Journal of Microbiology (2014) Vol. 52, No. 6, pp. 452–459
DOI 10.1007/s12275-014-3586-2

Comparative Study of the *marR* Genes within the Family *Enterobacteriaceae*[§]

Dan Wang, Changjiang Guo, Longjiang Gu, and Xiaohui Zhang*

State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, 210093, Nanjing, P. R. China

(Received Nov 7, 2013 / Revised Dec 11, 2013 / Accepted Dec 30, 2013)

***marR* genes are members of an ancient family originally identified in *Escherichia coli*. This family is widely distributed in archaea and bacteria. Homologues of this family have a conserved winged helix fold. MarR proteins are involved in non-specific resistance systems conferring resistance to multiple antibiotics. Extensive studies have shown the importance of MarR proteins in physiology and pathogenicity in *Enterobacteria*, but little is known about their origin or evolution. In this study, all the *marR* genes in 43 enterobacterial genomes representing 14 genera were identified, and the phylogenetic relationships and genetic parameters were analyzed. Several major findings were made. Three conserved *marR* genes originated earlier than *Enterobacteriaceae* and a gene-loss event was found to have taken place in *Yersinia pestis Antiqua*. Three functional genes, *rovA*, *hor*, and *slyA*, were found to be clear orthologs among *Enterobacteriaceae*. The copy number of *marR* genes in *Enterobacteriaceae* was found to vary from 2 to 11. These *marR* genes exhibited a faster rate of nucleotide substitution than housekeeping genes did. Specifically, the regions of marR domain were found to be subject to strong purifying selection. The phylogenetic relationship and genetic parameter analyses were consistent with conservation and specificity of *marR* genes. These dual characters helped MarR to maintain a conserved binding motif and variable C-terminus, which are important to adaptive responses to a number of external stimuli in *Enterobacteriaceae*.**

Keywords: *marR*, gene family, *Enterobacteriaceae*, evolution

Introduction

Bacteria have many methods of microbial resistance that allow them to respond and adapt to a variety of environmental conditions, one of which involves transcriptional regulation of specific genes under specific sets of circumstances. These include the *MarRAB* operon, which was mapped to 34 min on the chromosome of *Escherichia coli* (George and Levy, 1983). This operon was initially discovered by selection of *E. coli* resistant to low levels of tetracycline or chloramphenicol. Proteins encoded by MarRAB included MarR, a repressor of the operon (Cohen *et al.*, 1993a; Ariza *et al.*, 1994). Alteration of the wild-type *marR* gene can contribute to increases in multi-resistance in bacteria (Cohen *et al.*, 1993a; Kern *et al.*, 2000). MarA, a positive transcriptional activator that produces multidrug resistance when overexpressed (Gambino *et al.*, 1993), and MarB, a protein essential to resistance but whose function is unknown (Cohen *et al.*, 1993a).

MarR was first found to be a negative repressor and additional homologous proteins were identified as transcriptional regulators subsequently. For example, MprA [EmrR] acts as a negative regulator and SlyA is a cytolysin (del Castillo *et al.*, 1991; Ludwig *et al.*, 1995). RovA, SlyA, and PecS have all been shown to positively and negatively affect expression of several genes (Libby *et al.*, 1994; Reverchon *et al.*, 1994; Ludwig *et al.*, 1995; Praillet *et al.*, 1996; Nasser *et al.*, 1999; Heroven *et al.*, 2004; Rouanet *et al.*, 2004; Navarre *et al.*, 2005; Ellison and Miller, 2006). The MarR family originated before the divergence of archaea and bacteria (Perez-Rueda and Collado-Vides, 2001). There are currently 336 putative members of the *marR* family in 45 species of bacteria and 13 species of archaea listed in the Clusters of Orthologous Groups database (http://www.ncbi.nlm.nih.gov/COG/). The DNA binding domains of MarR proteins show a conserved winged helix (or winged helix-turn-helix, wHTH) fold (Kumarevel, 2012). Alignment of proteins within the MarR family show conserved residues to be distributed throughout the entire length of the alignment (Wilkinson and Grove, 2006). All proteins show the most pronounced strongest homology and similarity in hydropathy plots at a particular hydrophilic 50-amino-acid domain within MarR (amino acids 50–107) (Seoane and Levy, 1995). In addition, the highly conserved amino acids of DxRxxxxx(L/I)TxxG extending from residue 132 to residue 144 appear to be a signature of this family of proteins (Sulavik *et al.*, 1995).

In addition to significant structural homology and wide distribution, primary sequence similarity was also detected among proteins in the MarR family. For example, the amino acid sequence of RovA (*Yersinia*) is 77% identical to that of the *Salmonella typhimurium* SlyA (Wu *et al.*, 2003). Comparative analysis was used to better understand the *marR* locus between and within genera. On one hand, the *S. typhimurium mar* locus is structurally and functionally similar to *marRAB* in *E. coli* (Sulavik *et al.*, 1997). The amino acid sequences of MarR showed 91% identity. On the other hand,

*For correspondence. E-mail: xiaohuizhang@nju.edu.cn; Tel.: +86-25-83686406; Fax: +86-25-83592705
[§]Supplemental material for this article may be found at http://www.springerlink.com/content/120956.

99% sequence homology was detected at this locus between *Salmonella enteritidis* and *S. typhimurium* using DNA primers based on the MarRAB operon of *S. typhimurium* (Kunonga *et al.*, 2000), suggesting a conservation of MarRAB operon within members of the family *Enterobacteriaceae* (Cohen *et al.*, 1993b).

In contrast to the biological functions, crystal structures, and sequence similarity of MarR homolog in family *Enterobacteriaceae*, which have been the subject of considerable study, little attention has been paid to the origin and evolution of MarR family. There is ample evidence showing that intra-chromosome duplication and horizontal gene transfer (HGT) plays an adaptive role in bacterial evolution (Alm *et al.*, 2006; Serres *et al.*, 2009). This has been shown to be the case with antibiotic resistance (Andersson and Hughes, 2009; Treangen *et al.*, 2009). Whether the ability to resist multiple antibiotics was transferred among taxa or developed independently in each lineage is still not clear, even though there has been great interest in the isolation of antibiotic-resistant strains. Therefore, it is important to investigate how *marR* genes have evolved and how and why bacteria have acquired *marR* genes. Fortunately, the genome sequences of many bacteria have been completed. This provides an opportunity to perform genome-wide comparisons to investigate the patterns of gene number and evolutionary characters within bacterial MarR family. In this study, 43 bacterial strains from 14 genera within the family *Enterobacteriaceae* were chosen for a comprehensive set of analyses to determine the number of relationships among *marR* genes in *Enterobacteria*, fostering significant advances in our understanding of the diversification of *marR* gene family and its evolution in multiple bacterial genomes.

