BMC Evolutionary Biology

BioMed Central

Research article

Open Access

Evolution of glutamate dehydrogenase genes: evidence for lateral gene transfer within and between prokaryotes and eukaryotes

Jan O Andersson*[1,2] and Andrew J Roger[1]

Address: [1]The Canadian Institute for Advanced Research, Program in Evolutionary Biology, Department of Biochemistry & Molecular Biology, Dalhousie University, Sir Charles Tupper Medical Building, Halifax, Nova Scotia, B3H 1X5, Canada and [2]Current address: Institute of Cell and Molecular Biology, Uppsala University, Biomedical Center, Box 596, S-751 24 Uppsala, Sweden

Email: Jan O Andersson* - Jan.Andersson@icm.uu.se; Andrew J Roger - aroger@dal.ca

* Corresponding author

Published: 23 June 2003

BMC Evolutionary Biology 2003, **3**:14

Received: 26 March 2003
Accepted: 23 June 2003

This article is available from: http://www.biomedcentral.com/1471-2148/3/14

© 2003 Andersson and Roger; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Lateral gene transfer can introduce genes with novel functions into genomes or replace genes with functionally similar orthologs or paralogs. Here we present a study of the occurrence of the latter gene replacement phenomenon in the four gene families encoding different classes of glutamate dehydrogenase (GDH), to evaluate and compare the patterns and rates of lateral gene transfer (LGT) in prokaryotes and eukaryotes.

Results: We extend the taxon sampling of *gdh* genes with nine new eukaryotic sequences and examine the phylogenetic distribution pattern of the various GDH classes in combination with maximum likelihood phylogenetic analyses. The distribution pattern analyses indicate that LGT has played a significant role in the evolution of the four *gdh* gene families. Indeed, a number of gene transfer events are identified by phylogenetic analyses, including numerous prokaryotic intra-domain transfers, some prokaryotic inter-domain transfers and several inter-domain transfers between prokaryotes and microbial eukaryotes (protists).

Conclusion: LGT has apparently affected eukaryotes and prokaryotes to a similar extent within the *gdh* gene families. In the absence of indications that the evolution of the *gdh* gene families is radically different from other families, these results suggest that gene transfer might be an important evolutionary mechanism in microbial eukaryote genome evolution.

Background

Lateral gene transfer (LGT) is a significant evolutionary mechanism in prokaryotic genome evolution. Indeed, it may be the most important mechanism for evolutionary innovation in Eubacteria and Archaea [1,2]. However, gene transfer events do not necessarily produce novel functions in recipient lineages; many documented gene transfers are replacements of genes by homologs or analogs with the same function [3,4]. The occurrence of LGT has been far less studied in eukaryotes than prokaryotes, partly because of the lack of complete genome sequences available from diverse eukaryotes. Nevertheless, several individual cases of gene transfer between prokaryotes and eukaryotes have been published [for example: [5–8]]. We recently presented an analysis which showed a number of transfers involving eukaryotes, mostly in the prokaryote-to-eukaryote direction, but also between different eukaryotic lineages [9]. Collectively, these examples indicate that LGT does affect protists, although the quantitative importance of the process in eukaryotic genome evolution remains unclear [10,11]. We have selected the glutamate dehydrogenase (*gdh*) gene families to investigate the

relative frequency of gene transfer in prokaryotic versus eukaryotic genome evolution, to deepen our understanding of the extent to which gene transfer, in general, and gene replacement, in particular, affects eukaryotes.

GDH catalyzes the reversible oxidative deamination of glutamate to α-ketoglutarate and ammonia. These enzymes are very diverse and can be divided into four distinct classes. GDH-1 and GDH-2 are small hexameric enzymes with a broad taxonomic distribution that utilize either NAD^+ or $NADP^+$ as a coenzyme and function mainly in ammonia assimilation [12–14]. A class of larger (115 kDa) GDHs (herein referred to as GDH-3), that have previously been found only in fungi and protists, function in glutamate catabolism. Finally, enzymes of a fourth class (herein called GDH-4) have been recently discovered in Eubacteria [14] that are approximately 180 kDa in size and are NAD^+ specific.

