Research Article

Journal of Molecular Microbiology and Biotechnology

J Mol Microbiol Biotechnol 2010;18:37–51
DOI: 10.1159/000274310

Published online: January 6, 2010

Bacterial and Eukaryotic Phosphoketolases: Phylogeny, Distribution and Evolution

Borja Sánchez[a] Manuel Zúñiga[b] Fernando González-Candelas[c–e]
Clara G. de los Reyes-Gavilán[f] Abelardo Margolles[f]

[a]Université de Bordeaux, UMR 5248 CNRS, UBX1-ENITAB, ENITAB, Gradignan, France; [b]Instituto de Agroquímica y Tecnología de Alimentos, Consejo Superior de Investigaciones Científicas, [c]Instituto Cavanilles de Biodiversidad y Biología Evolutiva, Universidad de Valencia, [d]CIBER en Epidemiología y Salud Pública, [e]Area de Genómica y Salud, Centro Superior de Investigación en Salud Pública, Valencia, [f]Departamento de Microbiología y Bioquímica, Instituto de Productos Lácteos de Asturias, Consejo Superior de Investigaciones Científicas, Villaviciosa, Spain

Key Words
Phosphoketolase · Horizontal gene transfer · Phylogenetic study

Abstract
Phosphoketolases (XFPs) are glycolytic enzymes present in several organisms belonging to the *Eukarya* and *Bacteria* domains. A total of 151 putative *xfp* genes were detected in 650 complete genomes available in public databases. Elimination of redundant sequences and pseudogenes rendered a final data set of 128 phosphoketolases, which was analyzed by phylogenetic methods. The distribution of *xfp* genes was uneven in most taxonomic groups, with the exception of the taxonomical division Lactobacillaceae, in which all the species studied harbored a putative *xfp* gene. Putative *xfp* genes were also present predominantly in *Rhizobiales* and *Actinobacteria* divisions, in which 23 out of 28 genomes and 23 out of 41 genomes contained at least one putative *xfp* homolog, respectively. Phylogenetic analyses showed clear discordance with the expected order of organismal descent even in groups where *xfp* is prevalent, such as *Lactobacillaceae*. The presence of putative paralogs in some organisms cannot account for these discrepancies; instead, these paralogs are most possibly xenologs. The results of the phylogenetic analyses, the distribution of *xfp* genes and the location of some *xfp* genes in plasmids are independent pieces of evidence that point to horizontal gene transfer as a major driving force in the evolution of phosphoketolases.

Introduction

Phosphoketolases (XFPs) catalyze the phosphorolytic cleavage of fructose-6-phosphate and/or xylulose-5-phosphate into acetyl-phosphate plus erythrose-4-phosphate and/or acetyl-phosphate plus glyceraldehyde-3-phosphate. To date, two types of XFP activities have been described: fructose-6-phosphate phosphoketolase (F6PPK, EC 4.1.2.22), which catalyzes the conversion of fructose-6-phosphate to erythrose-4-phosphate and acetyl-phosphate, and xylulose-5-phosphate phosphoketolase (X5PPK, EC 4.1.2.9), which cleaves xylulose-5-phosphate rendering acetyl-phosphate and glyceraldehyde-3-

B.S. and M.Z. contributed equally to this work.

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

1464–1801/10/0181–0037$26.00/0

Accessible online at:
www.karger.com/mmb

Borja Sánchez
Université de Bordeaux, UMR 5248 CNRS, UBX1-ENITAB, ENITAB
1 cours du Général de Gaulle
FR–33175 Gradignan Cedex (France)
Tel. +34 985 892 131, Fax +34 985 892 233, E-Mail borja@ipla.csic.es

phosphate. These enzymes require divalent cations and thiamine pyrophosphate (TPP) for their activity [Yevenes and Frey, 2008], being the N-terminal part of the protein responsible for the TPP binding. X5PPK was first purified from *Lactobacillus plantarum* [Heath et al., 1958]. On the other hand, F6PPK was first described and purified from *Acetobacter xylinum* [Schramm et al., 1958]. Further research on F6PPKs has mainly focused on members of the genus *Bifidobacterium* where, in addition, a dual xylulose-5-phosphate/fructose-6-phosphate phosphoketolase able to act over both fructose-6-phosphate and xylulose-5-phosphate has been described [Grill et al., 1995; Meile et al., 2001].

X5PPK is the key enzyme enabling a catabolic variant of the pentose phosphate pathway (fig. 1). The pentose phosphate pathway is almost ubiquitous in bacteria and animals, and its main role is to provide the cell with anabolic requirements such as NADPH, ribose-5-phosphate and other metabolites [Portais and Delort, 2002; Wood, 1986]. The first series of nonreversible reactions (oxidative branch) of the pathway ensures the oxidative decarboxylation of glucose-6-phosphate into ribulose-5-phosphate. A second series of reactions (the nonoxidative branch) allows the conversion of ribulose-5-phosphate into other sugars, including ribose-5-phosphate, or glycolytic intermediates (fructose-6-phosphate and glyceraldehyde-3-phosphate). XFP action bypasses the nonoxidative branch by directly splitting xylulose-5-phosphate into glyceraldehyde-3-phosphate and acetyl-phosphate. Furthermore, F6PPK is the key enzyme of the so-called bifid shunt (fig. 1), the only metabolic pathway by which members of the genus *Bifidobacterium* convert hexoses into acetic and lactic acids [Meile et al., 2001]. In fact, the presence of F6PPK activity has been used for decades for the identification of *Bifidobacterium* species since this metabolic activity differentiated them from the rest of the intestinal microbiota [Vlkova et al., 2005].

XFPs are present in both eukaryotes and bacteria. Previous structural [Duggleby, 2006] and phylogenetic analyses [Costelloe et al., 2008] of proteins belonging to the TPP-dependent family have shown that XFPs share the same domain structure and are evolutionarily related to transketolases. However, no detailed phylogenetic study of XFPs has been carried out so far. In the present work, a phylogenetic analysis of *xfp* genes from 650 genomes has been carried out at both the amino acid and nucleotide levels. The results obtained indicate that horizontal gene transfer (HGT) has been a major force in the evolution of these enzymes.

Results

Distribution and Genomic Context of XFP-Encoding Genes

A total of 650 genomes were screened for genes encoding putative XFPs. This data set included 46 *Archaea*, 547 *Bacteria* and 57 *Eukarya* genomes (table 1 and online suppl. table 1, www.karger.com/doi/10.1159/000274310). A total of 151 putative *xfp* genes were identified in the screening from which redundant sequences and possible pseudogenes were excluded, thus obtaining a data set of 128 sequences. XFP-encoding genes were found in *Bacteria* and *Eukarya*, whereas no putative XFP-encoding gene was detected in *Archaea*. Putative XFPs in *Eukarya* were found only in some members of the *Ascomycota* and *Basiodiomycota* phyla, whereas they were not present in protists, animals or plants.

Similarly, the distribution of XFP in *Bacteria* is quite patchy (table 1). Putative XFP genes are prevalent in very few taxonomical divisions such as *Lactobacillaceae*, where they are present in all species included in the study (online suppl. table 1), *Rhizobiales* (23 out of 28 genomes harbor at least a putative *xfp* homolog; table 1) and *Actinobacteria* (23 out of 41 genomes; table 1). In contrast, in most other bacterial groups, putative *xfp* genes are present only in a few species or strains, while they are absent from other closely related ones. For example, *Pseudomonas syringae* pv. tomato strain DC3000 harbors a putative *xfp* gene, whereas the pathovars *syringae* and *phaseolicola* do not. Another example is the case of *Lactococcus lactis*: *L. lactis* subsp. *lactis* strain IL1403 encodes a putative *xfp* gene, whereas *L. lactis* subsp. *cremoris* strains MG1363 and SK11 do not.

Most organisms considered in this study harbor only one putative *xfp* gene; however, a few strains harbor two or in one case, *Nostoc* sp. PCC 7120, three putative *xfp* homologs (table 2). Again, if present, paralogs do not always appear in all members of some taxonomic groups. For example, in *Cyanobacteria Synechococcus* sp. JA-3–3Ab, *Thermosynechococcus elongatus* BP-1, *Synechocystis* sp. PCC 6803, *Anabaena variabilis* ATCC 29413, *Nostoc* sp. PCC 7120 and *Gloeobacter violaceus* PCC 7421 harbor paralogs, whereas *Synechococcus elongatus* PCC 6301, *Synechococcus* sp. WH 7803 and *Trichodesmium erythraeum* IMS101 harbor only one putative *xfp* gene. There is also heterogeneity in the location of putative *xfp* genes: in most cases *xfp* genes are located in the chromosome, although in some cases they are encoded by plasmids (online suppl. table 1; see also some examples in table 2).

