

Phylogenetic analysis of substrate-binding subunits in an osmoprotectant system

T.J.D. Coutinho, N.F. Rodrigues and S.T. Farias

Laboratório de Genética Evolutiva Paulo Leminsk,
Departamento de Biologia Molecular, Universidade Federal da Paraíba,
João Pessoa, PB, Brasil

Corresponding authors: T.J.D. Coutinho / S.T. Farias
E-mail: coutinho8178@hotmail.com / stfarias@yahoo.com.br

Genet. Mol. Res. 12 (4): 5743-5757 (2013)
Received December 3, 2012
Accepted March 22, 2013
Published November 22, 2013
DOI <http://dx.doi.org/10.4238/2013.November.22.1>

ABSTRACT. Substrate-binding subunits are important components of the solute importation system, known as the osmoprotectant system, which consists of a membrane protein belonging to the ABC superfamily. These molecules recognize specific substrates that have different physiological roles in prokaryotes, i.e., roles that contribute to the survival of these organisms in environments with high concentrations of salt. Using the MEGA software, this study performed a phylogenetic analysis of 431 nucleotide sequences of these subunits, orthologous to each other, collected from the <http://www.genome.jp/kegg/> database. This analysis allowed phylogenetic trees to be generated, clearly demonstrating that there was horizontal transfer of some genes through sharing by different organisms. Furthermore, two probable ancestral sequences were generated that showed homology with permeases that transport choline, glycine betaine, and carnitine, which are trimethylamines currently present in various prokaryotes. Therefore, this system probably arose in prokaryotic organisms with the basic function of capturing nutrients, and by performing this basal function and being shared with other organisms, it was fixed in the genome. However, because of prokaryote habitat diversification, this

system contributed decisively to the adaptation of these organisms to different environments, especially environments that had a high salt concentration, thus acting as an osmoprotection system, which is how they are currently categorized.

Key words: Transportation; Phylogenetic analysis; Glycine betaine; Osmoadaptation

INTRODUCTION

Water is the most important molecule related to life, and the capacity of an organism to survive may depend on its ability to cope with internal variations of water content. Water loss to the environment due to evaporation, excretion of waste, and osmosis is a common threat in various environments. With respect to water stress, organisms were traditionally divided into two groups: osmoconformers, which usually use compatible solutes to maintain intracellular osmotic pressure equal to that in the external environment, and osmoregulators, which use ion transport to homeostatically regulate intracellular osmotic pressure (Yancey, 2005).

The types of organic molecules characterized as compatible solutes include polyols and derivatives, sugars and derivatives, amino acids and derivatives, betaines, ecotins, and a few peptides. Bacteria and eukaryotes normally accumulate neutral solutes, while archaea tend to modify many of the solutes accumulated by bacteria and eukaryotes into negatively charged molecules. The fact that these molecules do not inhibit cellular functions, even though they may act by modulating the activity of individual enzymes, is why they were designated as compatible solutes and why their intracellular accumulation helps maintain cell turgidity, pressure, and volume (Roberts, 2005).

Compatible solutes may be introduced into the natural environment following the death of an organism or by efflux processes that allow these molecules to become accessible to other organisms and be used for osmoadaptation or as a carbon source. In certain cases, capture systems are essential to microorganisms that lack the biochemical machinery for synthesis of these compounds. Because of this, it is common to find transport systems (influx) that have high affinity for these molecules and allow them to be rapidly imported and their intracellular levels be readily restored, if necessary, or to respond quickly to changes in the salinity of the medium (Empadinhas and da Costa, 2008).

Environments in which frequent variations in water content occur as, for example, in the most superficial layers of soil, create a physiological challenge for microorganisms, and the survival of microorganisms in such environments requires adaptive mechanisms. One such mechanism is a system for importing compatible solutes called osmoprotectant uptake (Opu), which consists of five transporters, of which three belong to the ABC superfamily of proteins (ATP-binding cassette): OpuA (transports glycine betaine and proline betaine), OpuB (transports only choline), and OpuC (transports several compatible solutes such as glycine betaine, proline betaine, carnitine, choline, γ -butyrobetaine, choline-*O*-sulfate, and ectoine). These three transporters use energy released by ATP hydrolysis to import solutes (Du et al., 2011). The other transporters are OpuD, a transporter of the BCCT family (betaine-choline-carnitine transporter) that uses the energy of the driving

force of sodium symport transport for importing solutes (Ziegler et al., 2010), and OpuE, a transporter from the sodium/solute symport family that is specific for importing proline (Moses et al., 2012).

ABC transporters are integral membrane proteins found in all domains of life that have, as a rule, two transmembrane subunits or domains (TMDs) that form the translocation pathway, and two cytoplasmic nucleotide-binding subunits or domains (NBDs) that hydrolyze ATP. These transporters diverged very early in evolution, creating two categories related to the transport of substrate: exporters and importers. In the case of exporters, the TMDs and the NBDs are fused, while in the case of importers, besides the TMDs and NBDs being separate subunits, there is a third subunit that binds with high affinity to the substrates in order to deliver them to the TMDs (Oldham et al., 2008).

Of the subunits forming solute importers, the substrate-binding subunit (SBP, substrate-binding protein) is extremely important for prokaryotes and is widespread in their genomes (Thomas, 2010). In addition, it exhibits interesting structural diversity: in Gram-negative bacteria it is located in the periplasmic space between the outer and inner membranes; in Gram-positive bacteria, it can be bound to the lipoproteins of the extracellular face of the plasma membrane or to the TMD of the ABC transporter (Davidson et al., 2008), or it can be anchored in the lipids of the external face of the plasma membrane (Biemans-Oldehinkel et al., 2006). In archaea, on the other hand, they bind to lipoproteins of the plasma membrane or components of a transport system known as the bindosome.

Owing to these multiple variations in SBPs, their degree of specificity for substrates, and their distribution among the three domains of life, we aimed to construct a phylogenetic profile of these subunits, which will give us a better understanding of what happened over time with these molecules and how they were and are being selected by the environment.