A number of groups have previously investigated the evolution of the GDH enzyme families. A decade ago it was proposed that *gdh1* and *gdh2* originated via a single ancient gene duplication and therefore these paralogs could be used to root the universal tree of life [12]. However, as more *gdh* genes were collected, gene duplication scenarios required that multitudes of gene duplications and parallel loss events had to be invoked to explain the phylogenetic patterns observed. Brown and Doolittle suggested that it was more likely that at least part of the incongruity of GDH trees with organismal phylogeny were caused by other evolutionary processes such as LGT [13]. Phylogenetic analyses in a more recent study were unfortunately based on an alignment of all four classes of the enzyme, that are very distantly related, and did not include any bootstrap support values [14], making them difficult to interpret. Nevertheless, the phylogenetic distribution pattern of the different GDH classes between species and the phylogenetic analyses of the classes themselves clearly indicated that gene transfer was likely to be a significant evolutionary mechanism in the evolutionary history of these enzymes. Here we revisit these issues using up-to-date taxon-rich datasets that include novel sequences from eukaryotes. Our rigourous analyses of the phylogenetic relationships within these gene families have identified a number of cases of gene replacements, affecting both eukaryotes and prokaryotes. Analyses of the phylogenetic distribution of the four classes of GDH across the tree of life complement and further support the conclusion that LGT was relatively frequent in the evolutionary history of the *gdh* gene families.

Results and Discussion

Nine new eukaryotic GDH sequences

All available homologs of GDH were downloaded from public databases and some ongoing genome projects in order to study the evolution of *gdh* genes. As noted before [12,14], GDH-1 is found in eubacteria and eukaryotes, GDH-2 is found in all domains of life, and GDH-4 is only found in eubacteria. However, we identified two GDH-3 genes from the δ-proteobacteria *Desulfovibrio vulgaris* and *Geobacter sulfurreducens*; GDH-3 was previously found only in eukaryotes including fungi, euglenozoa and apicomplexa [14].

In order to study the evolution of eukaryotic GDHs in more detail, we also broadened the taxon sampling of *gdh* genes amongst eukaryotes. Seven new GDH-1 and two new GDH-2 sequences were obtained. GDH-1 and GDH-2 cDNA clones from the red alga *Porphyra yezoensis*, GDH-1 cDNA clones the oomycete *Phytophthora infestans*, the diplomonad *Spironucleus barkhanus*, and the parabasalid *Trichomonas vaginalis*, and a GDH-2 cDNA clone from the green alga *Chlamydomonas reinhardtii* were kindly made available from the various EST projects [15–17] and fully sequenced. GDH-1 sequences from the diplomonads *Spironucleus vortens* and *Hexamita inflata*, and the parabasalid *Monocercomonas* sp. were obtained using degenerate PCR.

Distribution of gdh genes among completely sequenced genomes

The phylogenetic distribution pattern of the genes within a gene family may provide indications of gene transfer events within the family. If a gene is present in distantly related organisms, but absent from many organisms that are more closely related, either extensive parallel gene losses or gene transfer events have to be inferred. The gene loss scenario requires that the gene was present in the ancestors of all the lineages that encode the genes.

We analyzed the pattern of the distribution of *gdh* genes in the three domains of life by analyzing the presence or absence of the four classes in all available completely sequenced genomes (Table 1). All classes of *gdh* have been found in eubacteria, all but *gdh4* have been found in eukaryotes, while only *gdh2* has been found in Archaea. *gdh* genes seem to be absent from some archaeal genomes as well as some of the smaller eubacterial and eukaryotic genomes. Among the organisms that do encode GDH, one or two classes are represented – no genome has yet been shown to encode three or all four classes (Table 1). Except for the unique presence of *gdh2* in Archaea, no strong pattern that is correlated with organismal phylogeny can be observed for the distribution of *gdh* genes (Table 1). For example, *gdh* genes encoding all four classes of the enzyme are present in various proteobacterial genomes. The distribution of the genes is scrambled even within this eubacterial group; both α- and γ-proteobacterial groups have members that encode *gdh1*, *gdh2* or *gdh4*, or a combination of two of them (Table 1 and Figs. 1,2,3).