Materials and Methods

Identification of *marR* genes

Complete sequences of 43 genomes (Supplementary data Table S1) from 14 genera in the family *Enterobacteriaceae* were downloaded from NCBI bacteria database (http://ftp.ncbi.nlm.nih.gov). Among the 43 bacterial strains, 13 species or subspecies were included in *Salmonella*; 4 species in *Shigella*; 3 species in *Erwinia*, *Pectobacterium*, and *Yersinia*, 2 species in *Citrobacter*, *Cronobacter*, *Dickeya*, *Edwardsiella*, *Enterobacter*, *Escherichia*, *Klebsiella*, *Pantoea*, and *Photorhabdus*, respectively. To identify MarR homologs in the bacterial genomes, tBLASTn searches were performed using amino acid sequences of the MarR domain against the coding sequences (CDS) of all 43 genomes with E-values of at least 1. Then the MarR domains were confirmed by Pfam database version 25.0 with E-values of at least 1 (http://pfam.sanger.ac.uk/) and SMART protein motif analyses (http://smart.embl-heidelberg.de/). All selected *marR* genes were used for further analyses.

Sequence alignment and data analysis

The nucleotide coding sequences (CDS) of *marR* genes were aligned in MEGA v5.0 according to the alignments of protein sequences with default options (Tamura *et al.*, 2011). The alignments were used to generate a phylogenetic tree using the bootstrap neighbor-joining (NJ) method with a p-distance model and MEGA v5.0. The stability of internal nodes was assessed using bootstrap analysis with 1000 replicates. Then the phylogeny was split into individual clades based on nucleotide diversity/divergence between orthologs and paralogs within the clades (0.3) combined with their bootstrap values (80%). Other criteria were also used for the division of all these *marR* genes. They were classified into 23 subfamilies according to more than 70% identity with 60% coverage as the cutoff. The subfamilies correspond closely with previous clades, suggesting that this classification was accurate. In the present study, they were considered the same. These clades were further classified into groups based on their highly conserved character and specificity among Enterobacteria.

After the group division of the phylogenic tree, the average nucleotide divergence of *marR* domain sequences in each clade was calculated. These values were estimated by *Dxy* with the Jukes and Cantor correction using MEGA 5.0 (Nei, 1987). The ratio of nonsynonymous to synonymous nucleotide substitutions (*dN/dS*, where dN and dS are defined as nonsynonymous substitutions synonymous substitutions respectively) can be used to detect the signature of selection. Then dN, dS, and *dN/dS* were calculated for each clade based on the Nei-Gojobori method with Jukes-Cantor correction (Nei and Gojobori, 1986). To detect the evolutionary constraints on the functional domains, the dN and dS of *marR* and regions outside this domain were also calculated. A sliding-window analysis of the ratio of nonsynonymous to synonymous substitution rates (*dN/dS*) was performed on the sequences of the three conserved subfamilies using DnaSP v5 (http://www.ub.edu/dnasp) (Librado and Rozas, 2009), with a window size of 30 bases and 5-base increments.

In addition, a housekeeping gene *rpoB*, encoding the bacterial RNA polymerase β-subunit is universal, and comparison of *rpoB* sequences has been used as basis for genetic analysis among bacteria (Pühler *et al.*, 1989). In our study, the same method was used to evaluate the *rpoB* gene to perform a comparative analysis. In contrast analysis, two different species within the same genus were defined as strain pair and used to assess all the *marR* and *rpoB* genes identified in these strains. Gene pairs were defined as any orthologous genes within any pair of strains. Then all *Dxy*, dN, dS, and *dN/dS* values were calculated for each gene pair. The average values of these parameters in each strain pair were calculated and these values were used for further comparison.

Results

Reconstructing the phylogenetic tree of *Enterobacteriaceae*

The phylogenetic relationship within family *Enterobacteriaceae* can be used to elucidate the evolutionary history of *marR* family. Based on 16S rDNA sequences, a previous study divided *Pectobacterium*, *Klebsiella*, and some other genera of the family *Enterobacteriaceae* into two groups (Groups A and B) (Spröer *et al.*, 1999). Herein, a species tree was reconstructed based on 16S rDNA sequences extracted from the genomes of 43 species/subspecies of *Enterobacteri-*

aceae (Fig. 1). The entire tree encompassed 14 genera, with 7 newly added. The phylogenetic tree indicated the relative relationships between different genera with strong support. Based on bootstrap value and previous study, seven clusters were delineated in the present study. Cluster 1 contained *Salmonella*, *Citrobacter*, *Escherichia*, *Shigella*, *Cronobacter*, *Enterobacter*, and *Klebsiella*, consistent with the statement that the genera *Escherichia*, *Klebsiella*, and *Enterobacter* should be integrated into a single genus based on DNA segmental homology (Murata and Starr, 1974). Cluster 2 was found to contain *Pantoea* and *Erwinia*. The remaining 5 clusters were placed in single genus. These clusters were consistent with previously identified phylogenetic relationships. Clusters 1, 2, 3, and 6 were found to correspond to groups A and B.

