Protist, Vol. 152, 193–201, September 2001 © Urban & Fischer Verlag
http://www.urbanfischer.de/journals/protist

Protist

**ORIGINAL PAPER**

# Diplonemid Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH) and Prokaryote-to-Eukaryote Lateral Gene Transfer

**Qing Qian[1] and Patrick J. Keeling[2]**

Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

Submitted February 22, 2001; Accepted May 18, 2001
Monitoring Editor: Michael Melkonian

**Lateral gene transfer refers to the movement of genetic information from one genome to another, and the integration of that foreign DNA into its new genetic environment. There are currently only a few well-supported cases of prokaryote-to-eukaryote transfer known that do not involve mitochondria or plastids, but it is not clear whether this reflects a lack of such transfer events, or poor sampling of diverse eukaryotes. One gene where this process is apparently active is glyceraldehyde-3-phosphate dehydrogenase (GAPDH), where lateral transfer has been implicated in the origin of euglenoid and kinetoplastid genes. We have characterised GAPDH genes from diplonemids, heterotrophic flagellates that are closely related to kinetoplastids and euglenoids. Two distinct classes of diplonemid GAPDH genes were found in diplonemids, however, neither class is closely related to any other euglenozoan GAPDH. One diplonemid GAPDH is related to the cytosolic *gapC* of eukaryotes, although not to either euglenoids or kinetoplastids, and the second is related to cyanobacterial and proteobacterial *gap3*. The bacterial *gap3* gene in diplonemids provides one of the most well-supported examples of lateral gene transfer from a bacterium to a eukaryote characterised to date, and may indicate that diplonemids have acquired a novel biochemical capacity through lateral transfer.**

## Introduction

Lateral, or horizontal, gene transfer is refers to the process where genetic information is transmitted between two genomes. The importance of lateral gene transfer is that foreign genetic information is introduced into the genome, potentially replacing components of existing biological systems or introducing novel functionality to the cell. With the exception of the transfer of organelle-derived genes, this process was once regarded as an extreme rarity in evolution, but the recent explosion of molecular data from diverse prokaryotic and eukaryotic genomes has shown that some proportion of genes in any given genome may not match our phylogenetic expectations. This has been interpreted as showing that the lateral transmission of genetic information is a frequent occurrence in evolution (Doolittle 1999; Ochman et al. 2000). While it is almost certain that

[1]Current address:
Department of Microbiology & Immunology, University of British Columbia, Vancouver, BC V6T 1Z3, Canada

[2]Corresponding author;
fax 1 604 822 6089
e-mail pkeeling@interchange.ubc.ca

1434-4610/01/152/03-193 $ 15.00/0

other processes account for some of these discrepancies (in particular, ancient gene duplications can mimic lateral transfer, and rapidly evolving genes can introduce substantial "noise" that may appear to represent lateral transfer until they are more carefully inspected: Eisen 2000; Roger 1999), it is likely that lateral transfer does occur, and is an important feature of genome evolution.

Currently, the majority of evidence for gene transfer comes from comparisons between prokaryotic genomes, and it has been argued that this process is much more limited in eukaryotes (Ochman et al. 2000). Conversely, it has also been argued that phagotrophy in eukaryotes sets up a gene transfer ratchet that should lead to an abundance of bacterial genes in eukaryotic genomes (Doolittle 1998). Current data from eukaryotic genomes seem to suggest that gene transfer is not as abundant in eukaryotes as in prokaryotes, but this may result from a severely skewed representation of eukaryotes in the current databases. The vast majority of eukaryotic molecular data presently comes from the "higher" eukaryotes (animals, fungi, and land plants) and these groups are generally endowed with various characteristics that could combine to impede lateral gene transfer. For instance, phagotrophy is typically either rare or absent in these organisms, reducing the exposure of a cell to foreign DNA. In addition, most of the multicellular "higher" eukaryotes have, in various fashions and to various degrees, separated their germ line from soma, if not in potential then often at least in practice (for instance, many cells in a plant can establish a germ line but never do). This reproductive strategy severely reduces the likelihood that a transferred gene would be fixed in a population, even if it were selectively advantageous, because it limits successful transfers to those involving cells which will ultimately and directly give rise to new reproducing organisms. These traits, or their combined effect in difference cells within an organism, do not make lateral transfer impossible, but they are expected to substantially limit its frequency. In contrast to "higher" eukaryotes, these characteristics are not very common among protist. In particular, since most protists are unicellular, any successfully integrated gene transfer is automatically passed on to the recipient's offspring, which favours the fixation of transferred genes, especially if they are advantageous. Protists also make up the majority of eukaryotic diversity, but molecular data from both protists as a whole and from most individual groups is embarrassingly sparse, so the prevalence of lateral transfer involving eukaryotes is potentially underestimated at present.

