Horizontal Gene Transfer in the Molecular Evolution of Mannose PTS Transporters

Manuel Zúñiga,*1 Iñaki Comas,†1 Raquel Linaje,* Vicente Monedero,* María Jesús Yebra,* Carlos David Esteban,* Josef Deutscher,‡ Gaspar Pérez-Martínez,* and Fernando González-Candelas†

*Instituto de Agroquímica y Tecnología de Alimentos, IATA-CSIC, 46100 Burjassot, Valencia, Spain; †Instituto Cavanilles de Biodiversidad y Biología Evolutiva, Universidad de Valencia, Apartado Oficial 22085, Valencia E-46071, Spain; and ‡Laboratoire de Microbiologie et Génétique Moléculaire, INRA-INAPG-CNRS, F-78850 Thiverval-Grignon, France

The phosphoenolpyruvate:carbohydrate phosphotransferase system (PTS) represents hitherto the only example of group translocation transport systems. PTS transporters are exclusively found in bacteria and can be grouped on the basis of sequence and structure into six classes. We have analyzed the evolution of mannose-class PTS transporters. These transporters have a limited distribution among bacteria being mostly harbored by species associated to animals. The results obtained indicate that these genes have undergone a complex evolutionary history, including extensive horizontal gene transfer events, duplications, and nonorthologous displacements. The phylogenetic analysis revealed an early diversification to specialize in different transport capabilities, but these events have also occurred relatively recently. In addition, these transporters can be further divided into seven groups and this division correlates with their transport capabilities. Finally, the consideration of the genomic context allowed us to propose putative functional roles for some uncharacterized PTS transporters. The functional role and distribution of mannose-class PTS transporters suggest that their expansion may have played a significant role in the establishment of symbiotic relationships between animals and some bacteria.

Introduction

Bacteria utilize different transport mechanisms for the translocation of solutes across the membrane, such as facilitated diffusion, primary and secondary active transport, and group translocation (Saier 2000). Group translocation results in chemical modification of the transported solute (Hays 1978), and it is exemplified by the bacterial phosphoenolpyruvate:carbohydrate phosphotransferase system (PTS) which couples sugar transport to sugar phosphorylation at the expense of phosphoenolpyruvate (PEP) (Kundig, Gosh, and Roseman 1964). PTS components have also been associated with bacterial virulence: signature-tagged mutagenesis screenings have identified PTS genes as virulence factors in some pathogenic streptococci and Salmonella (Turner et al. 1998; Jones, Knoll, and Rubens 2000; Lau et al. 2001; Hava and Camilli 2002).

The PTS consists of the general phosphotransferase proteins, Enzyme I (EI) and HPr, the carbohydrate-specific transporters, and proteins involved in signal transduction (fig. 1). Together they constitute a phosphorylation cascade that starts at EI, which can be autoprophosphorylated at a histidine residue by PEP. Phospho-EI transfers the phosphoryl group to HPr, which becomes phosphorylated at a conserved histidine-15 residue (for a review see Ginsburg and Peterkofsky 2002). P—His-HPr functions as a phosphoryl donor to the different PTS transporters. The PTS transporters consist of three functional subunits, IIA, IIB, and IIC, with the only exception of those belonging to the mannose family, which contain an additional subunit, IID. These subunits can occur as separate proteins or as domains of complex proteins. IIA receives the phosphoryl group from P—His-HPr at a histidyl residue and transfers it to histidyl or cystyl residues in IIB. Finally, the phosphorylated IIB transfers its phosphate to the incoming sugar, which enters the cell via the IIC (and IID) transmembrane transporters (Postma, Lengeler, and Jacobson 1993).

In addition to their role in sugar transport, PTS components participate in signal transduction, chemotaxis, and the regulation of essential physiological processes such as carbohydrate utilization, catabolite repression, and coordination of carbon and nitrogen metabolism (Postma, Lengeler, and Jacobson 1993). For example, the glucose-specific IIA from enteric bacteria can modulate the activity of the enzyme adenylate cyclase and thus control levels of cyclic adenosine monophosphate, which in turn affects transcription of a variety of catabolic genes. In low G+C gram-positive bacteria, HPr undergoes a second ATP-dependent phosphorylation at serine residue 46, catalyzed by a metabolically activated HPr kinase/phosphorylase (Poncet et al. 2004). P-Ser-HPr acts as coregulator of the catabolite global regulator CcpA, which mediates catabolite repression/activation of up to 10% of total genes (Deutscher et al. 1995). Also, many transcriptional regulators contain PTS-regulated domains, with domains IIA and IIB with fully conserved phosphorylation sites in the C-terminal part of activators. Phosphorylation of these domains plays a role in modulating the activity of these proteins (Greenberg, Stülke, and Saier 2002).

PTS transporters can be divided into six classes on the basis of their amino acid sequences and structural properties (Saier and Reizer 1992; Postma, Lengeler, and Jacobson 1993; Robillard and Broos 1999; Siebold et al. 2001): (1) the glucose class, which includes the glucose PTS from Escherichia coli and Bacillus subtilis and maltose and N-acetylglucosamine PTS from E. coli, (2) the mannitol class, which is represented by the mannitol and fructose PTS transporters from E. coli, (3) the lactose class, responsible for the transport of lactose in Lactobacillus casei or Staphylococcus aureus and cellobiose in E. coli, (4) the galactitol class, which is represented by the transporter characterized in
E. coli, (5) the sorbitol class, responsible for the transport of sorbitol, and (6) the mannose class, transporting mannose, sorbose, and fructose, respectively. Structural studies have shown that IIA and IIB domains of the different PTS families display completely different foldings (for a review, see Robillard and Broos 1999). Therefore, they appear to be mosaic systems that have been constituted by the recruitment of unrelated proteins (Saier 2000).