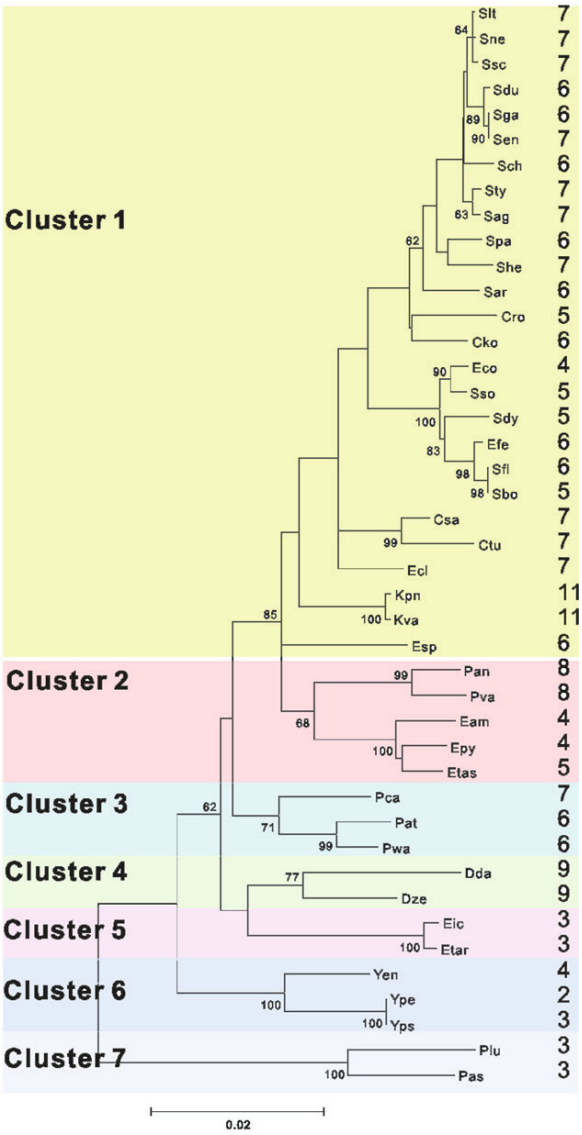

Fig. 1. Distribution of all identified *marR* genes among 43 *Enterobacteriaceae* in a clear phylogenetic background. This phylogenetic analysis was based on 16S rDNA. The numbers on the right are the copy numbers of MarR genes within each genome.

Number and domain organization of MarR proteins

tBLASTn and HMM searches were performed on the basis of sequence similarity and the conserved *marR* domain. Some 255 *marR* genes were identified in 43 sequenced bacterial genomes from 14 genera within the family *Enterobacteriaceae*.

The length of 255 identified sequences was highly variable, ranging from 378 to 1227 nucleotides. Domain organizations of MarR proteins varied. In addition to MarR domain, two types of other domains were detected in *Enterobacteriaceae* MarR homologous. Proteins with other domains (ROK and DypPerox) were found to be exclusive to each strain, suggesting a possible clear-ortholog relationship between species. There were many MarR proteins with as yet undefined (or unknown) C-terminal domains.

The average number of *marR* genes in all the 43 bacterial genomes was 5.9. The number in a single genome ranged from 2 in Ype to 11 in Kpn and Kva, respectively, with other strains falling in between (Fig. 1 and Supplementary data Table S1). Strains of the same genus were expected to have the same number of *marR* genes. However, exceptions were found in 9 of the 14 genera. For example, in *Yersinia*, Yen showed 4 *marR* genes, Yps showed 3, and Ype showed only 2 *marR* genes. In *Escherichia*, Eco showed 4 and Efe has 6 *marR* genes. The remaining 7 genera, differed by only one *marR*, but it was not always the same one. Copy number variations among closely related strains were here attributed to either gene gain or gene loss. Fewer *marR* genes were found in *Edwardsiella*, *Yersinia*, and *Photorhabdus* lineages than in other lineages, indicating an unequal distribution of *marR* genes among different groups.

Phylogeny and classification of *marR* genes

To evaluate the evolution of *marR* genes in family *Enterobacteriaceae*, all 255 MarR domain sequences from the 43 bacterial genomes were aligned and a phylogenetic tree was constructed in MEGA v5.0 (Supplementary data Fig. S1 and Fig. 2). Phylogenetic analyses allowed us to identify evolutionarily conservative and divergent *marR* genes. The tree was divided into 23 clades based on the nucleotide diversity/divergence between orthologs and paralogs within the clades (<30%) combined with their bootstrap values (>80%). The number of genes in these clades ranged from 2 to 43 with an average of 10.8. Three of the most conserved clades (clade 20, clade 22, and clade 23) were composed of sequences from all the 43 bacterial genomes excepting that Ype was not found in clade 23. This suggested that these three clades might have existed before the split of 14 genus bacteria and these genes may be related to some essential functions. No sequences from Ype were found in clade 23, indicating that

Table 1. Number of *marR* genes in four groups

Group name	Gene no.	Rate
Group1	128	0.502
Group2	48	0.188
Group3	35	0.137
Group4	38	0.149
Single gene	6	0.024

