Research Article

Parallel Evolution and Horizontal Gene Transfer of the pst Operon in Firmicutes from Oligotrophic Environments

Alejandra Moreno-Letelier,1, 2 Gabriela Olmedo,2 Luis E. Eguiarte,1 Leon Martinez-Castilla,3 and Valeria Souza1

1 Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Apdo. Postal 70-275, Ciudad Universitaria, 04510 México D. F., Mexico
2 Departamento de Ingeniería Genética, CINVESTAV Campus Guanajuato, Apdo. Postal 629, 36500 Irapuato, Mexico
3 Departamento de Bioquímica, Facultad de Química, Universidad Nacional Autónoma de México, Apdo. Postal 70-275, Ciudad Universitaria, 04510 México D. F., Mexico

Correspondence should be addressed to Valeria Souza, souza@servidor.unam.mx

Received 22 October 2010; Accepted 22 December 2010

Academic Editor: Hiromi Nishida

Copyright © 2011 Alejandra Moreno-Letelier et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The high affinity phosphate transport system (pst) is crucial for phosphate uptake in oligotrophic environments. Cuatro Cienegas Basin (CCB) has extremely low P levels and its endemic Bacillus are closely related to oligotrophic marine Firmicutes. Thus, we expected the pst operon of CCB to share the same evolutionary history and protein similarity to marine Firmicutes. Orthologs of the pst operon were searched in 55 genomes of Firmicutes and 13 outgroups. Phylogenetic reconstructions were performed for the pst operon and 14 concatenated housekeeping genes using maximum likelihood methods. Conserved domains and 3D structures of the phosphate-binding protein (PstS) were also analyzed. The pst operon of Firmicutes shows two highly divergent clades with no correlation to the type of habitat nor a phylogenetic congruence, suggesting horizontal gene transfer. Despite sequence divergence, the PstS protein had a similar 3D structure, which could be due to parallel evolution after horizontal gene transfer events.

1. Introduction

Phosphorus is an essential nutrient for multiple processes such as the synthesis of DNA, RNA, ATP, and many other pathways involving phosphorylation [1]. However, it is not an abundant element on the planet and can only be obtained form organic detritus or from tectonics and volcanism [2, 3], so, its availability is a limiting factor for all life forms. As growth rate and primary productivity are highly dependent on phosphorus [4–6], bacteria have different mechanisms for the uptake and storage of phosphates to be able to cope with this limitation [1, 7–9].

Some of the genes involved in phosphorus metabolism belong to the pho regulon that is induced by phosphorus starvation by a two-component regulatory system in several bacteria such as Escherichia coli, Bacillus subtilis, and Cyanobacteria [8, 10–13]. The pho regulon is comprised of 20 or so genes that include phosphatases, phosphate transport systems, and other enzymes used to assimilate phosphorus from other sources such as phosphonates [8]. Even though the pho regulon is found in both Eubacteria and Archaea, the number and identity of the genes are highly variable and not always congruent with the 16S rRNA gene phylogeny of the organisms [11, 14]. It is also to be expected that the genes involved in phosphate uptake and metabolism would be under strong selection.

Among the genes of the pho regulon, the high affinity phosphate transport system (pst) is thought to be responsible for phosphate uptake under nutrient stress [8, 10]. Pst is a typical ABC transport system encoded in 4 to 6 genes in a single operon [10, 15–17]. As an ABC transporter, the pst operon belongs to one of the largest gene families and is
found in all Eubacteria and Archaea and the level of sequence divergence indicates an ancient origin of each lineage of transporters [18, 19].

The genes of the pst operon are arranged in the following way: the pstS gene, coding for a periplasmic protein that binds phosphate with high affinity; pstC and pstA, coding for the two proteins proposed to form the inner membrane channel; pstB, coding for an ATPase that energizes the transport [18]. However, some variation exists in the number of genes in the operon. In Escherichia coli and Clostridium acetobutylicum, the gene phoU, coding for a repressor of the pho regulon, is also located in the operon [15, 17], while in B. subtilis and its close relatives there are no phoU orthologs. Also, the gene pstB is duplicated (pstBA and pstBB; [10]). The pst operon presents further variation in Cyanobacteria, where the genes pstS or pstB may be missing from the operon depending on the strain and environmental conditions [11], or additional pstS copies may be present although not associated to the operon [11, 20].