Previous studies have explored the evolution of PTS components. These studies focused on the identification of novel PTS components on newly available genome sequences (Saier 2001 and references therein) or the evolution of the components of the phosphorylation cascade (Reizer and Saier 1997; Greenberg, Stülke, and Saier 2002; Hu and Saier 2002). These studies have shown that PTS components are exclusively found in bacteria and their distribution among different organisms is very uneven: while PTS transporters are in general numerous in species belonging to lactic bacteria and enterobacteria, they are absent from many other species (Saier and Paulsen 1999). The distribution of, on one hand, components of the phosphorylation cascade and, on the other, PTS transporters is also uneven. Some bacteria possess complete phosphorylation cascades although they lack PTS transporters (Hu and Saier 2002). Considerable differences can be observed even among species belonging to the same genus: homofermentative lactobacilli possess numerous PTS transporters, while they are rare in heterofermentative lactobacilli (Reizer, Peterkofsky, and Romano 1988). These data suggest that the genes responsible for the phosphorylation cascade and the transporters have different evolutionary histories.

The aim of this work is to analyze the evolution of PTS transporters. Due to the complexity of this system, we have focused our interest on the mannose class of PTS transporters. Because this class includes transporters with varied transport specificities, we will use the term mannose class to refer to the complete group and mannose/glucose transporters specifically for those members that transport two sugars. This particular class is relevant for the following reasons: (1) it is the main PTS for glucose transport in important bacteria, such as lactobacilli or streptococci, (2) it possesses an additional IID domain, (3) a relative wide spectrum of substrates has been described for some members of this class, such as glucose, mannose, sorbose, galactosamine, or fructose, (4) genetic and biochemical analyses have revealed that mannose/glucose transporters are implicated in the regulation of some cellular processes in Firmicutes (Vadeboncoeur and Pelletier 1997). We have performed a phylogenetic analysis of the mannose-class PTS transporter components present in available complete genome sequences. The results obtained indicate an early diversification to specialize in different transport capabilities, but these events have also occurred relatively recently. In addition, the results obtained indicate that extensive lateral gene transfers rather than gene duplications can explain the presence of multiple transporters within the same organism.

Materials and Methods
Sequences

The complete genome sequences available from Entrez Genomes on October 3, 2003 were used in this study with the exception of the E. coli C sequence harboring the genes encoding the PTS transporters of N-acetylgalactosamine and galactosamine because this region has suffered a major deletion in other E. coli strains (Brinkkötter et al. 2000). BlastP and TBlastX (Altschul et al. 1990) were used to identify mannose PTS–encoding genes. The data set was refined by excluding incomplete PTS gene clusters and redundant sequences. Table 1 presents a summary of the sequences used in this analysis, their accession numbers, and the subunit structure. More detailed information is provided in Supplementary table 1 (Supplementary Material online). A possible frameshift in the gene encoding the IID subunit of the cPERMAN cluster was corrected (C in position 1008362 of GenBank accession number NC_003366 was removed), and the translated product was used in this analysis. Phylogenetic reconstructions were obtained for each of the four subunits, although subunits A and B are found in two different forms, independent or fused, depending on the species and gene family. This is also the case for the C and D subunits of the transporter of Thermoanaerobacter tengcongensis.
Alignment and Informational Analysis

Deduced amino acid sequences were aligned with ClustalW (Thompson, Higgins, and Gibson 1994) and extensive gaps and regions of uncertain homology were removed by using GBLOCKS (Castresana 2000). Sequences with fused domains (A + B or C + D) were split on the basis of these multiple alignments and each region was analyzed with the corresponding homologues. The alignments used for analysis are available as Supplementary Material online.

Phylogenetic information content in the sequences was analyzed with two different methods: likelihood mapping (Strimmer and von Haeseler 1997) and split decomposition (Bandelt and Dress 1992). Likelihood mapping, implemented in Tree-Puzzle 5.1 (Schmidt et al. 2002), evaluates the resolution in quartets generated from combinations of the different sequences under study. Values below 90% resolved quartets were considered to indicate a low phylogenetic signal. In this evaluation we used the JTT substitution model for amino acid sequences with a proportion of invariant sites, empirical amino acid frequencies, and eight categories in a discrete gamma distribution to account for rate heterogeneity across sites.

Another method that provides an indication of the possible complexity in phylogenetic reconstruction is split decomposition. This method can be used as an indicator of conflicting phylogenetic signal. The method is aimed at representing all possible relationships among sequences although they are discordant, non–tree-like. We used the implementation in SplitsTree 4.0 (Huson 1998) with the JTT substitution model and Neighbor-net (Bryant and Moulton 2002) for clustering. This is a distance-based clustering algorithm derived from Neighbor-Joining (Saitou and Nei 1987) that provides a good resolution for networks and, simultaneously, allows the identification of conflicting signals arising from processes such as horizontal gene transfer (HGT).

Functional annotation of genes associated to mannose-class PTS gene clusters was verified by Blast searches. In addition, functional domains were identified by using CDSearch (Marchler-Bauer and Bryant 2004). SignalP 3.0 (Dyrlov Bendtsen et al. 2004) was used for prediction of signal peptides.

Phylogeny Reconstruction

We have used three different methods for phylogeny reconstruction and assessment of the inferred evolutionary relationships based on amino acid sequences. First, we used maximum-likelihood inference as implemented in PHYML 2.1b (Guindon and Gascuel 2003), which optimizes the likelihood function by simultaneously adjusting the topology and branch lengths. We used the JTT (Jones, Taylor, and Thornton 1992) substitution matrix, with a discrete gamma function with eight categories plus invariant sites to account for substitution rate heterogeneity among sites and empirically estimated amino acid frequencies. Second, we employed Bayesian inference with MrBayes 3.0 (Ronquist and Huelsenbeck 2003) using the same substitution model as in PHYML. We used one cold and three heated chains, with random initial trees. Trees were generated for 1,000,000 generations, with sampling every 100 generations, and the first 25,000 generations were discarded.