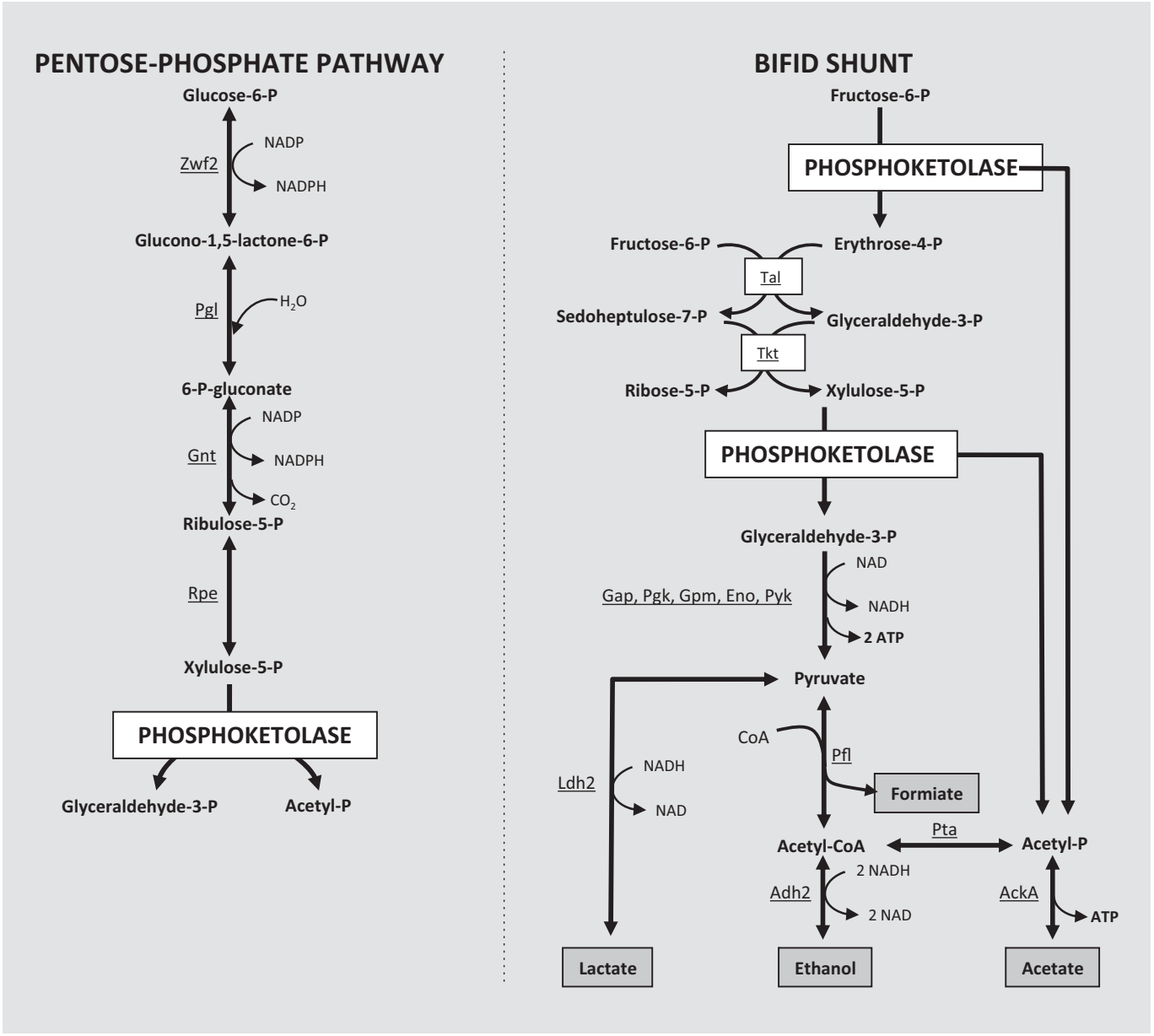


Fig. 1. Schematic representation of the reactions catalyzed by XFPs. Zwf2 = Glucose-6-phosphate dehydrogenase; Pgl = 6-phosphogluconolactonase; Gnt = 6-phosphogluconate dehydrogenase; Rpe = ribulose-phosphate 3-epimerase; Tal = transaldolase; Tkt = transketolase; Gap = glyceraldehyde-3-phosphate dehydrogenase; Pgk = phosphoglycerate kinase; Gpm = phosphoglycerate mutase; Eno = enolase, Pyk = pyruvate kinase; AckA = acetate kinase A; Ldh2 = L-lactate dehydrogenase; Pfl = pyruvate formate-lyase; Adh2 = alcohol dehydrogenase; Pta = phosphate acetyltransferase.

The genomic context of the putative *xfp* genes is highly variable and provides very limited help for understanding their evolutionary history. XFP-encoding genes can occur in monocystronic clusters or in putative operons with a great variety of genes (see fig. 2 for some examples), although no clear correlation among the genomic context and the phylogenetic reconstructions was observed except for very closely related genes. Notwithstanding, the association of *xfp* with genes encoding acetate kinases was relatively frequent, especially in *Cyano-*

Table 1. Distribution of *xfp* genes in *bacteria*

Class	Order	Screened genomes	Genomes harbouring xfp
Acidobacteria	*Acidobacteriales*	1	0
Solibacteres	*Solibacterales*	1	0
Actinobacteria	*Actinomycetales*	41	23
Actinobacteria	*Bifidobacteriales*	2	2
Actinobacteria	*Rubrobacterales*	1	0
Aquificae	*Aquificales*	1	0
Bacteroidia	*Bacteroidales*	6	0
Flavobacteria	*Flavobacteriales*	3	0
Sphingobacteria	*Sphingobacteriales*	2	0
Chlamydiae	*Chlamydiales*	11	1
Chlorobia	*Chlorobiales*	5	2
Chloroflexi	*Chloroflexales*	2	0
Dehalococcoidetes	–	3	0
Dehalococcoidetes	*Chroococcales*	12	6
Dehalococcoidetes	*Nostocales*	2	2
Dehalococcoidetes	*Oscillatoriales*	1	0
Gloeobacteria	*Gloeobacterales*	1	1
Gloeobacteria	*Prochlorales*	11	0
Unclassified *Cyanobacteria*	–	1	0
Deinococci	*Deinococcales*	2	0
Deinococci	*Thermales*	2	0
Bacilli	*Bacillales*	39	0
Bacilli	*Lactobacillales*	44	19
Clostridia	*Clostridiales*	20	2
Clostridia	*Thermoanaerobacterales*	4	0
Mollicutes	*Acholeplasmatales*	2	0
Mollicutes	*Entomoplasmatales*	1	0
Mollicutes	*Mycoplasmatales*	14	1
Fusobacteria	*Fusobacteriales*	1	0
Planctomycetacia	*Planctomycetales*	1	1
Alphaproteobacteria	*Caulobacterales*	1	0
Alphaproteobacteria	*Rhizobiales*	28	23
Alphaproteobacteria	*Rhodobacterales*	10	1
Alphaproteobacteria	*Rhodospirillales*	5	1
Alphaproteobacteria	*Rickettsiales*	16	0
Alphaproteobacteria	*Sphingomonadales*	5	0
Betaproteobacteria	*Burkholderiales*	33	3
Betaproteobacteria	*Hydrogenophilales*	1	1
Betaproteobacteria	*Methylophilales*	1	1
Betaproteobacteria	*Neisseriales*	4	0
Betaproteobacteria	*Nitrosomonadales*	3	3
Betaproteobacteria	*Rhodocyclales*	3	1
Deltaproteobacteria	*Bdellovibrionales*	1	0
Deltaproteobacteria	*Desulfobacterales*	1	0
Deltaproteobacteria	*Desulfovibrionales*	4	0
Deltaproteobacteria	*Desulfuromonadales*	5	0
Deltaproteobacteria	*Myxococcales*	3	2
Deltaproteobacteria	*Syntrophobacterales*	1	0
Epsilonproteobacteria	*Campylobacterales*	17	0
Epsilonproteobacteria	unclassified *Epsilonproteobacteria*	2	0
Gammaproteobacteria	*Aeromonadales*	2	0
Gammaproteobacteria	*Alteromonadales*	19	13
Gammaproteobacteria	*Cardiobacteriales*	1	0
Gammaproteobacteria	*Chromatiales*	3	1

Table 1 (continued)

Class	Order	Screened genomes	Genomes harbouring xfp
Gammaproteobacteria	*Enterobacteriales*	45	0
Gammaproteobacteria	*Legionellales*	6	0
Gammaproteobacteria	*Methylococcales*	1	1
Gammaproteobacteria	*Oceanospirillales*	1	1
Gammaproteobacteria	*Pasteurellales*	10	0
Gammaproteobacteria	*Pseudomonadales*	18	7
Gammaproteobacteria	*Thiotrichales*	8	0
Gammaproteobacteria	unclassified *Gammaproteobacteria*	4	0
Gammaproteobacteria	*Vibrionales*	9	0
Gammaproteobacteria	*Xanthomonadales*	8	0
unclassified *Proteobacteria*	–	1	0
Spirochaetes	*Spirochaetales*	10	0
Thermotogae	*Thermotogales*	5	1

bacteria and *Proteobacteria*. Some cases, such as the δ-*Proteobacteria Anaeromyxobacter* sp. Fw109-5 and *Anaeromyxobacter dehalogenans* 2CP-C, harbor genes encoding putative fusion acetate kinase-XFP proteins (fig. 2).

