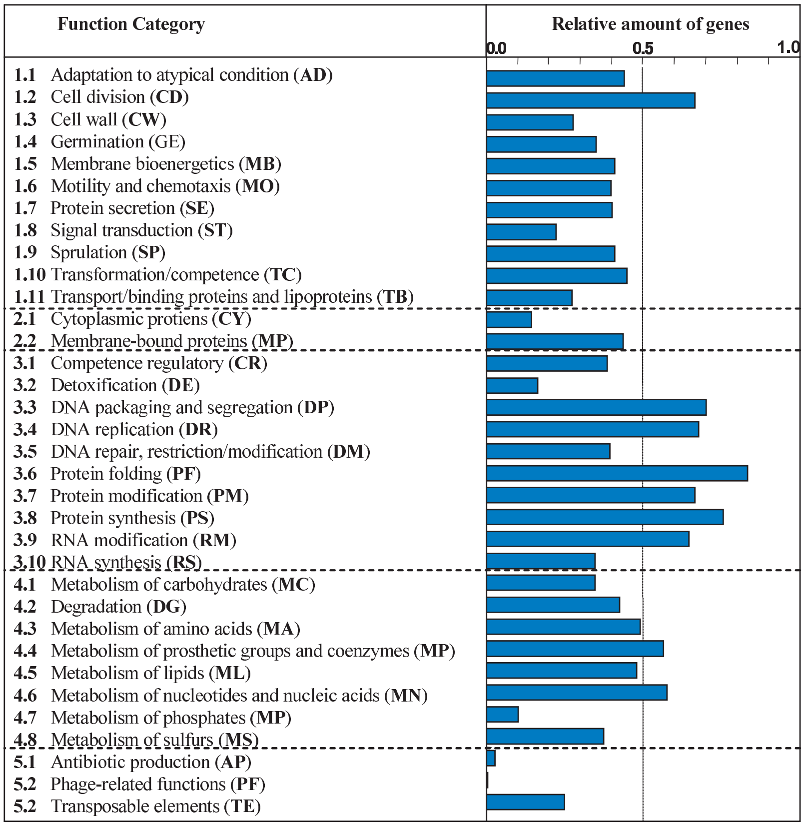

Fig. 4 Relative amount of genes in ORGroups corresponding to isomorphic group (5,10)-1with respect to all genes of *B. subtilis* calculated for individual function categories.

prises of 5 categories (1, cell envelope and cellular processes; 2, cytochrome; 3, genetic information pathway; 4, intermediary metabolism; and 5, other functions).

Figure 4 shows the relative amount of genes in ORGroups corresponding to isomorphic group (5,10)-1 compared to the total number of genes in *B. subtilis* calculated for individual function categories. Function categories for which the aforementioned relative amount is larger than 0.5 are as follows, (i) 1.2 Cell division, (ii) 3.3 DNA packaging and segregation, (iii) 3.4 DNA replication, (iv) 3.6 Protein foliding, (v) 3.7 Protein modification, (vi) 3.8 Protein synthesis, (vii) 3.9 RNA modification, (viii) 4.4 Metabolism of prosthetic groups and coenzymes, and (ix) 4.6 Metabolism of nucleotides and nucleic acids. Over half of genes in most of the categories concerning genetic information pathway tends to be classified into ORGroups with isomorphic configuration (5,10)-1. Only, 3.1 Competence regulatory, 3.2 Detoxification, 3.5 DNA repair restriction/modification, and 3.10 RNA synthesis are exceptions. This indicates that genes included in genetic information pathway tend to be highly conserved and remain as single copies across genomes, which is consistent with the fact that genes concerning genetic information pathway tend to be single orthologs in human, fly, worm and yeast that have not undergone duplication and elaboration in the various organisims [14]. Among intermediary metabolism categories, most of the genes classified into metabolism of prosthetic groups and coenzymes (category ID = 4.4) and nucleotides and nucleic acids (4.6) are also classified into the isomorphic group (5,10)-1. This indicates that the pathways for those metabolites are highly conserved in comparison to the other metabolic pathways. Thus, the ORGroups comprising of complete graphs that include genes from all the species are very highly conserved and associated with fundamental genes concerning house keeping.

