Folia Microbiol. **49** (3), 247–258 (2004) **http://www.biomed.cas.cz/mbu/folia/**

Evolutionary Relatedness between Glycolytic Enzymes Most Frequently Occurring in Genomes

A. OSLANCOVÁ, Š. JANEČEK*

Institute of Molecular Biology, Center of Excellence for Molecular Medicine, Slovak Academy of Sciences, 845 51 Bratislava, Slovakia
e-mail Stefan.Janecek@savba.sk

Received 10 December 2003

ABSTRACT. More than 100 sequenced genomes were searched for genes coding for the enzymes involved in glycolysis in an effort to find the most frequently occurring ones. Triosephosphate isomerase (TIM), glyceraldehyde-3-phosphate dehydrogenase (GAPD), phosphoglycerate kinase (PGK) and enolase (ENOL) were found to be present in 90 investigated genomes all together. The final set consisted of 80 prokaryotic and 10 eukaryotic genomes. Of the 80 prokaryotic genomes, 73 were from Bacteria, 7 from Archaea. Two microbial genomes were also from Eucarya (yeasts). Eight genomes of nonmicrobial origin were included for comparison. The amino acid sequences of TIMs, GAPDs, PGKs and ENOLs were collected and aligned, and their individual as well as concatenated evolutionary trees were constructed and discussed. The trees clearly demonstrate a closer relatedness between Eucarya and Archaea (especially the concatenated tree) but they do not support the hypothesis that eukaryotic glycolytic enzymes should be closely related to their α-proteobacterial counterparts. Phylogenetic analyses further reveal that although the taxonomic groups (*e.g.*, α-proteobacteria, γ-proteobacteria, firmicutes, actinobacteria, *etc*.) form their more or less compact clusters in the trees, the inter-clade relationships between the trees are not conserved at all. On the other hand, several examples of conservative relatedness separating some clades of the same taxonomic groups were observed, *e.g.*, *Buchnera* along with *Wigglesworthia* and the rest of γ-proteobacteria, or mycoplasmas and the rest of firmicutes. The results support the view that these glycolytic enzymes may have their own evolutionary history.

In the post-genome era when an enormous amount of sequence information has become available, comparative analysis of metabolic pathways in various genomes may offer essential information on (*i*) how the pathways have evolved, (*ii*) the targets important from the pharmacological point of view; and (iii) the biotechnological applications (Dandekar *et al.* 1999). Glycolysis, as the central metabolic and probably the most highly conserved pathway, enables one to compare the enzymes isolated from a wide spectrum of organisms (Fothergill-Gilmore 1986). It is thus very suitable for studying the protein evolution and the results concerning the individual glycolytic enzymes may add to our understanding of species evolution and their mutual relatedness.

Despite the virtually ubiquitous presence of glycolysis in living organisms, even the initial analyses of the first sequenced genomes have revealed that some enzymes may be missing especially in microorganisms, *e.g.*, pyruvate kinase in *Helicobacter pylori* (Tomb *et al.* 1997) and *Archaeoglobus fulgidus* (Klenk *et al.* 1997). The lack of such enzymes could be compensated by (*i*) suitable adaptation of related enzymes; (*ii*) connection to other parts of metabolism; and (*iii*) different metabolic alternatives (Dandekar *et al.* 1999). The eventuality of the existence of unique yet uncharacterized proteins in sequenced genomes, that would play the role of the missing glycolytic enzymes, should also be taken into account (Cordwell 1999; Galperin and Koonin 1999). Further examples of incompleteness and/or loss of important metabolic pathways may be the citrate cycle (Huynen *et al.* 1999) and glycogen metabolism (Henrissat *et al.* 2002).

The "classical" glycolytic reactions are catalyzed by the following 10 enzymes: hexokinase (EC 2.7.1.1), phosphoglucose isomerase (EC 5.3.1.9), phosphofructokinase (EC 2.7.1.11), fructose-bisphosphate aldolase (EC 4.1.2.13), triosephosphate isomerase (TIM; EC 5.3.1.1), glyceraldehyde-3-phosphate dehydrogenase (GAPD; EC 1.2.1.12), phosphoglycerate kinase (PGK; EC 2.7.2.3), phosphoglycerate mutase (EC 5.4.2.1), enolase (ENOL; EC 5.4.2.1), and pyruvate kinase (EC 2.7.1.40) (Erlandsen *et al.* 2000). Structurally the glycolytic enzymes are well studied, *i.e.*, at least one three-dimensional structure is available for each of the ten proteins (Muirhead and Watson 1992). All glycolytic enzymes belong to the class of α/β-proteins. The four of them – fructose-bisphosphate aldolase, TIM, ENOL, and pyruvate kinase – rank among the $(\beta/\alpha)_8$-barrel proteins (Pujadas and Palau 1999), the remaining six enzymes share a similar nucleotide–co-

*Corresponding author.

factor/substrate-binding domain (Muirhead and Watson 1992). Hexokinase, phosphofructokinase, PGK, and GAPD (with nucleotide cofactors) posses in the core of the binding domain a similar six-stranded parallel β-sheet, while phosphoglucose isomerase and phosphoglycerate mutase (without cofactors) contain a single antiparallel β-sheet (Muirhead and Watson 1992; Erlandsen *et al.* 2000).

The main goal of the present work was to contribute to the evolutionary picture of glycolysis by sequence comparison of glycolytic enzymes originating from a large and wide-spectral set of organisms. Although the genome-sequencing projects yield a number of available sequence data, the efforts aimed at collecting as complete as possible samples of glycolytic enzymes and their producers have been complicated by the fact that some of the enzymes exist in multiple forms not always present in an organism. The evolutionary studies focused on glycolysis are thus a compromise between the number of glycolytic enzymes and the number of organisms involved. This study therefore provides the evolutionary relationships between the 90 organisms belonging to Bacteria, Archaea and Eucarya as revealed by the amino acid sequence similarities and differences between the four most frequently occurring glycolytic enzymes, *i.e.* triosephosphate isomerase, glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase, and enolase.

MATERIALS AND METHODS

All amino acid sequences were retrieved from the complete (for Bacteria and Archaea) and both complete and unfinished (for Eucarya) sequenced genomes from the GenBank (Benson *et al.* 2000) as available before January 2003 *via* the ENTREZ system (Schuler *et al.* 1996). The SwissProt sequence database (Bairoch and Apweiler 2000) was also used. The genomes were screened for the individual glycolytic enzymes using the "search" function throughout the annotations. Thus of the 10 glycolytic enzymes TIM, GAPD, PGK, and ENOL were recognized to be present in 90 of 109 investigated genomes all together. These four enzymes were therefore selected for further analysis.

