5-4-2015

Environmental Processing in Meiothermus ruber: the Inorganic Phosphate ABC Transporter

Mich A. Gehrig Jr.
Augustana College - Rock Island

Emma M. Segura-Fernandez
Augustana College - Rock Island

Dr. Lori Scott
Augustana College - Rock Island

Research

Environmental Processing in *Meiothermus ruber*: the Inorganic Phosphate ABC Transporter

Mich A. Gehrig Jr.[a], Emma M. Segura-Fernandez[a], Dr. Lori Scott[a]

[a] *Augustana College, Rock Island, IL 61201, USA*

ARTICLE INFO

Article History:

Received 4 May 2015

Keywords

Meiothermus ruber
Escherichia Coli
Bioinformatics
Genome
Annotation
GENI-ACT
Phosphate
ABC Transporter
pstS
pstA
pstB
pstC

ABSTRACT

In this paper, we use *Escherichia coli* (*E. coli*) as a model organism to investigate if *Meiothermus ruber* (*M. ruber*) has genes that are comparable to the genes responsible for inorganic phosphate ABC transport in *E. coli* (pstS, pstC, pstA, pstB). Various bioinformatics tools were utilized to compare the similarity of function of the gene products of these two organisms, providing evidence of the cellular location, the protein domain, as well as the protein family to which they belong. These data are sufficient to conclude with confidence that the genes that code for the proteins of the phosphate ABC transport system in *E. coli* are comparable or homologues to the genes proposed to code for the proteins involved in the phosphate ABC transport system in *M. ruber*. We also searched for evidence of horizontal gene transfer (HGT) in all the four genes of *E. coli* as well as all the four putative genes in *M. ruber*. After analyzing the data obtained from the bioinformatics tools, it was concluded that the genes of *E. coli* do not show evidence or signs of HGT, while the comparable *M. ruber* genes showed moderate evidence of distant HGT, though not enough to conclude a recent HGT event.

1. INTRODUCTION

Thermophiles such as *Meiothermus ruber* (*M. ruber*)—a gram-negative, nonmotile, and nonsporulating bacteria—live in nutrient-deprived environments (Nobre et al., 1996). Consequently, they require efficient means to acquire nutrients from their environment, as failing to uptake nutrients even just once may spell death for the organism, since nutrients are sparse in these conditions.

The environmental processing of gram-negative bacteria differs from that of gram-positive bacteria, due to the difference in their effective cell walls. Gram-negative bacteria (including *E. coli*) have two cytoplasmic membranes—an outer membrane and an inner membrane. The space between these two membranes is called the periplasmic space, and it is thinly laced with peptidoglycans that provide structural rigidity. The outer membrane is

coated in porin complexes that permit the ATP-independent diffusion of smaller polar compounds into the periplasmic membrane from the environment. Once in the periplasm, there are various ways by which the cell can transport substrates into the cytoplasm; in particular, this study investigates the protein superfamily of ATP-Binding Cassette (ABC) transporters. In an ATP-dependent manner, ABC transporters facilitate the translocation of various substrates across the inner cytoplasmic membrane. Holistically, ABC transporters are 5-part systems, consisting of a periplasmic substrate-binding protein, two transmembrane proteins (inner membrane), and a homodimeric (two identical proteins which are linked) ATPase that is bound to the cytoplasmic domains of the two transmembrane proteins. The free-floating periplasmic protein binds specifically to its substrate and sequesters it to the appropriate ABC transporter, rooted in the inner cytoplasmic membrane. When the binding protein comes in contact with the transmembrane proteins, the homodimeric ATPase hydrolyses ATP into ADP at nucleotide binding domains. The chemical energy of ATP is translated into the mechanical movement of the two transmembrane proteins, effectively "opening a gate" by which the substrate bound to the binding protein can flow through the membrane and into the cytoplasm. The process is attenuated by the dissociation of the substrate binding protein and the regeneration of the ATPase. (Moussatova and Kandt, 2008)

Inorganic phosphate (P_i; PO_3^{2-}) is necessary in energy production, energy use, membrane formation, enzymatic function, DNA replication and thus proliferation of all cells. Phosphate esters are the basis of many pivotal biological structures, and are included in the structures of key compounds such as ATP and various phospholipids (Berg et al., 2012). In that light, the environmental processing and subsequent uptake of phosphate is a high priority for bacteria, especially for those who live in nutrient-poor environments such as thermophiles. It has been reported that *E. coli*, a highly studied gram-negative bacteria, regulates phosphate homeostasis with an inorganic phosphate ABC transporter. This transporter, dubbed the Pst system, is a typical 5-component ABC system: pstS (*b3728*) is the periplasmic phosphate-binding protein, pstA (*b3726*) and pstC (*b3727*) are the two transmembrane proteins, and pstB (*b3725*) is the homodimeric ATPase (Webb et al., 1992). In this study, we utilized *E. coli* as a model organism to determine if *M. ruber* has a comparable means of regulating phosphate intake. From preliminary investigations, it is believed that the genes Mrub_2518, Mrub_2520, Mrub_2519, and Mrub_2521 of *M. ruber* code for the equivalent pstS, pstA, pstC, and pstB genes of *E. coli*, respectively. Various bioinformatic tools were utilized in our assessment of the functional relatedness of the Pst system in *E. coli* and the proposed system in *M. ruber*.

2. MATERIALS & METHODS

A comparative genome analysis between the model organism *E. coli* and the analyte *M. ruber* was conducted, utilizing the following publically available online bioinformatics tools, all of which were accessed through the Guiding Education through Novel Investigation (GENI) – Annotation Collaboration Toolkit (ACT).

AVAILABLE MODULES ON GENI-ACT	BIOINFORMATICS PROGRAMS
Basic Information	GENI-ACT: http://geni-act.org/
Sequence-based Similarity Data	NCBI BLAST: http://blast.ncbi.nlm.nih.gov/Blast.cgi CCD: http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml T-Coffee: http://www.tcoffee.org/Projects/tcoffee/ WebLogo: http://weblogo.berkeley.edu/logo.cgi
Cellular Localization Data	TMHMM: http://www.cbs.dtu.dk/services/TMHMM-2.0/ SignalP: http://www.cbs.dtu.dk/services/SignalP/ LipoP: http://www.cbs.dtu.dk/services/LipoP/ PSORT-B: http://www.psort.org/psortb/ Phobius: http://phobius.sbc.su.se/
Alternative Open Reading Frame	JGI IMG/EDU 6-Frame viewer: http://img.jgi.doe.gov/cgi---bin/edu/main.cgi
Structure-based Evidence	TIGRFAM: http://blast.jcvi.org/web---hmm/ Pfam: http://pfam.xfam.org/search PDB: http://www.rcsb.org/pdb/home/home.do
Enzymatic Function	KEGG: http://www.genome.jp/kegg/ MetaCyc: http://metacyc.org/) ExPASy: http://enzyme.expasy.org/enzyme---search---ec.html
Duplication and Degradation	JGI IMG-EDU: http://img.jgi.doe.gov/cgi---bin/edu/main.cgi
Horizontal Gene Transfer	Phylogeny.fr: http://www.phylogeny.fr/), JGI IMG/EDU: http://img.jgi.doe.gov/cgi---bin/edu/main.cgi

3. RESULTS

The outputs from the various bioinformatic tools utilized in this investigation are summarized below in tables. For complete coverage of the data output, see the Appendix. Output values listed are first given for the *E. coli* gene and then the respective *M. ruber* gene.

3.1 E. coli b3725 pstB and M. ruber Mrub_2521 (TABLE 1)

Each gene product was predicted to have zero transmembrane helices (TMHMM) (Fig. A.1a), and low likelihoods of containing a signal peptide (D=0.140; D=0.110) (Fig. A.1b) or a cleavage site. The gene products were predicted to be housed in the cytoplasmic membrane by PSORT-B (p=10.00; p=10.00) and Phobius (Fig. A.1c) predicted non-cytoplasmic for both, but LipoP predicted that both are cytoplasmic proteins. A BLAST of the two amino acid sequences against each other yields an alignment with an expect value of 6e-102, a bit score of 290, and 53% identity (Fig. A.2a). BLASTed individually, the top hit for b3725 was the phosphate ABC transporter ATP-binding protein of *Shigella sonnei* (E=0.0) (Fig. A.2b); the top hit for Mrub_2521 was a hypothetical protein from *Meiothermus taiwanesis* (E=0.0) (Fig. A.2c). Both gene products belong to the same KEGG pathway (ABC Transporters) (Fig. A.3). Each gene product had the same top hit for the CDD COG analysis (COG1117; E=0.0; E=8.80e-168), Pfam (PF00005; E=1.3e-34; 5.0e-30), TIGRfam (TIGR00972; E=3.1e-188; E=1.1e-171), and PDB (2OLJ; E=8.3e-35; E=2.1e-38). The GC% for b3725 is 52% (genomic 51%), the GC% for Mrub_2521 is 56% (genomic 63%). Gene neighborhood maps (Fig. A.4a,b) and phylogenetic trees were constructed for b3725 and Mrub_2521 (Fig. A.5a,b).

Table 1: Summary of Bioinformatic Outputs for b3725 and Mrub_2521

Description of evidence collected	*E. coli* (b3725)	*M. ruber* (Mrub_2521) Replicate information not listed
Cellular Localization (summary of TMHMM, SignalP, LipoP, PSORT-B, Phobius)	Cytoplasmic membrane	Cytoplasmic membrane
BLAST *E. coli* against *M. ruber*	**E-value: 6e-102** Bit score: 290 bits Identities value: 181/258 (53%)	--
BLAST (Top Hit)	**Shigella sonnei (E = 0.0)** Phosphate ABC transporter ATP-binding protein	**Meiothermus taiwanensis (E = 0.0)** Hypothetical protein
KEGG pathway	**ABC Transporters:** Phosphate and amino acid transporters	**ABC Transporters**
CDD – protein domain (COG category)	**COG1117 (E = 0.0)** *PstB*: ABC-type phosphate transport system, ATPase component [Inorganic ion transport and metabolism]	**COG1117 (E = 8.80e-168)**
Pfam – protein domains	**PF00005 (E = 1.3e-34)** *Description*: ABC transporter *Family*: ABC_tran *Clan*: CL0023; P-loop_NTPase P-loop containing nucleoside triphosphate hydrolase superfamily	**PF00005 (E = 5.0e-30)**
TIGRfam – protein family	**TIGR00972 (E = 3.1e-188)** phosphate ABC transporter, ATP	**TIGR00972 (E = 1.1e-171)**
E.C. number	**3.6.3.27** Phosphate-transporting ATPase $ATP + H_2O + phosphate_{out}$ <=> $ADP + phosphate + phosphate_{in}$	No match
PDB	**2OLJ (E = 8.3e-35)** ABC Protein ArtP in complex with ADP/Mg2+	**2OLJ (E = 2.1e-38)**
Evidence of HGT?	No	**Yes**, but not recently

3.2 E. coli b3726 pstA and M. ruber Mrub_2520 (TABLE 2)

Each gene product was predicted to have six transmembrane helices (TMHMM) (Fig B.1a), and low likelihoods of containing a signal peptide (D=0.131; D=0.158) (Fig B.1b) or a cleavage site. The gene products were predicted to be a transmembrane protein by PSORT-B (p=10.00; p=10.00), Phobius (Fig B.1c), and LipoP. A BLAST of the two amino acid sequences against each other yields an alignment with an expect value of 4e-51, a bit score of 160, and 37% identity (Fig B.2a). BLASTed individually, the top hit for b3726 was the phosphate transporter permease subunit PstA of *Escherichia albertii* (E=0.0) (Fig B.2b); the top hit for Mrub_2520 was a hypothetical protein from *Meiothermus taiwanesis* (E=6.0e-171) (Fig B.2c). Both gene products belong to the same KEGG pathway (ABC Transporters) (Fig. A.3). Each gene product had the same top hit for the CDD COG analysis (COG0581; E=5.12e-104; E=1.97e-61), Pfam (PF00528; E=1.6e-30; 7.0e-21), and TIGRfam (TIGR00974; E=7.6e-103; E=7.0e-21). There were no PDB matches for either gene product with suitable E values (E<1). The GC% for b3726 is 56% (genomic 51%); the GC% for Mrub_2520 is 61% (genomic 63%). Gene neighborhood maps (Fig. A.4a,b) and phylogenetic trees (Fig. B.3a,b) were constructed for b3726 and Mrub_2520.

