African Journal of Microbiology Research Vol. 6(23), pp. 5062-5073, 21 June, 2012
Available online at http://www.academicjournals.org/AJMR
DOI: 10.5897/AJMR11.1364
ISSN 1996-0808 ©2012 Academic Journals

Full Length Research Paper

Networking clusters and sequence characteristics of clustered regularly interspaced short palindromic repeats (CRISPR) direct repeats and their evolutionary comparison with *cas1* genes in lactic acid bacteria

Kaibo Deng, Fei Liu, Chuntao Gu and Guicheng Huo*

Key Laboratory of Dairy Science, Ministry of Education, Northeast Agricultural University, Harbin, China.

Accepted 16 March, 2012

Clustered regularly interspaced short palindromic repeats (CRISPR) widely spread in Archaea and bacteria are an acquired immunity system, which resisted on the infection of exogenous chromosome according to a mechanism similar to RNA interference. In this study, evolutionary origin and sequence feature of totally 211 CRISPRs in 192 of 588 lactic acid bacteria genomes covered 18 genera of 5 phyla were employed, and comparative analysis of direct repeats (DRs), *cas1* (CRISPR-associated) genes and 16s rRNA were performed as well. In summary, 11 clusters of CRISPRs were identified based on DRs, and sequence similarity among genera even species were determined. In GC content investigation, complementary sequences and the symmetry in DRs of all clusters can opportunely construct the stem-loop secondary structure, moreover, the GC% level of spacers in one CRISPR locus was comparable, which suggested that foreign sequences with similar GC% were more likely to be inserted into the LAB genomes as a new spacer. *cas1* coevolved with DRs as a whole phylogenetic cassette, while it was slightly more conservative. Besides, the distribution of *cas1* and DRs was found very different with 16s rRNA in clusters, and it implied the possibility of horizontal gene transfer of LAB CRISPR loci.

Key words: Clustered regularly interspaced short palindromic repeats (CRISPR), lactic acid bacteria, network clustering, evolutionary comparison.

INTRODUCTION

Lactic acid bacteria (LAB) are regarded as probiotics for thousands of years with numerous functional advantages (Ouwehand et al., 1999; Guarner and Malagelada, 2003). In all LAB spread in 43 genera among 5 phyla (Garrity et al., 2004), many species are cultivated as starter culture and widely applied in milk fermentation industry (Axelsson, 2004). However, they are often attacked by phages which result into great loss (Jarvis, 1989; Moineau et al., 2002), although many strategies and measures have been taken (Josephsen and Neve, 2004).

*Corresponding author. E-mail: gchuo58@126.com. Tel: +86045155191807.

CRISPR is an adaptive immunity system against invasive DNA such as phages and plasmids (Jansen et al., 2002; Barrangou et al., 2007), and found in many prokaryotic genomes (Ishino et al., 1987; Haft et al., 2005; Lillestøl et al., 2006; Grissa et al., 2007), about 90% of Archaea and 40% of Bacteria including LAB by now (Barrangou and Horvath, 2009). CRISPR locus typically consists of non-contiguous, partially palindromic DNA DRs of 21-48 bp and interspaced by stretches of nonrepetitive spacers with similar length (Jansen et al., 2002), and it is usually adjacent to *cas* genes (Grissa et al., 2007; Barrangou and Horvath, 2009). To resist invasive chromosomes, DNA fragments from phages or plasmids are inserted into CRISPR locus as a new spacer. After transcription and modification, crRNA of CASCADE complex matches with

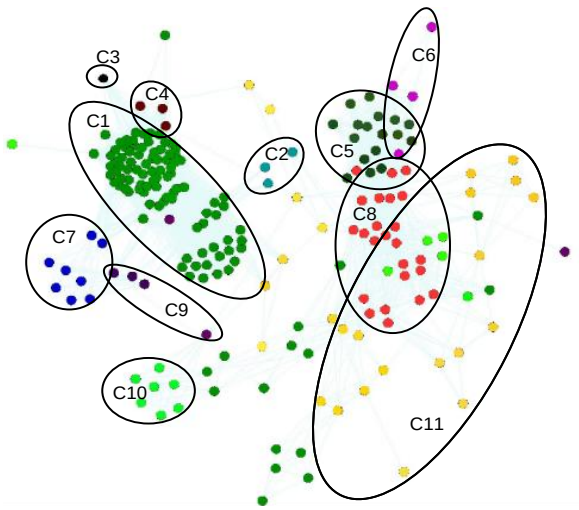

Figure 1. Network clusters of DRs in LAB by Biolayout Express[3D]. Each node denotes an individual DR, and the edges between them denote the evolutionary associations. Different colours denote clusters the nodes belonging to.

invasive mRNA through base-pairing potential and degraded them by endonuclease activity of Cas proteins (Barrangou and Horvath, 2009; Makarova et al., 2006). Cas proteins play a key role in the entire process (Jansen et al., 2002; Haft et al., 2005). In which, Cas1 is viewed as the best marker because of its general occurrence (Makarova et al., 2006).

Recent years some studies involving CRISPR evolution have been reported (Kunin et al., 2007; Horvath et al., 2008; Chakraborty et al., 2010; Rezzonico et al., 2011). With regard to LAB, there was only one, but the species was limited in only 49 genomes in Firmicutes and Actinobacteria phylum (Horvath and Barrangou, 2009). In this composition, 192 genomes spread all over 5 phyla were investigated. Larger range must provide more persuasive evidences for the evolutionary regularity.