The final complete set of investigated genomes consisted of 80 genomes from prokaryotes and 10 genomes from eukaryotes (Table I; for a detailed table showing the GenBank accession numbers as well as the taxonomic composition of the studied set, see the URL: **`http://imb.savba.sk/~janecek/papers/Glycolysis_04/`**). Of the 80 prokaryotic genomes, 73 were from Bacteria and 7 from Archaea. As to the Archaea, three taxa were from Crenarchaeota and four from Euryarchaeota. As far as the domain Bacteria is concerned, the taxonomic composition was the following: α-Proteobacteria – 7, β-Proteobacteria – 3, γ-Proteobacteria – 18, ε-Proteobacteria – 1, Firmicutes – 24, Chlamydiae – 5, Actinobacteria – 7, Cyanobacteria – 3, Spirochetes – 2, Deinococci – 1, Fusobacteria – 1, and Thermotogae – 1. Two microbial genomes were also from Eucarya (yeasts). Only 8 of 90 genomes of this study were of nonmicrobial origin: nematode, apicomplexa, microsporidium, plant, insect, and 3 mammals (mouse, rat and human). They were included in order to obtain as much as possible a relevant comparison of all the three domains of life: Bacteria, Archaea, and Eucarya.

The sequences were aligned using the program Clustal W (Thompson *et al.* 1994) and the computer-produced alignments were slightly manually tuned where applicable. The final alignments served for calculation by the neighbor-joining method (Saitou and Nei 1986) of the evolutionary trees. The Phylip format tree output was applied using the bootstrapping procedure (Felsenstein 1985); the number of bootstrap trials used was 1000. The trees were drawn with the program TreeView (Page 1996).

RESULTS AND DISCUSSION

The evolution of glycolysis has been extensively studied in the 1990s (Fothergill-Gilmore and Michels 1993; Cordwell 1999; Dandekar *et al.* 1999). The number of sequence data, accumulated in the last few years from the sequencing of genomes, requires to be analyzed from various points of view more frequently. It was hypothesized that eukaryotes might have arisen through symbiotic association between an α-proteobacterium and a methanogenic archaeon (Martin and Müller 1998). Although such a model could suggest that genes for glycolysis in modern eukaryotes descend from genes of the α-proteobacterial ancestor (Keeling and Doolittle 1997), the recent phylogenetic analyses do not support it (Canback *et al.* 2002). It would therefore be fruitful to shed more light on these views by analyzing as complete and comparable as possible samples of glycolytic enzymes.

Before studying the evolutionary relationships of the four glycolytic enzymes most frequently occurring in genomes it was necessary to align their amino-acid sequences and to analyze the alignments; generally, they confirmed that the sequences of glycolytic enzymes are well conserved although the large sample

Table I. Sources of the glycolytic enzymes used

Organism	Abbreviation
Bacteria	
Agrobacterium tumefaciens C58 (CEREON)	Agrtu C
Agrobacterium tumefaciens C58 (U. WASHINGTON)	Agrtu W
Bacillus anthracis A2012	Bacan
Bacillus halodurans C-125	Bacha
Bacillus subtilis	Bacsu
Bifidobacterium longum NCC2705	Biblo
Borrelia burgdorferi	Borbu
Brucella melitensis	Brume
Brucella suis 1330	Brusu
Buchnera aphidicola APS	Bucap APS
Buchnera aphidicola Sg	Bucap Sg
Campylobacter jejuni	Camje
Caulobacter crescentus	Caucr
Chlamydia muridarum	Chlmu
Chlamydia pneumoniae AR39	Chlpn AR
Chlamydia trachomatis	Chltr
Chlamydophila pneumoniae CWL029	Chppn CW
Chlamydophila pneumoniae J138	Chppn J
Clostridium acetobutylicum ATCC824	Cloac
Clostridium perfringens	Clope
Corynebacterium efficiens YS-314	Coref
Corynebacterium glutamicum ATCC 13032	Corgl
Deinococcus radiodurans R1	Deira
Escherichia coli K12	Escco K12
Escherichia coli O157:H7	Escco H7
Escherichia coli O157:H7 EDL933	Escco EDL
Fusobacterium nucleatum subsp. *nucleatum* ATCC 25586	Fusnu
Haemophilus influenzae Rd	Haein
Lactococcus lactis subsp. *lactis*	Lacla
Leptospira interrogans serovar lai 56601	Lepin
Listeria innocua Clip11262	Lisin
Listeria monocytogenes EGD	Lismo
Mesorhizobium loti	Meslo
Mycobacterium leprae TN	Mycle
Mycobacterium tuberculosis CDC1551	Myctu CDC
Mycobacterium tuberculosis H37Rv	Myctu Rv
Mycoplasma genitalium	Mypge
Mycoplasma pneumoniae	Myppn
Mycoplasma pulmonis	Myppu
Neisseria meningitidis serogroup A Z2491	Neime A
Neisseria meningitidis serogroup B MC58	Neime B
Nostoc sp. PCC 7120	Nossp
Oceanobacillus iheyensis	Oceih
Pseudomonas aeruginosa	Pseae
Ralstonia solanacearum	Ralso
Salmonella enterica subsp. *enterica* serovar Typhi	Salen
Salmonella typhimurium LT2	Salty
Shewanella oneidensis MR-1	Sheon
Shigella flexneri 2a 301	Shifl
Sinorhizobium meliloti 1021	Sinme
Staphylococcus aureus Mu50	Staau Mu
Staphylococcus aureus subsp. *aureus MW2*	Staau MW
Staphylococcus aureus subsp. *aureus N315*	Staau N
Streptococcus agalactiae	Stcag
Streptococcus agalactiae NEM316	Stcag N
Streptococcus mutans UA159	Stcmu
Streptococcus pneumoniae R6	Stcpn R
Streptococcus pneumoniae TIGR4	Stcpn T
Streptococcus pyogenes M1 GAS	Stcpy M
Streptococcus pyogenes MGAS315	Stcpy G
Streptococcus pyogenes MGAS8232	Stcpy S8
Streptomyces coelicolor A3(2)	Strco
Synechocystis sp. PCC 6803	Synsp
Thermoanaerobacter tengcongensis MB4T	Thbte
Thermosynechococcus elongatus BP-1	Thsel
Thermotoga maritima	Thtma
Vibrio cholerae	Vibch
Wigglesworthia brevipalpis	Wigbr
Xanthomonas axonopodis pv. *citri* 306	Xanax
Xanthomonas campestris pv. *campestris* TCC 33913	Xanca
Xylella fastidiosa 9a5c	Xylfa
Yersinia pestis CO92	Yerpe
Yersinia pestis KIM	Yerpe K
Archaea	
Aeropyrum pernix	Aerpe
Methanosarcina mazei Goe1	Metma
Pyrobaculum aerophilum IM2	Pybae
Pyrococcus abyssi	Pycab
Pyrococcus furiosus DSM 3638	Pycfu
Pyrococcus horikoshii	Pycho
Sulfolobus solfataricus	Sulso
Eucarya	
Arabidopsis thaliana	Arath
Caenorhabditis elegans	Caeel
Drosophila melanogaster	Drome
Encephalitozoon cuniculi GB-M1	Enccu
Homo sapiens	Homsa
Mus musculus	Musmu
Plasmodium falciparum	Plafa
Rattus norvegicus	Ratno
Saccharomyces cerevisiae	Sacce
Schizosaccharomyces pombe	Schpo