The pst phosphate uptake system is particularly crucial in oligotrophic environments such as the North Pacific, North Atlantic and the Eastern Mediterranean Sea [6, 21]. Metagenomic studies have shown that there are some functional adaptations for P uptake in such oligotrophic waters [7, 8, 20, 22]. Another example of an extreme oligotrophic environment is the Cuatro Ciéngas Basin (CCB), that presents very low levels of P in the ecosystem [4, 23, 24]. Phosphate concentrations range from 0.008 to 0.6 μM, in Pozas Azules and Rio Mezquites, respectively (E. Rebollar and F. García-Oliva pers. com.; [4]), but for most water systems P concentrations lie below the threshold concentration for the expression of the pho regulon in B. subtilis (0.1 mM; [10]).

CCB is an isolated oasis in the center of the Chihuahuan Desert, with water systems rich in microbial mats and stromatolites, and its microbiota exhibits ancestral marine affinities [9, 25–29]. Despite the extreme oligotrophy of the ecosystem, CCB has a high level of diversity and species endemism both at the macro- and microscopic levels [24, 25, 30–33]. We believe that this high rate of diversification is a consequence of the extreme oligotrophy of the ecosystem [24], where the lack of available P promotes both reproductive and geographic isolation, by limiting replication and the frequency of genetic exchange [24, 34–36]. Moreover, two of the newly sequenced taxa, Bacillus coahuilensis and Bacillus sp. m3–13, have particular adaptations to low P environments. Unlike Escherichia coli or B. subtilis, CCB and marine Bacillus lack the low affinity phosphate uptake system so they must rely on the high affinity transport system [9, 27].

There are some comparative studies about genes involved in phosphorus uptake in Cyanobacteria [8, 11, 20], but as far as we know, no studies exist in other bacterial groups. Hence, we believe that an analysis of the phosphorus uptake in the Firmicutes from CCB in comparison to sequenced Firmicutes from different environments could help us understand the evolution of the high affinity phosphate transport system. Firmicutes is a cosmopolitan and ancient lineage [37], and their diversification happened during a time in the Earth’s history where P was very scarce [3, 5]. We expected the pst operon of the Firmicutes from CCB to have a marine affinity and to be related (both in sequence and structure) to the pst operons of other marine Firmicutes that live in oligotrophic waters.

In this study we analyzed for the first time the evolutionary relationships, gene architecture, of the pst operons of 55 complete genomes of the main lineages of Firmicutes [38] with special emphasis on CCB and marine taxa, as well as the protein structure of PstS from a few Bacillus. To evaluate phylogenetic congruity between the phosphate uptake genes and housekeeping genes, expected to reflect vertical descent, we performed a phylogenetic reconstruction of the genes of the pst operon and of 14 proteins of the core genome of Firmicutes. We also compared the protein structure of phosphate-binding protein PstS of Bacillus from oligotrophic and eutrophic environments, to try to evaluate any association between protein sequence and structure to the environment in which the members of Firmicutes live.

2. Materials and Methods

2.1. Phylogenetic Reconstructions. We used the amino acid sequence of the substrate-binding protein gene pstS of Bacillus coahuilensis and Bacillus subtilis [9, 39] to identify the orthologs of the pst operon in the draft and complete genomes of the main lineages of Firmicutes (for accession numbers see Table S1 of Supplementary Material available online at doi: 10.4061/2011/781642). Searches were performed using psi-Blast, and the sequences identified with at least 30% of identity over a minimum of 70% in length, and e-value < 10−35 were considered as orthologs [27]. As the pstS gene can be duplicated in some genomes, the operon structure of the genes in the operon was manually checked in all cases, and only genes with the highest bit scores and lowest e-values were considered in cases of multiple hits in the same genome. For the cases with multiple hits of the entire operon, all those extra copies of the operon were also included in the analysis. We also included 11 sequences of the pst operon of non-Firmicutes that had Blast scores better than our threshold. As outgroups, we included 2 genomes of non-Firmicutes Thermotoga maritima (Thermotogae) and Pelobacter carbinolicus (ß-Proteobacteria), that gave the next best hitting scores below our threshold. Due to the high sequence variation of the pstS gene and the different number of copies of the other genes in the operon, the reconstruction was done with genes pstC, pstA, and pstB in a concatenated matrix. The phoU gene was excluded because it was missing in several Bacillus species. When both pstBA and pstBB were present, pstBB was used in the analysis as it is the ortholog of pstB; pstBA was not included, The pstBB gene was identified on the basis of genomic context, as it is the second gene coding for an ATP-binding protein in the operon. We compared the topology of the pst operon phylogeny with a phylogeny reconstructed from 14 concatenated amino acid sequences from genes from the core genome of Firmicutes. Those 14 genes were chosen from a set of genes already identified by Maughan [38] and Alcaraz et al. ([27]; GI from B. subtilis: 2632976, 2632269, 2632399, 2634021, 2636597, 16079910, 50812244, 50812227,
Four *Cyanobacteria, Chloroflexus aurantiacus* and *Thermotoga maritima* were used as outgroups (list of strain names and genome accession numbers in Table S1 of supplementary material). To establish a temporal frame of events, we dated the 14 gene phylogeny using a penalized likelihood method implemented by r8s [40]. The calibration of the tree was done using dates of geological events: the divergence of aerobic firmicutes was fixed at 2300 million years ago, a conservative date for the Great Oxidation Event [3, 37]. The divergence of CCB *Firmicutes* and their closest relatives was constrained to have a minimum age of 35 my, that corresponds to the uplift of the Sierra Madre Oriental that finally isolated CCB from the Gulf of Mexico [41].