MATERIAL AND METHODS

Data collection

We selected 431 SBP gene sequences from the online database <http://www.genome.jp/kegg/>. These sequences correspond to SBPs belonging to the trimethylamine transporters (glycine betaine, proline betaine, carnitine, choline, and proline) of prokaryotic organisms (bacteria, cyanobacteria, and archaea) present in the most diverse environments.

Bacillus subtilis 168 was used as the reference organism, and from its SBP gene sequence, a search for orthologous sequences was performed in the same database (Kanehisa and Goto, 2000; Kanehisa et al., 2012).

Sequence alignment and nucleotide substitution model

Sequence alignment was done using ClustalW, included in the Molecular Evolutionary Genetics Analysis software (MEGA 5.05) (Thompson et al., 1994; Tamura et al., 2011) and was used to test which evolutionary model was consistent with the data.

Among the models offered by the software, the best suited for phylogenetic reconstruction is general time reversible (GTR) with gamma distribution with 2 parameters.

Phylogenetic analysis

Phylogenetic analysis was performed with the MEGA 5.5 software (Tamura et al., 2011), using the maximum likelihood statistical method with 1000 bootstrap replications. All nucleotide substitution sites were used, the heuristic method was the nearest-neighbor-interchange (DasGupta et al., 2000), and all codons (1st + 2nd + 3rd + non-coding) were used. Additionally, 16S ribosomal RNA gene sequences of the organisms were used as the control.

Selection test

A Z-test of selection, with 1000 bootstrap replications, was conducted based on codons, using MEGA (Tamura et al., 2011), where the model was the Nei-Gojobori method (proportion) (Saitou and Nei, 1987) and the analysis involved synonymous- and non-synonymous-type substitutions.

The null hypothesis tested was $H_0: d_N = d_S$, where d_N is the number of non-synonymous substitutions per non-synonymous site, and d_S is the number of synonymous substitutions per synonymous site. For this hypothesis, three alternative hypotheses were tested: a) $d_N \neq d_S$ (test for neutrality), b) $d_N > d_S$ (positive selection), and c) $d_N < d_S$ (purifying selection). These are only valid if the P value is less than 0.05.

Ancestral sequences

Hypothetical ancestral sequences were generated from MEGA (Tamura et al., 2011) using the maximum likelihood statistical method and the GTR nucleotide substitution model with gamma distribution with 2 parameters (Tamura et al., 2011). Generated sequences were analyzed to find some probable similarity to an actual subunit, using the BLASTN online tool (for nucleotides) and BLASTP (for amino acids) found on the site <http://www.ncbi.nlm.nih.gov/>. Besides this analysis, three-dimensional models were also generated for comparison with current models, using the <http://swissmodel.expasy.org> site and standard parameters.

Analysis of codon usage

SBP sequences were analyzed and organism genomes selected with regard to codon usage, using the <http://www.kazusa.or.jp/codon/> database.

RESULTS

Phylogenetic analysis

The tree generated by sequence analysis ([Supplementary material](#)) shows the relationships between SBPs. At the end of each branch, the reference organisms for each sequence are shown and, for some, the annotation of the gene in the Kyoto Encyclopedia of Genes and Genomes (KEGG) and the type of solute for each SBP are shown (Figure 1).

In the analysis of this tree, two types of organism grouping were observed: a grouping of phylogenetically related organisms, which would be the most anticipated, and a grouping

of phylogenetically unrelated organisms. The presence of the second grouping suggests that horizontal gene transfer had taken place. An example of this multiple form of grouping occurs with a non-virulent soil bacterium from the family Rhizobiaceae, known as *Agrobacterium radiobacter* K84, which has a main chromosome (chromosome 1) and a secondary chromosome, chromosome 2, which can also be regarded as a mega plasmid (Slater et al., 2009).

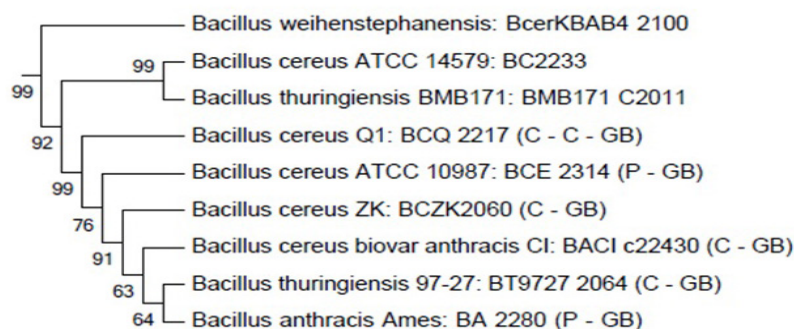


Figure 1. Organisms, gene annotation of the substrate-binding protein in KEGG and type of solute transported (P - GB: proline and glycine betaine; C: choline; C - GB: choline and glycine betaine; C - C - GB: carnitine, choline and glycine betaine; When the type of solute is not specified, it is because only glycine betaine is transported).

Chromosome 1 has the gene Arad_0755 (NCBI gene ID: 7368963), which grouped with phylogenetically related organisms (Figure 2). On the other hand, on chromosome 2, there are the genes Arad_8392 and Arad_8402, (NCBI 7366822 and 7366831, respectively), which clustered with phylogenetically unrelated organisms (Figures 3 and 4), indicating that there was probably horizontal transfer of chromosome 2.