Table 2: Summary of Bioinformatic Outputs for b3726 and Mrub_2520

Description of evidence collected	*E. coli* (b3726)	*M. ruber* (Mrub_2520) Replicate information not listed
Cellular Localization (summary of TMHMM, SignalP, LipoP, PSORT-B, Phobius)	Cytoplasmic membrane	Cytoplasmic membrane
BLAST *E. coli* against *M. ruber*	**E-value: 4e-51** Bit score: 160 bits Identities value: 92/250 (37%)	--
BLAST (Top Hit)	**Escherichia albertii (E = 0.0)** Phosphate transporter permease subunit PstA	**Meiothermus taiwanensis (E = 6.0e-171)** Hypothetical protein
KEGG pathway	**ABC Transporters:** Phosphate and amino acid transporters	**ABC Transporters**
CDD – protein domain (COG category)	**COG0581 (E = 5.12e-104)** *PstA:* ABC-type phosphate transport system	**COG0581 (E = 1.97e-61)**
Pfam – protein domains	**PF00528 (E = 1.6e-30)** *Description:* Binding-protein-dependent transport system inner membrane component *Family:* BPD_transp_1 *Clan:* CL0404; BPD transporter like bacterial binding protein-dependent transport system inner membrane component, which is an ATP dependent system involved in transport of a range of substrates	**PF00528 (E = 7.0e-21)**
TIGRfam – protein family	**TIGR00974 (E = 7.6e-103)** phosphate ABC transporter, permease	**TIGR00974 (E = 7.0e-21)**
E.C. number	**3.6.3.27** Phosphate-transporting ATPase $ATP + H_2O + phosphate_{out}$ <=> $ADP + phosphate + phosphate_{in}$	No match
PDB	No match with suitable E values	No match with suitable E values
Evidence of HGT?	No	**Yes**, but not recently

3.3 E. coli b3727 pstC and M. ruber Mrub_2519 (TABLE 3)

b3727 was predicted to have six transmembrane helices while Mrub_2519 was predicted to have eight transmembrane helices (TMHMM) (Fig C.1a); neither gene has a predicted signal peptide (D=0.0; D=0.106) (Fig C.1b) nor a cleavage site. The gene products were predicted to be housed in the cytoplasmic membrane by PSORT-B (p=10.00; p=10.00) and LipoP, while Phobius predicted both genes have six transmembrance helices (Fig C.1c). A BLAST of the two amino acid sequences against each other yields an alignment with an expect value of 1e-57, a bit score of 444, and 41% identity (Fig C.2a). BLASTed individually, the top hit for b3727 was the phosphate transporter permease subunit PstC of *Salmonella enterica* (E=0.0) (Fig C.2b); the top hit for Mrub_2519 was the phosphate ABC transporter permease from *Meiothermus cerbereus* (E=0.0) (Fig C.2c). Both gene products belong to the same KEGG pathway (ABC Transporters) (Fig. A.3). Each gene product had the same top hit for the CDD COG analysis (COG4149; E=2.20e-15; E=9.95e-13), Pfam (PF00528; E=6.0e-18; 8.4e-20), and TIGRfam (TIGR02138; E=2.4e-92; E=4.1e-80). The PDB hits for this pair of genes were weak: 3D31 for b3727 (E=0.99) and 3DHW for Mrub_2519 (E=0.22). The GC% for b3727 is 52% (genomic 51%), the GC% for Mrub_2519 is 57% (genomic 63%). Gene neighborhood maps (Fig. A.4a,b) and phylogenetic trees were constructed for b3727 and Mrub_2519 (Fig C.3a,b).

Table 3: Summary of Bioinformatic Outputs for b3727 and Mrub_2519

Description of evidence collected	*E. coli* (b3727)	*M. ruber* (Mrub_2519) Replicate information not listed
Cellular Localization (summary of TMHMM, SignalP, LipoP, PSORT-B, Phobius)	Cytoplasmic membrane	Cytoplasmic membrane
BLAST *E. coli* against *M. ruber*	**E-value: 1e-57** Bit score: 444 bits Identities value: 101/246 (41%)	--
BLAST (Top Hit)	**Salmonella enterica (E = 0.0)** phosphate transporter permease subunit PstC	**Meiothermus cerbereus (E = 0.0)** phosphate ABC transporter permease
KEGG pathway	**ABC Transporters:** Phosphate and amino acid transporters	**ABC Transporters**
CDD – protein domain (COG category)	**COG4149 (E = 2.20e-15)** ABC-type molybdate transport system, permease component [Inorganic ion transport and metabolism]	**COG4149 (E = 9.95e-13)**
Pfam – protein domains	**PF00528 (E = 6e-18)** Description: Binding-protein-dependent transport system inner membrane component Family: BPD_transp_1 Clan:CL0404 BPD transporter like	**PF00528 (E = 8.4e-20)**
TIGRfam – protein family	**TIGR02138 (E = 2.4e-92)** phosphate_pstC: phosphate ABC transporter	**TIGR02138 (E = 4.1e-80)**
E.C. number	**3.6.3.27** Phosphate-transporting ATPase $ATP + H_2O + phosphate_{out}$ <=> $ADP + phosphate + phosphate_{in}$	No match
PDB	**3D31(E= 0.99)** ModBC from Methanosarcina acetivorans	**3DHW (E= 0.22)** Crystal structure of methionine importer MetNI
Evidence of HGT?	No	**Yes**, but not recently

3.4 E. coli b3728 pstS and M. ruber Mrub_2518 (TABLE 4)

b3728 was predicted to have one transmembrane helix, while Mrub_2518 did not have any predicted (TMHMM) (Fig D.1a). There is a high probability that each gene product contains a signal peptide as predicted by SignalP (D=0.570 between position 25 and 26; D=0.779 between position 20 and 21) (Fig D.1b) and LipoP (cleavage between position 25 and 26 by SpI; cleavage between position 20 and 21 by SpI). The gene products were predicted to be housed in the periplasmic space by PSORT-B (p=10.00; p=9.76), and Phobius predicted that both are non-cytoplasmic with an initial signal peptide sequence (Fig D.1c). A BLAST of the two amino acid sequences against each other yields an alignment with an expect value of 5e-56, a bit score of 438, and 36% identity (Fig D.2a). BLASTed individually, the top hit for b3728 was the phosphate ABC transporter substrate-binding protein of *Shigella boydii* (E=0.0) (Fig D.2b); the top hit for Mrub_2518 was the phosphate ABC transporter substrate-binding protein from *Meiothermus taiwanesis* (E=0.0) (Fig D.2c). Both gene products belong to the same KEGG pathway (ABC Transporters) (Fig. A.3). Each gene product has the same top hit for the CDD COG analysis (COG0226; E=2.44e-88; E=1.52e-51), Pfam (PF12849; E=1.3e-55; E=5.7e-41), TIGRfam (TIGR00972; E=3.3e-96; E=1.3e-80). The PDB hits for this pair of genes were not identical, however: 1A40 for b3728 (E=0.0) and 2Z22 for Mrub_2518 (E=4.2e-61). The GC% for b3728 is 52% (genomic 51%) and the GC% for Mrub_2518 is 60% (genomic 63%). Gene neighborhood maps (Fig. A.4a,b) and phylogenetic trees (Fig D.3a,b) were constructed for b3728 and Mrub_2518.

Table 4: Summary of Bioinformatic Outputs for b3728 and Mrub_2518

Description of evidence collected	*E. coli* (b3728)	*M. ruber* (Mrub_2518) Replicate information not listed
Cellular Localization (summary of TMHMM, SignalP, LipoP, PSORT-B, Phobius)	Periplasmic space	Periplasmic space
BLAST *E. coli* against *M. ruber*	**E-value: 5e-56** Bit score: 438 bits Identities value: 107/295 (36%)	--
BLAST (Top Hit)	**Shigella boydii (E = 0.0)** Phosphate ABC transporter substrate-binding protein	**Meiothermus taiwanensis (E = 0.0)** phosphate ABC transporter substrate-binding protein
KEGG pathway	**ABC Transporters:** Phosphate and amino acid transporters	**ABC Transporters**
CDD – protein domain (COG category)	***COG0226* (E = 2.44e-88):** *PstS protein domain*	**COG0226 (E = 1.52e-51)**
Pfam – protein domains	**PF12849 (E = 1.3e-55)** *Description:* PBP superfamily domain Family: PBP_like_2 *Clan:* CL0177; Periplasmic binding protein clan	**PF12849 (E = 5.7e-41)**
TIGRfam – protein family	**TIGR00975 (E = 3.3e-96)** **phosphate ABC transporter**	**TIGR00975 (E = 1.3e-80)**
E.C. number	**3.6.3.27** Phosphate-transporting ATPase $ATP + H_2O + phosphate_{out}$ <=> $ADP + phosphate + phosphate_{in}$	No match
PDB	**1A40 (E = 0.0)** Phosphate-Finding Protein with ALA 197 Replaced with TRP	**2Z22 (E = 4.2e-61)** Crystal Structure of Phosphate Preplasmic binding Protein psts from Yersinia Pestis
Evidence of HGT?	No	**Yes**, but not recently

4. Discussion

Interestingly, there is no E.C. number for the hypothesized Pst system in *M. ruber*. However, since E.C. numbers confer enzymatic reactions (and all four components are involved in the same reaction), it's actually more surprising that there *is* a corresponding E.C. number to the *E. coli* Pst system. Rather, talks of E.C. numbers in reference to transport systems are seemingly irrelevant; nonetheless, the E.C. number for the Pst system in *E. coli* was included in the results as additional proof that there is functional evidence of the system.

4.1 b3725 pstB and Mrub_2521

The results from the comparison of these two ATPase genes overwhelmingly suggest that they are functionally similar. One discrepancy was in the determination of the cellular localization of both the *E. coli* and *M. ruber* gene products. LipoP predicted that both gene products are located in the cytoplasm, while PSORT-B and Phobius predicted that they both pertain to the cytoplasmic membrane. There were no predicted transmembrane helices nor signal peptides, so the true location must be among those that were predicted. The different determinations of the cellular localization seems to be related to the fact that the ATPase in the Pst system is a peripheral protein, bound to the cytoplasmic portions of the two transmembrane proteins of the system. So, since the gene products *are* located in the cytoplasm, though they are not free-floating like other cytoplasmic proteins, it is understandable why LipoP would predict that these gene products are cytoplasmic. In the end, the designation by these analytical programs is inconsequential in this particular case, since each program does not consider protein domains that may confer a peripheral localization on the cytoplasmic membrane—it is an inherent flaw of the bioinformatic tools.

Otherwise, there is an apparent similarity between b3725 and Mrub_2521. The BLAST of the two amino acid sequences against each other yielded a low E-value, suggesting their structural and thus functional relatedness. The top BLAST hit for Mrub_2521 corresponded to a hypothetical protein of *M. taiwanensis*, but the stark similarities in each of the other bioinformatic outputs makes the ambiguity—the fact that it's hypothetical—of this result unimportant. There were identical hits with acceptable expect values for KEGG, CDD, Pfam, and TIGRfam, providing overwhelming evidence for the functional relatedness of the two gene products. Moreover, the hits they were matched with were expected: the pstB protein domain (CDD), the ABC transporter protein family (Pfam), and the ATPase component of the phosphate ABC transporter (TIGRfam).

b3725 shows no sign of horizontal gene transfer (HGT), neither in its phylogenetic tree (only linked to organisms which are closely related to it by descent) nor its GC% (no remarkable difference from the genomic GC%). Mrub_2521, however, does show signs of distant HGT. Mrub_2521 is linked to organisms that are not closely related to *M. ruber*

(phylum Deinococcus-Thermus) by descent, such as *Spirochaeta Africana* (phylum spirochetes) and *Geobacter metallireducens* (phylum proteobacteria). Moreover, the stark difference in the GC% of Mrub_2521 (56%) and the genomic GC% (63%) is beyond the threshold of normal variance within the genome (generally ±5% of the genomic GC%). Lastly, the supposed Pst operon in *M. ruber* shows strong conservation in organisms that are distantly related to it by descent, suggesting that HGT occurred as a translocation of the entire operon (Fig E.1). These data holistically suggest that HGT occurred in *M. ruber* at some time in the past.

4.2 b3726 pstA and Mrub_2520

The results from the comparison of these two transmembrane pstA-like genes overwhelmingly suggest that they are functionally similar. There are no discrepancies to evaluate. The BLAST of the two amino acid sequences against each other yielded a low E-value, suggesting their structural and thus functional relatedness. The top BLAST hit for Mrub_2520 corresponded to a hypothetical protein of *M. taiwanensis*, but the stark similarities in each of the other bioinformatic outputs makes the ambiguity of this result unimportant. There were identical hits with acceptable expect values for KEGG, CDD, Pfam, and TIGRfam, providing overwhelming evidence for the functional relatedness of the two gene products. Moreover, the hits they were matched with were expected: the pstA protein domain (CDD), the inner membrane component of a binding-protein-dependent transport system (Pfam), and the permease component of the phosphate ABC transporter (TIGRfam).

b3726 shows no sign of HGT, neither in its phylogenetic tree (only linked to organisms which are closely related to it by descent) nor its GC% (no remarkable difference from the genomic GC%). Mrub_2520, however, does show signs of distant HGT. Mrub_2520 is linked to organisms that are not closely related to *M. ruber* (phylum Deinococcus-Thermus) by descent, such as *Spirochaeta Africana* (phylum spirochetes) and *Geobacter bemidjiens* and *Geobacter uraniireducens* (phylum proteobacteria). The GC% of Mrub_2520 (61%), however, does not significantly differ from the genomic GC% (63%) and as such is not evidence for HGT. Lastly, the supposed Pst operon in *M. ruber* shows strong conservation in organisms that are distantly related to it by descent, suggesting that HGT occurred as a translocation of the entire operon (Fig E.1). These data holistically offer conflicting views of whether HGT occurred for Mrub_2520.

4.3 b3727 pstC and Mrub_2519

The comparison between b3727 (pstC) and M_rub2519 strongly implies a functional similarity of the two gene products, thus providing evidence that M_rub2519 is homologous to b3727.

The BLAST of b3727 against Mrub_2519 has a low expect value and a relatively high bit score, which is evidence of strong sequence similarity, thus indicative of functional similarity. The BLAST top hit for the gene product of b3727 is the phosphate transporter permease subunit pstC of *Salmonella enterica*. These results show that b3727 has sequence similarities with another gene from a different organism that also codes for the same pstC protein in the phosphate transport system. The BLAST top hit gene product of Mrub_2519 is the phosphate ABC transporter permease of *Meiothermus cerbereus*. These results show that the top hits for the *E. coli* and *M. ruber* gene products are both involved in phosphate transport, which again, provides further evidence that these two genes code for a functionally similar protein.

To further confirm that these two genes are homologues to each other, CDD shows evidence that these two genes belong to the ABC-type molybdate transport system protein domain. Pfam predicts that both genes contain the same protein domain and pertain to the same domain family (BPD: binding protein dependent transport system); furthermore, TIGRfam predicts that both genes belong to the same protein family: phosphate pstC (phosphate ABC transport). All these data also show evidence that these two genes have similar functions, and are therefore homologues to each other. However, there is discrepancy in the PDB results (each gene has a different match), but this is not concerning since both of these hits have high expect values and, as such, are irrelevant to the system being analyzed. Overall, the majority of the data analyzed show evidence that b3727 and Mrub_2518 are homologous to each other.