All reconstructions were done using amino acid sequences aligned using MUSCLE [42] and a Maximum Likelihood approach, implemented by Raxml v7.0.4 [43]; (CIPRES portal: http://www.phylo.org/) with a LG substitution model chosen using ProtTest v 2.1 [44] with the Akaike Information Criterion, 4 substitution categories, and allowing Raxml to estimate the proportion of invariant sites and the gamma shape parameter. For both datasets 100 bootstrap replicates were performed.

### 2.2. PstS Protein Motifs and 3D Structure.

The main conserved motifs of the substrate-binding protein PstS were detected using the MEME suite ([45]; http://meme.sdsc.edu/meme/) using the default parameters and an alignment of the PstS amino acid sequence from all the *Firmicutes*, but including in this alignment only one sequence representative of the *Bacillus cereus* group.

The 3D structure of the PstS protein was modeled based on the 3-D structures of PstS from *Yersinia pestis* (PDB ID: 2x22) and PstS-1 from *Mycobacterium tuberculosis* (PDB ID: 1pc3; [46]). Only the PstS from *B. subtilis*, *B. coagulans*, *B. sp. m3-13*, and *B. sp. NRRL-14991* were modeled using the web-based module of MODELLER using the default settings (http://modbase.compbio.ucsf.edu/ModWeb20/html/modweb.html; [47]). Comparisons of 3D models were performed with TOPOFIT. This method only takes into account the geometric attributes of the proteins and not the sequence similarity, so it is able to find structure homology in highly variable proteins [48]. The quality of the models was evaluated with the r.m.s.d. value (root mean squared deviation) and the z-score (a measure of the energy separation between two protein folds). Images were prepared with CHIMERA (http://www.cgl.ucsf.edu/chimera).

### 3. Results

Using a psi-Blast search we were able to find orthologs of the *pst* operon in all members of *Firmicutes*, several *Cyanobacteria* and Archaea as well as in some *Bacteroidetes, Fusobacteria, Actinobacteria*, and *Planctomycetes*. In the search we observed that the gene architecture of the operon showed variation within and outside *Firmicutes* (Figure 1). Several taxa had a duplication in tandem of the gene *pstB*, as was the case for *Bacillus subtilis*, *Listeria monocytogenes*, and *Streptococcus pneumoniae*. *Cyanobacteria* generally lacked the gene *phoU* in the same operon, and it was missing in the *B. subtilis* group, and *Clostridium tetani*. *Synechococcus* sp. 7002 also lacked the *pstB* gene in the operon (Figure 1). Duplications of the *pstS* gene were found in the *Bacillus cereus* group, *Exiguobacterium spp.*, *Brevibacillus brevis*, *Bacillus sp. B14905*, *Enterococcus faecalis*, *Lactobacillus plantarum*, and *Geobacillus kaustophilus*, but the entire operon was only duplicated in *Streptococcus pneumoniae* and *Symbiobacterium thermophilus*. In the *B. cereus* group, an incomplete copy (*pstSCA*, lacking *pstB*) of the *pst* operon was found, similar to that of *B. subtilis*, thus it was not used for the concatenated phylogeny, but only for the PstS phylogeny (see Figure S1 in the supplementary material). All *Bacillus* from CCB and most marine *Bacillus* had just one copy of the *pstS* gene (the exception, *Bacillus sp. B14905*).

The phylogenetic reconstruction of the concatenated PstC, PstA, and PstB protein sequences showed two distinct and highly supported clades (Figure 2) that bear no relation to either the type of habitat or the phylogenetic relationships obtained with the amino acid sequences from housekeeping genes (Figure 3). Reconstructions made with each sequence independently, yielded the same basic topology, with minor differences in branch length (data not shown), so the phylogenetic signal was present in all three genes. We named one of the clades “*cereus-like*”, which consists of the *pstSCABU* operon structure (operon architecture A, in Figure 1), and it includes all members of the *B. cereus* group, most of *Bacillus* and *Staphylococcus, Exiguobacterium*, an anaerobic soil firmicute *Desulfobacterium hafniense*, and most noteworthy, several *Cyanobacteria* and *Archaea* (Figure 2). None of the members of that clade have the duplication of gene *pstB*, and only two taxa (*Desulfobacterium* and *Oceanobacillus*) lack the gene *phoU* in the operon (for the operon structure of all taxa in the dataset see Table S1 of Supplementary Material).