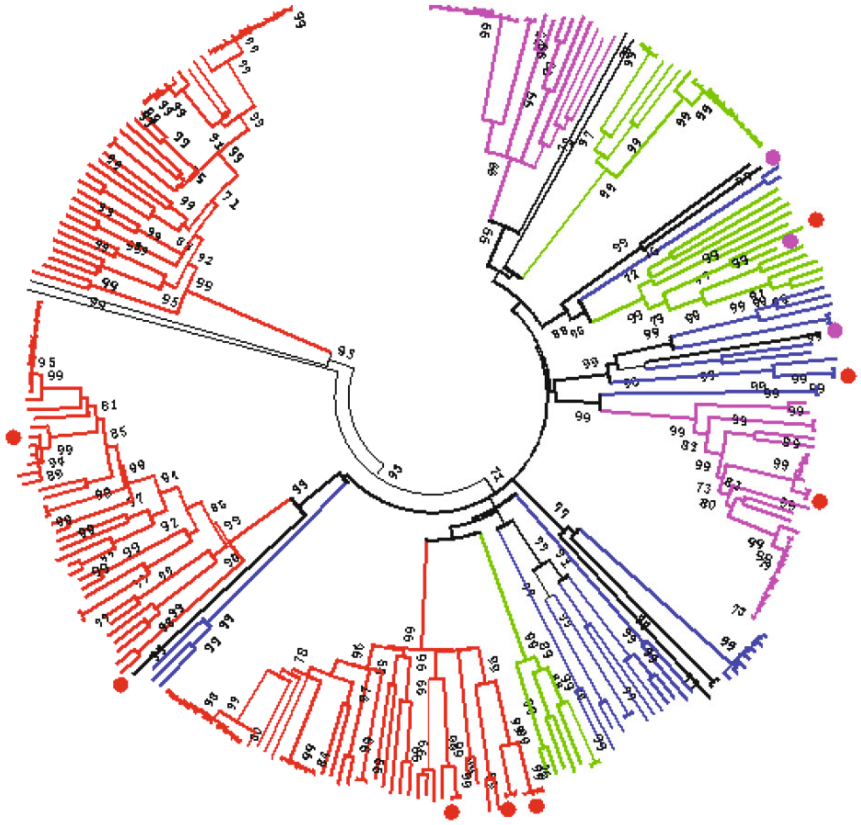

Fig. 2. Phylogenetic relationship among the MarR domain-related genes. Clades in red, purple, green, and blue correspond to groups 1 through 4, respectively. Red dots represent functional *marR* genes. Purple dots represent genes from outside *Enterobacteriaceae*.

a gene loss event occurred in the Ype strain at some point. However, in 15 cases, each clade was composed of sequences from the same genus, indicating genus-specific gene gains after split of different genera from the ancient common ancestor. The two types of clades above served as the conserved group (Group 1) and specific group (Group 4), respectively. The remaining 5 clades were in between. Based on the number of genera in which the clades' sequences were found, clades containing at least half of the 14 genera were classified into Group 2 and Group 3 was in between Groups 2 and 4. Group 2 was the moderately conserved group and Group 3 was the less conserved group. Among the identified 255 *marR* genes, 128 (50.2%), 48 (18.8%), 35 (13.7%), 38 (14.9%) genes were found in Groups 1–4, respectively (Table 1). Six genes that were not included in any clade were classified as species-specific genes.

As shown in Supplementary data Fig. S1, in Group 1, sequences from the 43 bacterial genomes formed the basal lineages of the three clades and their topology was generally consistent with the species tree based on 16S rDNA. This topology of clade 20, clade 22, and clade 23 strongly suggested that the division of the three subfamilies must have occurred earlier than the division of the family *Enterobacteriaceae*. Group 2 genes (clades 1 and 10) appeared to exist mainly in *E. coli* and its close relatives. They were not found in strains from other lineages. The results of phylogenetic analysis suggested a scenario in which most strains of *E. coli* and its close relatives of the two clades are orthologs, direct (vertical) descendants of their common ancestor. Within Group 4, incongruence was observed between *marR* gene clades and genome tree based on 16S rDNA. This is prima facie evidence of horizontal gene transfer. A best BLAST match targeting a distant taxon is widely considered sufficient to prove HGT (Ragan, 2001). An online Blast search in the Genbank database was performed. Results showed that 3 *marR* genes in *Serratia proteamaculans 568*, *Alcanivorax dieselolei B5*, and *Xanthomonas campestris* pv. *campestris str.* ATCC 33913 were closely related to the *marR* genes identified in *Enterobacteriaceae*, suggesting three HGT events (Supplementary data Fig. S1 and Fig. 2).

To investigate the relationship between clades and gene functions, eight functional and homologous *marR* genes were included in these phylogenetic relationships. The topology of phylogenetic tree showed that *rovA*, *hor* (Thomson *et al.*, 1997), *slyA*, *marR*, and *mprA(EmrR)* were located in Groups 1 and 2. *RovA*, *hor*, and *slyA* were classified into the same clade, suggesting a clear-ortholog relationship among them. The present results indicated that *marR* functional genes are prone to a conserved relationship in *Enterobacteriaceae.* However, three homologous genes were classified into very specific group. *PecS* was specific to *Dickeya* in the present study, which was reported to control enzymes that degrade the plant cell wall and affect hairpin and flagella synthesis in *Erwinia chrysanthemi* (Reverchon *et al.*, 1994; Rouanet *et al.*, 2004). The very high similarity and the fact that they share a function strongly suggest a lateral gene transfer event; *ohrR*, which was involved in the resistance of lipid hydroperoxide in *X. campestris* (Klomsiri *et al.*, 2005). The close relationship between *ohrR* and genes in *Enterobacteriaceae*, the conserved redox sensing Cys resi-

Table 2. Distribution of 23 clades in 4 groups

Nucleotide divergence (D)	D=0.0%		0.0<D≤4.0%		4.0%<D≤9%		9%<D≤16%		16<D≤26%		Total	
	C.No.	Gp. No.	C.No.	Gp. No.	C.No.	Gp. No.	C.No.	Gp. No.	C.No.	Gp. No.	C.No.	Gp. No.
Group1	0	5	0	22	3	15	0	19	0	9	3	70
Group2	0	0	0	11	2	1	0	2	0	2	2	16
Group3	0	0	0	3	1	2	2	5	0	3	3	13
Group4	1	1	4	4	1	2	6	12	2	2	14	21

C. No, clade number; Gp. No, gene pair number

dues and similar genetic organization indicated a HGT event (Reverchon *et al.*, 2010). Then 17 kDa, was obtained via HGT and deletion of this stretch did not affect the degree of piliation or the adhesion characteristics of the recombinant host (Marklund *et al.*, 1992).

In conclusion, conservatism and specificity were detected in *marR* genes based on phylogenetic relationships. The clear-ortholog relationship among *Enterobacteriaceae* was detected in three conserved clades and in three functional genes. This suggests the conserved nature and functional importance of *marR* genes to resistance to multiple antibiotics. The variable copy number among *Enterobacteriaceae* indicates frequent gene gain and loss events, especially for specific groups. In the present study, HGT events were identified in specific groups. Furthermore, some functional genes were found to be the result of HGT. This is consistent with its multiple functional characteristics in bacterial physiology and adaptive response to external stimuli.

Genetic variations comparison

Sets of orthologous protein sequences and corresponding nucleotide sequences shared by the 43 completely sequenced bacterial genomes were identified and aligned. Nucleotide sequences were used to calculate the divergence (*Dxy*) and synonymous (dS) and nonsynonymous (dN) substitution rates (see 'Materials and Methods'). To prevent the background effects of strains from different genera, all calculations were based on pairs of genes from different species within the same genus. In this way, clade 11, which contained strains of different subspecies, was excluded from the investigation.