Well-documented cases of lateral transfer to eukaryotes are known, but cases which do not involve endosymbiotic organelles such as the mitochondrion and plastid are still comparatively rare (for some examples, see: Boucher and Doolittle 2000; de Koning et al. 2000; Markos et al. 1993; Screen and St. Leger 2000). One instance where lateral transfer appears to be quite prevalent, however, is the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) of Euglenozoa. From previous work, it is clear that euglenozoan GAPDH evolution is very complex. The kinetoplastid Euglenozoa contain two enzyme types, a glycosomal version that is potentially derived from a bacterium, and a cytosolic version that is apparently restricted to *Trypanosoma* and *Leishmania* and seems to resemble a subclass of proteobacterial genes, which are themselves relatively closely related to eukaryotic homologues (Hannaert et al. 1998; Henze et al. 1995; Michels et al. 1991; Wiemer et al. 1995). Similarly, euglenoids contain a cytosolic enzyme related to the kinetoplastid glycosomal GAPDH, and a plastid enzyme related to the plastid GAPDH of red algae, green algae, and land plants (Henze et al. 1995). We have examined the GAPDH from a third lineage of Euglenozoa, the diplonemids. These heterotrophic flagellates have not been well studied at the molecular level, but are known to be closely related to kinetoplastids and euglenoids based on several ultrastructural and molecular characteristics. Some diplonemid flagellar shafts include a paraxonemal rod similar in morphology to other Euglenozoa, and their flagellar rootlet pattern and the ultrastructure of their tubular extrusomes are also extremely similar to examples found in kinetoplastids or euglenoids (Schuster et al. 1964; Simpson 1997). At the molecular level, diplonemids are known to use the rare nucleotide "J" and to use spliced-leaders at the 5' of mRNAs exactly as do euglenoids and kinetoplastids (Sturm et al. 2001; van Leeuwen et al. 1998). In addition, sequences of *Diplonema* SSU rRNA and mitochondrial *cox1* have been characterised, and phylogenetic analysis confirms that diplonemids are related to Euglenozoa (Maslov et al. 1999), although neither the molecular trees nor the morphological data clearly reveal the branching order of diplonemids, kinetoplastids, and euglenoids (Maslov et al. 1999; Simpson 1997). We find that diplonemid GAPDH genes, like those of kinetoplastids and euglenoids, fall into two classes, but neither are related to any other euglenozoan GAPDH. Instead, one diplonemid GAPDH is clearly eukaryotic-like but not strongly related to any known eukaryotic GAPDH, while the other is apparently derived from lateral gene transfer involving a bacterium, and potentially

utilises some substrate other than glyceraldehyde-3-phosphate. These results provide potential insights into euglenozoan GAPDH evolution, as well as one of the best-supported instances of bacteria-to-eukaryote gene transfer yet described.

## Results and Discussion

### Cloning and Sequencing of Diplonemid GAPDH Genes

Approximately 90% of the GAPDH coding region (between 301 and 316 codons) was amplified from three diplonemids, *Diplonema* ATCC 50224, *Diplonema* ATCC 50225, and *Rhynchopus* ATCC 50231. In all three cases products corresponding to the size expected for intron-free GAPDH genes were observed, and in *Diplonema* ATCC 50224 a larger product was also observed. All of these products were cloned and sequenced, revealing that all encode proteins homologous to GAPDH: the larger product from *Diplonema* ATCC 50224 was found to contain 2 short GT-AG spliceosomal introns of 77 and 120 base pairs, while the smaller products from all three taxa encoded intron-lacking genes (the two types are referred to hereafter as *gap1* and *gap3* respectively, for reasons that will be made clear in the phylogenetic analysis). The two introns in *Diplonema* ATCC 50224 *gap1* both include a CA-rich track at their extreme 3' end that is diagnostic of diplonemid spliceosomal introns characterised in alpha-tubulins, beta-tubulins, and actins (Qian 2000).

The *Diplonema* ATCC 50224 *gap1* gene was found to be extremely similar to other eukaryotic GAPDH genes based on sequence identity, and was found to align easily with *gapC* homologues (not shown). Conversely, the three *gap3* genes were all found to be comparably divergent in nature, but similar to one another based on sequence identity. Moreover, all three share a number of unique insertions that are similar in length and sequence (Fig. 1),