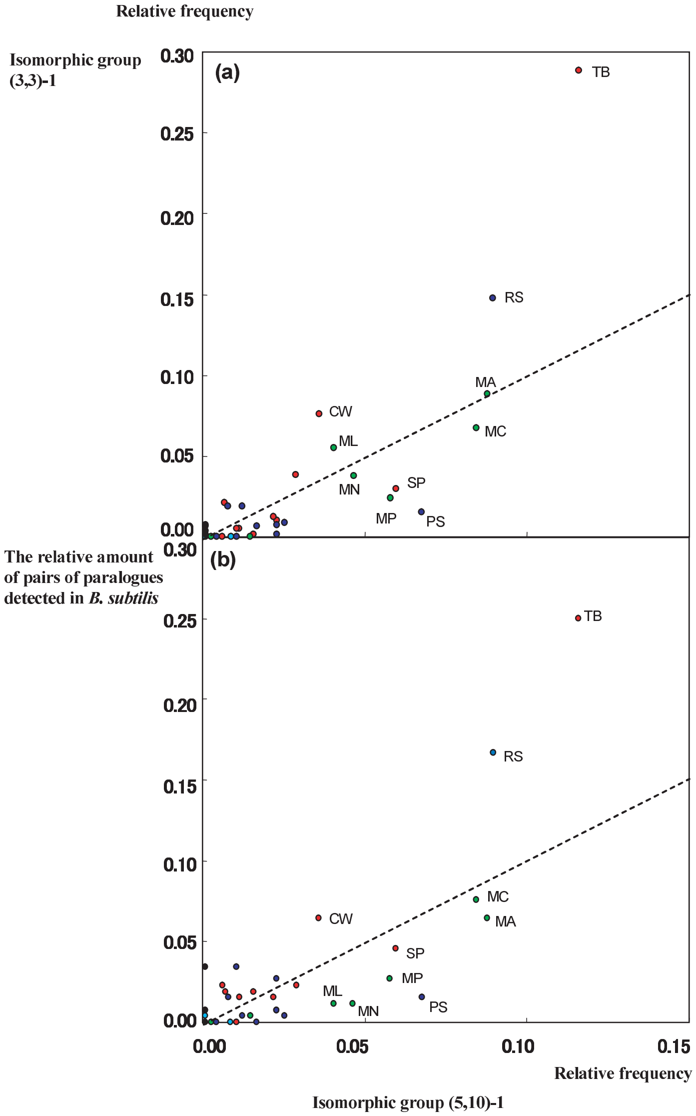

Fig. 5 Relation of relative frequencies of genes in function categories (a) between (5,10)-1 and (3,3)-1, and (b) between (5,10)-1 and multiple paralogues in *B. subtilis*. Equivalent relation between two frequencies (y = x) is represented by a broken line. Abbreviation of a function category corresponds to that in Fig. 4.

The largest number of ORGroups corresponds to the isomorphic group (3,3)-1. Among them, 92% of ORGroups consist of genes from three *Bacillus* species, BA, BT, and BC, which are classified as common group [10] with pathogenicity. Species-specific conserved functions for these three pathogenic species can be examined by comparing gene distributions in function categories in the context of (5,10)-1 and (3,3)-1. The relative amount of genes belonging to each functional category is calculated with respect to all genes in ORGroups of type (3,3)-1 and (5,10)-1 separately. **Figure 5** (a) shows the relative amounts of (3,3)-1 against those of (5,10)-1. The number of genes in the ORGroups of type (3,3)-1 which belong to transport/binding proteins and lipoprotein (TB), RNA synthesis (RS), Cell wall (CW), and Metabolism of lipid (ML) is much larger compared to those in the ORGroups of type (5,10)-1. This indicates that the genes under function categories TB, RS, CW, and ML are highly conserved in these three pathogenic *Bacillus* species which is not the case for all five *Bacil-*