The other highly supported clade was named “*subtilis-like*” (operon architecture C, in Figure 1) and it included the members of the *B. subtilis* group, marine *Bacillus, Bacillus marisflavi* and its sister species *Bacillus sp. CH108* from CCB, *Listeria, Clostridium*, some host-associated firmicutes, and
Figure 2: Maximum likelihood phylogenetic reconstruction of the concatenated PstC, PstA, and PstB (PstBB) protein sequences encoded by the pst operon. Branch colors indicate the two divergent clades: subtilis-like and cereus-like. Tag colors indicate the type of habitat where each species is found. Bootstrap values above 70% are indicated with an asterisk. The phylogeny of the individual proteins has a very similar topology (data not shown).
an Actinobacteria (Collinsella intestinalis; Figure 2). The gene architecture of the operon in the members of this clade is more variable. The members of the genus Bacillus have the gene pstB duplicated and lack the phoU gene in the operon or entirely (Figure 1). Listeria, Enterococcus and Streptococcus also have the pstB gene duplicated but the gene phoU is in the operon, and although the pst operon in Clostridium has an architecture similar to that of B. cereus, it is very different at sequence level, as seen from the fact that these two are located in different clades (Figure 2).

The high variation at the amino acid sequence level observed for PstS is common for substrate-binding proteins of ABC transporters [18]. In our case, PstS had a shape parameter of 1.0559 and the proteins used for the alignment were for the Gamma distribution for site rates of 0.7002 (Mega 4; [49]).

The pst operon was not monophyletic for marine Bacillus, even though marine Bacillus are mostly monophyletic, as determined from house-keeping genes (Figure 3) and from other reconstructions (Figure 3; [27]). The main incongruence observed in the tree obtained from the amino acid sequence of the proteins encoded in the pst operon is the position of the B. subtilis group in a clade with the Listeria and Streptococcus sequences, instead of grouping with the rest of Bacillus taxa (Figure 2). This contrasts with the house-keeping genes phylogeny, were B. subtilis and its close relatives are found well within the Bacillus clade (Figure 3). Also, the B. marisflavi-B.sp CH108 clade, that groups with other marine Bacillus in the house-keeping genes phylogeny (Figure 3), appears as a sister group of B. subtilis and close taxa in the pst operon (Figure 2). Also, Bacillus sp. m3-13 from CCB, appears within the B. subtilis clade in the house-keeping genes phylogeny, but is sister to Bacillus sp. SG-1 from the Gulf of Mexico in the pst phylogeny. Another main topological incongruence of the pst phylogeny compared to the one done with housekeeping genes, is that of sister taxa Bacillus halodurans, and Bacillus clausii that are found in different clades: B. clausii is found in a clade with B. subtilis, B. marisflavi, and Bacillus sp. CH108 while B. halodurans forms a monophyletic clade with some marine Bacillus and in turn, is sister to the clade of Desulfotobacterium hafniense, an anaerobic species that is found in a basal position in the Firmicutes clade obtained from housekeeping genes (Figure 3).

Regarding the encountered motifs of protein PstS (Figure 4), we observed a marked difference between the PstS of the cereus-group and that from the subtilis-group. Motifs 4 and 5 are located in the same region of the protein but are markedly different, while motif 3 is found in both sets of sequences, but in the subtilis-like PstS it had a lower e-value (Figure 4). Despite the marked difference at the sequence level, the PstS proteins of B. subtilis, B. sp. m3-13, and B. sp. NRRL-14911, the 3D structures of the proteins showed similarity with geometry-based alignments (low r.m.s.d. and high z-scores; Figures 5(a)–5(c)), with the exception of B. coahuilensis that showed the worse fit values of the three comparisons (Figure 5). This could be a product of the initial 3D model based on a more distantly-related PstS, because the PstS from B. coahuilensis also had bad fitting rmsd and z-score values with the PstS of B. sp. NRRL-14911 despite having high sequence similarity (74% identity). Thus, the 3D model of the PstS of B. coahuilensis can still be improved.