Phylogenetic Reconstructions

Previous phylogenetic analyses have suggested a common origin for XFPs, transketolases, 2-oxoisovalerate dehydrogenases and dihydroxyacetone synthases [Costelloe et al., 2008]. Although not shown in this study, our preliminary phylogenetic analyses confirmed that XFPs constitute a cluster clearly distinguishable from the other three enzymes.

The best evolutionary model for amino acid sequences under the Akaike Information Criterion (AIC) was rtREV [Dimmic et al., 2002] with a gamma distribution accounting for heterogeneities in evolutionary rates among sites. The likelihood mapping analysis of the 128 amino acid sequences alignment indicated that the dataset contained a fair amount of phylogenetically informative signal, with 86.9% fully resolved quartets. Only 5.2% of the quartets were completely unresolved; therefore, a completely resolved phylogeny was not expected. Notwithstanding, the phylogenetic reconstruction using maximum likelihood allowed distinguishing three highly supported groups (fig. 3). Group 1 encompassed 110 sequences (73.8%) including both bacterial and eukaryotic XFPs (fig. 4). Group 2 was constituted by sequences from *Proteobacteria* (fig. 5a), and group 3 contained sequences from *Cyanobacteria* and one sequence from the β-proteobacterium *Thiobacillus denitrificans* (fig. 5b). One of the putative XFPs encoded by *Nocardia farcinica* IFM 10152 plasmid pNF1 (gene pnf11130, nfarcin2) clustered apart, thus forming a fourth group (fig. 3). The large phylogenetic distances between the three groups indicate either an ancestral divergence or a high evolutionary rate for the sequences of groups 2 and 3. However, the deep branching of these groups (fig. 3) would be in agreement with an ancestral divergence. Unfortunately, no structural or functional information on the proteins belonging to groups 2 and 3, which might help to gain insight in this issue, is available.

The low similarity among the sequences of the different groups limited the utility of the joint phylogenetic analysis. Therefore, in order to gain resolution in the phylogenetic reconstructions, trees for each group identified in the previous analysis of amino acid sequences were constructed using their cognate DNA sequences (fig. 4 and 5). The best evolutionary model for the three groups was GTR. In order to determine whether saturation in substitutions at third codon positions would affect the phylogenetic reconstructions, these were also performed excluding third bases. The topologies of the corresponding trees were very similar, but higher support for the nodes was obtained in the trees from complete codons which were used in further analyses. The likelihood mapping analyses showed strong phylogenetic signals (93.9, 92 and 95.7% resolved quartets for groups 1, 2 and 3, respectively).

Overall, the phylogenetic reconstructions showed a clear incongruence with the expected order of organismal descent with the relevant exception of the cyanobacterial sequences in group 3. The phylogenetic reconstruc-

Table 2. Organisms harboring more than one *xfp* gene

Taxonomy	Species	Paralogs	Alias	Group	Location
Actinobacteria	*Nocardia farcinica*	nfa13300	nfarcin1	1	
Actinomycetales	IFM 10152	pnf11130	nfarcin2	4	plasmid pNF1
Nocardiaceae					
Cyanobacteria	*Synechococcus* sp.	CYA_1981	scoccus1	1	
Chroococcales	JA-3-3Ab	CYA_0447	scoccus2	3	
	Synechocystis sp. PCC	slr0453	synech1	1	
	6803	sll0529	synech2	3	
	Thermosynechococcus	tll1186	telong1	1	
	elongatus BP-1	tll1846	telong2	3	
Cyanobacteria	*Anabaena variabilis*	Ava_0496	avariab1	1	
Nostocales	ATCC 29413	Ava_4264	avariab2	3	
Nostocaceae	*Nostoc* sp. PCC 7120	all2567	nostoc1	1	
		all1483	nostoc2	1	
		alr1850	nostoc3	3	
Cyanobacteria	*Gloeobacter violaceus*		gviolac1	1	
	PCC	glr0997			
Gloeobacterales	7421	glr3073	gviolac2	3	
Firmicutes	*Lactobacillus plantarum*	xpk1	lbplan1	1	
Lactobacillales	WCFS1	xpk2	lbplan2	1	
Lactobacillaceae	*Lactobacillus salivarius* subsp.	LSL_1509	lbsali1	1	
	salivarius UCC118	LSL_1956	lbsali2	1	plasmid pMP118
Firmicutes	*Leuconostoc mesenteroides* subsp. *mesenteroides*	LEUM_1456	lmesen1	1	
Lactobacillales	ATCC 8293	LEUM_1961	lmesen2	1	
Leuconostocaceae	*Oenococcus oeni* PSU-1	OEOE_1183	ooeni1	1	
		OEOE_1812	ooeni2	1	
Alphaproteobacteria	*Nitrobacter hamburgensis* X14	Nham_4367	nhambur1	1	plasmid pPB12
Rhizobiales					
Bradyrhizobiaceae		Nham_1896	nhambur2	1	
Alphaproteobacteria	*Mesorhizobium* sp. BNC1	Meso_2236	mesorhi1	1	
Rhizobiales					
Phyllobacteriaceae		Meso_3878	mesorhi2	2	
Alphaproteobacteria	*Rhizobium leguminosarum* bv.	RL3902	rlegum1	1	
Rhizobiales	*viciae* 3841	RL066	rlegum2	1	
Rhizobiaceae	*Sinorhizobium meliloti* 1021	SMa1084	smelit1	1	plasmid pSymA
		SMc04146	smelit2	1	
Betaproteobacteria	*Acidovorax* sp. JS42	Ajs_1497	acidov1	2	
Burkholderiales					
Comamonadaceae		Ajs_2678	acidov2	2	
Betaproteobacteria	*Thiobacillus denitrificans* ATCC	tbd_0831	tdenitr1	1	
Hydrogenophilales					
Hydrogenophilaceae	25259	tbd_0049	tdenitr2	3	
Eurotiomycetes	*Aspergillus fumigatus*	AFUA_3G00370	afumigat1	1	
Eurotiales	Af293	AFUA_3G10760	NA		
Trichocomaceae	*Aspergillus terreus*	ATEG_07454	aterreu1	1	
	NIH2624	ATEG_04606	aterreu2	1	
Sordariomycetes	*Neurospora crassa*	NCU06123	ncrassa1	1	
Sordariales	OR74A				
Sordariaceae		NCU05151	ncrassa2	1	
Tremellomycetes	*Cryptococcus neoformans* var.	CNK00070	cneoform1	1	
Tremellales					
Tremellaceae	*neoformans* JEC21	CNE03100	cneoform2	1	

NA = Not included in this study.