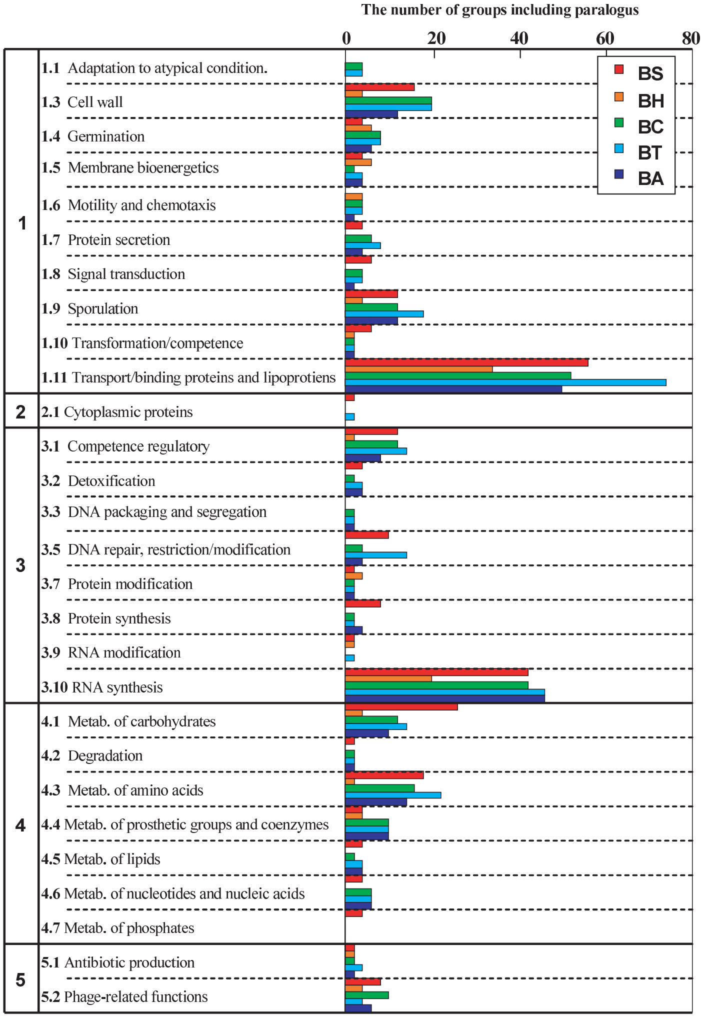

Fig. 6 Histogram showing distribution of paralogous genes with respect to function categories. Paralogous were not detected in the function categories, 1.2, 2.2, 3.4, 3.6, 4.8, and 5.2 in all five species.

lus species analyzed in the present study. On the other hand the relative amounts of genes belonging to some function categories are smaller. Similar situation can be observed in the relation between (5,10)-1 and paralogous genes detected in *B. subtilus* except for the case of ML. The relative amount of ML genes in ORGroups corresponding to (3,3)-1 is larger than that in ORGroups corresponding to (5,10)-1, but an opposite situation has been found in case of the relative number of pairs of paralogs in *B. subtilis*. The fatty acid composition of the cytoplasmic membrane lipids of bacterial cells plays key roles to adaptation phenomenon and diverged [15)~21)]. This indicates that the process of the fatty acid synthesis is diverged in the five *Bacillus* species, but, the fatty acid synthesis in three pathogens (BA, BC, BT) are conserved and their adaptation strategy may be very similar. Consequently it can be suggested that, species-specificity of five *Bacillus* species is derived from gene functions TB, RS and CW. This is reasonable that optimization of adaptation of atypical conditions for survival can be carried by Transport/binding protein and lipoproteins, and the transcription factors, which are main components in RS, and are highly diverged among species. This is also validated by **Fig. 6**, which shows histogram of the distribution of paralogous genes with respect to function categories. The total pairs of paralogous for BS, BH, BC, BT, and BA detected by the BBH relations are 172, 58, 171, 221 and 153, respectively. Again, relatively large numbers of paralogous genes for transport/binding proteins and lipoprotein (TB), RNA synthesis (RS), and Cell wall (CW) were obtained for all *Bacillus* species.

4. Conclusions

Gene classification based on orthologous relations is one of the important problems to understand species-universal or species-specific conservation of genes in genomes associated with phenotype in species. The proposed classification system of genes based on networks configured by bidirectional best-hit relations makes it possible to compare gene composition in multiple genomes and the highly conserved paralogs can be detected by constructing orthologous relation groups. Depending on the number of occurrences of complete ORGroups involving different combination of species, the closely related group of species can be determined. Furthermore classification of ORGroups to isomorphic groups configured by BBH makes it possible to estimate conservation and diversity of genes in multiple genomes. Also, relation between conservation and function categories of genes in the five *Bacillus* genomes has been realized and explained.