| | A B | C D |
|---|---|---|
| ***Diplonema* 50224 *gap3*** | **QSLVDLVQTKKNDMRRSRSGMLNLA** | **VNKLLEAASSSGPLSANSEHGSILGYETRPLV** |
| ***Diplonema* 50225 *gap3*** | **QSLVDMVNTKKNDLRRSRSGMLNLC** | **VNALLEKAAQEGPLAGTAEHGSILGYEARPLV** |
| ***Rhynchopus* 50231 *gap3*** | **QSLVDMVNTKKNDLRRSRSGMLNLA** | **VNALLQDAAAEGPLAATSEHGAILGFETRPLV** |
| *Synechocystis gap3* | QPTVDGPS-KK-DFRGGRGAAQNII | ICAAMKTAAE------GELKG-ILGYTADDVV |
| *Anabaena gap3* | QTLVDAP--HK-DLRRARATSLSLI | INALLKAASEQ-----APLQG-ILGYEERPLV |
| *Prochloron gap3* | QTIVDKP--HK-DLRRARSCLQSLI | VNSFLQGATEK-----EPLAG-ILGYEERPLV |
| *Rhodobacter gap3* | QTIVDRP--AK-DLRRARSALMNLI | VNSLLQTAAE------GGLNG-ILGFETRPLV |
| *Bacillus* | QQILDLP--HK-DYRRARAAAENII | VNAALKEAAE------GDLKG-ILGYSEEPLV |
| *Streptomyces* | QNLQDGP--HR-DPRRARAAAVNIV | VLDAYRAAAQ------GPLAG-VLEYSEDPLV |
| *Paracoccus gap1* | QPTLDTM--HK-DLYRARAAALSMI | INEAIRTAAN------GSLKG-ILGYTDEPLV |
| *Thermotoga* | QRVLDLP--HK-DLRRARAAAVNII | VNAVMKEATE------GRLKG-IIGYNDEPIV |
| *Trichomonas* | QVVADTM--HK-DLRRARAAGMNII | VNAALKKATEE-----GSLKG-IMTYVTDPIV |
| *Arabidopsis gap2* | QRLLDAS--HR-DLRRARAAALNIV | VNAAFRDSAE------KELKG-ILDVCDEPLV |
| ***Euglena gap2*** | **QMILDGR--HS-DLRRARAGAVNIV** | **VNAALKKAAD------GPLNG-VLGYTELPLV** |
| *Synechocystis gap2* | QRILDAS--HR-DLRRARAAAVNIV | VNGVLKEAAN------TSLKG-VLEYTDLELV |
| *Anabaena gap2* | QRLLDAS--HR-DLRRARAAAINIV | VNQALKDASE------GPLKG-ILDYSELQLV |
| *Prochloron gap2* | QRLLDAS--HR-DLRRARAAAINIV | VNDALKAAAE------GPLKG-ILRYNDLPLV |
| *Synechocystis gap1* | QPTVDGPS-KK-DFRGGRGAAQNII | ICAAMKTAAE------GELKG-ILGYTADDVV |
| *Anabaena gap1* | QPTVDGPS-KK-DWRGGRGAAQNII | ICAAMKQASE------GSLAG-ILGYTDEEVV |
| *Treponema palladum* | QKTVDGVS-LK-DWRGGRAAAVNII | IDKAIKKASE------SYLKG-VLAYCDEEIV |
| ***Trypanosoma* glycosome** | **QKTVDGVS-VK-DWRGGRAAALNII** | **IDAALKRASK------TYMKN-ILGYTDEELV** |
| ***Crithidia* glycosome** | **QKTVDGVS-IK-DWRGGRAAAVNII** | **IDAALKKASQ------TYMKG-ILGYTDEELV** |
| ***Euglena* cytosol** | **QKTVDGPS-KK-DWRGGRAAAINII** | **IDSLLKKASQ------TYLKG-ILGFTDEELV** |
| *Escherichia gapA* | QKTVDGPS-HK-DWRGGRGASQNII | IKAAVKAAAE------GEMKG-VLGYTEDDVV |
| ***Trypanosoma* cytosol** | **QKTVDGPS-QK-DWRGGRGAAQNII** | **ICDAIKAASE------GELKG-ILGYVDEEIV** |
| ***Leishmania* cytosol** | **QKTVDGPS-LK-DWRGGRGASQNII** | **ICAAIKAAAE------GEMKG-ILGYTDDEVV** |
| ***Diplonema* 50224 *gap1*** | **QKTVDAPS-KK-DWRGGRGILGNII** | **IDAAVLKASQS-----GKMAG-VIGFTNEDVV** |
| *Naegleria* | QPTVDAPS-KK-DWRGGRAAGYNII | IDEAMKKASES-----ERFKG-ILKYTDEEVV |
| *Acrasis* | QKTVDGPQ-RG-DWRFGRGAAFNII | IDKTLKEAANS-----ERWKN-IFAYTDDDVV |
| *Gonyaulux* cytosol | QLTVDGPSVAARTGVVAAAHPQNII | IVAAIKEAAA------GPMSG-VLDWTDEEVV |
| *Plasmodium* cytosol | QLVVDGPSKGGKDWRAGRCALSNII | VALEIKKAAE------GLLKG-VLGYTEDEVV |
| *Giardia* | QLPVDGPS-KK-DWRGGRSCGANVI | ICAEIKRGSE------NELKG-IMTYTNEDVV |
| *Gracilaria* | QKTVDGPS-QK-DWRGGRGAGANII | IKATMKAAAE------DSMKG-ILKYTEEAVV |
| *Arabidopsis* | QKTVDGPS-MK-DWRGGRAASFNII | IKKAIKEESE------GKLKG-ILGYTEDDVV |
| *Dictydistelium* | QKTVDGPS-GK-DWRAGRSALSNII | IKKVMKAASES-----DKYKR-YIGYTEDEVV |
| *Saccharomyces* | QKTVDGPS-HK-DWRGGRTASGNII | IKKAVKAAAE------GPMKG-VLGYTEDAVV |
| *Drosophila* | QKTVDGPS-GK-LWRDGRGAAQNII | IKAKVEEASK------GPLKG-ILGYTDEEVV |

**Figure 1.** Protein insertions in diplonemid *gap3* genes. Two blocks of the GAPDH sequence alignment corresponding to amino acids 186–207 (left block) and 258–283 (right block) of the *Anabaena gap3* gene. Diplonemid, kinetoplastid, and euglenoid sequences are shown in bold, and in cases where more than one gene is known the location or copy number is indicated by the name. Four insertions, conserved in position and sequence, are common to all three diplonemid *gap3* genes (labeled A, B, C, and D at top). Other insertions in other regions of the gene are also common to the three diplonemid *gap3* sequences, but these are not shown as they occur in less conserved regions of the gene, and are therefore less easily interpreted. The single amino acid, non-unique insertion common to *Diplonema* ATCC 50224 *gap1* and Heterolobosea is in region C.

altogether suggesting that these are closely related members of a relatively divergent class of enzyme. *gap1*-like genes were not found in either *Diplonema* ATCC 50225 or *Rhynchopus* ATCC 50231, although this of course in no way suggests that they do not exist.