MATERIALS AND METHODS

Genome sequences, DRs and *cas1* retrieval

All the completed and draft genome sequences of LAB were retrieved from NCBI ftp website. CRISPRs of completed genomes are published in CRISPR database (http://crispr.u-psud.fr/CRISPRHomepage.php), and CRISPRs of draft genomes were obtained by CRISPRs Finder (http://crispr.u-psud.fr/Server/). *cas1* and 16s rRNA sequences of completed genome species were acquired from CRISPR database and NCBI GenBank.

Network clusters analysis of DRs, *cas1* and 16s rRNA

Construction of networking clusters for DRs, *cas1* and 16s rRNA was performed through Biolayout Express[3D] (Theocharidis et al., 2009). The pair-wise alignment scores ranging below 1.0 were obtained from MEGA5 and applied to establish an expression file for visualizing the networking clusters.

Determination of multiple alignments and the secondary structures in DR clusters

Multiple alignments for DRs of every cluster were preformed to create for the consensus using sequence weblogo (http://weblogo.berkeley.edu/logo.cgi), and the folding scores were divided into 2 bits. The secondary structure prediction of DRs was preformed by Vienna RNA package (Hofacker, 2003).

Analysis of GC content for DR clusters

To verify the disparity of GC content for DR clusters, one random CRISPR locus was picked out from each repeats cluster, and the GC contents of first several repeat-spacer sequences were detected by DNAMAN, with the default windows was set 12.

Construction of phylogenetic tree of DRs and *cas1*

Phylogenetic tree for DRs and *cas1* were constructed using ClustalX and MEGA5 based on neighbor-joining method and the bootstrap test for 1000 replications.

RESULTS

Evolutionary inference based on clusters of DRs

LAB are widespread in 2 genera in *Thermotogae*, 30 in *Firmicutes*, 7 in *Actinobacteria*, 2 in *Bactero* and 2 in *Fusoaceteria* (Garrity et al., 2004). In all 588 genomes, 159 CRISPR loci in 70 completed genomes and 309 in 406 draft genomes were detected, which accounted for 81%. As the single nucleotide polymorphisms usually generated on the terminal repeat at the 3' end, the most frequent repeat in a CRISPR loci were selected (Rezzonico et al., 2011; Horvath and Barrangou, 2009). Eventually, 211 CRISPR DRs were divided into 11 clusters by Biolayout Express[3D], and the sequences with evolutionary distance below 2.5 were divided into the same cluster (Figure 1). The parameters (min correlation and squared correlation were set 0.3 and 0.005, respectively) were adjusted as well. The detailed information of strains was listed in Table 1.

In the 11 clusters, cluster 1 contained 105 sequences for the most, and cluster 3 had only one to the contrary. Majority edges of cluster 1 were short, which indicated closer evolutionary association; inversely the discrete nodes in cluster 11 indicated further evolutionary association. Generally, there was one kind of DR sequence effective in individual strain, but 37 strains had two different DRs. Here, DRs in 16 strains were divided into different clusters, and others were in the same cluster. It was worth noting that 7 *Streptococcus pyogenes* had the same two kinds of DR and belonged to Cluster 1 and 5 simultaneously. Furthermore there

Table 1. Information of the LAB strains and CRISPRs assignments to clusters.

Genus	Specie	Strain	CRISPR id*	Cas1 gene	Cluster	Genus	Specie	Strain	CRISPR id*	*Cas1 gene*	Cluster
Bacillus	*Cereus*	F65185			11	*Lactobacillus*	*Jensenii*	SJ-7A-US			1
	Cerius subsp.cytotoxis	NVH 391-98	NC_009674_3	Yes	10		*Paracasei subsp.paracasei*	8700:2			7
	Clausii	KSM-K16	NC_006582_6		11		*Rhamnosus*	GG (ATCC 53103)	NC_013198_2	*Yes*	7
	Coagulans	36D1			9			HN001			7
	Coahuilensis	m4-4			11		*Salivarius*	ACS-116-V-Col5a			7
	Halodurans	C-125	NC_002570_4		11		*Salivarius subsp.salivarius*	(=UCC118)	NC_007929_1	*Yes*	7
	Mycoides	Rock3-17			4	*Leptotrichia*	*Buccalis*	DSM1135 (C-1013-b)	NC_013192_2 NC_013192_6	*Yes*	2 1
	Pseudomycoides	DSM12442			11		*Hofstadii*	F0254			9
Bacteroides	*Capillosus*	ATCC29799			11	*Listeria*	*Innocua*	Clip11262	NC_003212_2		1
	Cellulosilyticus	DSM14838			1		*Monocytogenes*	08-5578	NC_013766_2		1
	Coprophilus	DSM18228 (CB42 =JCM 13818)	1 2		11 1			08-5923	NC_013768_2		1
	Dorei	5_1_36/D4			11			10403S	1 2		1 1
		DSM17855			10			EGD-e	NC_003210_2		1
	Fragilis	NCTC9343	NC_003228_3 NC_003228_4		9 10			F6900	1 2		1 1
		YCH46	NC_006347_2		9			Finland1988			1
	Sp.	2_1_16			10			FSL F2-515			1
		20_3			10			FSL J1-194			1
		3_1_33FAA			10			FSL J2-003			1
		4_3_47FAA			11			FSL J2-071			1
		9_1_42FAA			10			FSL N1-017	1 2		1 1
		D2			10			FSL R2-503	1 2		1 1

Table 1. Contd.