References

1) Tatsusov, R.L., Koonin, E.V. and Lipman, D.J.: A genomic perspective on protein families, *Sciene*, Vol.273, pp.631–637 (1997).
2) Jordan, I.K., Rogozin, I.B., Wolf, Y.I. and Koonin, E.V.: Essentail genes are more evolutionarily conserved than are non essentaial genes in bacteria, *Genome Res.*, Vol.12, pp.962–968 (2001).
3) Lespinet, O., Wolf, Y., Koonin, E. and Aravind, L.: The role of lineage-specific gene family expansion in evolution of eukaryotes, *Genome Res.*, Vol.12, pp.1048–1059 (2002).
4) Remm, M., Storm, C. and Sonhammer, E.: Automatic clustering of orthologs and inparalogs from pairwise species comparisons, *J. Mol. Biol.*, Vol.314, pp.1041–1052 (2001).
5) Sonnhammer, E. and Koonin, E.: Orthology, paralogy and proposed classification for paralog subtypes, *Trends Genet.*, Vol.18, pp.619–620 (2002).
6) Uchiyama, I.: Hierarchical clustering algorithm for comprehensive orthologous-domain classification in multiple genomes, *Nucleic Acids Res.*, Vol.34, pp.647–658 (2006).
7) Abascal, F. and Valencia, A.: Clustering of proximal sequence space for the identification of protein families, *Bioinformatics*, Vol.18, pp.908–921 (2002).
8) Kunst, F., et al.: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*, *Nature*, Vol.390, pp.249–256 (1997).
9) Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., Fuji, F., Hirama, C., Nakamura, Y., Ogasawara, N., Kuhara, S. and Horikoshi, K.: Complete genome sequence of the alkaliphilic bacterium *Bacillus* halodurans and genomic sequence comparison with *Bacillus subtilis*, *Nucleic Acids Res.*, Vol.28, pp.4317–4331 (2000).
10) Jensen, G.B., Hansen, B.M., Eilenber, J. and Mahillon, J.: The hidden lifestyles of *Bacillus cereus* and relatives, *Environ. Microbiol.*, Vol.5, pp.631–640 (2003).
11) Rasko, D.A., Ravel, J., Okstad, O.A., Helgason, E., Cer, R.Z., Jiang, L., Shores, K.A., Fouts, D.E., Tourasse, N.J., Angiuoli, S.V., Kolonay, J., Nelson, W.C., Kolsto, A., Fraser, C.M. and Read, T.D.: The genome sequence of *Bacillus cereus* ATCC 10987 reveals metabolic adaptations and a large plasmid related to *Bacillus* anthracis pXO1, *Nucleic Acid Res.*, Vol.32, pp.977–988 (2004).
12) Kobayashi, K., Ehrlich, S.D., Albertini, A., et al.: Essential *Bacillus subtilis* genes, *Proc. Natl. Acad. Sci. USA*, Vol.100, pp.4678–4683 (2003).
13) Sonenshein, A.L., Hoch, J.A. and Losick, R.: *Bacillus subtilis and its closest relatives from genes to cells*, ASM Press (2002).
14) Consortium, I.H.G.S.: Initial sequencing and analysis of the human genome, *Nature*, Vol.409, pp.860–921 (2001).
15) Ingram, L.O.: Adaptation of membrane lipids to alcohols, *J. Bacteriol.*, Vol.125, pp.670–678 (1976).
16) Bygraves, J.A. and Russell, N.J.: Solute tolerance and membrane lipid composition in some halotolerant food spoilage bacteria, *Food Microbiol.*, Vol.5, pp.109–116 (1988).
17) Juneja, V.K. and Davidson, P.M.: Influence of altered fatty acid composition on resistance of Listeria-monocytogenes to antimicrobials., *J. Food Prot.*, Vol.56, pp.302–305 (1993).
18) Heipieper, H.J. and Bont, J.A.M.D.: Adaptation of Pseudomonaus putida S12 to ethanol and toluene at the level of fatty acid compostion of membranes, *App. Environ. Microbiol.*, Vol.60, pp.4440–4444 (1994).
19) Russell, N.J., Evans, R.I., ter Steeg, P.F., Hellemons, L., Verheul, A. and Abee, T.: Membranes as a target from stress adaptation, *Int.*