## Diplonemid GAPDH Phylogeny

A global phylogeny of GAPDH showing the general features of GAPDH evolution is shown in Figure 2. There have been a number of attempts to make sense of the overall GAPDH tree and to classify the various sublineages (Figge et al. 1999; Hannaert et al. 1998; Henze et al. 1995; Michels et al. 1991; Wiemer et al 1995), which can be quite complex. For simplicity, the GAPDH tree may be divided into two major lineages, the bacterial GapA/B group and the eukaryotic cytosolic GapC group, with the interface between these groups including a number of poorly resolved bacterial and eukaryotic lineages. Within both GapA/B and GapC, a number of strongly supported groups are found, but the branching order of these groups is generally poorly resolved.

Diplonemid GAPDH sequences can been seen to occupy two very distinct positions in the tree. First, the single representative of diplonemid *gap1* (from *Diplonema* ATCC 50224) branches as one would expect with eukaryotic *gapC* genes, but unexpectedly does not branch with other Euglenozoa. Instead, this gene branches very weakly with the Heterolobosea (*Naegleria* and *Acrasis*) in Figure 2, and in slightly different positions in other analyses (not shown). This relationship with Heterolobosea is very weak, but intriguing nonetheless since numerous gene trees are now showing a close relationship between Heterolobosea and Euglenozoa (Baldauf et al. 2000; Keeling et al. 1998; Van de Peer et al. 2000), and both heterolobosean GAPDH genes as well as *Diplonema* ATCC 50224 *gap1*share a non-unique single amino acid insertion at a variable position in the alignment (see region C of Fig. 1). If this weakly-supported relationship is true, it would most likely indicate that the *Diplonema* ATCC 50224 *gap1* represents the ancestral GAPDH in Euglenozoa, since it is the only euglenozoan GAPDH that shows the expected relationship to Heterolobosea. Unfortunately, the position of *Diplonema* ATCC 50224 *gap1* within the GapC clade is so unstable that its putative relationship to Heterolobosea will likely never be clarified, a condition exemplified by Kishino-Hasegawa tests which fail to reject 8 alternative positions for this gene within GapC. Nevertheless, trees placing *Diplonema* ATCC 50224 *gap1* with either the kinetoplastid cytosolic clade or the kinetoplastid glycosomal-euglenoid cytosolic clade are both rejected at the 1% level, confirming that the diplonemid *gap1* is indeed a lineage distinct from all other characterised euglenozoan GAPDH genes. Similarly, all trees in which this gene is placed in the GapA/B clade, including with the diplonemid *gap3* clade or the *Euglena* plastid gene, are also rejected at the 1% level, confirming that this gene is clearly a GapC.

In contrast to the diplonemid *gap1*, the highly divergent *gap3* genes from *Diplonema* and *Rhynchopus* do not branch with other eukaryotes, but instead branch specifically with a cluster of genes from cyanobacteria and proteobacteria referred to as Gap3 (named for the cyanobacterial *gap3* genes). This grouping is very strongly supported in all analyses (with consistent bootstrap support of 100%) and is quite clearly discerned from a visual inspection of the alignment as these genes share a number of unique substitutions (not shown). In support of this, when these genes were moved to 12 alternative positions and the resulting trees compared by Kishino-Hasegawa tests, all alternative trees, including any tree in which diplonemid *gap3* branched with any other euglenozoan GAPDH, were rejected at the 1% level.

This relationship appears to represent a relatively recent lateral gene transfer of a bacterial *gap3* to diplonemids, presumably occurring after the divergence of diplonemids from euglenoids and kinetoplastids. The relationship between diplonemid and cyanobacterial GAPDH genes should not be interpreted as indicating a cryptic photosynthetic ancestry of diplonemids, since the cyanobacterial *gap3* is not the paralogue that has been retained in plastids (cyanobacterial *gap2* is the plastid enzyme) and *gap3* genes contain signature residues expected of a catabolic enzyme (Clermont et al. 1993), suggesting they are not involved in Calvin cycle. It is also unlikely that the diplonemid *gap3* genes are bacterial contaminants, since the cultures were axenic and three distinct but related genes were isolated from the three distinct but related diplonemids, demanding that three different but related bacterial contaminants would have to have been introduced. One possible exception would be if the diplonemids harboured a long-standing endosymbiotic bacterium, since such an endosymbiont could diverge and evolve in parallel with the host to some extent, and lead to the results observed here. However, ultrastructural investigations on all three species examined here have not revealed any such endosymbiont (T. Nerad, personal communication), so this remains an unlikely alternative to lateral gene transfer.

Since the branch uniting diplonemid *gap3* and other Gap3 genes is quite long, the position of the