The average of *Dxy* was 0.091 with a standard deviation of 0.064. The *Dxy* value among the 23 clades ranged from 0 in clade 6 to 0.234 in clade 15. Clades 6 and 15 also showed the highest and lowest values of dS and dN, respectively (data not shown). The highest and the lowest divergence values among the 22 clades were both observed in Group 4. The average nucleotide divergence values were 0.077, 0.055, 0.103, and 0.096 for Groups 1–4, respectively (Table 3). A significantly more pronounced level of volatility was detected in Groups 3 and 4 than in Groups 1 and 2 ($P<0.01$), suggesting an alternative evolutionary pattern in *marR* genes.

More stringent criteria were used to identify polymorphisms among orthologs within each genus. Specifically, 60.0% of the orthologs in Group 1 had ≤9% divergence. The proportions in Groups 2–4 were 75.0%, 38.5%, and 33.3% respectively (Table 2). Besides, *Dxy* with≤16%, ≤9% or ≤4% among orthologs, Group 2 still had the highest proportion of orthologs. And group3 and group4 have the highest proportions of orthologs with *Dxy* ranging from 16–26% and 9–16%, respectively.

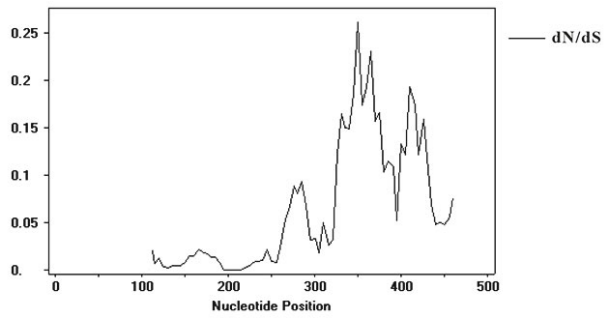

Fig. 3. Sliding window analysis of *dN/dS* and divergence values of clade 20. Nucleotide positions 136–306 correspond to the MarR domain. *dN/dS* ratios were calculated using DnaSP with a sliding window of 30 bases and 5-base increments.

Conservation of functional domains in *marR* genes

To evaluate the evolution of *marR* genes, and especially to elucidate the nucleotide substitution patterns among different regions of *marR* genes, a sliding window analysis of dN, dS, and *dN/dS* was carried out for three highly conserved clades using DnaSP v5.0. Interestingly, the dN of *marR* domains were consistently significantly lower than ROK domains and regions outside the domains, especially for clade 20, into which *rovA*, *slyA,* and *hor* were classified (Fig. 3). The present results suggested that MarR domain corresponding to nucleotide positions 136–306 had low *dN/dS* values. However, the dS value of MarR domains was similar to that of complete genes. MarR domain is likely to be responsible for DNA binding, suggesting a strong purifying selection on this section to keep a conserved winged-helix DNA binding motif.

More elaborate results were acquired to identify the exhaustive polymorphisms of MarR homologues. Genetic parameters were calculated in strain pairs within each genus to reduce the effects of the relationships between the strains.

Table 3. Divergence of four groups of *marR* gene clades and *rpoB* gene

Group	*Dxy*	dN	dS	*dN/dS*
Group1	0.077	0.018	0.424	0.094
Group2	0.055	0.02	0.287	0.082
Group3	0.103	0.051	0.449	0.108
Group4	0.096	0.036	0.502	0.095
rpoB	0.033	0.003	0.147	0.015

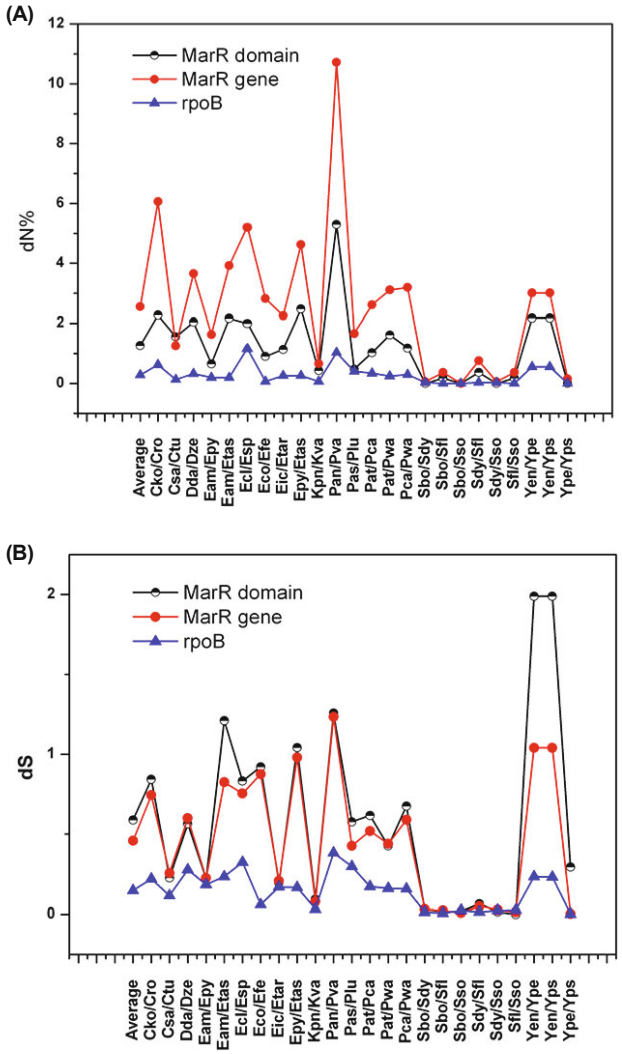

Fig. 4. Non-synonymous (A) and synonymous (B) substitutions of MarR domain, *marR* gene, and *rpoB* gene among pairs of strains.