Despite the structure similarities among the PstS of B. subtilis, B. sp. m3-13, and B. sp. NRRL-14911, the active site showed some striking differences in amino acid composition. B. subtilis and most of the taxa that grouped in the subtilis-like clade have an arginine as the first residue of the active site, just like Y. pestis and M. tuberculosis, while the firmicutes of the cereus-like clade have a proline in the same position (Figure 5(d)). Also, some members of the cereus-like clade, like B. sp. m3-13, B. halodurans, Desulfotobacterium spp., O. iheyensis, B. sp. SG-1, and the Staphylococcus group had also a histidine in the second residue of the active site, while all the other taxa had a serine (Figure 5(d)). In view of these changes in amino acid composition an additional codon-based Z-test of selection was made for the pstS gene with Mega 4 (Tables S2 and S3 of supplementary material; [49]). In all cases, dS (synonymous substitutions) was significantly higher than dN (non-synonymous substitutions), suggesting purifying selection.

4. Discussion

As expected, the pst operon was found in all Firmicutes. However, not so expected was the finding of two types of operons in these Gram positives. We describe them as subtilis-like and cereus-like operons, after the best known members of Bacillus. Even more interesting, these operons were not shared by descent in the monophyletic groups of Bacillus, neither they were operons related to the particular habitat of the strains. Both operons were very divergent from each other at the amino acid sequence level, suggesting independent parallel evolution.

The high divergence of the two types of pst-operons in Firmicutes and their incongruence with species phylogeny is most noteworthy. Contrary to what was expected, the pst operon of marine Bacillus is not monophyletic, even though marine and CCB Bacillus are resolved as a monophyletic group in the 14 housekeeping gene reconstruction. Therefore, we cannot argue for a common origin due to shared environmental conditions. The patchy distribution of either type of operon within the phylogeny of Firmicutes suggests horizontal gene transfer, especially considering closely related species with entirely different pst operons (B. clausii and B. halodurans, B. subtilis clade, and B. sp m3-13; Figure 2). Another possible explanation of these divergent operons, would be an ancient duplication. However, it is a hypothesis difficult to test, since only very few taxa have both kinds of operons or at least partial copies. At least in the case of the B. cereus group, the subtilis-like pstS copy is more closely related to that of Clostridium, while the pstS of B. subtilis is closely related to Listeria, suggesting an independent acquisition (see Figure S1 in supplementary material). If this partial operon or any of the extra copies of pstS found in different organisms are functional and
expressed, is still not known and would require experimental validation.

The relative conservation of gene architecture in Firmicutes as opposed to what is seen in Cyanobacteria suggests fewer rearrangements due to phages or some sort of constraint for the transcription and/or regulation of the pst operon that has kept the gene architecture fairly constant since the divergence of Cyanobacteria and Firmicutes around

---

**Figure 3:** Maximum likelihood phylogeny of Firmicutes based on the concatenated amino acid sequence of 14 housekeeping genes and dated with a penalized likelihood method. The branch colors indicate the type of operon present in each taxon. Tag colors refer to the type of habitat. Clade I corresponds to aerobic Firmicutes and clade II includes CCB and marine Bacillus. Clade I had a fixed age of 2300 my and clades III and IV had a fixed minimum age of 35 my. Bootstrap values above 70% are denoted with an asterisk. Clades with branch lengths of 0 were collapsed (D. hafniense DCB-2-D. hafniense Y51 and G. thermodenitrificans NG80-2-G. sp. G11MC16).
Figure 4: Conserved motifs of the PstS protein for both cereus-like and subtilis-like operons. Red stars on top of residues on motif 2 indicate the binding site of phosphate. Motifs 4 and 5 are aligned (black lines) to show homologous amino acid positions. The height of the blocks is proportional to the e-value of each motif. Motif 3 for subtilis-like PstS has an e-value < 10^{-10}.

Figure 5: Comparison of 3D structures with TOPOFIT of PstS from B. subtilis (red), with (a) B. sp. NRRL-14911 (cyan; r.m.s.d. = 1.02, z-score = 39.89), (b) Bacillus sp. m3-13 (green; r.m.s.d. = 1.53, z-score = 20.24) and (c) B. coahuilensis m4-4 (yellow; r.m.s.d. = 1.49, z-score = 9.65). Residues involved in phosphate binding are highlighted in purple. (d) shows an alignment of the active site of PstS and the different amino acids involved in phosphate binding are highlighted in gray.

3 billion years ago ([37]; this study). Even though we observed a constant operon architecture, we also observed two taxa with both types of operon (S. thermophilum and S. pneumoniae), one early divergent and the other more derived (Figure 3), however, the protein sequence of either of them is very divergent from the other taxa.