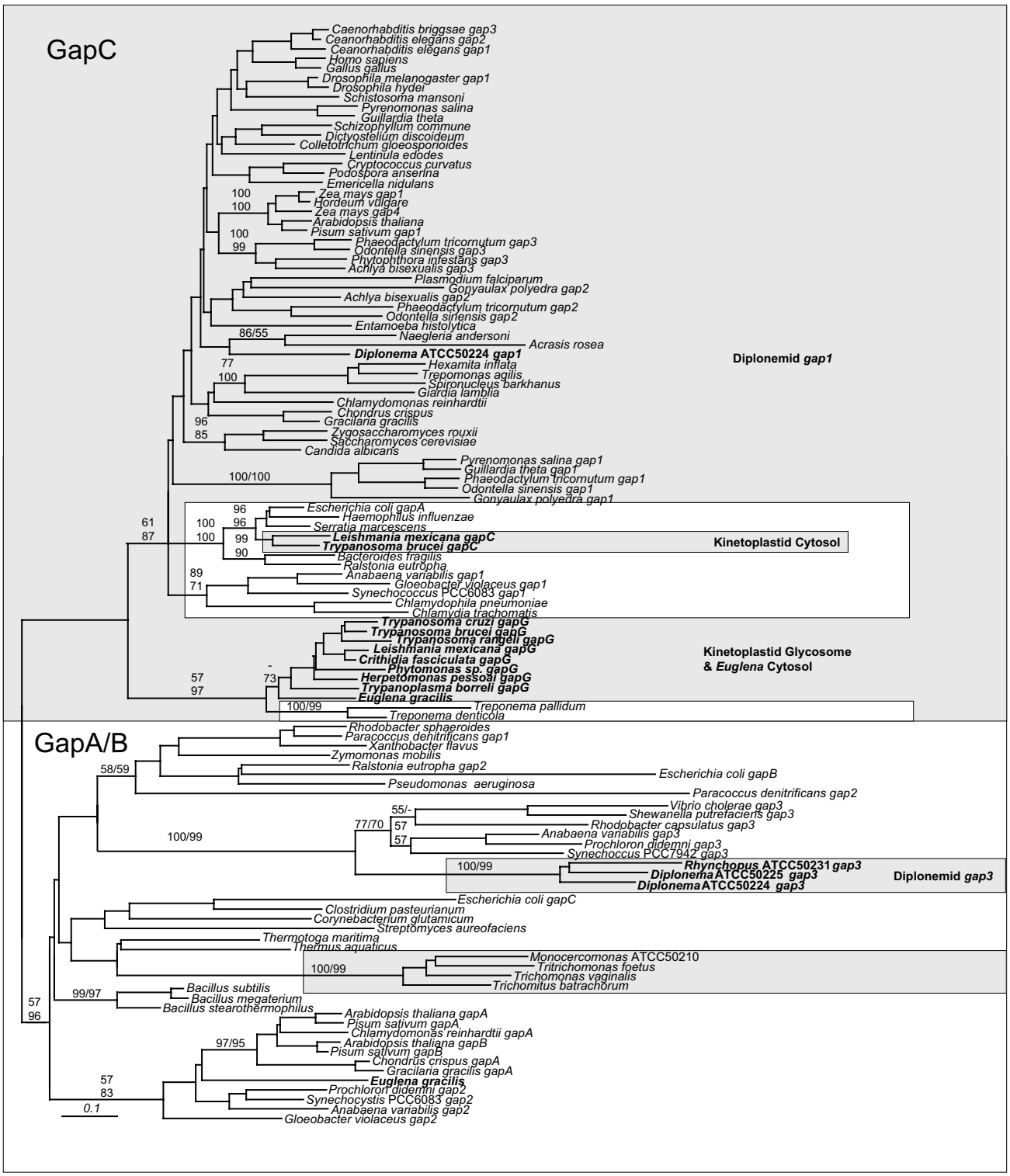


**Figure 2.** GAPDH phylogeny shaded to show instances of lateral gene transfer. Weighted neighbor-joining tree of 115 sequences based on gamma-corrected distances. Numbers at nodes indicate bootstrap support from weighted neighbor-joining (top) and Fitch-Margoliash (bottom). Scale bar indicates 0.1 (corrected) substitutions per site. In cases where more than one gene is known, the copy number is indicated by the name (e.g. *gap1*, or *gapC*). Eubacterial genes (including cyanobacteria-derived plastid-targeted sequences from plants and algae) are shown in white while eukaryotic genes are shaded gray, and genes from kinetoplastids, *Euglena*, and diplonemids are shown in bold, and the various classes labeled to the right.

root of the Gap3 clade is questionable, and one therefore cannot say with any certainty whether the diplonemid genes are specifically related to cyanobacteria, proteobacteria, or neither. To attempt to address this, the phylogeny of the nine Gap3 genes was inferred independently. Here (Fig. 3), it can be seen that different methods do not agree on the internal branching order of the Gap3 clade: weighted neighbor-joining and Fitch-Margoliash distance trees (Fig. 3A) show the diplonemids, cyanobacteria, and proteobacteria are three monophyletic lineages, while protein maximum likelihood (Fig. 3B) shows diplonemid *gap3* genes specifically related to the proteobacterium, *Rhodobacter capsulatus*. However, neither analysis is very well supported, so no firm conclusion on the exact source of the genes can be made.