The same was done to *marR* genes, nucleotide sequences of MarR domain, and a housekeeping gene *rpoB*. The average nucleotide divergence was 8.02%, 8.23%, and 3.27% for MarR domains, *marR* genes and *rpoB* genes respectively. Significant more genetic divergence were detected in nucleotide sequences corresponding to the *marR* genes and MarR domains than those of the *rpoB* genes ($P<0.05$). There is not much difference in divergence or synonymous substitutions between the complete gene and MarR domain regions. However, the average dN value was 1.26% and 2.55% for MarR domains and *marR* genes respectively. Significantly fewer non-synonymous substitutions and lower *dN/dS* values were detected in nucleotide sequences of MarR domain than the complete *marR* genes (Fig. 4) ($P<0.05$). The fact that closely related genes were nearly the same between strain pairs in *Shigella* suggest that the strains differentiated recently.

Discussion

Conserved nature and specificity of *marR* genes

MarR was first found to be a negative repressor in *E. coli* (George and Levy, 1983), and additional MarR homologs were later found. For example, SlyA in *S. typhimurium*, PecS in *E. chrysanthemi*, and RovA in *Yersinia enterocolitica* (Revell and Miller, 2000; Reverchon *et al.*, 2002). MarR have a wide distribution in both archaea and bacteria. However, the origin and evolution of this superfamily remain ambiguous. In the present study, a total of 255 *marR* genes were identified among 43 strains in *Enterobacteriaceae*. These could be divided into 23 clades based on their phylogenetic traits. The fact that the branching patterns of three conserved clades are congruent to the species tree based on 16S rDNA indicates that the three conserved genes were present in a common ancestor of *Enterobacteriaceae*, this is consistent with the conservation of MarRAB operon within family *Enterobacteriaceae* (Cohen *et al.*, 1993b). However, the two clades in group 2 were found to be present only in Cluster 1, indicating an unequal distribution of *marR* genes among different clusters of *Enterobacteriaceae*. The absence of two clades of *marR* genes from other lineages might be attributable to independent gene loss events that occurred during lineage-specific evolution. The results of phylogenetic analysis suggested a scenario in which Cluster 1 of the two clades are orthologs were direct (vertical) descendants of their common ancestor.

Multiple *Klebsiella* strains were isolated from hostile environments, for example, from hydrocarbon-contaminated estuaries or a heavily polluted environment (Rodrigues *et al.*, 2009; Xu *et al.*, 2010). Sophisticated systems may have evolved in these pathogenic bacteria to sense hostile environments and trigger compensatory gene expression in order to survive within the host. In our result, more genus-specific genes were detected in *Klebsiella* than in other genera. These MarR family homologues may play an important role in adaptive responses to a number of external stimuli.

Unequal distribution of *marR* genes among different groups was apparent. Variable copy number among *Enterobacteriaceae* indicates frequent gene gain and gene loss events, especially for specific groups. According to the present phylogenetic tree, fifteen genus-specific clades were identified. In many cases, the genus-specificity of these clades can be attributed either to the acquisition of genes through HGT or to the loss of ancestral genes. The present phylogenetic analysis revealed that 3 of the *marR* genes had been acquired from non-*Enterobacteriaceae* species, suggesting HGT playing an important character in rapid gene gain (Alm *et al.*, 2006; Serres *et al.*, 2009), especially to the case with antibiotic resistance (Andersson and Hughes, 2009; Treangen *et al.*, 2009). Furthermore, three functional genes were found to be the result of HGT, reflecting a scenario that the ability to resist multiple antibiotics can be transferred among taxa.

In addition to the different evolutionary patterns for conserved and specific clades, the present results suggest that clade type is relevant to gene function. Most of the functional genes are conserved within clades and lineages, suggesting a functional importance in *Enterobacteriaceae*. It has been suggested that function of uncharacterized proteins can be

predicted by their corresponding positions in the MarR family tree. In addition to, *rovA*, which is important to pathogenesis and virulence (Heroven *et al.*, 2004); *Hor*, which controls the production of antibiotics, of exoenzymes, and of determinants of virulence in phytopathogens (Thomson *et al.*, 1997); and *slyA*, which is important to the development of soft-rot disease on plants (Rouanet *et al.*, 2004), have also been shown to be clear orthologs among *Enterobacteriaceae*. This indicates the functional importance and a conserved evolutionary strategy of these genes. On the contrary, three specific *marR* genes, *pecS*, *ohrR*, and 17 kDa, were obtained through horizontal gene transfer and their functions usually involve responses to special conditions or play redundant roles in *Enterobacteriaceae* (Marklund *et al.*, 1992; Reverchon *et al.*, 1994; Klomsiri *et al.*, 2005).

The genetic parameters also reflected conservatism and specificity of *marR* genes. The average nucleotide divergence values were 0.077, 0.055, 0.103, and 0.096 for groups 1 through 4, respectively. The highest and lowest divergence values were both found in Group 4. The level of polymorphism was found to be significantly more volatile in Groups 3 and 4 than in Groups 1 and 2 ($P<0.01$). This suggests an alternative evolutionary pattern for *marR* genes.

Evolutionary patterns of *marR* genes

Significantly more nucleotide polymorphisms were detected in both the domain and complete gene of *marR* than in the housekeeping gene *rpoB*, suggesting a fast ratio of evolution for both domain region and complete gene of *marR*. Not much difference in divergence or synonymous substitutions was found between the complete gene and MarR domain. However, the average dN values were 1.26% and 2.55% for *marR* domain and *marR* genes respectively. Significant less nonsynonymous substitutions were detected in *marR* domain regions ($P<0.05$), suggesting a severe mutation constraint on these sections. Inconsistent substitutions rate within the same gene tend to correspond to different types of function. The MarR domain (β1-a3-a4-β2-W1(wing)-β3) is a dimer with each subunit containing a winged-helix DNA binding motif (Alekshun *et al.*, 2001). The N-terminus was shown to be essential to repressor activity (Seoane and Levy, 1995). Mutations in both a4 and W1 affect the DNA binding activity of MarR because amino acids from each of these regions contribute to the DNA binding activity of the protein. For example, mutations in these regions have been shown to abolish MarR DNA binding activity in whole cells and *in vitro* (Alekshun *et al.*, 2000). Our sliding window analysis demonstrated that *marR* domains often exhibit very low *dN/dS* values. The level of conservation of MarR homologues is particularly high in the N-terminus, but it is relatively low in the C-terminus (Thomson *et al.*, 1997). The DNA binding activity endows the MarR domain with a particular feature. For example, it may undergo severe purifying selection to preserve a conserved winged-helix DNA binding motif. The variable C-terminus may be largely responsible for the modulation of the relative specificity of these closely homologous genes (Ludwig *et al.*, 1995; Thomson *et al.*, 1997). These opposite evolutionary patterns have been observed both within and outside functional domains of *marR* genes. This likely reflects a sophisticated strategy in the evolution of repressors of multiple antibiotic resistance. The findings of the present and previous studies suggested that negative selection pressure allows the MarR domain to maintain the conserved functions and positive selection pressure that drive the evolution of C-terminus, which then helps modulate the relative specificity of these closely homologous genes. These dual characters helped MarR to maintain a conserved binding motif and variable C-terminus, which are important to adaptive responses to a number of external stimuli in *Enterobacteriaceae*.