and wide diversity of sources led to the number of invariant residues being low (Table II). Most of these residues are the residues belonging to the enzyme active sites (*the alignments are not shown*).

Sequence analysis. In TIMs all the three catalytic residues, Lys-12, His-95 and Glu-165 (*Saccharomyces cerevisiae* TIM numbering) (Lolis *et al.* 1990), are conserved invariantly. The Ser-96, which forms a hydrogen bond with Glu-165 in the yeast TIM, was found to be replaced by an alanine in two bacterial TIMs from *Listeria* sp. and by a phenylalanine in the enzyme from the malaria-causing parasite *Plasmodium falciparum*. It has been indicated, however, that the hydrogen bonds between the catalytic residues and residue 96 might

Table II. Sequence characteristics of the studied glycolytic enzymes

Parameter	TIM	GAPD	PGK	ENOL
Minimum length	222	302	374	393
Maximum length	274	359	500	461
Invariant residues	11	7	24	40
Consensus length	305	428	606	526
Identity, %	3.61	1.64	3.96	7.61

not be essential for catalysis (Velanker *et al.* 1997). Nevertheless, the *P. falciparum* TIM is the only one from the eukaryotes (including plant, insect and mammals) without the ability to form the above-mentioned hydrogen bond. The studied archaeal TIMs have also serine in the position equivalent to yeast Ser-96. All the TIMs from Archaea are, however, by approximately 20 amino acid residues shorter than those from Bacteria and Eucarya (Kohlhoff *et al.* 1996). The four short deletions are located outside the active-site regions so that the sequences of archaeal TIMs go well with those of both bacterial and eukaryotic origin. The TIM from *Methanosarcina mazei* with 222 residues is the shortest one (Table II). The average length of a bacterial or a eukaryotic TIM is about 250 residues. The only exception is the TIM from *Campylobacter jejuni* with 223 residues.

GAPD sequences were found to be of two types: (*i*) bacterial and eukaryotic, and (*ii*) archaeal, resulting in a very low degree of sequence identity (Table II). While the GAPDs from Bacteria and Eucarya are similar throughout their entire length, the archaeal GAPDs differ from them substantially. Two catalytic residues were reported for GAPD from *Bacillus stearothermophilus* (Skarzynski *et al.* 1987): Cys-139 and His-176. Although the cysteine is conserved invariantly in all studied GAPDs, in all archaeal enzymes there is no histidine residue at the position equivalent to His-176 of *B. stearothermophilus*. However, all GAPDs from Archaea contain another conserved histidine (His-219 in the enzyme from *Sulfolobus solfataricus*; Isupov *et al.* 1999) with its imidazole group at about the same position. The ancestral GAPD thus had probably a low turnover and broad specificity using only a cysteine for catalysis (Isupov *et al.* 1999) which, together with the invariant preceding serine residue, is found in all present-day GAPDs. Despite the low sequence identity between bacterial and/or eukaryotic and archaeal GAPDs, their lengths are comparable (330–340 residues) and the sequence differences are not reflected in the higher types of structure (Isupov *et al.* 1999; Antonyuk *et al.* 2003).

The sequences of PGKs are substantially similar. Although the active site residues vary slightly depending on the source of PGK, the majority of them are conserved in all PGKs, such as Lys-218, Glu-342, Asn-335, and Asp-373 (*Saccharomyces cerevisiae* PGK numbering) (Watson *et al.* 1982; Bernstein *et al.* 1998). The lysine has, however, not always an equivalent in the archaeal PGKs and may be replaced by an argininine or a methionine (Fleming and Littlechild 1997). From the studied set of PGKs the longest enzyme was the one from *Mycoplasma pulmonis* (771 residues) but its C-terminal extension from the residue 416 was deleted. Most of the lengths of PGK sequences lay in the range of 390–420 residues. The shortest PGK from *Synechocystis* sp. PCC 6803 (374 residues, Table II) apparently lacks the N-terminal segment present in all other PGKs. On the other hand, the longest PGK from *Xylella fastidiosa* with its 500-residue long sequence exhibits an additional N-terminal region, the feature present also in the remarkably long PGK from the plant *Arabidopsis thaliana*.

The sequences of ENOLs are the best conserved of all the four glycolytic enzymes studied here (Table II). In addition to residues involved in the active site (Ser-39, Glu-168, Asp-246, Glu-295, Asp-320, Asp-321, Lys-345, Arg-374, and Lys-396; *Saccharomyces cerevisiae* ENOL numbering) (Lebioda *et al.* 1989; Stec and Lebioda 1990; Wedekind *et al.* 1994) many other residues are conserved invariantly. There are no apparent sequence differences between the ENOLs belonging either to Bacteria or Archaea or Eucarya. The average length of an ENOL varies from 420 to 440 amino acid residues. The shortest and longest ENOLs are those from *Agrobacterium tumefaciens* U. WASHINGTON (393 residues) and *Xylella fastidiosa* (461 residues), respectively. The former lacks the usual ENOL N-terminal segment, while the latter possesses an extra N-terminal extension.