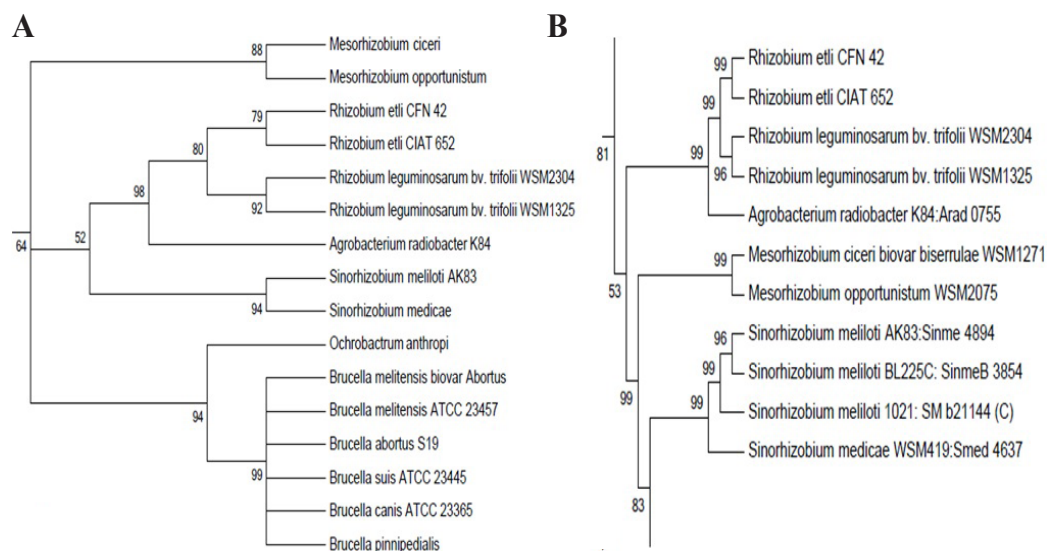


Figure 2. A. Phylogeny based on RNAr 16S. **B.** Grouping based on sharing gene Arad_0755 present on chromosome 1 of *Agrobacterium radiobacter* K84.

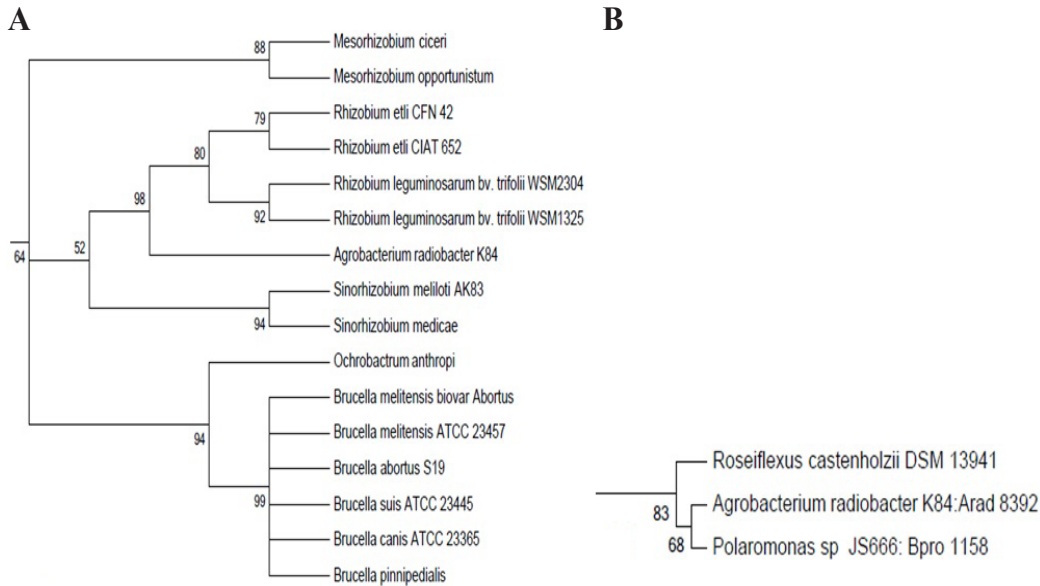


Figure 3. A. Phylogeny based on RNAr 16S. **B.** Grouping based on sharing gene Arad_8392 present on chromosome 2 of *Agrobacterium radiobacter* K84.

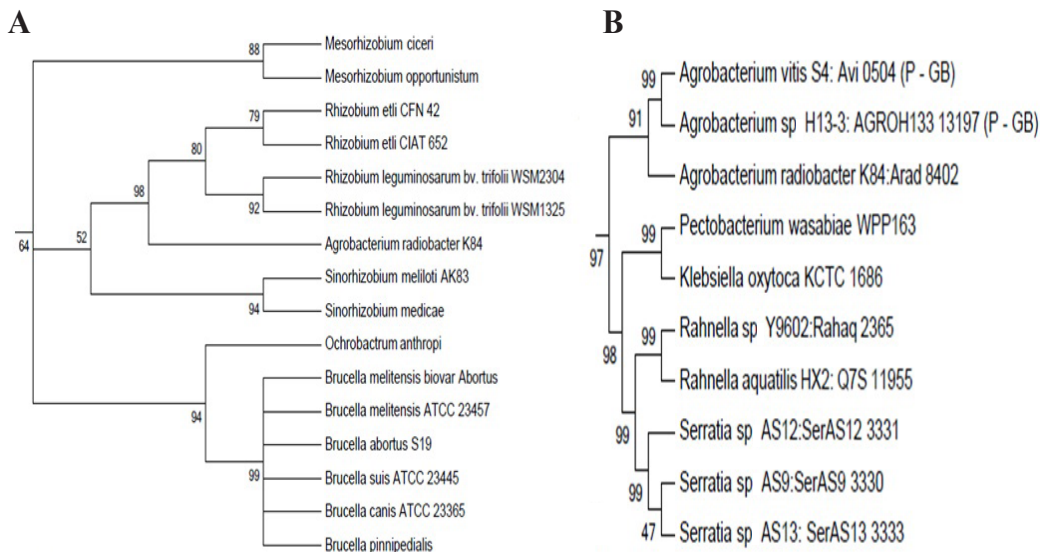


Figure 4. A. Phylogeny based on RNAr 16S. **B.** Grouping based on sharing gene Arad_8402 present on chromosome 2 of *Agrobacterium radiobacter* K84.

Selection test

None of the P values were significant, suggesting that any type of difference or correlation seen in this data set was due to chance, although it was perceived that there is a tendency for purifying selection (Table 1).

Table 1. Codon-based Z-test of selection.