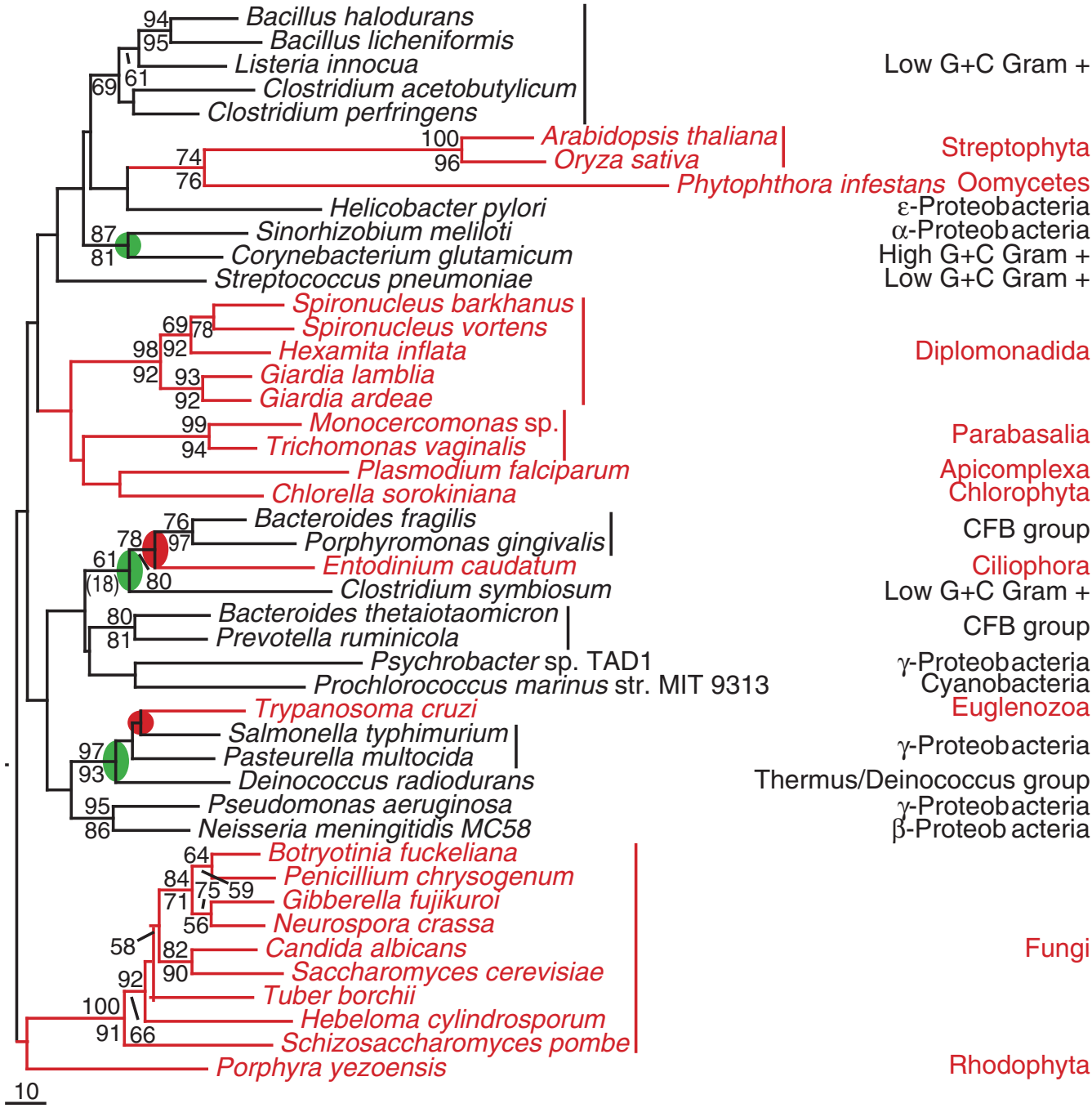

Figure 1

Maximum likelihood tree of GDH-1 The phylogenetic tree is based on 380 unambiguously aligned amino acid positions. The Γ shape parameter, α, was estimated to 0.76 with no invariable sites detected (P_{inv} = 0). The tree is arbitrarily rooted. Eukaryotes are labeled red and Eubacteria black. Potential inter-domain and intra-domain LGTs supported by a maximum likelihood bootstrap support value > 50% are indicated by red and green ovals, respectively. Protein maximum likelihood bootstrap values are shown above the branches and protein maximum likelihood distance bootstrap values are shown below the branches. Only values >50% for bipartitions are shown, except that maximum likelihood distance bootstrap values <50% are shown in parentheses at nodes where potential transfers are indicated.