In determining whether these genes were acquired by HGT or not, it is immediately noticeable that *E. coli* pstC does not show evidence of HGT. A quick look at the phylogenetic tree shows that all the organisms that surround *E. coli* belong to the same phylum, and therefore this provides no evidence of HGT. Also, b3727's GC% (55%) does not differ significantly from the genomic GC content, which further refutes the possibility of HGT.

On the other hand, M_rub2519 (pstC) shows strong evidence of distant HGT. The reason being is that there are organisms below and above Mrub_2518 in its phylogenetic tree that belong to different phyla. Although there are two organisms fairly close to *M. ruber* that could potentially concede recent HGT (*Desulfurivibrio alkaliphilus* and *Spirochaeta Africana*), we suggest that these two organisms were recent recipients of HGT, possibly from the nearby thermus organisms—instead of *M. ruber* being a recipient of HGT from these organisms. Furthermore, the GC% of the Mrub_2519 (57%) differs significantly from the genomic GC content, providing evidence for HGT. These data holistically offer considerable evidence for the distant HGT of Mrub_2519.

4.4 b3728 pstS and Mrub_2518

The comparison between the genes Mrub_2518 and b3728, which encodes for the binding protein pstS in the inorganic phosphate ABC transport system, strongly implies a functional similarity of the two.

The BLAST of b3728 against Mrub_2518 has a low expect value and a relatively high bit score, which is evidence of strong sequence similarity—indicative of functional similarity. The BLAST top hit gene product of b3728 is the phosphate ABC transporter substrate-binding protein of *Shigella boydii*, an organism that belongs to the same phylum as *E. coli*. These results show that b3728 has sequence similarities with another gene from another organism that also codes for the same binding protein in the phosphate ABC transport system. The BLAST top hit gene product of Mrub_2518 is also a phosphate ABC transporter substrate-binding protein, belonging to *Meiothermus Taiwanesis*. These results show that the top hits for these *E. coli* and *M. ruber* gene products are the same for both organisms, which again, further proves that these two genes code for the functionally similar proteins.

To further confirm that these two genes are homologues to each other, CDD shows evidence that these two genes belong to the pstS protein domain. Pfam predicts the same PBP (periplasmic binding protein) family for both genes. Both genes belong to the same phosphate ABC transport TIGRfam. All these data show evidence that these two genes have similar functions, and are therefore homologues to each other. Nevertheless, there is a little discrepancy in the PDB results. Both genes have different matches but both of them belong to a phosphate binding protein; the reason for this discrepancy might be due to the fact that the PDB findings for Mrub_2518 belong to an organism more closely related to *M. ruber*, while the PDB hit for b3728 pertains to an organism that is more closely related to *E. coli* than it is to *M. ruber*. Overall the majority the data analyzed provides evidence that b3728 and Mrub_2518 are homologous to each other.

When looking for evidence of HGT in these two genes, *E. coli* pstS does not show evidence of HGT since in its phylogenetic tree it is surrounded by organisms which belong to the same phylum. The GC percentage is a further indication that there is no HGT. The GC% of the genome is 51% and the GC% of the gene is 51%. The lack of a difference in GC% is strong evidence against HGT.

On the other hand, *M. ruber* pstS shows substantial evidence of distant HGT. This is predicted from the phylogenetic tree, which shows *M. ruber* closely surrounded by organisms that belong to the same Deinococcus-Thermus phylum. But if we look at the organisms above *Meithermus Silvanus*, none of them belong to the Deinococcus-Thermus phylum. For example, *Spirochaeta Africana* belongs to the Spirochaeta phylum, and *Desulfurivibrio alkaliphilus* belongs to the proteobacteria phylum, so this shows evidence of distant HGT. To further prove that there may be distant HGT we looked at the GC% of the gene. The GC% for the entire genome is 63%, and the GC% of this specific gene is 60%. The

difference in percentages is less than 5%, so this provides evidence against recent HGT. Therefore, it was concluded that there is no recent HGT, but there is still a possibility of distant HGT.

4.4 Final Note on HGT

The pstB and pstC genes of *M. ruber* provide significantly greater evidence of HGT than do the pstA and pstS genes, as noted by their stark variations in GC% from the genomic average and the greater variation in their respective phylogenetic trees. Since HGT is more likely to occur as the transfer of multiple genes (such as an operon) as compared to singular genes, and considering the fact that the gene neighborhood map of the *M. ruber* Pst system shows conservation of the operon across unrelated species, it is proposed that at some point in the past *M. ruber* obtained the Pst operon via HGT. It appears that this did not happen too far into the past, however, as the operon is still in the process of being accommodated to a suitable GC content—all of the genes of the Pst operon have below average GC%.

4.5 Phosphate Homeostasis is More Nuanced Than Originally Thought

Although it has been concluded that the two Pst systems in *M. ruber* and *E. coli* are functionally similar, they do not seem to be regulated in the same way. The conserved operon in *E. coli* consists of a fifth gene, phoU, that encodes a repressor protein for the pho (phosphate) regulon, to which the Pst system is a part of (the yellow gene in Fig A.4a)—the regulon also consists of two other genes, phoR and phoB. A regulon is a system of multiple operons that are regulated by the same regulatory protein, which in this case is phoU. However, phoU is not in the Pst operon in *M. ruber*, which is alarming. This potentially means that the Pst system in *M. ruber* is not regulated in the same manner as it is in *E. coli*. Even if the phoU repressor gene were to be located at a different locus, the entirely of its mechanism of inhibition would not be the same as it is in *E. coli* since it is not coordinately transcribed with the Pst operon. However, the real question, then, is if *M. ruber* truly needs to regulate its phosphate intake. It seems that, given that is grows in a nutrient-poor environment, *M. ruber* would have need to ever *refuse* the intake of nutrients as precious as phosphates. (Webb et al., 1992)

All things considered, this is still a simplification of phosphate homeostasis, as *E. coli* contains a second system (the Pit system) that is involved (Webb et al., 1992). As such, more research needs to be carried out if we are to fully understand phosphate homeostasis in *M. ruber*.

5. CONCLUSION

All four *M. ruber* genes (Mrub_2518 (pstS), Mrub_2519 (pstC), Mrub_2520 (pstA), and Mrub_2521 (pstB)) are homologues of their respective *E. coli* genes (b3728, b3727, b3726, b3725, respectively); therefore, all four of the *M. ruber* genes code for the proposed proteins in the phosphate ABC transport system (PstA,B,C,S). Insofar as the Pst system, *M. ruber* transports P_i into its cytoplasm in the same manner as *E. coli.* However, we are left unsure as to the relevance of the pho regulon in *M. ruber***.** With respect to signs of HGT, the four genes in *E. coli* do not show evidence of HGT whatsoever. In *M. ruber*, even though there is no strong evidence suggesting recent HGT, there is moderate evidence that suggests distant HGT for the entire operon. Overall, it is suspected that there was a HGT of the Pst operon in the past, and the genes in the operon are in the process of accommodation (to a more suitable GC content).

References

(1) Nobre, M.F., Trüper, H.G., Da Costa, M.S. Transfer of *Thermus ruber* (Loginova et al. 1984), *Themus silvanus* (Tenreiro et al. 1999, and *Themus chliarophilus* (Tenreiro et al. 1995) to *Meiothemzus* gen. nov. as *Meiothermus ruber* comb. nov., *Meiothermus silvanus* comb. nov., and *Meiothermus chliarophilus* comb. nov., Respectively, and Emendation of the Genus *Thermus*. *Int. J. Syst. Evol. Microbiol.* **1996** Apr; 46(2): 604-6.

(2) Webb DC, Rosenberg H, Cox. GB. Mutational analysis of the Escherichia coli phosphate-specific transport system, a member of the traffic ATPase (or ABC) family of membrane transporters. A role for proline residues in transmembrane helices. *J Biol Chem*. **1992** Dec 5; 267(34): 24661-8.

(3) Moussatova A, Kandt C. ATP-binding cassette transporters in Escherichia coli. *Biochimica et Biophysica Acta*. **2008** January 9.

(4) Berg, J.M., Tymoczko J.L., Stryer, L. *Biochemistry*, 7th ed.; W.H. Freeman and Company: New York, 2012.

Appendix A

b3725 and Mrub_2521

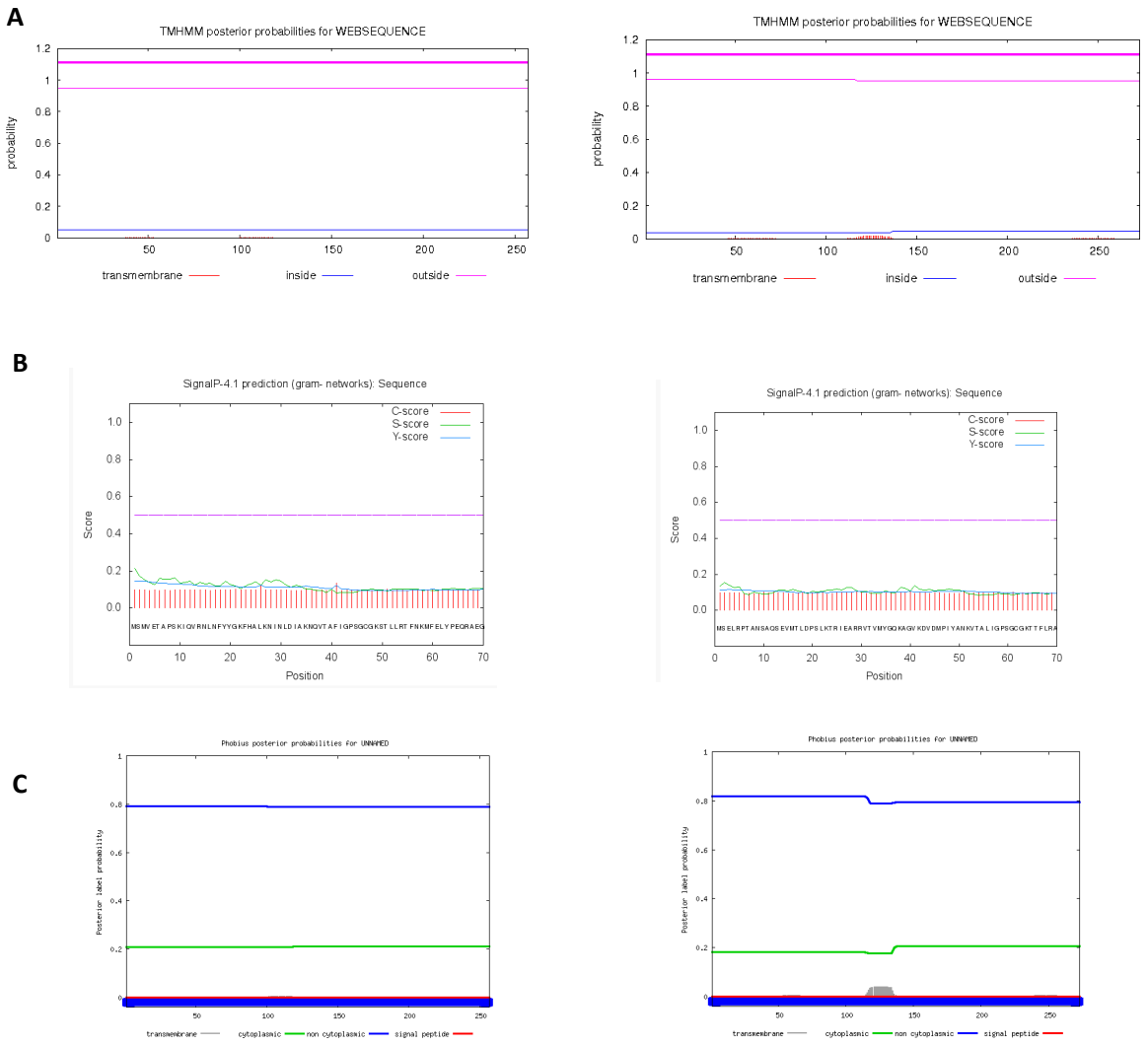


Fig. A.1 Cellular localization bioinformatics tools predict that b3725 (left) and Mrub_2521 (right) are either cytoplasmic or non-cytoplasmic proteins. (a) TMHMM predicts that neither gene has transmembrane helices, (b) SignalP predicts that neither gene contain a signal peptide sequence, (c) Phobius predicts that both gene products are non-cytoplasmic. The discrepancy observed can be explained by the gene products being peripheral proteins that are neither free-floating cytoplasmic proteins nor transmembrane helices-containing integral proteins.