References

Alekshun, M.N., Kim, Y.S., and Levy, S.B. 2000. Mutational analysis of MarR, the negative regulator of marRAB expression in *Escherichia coli*, suggests the presence of two regions required for DNA binding. *Mol. Microbiol.* **35**, 1394–1404.

Alekshun, M.N., Levy, S.B., Mealy, T.R., Seaton, B.A., and Head, J.F. 2001. The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 Å resolution. *Nat. Struct. Mol. Biol.* **8**, 710–714.

Alm, E., Huang, K., and Arkin, A. 2006. The evolution of two-component systems in bacteria reveals different strategies for niche adaptation. *PLoS Comput. Biol.* **2**, e143.

Andersson, D.I. and Hughes, D. 2009. Gene amplification and adaptive evolution in bacteria. *Annu. Rev. Genet.* **43**, 167–195.

Ariza, R., Cohen, S., Bachhawat, N., Levy, S., and Demple, B. 1994. Repressor mutations in the marRAB operon that activate oxidative stress genes and multiple antibiotic resistance in *Escherichia coli*. *J. Bacteriol.* **176**, 143–148.

Cohen, S.P., Hächler, H., and Levy, S. 1993a. Genetic and functional analysis of the multiple antibiotic resistance (mar) locus in *Escherichia coli*. *J. Bacteriol.* **175**, 1484–1492.

Cohen, S.P., Yan, W., and Levy, S.B. 1993b. A multidrug resistance regulatory chromosomal locus is widespread among enteric bacteria. *J. Infect. Dis.* **168**, 484–488.

del Castillo, I., González-Pastor, J.E., San Millán, J.L., and Moreno, F. 1991. Nucleotide sequence of the *Escherichia coli* regulatory gene *mprA* and construction and characterization of mprA-deficient mutants. *J. Bacteriol.* **173**, 3924–3929.

Ellison, D.W. and Miller, V.L. 2006. Regulation of virulence by members of the MarR/SlyA family. *Curr. Opin. Microbiol.* **9**, 153–159.

Gambino, L., Gracheck, S.J., and Miller, P.F. 1993. Overexpression of the MarA positive regulator is sufficient to confer multiple antibiotic resistance in *Escherichia coli*. *J. Bacteriol.* **175**, 2888–2894.

George, A. and Levy, S. 1983. Gene in the major cotransduction gap of the *Escherichia coli* K-12 linkage map required for the expression of chromosomal resistance to tetracycline and other antibiotics. *J. Bacteriol.* **155**, 541–548.

George, A.M. and Levy, S.B. 1983. Amplifiable resistance to tetracycline, chloramphenicol, and other antibiotics in *Escherichia coli*: involvement of a non-plasmid-determined efflux of tetracycline. *J. Bacteriol.* **155**, 531–540.

Heroven, A.K., Nagel, G., Tran, H.J., Parr, S., and Dersch, P. 2004. RovA is autoregulated and antagonizes H-NS-mediated silencing of invasin and rovA expression in *Yersinia pseudotuberculosis*. *Mol. Microbiol.* **53**, 871–888.

Kern, W., Oethinger, M., Jellen-Ritter, A., and Levy, S. 2000. Non-target gene mutations in the development of fluoroquinolone resistance in *Escherichia coli*. *Antimicrob. Agents Chemother.* **44**, 814–820.

Klomsiri, C., Panmanee, W., Dharmsthiti, S., Vattanaviboon, P., and Mongkolsuk, S. 2005. Novel roles of *ohrR-ohr* in *Xanthomonas* sensing, metabolism, and physiological adaptive response to lipid hydroperoxide. *J. Bacteriol.* **187**, 3277–3281.

Kumarevel, T. 2012. The MarR family of transcriptional regulators–a structural perspective, pp. 403–418. *In* Pana, D.M. (ed.), InTech, Rijeka, Croatia.

Kunonga, N.I., Sobieski, R.J., and Crupper, S.S. 2000. Prevalence of the multiple antibiotic resistance operon (marRAB) in the genus *Salmonella*. *FEMS Microbiol. Lett.* **187**, 155–160.

Libby, S.J., Goebel, W., Ludwig, A., Buchmeier, N., Bowe, F., Fang, F.C., Guiney, D.G., Songer, J.G., and Heffron, F. 1994. A cytolysin encoded by Salmonella is required for survival within macrophages. *Proc. Natl. Acad. Sci. USA* **91**, 489–493.

Librado, P. and Rozas, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452.

Ludwig, A., Tengel, C., Bauer, S., Bubert, A., Benz, R., Mollenkopf, H.J., and Goebel, W. 1995. SlyA, a regulatory protein from *Salmonella typhimurium*, induces a haemolytic and pore-forming protein in *Escherichia coli*. *Mol. Gen. Genet.* **249**, 474–486.

Marklund, B.l., Tennent, J.M., Garcia, E., Hamers, A., Baga, M., Lindberg, F., and Gaastra, W. 1992. Horizontal gene transfer of the *Escherichia coli* pap and prs pili operons as a mechanism for the development of tissue-specific adhesive properties. *Mol. Microbiol.* **6**, 2225–2242.

Murata, N. and Starr, M.P. 1974. Intrageneric clustering and divergence of Erwinia strains from plants and man in the light of deoxyribonucleic acid segmental homology. *Can. J. Microbiol.* **20**, 1545–1565.