Bifidobacterium	*Adolescentis*	ATCC15703	NC_008618_4	Yes	11			FSL R2-561			1
		L2-32			11			HCC23	NC_011660_1	*Yes*	1
	Angulatum	DSM20098			8			J0161	1 2		1 1
	Animalis subsp. lactis	AD011	NC_011835_4	Yes	8			J2818	1 2		1 1
		BI-04	NC_012814_4	Yes	8			str.1/2a F6854	1 2		1 1
		DSM10140	NC_012815_4	Yes	8		*Seeligeri*	serovar 1/2b str. SLCC3954	NC_013891_3 NC_013891_5	*Yes*	1 1
	Bifidum	S17	NC_014616_1	Yes	7	*Mobiluncus*	*Curtisii*	ATCC43063	NC_014246_1	*Yes*	8
	Catenulatum	DSM16992			8 8		*Curtisii subsp. curtisii*	ATCC35241			8
	Dentium	ATCC27678			11		*Mulieris*	ATCC35243			8
		Bd1	NC_013714_9 NC_013714_1	Yes Yes	11 11	*Parascardovia*	*Denticolens*	F0305			11
	Gallicum	DSM20093	1 2		8 8	*Rothia*	*Dentocariosa*	ATCC17931	NC_014643_6	*Yes*	6
	Longum	DJO10A	NC_010816_5		11		*Mucilaginosa*	ATCC25296			6
	Longum subsp. longum	BBMN68	NC_014656_1		8			DY-18	NC_013715_6 NC_013715_7		6 6
	Pseudocatenulatum	DSM20438			11	*Ruminococcus*	*Albus*	7			5
Enterococcus	*Faecalis*	AR01/DG			1			8			5
		ATCC4200			1		*Flavefaciens*	FD-1			5
		ATCC29200			1		*Gnavus*	ATCC29149			5
		D6			1		*Lactaris*	ATCC29176			5
		DS5			1	*Scardovia*	*Inopinata*	F0304			10
		JH1			1	*Sebaldella*	*Termitidis*	ATCC33386	NC_013517_2 NC_013517_3		4 4

Table 1. Contd.

		Merz96			1	*Staphylococcus*	*Epidermidis*	RP62A	NC_002976_2	*Yes*	1
		OG1RF			1		*Lugdunensis*	HKU09-01	NC_013893_1	*Yes*	1
		S613			1	*Streptococcus*	*Agalactiae*	18RS21			1
		T1			1			2603V/R	NC_004116_1		1
		T3			1			515			1
		T8			1			A909	NC_007432_1	*Yes*	1
		TUSoD Ef11			1			CJB111			1
		TX0470			1			COH1			1
		TX1322			1			H36B			1
		TX2134			1			NEM316	NC_004368_1		1
		TX4248			1		*Bovis*	ATCC 700338			1
		X98			1			TX20005			1
	Faecium	1141733			1		*Equi subsp. zooepidemicus*	(H70)	NC_012470_3	*Yes*	5
		1231408			1			MGCS10565	NC_011134_1	*Yes*	5
									NC_011134_4	*Yes*	1
		Com12			1		*Gallolyticus*	UCN34	NC_013798_1	*Yes*	1
									NC_013798_2	*Yes*	1
		PC4.1			1		*Gordonii Str.Challis substr.*	CH1 (ATCC 35105)	NC_009785_2	*Yes*	1
		TX1330			1		*Infantarius subsp. infantarius*	ATCC BAA-102			1
Exiguobacterium	*Sibiricum*	255-15	NC_010556_1	Yes	11		*Mitus*	ATCC 6249	1		2
									2		2
Fervidobacterium	*Nodosum*	Rt17-B1	NC_009718_1	Yes	3			SK321			1
			NC_009718_2	Yes	9						
Lactobacillus	*Acidophilus*	ATCC4796			8		*Mutans*	NN2025	NC_013928_1		8
									NC_013928_2		1
		NCFM	NC_006814_1		8			UA159 (ATCC700610)	NC_004350_1		1
	Antri	DSM16041T (LMG 22111T)			8		*Oralis*	ATCC35037			11
	Brevis	ATCC367	NC_008497_1		8		*Parasanguinis*	ATCC15912			5
	Buchneri	ATCC11577			8		*Pyogenes*	M1 GAS	NC_002737_1		1
									NC_002737_4		5

Table 1. Contd.