Alternative hypothesis	P values
$d_N \neq d_S$ (test for neutrality)	0.271
$d_N > d_S$ (positive selection)	1.000
$d_N < d_S$ (purifying selection)	0.129

Ancestral sequences

Two ancestral sequences were generated and were derived from the first cladogenesis; these sequences were probably the source of all current SBP sequences. The ancestral sequences were identified as Ancestral_1 and Ancestral_2 ([Supplementary material](#)). Data analysis using the BLASTN tool showed that Ancestral_1 was similar to the periplasmic subunit of *Desulfosporosinus orientis* DSM 765 (Table 2) that binds to glycine betaine and choline, whereas the amino acid sequence showed similarity to trimethylamine transporters, especially permeases (Table 3), with the characteristic domains of choline transporters, that is, OpuBB and OpuBC, conserved (Marchler-Bauer et al., 2011; Pittelkow et al., 2011) (Figure 5). As for Ancestral_2, the nucleotide sequence showed no similarity with any transporter, but the amino acid sequence showed similarity to permeases (Table 4) and conserved domains (Figure 6), similar to Ancestral_1.

Table 2. BLASTN of Ancestral_1-significant sequence alignments.

Accession	Description	Max. score	Total score	Query coverage	E value	Max. ident.
CP003108.1	<i>Desulfosporosinus orientis</i> DSM 765 complete genome	152	152	18%	1e-32	75%

Table 3. BLASTP of Ancestral_1-significant sequence alignments.

Accession	Description	Max. score	Total score	Query coverage	E value	Max. ident.
YP_003967851.1	Binding-protein-dependent transport system inner membrane	96.7	96.7	28%	2e-19	31%
YP_001425237.1	Glycine betaine transport system permease [<i>Coxiella burnetii</i> Dugway 5J108-111]	98.6	98.6	30%	1e-18	35%
ZP_01946271.2	ABC transporter, quaternary amine uptake transporter (QAT)	98.6	98.6	30%	1e-18	35%
YP_001596136.1	Quaternary amine ABC transporter substrate-binding protein [<i>Coxiella burnetii</i> RSA 331]	95.5	95.5	30%	9e-18	35%
YP_002304237.1	Glycine betaine transport system permease protein [<i>Coxiella burnetii</i> CbuG_Q212]	95.1	95.1	30%	1e-17	35%

Three-dimensional models

Using the Swiss-Model, two possible models were generated for Ancestral_1, of which only model 1 showed similarity to the ABC transporters of compatible solutes, specifically with choline-glycine betaine-carnitine transporters (Figure 7). Five possible models were generated for Ancestral_2, of which only number 5 was compatible with ABC transporters of compatible solutes that, like Ancestral_2, showed similarity to choline-glycine betaine-carnitine transporters (Figure 8).

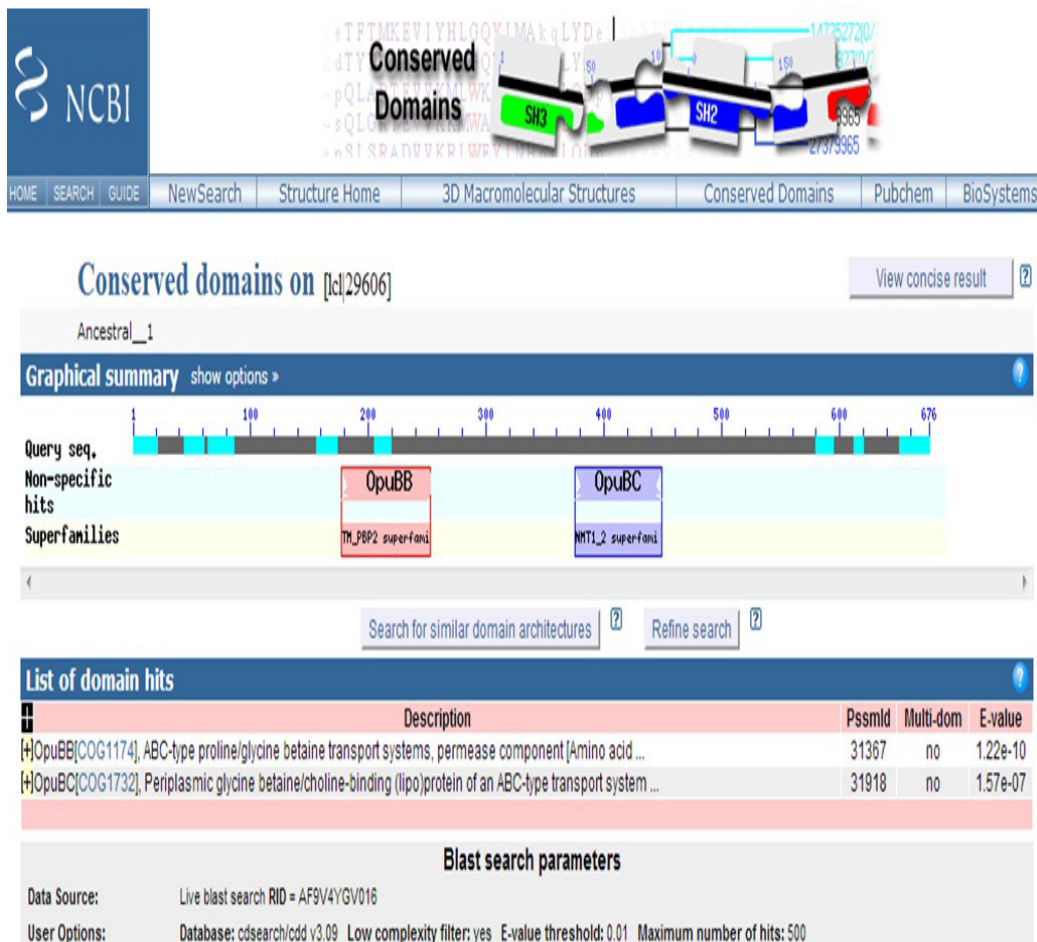


Figure 5. Conserved domains in Ancestral_1.

Table 4. BLASTP of Ancestral_2-significant sequence alignments.