Table 1
Summary of Sequences Used in this Analysis

<table>
<thead>
<tr>
<th>Organism</th>
<th>Accession number</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>A + B</th>
<th>C + D</th>
<th>A + B + C + D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacillus subtilis</td>
<td>NC_000964</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Candidatus Blochmannia floridanus</td>
<td>NC_005061</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Clostridium acetobutylicum ATCC824</td>
<td>NC_003030</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Clostridium perfringens</td>
<td>NC_003366</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Enterococcus faecalis V583</td>
<td>NC_004668</td>
<td>11</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>Escherichia coli O157:H7</td>
<td>NC_002695</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Escherichia coli CFT073</td>
<td>NC_004431</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Escherichia coli O157:H7 EDL933</td>
<td>NC_002655</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Faustobacterium nucleatum subsp. Nucleatum</td>
<td>NC_003454</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Haemophilus ducreyi 35000HP</td>
<td>NC_002940</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Lactobacillus plantarum</td>
<td>NC_004567</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Lactococcus lactis</td>
<td>NC_002662</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Listeria monocytogenes EGD-e</td>
<td>NC_003210</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Oceanobacillus iheyensis</td>
<td>NC_004193</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Pasteurella multocida</td>
<td>NC_002663</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Photorhabdus luminescens subsp. Laumondii TTO1</td>
<td>NC_005126</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Salmonella typhimurium LT2</td>
<td>NC_003197</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Streptococcus agalactiae 2603VIR</td>
<td>NC_004116</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Streptococcus mutans UA159</td>
<td>NC_004350</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Streptococcus pneumoniae R6</td>
<td>NC_003098</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Streptococcus pyogenes MGA58232</td>
<td>NC_003485</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Thermoanaerobacter tengcongensis</td>
<td>NC_003869</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Vibrio vulnificus CMCP6</td>
<td>NC_004460</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Yersinia pestis KIM</td>
<td>NC_004088</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

NOTE.—From each complete genome sequence (accession number) all pts sequences were extracted and classified into one of three possible organization schemes: A B C D = separate genes for each subunit, A + B C D = one gene for fused subunits A and B, A B C + D = one gene for fused subunits C and D.
as “burnin.” The final tree was obtained as the consensus of 9,750 trees. Third, we used quartet-puzzling–based maximum likelihood as implemented in Tree-Puzzle 5.1 (Schmidt et al. 2002). In this case, we used 25,000 puzzling steps and the same evolutionary model as above except for the amino acid substitution matrix which in this case was WAG (Whelan and Goldman 2001), more appropriate for distantly related sequences.

Because the topologies obtained with these different methods were not coincident, they were compared using the Shimodaira and Hasegawa (1999) test and expected-likelihood weights (Strimmer and Rambaut 2002) by resampling–estimated log likelihood bootstrapping using 1,000 replicates. The best topology in each case was further evaluated by bootstrap resampling using Seqboot from the PHYLIP 3.6 package (Felsenstein 1993) with 500 pseudoreplicates.

From the resulting topologies we were able to clearly differentiate seven groups in each subunit. Group 1 is the largest one, including 21 sequences from different species and we paid special attention to it. New phylogenetic trees using only sequences from this group were obtained and they were compared to the species tree. This tree was obtained after concatenating 60 homologous, informational genes from the 21 species obtained in the Microbial Genome Database (Uchiyama 2003) (see table 2). Both the subunit IIA sequences and the concatenated polypeptide were aligned using ClustalW (Thompson, Higgins, and Gibson 1994) and the resulting alignments were processed with GBLOCKS (Castresana 2000) to eliminate sites with uncertain homology. Phylogenetic trees for group 1 subunits were obtained by Bayesian inference (Huelsenbeck and Ronquist 2001), with four chains and 250,000 generations, with the initial 25,000 generations used for burning. For the concatenated sequences, a species phylogenetic tree was obtained using the quartet-puzzling method implemented in Tree-Puzzle 5.1 (Schmidt et al. 2002) with 4,000 puzzling steps. In both cases we used the same evolutionary model described above for the subunits tree.

Results

Distribution of Mannose-Class PTS Transporters

The screening of available completed genome sequences revealed a limited distribution of clusters of genes encoding mannose-class PTS transporters (A, B, C, and D subunits). Most transporters were found in species belonging to Firmicutes and γ-Proteobacteria. The unique exception is a PTS transporter found in Fusobacterium nucleatum. Most clusters spanned the four domains, but a few incomplete clusters have also been found (see Supplementary table 1, Supplementary Material online). This screening led to the detection of two complete clusters in Enterococcus faecalis where the genes encoding the IIA and IIB subunits were incorrectly annotated as the result of frameshifts and therefore they were not included in the annotation of the genome sequence (clusters eFAMAEn11 and eFAMAEn12; Supplementary table 1, Supplementary Material online). In addition, mannose-class IIA–encoding sequences either as isolated genes or as domains of complex proteins were found. These included IIA subunits of dihydroxyacetone kinases (see Gutnkecht et al. 2001), IIA genes associated to other phosphoryl transfer proteins (Reizer and Saier 1997; Hu and Saier 2002), and IIA domains located within transcriptional regulators (Greenberg, Stülke, and Saier 2002; Boël et al. 2003). Some of these genes or domains were harbored by species that otherwise lack mannose PTS transporters (not shown). Because this study was focused on the evolution of mannose PTS transporters, only complete clusters have been analyzed.