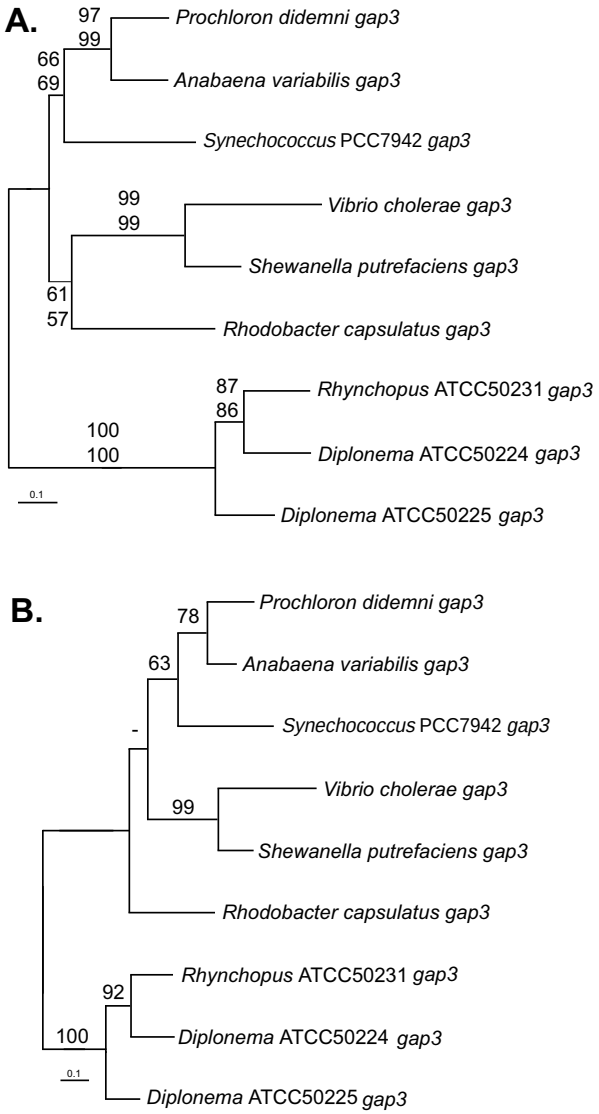


**Figure 3.** Unrooted phylogeny of diplonemid, cyanobacterial, and proteobacterial *gap3* genes using (A) protein distances and (B) protein maximum likelihood. Numbers at nodes correspond to bootstrap support from weighted neighbor-joining (upper) and Fitch-Margoliash (lower) in part A, and protein maximum likelihood in part B. Trees A and B differ only in the position of diplonemids: in tree A diplonemids, cyanobacteria, and proteobacteria are three monophyletic groups, while in tree B the diplonemids are specifically related to *Rhodobacter*. Neither topology is strongly supported so the exact taxonomic origin of the diplonemid gene cannot be inferred with the present data.

## Functional Implications of Diplonemid *gap3* Lateral Transfer

While the presence of a plastid makes it obvious why *Euglena* has two GAPDH enzymes, it is not as clear why kinetoplastids and diplonemids would also have more than one GAPDH. In the case of diplonemids, a clue may come from the fact that *gap3* is also seemingly redundant in the bacteria where it is found. This may be significant as other apparently redundant GAPDH enzymes in other bacteria have been shown not to be GAPDH at all. In *E. coli*, the highly divergent *gapB* gene has been shown not to use glyceraldehyde-3-phosphate as a substrate, but rather to use erythrose-4-phosphate as part of the pyridoxal-5'-phosphate biosynthetic pathway (Zhao et al. 1995). The same now seems to be the case for at least some of the related "GAPDH" homologues in other bacteria (Carroll et al. 1997), raising questions about the role of various genes in genomes with multiple distant homologues. The function of *gap3* is completely unknown, but it is apparently redundant when present and is comparatively divergent, so it seems reasonable to suppose that it may also utilise some other substrate. If so, then the transfer of this enzyme to diplonemids takes on an even greater interest since it would not just be replacing or backing-up some existing biochemical function, but would be adding variability to the biochemical capacity of diplonemids.

## Trans-Domain Lateral Gene Transfer to Eukaryotes

The role of lateral gene transfer in evolutionary considerations has been transformed in recent years, from a minor but interesting quirk to an important and universal force. However, the strength of evidence for different types of lateral transfer varies considerably, so it is still difficult to generalise on the

importance of the process in nature. It is important to remember that the likelihood of any two genes being successfully transferred will almost certainly vary depending on the activities and interactions of the gene products. Similarly, there can be transfer between members of each of the three domains, or trans-domain transfers between any two domains, and each of which will have different molecular, cellular, and ecological impediments. In short, not all transfers should be considered equally.

In prokaryotes, there is now evidence for substantial transfer between closely related organisms (Ochman et al. 1999), and there is mounting evidence for a moderate or even high frequency of transfer between more distantly related taxa, including trans-domain transfers between eubacteria and archaebacteria (Boucher and Doolittle 2000; Doolittle 1999), suggesting that these kinds of transfers are not as rare as once thought.

Compared to prokaryotes, few eukaryotic genomes have been completely sequenced, but from those that have been completed there is little evidence of rampant gene exchanges (for instance, a relatively small number of putatively bacterial genes have been identified in human, and the evidence that these are really bacterial is generally weak: International Human Genome Sequencing Consortium 2001). As we have pointed out, however, this could be a reflection of the sampling of eukaryotic genomes since many animals, fungi, and land plants arguably stand a poorer-than-average chance of successfully integrating a foreign gene into their genome, and these represent the source of the only completed genomes, as well as most other molecular data. This raises the question: is there a fundamental difference between the importance of lateral transfer in prokaryotes and eukaryotes, or have we inadequately sampled eukaryote molecular diversity to detect this process in abundance? Several recent molecular studies seem to support the latter, for instance N-acetylneramate lyase of parabasalia (de Koning et al. 2000), HMG-CoA reductase of diplomonads (Boucher and Doolittle 2000), and chymotrypsin of fungi (Screen and St. Leger 2000). In GAPDH phylogeny alone there appear to be no less than three individual cases of lateral transfer from a bacterium to a eukaryote: involving parabasalia (Markos et al. 1993), kinetoplastids (Figge et al. 1999; Hannaert et al. 1998; Henze et al. 1995; Michels et al. 1991; Wiemer et al 1995), and the current example involving diplonemids. While the current data are not sufficient to say how common prokaryote-to-eukaryote lateral transfer is, these and other recent studies of isolated cases do suggest that this process does take place and could potentially be important. On the other hand, in the many completed prokaryotic genomes there is little evidence for substantial numbers of genes derived from eukaryotes (with a few possible exceptions: e.g. Brown and Doolittle 1999; Olendzenski et al. 2000; Subramanian et al. 2000), suggesting that trans-domain transfer in this direction is impeded in some way.