Accession	Description	Max. score	Total score	Query coverage	E value	Max. ident.
YP_003967851.1	Binding-protein-dependent transport system inner membrane protein [<i>Ilyobacter polytropus</i> DSM 2926]	108	108	30%	3e-23	31%
ZP_01946271.2	ABC transporter, quaternary transporter (QAT) amine uptake	105	105	30%	8e-21	34%
YP_001425237.1	Glycine betaine transport system permease [<i>Coxiella burnetii</i> Dugway 5J108-111]	104	104	30%	1e-20	34%
YP_002507602.1	Glycine betaine ABC transporter substrate-binding protein [<i>Clostridium cellulolyticum</i> H10]	103	103	59%	3e-20	28%
ZP_08192210.1	Substrate-binding region of ABC-type glycine betaine transport system [<i>Clostridium papyrosolvans</i> DSM 2782]	103	103	34%	3e-20	34%

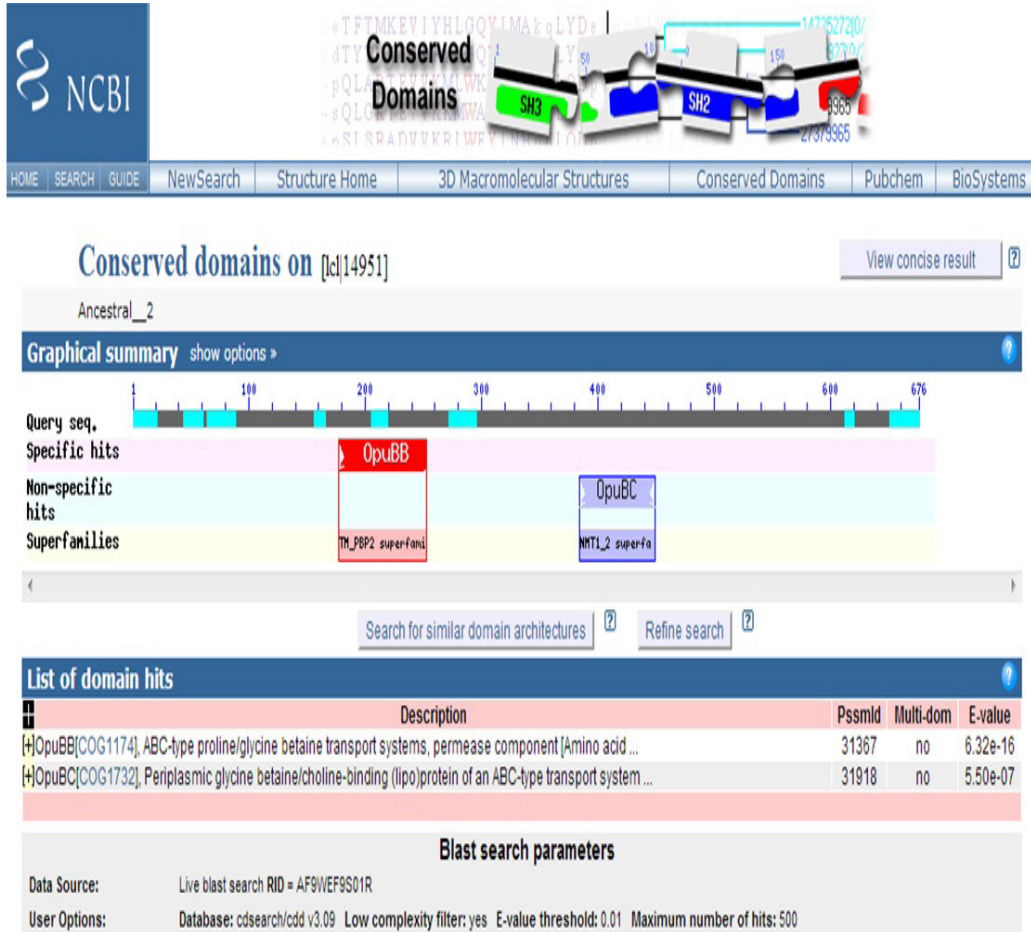


Figure 6. Conserved domains in Ancestral_2.

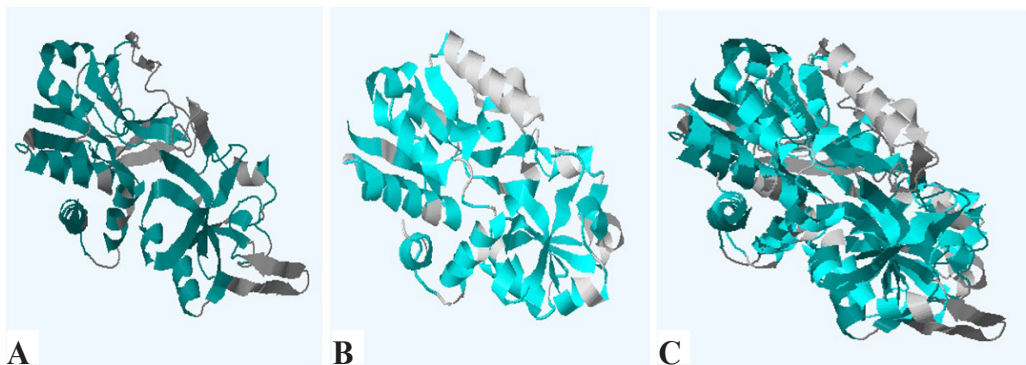


Figure 7. **A.** Three-dimensional model of Ancestral_1. **B.** Crystal structure of glycine betaine/carnitine/choline ABC transporter. **C.** Overlay of Ancestral_1 with crystal structure of glycine betaine/carnitine/choline ABC transporter.