Even if the general architecture of the operon suggests less recombination than in the Cyanobacteria lineages, in Firmicutes several other taxa had more than one copy of the pstS gene or presented an incomplete extra copy of the operon. A phylogenetic analysis of PstS (see Figure S1 in Supplementary Material available online at doi:10.406/2011/781642)
suggests that these extra copies of the *pstS* gene were acquired later by HGT rather than acquired by duplications, since different copies of the gene belonging to the same organism are found in different places in the phylogeny. Even though the high sequence variation present in *pstS* at both nucleotide and protein levels make it hard to obtain a well-supported phylogenetic reconstruction. From our analysis it is evident that the *pstS* gene is evolving at a faster rate than the rest of the genes of the operon, where the purifying selection maintains the folding of the protein instead of the amino acid sequence [19]. Since we found no significant positive selection for *pstS*, the high level of sequence divergence could be due to the accumulation of repeated mutations after the ancient split between the *cereus*-like and *subtilis*-like operon [50].

Interestingly, despite the marked sequence divergence of the genes of the *pst* operon of firmicutes and particularly the *pstS* gene (Figure 4), the 3D structures of PstS from the *subtilis*-like and *cereus*-like operons were surprisingly similar (Figures 5(a)–5(c)). However, it is still to be determined how the changes of particular amino acids in the phosphate binding site (Figure 5(d)) would affect the formation of the hydrogen bonds necessary for phosphate uptake [51]. In particular, it is possible that the presence of a proline (P) instead of serine in the active site of the *cereus*-like PstS (Figure 5(d)) could have some relevance in the discrimination between the mono- and dibasic forms of phosphate, since this amino acid only acts as a hydrogen bond acceptor but not as a donor [46]. The difference of affinity to phosphate and the potential selectivity for either of the phosphate species should be investigated experimentally, as the protein backbone is also involved in the hydrogen bond formation and not only the side chains of the residues [46, 51]. The amino acid changes in the active site of the PstS protein can have an effect on the efficiency of the acceptor under different pH conditions [46, 51], and maybe some of those substitutions could be related to habitat. Nevertheless, this idea is not sustained within the *cereus*-like clade, were it can be observed that the lineage with *B. sp m3-13* (Figure 2), the *Staphylococcus* group and the clade with *Anoxybacillus flavithermus* as well as two species of *Geobacillus* have a histidine instead of a serine in the second residue of the active site (Figure 5(d)), and all those taxa actually live in environments with a wide range of pH conditions, as is the case in the rest of *Firmicutes*.

It has been previously noted that protein structure is fairly conserved in nature, and that proteins with only 8% similarity at sequence level can have a much higher similarity in their structural features [52, 53]. In our case, some of the PstS proteins had a sequence similarity as low as 17% when comparing those from the *subtilis*- or *cereus*-like operons, yet their structure was fairly conserved (Figures 5(a)–5(c)). This could be due to natural selection acting on protein structure, thus allowing for changes in amino acids that would not alter the basic features of the protein [54]. The fact that these similar protein structures occur on lineages that have such deep divergences such as *Cyanobacteria* and *Firmicutes* (ca. 3 billion years ago), favors the idea of parallel evolution from a common ancestor [37, 54, 55]. This very ancient divergence produced one *pst* operon mostly found in anaerobic *Firmicutes* and some pathogenic groups (*Listeria*), while a *pst* operon similar to that of *Cyanobacteria* is found mostly in oligotrophic *Firmicutes* and in some other pathogenic groups (*Staphylococcus*), with various cases of ancient HGT between either group (i.e., *B. halodurans*, *B. clausii*, and *D. hafniense*).

Therefore, we should reconsider the environmental constrain hypothesis: *Bacillus* sp CH108 from the Churince water system in the present shares similar environmental conditions to other *Bacillus* from CCB, but has a *subtilis*-like *pst* operon instead of the *cereus*-like operon that is common to other CCB species. The sister species of *B. sp* CH108, *B. marisflavi* from the Yellow Sea of Korea [56], also has a *subtilis*-like operon. This suggests that acquisition of the operon predates the divergence of the two taxa (~92 my ago; Figure 3), which in turn is older than the last time CCB was connected to the ocean, ca. 45 my ago. This may leads to the idea that this particular arrangement is not specific to the actual oligotrophic conditions in Cuatro Cienegas [41, 57] but it is an adaptation to an ancient sea.

5. Conclusions

The *pst* operon in *Firmicutes* showed a very high sequence divergence that is not correlated to either phylogenetic relationships among taxa, the type of habitat, or the phosphorus availability where these organisms currently live. Thus, it is likely that the current distribution the *pst* operon was determined by a very early divergence and repeated events of HGT of the phosphate transporter genes followed by parallel evolution that lead to similar 3D structures. Unlike what was observed in *Cyanobacteria*, most *Firmicutes* only have one or a couple of copies of the PstS protein, so it is crucial for phosphate uptake that both function and affinity are conserved in the substrate binding protein.