Casei	ATCC334	NC_008526_2	Yes	8			M49 591	1 2		1 5
	BL23	NC_010999_1	Yes	7			MGAS10270	NC_008022_1 NC_008022_3		1 5
	str. Zhang	NC_014334_6	Yes	7			MGAS10750	NC_008024_2		5
Crispatus	ST1	NC_014106_1	Yes	8			MGAS2096	NC_008023_2		5
Delbrueckii subsp. bulgaricus	ATCC11842	NC_008054_2	Yes	8			MGAS5005	NC_007297_1 NC_007297_4		1 5
	ATCC BAA-365	NC_008529_2	Yes	8			MGAS6180	NC_007296_2 NC_007296_3		1 5
	ND02	NC_014727_1	Yes	8			MGAS9429	NC_008021_2		5
	PB2003/044-T3-4			8			NZ131	NC_011375_1 NC_011375_3		1 5
Fermentum	IFO3956	NC_010610_3		8		Sanguinis	SK36	NC_009009_1		11
Helveticus	DPC4571	NC_010080_2 NC_010080_3		8 11		Sp.	M143			11
Iners	LactinV 01V1-a			10		Sp.oral taxon	071 str. 73H25AP	1 2		11 1
	LEAF 2052A-d			10		Suis	89/1591			1
	LEAF 2053A-b			10		Thermophilus	CNRZ1066	NC_006449_1		1
	LEAF 3008A-a			10			LMD-9 (=ATCC BAA-491)	NC_008532_2 NC_008532_4 NC_008532_5	*Yes* *Yes* *Yes*	1 11 9
	SPIN 2503V10-D			1			LMG18311	NC_006448_1 NC_006448_2		1 11
Jensenii	1153			1	*Thermoanaerobacterium*	Thermosaccharolyticum	DSM571	NC_014410_1 NC_014410_5	 *Yes*	1 1

Table 1. Contd.

115-3-CHN	1	*Thermotoga*	Lettingae	TMO (DSM14385)	NC_009828_1 NC_009828_2	*Yes*	1 1
208-1	1		Naphthophila	RKU-10	NC_013642_1 NC_013642_5	 *Yes*	1 1
269-3	1		Neapolitana	DSM4359	NC_011978_1 NC_011978_5	 *Yes*	1 1
27-2-CHN	1		Petrophila	RKU-1	NC_009486_4 NC_009486_5	 *Yes*	1 1
JV-V16	1			sp. RQ2	NC_010483_5	*Yes*	9

*The single numbers were the CRISPR id in draft LAB genomes designated according to the order of results by CRISPRs Finder.

usually had homologous DR sequence in the same genus and even in the same specie. This may suggest the sequence conservation of CRISPR loci in genus.

Multiple alignments and the second structures prediction of DRs

The DR multiple alignments for 11 clusters were conducted (Figure 2). Consistent to the previous report (Kunin et al., 2007), the bases with high folding scores had the pattern of base pairing, and this characteristic was possessed in majority DRs of all clusters. These paring bases may form the stem-loop structure. Meanwhile, we sampled one repeat from each cluster and predicted its secondary structure, and also the stem-loops were illustrated. The stem length, the loop size, and the pair-wise occurrence were not consistent, but all structures had flanking strands on two sides. With regard to the 5' end flanking sequence deficiency of *Bacillus mycoides* Rock3-17 and *Bacteroides fragilis* YCH46 CRISPR repeats, the 5' end base-pair probabilities of stem were intermediate, so the combination was not permanent.

Analysis of GC contents of CRISPRs

GC contents of CRISPR segments were displayed in Figure 3. In the 11 CRISPR samples, each repeat-spacer unit was 61-74 bp, and the repeat and the spacer were in the limit of 29-36 bp and 30-38 bp, respectively, which had little change. This was inconsistent with previous researches involved in the whole range of prokaryotes (Jansen et al., 2002). The fluctuation of GC content for each repeat-spacer unit was obvious in every graph.

DR sequences had great distinction on GC content in different segments, but they showed symmetrical appearances from the tendency of curves, especially on the stem domains. This provided another evidence of the stem-loop structure of CRISPR repeats (Kunin et al., 2007).

Meanwhile, we observed GC contents of spacers. Generally in the same CRISPR segment, they always kept a relative consistent level, and would not change sharply in a single spacer. When forming a new spacer, the foreign DNA fragment was often inserted into the front of CRISPR locus (Barrangou et al., 2007), so this characteristic suggested that there may have some pattern about the selection of DNA fragment from invasive chromosomes.

Comparisons on phylogenetic disparity between DRs, *cas1* and 16s rRNA

cas1 of 32 CRISPRs were obtained in 26 LAB strains from CRISPR database, among which two *cas1* were detected in each of four genomes, and three in one genome, respectively (Table 1). Since *cas* usually exist on the flanking sequences of CRISPR (Grissa et al., 2007), DRs in the range of 10 kb upstream and downstream of *cas1* were adopted to create the phylogenetic tree as