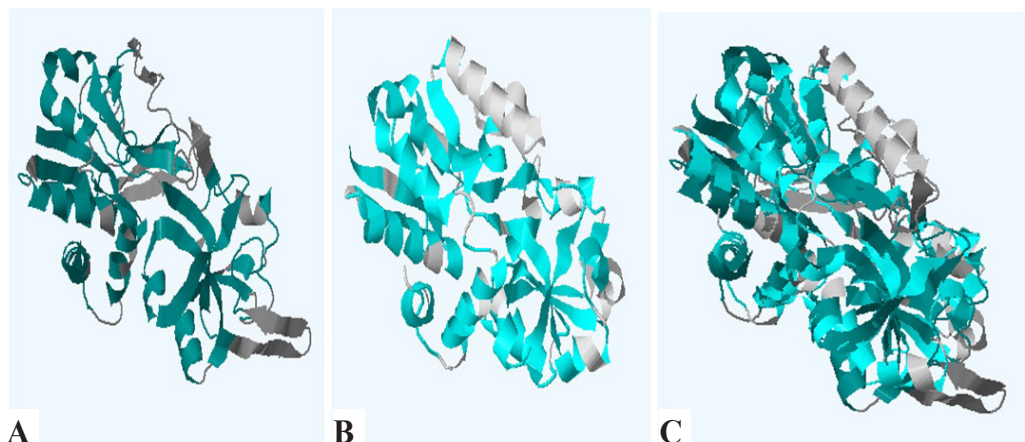


Figure 8. **A.** Three-dimensional model of Ancestral_2. **B.** Structure of the substrate-binding protein provides insights into the multiple compatible solute-binding specificities of *Bacillus subtilis* ABC transporter OpuC. **C.** Overlay of Ancestral_2 with the structure of the substrate-binding protein that provides insights into the multiple compatible solute-binding specificities of *B. subtilis* ABC transporter OpuC.

Codon usage

According to this analysis, codon usage of SBPs corresponds to that of the genomes of the organisms in which they are inserted, suggesting that this system is important for cell physiology, and therefore, its sequence has changed little over time (Table 5).

DISCUSSION

From the analysis of the SBP sequence tree, it was noted that the trimethylamine and proline import system, formed by members of the ABC family of proteins, is well distributed in the prokaryotic world, i.e., these subunits are present in organisms that live in very different environments where they are not necessarily subjected to adverse salt conditions. This analysis allows us to infer other possible physiological interactions of the osmoprotectant system, since it is widely distributed in different environments and could therefore be involved with other functions besides being responsible for the osmotic response.

When subjected to osmotic stress, prokaryotic organisms activate a particular transcription factor, σ^B , which is responsible for the physiological response to stress. This factor triggers the transcription of approximately 150 genes, some directly and others indirectly. Under such conditions, the transcription levels of osmoprotectant uptake genes (*opuA-opuE*) are raised, rapidly increasing protein synthesis of this system in the first moments after osmotic stress is encountered (Hahne et al., 2010). However, there is no consensus on what is the primary activation of the osmoprotectant system by σ^B (Höper et al., 2005). There are studies indicating that the osmoprotectant system, or at least some of its components, is activated by other transcription factors (Petersohn et al., 2001; Hecker et al., 2007).

Another factor that contributes to a better understanding of the possible physiological roles played by the osmoprotectant system is the fact that it has a tendency to purifying selection. This type of selection plays an important role in maintaining the long-term stability

of biological structures by removing deleterious mutations, thereby allowing the system to acquire improvements and not risk their subsequent loss (Loewe, 2008). In our view, because of its tendency to purifying selection, the system must originally have been fixed having a different purpose from osmoprotection, but still carrying out that function. This is why it is found in organisms in different habitats and is apparently not directly controlled by the σ^B transcription factor.

Table 5. Codon usage.

Organism	Genome codon usage		SBP codon usage	
<i>Rhizobium etli</i> CFN 42	UUC	UAU	UUC	UAU
	CUG	CAG	CUG	CAG
	AUC	AAG	AUC	AAG
	GUC	GAC	GUC	GAC
	UCG	UGG	UCG	UGG
	CCG	CGC	CCG	CGC
	ACC	AGC	ACC	AGC
	GCC	GGC	GCC	GGC
<i>Bacillus subtilis</i>	UUU	UAU	UUU	UAU
	CUG	CAA	CUG	CAA
	AUU	AAA	AUU	AAA
	GUU	GAA	GUU	GAA
	UCA	UGG	UCA	UGG
	CCG	CGC	CCG	CGC
	ACA	AGC	ACA	AGC
	GCA	GGC	GCG	GGC
<i>Agrobacterium tumefaciens</i> C58	UUC	UAU	UUC	UAU
	CUG	CAG	CUG	CAG
	AUC	AAG	AUC	AAG
	GUC	GAA	GUC	GAA
	UCG	UGC	UCG	UGC
	CCG	CGC	CCG	CGC
	ACC	AGC	ACC	AGA
	GCC	GGC	GCC	GGC
<i>Leifsonia xyli</i> subsp <i>xyli</i>	UUC	UAC	UUC	UAC
	CUC	CAG	CUC	CAG
	AUC	AAG	AUC	AAG
	GUC	GAG	GUC	GAG
	UCC	UGG	UCC	UGG
	CCG	CGC	CCG	CGC
	ACC	AGC	ACC	AGC
	GCC	GGC	GCC	GGC
<i>Yersinia pseudotuberculosis</i> IP 32953	UUU	UAU	UUU	UAU
	CUG	CAG	CUG	CAG
	AUU	AAA	AUU	AAA
	GUG	GAU	GUG	GAU
	UCA	UGG	UCU	UGG
	CCG	CGU	CCG	CGU
	ACC	AGC	ACC	AGC
	GCC	GGU	GCU	GGU
<i>Lactococcus lactis</i> subsp <i>lactis</i> IL 1403	UUU	UAU	UUU	UAU
	CUU	CAA	CUU	CAA
	AUU	AAA	AUU	AAA
	GUU	GAA	GUU	GAA
	UCA	UGG	UCA	UGG
	CCA	CGU	CCA	CGU
	ACA	AGU	CCA	CGU
	GCU	GGA	GCU	GGA

SBP = substrate-binding protein.

In the codon usage analysis, it was noticed that usage in the SBP sequences coincides with that of the rest of the genomes in which they occur. This information allows us to infer that the system is comprised of genes that are either native to the genome or that were horizontally transferred a long time ago, and therefore tend to have codon usage similar to the host genome (Davis and Olsen, 2011).