Phylogenetic Analysis of Mannose PTS–Encoding Sequences

A total of 58 protein sequences from 24 bacterial species were used for the alignment of the different subunits that constitute the PTS transport system. Subunit IIA is the smallest, with only 104 amino acids after removal of positions with gaps in the multiple alignment. Subunits IIB, IIC, and IID have 152, 282, and 244 residues in the corresponding alignments. Likelihood-mapping analyses showed a clear relationship between sequence length and phylogenetic information content. None of the four subunits reached a resolution of quartets higher than 90%. Subunit IIA rendered the poorest signal with almost 30% of unresolved quartets, followed by subunit IIB (table 2).

This low phylogenetic signal was also reflected in the networks obtained with SplitsTree. The networks showed a great deal of conflicting phylogenetic signal (see for instance supplementary fig. 1 [Supplementary Material online] for subunit IIA). This suggested that there was a large amount of information shared by the sequences that was not compatible with a single, tree-like topology. Nevertheless, the network was still capable of showing congruence with the groups derived in the phylogenetic reconstructions (see below).

Both likelihood-mapping and phylogenetic networks gave hints on the difficulties to obtain a single, “best” topology to explain the course of evolution for each subunit. To partially circumvent this problem, we used three different approaches as described in Materials and Methods. SH test rejected consistently the topologies derived by quartet puzzling (table 3), a result that can be explained mainly by the low resolution of the nodes (data not shown). Consequently, we continued the analyses with the maximum-likelihood topologies derived by using PHYML, as they were not rejected in any of the corresponding SH tests. Figures 2–5 present the phylogenetic trees of each subunit with support values for the nodes from Bayesian analysis (with a posteriori probability larger than 0.95) and bootstrap resampling (bootstrap support larger than 90%). In general, Bayesian posterior probability is always larger than the corresponding

### Table 2

<table>
<thead>
<tr>
<th>Subunit</th>
<th>Resolved quartets</th>
<th>Partially resolved quartets</th>
<th>Unresolved quartets</th>
</tr>
</thead>
<tbody>
<tr>
<td>IIA</td>
<td>65.4</td>
<td>6</td>
<td>28.6</td>
</tr>
<tr>
<td>IIB</td>
<td>77.9</td>
<td>5.2</td>
<td>19.9</td>
</tr>
<tr>
<td>IIC</td>
<td>84.8</td>
<td>4.6</td>
<td>10.7</td>
</tr>
<tr>
<td>IID</td>
<td>85.8</td>
<td>4.6</td>
<td>9.6</td>
</tr>
</tbody>
</table>

Downloaded from https://academic.oup.com/mbe/article-abstract/22/8/1673/1042822 by guest on 11 March 2019
bootstrap support, and both values are larger in the trees for subunits IIC and IID than for subunits IIA and IIB.

A comparison of the common groupings from the four topologies allowed us to divide the corresponding sequences into seven groups. All the sequences with fused IIA-IIIB domains belong to group 1, with the exception of one E. faecalis copy, which always appeared in group 5. The remaining sequences, except for T. tengcongensis that encodes a fused IICD subunit, have separate domains and were distributed in the six remaining groups. The seven identified groups were almost invariable across the four subunits. There were only seven cases in which incongruities in the grouping of the different subunits were observed. Among these, six exceptions were found for subunit IIA and one for subunit IIB (figs. 2 and 3). Despite the constancy in the composition of these groups in the four phylogenies, the evolutionary relationships deduced for them were different for each subunit, and it was not possible to decide whether this was due to the aforementioned difficulty to derive the deepest phylogenetic relationships or to a different course of the evolution of the subunits.

Group 1 was the largest one and it included the mannose/glucose transporters, which also share the characteristic IIA-IIIB fusion. Besides, this was the only group in which the phylogenetic relationships of the included species were almost constant for the four subunits. In consequence, we proceeded to construct a species tree as described in Material and Methods and we compared this tree with the one derived for each subunit. Figure 6 shows the comparison of both phylogenies. The species tree presented two main monophyletic groups, separating \( \gamma \)-Proteobacteria from Firmicutes. This same division could be observed in the subunit trees with three exceptions. These three sequences corresponded to the sorbose transporter of E. coli O157:H7, one of the Listeria monocytogenes copies, and the only sequence from B. subtilis, which usually transports fructose although it can also transport mannose. Interestingly, although an acceptable agreement was observed between the two trees for \( \gamma \)-Proteobacteria and streptococci, the topology of the trees varied significantly for the remaining Firmicutes. We will further discuss this point below.

Discussion

Previous studies have noted the limited distribution of PTS transporters within bacteria (Saier and Paulsen 1999). In particular, mannose-class PTS transporters have been found only in Firmicutes and \( \gamma \)-Proteobacteria with the exception of the transporter found in F. nucleatum. Noticeably, most bacteria that harbor mannose-class PTS transporters may thrive in mucosal surfaces of animals either as saprophytic or pathogenic organisms. Moreover, some of these transporters are clustered with genes that encode proteins involved in polysaccharide degradation (see below). These facts suggest that PTS transporters may have played an important role in the process of adaptation of these bacteria to live in this habitat.

In an evolutionary context, this limited distribution suggests either that they have arisen relatively late during the bacterial evolution or that they have been lost in many bacterial lineages. Our results indicate that HGT has been prevalent in the evolution and distribution of these transporters. Several pieces of evidence support this conclusion. First, the phylogenetic analysis of the four domains/genes of the mannose-class PTS clearly shows that the evolution of these genes does not agree with the order of organismal descent, inferred from the analyses of either ribosomal RNAs (Woese 1987) or other protein markers such as RecA (Eisen 1995). Second, in organisms that possess multiple transporters, these are more related to transporters present in other species than to those present in the same organism, even in the presence of several transporters of the same group. This prevalence of HGT fits poorly with the hypothesis of gene loss in several bacterial lineages and points to a relatively late origin of this transport system.