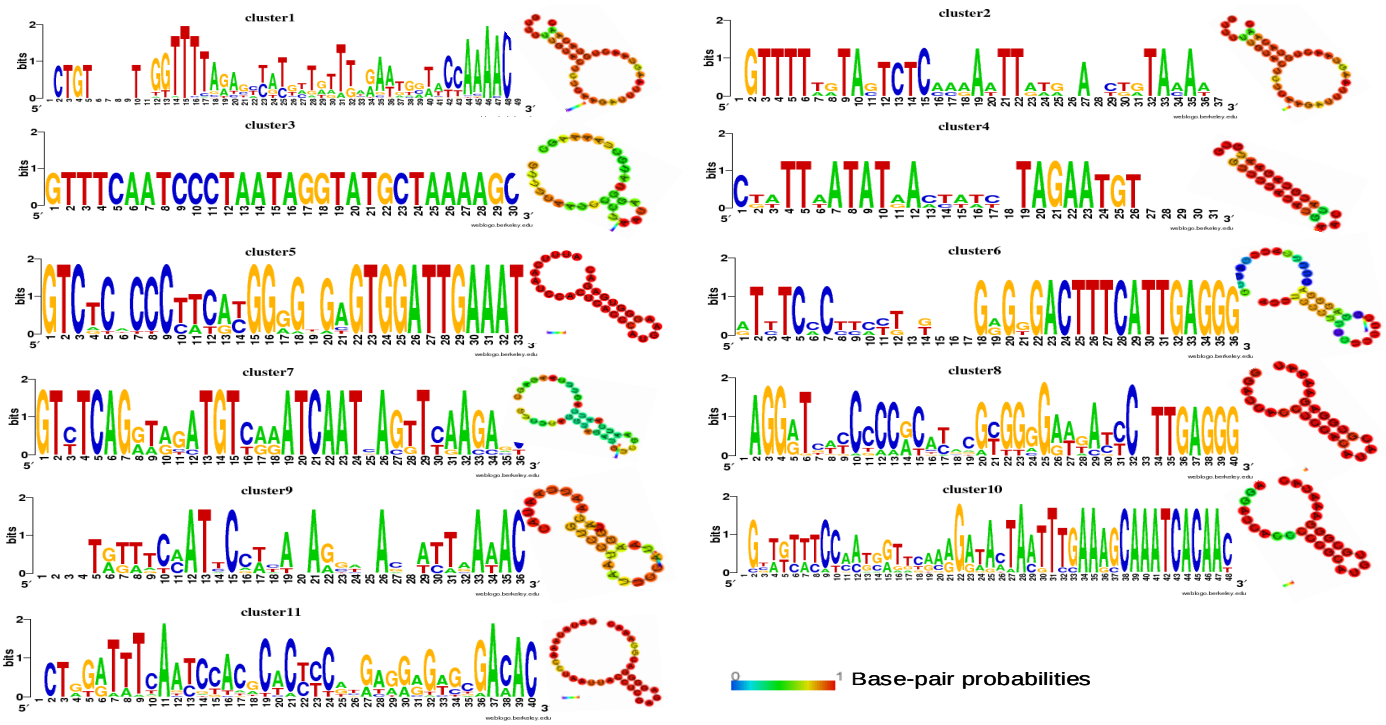

Figure 2. Sequence logo of DRs (left) and secondary structure prediction of samples (right) of 11 clusters. The sample repeats were randomly chosen in every cluster. They are in turns Cluster 1-*Lactobacillus jensenii* JV-V16, Cluster 2-*Streptococcus mitus* ATCC6249, Cluster 3-*Fervidobacterium nodosum* Rt17-B1(1), Cluster 4-*Bacillus mycoides* Rock3-17, Cluster 5-*Streptococcus pyogenes* MGAS2096, Cluster 6-*Rothia mucilaginosa* DY-18(7), Cluster 7-*Lactobacillus casei* BL23, cluster 8-*Lactobacillus acidophilus* NCFM, cluster 9-*Bacteroides fragilis* YCH46, cluster 10-*Lactobacillus iners* LEAF 3008A-a and Cluster 11-*Streptococcus thermophilus* LMD18311(2). The numbers in brackets denote CRISPR id. The colorized bar represents the base-pair probabilities of secondary structures.

comparison with the latter (Figure 4).

When the sequences with the bootstrap value below 50 were regarded as different clades, there were 4 clades for *cas1* tree, whereas 18 clades for DR tree. This suggested that *cas1* was more conservative than DRs. When the cut-off values were set 0.3 and 0.32, there were 4 and 3 clades for *cas1* and DR tree, respectively. Here interestingly, all strains in clade 3 of *cas1* tree were found in clade 1 of DR tree, likewise, the strains in clade 2 of DR tree were found in clade 4 of *cas1* tree, even in the same sub-clade.

This co-evolution distribution of *cas1* and DRs can also be validated in the network clustering graphs (Figure 5), nevertheless, different evolutionary patterns were identified in the comparison with 16s rRNA. 16s rRNA sequences of *Lactobacillus* in cluster 2 and *Streptococcus* in cluster 4 were scattered across the cluster 1 and cluster3 of *cas1* network graph, and the cluster 4 of *cas1* were constituted by multiple strains in cluster 1, 3, 4 and 5 of 16s rRNA.

DISCUSSION

In this study, 211 CRISPR loci in 192 strains were investigated. Compared with 66 CRISPRs of 49 LAB genomes in Horvath's research (Horvath and Barrangou, 2009), there was approximate 3.1 times in fact, so the consequences were more convincing. Among 588 genomes surveyed, CRISPR loci were found in 249 strains including those the encoding strand can not be determined, and the ratio of 40.6% was much lower than that in Horvath's research (46.1%) and the documented CRISPRs in bacterial genomes announced on CRISPR database (45.0%, 537 out of 1193 genomes) (Grissa et al., 2007).

cas1 and DRs were hypothesized to evolve as a complete cassette in CRISPR/Cas system of Bacteria (Chakraborty et al., 2010), and also displayed in the CRISPRs of LAB. However, *cas1* encodes a highly conserved protein Cas1, which displays the activity of nuclease or integrase as a member of CRISPR/Cas system (Makarova et al., 2006). *cas1* exhibited slightly more conservative than DRs, which seemed to suggest that this coevolution was not absolutely synchronous.