A

Range 1: 16 to 273 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
290 bits(741)	6e-102	Compositional matrix adjust.	137/258(53%)	181/258(70%)	1/258(0%)

```
Query  1    MSMVETAPSKIQVRNLNFYYGKFHALKNINLDIAKNQVTAFIGPSGCGKSTLLRTFNKMF  60
            M++  +  ++I+ R +   YG+   +K++++ I  N+VTA IGPSGCGK+T LR  N+M
Sbjct  16   MTLDPSLKTRIEARRVTVMYGQKAGVKDVDMPIYANKVTALIGPSGCGKTTFLRALNRMH  75

Query  61   ELYPEQRAEGEILLDGDNILTNSQDIALLRAKVGMVFQKPTPFP-MSIYDNIAFGVRLFE  119
            +L P  R  GE+LLDG N+     D   +R K+GMVFQKP PFP +SIY N+  G+RL
Sbjct  76   DLTPSARVTGEVLLDGINVYAPGVDPVEVRRKIGMVFQKPNPFPTLSIYGNVVAGLRLVG  135

Query  120  KLSRADMDERVQWALTKAALWNETKDKLHQSGYSLSGGQQQRLCIARGIAIRPEVLLLDE  179
              ++ +DE V+ AL++AALW+E KD+L+    SLSGGQQQRLCIAR +A+ PEVLL+DE
Sbjct  136  IRKKSILDEAVERALSQAALWDEVKDRLNAPSMSLSGGQQQRLCIARALAVEPEVLLMDE  195

Query  180  PCSALDPISTGRIEELITELKQDYTVVIVTHNMQQAARCSDHTAFMYLGELIEFSNTDDL  239
            P SALDPIST  IE+L+ ELK   T+VIVTHNMQQA R SD T +   GE++EF  T  L
Sbjct  196  PTSALDPISTQSIEDLLNELKNHVTIVIVTHNMQQAGRVSDFTGYFLNGEMVEFGPTSLL  255

Query  240  FTKPAKKQTEDYITGRYG  257
            FT P  K+TE YITGR+G
Sbjct  256  FTTPKDKRTEAYITGRFG  273
```

B

phosphate ABC transporter ATP-binding protein [Shigella sonnei]

Sequence ID: ref|WP_000063122.1| **Length:** 257 **Number of Matches:** 1

▶ See 2 more title(s)

Range 1: 1 to 257 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
536 bits(1381)	0.0	Compositional matrix adjust.	256/257(99%)	256/257(99%)	0/257(0%)

```
Query  1    MSMVETAPSKIQVRNLNFYYGKFHALKNINLDIAKNQVTAFIGPSGCGKSTLLRTFNKMF  60
            MSMVETAPSKIQVRNLNFYYGKFHALKNINLDIAKNQVTAFIGPSGCGKSTLLRTFNKMF
Sbjct  1    MSMVETAPSKIQVRNLNFYYGKFHALKNINLDIAKNQVTAFIGPSGCGKSTLLRTFNKMF  60

Query  61   ELYPEQRAEGEILLDGDNILTNSQDIALLRAKVGMVFQKPTPFPMSIYDNIAFGVRLFEK  120
            ELYPEQRAEGEILLD DNILTNSQDIALLRAKVGMVFQKPTPFPMSIYDNIAFGVRLFEK
Sbjct  61   ELYPEQRAEGEILLDDDNILTNSQDIALLRAKVGMVFQKPTPFPMSIYDNIAFGVRLFEK  120

Query  121  LSRADMDERVQWALTKAALWNETKDKLHQSGYSLSGGQQQRLCIARGIAIRPEVLLLDEP  180
            LSRADMDERVQWALTKAALWNETKDKLHQSGYSLSGGQQQRLCIARGIAIRPEVLLLDEP
Sbjct  121  LSRADMDERVQWALTKAALWNETKDKLHQSGYSLSGGQQQRLCIARGIAIRPEVLLLDEP  180

Query  181  CSALDPISTGRIEELITELKQDYTVVIVTHNMQQAARCSDHTAFMYLGELIEFSNTDDLF  240
            CSALDPISTGRIEELITELKQDYTVVIVTHNMQQAARCSDHTAFMYLGELIEFSNTDDLF
Sbjct  181  CSALDPISTGRIEELITELKQDYTVVIVTHNMQQAARCSDHTAFMYLGELIEFSNTDDLF  240

Query  241  TKPAKKQTEDYITGRYG  257
            TKPAKKQTEDYITGRYG
Sbjct  241  TKPAKKQTEDYITGRYG  257
```

C

hypothetical protein [Meiothermus taiwanensis]

Sequence ID: ref|WP_027887392.1| **Length:** 258 **Number of Matches:** 1

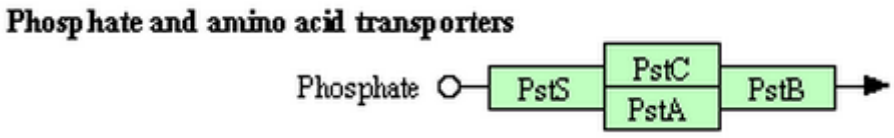
Range 1: 1 to 258 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
510 bits(1314)	0.0	Compositional matrix adjust.	247/258(96%)	251/258(97%)	0/258(0%)

```
Query  16   MTLDPSLKTRIEARRVTVMYGQKAGVKDVDMPIYANKVTALIGPSGCGKTTFLRALNRMH  75
            MTLDPSLKTRIE R VTV+YGQKAGVKDV MPIYANKVTALIGPSGCGKTTFLR+LNRMH
Sbjct  1    MTLDPSLKTRIETRHVTVIYGQKAGVKDVTMPIYANKVTALIGPSGCGKTTFLRSLNRMH  60

Query  76   DLTPSARVTGEVLLDGINVYAPGVDPVEVRRKIGMVFQKPNPFPTLSIYGNVVAGLRLVG  135
            DLTPSARVTGEVLLDGINVYA GVDPVEVRRKIGMVFQKPNPFPTLSIYGNVVAGLRLVG
Sbjct  61   DLTPSARVTGEVLLDGINVYAAGVDPVEVRRKIGMVFQKPNPFPTLSIYGNVVAGLRLVG  120

Query  136  IRKKSILDEAVERALSQAALWDEVKDRLNAPSMSLSGGQQQRLCIARALAVEPEVLLMDE  195
            IRKKS+LDEA ERAL QAALWDEVKDRL+APSMSLSGGQQQRLCIARALAVEPEVLLMDE
Sbjct  121  IRKKSLLDEAAERALRQAALWDEVKDRLHAPSMSLSGGQQQRLCIARALAVEPEVLLMDE  180

Query  196  PTSALDPISTQSIEDLLNELKNHVTIVIVTHNMQQAGRVSDFTGYFLNGEMVEFGPTSLL  255
            PTSALDPISTQSIEDLL ELKNHVTIVIVTHNMQQAGRVSDFTGYFLNGEMVEFGPTSLL
Sbjct  181  PTSALDPISTQSIEDLLTELKNHVTIVIVTHNMQQAGRVSDFTGYFLNGEMVEFGPTSLL  240

Query  256  FTTPKDKRTEAYITGRFG  273
            FTTPKDKRTEAYITGRFG
Sbjct  241  FTTPKDKRTEAYITGRFG  258
```

Fig. A.2 BLAST alignments for b3725 and Mrub_2521. (a) NCBI BLAST of b3725 against Mrub_2521, (b) top BLAST hit for b3725, (c) top BLAST hit for Mrub_2521.

A



B

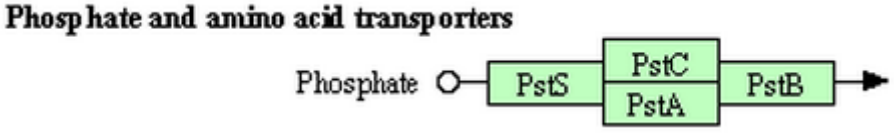
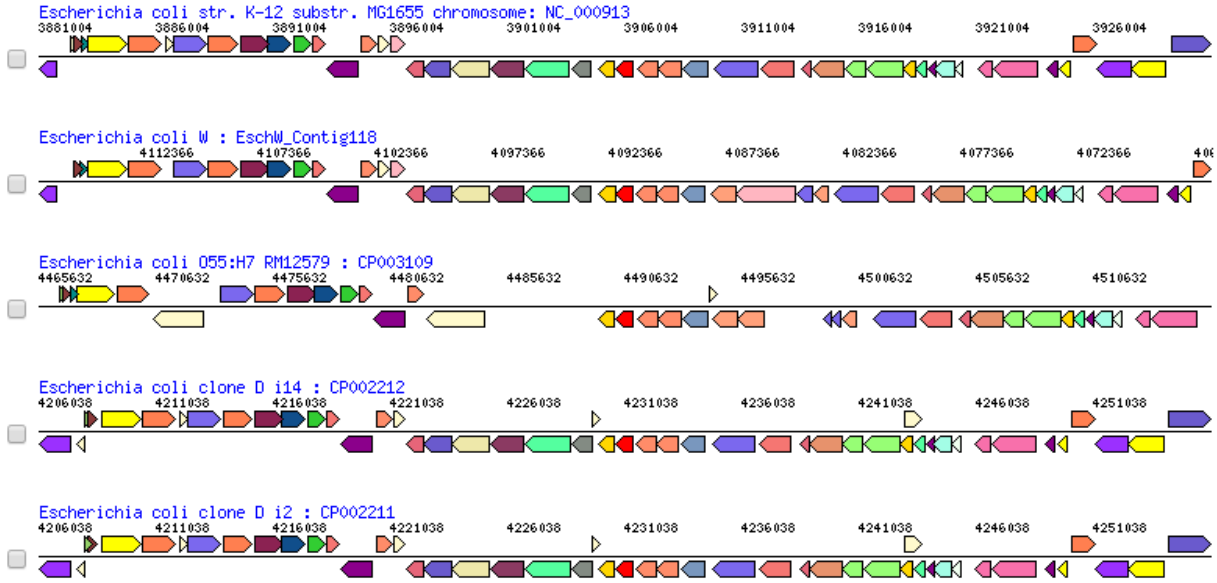


Fig. A.3 All of the genes analyzed belong to the same KEGG pathway. (a) KEGG pathway ecd02010 for *E. coli* and (b) KEGG pathway mrb02010 for *M. ruber*. b3725 and pstB and M. ruber corresponnd to pstB; b3726 and Mrub_2520 correspond to pstA; b3727and Mrub_2519 correspond to pstC; b3728 and Mrub_2518 correspond to pstS.

A



B

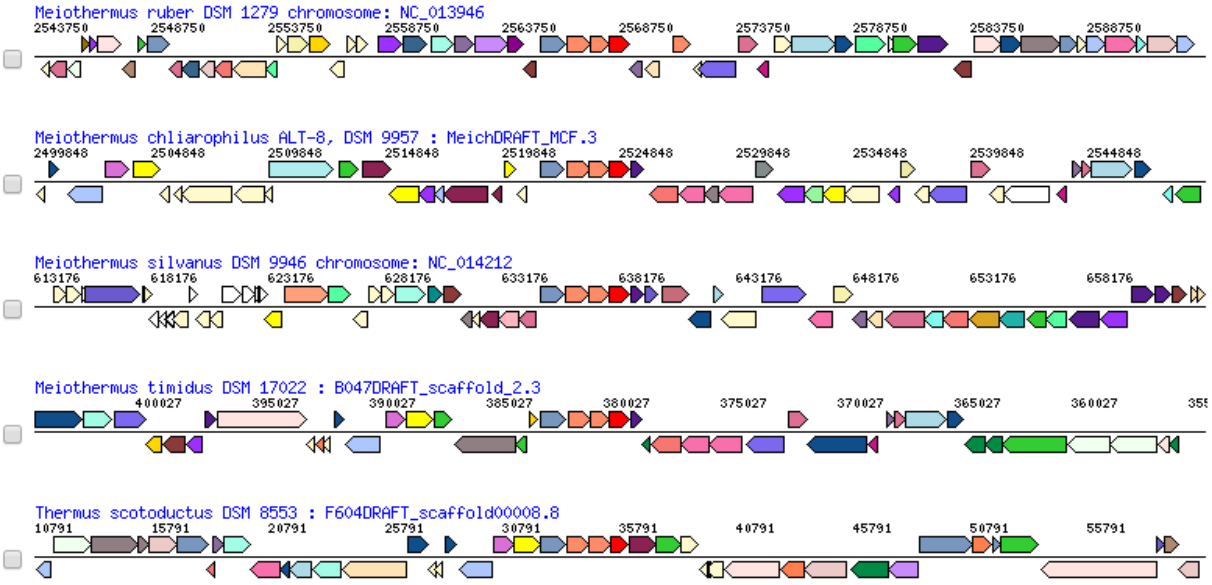
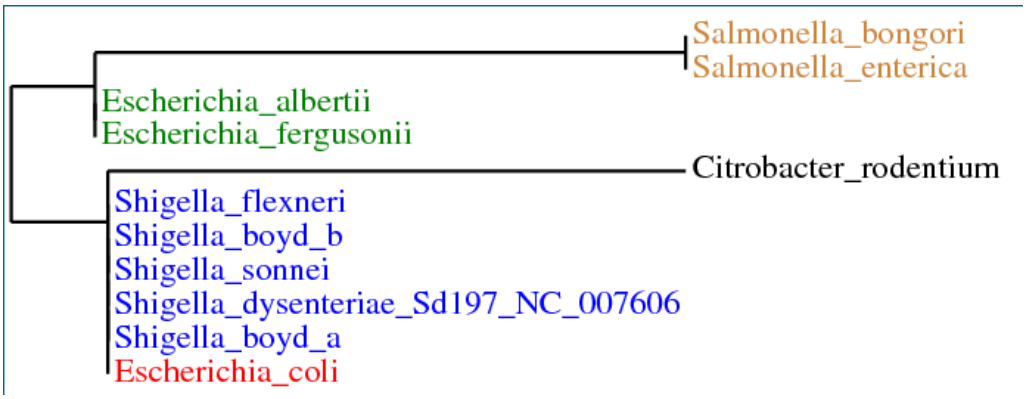


Fig. A.4 Gene neighborhood maps (NCBI BLAST) show that the Pst system is a conserved operon in *M. ruber* and an operon that is part of a regulon in *E. coli*. (a) Gene neighborhood map for the Pst system in *E. coli* (with b3725 in red). All four genes of the Pst system are contained in this operon, plus an additional gene (yellow), phoU, which is part of the pho regulon. (b) Gene neighborhood map for the Pst system in *M. ruber* (with Mrub_2521 in red). All four putative genes of the Pst system are contained in this operon.