Acknowledgments

This project was funded by CONACYT-SEP Grant no. 57507 and CONACyT-Semarnat 2006-C01-23459 awarded to V. Souza and a CINVESTAV Multidisciplinary awarded to G. Olmedo. V. Souza and L.E. Eguiarte worked in this paper during a sabbatical leave supported by DGAPA and UC-Mexus, respectively. Thanks are due to L. D. Alcaraz, E. Rebollar, F. García-Oliva, D. Ortega-Del Vecchyo, I. Hernandez, V. Lopez, G. Moreno-Hagelsieb, B.S. Gaut, M. Tenaillon, O. Tenaillon, A. Vázquez-Lobo, and D. Piñero for all their insights and comments.

References

[1] S. G. Tetu, B. Brahamsha, D. A. Johnson et al., “Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus sp. WH8102*,” ISME Journal, vol. 3, no. 7, pp. 835–849, 2009.
[2] P. G. Falkowski, T. Fenchel, and E. F. Delong, “The microbial engines that drive earth's biogeochemical cycles,” *Science*, vol. 320, no. 5879, pp. 1034–1039, 2008.
[3] D. Papineau, “Global biogeochemical changes at both ends of the Proterozoic: insights from phosphorites,” *Astrobiology*, vol. 10, no. 2, pp. 165–181, 2010.
Y. Qi, Y. Kobayashi, and F. M. Hulett, “The pst operon of Bacillus subtilis is dependent on phosphate concentration and pH,” *Journal of the American Chemical Society*, vol. 130, no. 7, pp. 2139–2142, 2008.

A. C. Martiny, Y. Huang, and W. Li, “Occurrence of phosphate acquisition genes in Prochlorococcus cells from different ocean regions,” *Environmental Microbiology*, vol. 11, no. 6, pp. 1340–1347, 2009.

R. Feingersch, M. T. Suzuki, M. Shmoish et al., “Microbial community genomics in eastern Mediterranean Sea surface waters,” *ISME Journal*, vol. 4, no. 1, pp. 78–87, 2010.

B. A. S. Van Mooy, H. F. Fredricks, B. E. Pedler et al., “Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity,” *Nature*, vol. 458, no. 7234, pp. 69–72, 2009.

J. J. Elser, J. H. Schampel, F. Garcia-Pichel et al., “Effects of phosphorus enrichment and grazing snails on modern stromatolitic microbial communities,” *Freshwater Biology*, vol. 50, no. 11, pp. 1808–1825, 2005.

V. Souza, L. E. Eguiarte, J. Siefert, and J. J. Elser, “Microbial endemism: does phosphorus limitation enhance speciation?” *Nature Reviews Microbiology*, vol. 6, no. 7, pp. 559–564, 2008.

V. Souza, L. Espinoza-Azuar, A. E. Escalante et al., “An endangered oasis of aquatic microbial biodiversity in the Chihuahuan desert,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 17, pp. 6565–6570, 2006.

C. Desnues, B. Rodriguez-Brito, S. Rayhawk et al., “Biodiversity and biogeography of phages in modern stromatolites and thrombolites,” *Nature*, vol. 452, no. 7185, pp. 340–343, 2008.

L. D. Alcaraz, G. Moreno-Hagelsieb, L. E. Eguiarte, V. Souza, L. Herrera-Estrella, and G. Olmedo, “Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics,” *BMC Genomics*, vol. 11, no. 1, article no. 332, 2010.

M. Breitbart, A. Hoare, A. Nitti et al., “Metagenomic and stable isotopic analyses of modern freshwater microbiolalites in Cuatro Ciénagas, Mexico,” *Environmental Microbiology*, vol. 11, no. 1, pp. 16–34, 2009.

B. M. Winsborough, E. Theriot, and D. B. Czarnecki, *Diatoms on a Continental Island: Lazeres Species, Marine Disjuncts and other Endemic Diatoms of the Cuatro Ciénagas basin, Coahuila, México*, University of Texas, Austin Tex, USA, 2008.

W. L. Minkley, “Cuatro Ciénegas fishes: research reviewed and a local test of diversity versus habitat size,” *Journal of the Arizona-Nevada Academy of Science*, vol. 19, pp. 13–21, 1984.

E. W. Carson and T. E. Dowling, “Influence of hydrogeomorphic history and hybridization on the distribution of genetic variation in the pupfishes *Cyprinodon atrorus* and *C. bifasciatus*,” *Molecular Ecology*, vol. 15, no. 3, pp. 667–679, 2006.

S. G. Johnson, “Age, phylogeography and population structure of the microendemic banded spring snail, *Mexypyrus chiricuanus*,” *Molecular Ecology*, vol. 14, no. 8, pp. 2299–2311, 2005.