Analysis of the hypothetical ancestral sequences indicated that the osmoprotectant system must have arisen in prokaryotic organisms as a permease-choline transporter, owing to its similarity to current permeases and its conservation of protein domains characteristic of choline transporters. As for the three-dimensional models generated from these same sequences, there was a similarity with trimethylamine transporters, suggesting that the process of diversification to include the transport of other solutes must have started with choline, and later, small structural changes allowed other solutes to be transported. However, these changes did not affect the basic characteristics of the system.

According to our studies, the osmoprotectant system arose with the function of nutrient uptake, principally choline. This is an important molecule for the biosynthesis of membrane phospholipids (Aktas et al., 2010) and acts as a precursor of glycine betaine (Burkhardt et al., 2009) and participates in metabolic pathways for the synthesis of amino acids such as methionine (Barra et al., 2006). This system has been modified over time, allowing transport of a variety of molecules; furthermore, it has been structurally modified so that it now has morphological diversity, making it very interesting with regard to its SBPs.

The transport of trimethylamines seems to be the hallmark of this system such that, even though the system has undergone some changes along the evolutionary process, rather than losing specificity for these molecules, the specificity has been enhanced, with choline and betaine being prominent among the trimethylamines transported by this system.

This transport system is currently targeted much more at glycine betaine than choline, especially when considered as a cellular osmoprotection system. However, as suggested above, the osmoprotection function was not responsible for fixing the system. Because of the need to capture nutrients at the beginning of the diversification of this and other systems, organisms that could directly capture glycine betaine had an advantage over those that could only capture choline from which glycine betaine could be synthesized. This advantage can be explained by the fact that transporting glycine betaine is more energetically advantageous, since it makes it unnecessary for the organism to synthesize the two enzymes and perform the two chemical reactions necessary for the catabolism of choline and glycine betaine (Burkhardt et al., 2009) (Figure 9). This gain in terms of energy efficiency must have been fundamental to the successful establishment of glycine betaine transport and the sharing of this transport system among various prokaryotes.

At least three membrane proteins are associated with the transport of glycine betaine: OpuA (osmoprotectant system), OpuC (osmoprotectant system), and OpuD (Pittelkow et al., 2011). However, the only one that is synthesized following the activation of σ^B is OpuD (Hecker et al., 2007; Hahne et al., 2010), suggesting that OpuD acts directly in response to osmotic stress, whereas members of the osmoprotectant system act in a complementary manner.

This complementary action can be fully understood by recognizing that, under salt stress conditions, especially hyperosmotic conditions, the organism will use all available metabolic resources for homeostasis. Thus, when the osmoprotection system must act, it does so as a secondary function of what is really a nutrient uptake system. This is because glycine betaine, in particular, has the required physicochemical properties of an osmoprotectant and

also acts as a raw material for amino acid synthesis. Under stress, synthesis of the proteins necessary for glycine betaine uptake is increased to maintain cell physiology, i.e., the osmoprotectant system is involved in some way to the response to osmotic stress, but as part of a generalized stress response rather than a specific reaction to osmotic stress (Figure 10).

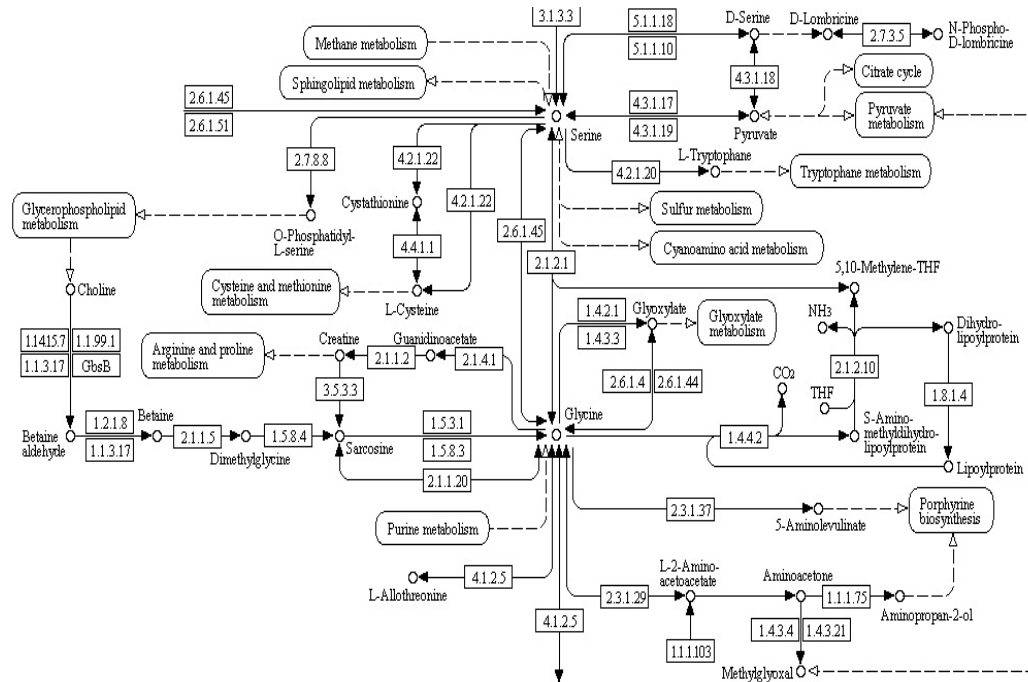


Figure 9. Metabolic map of the glycine betaine synthesis pathway from choline. Adapted from KEGG.