A



B

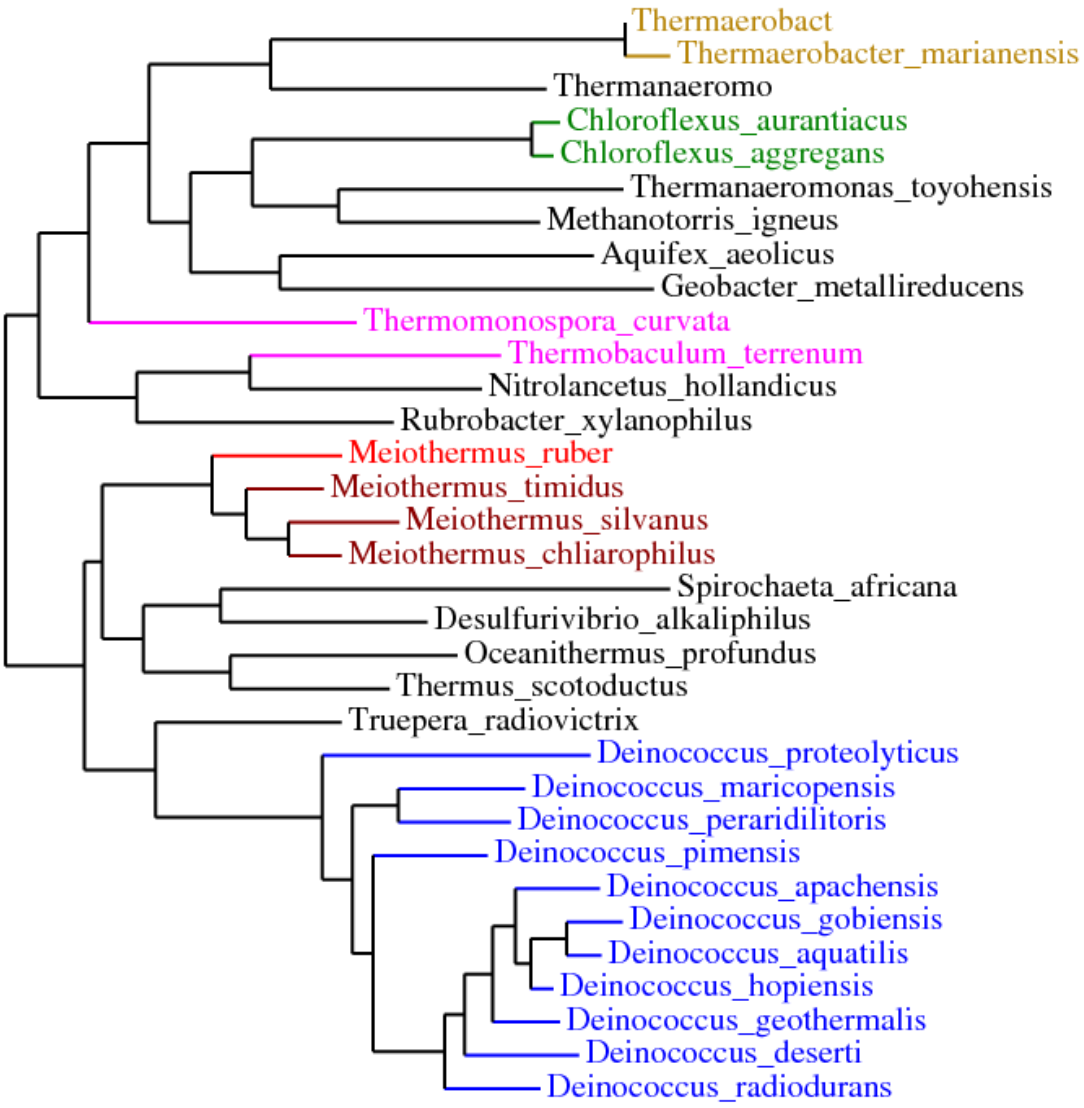


Fig. A.5 Phylogenetic tree (Phylogeny.fr) for b3725 (a) shows no sign of HGT, while the tree for Mrub_2521 hints at the possibility of distant HGT.

Appendix B

b3726 and Mrub_2520

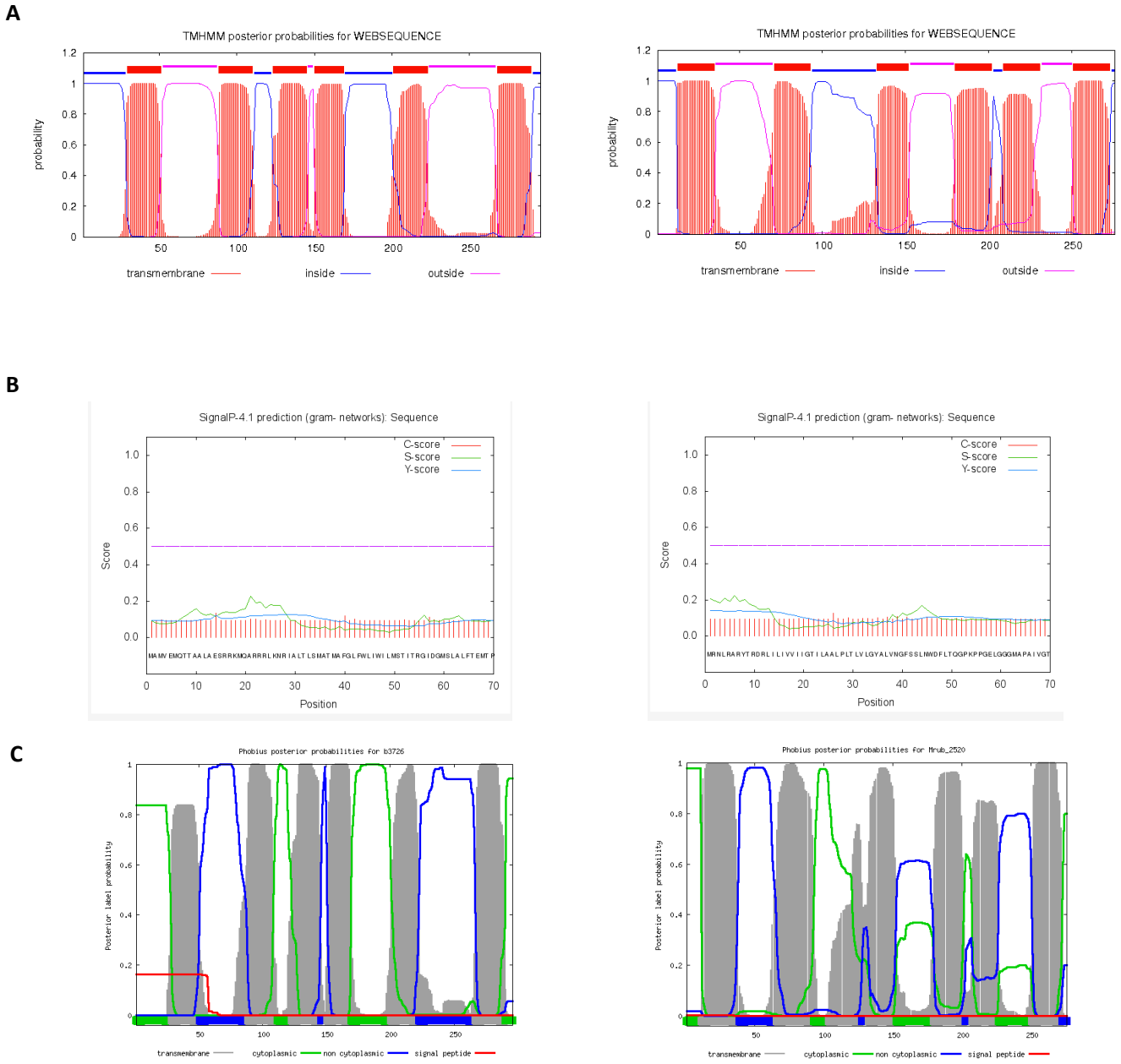


Fig. B.1 Cellular localization bioinformatics tools predict that b3726 (left) and Mrub_2520 (right) are transmembrane proteins. (a) TMHMM predicts that both genes have six transmembrane helices, (b) SignalP predicts that neither gene contain a signal peptide sequence, (c) Phobius predicts that both gene products are non-cytoplasmic and contain six transmembrane helices.

A

unnamed protein product

Sequence ID: lcl|20075 Length: 276 Number of Matches: 1

Range 1: 2 to 249 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
160 bits(405)	4e-51	Compositional matrix adjust.	92/250(37%)	152/250(60%)	2/250(0%)

```
Query  17   RKMQARRRLKNRIALTLSMATMAFGLFWLIWILMSTITRGIDGMSLALFTEMTPPPNTEG  76
            R ++AR   ++R+ L + +         L  +L   +  G   ++    T+   PP   G
Sbjct  2    RNLRARYT-RDRLILIVVIIGTILAALPLTLVLGYALVNGFSSLNWDFLTQGPKPPGELG  60

Query  77   GGLANALAGSGLLILWATVFGTPLGIMAGIYLAEYGRKSWLAEVIRFINDILLSAPSIVV  136
            GG+A A+ G+ ++     +  TP GI AGI LAEY     L   +R ++D L   P+I+
Sbjct  61   GGMAPAIVGTLIITGAGLLMATPFGIGAGILLAEYPDNK-LNPTLRLLSDTLNGMPAILK  119

Query  137  GLFVYTIVVAQMEHFSGWAGVIALALLQVPIVIRTTENMLKLVPYSLREAAYALGTPKWK  196
            GL  Y +VV     FSG +G +A+A + +PI+ +TTE++LKLVP+++REA  ALG P+W+
Sbjct  120  GLLAYVLVVKAQGSFSGLSGALAMAFIMIPIIAKTTESVLKLVPWNIREAGLALGLPRWR  179

Query  197  MISAITLKASVSGIMTGILLAIARIAGETAPLLFTALSNQFWSTDMMQPIANLPVTIFKF  256
            +I ++ L A+ +G++TG+LLA AR AGE APL+FTA  N   + +++QP+  LP+ ++ +
Sbjct  180  VIMSLVLPAARAGVVTGLLLATARAAGEAAPLIFTAFGNSLLTYNLLQPMDALPLRLYAY  239

Query  257  AMSPFAEWQQ  266
            A+SP+ +W +
Sbjct  240  AISPYEDWHR  249
```

B

phosphate transporter permease subunit PtsA [Escherichia albertii]

Sequence ID: ref|WP_001251990.1| Length: 296 Number of Matches: 1

▶ See 3 more title(s)

Range 1: 1 to 296 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
583 bits(1503)	0.0	Compositional matrix adjust.	294/296(99%)	294/296(99%)	0/296(0%)

```
Query  1    MAMVEMQTTAALAESRRKMQARRRLKNRIALTLSMATMAFGLFWLIWILMSTITRGIDGM  60
            MAMVEMQTTAALAESRRKMQARRRLKNRIALTLSMATMAFGLFWLIWILMSTITRGIDGM
Sbjct  1    MAMVEMQTTAALAESRRKMQARRRLKNRIALTLSMATMAFGLFWLIWILMSTITRGIDGM  60

Query  61   SLALFTEMTPPPNTEGGGLANALAGSGLLILWATVFGTPLGIMAGIYLAEYGRKSWLAEV  120
            SLALFTEMTPPPNTEGGGLANALAGSGLLILWATVFGTPLGIMAGIYLAEYGRKSWLAEV
Sbjct  61   SLALFTEMTPPPNTEGGGLANALAGSGLLILWATVFGTPLGIMAGIYLAEYGRKSWLAEV  120

Query  121  IRFINDILLSAPSIVVGLFVYTIVVAQMEHFSGWAGVIALALLQVPIVIRTTENMLKLVP  180
            IRFINDILLSAPSIVVGLFVYTIVVAQMEHFSGWAGVIALALLQVPIVIRTTENMLKLVP
Sbjct  121  IRFINDILLSAPSIVVGLFVYTIVVAQMEHFSGWAGVIALALLQVPIVIRTTENMLKLVP  180

Query  181  YSLREAAYALGTPKWKMISAITLKASVSGIMTGILLAIARIAGETAPLLFTALSNQFWST  240
             SLREAAYALGTPKWKMISAITLKASVSGIMTGILLAIARIAGETAPLLFTALSNQFWST
Sbjct  181  DSLREAAYALGTPKWKMISAITLKASVSGIMTGILLAIARIAGETAPLLFTALSNQFWST  240

Query  241  DMMQPIANLPVTIFKFAMSPFAEWQQLAWAGVLIITLCVLLLNILARVVFAKNKHG  296
            DMMQPIANLPVTIFKFAMSPFAEWQQLAWAGVLIITLCVLLLNILARVVFAK KHG
Sbjct  241  DMMQPIANLPVTIFKFAMSPFAEWQQLAWAGVLIITLCVLLLNILARVVFAKKKHG  296
```

C

hypothetical protein [Meiothermus taiwanensis]

Sequence ID: ref|WP_027887393.1| Length: 276 Number of Matches: 1

Range 1: 1 to 276 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
486 bits(1250)	6e-171	Compositional matrix adjust.	259/276(94%)	270/276(97%)	0/276(0%)

```
Query  1    MRNLRARYTRDRLILIVVIIGTILAALPLTLVLGYALVNGFSSLNWDFLTQGPKPPGELG  60
            MR+L+ARYTRDRLILIVVI GTILAALPLTLVLGYALVNGFSSLNWDF T+GPKPPGELG
Sbjct  1    MRSLQARYTRDRLILIVVIAGTILAALPLTLVLGYALVNGFSSLNWDFFTKGPKPPGELG  60

Query  61   GGMAPAIVGTLIITGAGLLMATPFGIGAGILLAEYPDNKLNPTLRLLSDTLNGMPAILKG  120
            GGMAPAIVGTLIITG GLL+ATPFGI AGILLAEYP+NKLNPTLRLLSDTLNGMPA+LKG
Sbjct  61   GGMAPAIVGTLIITGTGLLIATPFGIAAGILLAEYPENKLNPTLRLLSDTLNGMPAVLKG  120

Query  121  LLAYVLVVKAQGSFSGLSGALAMAFIMIPIIAKTTESVLKLVPWNIREAGLALGLPRWRV  180
            LLAYVLVVKAQGSFSGLSGA+AMAFIM+PIIAKTTESVLKLVPWNIREAGLALGLPRWRV
Sbjct  121  LLAYVLVVKAQGSFSGLSGAVAMAFIMVPIIAKTTESVLKLVPWNIREAGLALGLPRWRV  180

Query  181  IMSLVLPAARAGVVTGLLLATARAAGEAAPLIFTAFGNSLLTYNLLQPMDALPLRLYAYA  240
            I+SLVLPAARAGVVTGLLLATARAAGEAAPLIFTAFGNSLLTYNLLQPMD LPLRLYAYA
Sbjct  181  ILSLVLPAARAGVVTGLLLATARAAGEAAPLIFTAFGNSLLTYNLLQPMDTLPLRLYAYA  240

Query  241  ISPYEDWHRQAWAAALVLLALIVLTSLVARWVTRGR  276
            ISPYEDWHRQAWAAA+VLL LIVLTSL+ARWVTRGR
Sbjct  241  ISPYEDWHRQAWAAAVVLLGLIVLTSLLARWVTRGR  276
```

Fig. B.2 BLAST alignments for b3726 and Mrub_2520. (a) NCBI BLAST of b3726 against Mrub_2520, (b) top BLAST hit for b3726, (c) top BLAST hit for Mrub_2520.