In the phylogenetic analysis, the same DR was identified to have high possibility to present in relatively distant genera. *cas1* had been reported to have the characteristic of horizontal gene transfer (HGT) in *Proteobacteria*, *Firmicutes* (Godde and Bickerton, 2006),

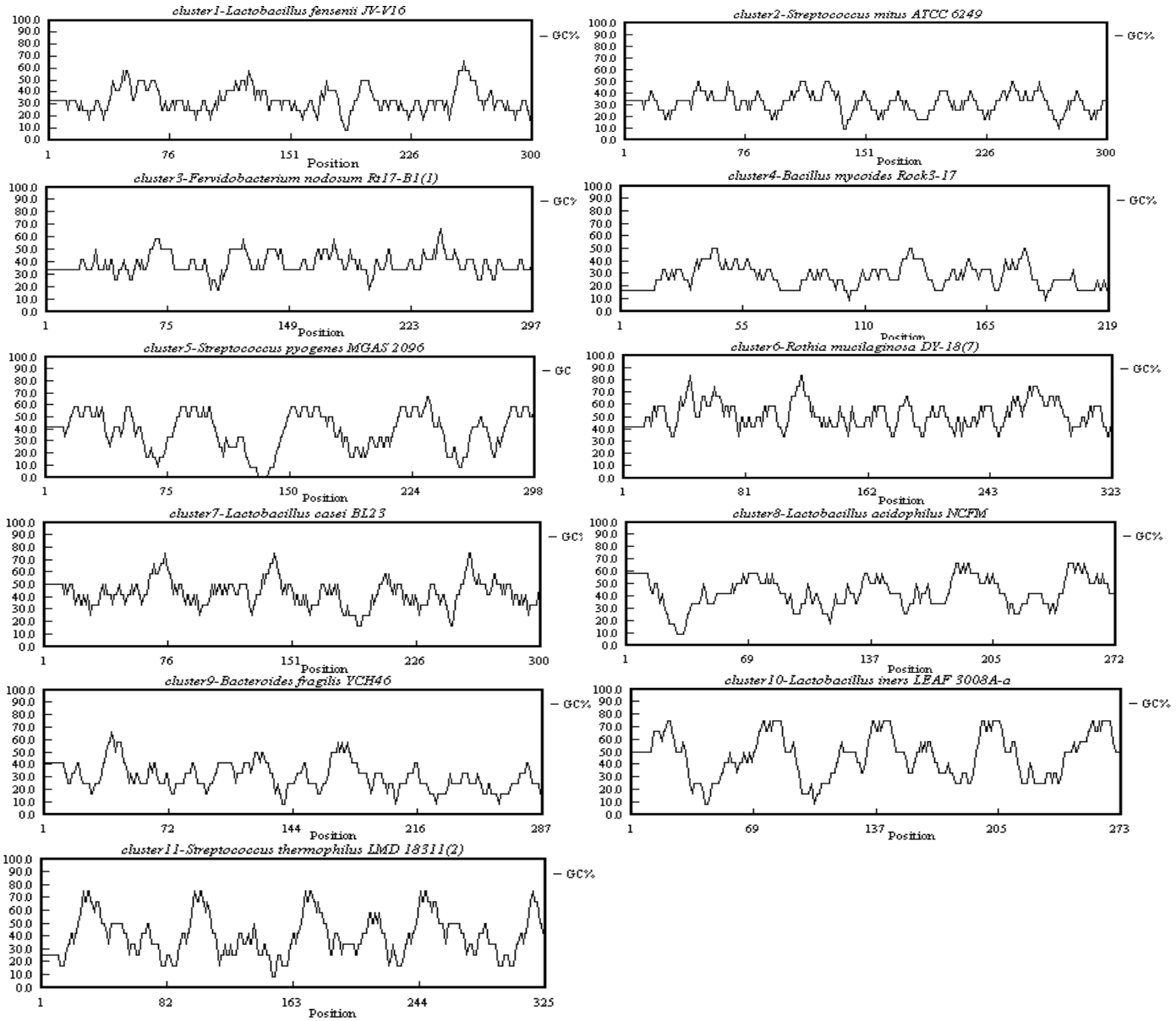

Figure 3. GC content tendency of CRISPR segments in samples of 11 clusters. Except the first four DRs for sample of cluster 4, the first five DRs were obtained for all the other clusters. Sample names were marked on the graphs, followed by CRISPR id in brackets.