M. Tobler and E. W. Carson, “Environmental variation, hybridization, and phenotypic diversification in Cuatro Ciénegas pupfishes,” *Journal of Evolutionary Biology*, vol. 23, no. 7, pp. 1475–1489, 2010.

R. Cerritos, P. Vinuesa, L. E. Eguiarte et al., “Bacillus coahuilensis sp. nov., a moderately halophilic species from a desiccation lagoon in the Cuatro Ciénegas Valley in Coahuila, Mexico,” *International Journal of Systematic and Evolutionary Microbiology*, vol. 58, no. 4, pp. 919–923, 2008.

R. Cerritos, L. E. Eguiarte, M. Avitia et al., “Diversity of cultivable thermo-resistant aquatic bacteria along an environmental gradient in Cuatro Ciénegas, Coahuila, Mexico,” *Antonio Van Leeuwenhoek*, vol. 99, no. 2, pp. 303–318, 2010.
[36] A. E. Escalante, L. E. Eguiarte, L. Espinosa-Asuar, L. J. Forney, A. M. Noguez, and V. Souza Saldivar, “Diversity of aquatic prokaryotic communities in the Cuatro Ciéneegas basin,” FEMS Microbiology Ecology, vol. 65, no. 1, pp. 50–60, 2008.

[37] F. U. Battistuzzi, A. Feijao, and S. B. Hedges, “A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land,” BMC Evolutionary Biology, vol. 4, article no. 44, 2004.

[38] H. Maughan, “Rates of molecular evolution in bacteria are relatively constant despite spore dormancy,” Evolution, vol. 61, no. 2, pp. 280–288, 2007.

[39] F. Kunst, N. Ogasawara, I. Moszer et al., “The complete genome sequence of the gram-positive bacterium Bacillus subtilis,” Nature, vol. 390, no. 6657, pp. 249–256, 1997.

[40] M. J. Sanderson, “Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach,” Molecular Biology and Evolution, vol. 19, no. 1, pp. 101–109, 2002.

[41] I. Ferrusquía-Villafranca, “Geología de México: una sinopsis,” in Diversidad biológica de México: orígenes y distribución, T. P. Ramamoorthy et al., Ed., Instituto de Biología UNAM, México D.F., 1998.

[42] R. C. Edgar, “MUSCLE: a multiple sequence alignment method with reduced time and space complexity,” BMC Bioinformatics, vol. 5, article no. 113, 2004.

[43] A. Stamatakis, P. Hoover, and J. Rougemont, “A rapid bootstrap algorithm for the RAxML web servers,” Systematic Biology, vol. 57, no. 5, pp. 758–771, 2008.

[44] F. Abascal, R. Zardoya, and D. Posada, “ProtTest: selection of best-fit models of protein evolution,” Bioinformatics, vol. 21, no. 9, pp. 2104–2105, 2005.

[45] T. L. Bailey and C. Elkan, “Fitting a mixture model by expectation maximization to discover motifs in biopolymers,” in Proceedings of the 2nd International Conference on Intelligent Systems for Molecular Biology, vol. 2, pp. 28–36, 1994.

[46] M. Tanabe, O. Mirza, T. Bertrand et al., “Structures of OppA and PstS from Yersinia pestis indicate variability of interactions with transmembrane domains,” Acta Crystallographica Section D, vol. 63, no. 11, pp. 1185–1193, 2007.

[47] N. Eswar, B. Webb, M. A. Marti-Renom et al., “Comparative protein structure modeling using Modeller,” Current Protocols in Bioinformatics, vol. 5, pp. 5.6.1–5.6.30, 2006.

[48] V. A. Ilyin, A. Abyzov, and C. M. Leslin, “Structural alignment of proteins by a novel TOPOFIT method, as a superimposition of common volumes at a topomax point,” Protein Science, vol. 13, no. 7, pp. 1865–1874, 2004.

[49] K. Tamura, J. Dudley, M. Nei, and S. Kumar, “MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0,” Molecular Biology and Evolution, vol. 24, no. 8, pp. 1596–1599, 2007.

[50] A. L. Hughes, “Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level,” Heredity, vol. 99, no. 4, pp. 364–373, 2007.

[51] H. Luecke and F. A. Quiocho, “High specificity of a phosphate transport protein determined by hydrogen bonds,” Nature, vol. 347, no. 6291, pp. 402–406, 1990.

[52] E. V. Koonin, Y. I. Wolf, and G. P. Karev, “The structure of the protein universe and genome evolution,” Nature, vol. 420, no. 6912, pp. 218–223, 2002.

[53] R. A. Goldstein, “The structure of protein evolution and the evolution of protein structure,” Current Opinion in Structural Biology, vol. 18, no. 2, pp. 170–177, 2008.