Nasser, W., Shevchik, V.E., and Hugouvieux-Cotte-Pattat, N. 1999. Analysis of three clustered polygalacturonase genes in *Erwinia chrysanthemi* 3937 revealed an anti-repressor function for the PecS regulator. *Mol. Microbiol.* **34**, 641–650.

Navarre, W.W., Halsey, T.A., Walthers, D., Frye, J., McClelland, M., Potter, J.L., Kenney, L.J., Gunn, J.S., Fang, F.C., and Libby, S.J. 2005. Co-regulation of *Salmonella enterica* genes required for virulence and resistance to antimicrobial peptides by SlyA and PhoP/PhoQ. *Mol. Microbiol.* **56**, 492–508.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York, USA.

Nei, M. and Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418–426.

Perez-Rueda, E. and Collado-Vides, J. 2001. Common history at the origin of the position-function correlation in transcriptional regulators in Archaea and Bacteria. *J. Mol. Evol.* **53**, 172–179.

Praillet, T., Nasser, W., Robert-Baudouy, J., and Reverchon, S. 1996. Purification and functional characterization of PecS, a regulator of virulence-factor synthesis in *Erwinia chrysanthemi*. *Mol. Microbiol.* **20**, 391–402.

Pühler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H.P., Lottspeich, F., Garrett, R.A., and Zillig, W. 1989. Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. *Proc. Natl. Acad. Sci. USA* **86**, 4569–4573.

Ragan, M.A. 2001. Detection of lateral gene transfer among microbial genomes. *Curr. Opin. Genet. Dev.* **11**, 620–626.

Revell, P.A. and Miller, V.L. 2000. A chromosomally encoded regulator is required for expression of the *Yersinia enterocolitica inv* gene and for virulence. *Mol. Microbiol.* **35**, 677–685.

Reverchon, S., Nasser, W., and Robert-Baudouy, J. 1994. pecS: a locus controlling pectinase, cellulase and blue pigment production in *Erwinia chrysanthemi*. *Mol. Microbiol.* **11**, 1127–1139.

Reverchon, S., Nasser, W., and Robert-Baudouy, J. 1994. pecS: a locus controlling pectinase, cellulase and blue pigment production in *Erwinia chrysanthemi*. *Mol. Microbiol.* **11**, 1127–1139.

Reverchon, S., Rouanet, C., Expert, D., and Nasser, W. 2002. Characterization of indigoidine biosynthetic genes in *Erwinia chrysanthemi* and role of this blue pigment in pathogenicity. *J. Bacteriol.* **184**, 654–665.

Reverchon, S., Van Gijsegem, F., Effantin, G., Zghidi-Abouzid, O., and Nasser, W. 2010. Systematic targeted mutagenesis of the MarR/SlyA family members of *Dickeya dadantii* 3937 reveals a role for MfbR in the modulation of virulence gene expression in response to acidic pH. *Mol. Microbiol.* **78**, 1018–1037.

Rodrigues, D.F., Sakata, S.K., Comasseto, J.V., Bicego, M.C., and Pellizari, V.H. 2009. Diversity of hydrocarbon-degrading *Klebsiella* strains isolated from hydrocarbon-contaminated estuaries. *J. Appl. Microbiol.* **106**, 1304–1314.

Rouanet, C., Reverchon, S., Rodionov, D.A., and Nasser, W. 2004. Definition of a consensus DNA-binding site for PecS, a global regulator of virulence gene expression in *Erwinia chrysanthemi* and identification of new members of the PecS regulon. *J. Biol. Chem.* **279**, 30158–30167.

Seoane, A.S. and Levy, S.B. 1995. Characterization of MarR, the repressor of the multiple antibiotic resistance (mar) operon in *Escherichia coli*. *J. Bacteriol.* **177**, 3414–3419.

Serres, M., Kerr, A., McCormack, T., and Riley, M. 2009. Evolution by leaps: gene duplication in bacteria. *Biol. Direct.* **4**, 46.

Spröer, C., Mendrock, U., Swiderski, J., Lang, E., and Stackebrandt, E. 1999. The phylogenetic position of *Serratia*, *Buttiauxella* and some other genera of the family *Enterobacteriaceae*. *Int. J. Syst. Evol. Microbiol.* **49**, 1433–1438.

Sulavik, M., Gambino, L., and Miller, P. 1995. The MarR repressor of the multiple antibiotic resistance (mar) operon in *Escherichia coli*: prototypic member of a family of bacterial regulatory proteins involved in sensing phenolic compounds. *Mol. Med.* **1**, 436.

Sulavik, M.C., Dazer, M., and Miller, P.F. 1997. The *Salmonella typhimurium mar* locus: molecular and genetic analyses and assessment of its role in virulence. *J. Bacteriol.* **179**, 1857–1866.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.

Thomson, N., Cox, A., Bycroft, B., Stewart, G., Williams, P., and Salmond, G. 1997. The rap and hor proteins of *Erwinia*, *Serratia* and *Yersinia*: a novel subgroup in a growing superfamily of proteins regulating diverse physiological processes in bacterial pathogens. *Mol. Microbiol.* **26**, 531–544.

Treangen, T.J., Abraham, A.L., Touchon, M., and Rocha, E.P. 2009. Genesis, effects and fates of repeats in prokaryotic genomes. *FEMS Microbiol. Rev.* **33**, 539–571.

Wilkinson, S.P. and Grove, A. 2006. Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins. *Curr. Issues Mol. Biol.* **8**, 51.

Wu, R.-y., Zhang, R.-g., Zagnitko, O., Dementieva, I., Maltzev, N., Watson, J.D., Laskowski, R., Gornicki, P., and Joachimiak, A. 2003. Crystal structure of *Enterococcus faecalis* SlyA-like transcriptional factor. *J. Biol. Chem.* **278**, 20240–20244.

Xu, J., Li, W., Chen, X., and Zhou, Y. 2010. *Klebsiella alba* sp. nov., a novel pesticide-tolerant bacterium from a heavily polluted environment. *J. Gen. Appl. Microbiol.* **56**, 241–247.