Evolutionary relationships. Based on the above-mentioned alignments of amino-acid sequences of 90 TIMs, GAPDs, PGKs and ENOLs, their evolutionary trees were constructed (Fig. 1). Since the set of sources for each of the four glycolytic enzymes was the same, it was possible to compare the individual trees. All taxonomic groups form, in fact, their own clades separated from each other. However, exceptions which do not reflect taxonomy exactly can be found in all trees. The α-proteobacteria and chlamydiae seem

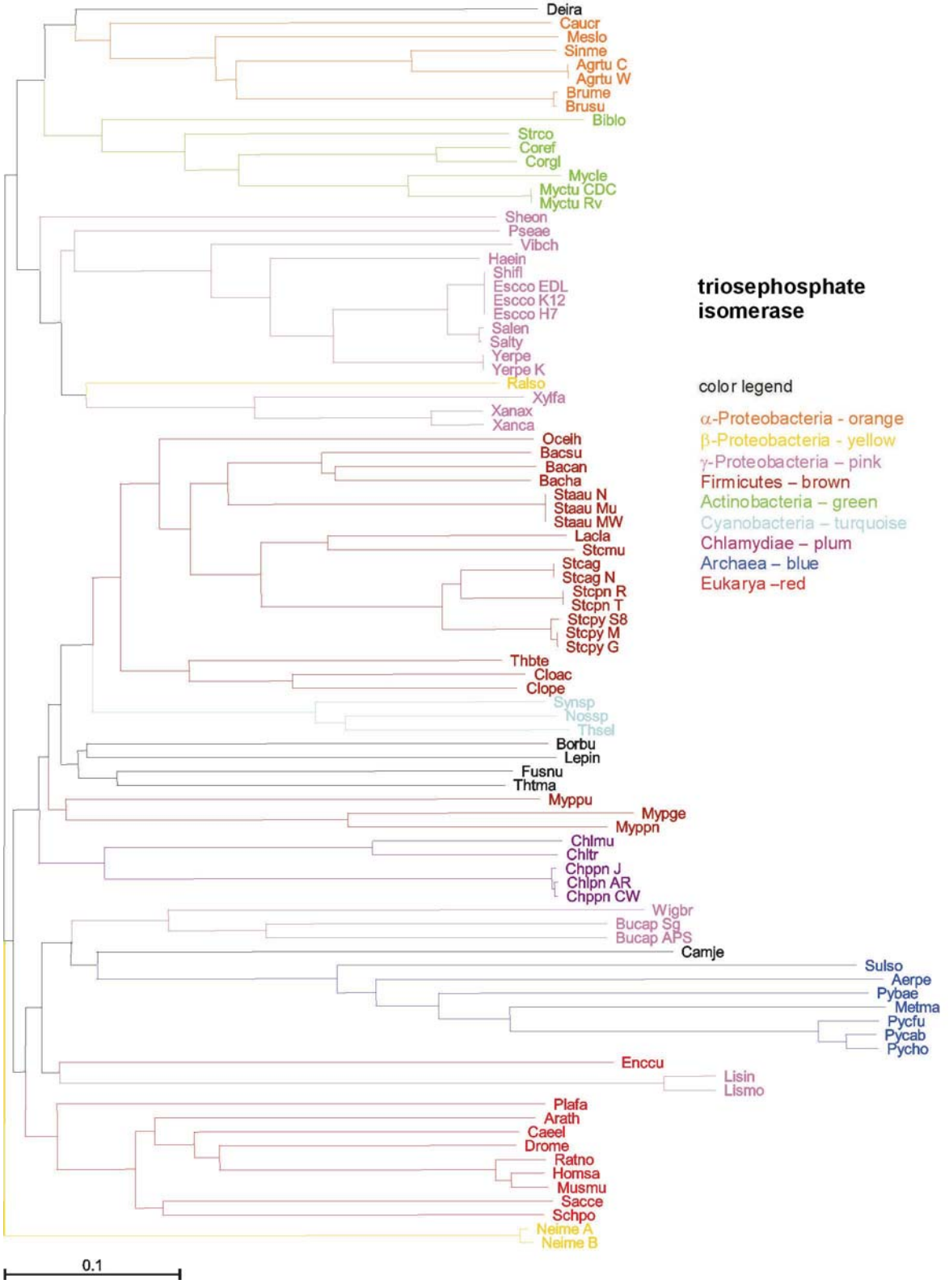
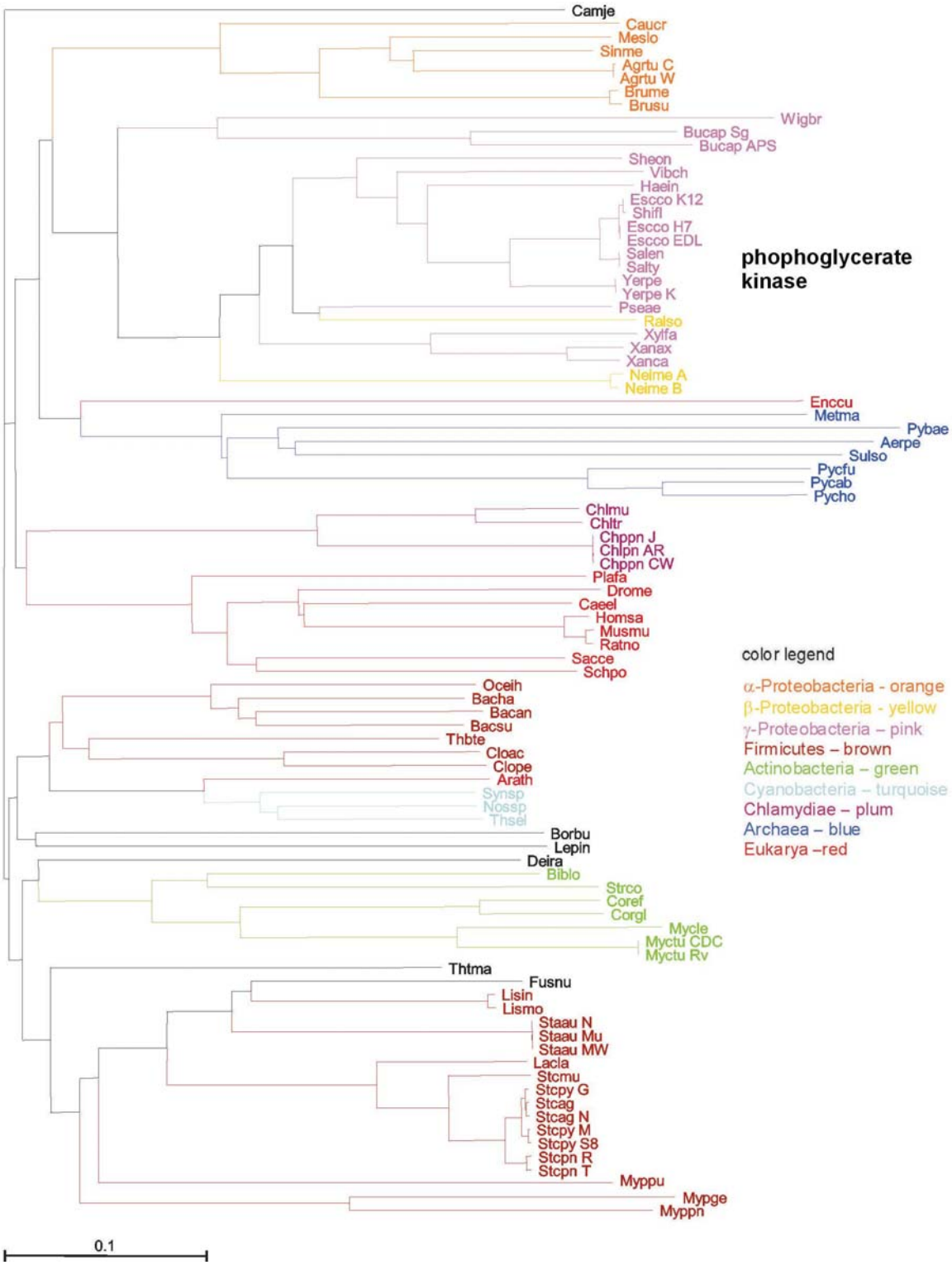