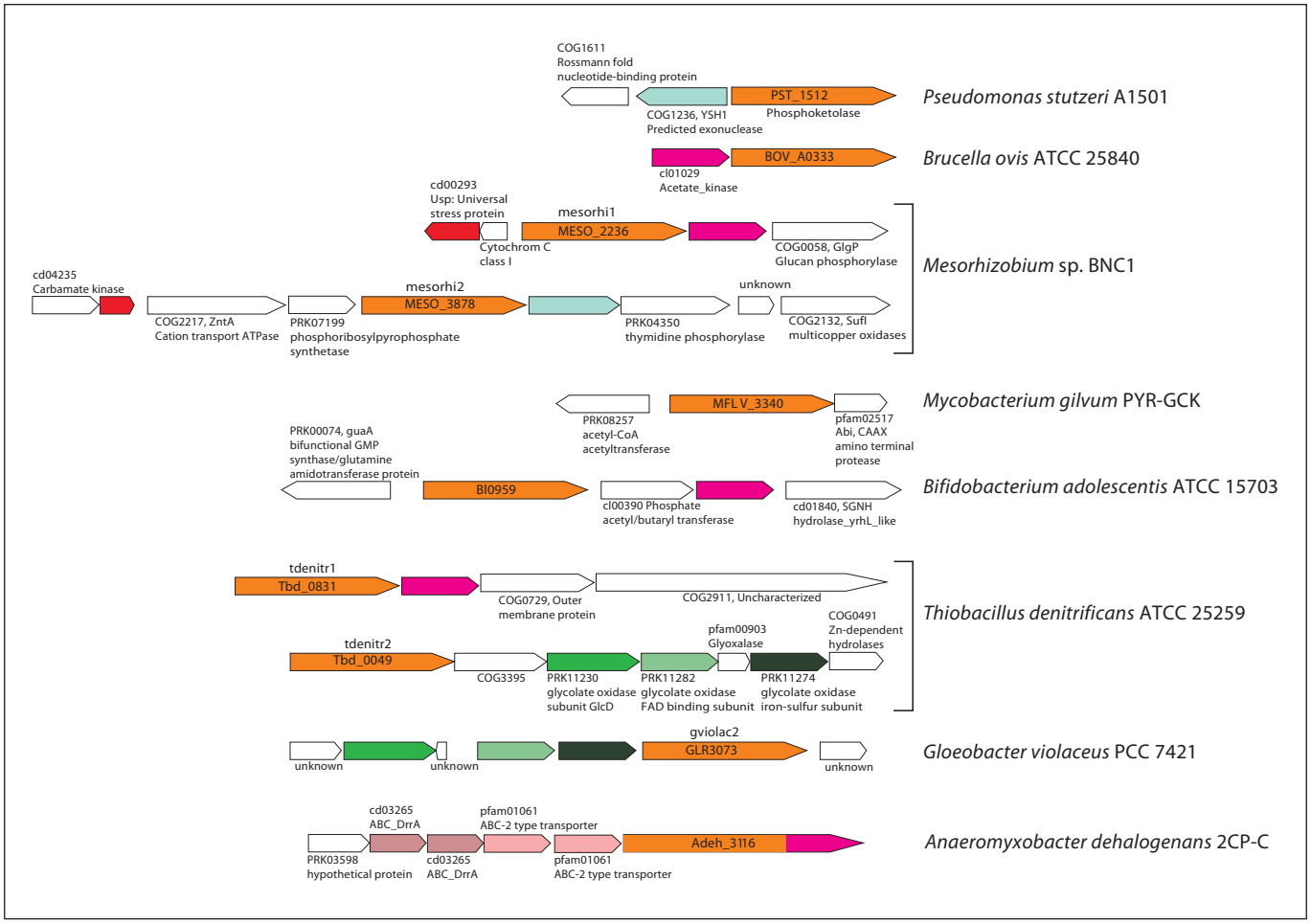


Fig. 2. Schematic representation of selected clusters containing *xfp* genes. Colors indicate homologous genes. Unknown indicates that no hit was found in domain databases.

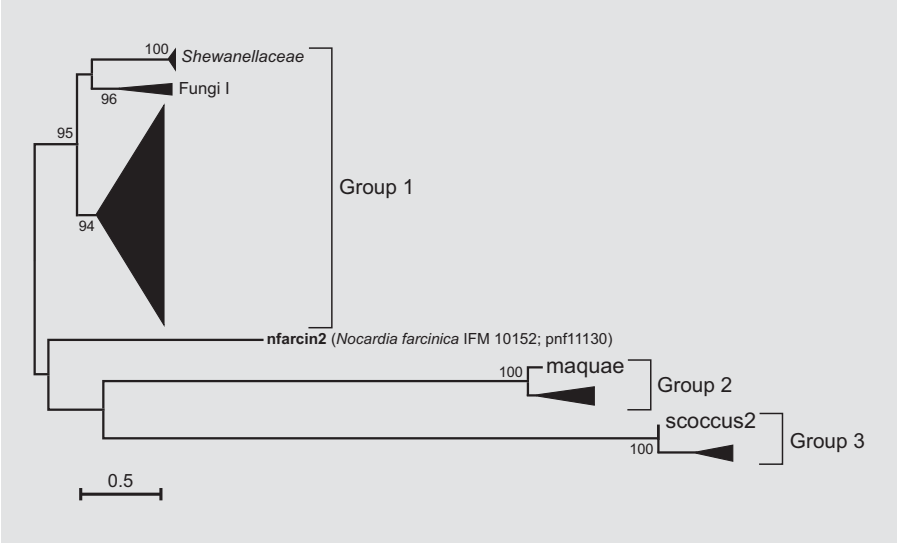
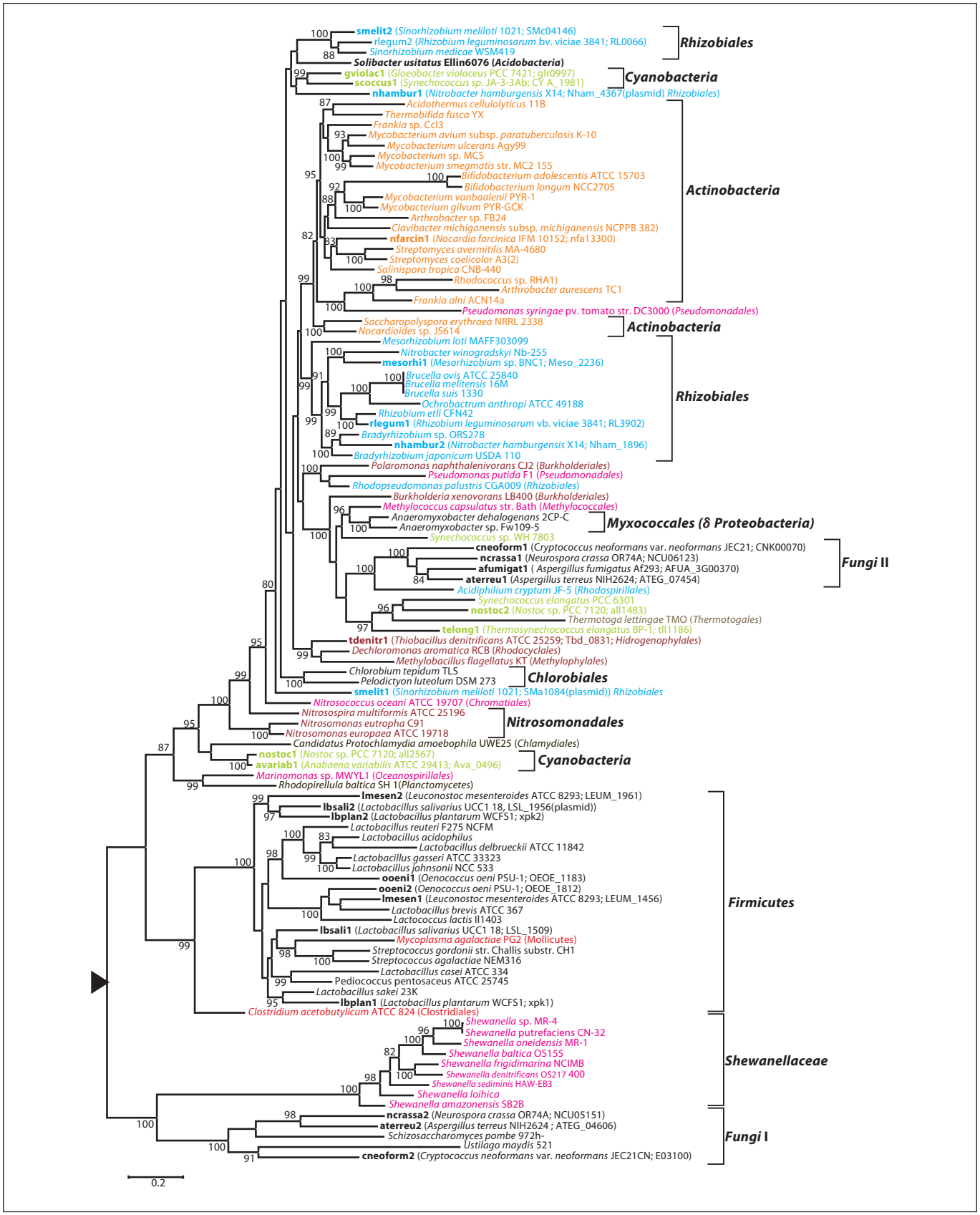


Fig. 3. Maximum likelihood topology derived from the alignment of amino acid sequences of the XFPs used in this study. The tree is arbitrarily rooted. The main groups derived from the analysis are indicated as well as the support values for the bootstrap analysis higher than 90%. The sequence maquae corresponds to *Marinobacter aquaeolei* VT8 gene Maqu_3607; scoccus2, *Synechococcus* sp. JA-3-3Ab gene CYA_0447.