The discrepancies between prokaryotic and eukaryotic data *may* suggest a fundamental difference between the role of lateral transfer in these organisms, but until more data from protists are available it would be premature to make any firm conclusions. Clearly, however, transfer involving different genes, different groups of organisms, and different directions of transfer will have to be considered separately in any global assessment of what impact lateral transfer has on genome evolution.

## Methods

**Strains, culture conditions:** *Diplonema* ATCC 50224 (sp. 2 strain IIIGPC), *Diplonema* ATCC 50225 (sp. 3 strain VIGPC), and *Rhynchopus* ATCC 50231 (sp. 3 strain VB) were cultivated in enriched *Isonema* medium (ATCC 1728) with 10% heat inactivated horse serum. Cultures were harvested by centrifugation, and genomic DNA was purified by multiple extractions with CTAB and chloroform (Clark 1992).

**Amplification and sequencing:** Diplonemid GAPDH genes were amplified using the primers CCAAGGTCGGNATHAAYGGNTTYGG and CGAGTAGCCCCAYTCRTTRTCRTACCA. PCR products were separated by gel electrophoresis and isolated using GenClean II (Bio101, Vista, CA). Purified products were cloned using TOPO TA cloning and the pCR2.1 vector (Invitrogen, Carlsbad, CA). Multiple clones of each product were sequenced on both strands, and in all cases a single coding region was found, with the exception of *Diplonema* ATCC 50224, where two classes of GAPDH were found, as described. New sequences have been deposited in GenBank as accession numbers AY033583-AY033586.

**Phylogenetic analysis:** Conceptual translations of new diplonemid GAPDH genes were added to an existing alignment, and phylogenetic trees constructed from 290 clearly alignable positions (the alignment is available upon request). Distances were calculated with PUZZLE 5.0 (Strimmer and von Haeseler 1996) using the WAG substitution frequency matrix and amino acid usage estimated from the data. Site-to-site rate variation was modeled on a gamma distribution with 8 rate categories and the

shape parameter estimated from the data using exact estimates (the estimated gamma parameter was 0.97) . Trees were constructed using WEIGHBOR (Bruno et al. 2000), Fitch-Margoliash (Felsenstein 1993), and BioNJ (Gascuel 1997), which did not produce significant differences. Bootstrap analyses were carried out using the methods described above, except that the gamma shape parameter was estimated from the original data set, using puzzleboot (script by M. Holder and A. Roger: http://www.tree-puzzle.de/). Trees restricted to *gap3* were constructed as above, and also by protein maximum likelihood using ProML (Felsenstein 1993) with global rearrangements, 10 input order jumbles and using the –R option for site-to-site rate variation (in this case with 9 categories corresponding to invariant sites and 8 rate categories estimated by PUZZLE). Similar analyses were also performed on the GapC and GapA/B divisions independently, but the relative positions of the diplonemid genes were not found to vary. Kishino-Hasegawa tests (Kishino and Hasegawa 1989) were carried out using PUZZLE and the parameters described above by holding the topology in Figure 1 constant and moving the diplonemid *gap3* cluster or the *Diplonema* ATCC 50224 *gap1* gene to 24 alternative positions (12 each), including all positions with one another or with other euglenozoan GAPDH clusters.

## Acknowledgements

We thank J. Archambault and E. E. Chao for technical assistance, N. M. Fast for providing the GAPDH alignment, critical reading of the manuscript and general guidance, and T. Cavalier-Smith for support and guidance to Q.Q. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (227301-00). P.J.K. is a scholar of the Canadian Institute for Advanced Research.

## References

**Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF** (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. Science **290**: 972–977

**Boucher Y, Doolittle WF** (2000) The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. Mol Microbiol **37**: 703–716

**Brown JR, Doolittle WF** (1999) Gene descent, duplication, and horizontal transfer in the evolution of glutamyl- and glutaminyl-tRNA synthetases. J Mol Evol **49**: 485–495

**Bruno WJ, Socci ND, Halpern AL** (2000) Weighted neighbor joining: A fast approximation to maximum-likelihood phylogeny reconstruction. Mol Biol Evol **17**: 189–197

**Carroll PA, Zhao G, Boyko SA, Winkler ME, Calderwood SB** (1997) Identification, sequencing, and enzymatic activity of the erythrose-4-phosphate dehydrogenase gene of *Vibrio cholerae*. J Bacteriol **179**: 293–296

**Clark CG** (1992) DNA Purification from Polysaccharide-Rich Cells. In Lee JJ, Soldo AT (eds) Protocols in Protozoology. Society of Protozoology, Lawrence, KA, pp D3.1–D3.2

**Clermont S, Corbier C, Mely Y, Gerard D, Wonacott A, Branlant G** (1993) Determinants of coenzyme specificity in glyceraldehyde-3-phosphate dehydrogenase: role of the acidic residue in the fingerprint region of the nucleotide binding fold. Biochemistry **32**: 10178–10184