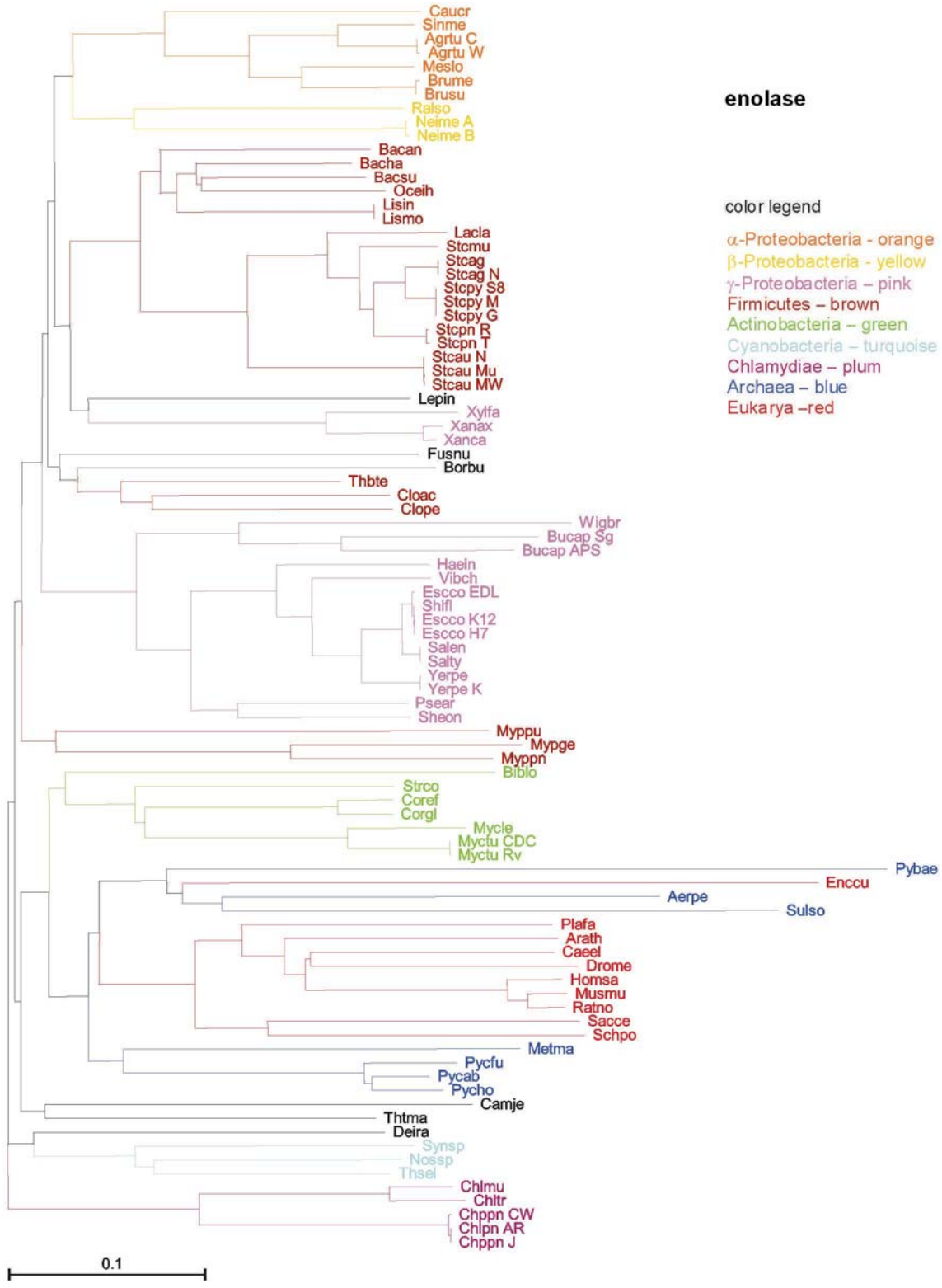
Fig. 1 (*four pages*)**.** The evolutionary trees of the four glycolytic enzymes most frequently occurring in genomes; the abbreviations of the sources are explained in Table I; for further details *see* also the Table I at the URL: **`http://imb.savba.sk/~janecek/papers/Glycolysis_04/`**.