The phylogenetic analyses of the four domains of the mannose-class PTS transporters has allowed us to consistently divide them into seven groups and has evidenced that
each domain originated from a common ancestor. However, no phylogenetic relationship is discernible between the different constituents of these transporters, in agreement with the view of this system as a mosaic (Saier 2000). This phenomenon has occurred to some extent in the late evolution of these transporters. Although deep evolutionary relationships between the different groups of mannose-class PTS transporters cannot be clearly elucidated, these results suggest that the evolution of transporters with novel capabilities occasionally involved the recruitment of genes from diverse groups by nonorthologous displacement. This mostly affected IIA domains (see fig. 2). In some cases, recent duplications are still discernible (for example, EFAECMAN3 and EFAECMAN4 and EFAECMAN6 and EFAECMAN7). Because IIA must interact with HPr and it may also play a regulatory role, we speculate that nonorthologous displacements have helped to fit recently acquired PTS transporters with the regulatory machinery of the new host. We will discuss next the results obtained for each group of transporters.

Group 1

Group 1 includes the characterized transporters of mannose/glucose of E. coli (Emi, Zanolari, and Kocher 1987) and Streptococcus mutans (Abranches, Chen, and Burne 2003). Moreover, transport of glucose and mannose by a mannose-class PTS transporter has been described in Pasteurella multocida (Binet and Bouvet 1998). Other characterized mannose/glucose transporters not included
in this study are those of *Streptococcus salivarius* (Lortie et al. 2000) and *Streptococcus thermophilus* (Cochu et al. 2003). The biochemical characterization of these transporters revealed that they have broad substrate specificity, being able to transport mannose, glucose, N-acetylglucosamine, fructose, and 2-deoxyglucose. These transporters are characterized by the fusion of IIA and IIB domains. Apart from these, only the *E. faecalis* transporter EFAEMAN1 possesses fused IIA and IIB domains, but the phylogenetic analysis does not support a close relationship of this transporter with group 1.

Additionally, group 1 includes the fructose transporter of *B. subtilis* (BSUBLEV), the sorbose transporter of *E. coli* (ECOLHSOR), and an uncharacterized transporter of *L. monocytogenes* (LMONOMAN2). In contrast to the genuine mannose/glucose transporters, these do not display the characteristic fusion of IIA and IIB subunits. The phylogenetic analysis suggests that these transporters constitute paralogs of mannose/glucose transporters that have evolved as specialized variants of mannose/glucose PTS ancestors of broad specificity. Nevertheless, it is not clear why the fusion of IIA and IIB subunits was not conserved, although at least for the sorbose transporter, a nonorthologous displacement by a group 5 IIA–encoding gene may have occurred (see fig. 2).

The occurrence of nonfused exceptions could indicate that there are some selection pressures that prevent the possibility of a fusion from being lost, possibly related to the biochemical and regulatory advantages of coupling two reactions (Yanai, Derti, and DeLisi 2001). In fact, recent studies have reported that it is easier to gain than to lose a fusion (Kummerfeld and Teichmann 2005).

There are two possible scenarios for the achievement of the fusion by each sequence. One possibility is that multiple HGT events spread the fusion from one genome to another, as depicted in the phylogenetic trees. The trees show the evolutionary relationships among the transporters, with the fusion of IIA and IIB domains appearing multiple times across different species. The trees also indicate that the fusion may have been acquired through horizontal gene transfer (HGT), as evidenced by the distribution of gene sequences across different bacterial genomes.

The phylogenetic trees (figs. 4 and 5) are constructed using the maximum-likelihood method with PHYML, as indicated at the bottom of each figure. These trees illustrate the evolutionary relationships among the transporter genes and provide insights into the possible origins and spread of the fusion events.
many others, and the other is its vertical inheritance (Yanai, Wolf, and Koonin 2002). In the case of the A and B subunits, their distribution is restricted to group 1 except for one of the copies of \textit{E. faecalis} (EFAEMAN1). This copy may have originated by an event unrelated to the fusions in group 1. Most sequences from this group maintain the most accepted species tree structure with the exception of some \textit{Firmicutes} species as we discuss below. Consequently, we think that the A and B fusions of the group 1 species were vertically inherited from a common ancestor and successive HGT events introduced the fusion in non-streptococcal \textit{Firmicutes}.

The phylogenetic analyses showed that sequences of mannose/glucose transporters of \gamma-Proteobacteria and streptococci constituted monophyletic groups in good agreement with the species tree, but other \textit{Firmicutes} sequences did not. These results suggest that mannose transporters in these groups may have been acquired by HGT. The distribution of transporters also points to this possibility: \textit{Oceanobacillus iheyensis} is, so far, the only member of the family Bacillaceae possessing a putative PTS transporter of mannose. In \textit{Clostridium acetobutylicum} the mannose transporter is located in plasmid pSOL1 (Nolling et al. 2001).

Furthermore, an additional gene, denoted \textit{manO} in \textit{S. salivarius} and \textit{S. thermophilus}, is commonly found in \textit{Firmicutes} associated to this cluster (see supplementary fig. 2, Supplementary Material online). Interestingly, \textit{manO} homologs are only found in \textit{Firmicutes} strains encoding a putative mannose/glucose PTS transporter. In \textit{Lactobacillus plantarum}, \textit{S. mutans}, and \textit{Streptococcus pneumoniae} this gene is separated by large insertions from the genes encoding the mannose PTS transporters, while \textit{Lactococcus lactis} possesses two copies (supplementary fig. 2, Supplementary Material online). Although the function of the product of this gene is unknown, it has been shown to be cotranscribed along with the genes encoding for the mannose/glucose transporter in \textit{S. salivarius} and \textit{S. thermophilus} (Lortie et al. 2000; Cochu et al. 2003). Together with its conservation in \textit{Firmicutes}, this suggests that the product of this gene may play a role related to the function of the mannose/glucose PTS transporter. Interestingly, in addition to its role in transport of mannose and glucose, the mannose/glucose PTS transporters play an important role in the regulation of the metabolism of carbon sources in \textit{Firmicutes} (for a review see Titgemeyer and Hillen 2002). Therefore, it is tempting to speculate that \textit{manO} homologs are involved in the regulatory functions of the mannose/glucose PTS transporter in \textit{Firmicutes}.