Color version available online

tion of these sequences was congruent with that based on their 16S rRNA ($p = 0.49$ in the Shimodaira-Hasegawa test; fig. 6c). Within group 1 (fig. 4), three well-supported subgroups could be distinguished: the first spanned sequences of *Actinobacteria*, *Cyanobacteria*, *Proteobacteria* and other bacteria as well as some fungal *xfp*s; the second one was a monophyletic cluster encompassing all XFPs from *Firmicutes*, and the third one was constituted by the putative XFPs of *Shewanella* and the second cluster of fungal XFPs. Even monophyletic clusters such as *Firmicutes* or *Actinobacteria* (excepting the putative XFP gene harbored by *P. syringae*) showed topologies that were not congruent with their counterparts based on 16S rRNA genes ($p < 0.05$ in the SH test; fig. 6a). These cases of lack of congruence cannot be explained as artifacts in the phylogenetic reconstructions since XFPs sequences within each group are highly conserved (for example, two of the more distant sequences in group 1, *Ustilago maydis* and *Thermotoga lettingae* share 39% identical amino acids; fig. 4) and most nodes had bootstrap support values higher than 80%.

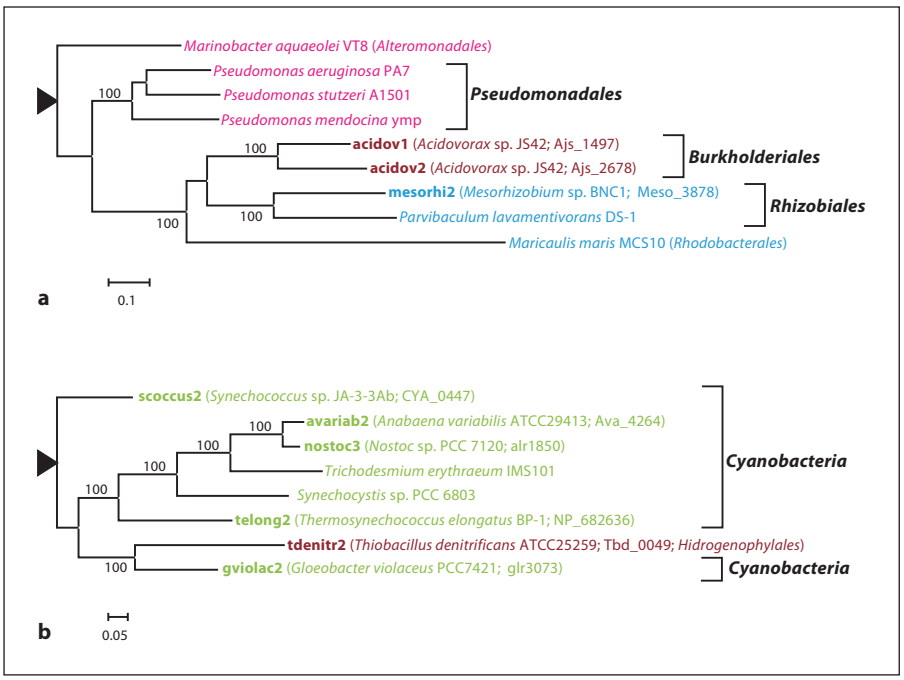
Discussion

The physiological role of XFP is unclear in some organisms since this enzyme is not essential for pentose utilization. For example, model organisms such as *Escherichia coli* [Fraenkel, 1987] or *Bacillus subtilis* [Stulke and Hillen, 2000] can grow on pentoses without XFP. However, a recent study on *Aspergillus nidulans* suggests that XFP provides greater flexibility to central metabolism: higher specific growth rates and growth yields were observed with xylose, glycerol or ethanol when XFP was overexpressed in *A. nidulans*. Moreover, the overexpression of XFP resulted in high conversion yields of sugars to secondary metabolites originating from acetyl-CoA at the expense of a lower flux through glycolysis [Panagiotou et al., 2008]. Theoretically, the degradation of pentoses through XFP would be energetically advantageous over the classical pentose phosphate pathway since an additional ATP can be gained from degradation of acetyl-P to acetate [Wolfe, 2005]. In this sense, the association of *xfp* to genes encoding acetate kinases may be an indication of the assembling of this pathway in some *Proteobacteria* and *Cyanobacteria*. The functional association between XFP and acetate kinase has also been evidenced in *Bifidobacterium lactis*, in which increases in acetic acid production through XFP overproduction were suggested as one of the most plausible mechanisms by which this organism optimizes its energetic metabolism in the presence of bile salts [Sánchez et al., 2005]. In summary, the functional information currently available indicates that XFP provides an alternative, energetically more efficient, catabolic pathway to the classical pentose phosphate pathway. This characteristic of being alternative to a central glycolytic pathway helps to understand the nonverticality of the evolution of the *xfp* gene.

The phylogenetic analysis of putative *xfp* genes shows a clear disagreement with the expected order of organismal descent. The strong phylogenetic signal and support for most nodes in the phylogenetic reconstructions allow us to rule out methodological artifacts. Therefore, the observed phylogenetic reconstructions and distribution of putative *xfp* genes must be explained by considering other phenomena such as recent and ancestral duplications, lineage-specific gene losses and HGTs. HGT is now accepted as a major factor in the evolution of prokaryotes. HGT among closely related bacteria is well characterized and the mechanisms that mediate gene transfers are known. Furthermore, a number of studies have also shown that HGT can occur among distantly related organisms [Beiko et al., 2005; Lerat et al., 2005; Nakamura et al., 2004; Raymond et al., 2002]. In a rigorous analysis of 144 prokaryotic genomes, Beiko et al. [2005] found that genes involved in sugar metabolism are among the most prone to HGT. Recently, Dagan et al. [2008] analyzed 188 prokaryotic genomes and estimated that at least 81 of the genes in each genome were involved in HGTs at some point in their history.

Several species included in this study harbored putative *xfp* paralogs. With the possible exception of *Acidovorax* sp. (fig. 5a), these paralogs do not seem to result from recent duplication events since paralogs from the same species regularly appear as distantly related to each other. However, several different situations can be distinguished: duplicated *xfp*s can be located in different clusters within the same group (for example fungal *xfp*s)

Fig. 4. Maximum likelihood topology derived from the alignment of nucleotide sequences of the *xfp* genes belonging to group 1. The tree is rooted according to the phylogenetic reconstruction of amino acid sequences (fig. 3). Bootstrap support values for nodes with bootstrap higher than 80% are shown. Relevant taxonomic groups are indicated. Coloring scheme: blue, α-*Proteobacteria*; brown, β-*Proteobacteria*; green, *Cyanobacteria*; orange, *Actinobacteria*; pink, γ-*Proteobacteria*; red, *Firmicutes* (excluding *Lactobacillales*). Paralogs are indicated by their corresponding aliases.



Color version available online

Fig. 5. Maximum likelihood topology derived from the alignment of nucleotide sequences of the *xfp* genes belonging to groups 2 (**a**) and 3 (**b**). The trees are rooted according to the phylogenetic reconstruction of amino acid sequences (fig. 3). Bootstrap support values for nodes with bootstrap higher than 80% are shown. Relevant taxonomic groups are indicated. Coloring scheme as indicated in figure 4. Paralogs are indicated by their corresponding aliases.

or in different groups, such as some *xfp*s from *Cyanobacteria*, *Proteobacteria* and *N. farcinica* IFM 10152. In some cases, one of the copies was chromosomally encoded whereas the other was located in a plasmid (table 2), suggesting that some of these paralogs are actually xenologs. Although direct evidence of the involvement of any of these plasmids in HGT is not available, there are some indications suggesting that at least some of them have been involved in HGT. The analysis of *Synorhizobium meliloti* 1021 megaplasmid pSymA evidenced a mosaic structure possibly due to recombination with closely related bacteria [Guo et al., 2007]. Furthermore, a cluster of genes putatively involved in conjugal transfer and a putative *oriT* can also be found in this megaplasmid [Galibert et al., 2001]. *Nitrobacter hamburgensis* plasmid pPB12 (encoding the *xfp* gene nhambur1) also harbors a gene cluster for conjugal transfer [Starkenburg et al., 2008].