color legend

α-Proteobacteria - orange
β-Proteobacteria - yellow
γ-Proteobacteria – pink
Firmicutes – brown
Actinobacteria – green
Cyanobacteria – turquoise
Chlamydiae – plum
Archaea – blue
Eukarya –red



to be the best conserved bacterial groups since these two form taxonomically pure clusters in all four trees. Also actinobacteria as well as cyanobacteria go well together as compact clades, except for the GAPD tree. With regard to Archaea and Eucarya, these two exhibit a mutually closer evolutionary relatedness but again this is not true in the GAPD tree, where the separation of Archaea is due to pronounced sequence differences concerning the catalytic histidine residues discussed above (Isupov *et al.* 1999; Antonyuk *et al.* 2003). Although there is a tendency to group the Crenarchaeota and Euryarchaeota separately in both TIM and ENOL trees, and perhaps also in the PGK tree, the division is not unambiguous. Eucarya also behave auto-



nomously, the microsporidian seems to be the eukaryote most distantly related to the rest. It is necessary to point out, however, and this concerns all studied phyla, that although the individual taxonomic groups are more or less conserved, their mutual interrelationships are not conserved at all. Such an evolutionary behavior, *i.e.* their own evolutionary history, was recently inferred for the $(\beta/\alpha)_8$-barrel glycolytic enzymes upon analysis of 43 genomes (Kováčová and Janeček 2002).

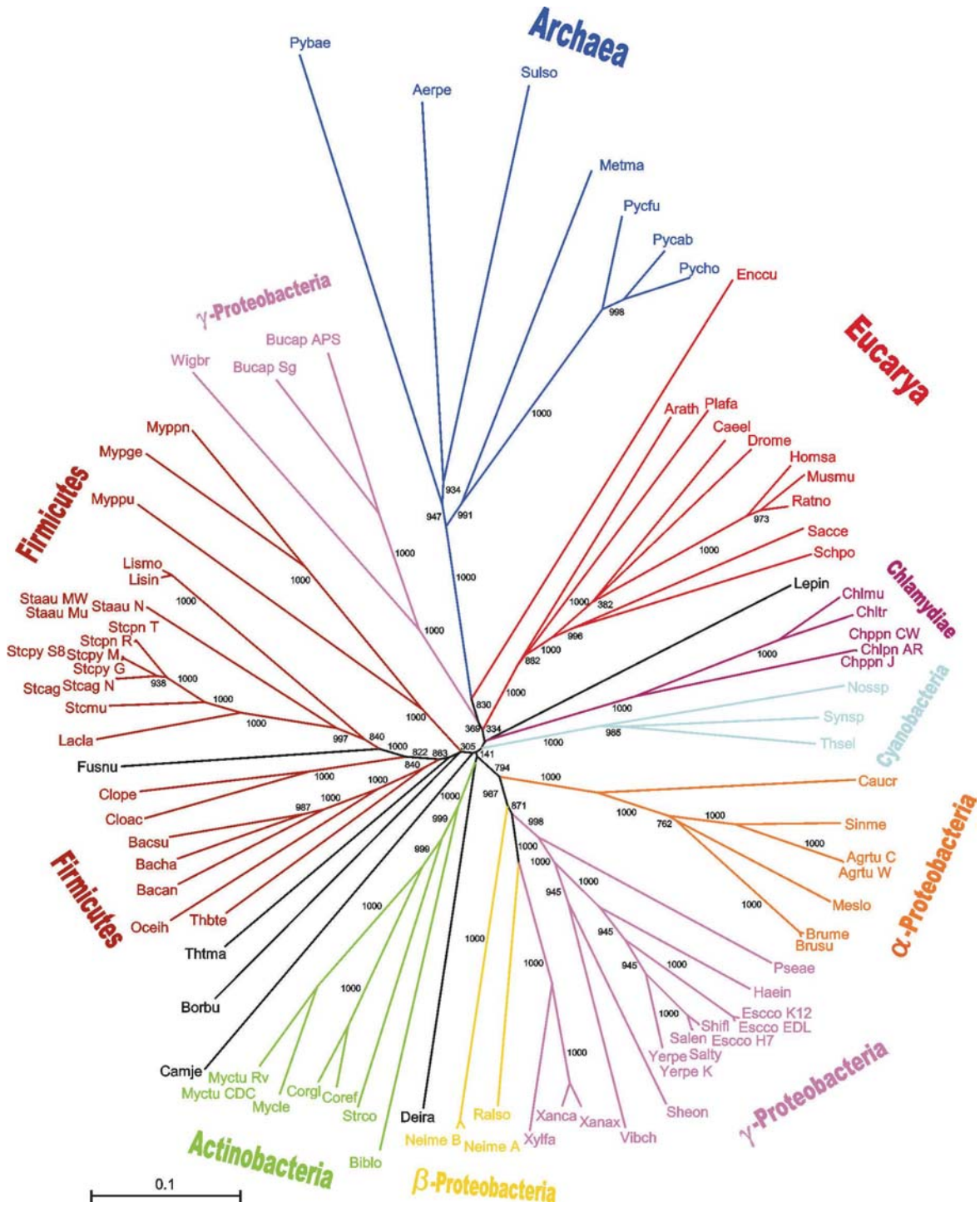


Fig. 2. The concatenated evolutionary tree for TIM, GAPD, PGK and ENOL; for abbreviations of the sources *see* Table I.

There are several interesting observations specific for each of the four trees. Thus, for instance, in the TIM tree actinobacteria go well together with the α-proteobacteria, while with firmicutes in both GAPD and PGK trees, and even with Archaea and Eucarya in the ENOL tree. Other findings of interest are the grouping of all chlamydiae with eukaryotes in the PGK tree or the clustering of five firmicutes (Wigbr, Bucap-Sg, Bucap-APS, Lisin and Lismo) among the Archaea and Eucarya in the TIM tree. The only plant representative (Arath; Table I) is grouped in the PGK tree with cyanobacteria and remarkably with some firmicutes, a feature that is not seen in any of the remaining three trees. When comparing the glycolytic enzymes these and other anomalous phylogenetic arrangements are not, however, rare (Bernstein *et al.* 1998; Hannaert *et al.* 2000; Canback *et al.* 2002; Kováčová and Janeček 2002; Ronimus and Morgan 2003).

Some specific features associated with taxonomy are found as conserved within the groups. For example, Wigbr, Bucap-Sg, Bucap-APS and the rest of γ-proteobacteria, or Mypge, Myppn, Myppu and the rest of firmicutes. Interestingly the enzymes from the two spirochetes (Borbu and Lepin) go well together only in the TIM and PGK trees (Fig. 1). For GAPD, spirochetes have been known to exhibit an extreme diversity (Figge and Cerff 2001). Although the GAPD, PGK and TIM are encoded by genes clustered into an operon in *Borrelia burgdorferi*, in the corresponding evolutionary trees spirochetes showed relationships to different phyla (Gebbia *et al.* 1997) supporting the idea of an independent evolutionary history for each of the glycolytic enzymes (Kováčová and Janeček 2002).