**de Koning AP, Brinkman FS, Jones SJ, Keeling PJ** (2000) Lateral gene transfer and metabolic adaptation in the human parasite *Trichomonas vaginalis*. Mol Biol Evol **17**: 1769–1773

**Doolittle WF** (1998) You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet **14**: 307–311

**Doolittle WF** (1999) Lateral genomics. Trends Cell Biol **9**: M5–8

**Eisen JA** (2000) Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. Curr Opin Genet Dev **10**: 606–611

**Felsenstein J** (1993) PHYLIP (Phylogeny Inference Package). Seattle: J. Felsenstein, University of Washington

**Figge RM, Schubert M, Brinkmann H, Cerff R** (1999) Glyceraldehyde-3-phosphate dehydrogenase gene diversity in eubacteria and eukaryotes: evidence for intra- and inter-kingdom gene transfer. Mol Biol Evol **16**: 429–440

**Gascuel O** (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol Biol Evol **14**: 685–695

**Hannaert V, Opperdoes FR, Michels PA** (1998) Comparison and evolutionary analysis of the glycosomal glyceraldehyde-3-phosphate dehydrogenase from different Kinetoplastida. J Mol Evol **47**: 728–738

**Henze K, Badr A, Wettern M, Cerff R, Martin W** (1995) A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. Proc Natl Acad Sci USA **92**: 9122–9126

**International Human Genome Sequencing Consortium** (2001) Initial sequencing and analysis of the human genome. Nature **409**: 860–921

**Keeling PJ, Deane JA, McFadden GI** (1998) The phylogenetic position of alpha- and beta-tubulins from the *Chlorarachnion* host and *Cercomonas* (Cercozoa). J Eukaryot Microbiol **45**: 561–570

**Kishino H, Hasegawa M** (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J Mol Evol **29**: 170–179

**Markos A, Miretsky A, Muller M** (1993) A glyceraldehyde-3-phosphate dehydrogenase with eubacterial features in the amitochondriate eukaryote, *Trichomonas vaginalis*. J Mol Evol **37**: 631–643

**Maslov DA, Yasuhira S, Simpson L** (1999) Phylogenetic affinities of *Diplonema* within the Euglenozoa as inferred from the SSU rRNA gene and partial COI protein sequences. Protist **150**: 33–42

**Michels PA, Marchand M, Kohl L, Allert S, Wierenga RK, Opperdoes FR** (1991) The cytosolic and glycosomal isoenzymes of glyceraldehyde-3-phosphate dehydrogenase in *Trypanosoma brucei* have a distant evolutionary relationship. Eur J Biochem **198**: 421–418

**Ochman H, Lawrence JG, Groisman EA** (2000) Lateral gene transfer and the nature of bacterial innovation. Nature **405**: 299–304

**Olendzenski L, Liu L, Zhaxybayeva O, Murphey R, Shin DG, Gogarten JP** (2000) Horizontal transfer of archaeal genes into the deinococcaceae: detection by molecular and computer-based approaches. J Mol Evol **51**: 587–599

**Qian Q** (2000) The evolutionary implications of diplonemids and their spliceosomal introns, MSc Thesis, University of British Columbia

**Roger AJ** (1999) Reconstructing early events in eukaryotic evolution. Am Nat **154**: S146–S163

**Schuster FL, Goldstein S, Hershenov B** (1968) Ultrastructure of the flagellate *Isonema nigrigricans* nov. gen. nov. from a polluted marine habitat. Protistologica **4**: 141–149

**Screen SE, St Leger RJ** (2000) Cloning, expression, and substrate specificity of a fungal chymotrypsin. Evidence for lateral gene transfer from an actinomycete bacterium. J Biol Chem **275**: 6689–6694

**Simpson AGB** (1997) The identity and composition of the Euglenozoa. Arch Protistenkd **148**: 318–328

**Strimmer K, Von Haeseler A** (1996) Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. Mol Biol Evol **13**: 964–969

**Sturm NR, Maslov DA, Grisard EC, Campbell DA** (2001) *Diplonema* spp. possess spliced leader RNA genes similar to the Kinetoplastida. J Eukaryot Microbiol **48**: 325–331

**Subramanian G, Koonin EV, Aravind L** (2000) Comparative genome analysis of the pathogenic spirochetes *Borrelia burgdorferi* and *Treponema pallidum*. Infect Immun **68**: 1633–1648

**Van de Peer Y, Baldauf SL, Doolittle WF, Meyer A** (2000) An updated and comprehensive rRNA phylogeny of (Crown) eukaryotes based on rate-calibrated evolutionary distances. J Mol Evol **51**: 565–576

**van Leeuwen F, Taylor MC, Mondragon A, Moreau H, Gibson W, Kieft R, Borst P** (1998) beta-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. Proc Natl Acad Sci USA **95**: 2037–2038

**Wiemer EA, Hannaert V, van den Ijssel PR, Van Roy J, Opperdoes FR, Michels PA** (1995) Molecular analysis of glyceraldehyde-3-phosphate dehydrogenase in *Trypanoplasma borelli*: an evolutionary scenario of subcellular compartmentation in kinetoplastida. J Mol Evol **40**: 443–454

**Zhao G, Pease AJ, Bharani N, Winkler ME** (1995) Biochemical characterization of *gapB*-encoded erythrose 4-phosphate dehydrogenase of *Escherichia coli* K-12 and Its possible role in pyridoxal 59-phosphate biosynthesis. J Bacteriol **177**: 2804–2812