There are additional indications of the likely role of HGT in the evolution of putative *xfp* genes: several *xfp* genes are annotated as putatively acquired by HGT in the Horizontal Gene Transfer database (HGT-DB) [García-Vallvé et al., 2003] such as *Mesorhizobium* sp. BNC1 mesorhi2, *N. hamburgensis*, nhambur2, *Nitrosococcus oceani* noc_2717 and *Rhizobium leguminosarum* rlegum2, although not all genomes studied here are included in the HGT-DB. This database compiles statistical parameters such as G+C content, codon and amino acid usage and identifies genes putatively acquired by HGT on the basis of deviations of these parameters from the values of their cognate genomes. Furthermore, the position of some sequences in otherwise monophyletic clusters such as *P. syringae* pv. tomato (fig. 4) or *T. denitrificans* tdenitr2 (fig. 5b) also points to HGT events.

Fungal xfp *Genes*

Fungal sequences appear in two clusters, Fungi I together with sequences of *Shewanella* and Fungi II with *Acidiphilium cryptum* (fig. 4). Organisms harboring Fungi II sequences also harbor an additional *xfp* gene from Fungi I (in the case of *Aspergillus fumigatus* the corresponding sequence of Fungi I was not included in the study; see online suppl. table 1). Fungi II sequences appear as a monophyletic cluster branching within a large group of bacterial sequences. Their branch lengths are similar to their bacterial counterparts. Taken together, these observations strongly suggest that Fungi II sequences are of bacterial origin and that the transference from bacteria to fungi occurred only once. However, the position of Fungi I is more debatable: the branching order of Fungi I sequences, although they include only five species, is in agreement with the expected order of organism descent, suggesting that this gene evolved or was acquired before the differentiation of *Ascomycota* and

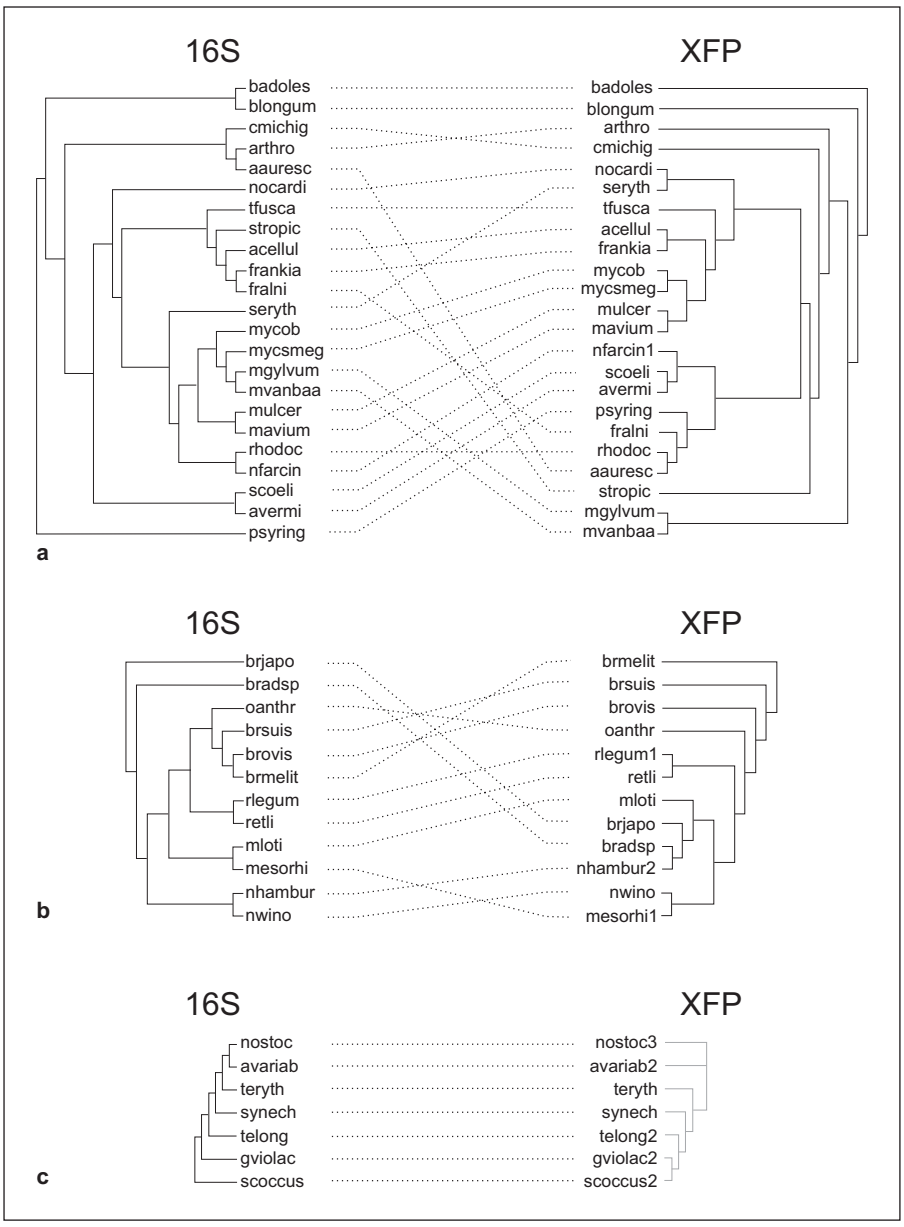


Fig. 6. Comparison between topologies of the maximum likelihood trees derived for *xfp* genes and their corresponding 16S rRNAs. **a** Comparison of actinobacterial sequences. **b** Comparison of rhizobial sequences. **c** Group 3 cyanobacterial sequences. See online suppl. table 2 (www.karger.com/doi/10.1159/000274310) for aliases of the sequences.

Basidiomycota. The close phylogenetic relationship with sequences of *Shewanella* suggests a possible transfer, although our results do not allow determining its direction.

Incongruence of Group 1 Monophyletic Clusters with the Expected Order of Organismal Descent

Within group 1, three large, supported monophyletic clusters were observed (fig. 4). The first group corresponded to *Actinobacteria*, with the exception of the sequence from *P. syringae* pv. tomato); the second group encompassed *Firmicutes* and the third group included *Rhizobiales*, although additional *xfp* genes from this family can be found in other clusters. Remarkably, none of these clusters agrees with the phylogenies derived from 16S rRNA ($p < 0.05$ in the SH test), thus indicating that each cluster originated from a single common ancestor and that additional events apart from vertical inheritance have occurred during their subsequent evolution.

In *Firmicutes*, paralogy (table 2; fig. 4) can account for some discrepancies but it cannot explain the positions of *Mycoplasma agalactiae*, *L. lactis* IL1403 and streptococ-

cal sequences within a cluster of *Lactobacillaceae* sequences (fig. 4). On the other hand, the basal position of *Clostridium acetobutylicum* suggests that this sequence was not transferred from lactobacilli or that the transfer occurred at an early stage in the evolution of these groups. Therefore, the phylogenetic analysis indicates that putative *xfp* genes of *Firmicutes* evolved from a common ancestor, although it cannot be established whether this sequence was present in the last common ancestor of *Firmicutes* and subsequently lost in most lineages orwhether it was acquired by a member of *Lactobacillaceae* and transferred to other *Firmicutes*. In any case, this putative ancestral *xfp* gene was clearly differentiated from the ancestors of the other subgroups within group 1.

A clear incongruence between the phylogenetic reconstruction of actinobacterial *xfp* genes and the presumed order of organismal descent is also observed (fig. 6a). With the exception of *N. farcinica*, no *Actinobacteria* harbor putative *xfp* paralogs. Although hidden paralogy might explain the observed phylogenetic relationships, several paralogs should have been present in the last common ancestor of *Actinobacteria*. The optimal reconstruction using TreeMap in order to reconcile the *xfp* and 16S trees required eight HGT events and seventeen sorting events (that is, lineage-specific gene losses or segregation of paralogs). HGT may also explain the phylogenetic reconstruction: under this view, the putative *xfp* gene could have been transferred once from a nonactinobacterial donor (possibly a proteobacteria considering the position of the actinobacterial cluster; fig. 4) to an *Actinobacteria* receptor and subsequently disseminated among actinobacteria by HGT.