In order to contribute to the whole evolutionary picture of the glycolytic pathway, a concatenated evolutionary tree was constructed (Fig. 2). This tree was based on the alignment composed from the alignments of TIMs, GAPDs, PGKs and ENOLs so that the sequence features of all four enzymes are expressed in a single tree at once. Although there is some uncertainty in the center of the tree (smaller bootstrap values), it clearly demonstrates that taxonomy is well respected and that partial scattering of the phyla observed in the GAPD tree was not retained. All taxonomic groups and domains keep, in fact, their integrity. The three γ-proteobacteria (Wigbr and the two from *Buchnera*) present the most striking exception with the position on a branch between Archaea and Eucarya. The archaeal cluster consists of two parts: the Crenarchaeota (Aerpe, Pybae and Sulso) and the Euryarchaeota (Metma, Pycab, Pycfu and Pycho). Eucarya are all in one group except for the microsporidian (Enccu) which is on its own long separated branch closer to Archaea. The remaining part of the tree is occupied by bacteria separated into fimicutes, actinobacteria and α-, β- and γ-proteobacteria (with the exception of the three above-mentioned Wigbr, Bucap-Sg, Bucap-APS). Cyanobacteria and chlamydiae form their own groups positioned between the rest of Bacteria on the one side, and both Eucarya and Archaea, on the other side.

One of the most striking observations is that neither the trees for the individual glycolytic enzymes (Fig. 1), nor the concatenated tree (Fig. 2) support the view about the close relationships between glycolytic enzymes from α-proteobacteria and eukaryotes. This model was suggested by: (*i*) the hypothesis on how the evolution of a eukaryotic cell might initiate (Martin and Müller 1998); and (*ii*) the idea that essential functions of modern eukaryotic cells are encoded by the genes descended from the bacterial ancestors of modern mitochondria (Gogarten *et al.* 1996; Keeling and Doolittle 1997). The views concerning the evolution of eukaryotic genes of intermediary metabolism have started to be re-evaluated recently (Canback *et al.* 2002). This study, together with our previous results (Kováčová and Janeček 2002), strongly supports the idea to do that re-evaluation. In fact, the evolution of enzymes as well as pathways is still under debate (Schmidt *et al.* 2003). Neither the strong conservation of enzymes nor their clustering into an operon in a genome warrants the conserved relationships. One possibility is to focus on Archaea that have developed novel enzymes and/or alternative pathways (van der Oost *et al.* 2002; Verhees *et al.* 2003). The continuously increasing number of sequenced genomes from all the three domains of life offers opportunities for approaching the relevant solutions.

This work was supported by the VEGA grant no. 2/2057/23 from the *Slovak Grant Agency*.

REFERENCES

ANTONYUK S.V., EADY R.R., STRANGE R.W., HASNAIN S.S.: The structure of glyceraldehyde-3-phosphate dehydrogenase from *Alcaligenes xylosoxidans* at 1.7 Å resolution. *Acta Crystallogr.* **D59**, 835–842 (2003).

BAIROCH A., APWEILER R.: The SwissProt protein sequence database and its supplement TrEMBL in 2000. *Nucl.Acids Res.* **28**, 45–48 (2000).

BENSON D.A., KARSCH-MIZRACHI I., LIPMAN D.J., OSTELL J., RAPP B.A., WHEELER D.L.: GenBank. *Nucl.Acids Res.* **28**, 15–18 (2000).

BERNSTEIN B.E., WILLIAMS D.M., BRESSI J.C., KUHN P., GELB M.H., BLACKBURN G.M., HOL W.G.: A bisubstrate analog induces unexpected conformational changes in phosphoglycerate kinase from *Trypanosoma brucei*. *J.Mol.Biol.* **279**, 1137–1148 (1998).

CANBACK B., ANDERSSON S.G., KURLAND C.G.: The global phylogeny of glycolytic enzymes. *Proc.Nat.Acad.Sci.USA* **99**, 6097–6102 (2002).

CORDWELL S.J.: Microbial genomes and "missing enzymes": redefining biochemical pathways. *Arch.Microbiol.* **172**, 269–279 (1999).

DANDEKAR T., SCHUSTER S., SNEL B., HUYNEN M., BORK P.: Pathway alignment: application to the comparative analysis of glycolytic enzymes. *Biochem.J.* **343**, 115–124 (1999).

ERLANDSEN H., ABOLA E.E., STEVENS R.C.: Combining structural genomics and enzymology: completing the picture in metabolic pathways and enzyme active sites. *Curr.Opin.Struct.Biol.* **10**, 719–730 (2000).

FELSENSTEIN J.: Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791 (1985).

FIGGE R.M., CERFF R.: GAPDH gene diversity in spirochetes: a paradigm for genetic promiscuity. *Mol.Biol.Evol.* **18**, 2240–2249 (2001).

FLEMING T., LITTLECHILD J.: Sequence and structural comparison of thermophilic phosphoglycerate kinases with a mesophilic equivalent. *Comp.Biochem.Physiol.* **A118**, 439–451 (1997).

FOTHERGILL-GILMORE L.A.: The evolution of glycolytic pathway. *Trends Biochem.Sci.* **11**, 47–51 (1986).

FOTHERGILL-GILMORE L.A., MICHELS P.A.M.: Evolution of glycolysis. *Progr.Biophys.Mol.Biol.* **59**, 105–235 (1993).

GALPERIN M.Y., KOONIN E.V.: Functional genomics and enzyme evolution. Homologous and analogous enzymes encoded in microbial genomes. *Genetica* **106**, 159–170 (1999).

GEBBIA J.A., BACKENSON P.B., COLEMAN J.L., ANDA P., BENACH J.L.: Glycolytic enzyme operon of *Borrelia burgdorferi*: characterization and evolutionary implications. *Gene* **188**, 221–228 (1997).

GOGARTEN J.P., OLENDZENSKI L., HILARIO E., SIMON C., HOLSINGER K.E.: Dating the cenancester of organisms. *Science* **274**, 1750–1751 (1996).