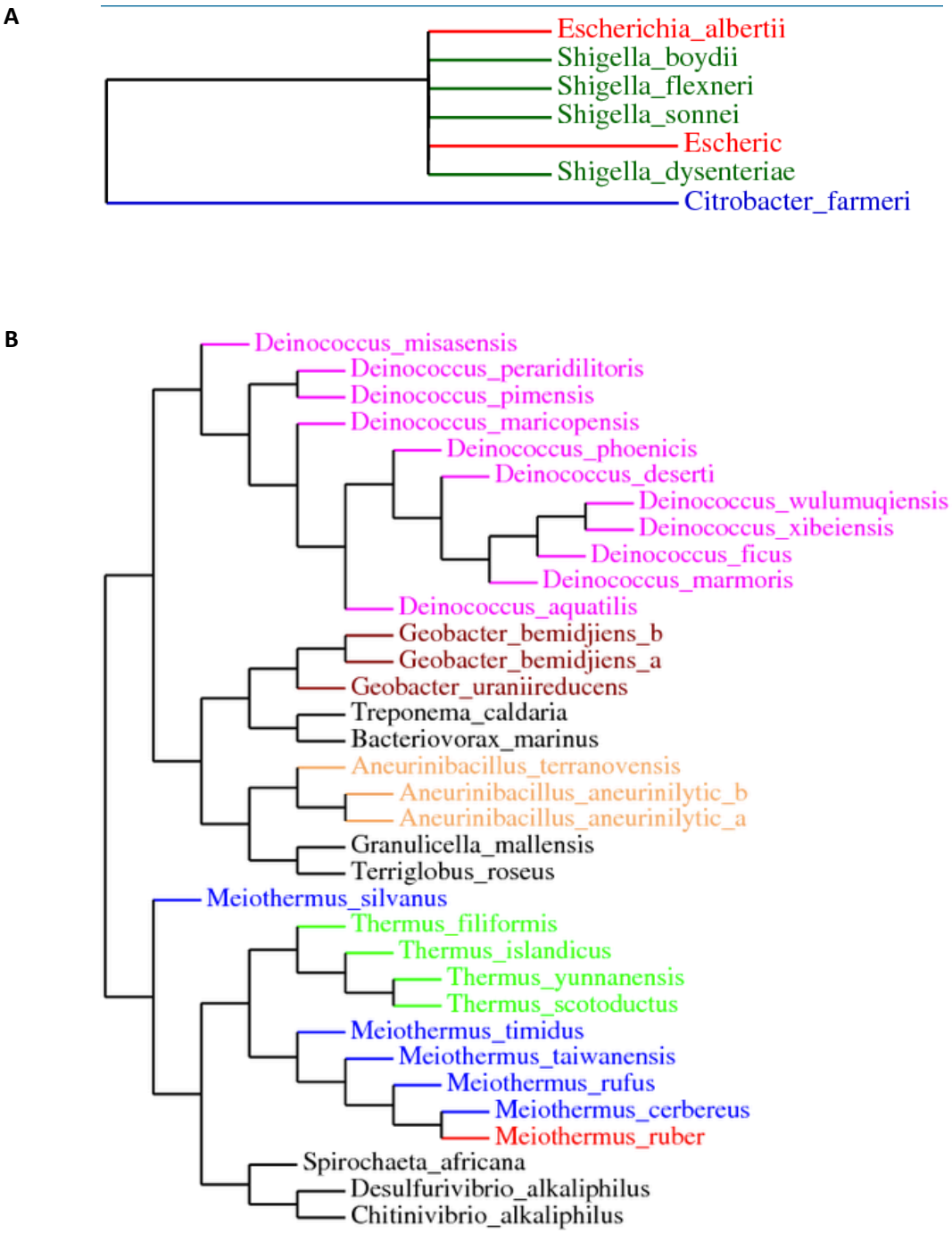


Fig. B.3 Phylogenetic tree (Phylogeny.fr) for b3726 (a) shows no sign of HGT, while the tree for Mrub_2520 hints at the possibility of distant HGT.

Appendix C

b3727 and Mrub_2519

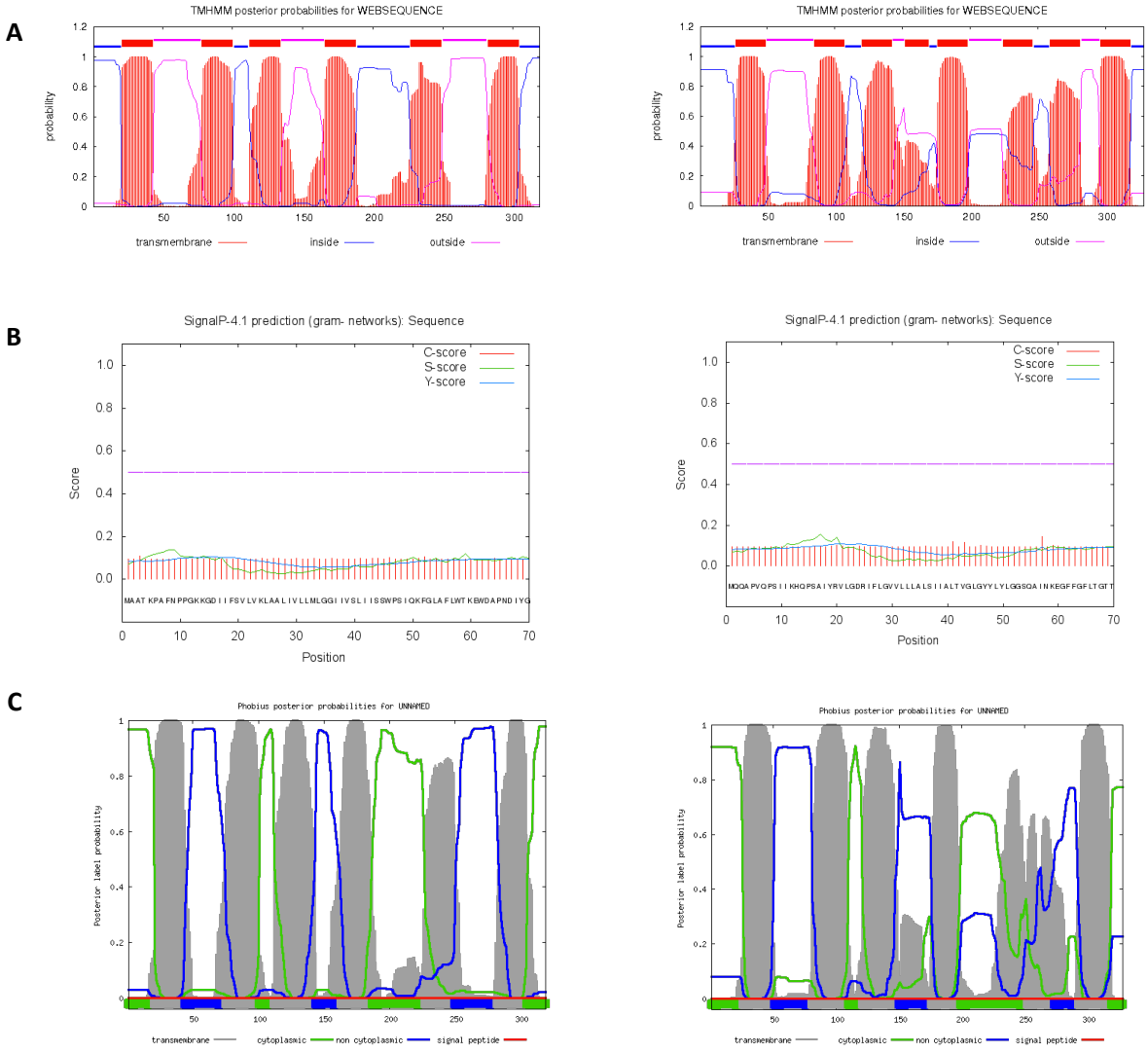


Fig. C.1 Cellular localization bioinformatics tools predict that b3727 (left) and Mrub_2519 (right) are transmembrane proteins. (a) TMHMM predicts that b3727 has six transmembrane helices and Mrub_2519 has eight transmembrane helices, (b) SignalP predicts that neither gene contain a signal peptide sequence, (c) Phobius predicts that both b3727 and Mrub_2519 have six transmembrane helices.

A

Score	Expect	Method	Identities	Positives	Gaps	Frame
175 bits(444)	1e-57()	Compositional matrix adjust.	101/246(41%)	150/246(60%)	4/246(1%)	

Features:

```
Query  11   IYGTLVTSFIALLIAVPVSFGIALFLTELAPGWLKRPLGIAIELLAAIPSIVYGMWGLFI  70
            + GT++TS  AL+++VPV+   A+F  E AP WL   +   ++L+AA+PS+VYG+WG+F+
Sbjct  21   VLGTIITSLAALVLSVPVALAAAIFTAEYAPRWLAGFINYLVDLMAAVPSVVYGIWGIFV  80

Query  71   FAPLFAVYFQEPVG-NIMSNIPIVGALFSGPAFGIGILAAGVILAIMIIPYIAAVMRDVF  129
              AP     F  P       N P +      PA G G+  A VIL+ M+IP+ AA+ RD
Sbjct  81   LAPFLREVFYMPFFLWAAENAPWLSRYLGNPA-GYGMFTAIVILSSMVIPFTAALSRDAI  139

Query  130  EQTPVMMKESAYGIGCTTWEVIWRIVLPFTKNGVIGGIMLGLGRALGETMAVTFIIGNTY  189
               PV  +E AY +G T WEV+  ++LP+ + G+  G ML LGRALGETMAV  +IGN
Sbjct  140  ALVPVAQREGAYALGATRWEVMQTVILPYARGGIFAGAMLALGRALGETMAVAMVIGNGN  199

Query  190  QLDSASLYMPGNSITSALANEFAEAESGLHVAALMELGLILFVITFIVLAASKFMIMRLA  249
              L   +L+ P +++ + +A E  EA   LH +A++ +G  LF+I FIV AA+ +++ +L
Sbjct  200  ILPY-TLFGPASTMPAVIALELKEAVEDLHYSAIIGVGFYLFLIAFIVNAAASYLLNKL-  257

Query  250  KNEGAR  255
            K  G R
Sbjct  258  KVGGQR  263
```

B

phosphate transporter permease subunit PstC [Salmonella enterica]

Sequence ID: ref|WP_000741613.1| Length: 319 Number of Matches: 1

▷ See 1 more title(s)

Range 1: 1 to 319 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
613 bits(1580)	0.0	Compositional matrix adjust.	313/319(98%)	315/319(98%)	0/319(0%)

```
Query  1    MAATKPAFNPPGKKGDIIFSVLVKLAALIVLLMLGGIIVSLIISSWPSIQKFGLAFLWTK  60
            MAATKPAFNPPGKKGDIIFS LVKLAALIVLLMLGGIIVSLIISSWPSIQKFG +FLWTK
Sbjct  1    MAATKPAFNPPGKKGDIIFSALVKLAALIVLLMLGGIIVSLIISSWPSIQKFGFSFLWTK  60

Query  61   EWDAPNDIYGALVPIYGTLVTSFIALLIAVPVSFGIALFLTELAPGWLKRPLGIAIELLA  120
            EWDAPNDIYGALVPIYGTLVTSFIALLIAVPVSFGIALFLTELAPGWLKRPLGIAIELLA
Sbjct  61   EWDAPNDIYGALVPIYGTLVTSFIALLIAVPVSFGIALFLTELAPGWLKRPLGIAIELLA  120

Query  121  AIPSIVYGMWGLFIFAPLFAVYFQEPVGNIMSNIPIVGALFSGPAFGIGILAAGVILAIM  180
            AIPSIVYGMWGLFIFAPLFA YFQEPVGNI+SNIP VGALFSGPAFGIGILAAGVILAIM
Sbjct  121  AIPSIVYGMWGLFIFAPLFATYFQEPVGNILSNIPFVGALFSGPAFGIGILAAGVILAIM  180

Query  181  IIPYIAAVMRDVFEQTPVMMKESAYGIGCTTWEVIWRIVLPFTKNGVIGGIMLGLGRALG  240
            IIPYIAAVMRDVFEQTPVMMKESAYGIGCTTWEVIWRIVLPFTKNGVIGGIMLGLGRALG
Sbjct  181  IIPYIAAVMRDVFEQTPVMMKESAYGIGCTTWEVIWRIVLPFTKNGVIGGIMLGLGRALG  240

Query  241  ETMAVTFIIGNTYQLDSASLYMPGNSITSALANEFAEAESGLHVAALMELGLILFVITFI  300
            ETMAVTFIIGNTYQLDSASLYMPGNSITSALANEFAEAESGLHVAALMELGLILFVITFI
Sbjct  241  ETMAVTFIIGNTYQLDSASLYMPGNSITSALANEFAEAESGLHVAALMELGLILFVITFI  300

Query  301  VLAASKFMIMRLAKNEGAR  319
            VLAASKFMIMRLAKNEGAR
Sbjct  301  VLAASKFMIMRLAKNEGAR  319
```