Groups 2 and 7

Group 2 includes the gluconate PTS transporter of \textit{E. faecalis}, encoded by cluster EFAEMAN10 (Hengstenberg et al. 1989; W. Hengstenberg, M. Skopnik and A. Brockmeier, personal communication), and three additional uncharacterized clusters harbored by \textit{E. faecalis} (EFAEMANS), \textit{Salmonella typhimurium} (STYHIMAN), and \textit{T. tengcongensis}. The presence of a gluconate dehydrogenase and the close phylogenetic relationship among subunits IIB, IIC, and IID of \textit{T. tengcongensis} and EFAEMANS suggest that this is probably also a gluconate PTS transporter. On the other hand, the IIA subunits were probably independently recruited in...
each cluster. Clusters EPAECM2 and STYPHMAN3 show a close phylogenetic relationship for every subunit. Moreover, these genes are flanked by homologous genes in both hosts (supplementary fig. 3, Supplementary Material online), suggesting that they may constitute a cluster that has been transferred between these species. Nevertheless, the available data do not allow inferring what substrates are recognized by these transporters.

Group 7 includes three uncharacterized PTS transporters from C. acetobutylicum, S. mutans, and E. faecalis. The comparison of the neighboring genes revealed the presence of a putative sugar sensor system shared by C. acetobutylicum and S. mutans which is absent from E. faecalis (see supplementary fig. 3, Supplementary Material online). This sensor system is unique to streptococci while a number of paralogs can be found in the genome of C. acetobutylicum (Nolling et al. 2001). In addition, two extracellular sugar-binding protein-encoding genes flank the sensor system. On the other hand, the phylogenetic relationships between the three sequences vary from one domain to another (see figs. 2–5). Taken together, these facts suggest that both S. mutans and E. faecalis acquired this PTS system from a clostridial donor. Possibly, E. faecalis subsequently lost the sensor system.

The phylogenetic analysis shows a close relationship of IIC and IID subunits of group 2 and group 7 indicating that they evolved from a common ancestor while IIA and IIB domains were recruited independently. Moreover, while IIA and IIB domains of group 7 cluster together, indicating that the assembly occurred once, IIA sequences of group 2 do not cluster suggesting that nonorthologous displacements also occurred in this group.

Group 3

Group 3 comprises a number of uncharacterized transporters, including the F. nucleatum transporter, the only species known so far to harbor a mannose-class PTS transporter without belonging to Firmicutes or γ-Proteobacteria. This fact, along with the evidence provided by the phylogenetic analysis, suggests that F. nucleatum acquired this transporter by HGT. In addition, the presence of a putative plasmidic replication protein next to the PTS cluster (supplementary fig. 4, Supplementary Material online) further supports this hypothesis. Group 3 can be divided into two subgroups: subgroup 3A is constituted by four closely related clusters (figs. 2–5) and their phylogenetic relationships are well conserved in the four domains. The phylogenetic relationships of sequences belonging to subgroup 3B are also conserved although they have a poorer resolution. This subdivision is further supported when the gene content of these clusters is considered: all 3B clusters are associated to two putative phosphosugar isomerases (supplementary fig. 4, Supplementary Material online) with the exception of cluster EPAECM2. The phylogenetic relationships between the two subgroups also vary depending on the domains considered: IIA and IID sequences of subgroup 3A are located within subgroup 3B while IIB and IIC are separated. In fact, the phylogenetic analysis did not cluster IIC sequences of 3A and 3B subgroups (fig. 4). Because the gene order ABCD is conserved in all group 3 clusters with the exception of LMONOMAN3, it may be hypothesized that 3A subgroup arose from a nonorthologous displacement that substituted IIB and IIC domains of an ancestral 3B cluster. Nevertheless, the poor resolution of deep branches of the phylogenetic trees makes it difficult to determine the origin of these sequences. In addition, the gene encoding LMONOMAN3 IIA constitutes another example of nonorthologous displacement because this sequence is located within group 1 and its position within its cognate gene cluster varies with respect to all other group 3 clusters. Also, in this group two copies of PTS transporters in the enteric bacteria Salmonella were found (STYPHMAN1 and STYPHMAN3). However, we have not been able to establish whether their origin is from two independent transfer events or from an ancient duplication whose signal has been lost.

Group 4

Group 4 is constituted by the Clostridium perfringens transporter CPERFMAN3 and the S. pneumoniae transporter SPINFENC. In both organisms, the genes encoding the four subunits of the transporter are located within complex clusters (see supplementary fig. 5, panel A, Supplementary Material online) which include the genes fucK, fucI, and fuc encoding fucose kinase, fucose isomerase, and fuculose phosphate aldolase, respectively. They catalyze a fucose utilization pathway that results in the degradation of fucose into dihydroxyacetone-P and 2-hydroxypropionaldehyde (supplementary fig. 5, panel B, Supplementary Material online). The transcription of the S. pneumoniae cluster has been partially characterized showing that the expression of these genes is induced by fucose although this organism is unable to grow with fucose as the sole carbon source (Chan et al. 2003). In addition, the cluster of C. perfringens spans lacA, lacB, and lacC homologs involved in galactose utilization by the tagatose pathway and three putative glycosidases. Gene cep0329 encodes an endo-β-galactosidase that has been described as active on blood group A and B glycans (Anderson et al. unpublished results; accession number AAR84225). The product of this gene contains a putative signal peptide and is probably secreted. Gene cep0325 encodes an α-N-acetylgalactosaminidase that cleaves the terminal α-1,3-linked N-acetylgalactosamine residues from various substrates including the blood type A epitope of erythrocytes (Levy and Aminoff 1980; Calcutt et al. 2002). Gene cep0324 encodes a protein with significant similarity to characterized broadly specific α-L-fucosidases (E.C. 3.2.1.51) belonging to the glycosyl hydroxylase family 29 (Sulzenbacher et al. 2004). Neither cep0324 nor cep0325 encode putative signal peptides.