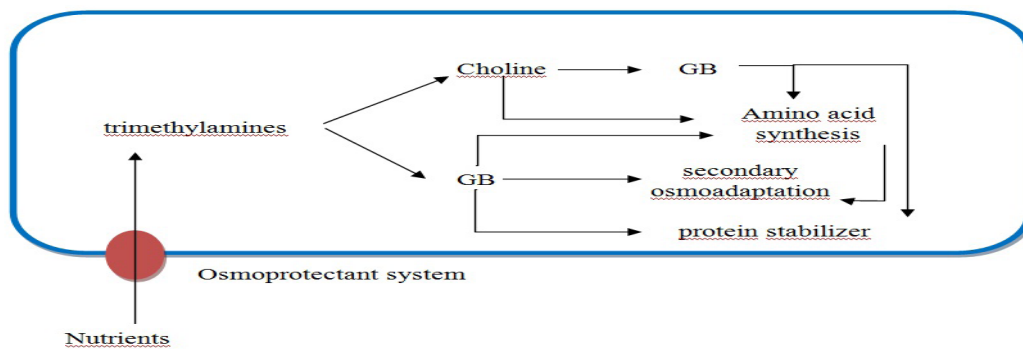


Figure 10. Physiological diversity of glycine betaine and choline. The glycine betaine (GB) absorption became more energetically advantageous due to multiple activities.

Therefore, according to our understanding, the osmoprotectant system was fixed by natural selection because it plays a fundamental physiological function, that is, to capture nutrients. Structural changes, among other aspects, allowed the system to transport different types of molecules without loss of function.

This system probably arose and was horizontally transferred before prokaryotic adaptation to diverse habitats. Because of this, its codon usage pattern matches that of the rest of the genome, and it is currently present in organisms inhabiting different environments.

The osmoprotection property must have arisen secondarily, as the σ^B transcription factor seems to have no direct effect on the transcription of the system. However, in environments with varying salinity, this system has proven to be important for organism homeostasis.

[Supplementary material](#)

REFERENCES

- Aktas M, Wessel M, Hacker S, Klusener S, et al. (2010). Phosphatidylcholine biosynthesis and its significance in bacteria interacting with eukaryotic cells. *Eur. J. Cell Biol.* 89: 888-894.
- Barra L, Fontenelle C, Ermel G, Trautwetter A, et al. (2006). Interrelations between glycine betaine catabolism and methionine biosynthesis in *Sinorhizobium meliloti* strain 102F34. *J. Bacteriol.* 188: 7195-7204.
- Biemans-Oldehinkel E, Doeven MK and Poolman B (2006). ABC transporter architecture and regulatory roles of accessory domains. *FEBS Lett.* 580: 1023-1035.
- Burkhardt J, Sewald X, Bauer B, Saum SH, et al. (2009). Synthesis of glycine betaine from choline in the moderate halophile *Halobacillus halophilus*: co-regulation of two divergent, polycistronic operons. *Environ. Microbiol. Rep.* 1: 38-43.
- DasGupta B, He X, Jiang T, Li M, et al (2000). On Computing the Nearest Neighbor Interchange Distance. In: Proceedings of the DIMACS Workshop on Discrete Problems with Medical Applications, DIMACS Series in Discrete Mathematics and Theoretical Computer Science (Du DZ, Pardalos PM and Wang J, eds.). *Am. Math. Soc.* 55: 125-143.
- Davidson AL, Dassa E, Orelle C and Chen J (2008). Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol. Mol. Biol. Rev.* 72: 317-64.
- Davis JJ and Olsen GJ (2011). Characterizing the native codon usages of a genome: an axis projection approach. *Mol. Biol. Evol.* 28: 211-221.
- Du Y, Shi WW, He YX, Yang YH, et al. (2011). Structures of the substrate-binding protein provide insights into the multiple compatible solute binding specificities of the *Bacillus subtilis* ABC transporter OpuC. *Biochem. J.* 436: 283-289.
- Empadinhas N and da Costa MS (2008). Osmoadaptation mechanisms in prokaryotes: distribution of compatible solutes. *Int. Microbiol.* 11: 151-161.
- Hahne H, Mader U, Otto A, Bonn F, et al. (2010). A comprehensive proteomics and transcriptomics analysis of *Bacillus subtilis* salt stress adaptation. *J. Bacteriol.* 192: 870-882.
- Hecker M, Pane-Farre J and Völker U (2007). SigB-dependent general stress response in *Bacillus subtilis* and related Gram-positive bacteria. *Annu. Rev. Microbiol.* 61: 215-236.
- Höper D, Völker U and Hecker M (2005). Comprehensive characterization of the contribution of individual SigB-dependent general stress genes to stress resistance of *Bacillus subtilis*. *J. Bacteriol.* 187: 2810-2826.
- Kanehisa M and Goto S (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28: 27-30.
- Kanehisa M, Goto S, Sato Y, Furumichi M, et al. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40: D109-D114.
- Loewe L (2008). Negative Selection. Nature Education, Cambridge.
- Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, et al. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* 39: 225-229.
- Moses S, Sinner T, Zapras A, Stoveken N, et al. (2012). Proline utilization by *Bacillus subtilis*: uptake and catabolism. *J. Bacteriol.* 194: 745-758.
- Oldham ML, Davidson AL and Chen J (2008). Structural insights into ABC transporter mechanism. *Curr. Opin. Struct. Biol.* 18: 726-733.
- Petersohn A, Brigulla M, Haas S, Hoheisel JD, et al. (2001). Global analysis of the general stress response of *Bacillus subtilis*. *J. Bacteriol.* 183: 5617-5631.
- Pittelkow M, Tschapek B, Smits SH, Schmitt L, et al. (2011). The crystal structure of the substrate-binding protein OpuBC from *Bacillus subtilis* in complex with choline. *J. Mol. Biol.* 411: 53-67.
- Roberts MF (2005). Organic compatible solutes of halotolerant and halophilic microorganisms. *Saline Systems* 1: 5.

- Saitou N and Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406-425.
- Slater SC, Goldman BS, Goodner B, Setubal JC, et al. (2009). Genome sequences of three *Agrobacterium* biovars help elucidate the evolution of multichromosome genomes in bacteria. *J. Bacteriol.* 191: 2501-2511.
- Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739.
- Thomas GH (2010). Homes for the orphans: utilization of multiple substrate-binding proteins by ABC transporters. *Mol. Microbiol.* 75: 6-9.
- Thompson JD, Higgins DG and Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680.
- Yancey PH (2005). Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. *J. Exp. Biol.* 208: 2819-2830.
- Ziegler C, Bremer E and Kramer R (2010). The BCCT family of carriers: from physiology to crystal structure. *Mol. Microbiol.* 78: 13-34.