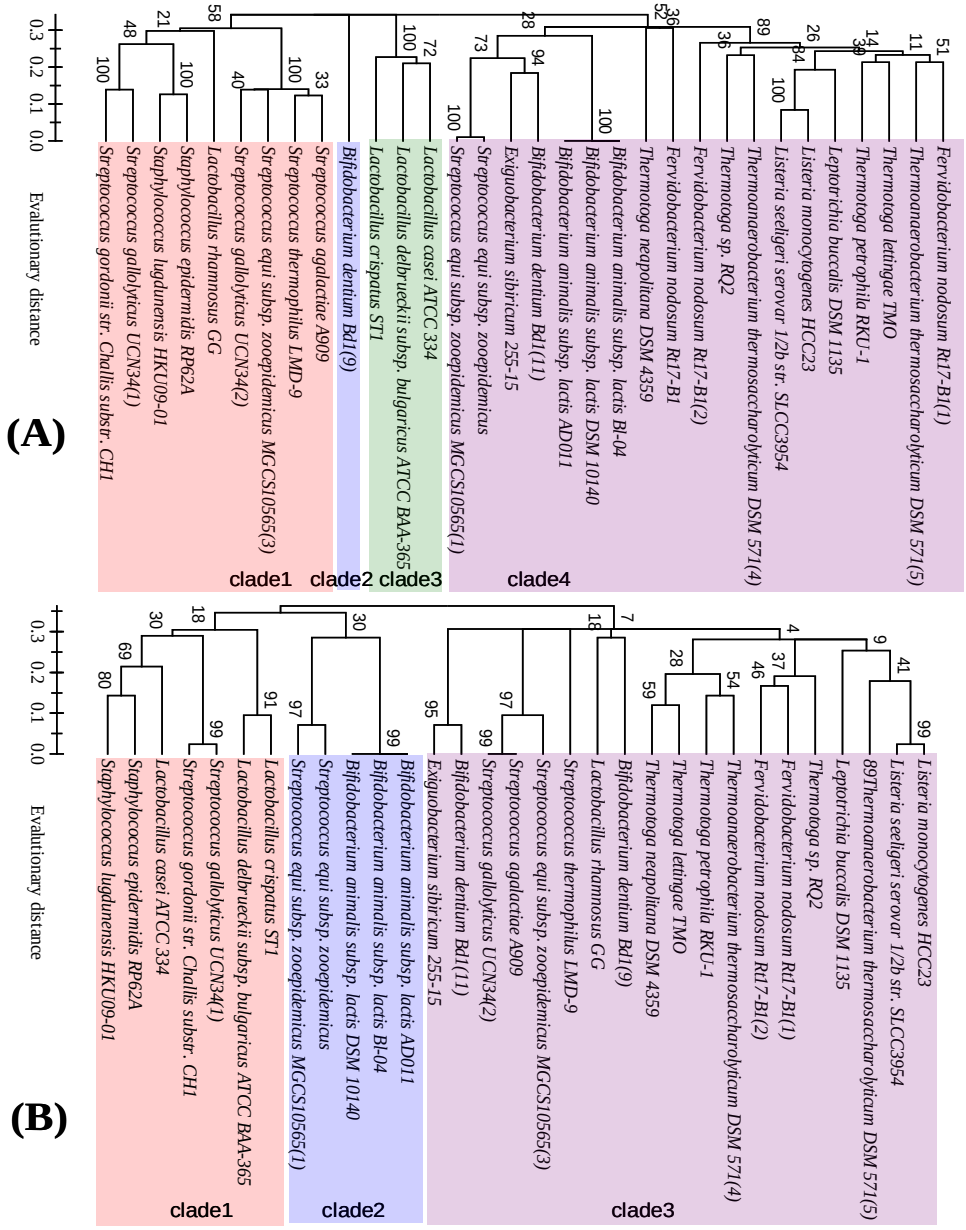

Figure 4. Phylogenetic trees of *cas1* (A) and related DRs (B) built by ClusterX and MEGA5 based on Neighbor-Joining method and clustered in bootstrap test (1000 replication). Evolutionary distances were computed in p-distance model.

and also in Archaea (Lillestøl et al., 2009). Due to the fact of DRs evolved together with *cas1* as a package, we can hypothesize that the DRs transfer across species in LAB genomes, namely CRISPR may possess the HGT pattern as well. 16s rRNA is acknowledged to be the target sequence for species identification (Muyzer et al., 1993). Therefore, *cas1* and DRs exhibited different evolutionary pathways with the phylogeny of strains.

In conclusion, 211 CRISPRs found in 588 LAB genomes were grouped into 11 clusters according to DR similarity, and as well as *cas1* gene, they were proved to have a particular evolutionary course compared to the species. The creation of varietal DRs may certificate the different evolutionary origins such as suffered from different environments. However, in the research of their DR sequences, it was found that they all can form a similar secondary structure, so it decided the same function when CRISPRs produce an effect (Kunin et al., 2007). CRISPRs have been known to be a potential mean for LAB resisting phages (Barrangou et al., 2007; Sorek et al., 2008), so the comparative evolutionary research of CRISPRs in LAB will be helpful for concluding the