The S. pneumoniae cluster also contains two putative glycosidases: gene spr1965 encodes a putatively secreted protein whose N-terminal part is significantly similar to the endo-β-galactosidase encoded by gene cep0329 (34% of identical residues, 53% conserved, in a 592-aa sequence span) and its C-end contains three repeats of a fucolectin domain (Honda et al. 2000). Gene spr1966 shares significant similarity to the catalytic domain of the specific α-(1→2) fucosidase of Bifidobacterium longum (26% of identical residues, 43% conserved, in a 891-aa sequence span; Katayama et al. 2004). The presence of these glycosidases...
therefore suggests that these clusters enable C. perfringens and S. pneumoniae to utilize glycans such as blood group A antigen (supplementary fig. 5, panel B, Supplementary Material online). The cleavage by endo-β-galactosidase of the galactose-β-(1→3)-N-acetylglucosamine bond would render a fucose-galactose-N-acetylglucosamine trisaccharide that would be subsequently degraded by the α-N-acetylgalactosaminidase and 1-fucosidase. Because the latter glycosidases apparently are not secreted, we hypothesize that the PTS transporter would internalize the trisaccharide and probably phosphorylate its galactose moiety. Nevertheless, the apparent lack of α-N-acetylgalactosaminidase by S. pneumoniae suggests that the disaccharide fucose-α-(1→2)-galactose may also be internalized. In both organisms, fucose will be phosphorylated and degraded as suggested before, while galactose-P would be directed towards the tagatose pathway by galactose-6-P isomerase (encoded by lacA and lacB) and tagatose-6-P kinase (lacC). Although C. perfringens lacks a lacD homolog, a putative tagatose/fructose aldolase (gaT) can be found close to the cperfMAN2 cluster. Streptococcus pneumoniae R6 harbors the genes encoding the tagatose pathway (lacABCDF) upstream of an operon encoding a putative lactose PTS transporter and genes required for lactose utilization and regulation (lacTPEG and lacR). Therefore, both organisms possess a complete set of genes for the utilization of fucose and galactose resulting from the degradation of the fucose-galactose disaccharide.

Group 5

This group includes a number of uncharacte rized PTS transporters, four of them harbored by E. faecalis (EFAEMAN1, EFAECMAN3, EFAECMAN4, and EFAECMAN8). These transporters are found within complex gene clusters (supplementary fig. 6, Supplementary Material online). The gene content of these clusters suggests that group 5 includes oligosaccharide transporters with different specificities, EFAEMAN1, EFAECMAN8, and SPNELAC are associated to putative nonsecreted β-galactosidases (encoded by genes EF0813, EF1805, and bgAC, respectively) while EFAECMAN4 and SMUTANMAN1 are associated to putative glycosidases belonging to family 31 such as α-glucosidases, α-galactosidases, and α-xyllosidases. Cluster EFAEMAN1 also includes a putative unsaturated glucuronyl hydrolase similar to the characterized protein of Bacillus GL1 (41% identical residues, 61% conserved, in a 364-aa sequence span). These enz ymes act on oligosaccharides produced through the reactions of polysaccharide lyases by specifically hydrolyzing the glycosidic bond between the unsaturated glucuronyl residue at the nonreducing terminus and the saccharide linked to the residue (Hashimoto et al. 1999). In agreement with this assumption, this cluster also includes a putative secreted polysaccharide lyase belonging to family 8 (EF0818) and exhibiting significant similarity to the xanthan lyase of Bacillus GL1 (31% identical residues, 49% conserved, in a 800-aa sequence span; see Hashimoto et al. 2001). Cluster EFAECMAN8 includes gene EF1800 that encodes a putative secreted protein of unknown function that shares some similarity to the N-terminal part of EF0818 (supplementary fig. 6, Supplementary Material online). Finally, a surface-located beta-N-acetylglucosaminidase (gene strH) is located close the SPNELAC cluster in S. pneumoniae (supplementary fig. 6, Supplementary Material online). This protein cleaves preferentially terminal N-acetylglucosamine residues β-(1→3) and β-(1→6) linked to galactose and β-(1→2) linked to mannose (Clarke, Platt, and Butters 1995). Therefore, it seems that at least some of these transporters internalize oligosaccharides resulting from the action of secreted polysaccharide degrading enzymes. Transporters EFAEMAN1, EFAECMAN8, and SPNELAC probably transport galactose-containing oligosaccharides. The presence of genes involved in xylose metabolism together with an α-glycosidase in EFAECMAN4 suggests that this transporter may internalize α-linked xylose-containing oligosaccharides such as isoprimeverose (α-D-xylopyranosyl-(1→6)-α-D-glucopyranose).

The glycosidase complement of these clusters corresponds to some extent to the phylogenetic relationships observed. The positions of EFAEMAN1, EFAECMAN8, and SPNELAC (all of them associated to β-galactosidases) are conserved in the four domains. The domains IIIC and IIC of SMUTANMAN1 and EFAECMAN8, both associated to glycosidases of family 31, also cluster together. In addition, some nonorthologous displacements have probably occurred within this group. CPERFMAN2 domains IIa and IIB are not related to their group 5 counterparts and apparently were recruited independently because their phylogenetic positions are very different: CPERFMAN2 IIa clusters with the gluconate transporter EFAEMAN10 IIa subunit (fig. 2) while the corresponding IIB domain occupies an intermediate position between groups 2 and 3. Also, domain IIA of SMUTANMAN1 is included in group 3.