J. Food Microbiol., Vol.28, pp.225–261 (1995).
20) Sikkema, J., Bont, J.A.M.D. and Poolman, B.: Mechanisms of membrane toxicity of hydrocarbons, *Microbiol.*, Vol.59, pp.201–222 (1995).
21) Bont, J.A.M.D.: Solvent-tolerant bacteria in biocatalysis, *Trends Biotechnol.*, Vol.16, pp.493–499 (1988).

(Received August 28, 2006)
(Accepted September 4, 2006)

(Communicated by *Chiemi Watanabe*)

Hijiri Maeno received B.S. from Kansai University in 2001, M.S. from Bioscience Department of Cell Biology, Graduate School of Biological Science, Nara Institute of Science and Technology, and Doctoral Course in Engineering (leaving after completing a course), Department of Bioinformatics and Genomics, Graduate School of Information Science, Nara Institute of Science and Technology. In 2006, Technical Staff at Nagahama Institute of Bio-Science and Technology.

Md. Altaf-Ul-Amin received B.Sc. in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology (BUET) in 1993, Master of Science in Electrical, Electronic and Systems Engineering from University Kebangsaan Malaysia (UKM) in 1999 and Ph.D. from Nara Institute of Science and Technology (NAIST) in 2003. Presently he is working as an Assistant Professor in Comparative Genomics Lab. of NAIST. His research interest includes application of Network theory and algorithms and self organizing mapping to bioinformatics.

Yoko Shinbo received Bachelor of Agriculture in Agricultural Science from Kagoshima University in 2002, M.Sc. in Department of Bioinformatics, Graduate School of Information Science, Nara Institute of Science and Technology (NAIST) in 2004. From 2004, she has been working as research fellow in Comparative Genomics Lab. of NAIST. She is doing research on database construction of secondary metabolite and biosyntheses pathway.

Ken Kurokawa received Ph.D. in Environmental Science & Microbiology from Graduate School of Pharmacy, Osaka University in 1995. M.Sc. in Earth Physics from Institute of Geology and Paleontology, Graduate School of Science, Tohoku University in 1993, B.Sc. in Structural Geology and Volcanology from Institute of Geology and Paleontology, Faculty of Science, Tohoku University in 1990. He was Associate Professor in Bioinformatics at Graduate School of Information Science, NAIST from 2004, Research Associate in Bioinformatics, Research Institute of Microbial Diseases, Osaka University 2001–2004. Post-Doctoral Fellow in Bioinformatics Genome Information Research Center, Osaka University 1998–2001. His interests are origins and evolution of genome organization. Particularly interested in the correlation between genome information and biosystem evolution. He seeks out the laws of self-organization in genome, and dynamic modeling of biosystems with network thermodynamics.

Naotake Ogasawara received Ph.D. from Nagoya University. In 1993, Professor in Graduate School of Biological Sciences, Nara Institute of Science and Technology. In 2001, Dean of Graduate School of Biological Sciences, Nara Institute of Science and Technology. In 2002, Professor in Graduate School of Information Science, Nara Institute of Science and Technology. In 2005–present, Dean of the Graduate School of Biological Sciences in Nara Institute of Science and Technology.

Shigehiko Kanaya received B.S. from Department Bioscience, Faculty of Science, Science Univercity of Tokyo in 1985, Ph.D. from Department Material and System Engineering, Toyohashi University of Technology in 1990. In 1990, Assistant Researcher in Information Engineering at Yamagata University, with a research program on genome informatics concerning species-specific codon usage on the basis of multivariate analysis. In 1996, Guest Associate Professor at National Institute of Genetics. In 1999 Associate Professor, Electronic and Information Engineering, and in 2000 Associate Professor, Applied Biosystem Engineering at Yamagata University. In 2004–present, Professor at Comparative Genomics, Department of Bioinformatics, Graduate School of Information Science, Nara Institute of Science and Technology.