C

Meiothermus cerbereus: phosphate ABC transporter permease

Score	Expect	Method	Identities	Positives	Gaps
553 bits(1424)	0.0	Compositional matrix adjust.	300/328(91%)	313/328(95%)	0/328(0%)

```
Query  1    MQQAPVQPSIIKHQPSAIYRVLGDRIFLGVVLLLALSIIALTVGLGYYLYLGGSQAINKE  60
            MQ+ P+Q  +IK +PSAIYRVLGDRIFL +VL LALSI+ LT+GLGYYLYLGGS AINKE
Sbjct  1    MQEVPIQSPLIKRKPSAIYRVLGDRIFLVLVLALALSIVLLTLGLGYYLYLGGSLAINKE  60

Query  61   GFFGFLTGTTWDPALKLEFGIWPYVLGTIITSLAALVLSVPVALAAAIFTAEYAPRWLAG  120
            G FGFLTGTTWDPALKLEFGIWPYV+GT++TSLAALVLSVPVALAAA+FTAEYAPRWLAG
Sbjct  61   GVFGFLTGTTWDPALKLEFGIWPYVVGTLVTSLAALVLSVPVALAAAVFTAEYAPRWLAG  120

Query  121  FINYLVDLMAAVPSVVYGIWGIFVLAPFLREVFYMPFFLWAAENAPWLSRYLGNPAGYGM  180
            FINYLVDLMAAVPSVVYGIWGIFVLAPFLRE+FYMP FLWAAENAPWLSRYLGNP+GYGM
Sbjct  121  FINYLVDLMAAVPSVVYGIWGIFVLAPFLREIFYMPVFLWAAENAPWLSRYLGNPSGYGM  180

Query  181  FTAIVILSSMVIPFTAALSRDAIALVPVAQREGAYALGATRWEVMQTVILPYARGGIFAG  240
            FTAIVILSSMVIPFTAALSRDAI LVP AQREGAYALGATRWEVMQ VILPYARGGIFAG
Sbjct  181  FTAIVILSSMVIPFTAALSRDAINLVPAAQREGAYALGATRWEVMQMVILPYARGGIFAG  240

Query  241  AMLALGRALGETMAVAMVIGNGNILPYTLFGPASTMPAVIALELKEAVEDLHYSAIIGVG  300
            AMLALGRALGETMAVAMVIGNGNILPYTLFGPASTMPAVIALELKEAVEDLHYSAIIGVG
Sbjct  241  AMLALGRALGETMAVAMVIGNGNILPYTLFGPASTMPAVIALELKEAVEDLHYSAIIGVG  300

Query  301  FYLFLIAFIVNAAASYLLNKLKVGGQRA  328
            FYLFLIAFIVNA ASYLLNKLKVG QRA
Sbjct  301  FYLFLIAFIVNATASYLLNKLKVGEQRA  328
```

Fig. C.2 BLAST alignments for b3727 and Mrub_2519. (a) NCBI BLAST of b3727 against Mrub_2519, (b) top BLAST hit for b3727, (c) top BLAST hit for Mrub_2519.

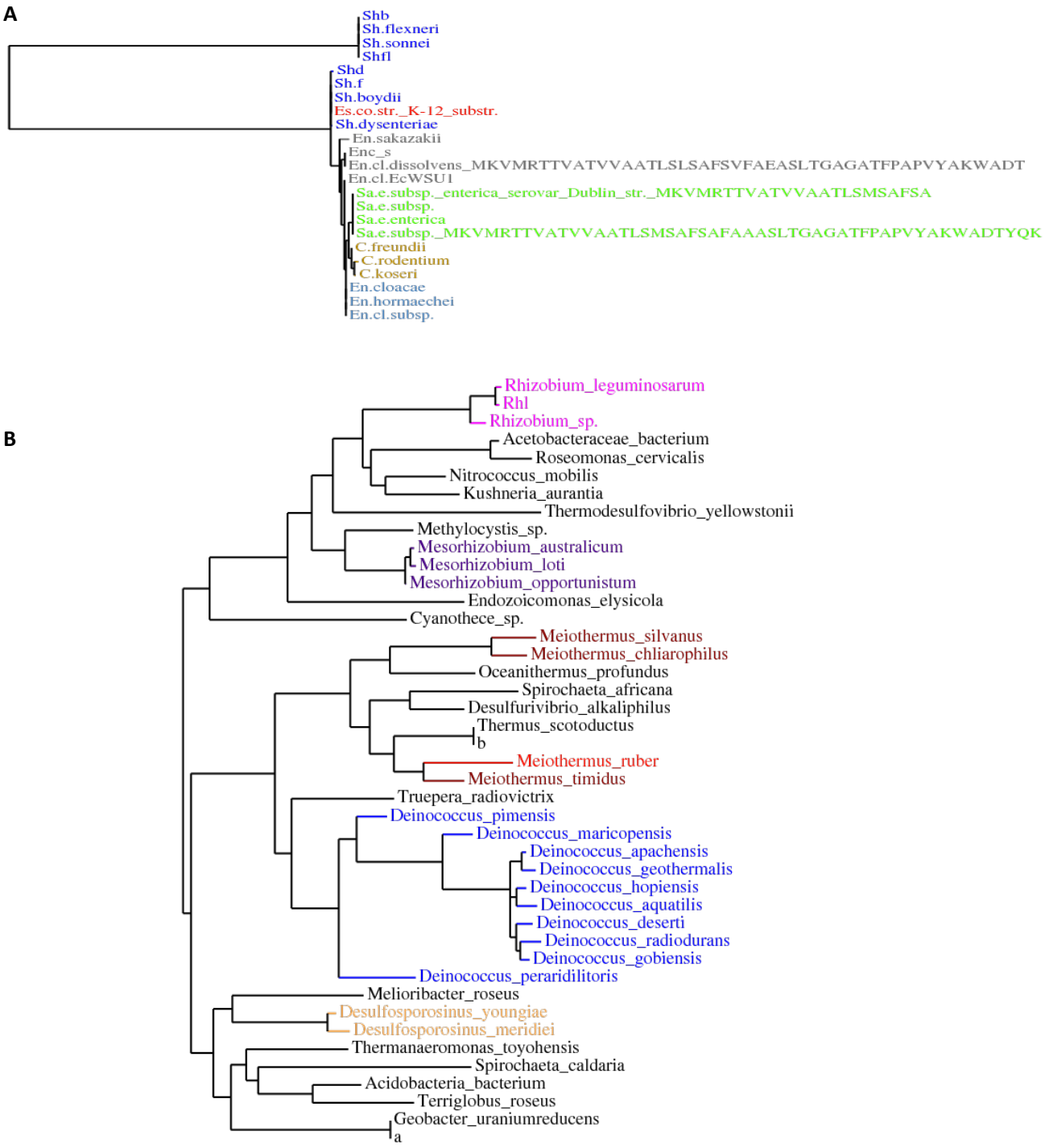


Fig. C.3 Phylogenetic tree (Phylogeny.fr) for b3727 (a) shows no sign of HGT, while the tree for Mrub_2519 strongly hints at the possibility of distant HGT.

Appendix D

b3728 and Mrub_2518

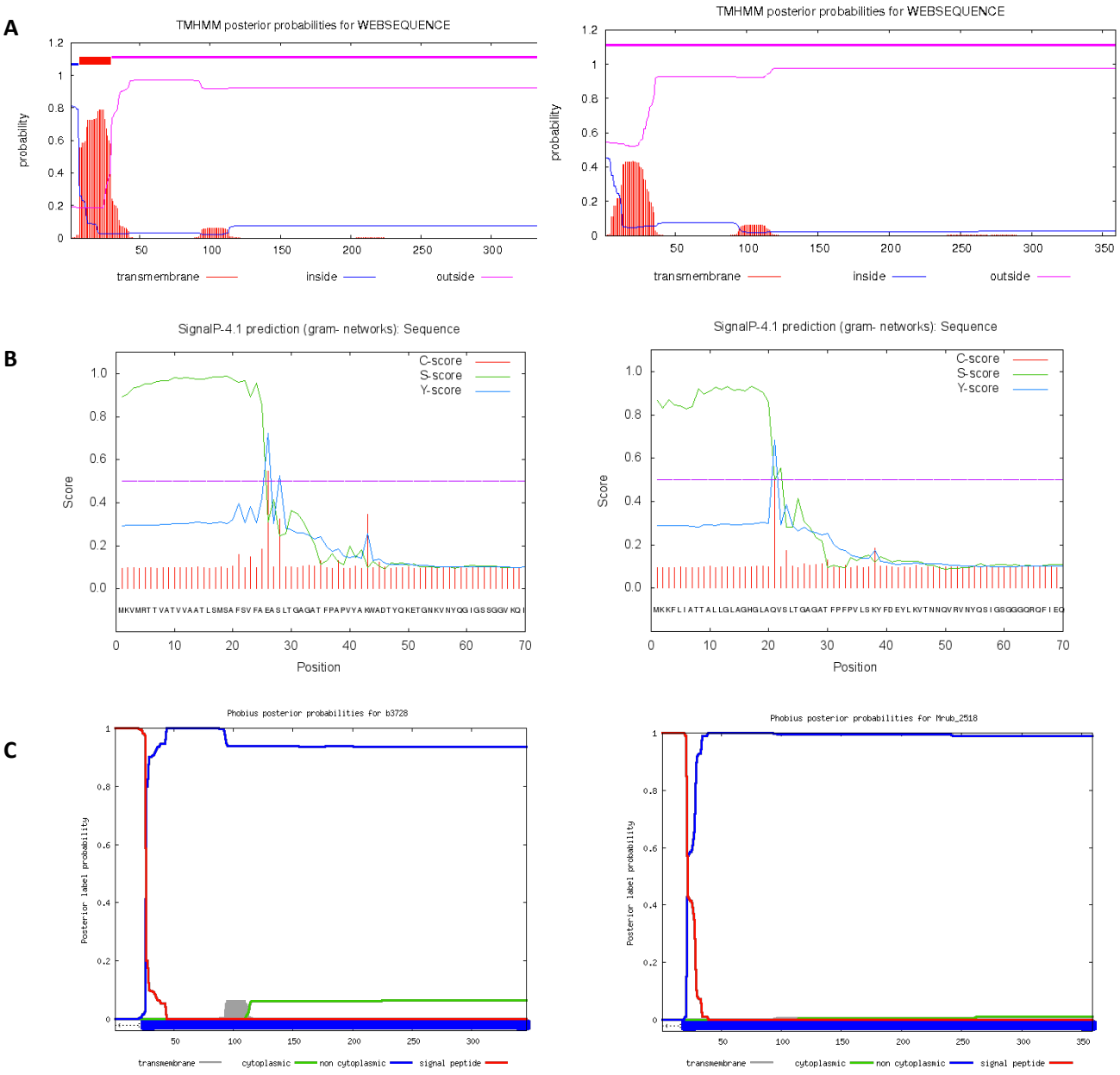


Fig. D.1 Cellular localization bioinformatics tools predict that b3728 (left) and Mrub_2518 (right) are periplasmic proteins. (a) TMHMM predicts that b3728 has a signal peptide and Mrub_2518 shows inconclusive evidence of a signal peptide sequence, (b) SignalP predicts that both genes contain a signal peptide sequence, (c) Phobius predicts that both genes have a signal peptide sequence.

A

Score	Expect	Method	Identities	Positives	Gaps	Frame
173 bits(438)	5e-56()	Compositional matrix adjust.	107/295(36%)	156/295(52%)	24/295(8%)	

Features:

```
Query  4    KQIIANTVDFGASDAPLSDEKLAQ------EGLFQFPTVIGGVVLAVNIPGLKSGELVLD  57
            +Q I  TV FGASD P +D+++A+            P V+G VV   N+PG+ +  L
Sbjct  1    RQFIEQTVHFGASDNPFNDQQMAEIRRNTGSPALNIPFVLGAVVPTYNLPGV-TQRLNFT  59

Query  58   GKTLGDIYLGKIKKWDDEAIAKLNPGLKLPSQNIAVVRRADGSGTSFVFTSYLAKVNEEW  117
            G+ L DI+LG IK W+D AIA+LN G++LP    I VV R+DGSGT++V+T YL KV+ EW
Sbjct  60   GEVLADIFLGNIKTWNDPAIARLNEGVRLPPLPITVVHRSDGSGTTYVWTDYLTKVSPEW  119

Query  118  KNNVGTGSTVKW--PIGLGGKGNDGIAAFVQRLPGAIGYVEYAYAKQNNLAYTKLISADG  175
             VG G++V W  P  +GG+GN+G+A  V+  PGAIGY E  YA QN + +  + +  G
Sbjct  120  AQKVGRGNSVNWLAPNKVGGRGNEGVAGVVRNTPGAIGYNEVTYAVQNRILFGAVQNRAG  179

Query  176  KPVS---PTEENFANAAKGADWSKTFAQDLTNQKGEDAWPITSTTFILIH------KDQK  226
            + +    P     AN     D   +    LTN +  D +PI S +++L++       K  K
Sbjct  180  RFMVADLPAIAAAANVVLPGDARVS----LTNTQAPDGYPIASFSYLLVYEQLDKNKAFK  235

Query  227  KPEQGTEVLKFFDWAYKTGAKQANDLDYASLPDSVVEQVRAAWKTNIKDSSGKPL  281
               +    ++   W    G K    L Y  L ++   Q RA    +     GKP+
Sbjct  236  SEAEARAFVQLLKWIVTEGQKYNEPLTYGRLTETA--QARALALISRITYQGKPI  288
```