Group 6

Group 6 includes the characterized N-acetylglactosamine transporter of E. coli ECOLNAG (Brinkkötter et al. 2000). The comparison of the gene clusters suggests that transporters EFAECMAN7, PLUMIAGA, VVULNMAN1, and YPESTM AN1 also internalize N-acetylglactosamine (supplementary fig. 7, Supplementary Material online). It is noteworthy that the cluster harbored by Photobacterium luminescens also includes a gene, plu0840, encoding a putative secreted protein that exhibits significant similarity to the C. perfringens α-N-acetylgalactosaminidase (27% identical residues, 46% conserved in a 460-aa sequence span; see discussion of group 4). On the other hand, the group 6 PTS transporter encoded by Yersinia pestis KIM is located within a region that contains genes involved in the utilization of sulfated polysaccharides such as chondroitin sulfate (supplementary fig. 7, Supplementary Material online). Although it was not included in this study, preliminary analyses showed that the galactosamine transporter of E. coli also belongs to this group, being most closely related to LPLANTNAG (results not shown). The phylogenetic relationships are relatively well conserved between the four subunits with the exception of the IIA subunit of EFAECMAN7 (fig. 2) that probably originated from a nonorthologous replacement by a duplicated copy of EFAECMAN6.

The phylogenetic analysis and the comparison of gene clusters indicate that transporters EFAECMAN6, SAGALMAN2, SPNEGLEN, and SPYONAG constitute a distinct subgroup.
within group 6. This subgroup includes genes involved in glucuronate metabolism, putative unsaturated glucuronol hydrolases, and putative polysaccharide lyases (distantly related to heparinase III of Pedobacter heparinus). The unsaturated glucuronol residues released by the action of the unsaturated glucuronol hydrolases are nonenzymatically converted to α-keto acids (Mori et al. 2003). The presence of genes involved in glucuronate metabolism suggests that these α-keto acids are degraded via 2-keto-3-deoxygluconate kinase and 2-keto-3-deoxyx phosphogluconate aldolase to pyruvate and glyceraldehyde-3-P. Therefore, we hypothesize that these transporters translocate disaccharides formed by an uronic acid and either galactosamine or N-acetylgalactosamine resulting from the degradation of polysaccharides by the action of the associated polysaccharide lyases.

**Conclusion**

Biological information is hierarchically organized and it is also hierarchically interpreted and used by living organisms (Weiss 2005). Hence, it should also be hierarchically analyzed and interpreted. In this work we have shown the interest and benefits of combining sequence analysis at two different levels to the study of the mannose-class PTS transporters. Analysis at the phylogenetic level allowed us to infer evolutionary scenarios and groups of sequences with possibly similar functions. The presence in many genomes of multiple copies of different origins reveals that their evolution is driven by HGT events. Their clustering with copies from other genomes in the phylogenetic trees suggests that the mannose-class PTS transporters group according to sugar specificity: hexoses, aminohexoses, gluconate, and different oligosaccharides.

On the other hand, the additional consideration of the genomic context at the functional and neighborhood levels has allowed us to propose putative functional roles for some of the uncharacterized PTS transporters. This same kind of genome-context studies has been used to improve functional prediction in prokaryotes (Overbeek et al. 1999; Huynen et al. 2000). The combination of both analyses confirms the functional flexibility of the PTS system based on its actual mosaic structure, a consequence of multiple large- and small-scale gene transfer events. The distribution of mannose-class PTS transporters suggests that they may play a significant role in adaptation of bacteria to life in association with animals, particularly in epithelial surfaces, and that their expansion and differentiation probably paralleled the process of establishment of symbiotic relationships with animals. The presence of multiple transporters in some bacterial species, best exemplified by *E. faecalis*, indicates that the availability of efficient transport systems to utilize a variety of sugar carbon sources may be an important trait for survival in the intestinal habitat.

**Supplementary Material**

Multiple alignments of mannose PTS transporter sequences used in the study and the following Supplementary table 1 and supplementary figs. 1–7 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Supplementary table 1. Detailed information of sequences used in this study.

Supplementary Figures:

Fig. 1.—Phylogenetic networks obtained with Split-Trees.

Fig. 2.—Genetic organization of mannose-class PTS gene clusters belonging to group 1.

Fig. 3.—Genetic organization of mannose-class PTS gene clusters belonging to groups 2 and 7.

Fig. 4.—Genetic organization of mannose-class PTS gene clusters belonging to group 3.

Fig. 5.—Panel A. Genetic organization of mannose-class PTS gene clusters belonging to group 4.

Fig. 6.—Panel A. Genetic organization of mannose-class PTS gene clusters belonging to group 5.

Fig. 7.—Genetic organization of mannose-class PTS gene clusters belonging to group 6.

**Acknowledgments**

This research was sponsored by Conselleria de Sanitat (Generalitat Valenciana), project Grupos03/204 from Agencia Valenciana de Ciencia y Tecnologia-Programa AGGV and project BMC2001–3096 from Ministerio de Ciencia y Tecnología (Spain). I.C. is recipient of a graduate student fellowship from the Spanish Ministerio de Educación y Cultura. M.J.Y. and M.Z. are researchers of the Ramón y Cajal Program of the Spanish Ministerio de Ciencia y Tecnología.

**Literature Cited**


Felsenstein, J. 1993. PHYLIP (phylogenetic inference package), Department of Genetics, University of Washington, Seattle.


Takashi Gojobori, Associate Editor

Accepted April 14, 2005