Cases of Proteobacterial and Cyanobacterial Sequences Point to HGT as a Major Driving Force in xfp *Evolution*

In contrast to the situation of *Actinobacteria* and *Firmicutes*, proteobacterial and cyanobacterial *xfp* genes appear in most clusters of the trees in clear contradiction with the expected order of organismal descent. This is particularly illustrated by gamma proteobacterial *xfp* genes (fig. 4 and 5b). Even in a monophyletic cluster such as that constituted by some sequences of *Rhizobiales* (fig. 4), the branching order is clearly incongruent with that derived from 16S rRNA sequences (SH < 0.05; fig. 6b). The presence of paralogs may be invoked as a source of incongruence; however, the phylogenetic reconstructions suggest that most paralogs, if not all, are actually xenologs. The optimal reconstruction using TreeMap required six duplications, eight HGT events and forty-five sorting events. The phylogenetic reconstruction indicates that the large cluster within group 1 which includes actinobacterial and most proteobacterial sequences evolved from a common ancestor (fig. 4). Considering only the most basal nodes of this cluster, at least six paralogs should be postulated in the last common ancestor of the bacteria present in the subgroup. Subsequently, the evolution of these bacteria would have been accompanied by the loss of most paralogs, and only in a very limited number of taxa some of them would have been conserved. For example, to explain the clustering of *Marinomonas* sp. MWYL1 and *Rhodopirellula baltica* SH 1 (fig. 4), a paralog present in the last common ancestor of both organisms should be postulated which would have been subsequently lost in all derived taxa except in these two species.

Furthermore, the phylogenetic positions of some sequences clearly point to HGT events as a better alternative explanation of the observed distribution. In addition to the cases already discussed, group 3 includes a putative XFP-encoding gene of the β-*Proteobacterium T. denitrificans* (tdenitr2; fig. 5b). As indicated above, the phylogenetic reconstruction of the cyanobacterial sequences of this group agrees with the expected order of organismal descent (fig. 6c), thus strongly suggesting that this gene has been inherited vertically within this group and that *T. denitrificans* acquired this gene from a cyanobacterial donor. The comparison of the genetic context of tdenitr2 and its most closely related counterpart, gviolac2, also shows remarkable similarities (fig. 2): both genes are clustered together with genes encoding the subunits GlcD, GlcE and GlcF of glycolate dehydrogenase [Eisenhut et al., 2006]. This arrangement is not found for any other putative XFP-encoding gene.

Summarizing the evidence discussed above, the phylogenetic reconstructions, the distribution of putative *xfp* genes, the location of some *xfp* genes in plasmids and, in some cases, the genetic context all point to HGT as a major mechanism explaining the distribution and evolution of *xfp* genes.

Xylulose-5-P and Fructose-6-P Phosphoketolases Cannot Be Distinguished on the Basis of Their Phylogenetic Relationships

Functional information on XFPs is limited to a few bacteria, mostly lactobacilli and bifidobacteria, and fungi. The available evidence has shown that some XFPs act over both fructose-6-phosphate and xylulose-5-phosphate [Meile et al., 2001; Schramm et al., 1958], whereas in some cases the use of only one of the two possible sub-

strates has been reported, like the XFP of *Thiobacillus novelus* [Greenley and Smith, 1978] or *Fibrobacter succinogenes* [Matheron et al., 1997], which can use only xylulose-5-phosphate. Unfortunately, the sequences of these XFPs are not available. Also, the existence of two types of XFPs, a fructose-6-phosphate-specific enzyme present in human-associated species like *Bifidobacterium dentium*, and a less specific xylulose 5-phosphate/fructose 6-phosphate phosphoketolase present in animal-associated species like *Bifidobacterium animalis* and *Bifidobacterium globosum* has been described [Grill et al., 1995; Meile et al., 2001; Sgorbati et al., 1976]. However, nucleotide sequences of different *Bifidobacterium* sp. *xfp*s do not allow differentiating between the two types [Yin et al., 2005] and both bifidobacterial sequenced genomes harbor only one *xfp* gene. It is also possible that different experimental approaches may have led to contradictory results. For example, the XFP of *L. plantarum* was originally described as specific for xylulose-5-phosphate [Heath et al., 1958]; however, recent results indicate that both XFPs of *L. plantarum* have dual activity [Yevenes and Frey, 2008]. In summary, differences in substrate preference cannot be correlated to phylogenetic clusters and possibly most enzymes have a dual activity.

Conclusions

The analysis of putative *xfp* genes shows that HGT has been a major driving force in the evolution of XFPs. In relation to their functional role, we hypothesize that the acquisition of *xfp*s enabled the assembly of an efficient catabolic pathway for pentose utilization in many organisms. Furthermore, although limited, the available functional information indicates that XFP has been inserted in different metabolic pathways by different host organisms. For example, in bifidobacteria the acquisition of *xfp* possibly allowed the organization of a particular glycolytic pathway. The presence of paralogs in some organisms points to the involvement of XFP in more than one pathway. Therefore, the acquisition of functional information about these XFPs would greatly help to understand the evolution of this enzyme.

Experimental Procedures

Sequences

Six hundred and fifty genomes from 435 different species available at the NCBI repository (on October 2007) were screened for genes encoding putative XFPs by using PSI-BLAST and TBLASTN [Altschul et al., 1990, 1997]. The *Bifidobacterium longum* sequence (Acc. No. NP_696135) was used to query the GenBank database. The search was iterated until no additional hits were retrieved. The sequences retrieved were then selected attending to the presence of their characteristic domains (pfam 09364, XPK_N; pfam03894, XPK; pfam09363, XPK_C) or coverage of at least 75% of the query sequence. Subsequently, additional PSI-BLAST searches using the most distant sequences among those selected were performed in order to retrieve possible homologs not identified in the first search. Finally, TBLASTN searches were performed against the available complete genome sequences using the same query sequences as for the PSI-BLAST searches. In a few cases, the TBLASTN search detected significant similarity upstream their annotated translational start sites. These sequences, suspected to be possible pseudogenes or containing sequencing or annotation mistakes, were excluded from the analysis (see online suppl. table 1). The data set was subsequently refined by excluding redundant sequences. Finally, 128 sequences were included in the analysis.

Alignment and Phylogeny Reconstruction

A multiple alignment of amino acid sequences was obtained using ClustalW [Thompson et al., 1994] and manually corrected where necessary. The MEGA 4 package [Tamura et al., 2007] was used to derive the multiple alignments of nucleotide sequences by introducing gaps according to the previously derived amino acid alignment. Positions of doubtful homology or introducing phylogenetic noise due to an excessive number of gaps were removed using Gblocks [Castresana, 2000]. The final multiple alignments used for the analyses can be found in online suppl. table 1.

In order to obtain accurate phylogenies, the best fit model of amino acid or nucleotide substitution was selected using the programs ProtTest [Abascal et al., 2005] and ModelTest [Posada and Crandall, 1998], respectively. The AIC, which allows for a comparison of likelihoods from nonnested models, was adopted to select the best models [Akaike, 1974]. For the protein data set, the model chosen was rtREV [Dimmic et al., 2002], and GTR [Lanave et al., 1984] for the nucleotide data sets. The selected models were implemented in PHYML [Guindon and Gascuel, 2003] to obtain maximum likelihood trees for the different alignments. Bootstrap support values were obtained from 1,000 pseudorandom replicates. The phylogenetic signal contained in the different data sets was assessed by likelihood mapping [Strimmer and von Haeseler, 1997] using Tree-Puzzle 5.2 [Schmidt et al., 2002]. Since the rtREV model is not implemented in Tree-Puzzle, the WAG model [Jones et al., 1992] of amino acid evolution (the second best model selected by ProtTest) was used. For amino acid sequences, the rtREV model of amino acid evolution with a discrete gamma distribution to account for heterogeneity in evolutionary rates among positions in the multiple alignments was used. For nucleotide sequences, the GTR model of nucleotide evolution with gamma correction and an estimation of the proportion of invariant sites was used.

Phylogenetic trees of sequences encoding 16S rRNA were obtained using the tools implemented in the Ribosomal Database Project II [Cole et al., 2007] and were considered as standard reference trees. Comparisons between each maximum likelihood-derived tree for putative *xfp* genes and the 16S rRNA topology were carried out for selected groups. Shimodaira-Hasegawa's test [Shimodaira and Hasegawa, 1999] implemented in the program

TreePuzzle 5.2 was used to determine whether the likelihood of the data associated with the two test trees was significantly different at an α-level of 0.05 (a value above the threshold indicating a nonsignificant difference). Congruence between pairs of topologies was also evaluated and graphically represented using TreeMap [Page, 1994].

Acknowledgements

Borja Sánchez was the recipient of a Clarín postdoctoral contract from the Gobierno del Principado de Asturias funded by the Plan de Ciencia, Tecnología e Innovación de Asturias 2006–2009. This work was partially financed by FEDER funds (EU) and the Spanish Plan Nacional de I+D+i through projects AGL2007-61805 and BFU2008-03000.