B

phosphate ABC transporter substrate-binding protein [*Shigella boydii*]

Score	Expect	Method	Identities	Positives	Gaps
696 bits(1796)	0.0	Compositional matrix adjust.	345/346(99%)	346/346(100%)	0/346(0%)

```
Query  1    MKVMRTTVATVVAATLSMSAFSVFAEASLTGAGATFPAPVYAKWADTYQKETGNKVNYQG  60
            MKVMRTTVATVVAATLSMSAFSVFAEASLTGAGATFPAPVYAKWADTYQKETGNKVNYQG
Sbjct  1    MKVMRTTVATVVAATLSMSAFSVFAEASLTGAGATFPAPVYAKWADTYQKETGNKVNYQG  60

Query  61   IGSSGGVKQIIANTVDFGASDAPLSDEKLAQEGLFQFPTVIGGVVLAVNIPGLKSGELVL  120
            IGSSGGVKQIIANTVDFGASDAPLSDEKLAQEGLFQFPTVIGGVVLAVNIPGLKSGELVL
Sbjct  61   IGSSGGVKQIIANTVDFGASDAPLSDEKLAQEGLFQFPTVIGGVVLAVNIPGLKSGELVL  120

Query  121  DGKTLGDIYLGKIKKWDDEAIAKLNPGLKLPSQNIAVVRRADGSGTSFVFTSYLAKVNEE  180
            DGKTLGDIYLGKIKKWDDEAIAKLNPGLKLPSQNIAVVRRADGSGTSFVFTSYLAKVNEE
Sbjct  121  DGKTLGDIYLGKIKKWDDEAIAKLNPGLKLPSQNIAVVRRADGSGTSFVFTSYLAKVNEE  180

Query  181  WKNNVGTGSTVKWPIGLGGKGNDGIAAFVQRLPGAIGYVEYAYAKQNNLAYTKLISADGK  240
            WKNNVGTGSTVKWPIGLGGKGNDGIAAFVQRLPGAIGYVEYAYAKQNNLAYTKLISADGK
Sbjct  181  WKNNVGTGSTVKWPIGLGGKGNDGIAAFVQRLPGAIGYVEYAYAKQNNLAYTKLISADGK  240

Query  241  PVSPTEENFANAAKGADWSKTFAQDLTNQKGEDAWPITSTTFILIHKDQKKPEQGTEVLK  300
            PVSPTEENFANAAKGADWSKTFAQDLTNQKGEDAWPITSTTFILIHKDQKKPEQGTEV+K
Sbjct  241  PVSPTEENFANAAKGADWSKTFAQDLTNQKGEDAWPITSTTFILIHKDQKKPEQGTEVMK  300

Query  301  FFDWAYKTGAKQANDLDYASLPDSVVEQVRAAWKTNIKDSSGKPLY  346
            FFDWAYKTGAKQANDLDYASLPDSVVEQVRAAWKTNIKDSSGKPLY
Sbjct  301  FFDWAYKTGAKQANDLDYASLPDSVVEQVRAAWKTNIKDSSGKPLY  346
```

C

phosphate ABC transporter substrate-binding protein [Meiothermus taiwanensis]

Score	Expect	Method	Identities	Positives	Gaps	Frame
565 bits(1456)	0.0()	Compositional matrix adjust.	273/295(93%)	286/295(96%)	0/295(0%)	

Features:

```
Query  1    RQFIEQTVHFGASDNPFNDQQMAEIRRNTGSPALNIPFVLGAVVPTYNLPGVTQRLNFTG  60
            RQFIEQTVHFG SDNPFNDQQMA+IRRNTGSPALNIPFVLGAVVPTYNLPGVTQRLNFTG
Sbjct  65   RQFIEQTVHFGVSDNPFNDQQMADIRRNTGSPALNIPFVLGAVVPTYNLPGVTQRLNFTG  124

Query  61   EVLADIFLGNIKTWNDPAIARLNEGVRLPPLPITVVHRSDGSGTTYVWTDYLTKVSPEWA  120
            EVLADIFLGNIKTWNDPAIARLNEGVRLPPLPITVVHRSDGSGTT+VWTDYL+KVSPEWA
Sbjct  125  EVLADIFLGNIKTWNDPAIARLNEGVRLPPLPITVVHRSDGSGTTFVWTDYLSKVSPEWA  184

Query  121  QKVGRGNSVNWLAPNKVGGRGNEGVAGVVRNTPGAIGYNEVTYAVQNRILFGAVQNRAGR  180
            QKVGRGNSVNWL  NKVG RGNEGVAGVVRNTPGAIGYNEVTYA+QNRI FGAVQNRAGR
Sbjct  185  QKVGRGNSVNWLPANKVGARGNEGVAGVVRNTPGAIGYNEVTYAIQNRIQFGAVQNRAGR  244

Query  181  FMVADLPAIAAAANVVLPGDARVSLTNTQAPDGYPIASFSYLLVYEQLDKNKAFKSEAEA  240
            F+VA+LPAI AAANVVLPGDARVSLT+TQAPDGYPI+SF+Y+LVYEQLDKNKAFKSEAEA
Sbjct  245  FIVAELPAITAAANVVLPGDARVSLTDTQAPDGYPISSFAYMLVYEQLDKNKAFKSEAEA  304

Query  241  RAFVQLLKWIVTEGQKYNEPLTYGRLTETAQARALALISRITYQGKPIGREIVGR  295
            RAFVQLLKW+VT+ QK+NEPLTY RLTE AQARALALISRITYQGKPIGREIVGR
Sbjct  305  RAFVQLLKWMVTDAQKFNEPLTYARLTEVAQARALALISRITYQGKPIGREIVGR  359
```

Fig. D.2 BLAST alignments for b3728 and Mrub_2518. (a) NCBI BLAST of b3728 against Mrub_2518, (b) top BLAST hit for b3728, (c) top BLAST hit for Mrub_2518.

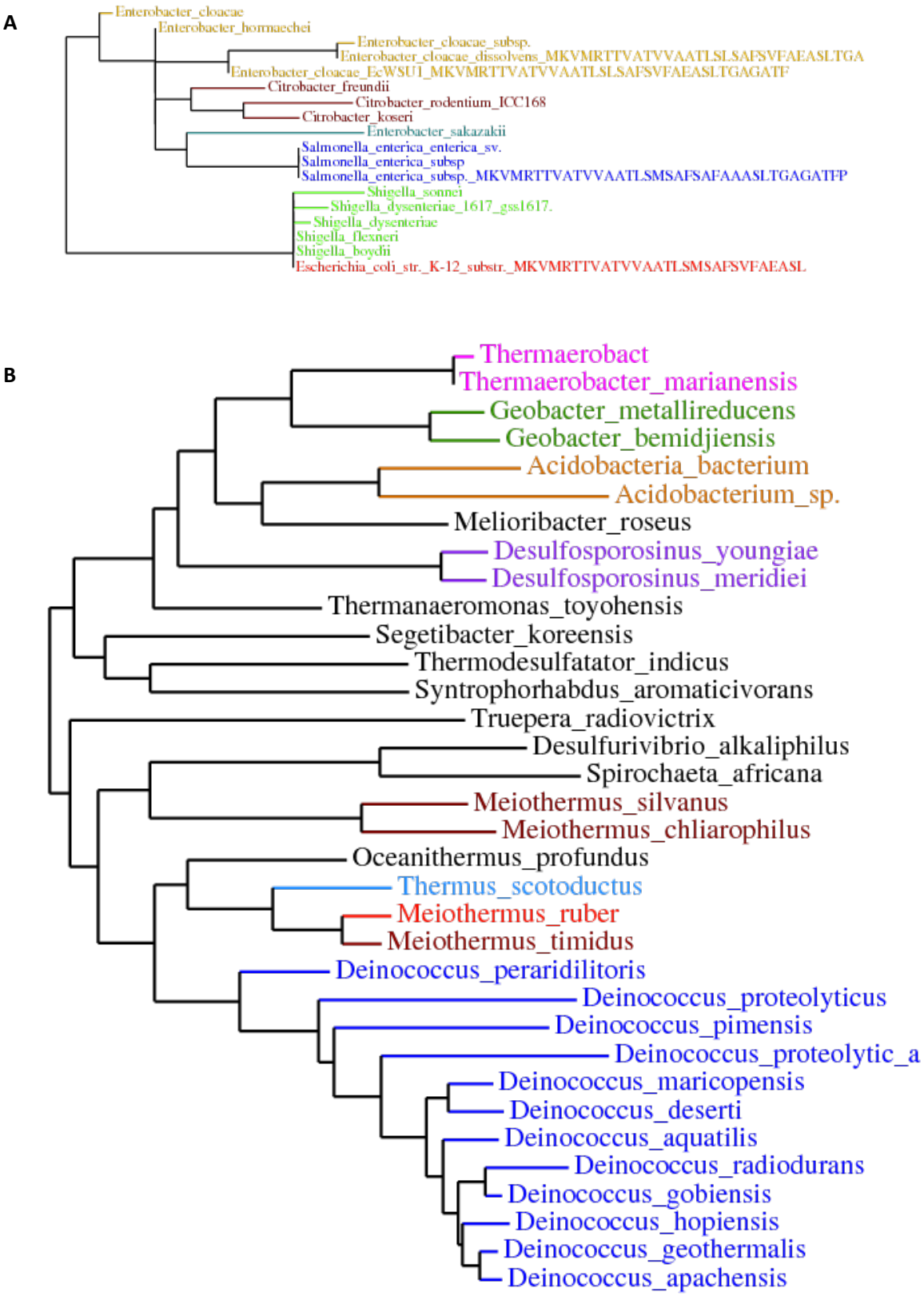


Fig. D.3 Phylogenetic tree (Phylogeny.fr) for b3728 (a) shows no sign of HGT; the tree for Mrub_2518 hints at the possibility of distant HGT.

Appendix E

Miscellaneous

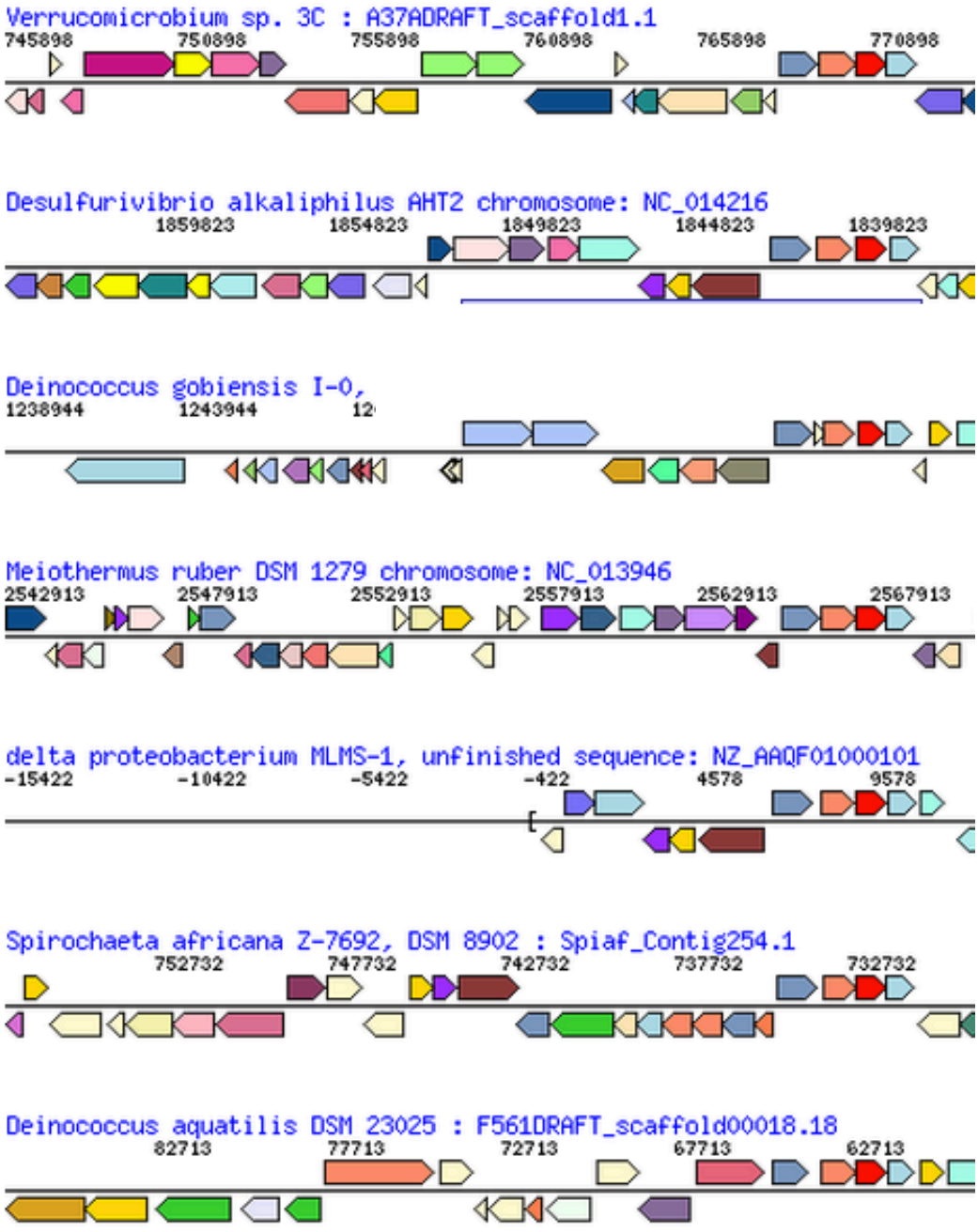


Fig. E.1 Gene neighborhood map (NCBI BLAST) showing strong conservation of the Pst operon (right side of image) in species distantly and closely related to *M. ruber*. These data offer additional evidence for HGT.