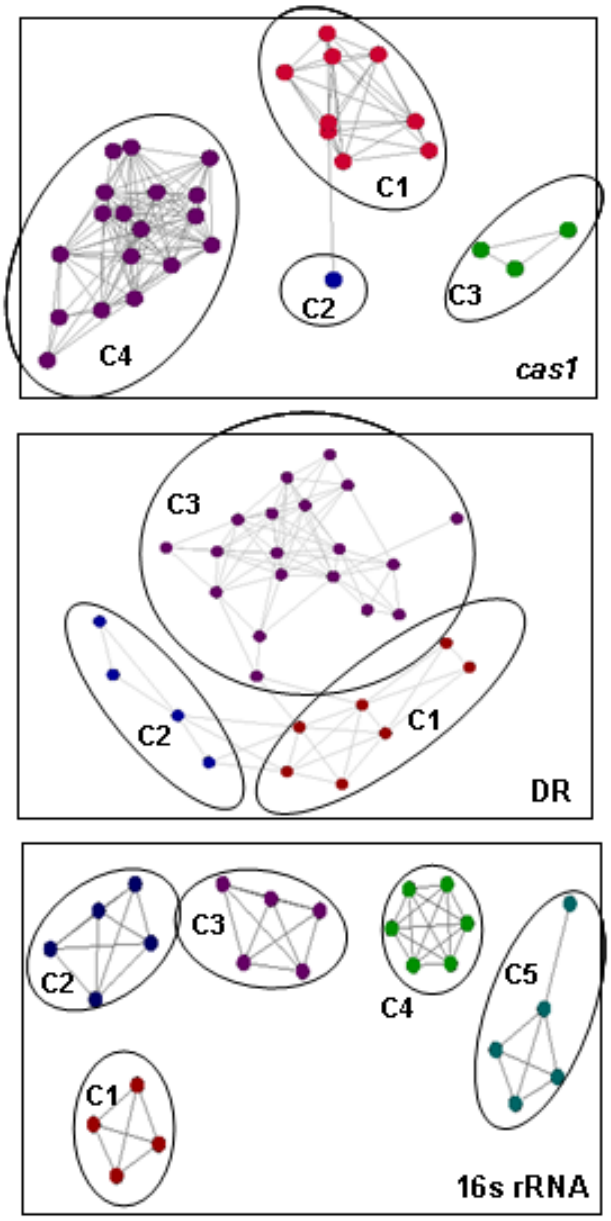

Figure 5. Network clusters for *cas1*, DRs and 16s rRNA. The min correlation was set 0.1, 0.2 and 0.5, and the squared correlation was set 0.001, 0.021 and 0.838 for them, respectively.

relationship of microbial community, and furthermore useful for developing superior starter cultures with high anti-phage activity during industrial process.

ACKNOWLEDGEMENT

This research was supported by National Natural Science funds of China (31171717), Program for Changjiang Scholars and Innovative Research Team in Northeast Agricultural University (IRT0959).

REFERENCES

Axelsson L (2004). Lactic acid bacteria: classification and physiology. In: Salminen S et al. (eds) Lactic acid bacteria microbiological and functional aspects, Marcel Dekker, Inc., USA, pp. 1-66.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science, 315 : 1709-1712.

Barrangou R, Horvath P (2009). The CRISPR system protects microbes against phages, plasmids. Microbe, 4: 224-230.

Chakraborty S, Snijders AP, Chakravorty R, Ahmed M, Tarek AM, Hossain MA (2010). Comparative network clustering of direct repeats (DRs) and *cas* genes confirms the possibility of the horizontal transfer of CRISPR locus among bacteria. Mol. Phylogenet. Evol., 56 : 878-887.

Garrity GM, Bell JA, Lilburn TG (2004). Bergey's manual of systematic bacteriology, 2nd edition, Springer Verlag, USA.

Godde JS, Bickerton A (2006). The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. J. Mol. Evol., 62 : 718-729.

Grissa I, Vergnaud G, Pourcel C (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinform., 8 : 172-182.

Guarner F, Malagelada JR (2003). Gut flora in health and disease. Lancet, 361 : 512-519.

Haft DH, Selengut J, Mongodin EF, Nelson KE (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput. Biol., 1: 474-483.

Hofacker IL (2003). Vienna RNA secondary structure server. Nucleic Acids Res., 31 : 3429-3431

Horvath P, Barrangou R (2009). CRISPR/Cas, the immune system of bacteria and archaea. Science, 327 : 167-170.

Horvath P, Romero DA, Coûté-Monvoisin AC, Richards M, Deveau H, Moineau S, Boyaval P, Fremaux C, Barrangou R (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. J. Bacteriol., 190 : 1401-1412.

Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A (1987). Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. J. Bacteriol., 169 : 5429-5433.

Jansen R, van Embden JDA, Gaastra W, Schouls LM (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol., 43 : 1565-1575.

Jarvis AW (1989). Bacteriophages of lactic acid bacteria. J. Dairy. Sci., 72 : 3406-3428.

Josephsen J, Neve H (2004). Bacteriophage and antiphage mechanisms of lactic acid bacteria. In: Salminen S et al. (eds) Lactic acid bacteria microbiological and functional aspects, Marcel Dekker, Inc. USA, pp. 295-350.

Kunin V, Sorek R, Hugenholtz P (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol., 8 : R61.

Lillestøl RK, Redder P, Garrett RA, Brügger K (2006). A putative viral defence mechanism in archaeal cells. Archaea, 2 : 59-72.

Lillestøl RK, Shah SA, Brügger K, Redder P, Phan H, Christiansen J, Garrett RA (2009). CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. Mol. Microbiol., 72 : 259-272.

Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol. Direct, 1 : 7.

Moineau S, Tremblay D, Labrie S (2002). Phage of lactic acid bacteria: from genomics to industrial applications. ASM News, 68 : 388-393.

Muyzer G, de Waal EC, Uitterlinden AG (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis

analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environ. Microb., 59 : 695-700.
Ouwehand AC, Kirjavainen PV, Shortt C, Salminen S (1999). Probiotics: mechanisms and established effects. Int. Dairy. J., 9: 43-52.
Rezzonico F, Smits THM, Duffy B (2011). Diversity, evolution and functionality of CRISPR regions in the fire blight pathogen Erwinia amylovora. Appl. Environ. Microbiol., 77 : 3819-3829.
Sorek R, Kunin V, Hugenholtz P (2008). CRISPR-a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat. Rev. Microbiol., 6 : 181-186.
Theocharidis A, van Dongen S, Enright AJ, Freeman TC (2009). Network visualization and analysis of gene expression data using BioLayout Express^{3D}. Nat. protoc., 4 : 1535-1550.