References

Abascal F, Zardoya R, Posada D: ProtTest: selection of best-fit models of protein evolution. Bioinformatics 2005;21:2104–2105.

Akaike H: A new look at the statistical model identification. IEEE Trans Automat Contr AC 1974;19:716–723.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. J Mol Biol 1990;215:403–410.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.

Beiko RG, TJ Harlow, Ragan MA: Highways of gene sharing in prokaryotes. Proc Natl Acad Sci USA 2005;102:14332–14337.

Castresana J: Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 2000;17:540–552.

Cole JR, Chai B, Farris RJ, Wang Q, Kulam-Syed-Mohideen AS, McGarrell DM, Bandela AM, Cardenas E, Garrity GM, Tiedje JM: The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. Nucleic Acids Res 2007;35:D169–D172.

Costelloe SJ, Ward JM, Dalby PA: Evolutionary analysis of the TPP-dependent enzyme family. J Mol Evol 2008;66:36–49.

Dagan T, Artzy-Randrup Y, Martin W: Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. Proc Natl Acad Sci USA 2008;105:10039–10044.

Dimmic MW, Rest JS, Mindell DP, Goldstein RA: rtREV: an amino acid substitution matrix for inference of retrovirus and reverse transcriptase phylogeny. J Mol Evol 2002;55:65–73.

Duggleby RG: Domain relationships in thiamine diphosphate-dependent enzymes. Acc Chem Res 2006;39:550–557.

Eisenhut M, Kahlon S, Hasse D, Ewald R, Lieman-Hurwitz J, Ogawa T, Ruth W, Bauwe H, Kaplan A, Hagemann M: The plant-like C2 glycolate cycle and the bacterial-like glycerate pathway cooperate in phosphoglycolate metabolism in cyanobacteria. Plant Physiol 2006;142:333–342.

Fraenkel DG: Glycolysis, pentose phosphate pathway, and Entner-Doudoroff pathway; in Neidhardt FC, Ingraham JL, Low KB, Magasanik B, Schaechter M, Umbarger HE (eds): *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology. Washington, American Society for Microbiology, 1987, vol 1, pp 142–150.

Galibert F, Finan TM, Long SR, Puhler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, Bothe G, Boutry M, Bowser L, Buhrmester J, Cadieu E, Capela D, Chain P, Cowie A, Davis RW, Dreano S, Federspiel NA, Fisher RF, Gloux S, Godrie T, Goffeau A, Golding B, Gouzy J, Gurjal M, Hernandez-Lucas I, Hong A, Huizar L, Hyman RW, Jones T, Kahn D, Kahn ML, Kalman S, Keating DH, Kiss E, Komp C, Lelaure V, Masuy D, Palm C, Peck MC, Pohl TM, Portetelle D, Purnelle B, Ramsperger U, Surzycki R, Thebault P, Vandenbol M, Vorholter FJ, Weidner S, Wells DH, Wong K, Yeh KC, Batut J: The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science 2001;293:668–672.

García-Vallvé S, Guzmán E, Montero MA, Romeu A: HGT-DB: a database of putative horizontally transferred genes in prokaryotic complete genomes. Nucleic Acids Res 2003;31:187–189.

Greenley DE, Smith DW: The presence of phosphoketolase EC 4.1.2.9 in *Thiobacillus novellus*. Abstr Annu Meet Am Soc Microbiol 1978;78.

Grill JP, Crociani J, Ballongue J: Characterization of fructose 6 phosphate phosphoketolases purified from *Bifidobacterium* species. Curr Microbiol 1995;31:49–54.

Guindon S, Gascuel O: A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 2003;52:696–704.

Guo X, Flores M, Morales L, García D, Bustos P, González V, Palacios R, Dávila G: DNA diversification in two *Sinorhizobium* species. J Bacteriol 2007;189:6474–6476.

Heath EC, Hurwitz J, Horecker BL, Ginsburg A: Pentose fermentation by *Lactobacillus plantarum*. I. The cleavage of xylulose 5-phosphate by phosphoketolase. J Biol Chem 1958;231:1009–1029.

Jones DT, Taylor WR, Thornton JM: The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 1992;8:275–282.

Lanave C, Preparata G, Saccone C, Serio G: A new method for calculating evolutionary substitution rates. J Mol Evol 1984;20:86–93.

Lerat E, Daubin V, Ochman H, Moran NA: Evolutionary origins of genomic repertoires in bacteria. PLoS Biol 2005;3:e130.

Matheron C, Delort AM, Gaudet G, Forano E: Re-investigation of glucose metabolism in *Fibrobacter succinogenes* using NMR spectroscopy and enzymatic assays. Evidence for pentose phosphates phosphoketolase and pyruvate formate lyase activities. Biochim Biophys Acta 1997;1355:50–60.

Meile L, Rohr LM, Geissmann TA, Herensperger M, Teuber M: Characterization of the D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase gene (xfp) from *Bifidobacterium lactis*. J Bacteriol 2001;183:2929–2936.

Nakamura Y, Itoh T, Matsuda H, Gojobori T: Biased biological functions of horizontally transferred genes in prokaryotic genomes. Nat Genet 2004;36:760–766.

Page RDM: Parallel phylogenies – Reconstructing the history of host-parasite assemblages. Cladistics 1994;10:155–173.

Panagiotou G, Andersen MR, Grotkjær T, Regueira TB, Hofmann G, et al: Systems analysis unfolds the relationship between the phosphoketolase pathway and growth in *Aspergillus nidulans*. PLoS ONE 2008;3:e3847.

Portais JC, Delort AM: Carbohydrate cycling in micro-organisms: what can (13)C-NMR tell us? FEMS Microbiol Rev 2002;26:375–402.

Posada D, Crandall KA: MODELTEST: testing the model of DNA substitution. Bioinformatics 1998;14:817–818.

Sánchez B, Champomier-Verges MC, Anglade P, Baraige F, de los Reyes-Gavilán CG, Margolles A, Zagorec M: Proteomic analysis of global changes in protein expression during bile salt exposure of *Bifidobacterium longum* NCIMB 8809. J Bacteriol 2005;187:5799–5808.

Schmidt HA, Strimmer K, Vingron M, von Haeseler A: TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 2002;18:502–504.

Schramm M, Klybas V, Racker E: Phosphorolytic cleavage of fructose-6-phosphate by fructose-6-phosphate phosphoketolase from *Acetobacter xylinum*. J Biol Chem 1958;233: 1283–1288.

Sgorbati B, Lenaz G, Casalicchio F: Purification and properties of two fructose-6-phosphate phosphoketolases in *Bifidobacterium*. Antonie Van Leeuwenhoek 1976;42:49–57.

Shimodaira H, Hasegawa M: Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 1999; 16:1114–1116.

Starkenburg SR, Larimer FW, Stein LY, Klotz MG, Chain PS, Sayavedra-Soto LA, Poret-Peterson AT, Gentry ME, Arp DJ, Ward B, Bottomley PJ: Complete genome sequence of *Nitrobacter hamburgensis* X14 and comparative genomic analysis of species within the genus *Nitrobacter*. Appl Environ Microbiol 2008;74:2852–2863.

Strimmer K, von Haeseler A: Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc Natl Acad Sci 1997;94:6815–6819.

Stulke J, Hillen W: Regulation of carbon catabolism in Bacillus species. Annu Rev Microbiol 2000;54:849–880.

Tamura K, Dudley J, Nei M, Kumar S: MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 2007;24:1596–1599.

Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–4680.

Vlkova E, Nevoral J, Jencikova B, Kopecny J, Godefrooij J, Trojanova I, Rada V: Detection of infant faecal bifidobacteria by enzymatic methods. J Microbiol Methods 2005;60:365–373.

Wolfe AJ: The acetate switch. Microbiol Mol Biol Rev 2005;69:12–50.

Wood T: Physiological functions of the pentose phosphate pathway. Cell Biochem Funct 1986;4:241–247.

Yevenes A, Frey PA: Cloning, expression, purification, cofactor requirements, and steady state kinetics of phosphoketolase-2 from *Lactobacillus plantarum*. Bioorg Chem 2008; 36:121–127.

Yin X, Chambers JR, Barlow K, Park AS, Wheatcroft R: The gene encoding xylulose-5-phosphate/fructose-6-phosphate phosphoketolase (xfp) is conserved among *Bifidobacterium* species within a more variable region of the genome and both are useful for strain identification. FEMS Microbiol Lett 2005;246: 251–257.