HANNAERT V., BRINKMANN H., NOWITZKI U., LEE J.A., ALBERT M.A., SENSEN C.W., GAASTERLAND T., MULLER M., MICHELS P., MARTIN W.: Enolase from *Trypanosoma brucei*, from the amitochondriate protist *Mastigamoeba balamuthi*, and from the chloroplast and cytosol of *Euglena gracilis*: pieces in the evolutionary puzzle of the eukaryotic glycolytic pathway. *Mol. Biol.Evol.* **17**, 989–1000 (2000).

HENRISSAT B., DELEURY E., COUTINHO P.M.: Glycogen metabolism loss: a common marker of parasitic behavior in bacteria? *Trends Genet.* **18**, 437–440 (2002).

HUYNEN M.A., DANDEKAR T., BORK P.: Variation and evolution of the citric-acid cycle: a genomic perspective. *Trends Microbiol.* **7**, 281–291 (1999).

ISUPOV M.N., FLEMING T.M., DALBY A.R., CROWHURST G.S., BOURNE P.C., LITTLECHILD J.A.: Crystal structure of the glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J.Mol.Biol.* **291**, 651–660 (1999).

KEELING P.J., DOOLITTLE W.F.: Evidence that eukaryotic triosephosphate isomerase is of α-proteobacterial origin. *Proc.Nat.Acad.Sci. USA* **94**, 1270–1275 (1997).

KLENK H.P., CLAYTON R.A., TOMB J.F., WHITE O., NELSON K.E., KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D., RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C., FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S., KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B., PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L., OVERBEEK R., GOCAYNE J.D., WEIDMAN J.F., MCDONALD L., UTTERBACK T., COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M., SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A., MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390**, 364–370 (1997).

KOHLHOFF M., DAHM A., HENSEL R.: Tetrameric triosephosphate isomerase from hyperthermophilic Archaea. *FEBS Lett.* **383**, 245–250 (1996).

KOVÁČOVÁ A., JANEČEK Š.: Evolutionary relationships of glycolytic $(\beta/\alpha)_8$-barrel enzymes present in completely sequenced genomes. *Biologia (Bratislava)* **57**, 283–288 (2002).

LEBIODA L., STEC B., BREWER J.M.: The structure of yeast enolase at 2.25-Å resolution. An 8-fold β+α-barrel with a novel ββαα $(\beta\alpha)_6$ topology. *J.Biol.Chem.* **264**, 3685–3693 (1989).

LOLIS E., ALBER T., DAVENPORT R.C., ROSE D., HARTMAN F.C., PETSKO G.A.: Structure of yeast triosephosphate isomerase at 1.9-Å resolution. *Biochemistry* **29**, 6609–6618 (1990).

MARTIN W., MÜLLER M.: The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).

MUIRHEAD H., WATSON H.: Glycolytic enzymes: from hexose to pyruvate. *Curr.Opin.Struct.Biol.* **2**, 870–876 (1992).

VAN DER OOST J., HUYNEN M.A., VERHEES C.H.: Molecular characterization of phosphoglycerate mutase in archaea. *FEMS Microbiol.Lett.* **212**, 111–120 (2002).

PAGE R.D.: TreeView: an application to display phylogenetic trees on personal computers. *Comput.Applic.Biosci.* **12**, 357–358 (1996).

PUJADAS G., PALAU J.: TIM barrel fold: structural, functional and evolutionary characteristics in natural and designed molecules. *Biologia (Bratislava)* **54**, 231–254 (1999).

RONIMUS R.S., MORGAN H.W.: Distribution and phylogenies of enzymes of the Embden–Meyerhof–Parnas pathway from archaea and hyperthermophilic bacteria support a gluconeogenic origin of metabolism. *Archaea* **1**, 199–221 (2003).

SAITOU N., NEI M.: The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol.Biol.Evol.* **4**, 406–425 (1987).

SCHMIDT S., SUNYAEV S., BORK P., DANDEKAR T.: Metabolites: a helping hand for pathway evolution? *Trends Biochem.Sci.* **28**, 336–341 (2003).

SCHULER G.D., EPSTEIN J.A., OHKAWA H., KANS J.A.: Entrez: molecular biology database and retrieval system. *Meth.Enzymol.* **266**, 141–162 (1996).

SKARZYNSKI T., MOODY P.C., WONACOTT A.J.: Structure of holo-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus* at 1.8 Å resolution. *J.Mol.Biol.* **193**, 171–187 (1987).

STEC B., LEBIODA L.: Refined structure of yeast apo-enolase at 2.25 Å resolution. *J.Mol.Biol.* **211**, 235–248 (1990).

THOMPSON J.D., HIGGINS D.G., GIBSON T.J.: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment trough sequence weighting, position specific gap penalties and weight matrix choice. *Nucl.Acids Res.* **22**, 4673–4680 (1994).

TOMB J.F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G., FLEISCHMANN R.D., KETCHUM K.A., KLENK H.P., GILL S., DOUGHERTY B.A., NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S., LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A., MCKENNEY K., FITZEGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K., BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M., COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E., HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M., VENTER J.C.: The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**, 539–547 (1997).

VELANKER S.S., RAY S.S., GOKHALE R.S., SUMA S., BALARAM H., BALARAM P., MURTHY M.R.N.: Triosephosphate isomerase from *Plasmodium falciparum*: the crystal structure provides insights into antimalarial drug design. *Structure* **5**, 751–761 (1997).

VERHEES C.H., KENGEN S.W.M., TUININGA J.E., SCHUT G.J., ADAMS M.W.W., DE VOS W.M., VAN DER OOST J.: The unique features of glycolytic pathways in Archaea. *Biochem.J.* **375**, 231–246 (2003).

WATSON H.C., WALKER N.P., SHAW P.J., BRYANT T.N., WENDELL P.L., FOTHERGILL L.A., PERKINS R.E., CONROY S.C., DOBSON M.J., TUITE M.F., KINGSMAN A.J., KINGSMAN S.M.: Sequence and structure of yeast phosphoglycerate kinase. *EMBO J.* **1**, 1635–1640 (1982).

WEDEKIND J.E., POYNER R.R., REED G.H., RAYMENT I.: Chelation of serine-39 to Mg^{2+} latches a gate at the active site of enolase: structure of the bis(Mg^{2+}) complex of yeast enolase and the intermediate, phosphonoacetohydroxamate, at 2.1 Å resolution. *Biochemistry* **33**, 9333–9342 (1994).