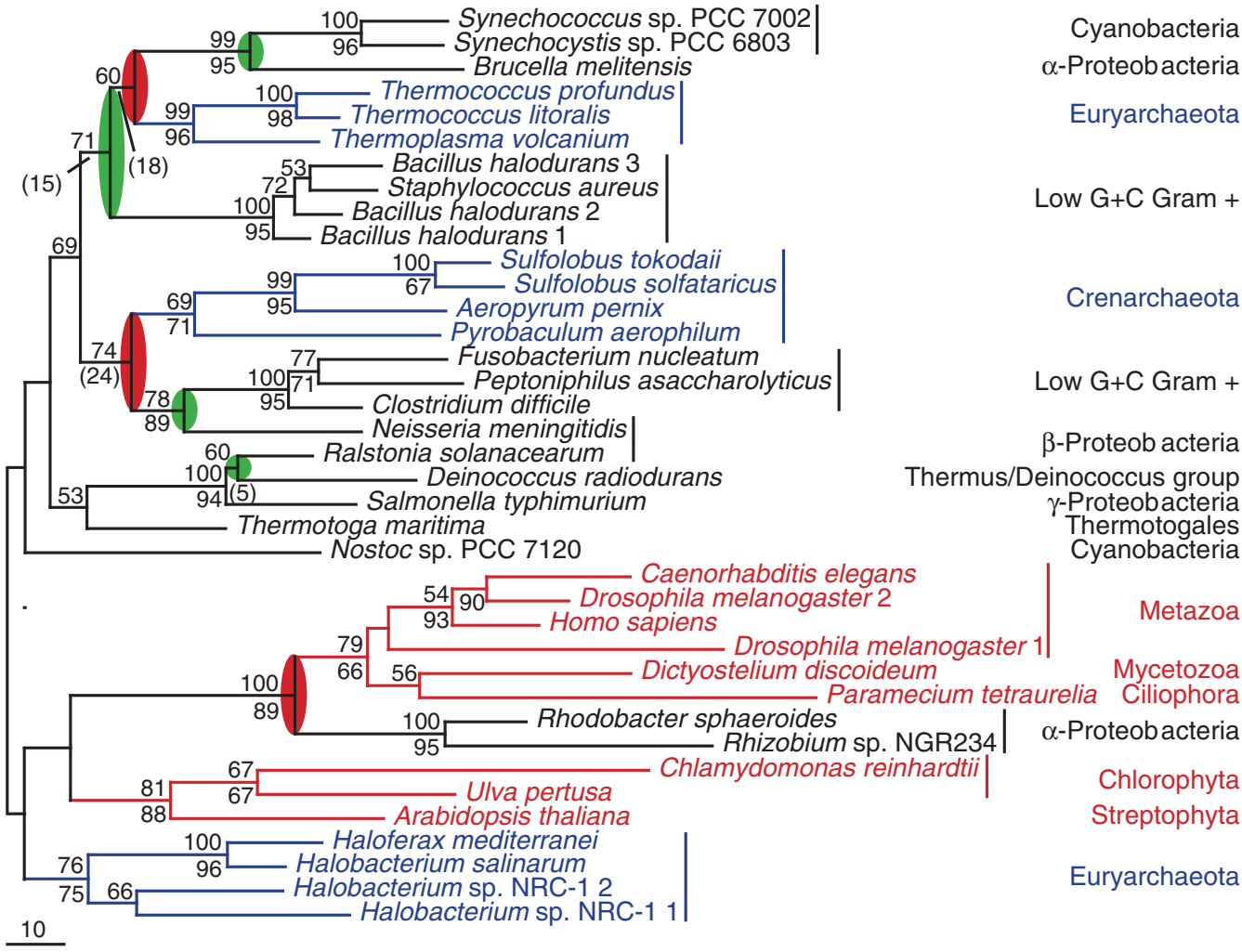

Figure 2
Maximum likelihood tree of GDH-2 The phylogenetic tree is based on 305 unambiguously aligned amino acid positions. The Γ shape parameter, α, and the fraction of invariable sites, P_{inv}, were estimated to 1.10 and 0.08, respectively. The tree is arbitrarily rooted. Labeling as in Figure 1 with the addition that Archaea are labeled in blue.

In the absence of gene transfer, this would require that the ancestral proteobacterium encoded all four classes, and the ancestors of α- and γ-proteobacteria encoded three classes which have been differentially lost in different lineages. This scenario seems very unlikely given that no sequenced extant genome contains more than two classes (Table 1). On the other hand, LGT events from outside proteobacteria, in combination with differential gene losses, could easily explain the gene distribution pattern within proteobacteria. For example, *Salmonella typhimurium* encodes both *gdh1* and *gdh2*, while no other γ-proteobacteria encode *gdh2* (Table 1). In this case, a recent transfer event to the *Salmonella* lineage is a more parsimonious explanation than a large number of parallel losses of *gdh2* within the γ-proteobacterial clade. Indeed, this interpretation is supported by phylogenetic analysis – the *Salmonella* sequence do not branch with any of the proteobacterial groups, which would be expected if the unique presence of *gdh2* in the *Salmonella* lineage amongst the γ-proteobacteria were due to losses in other lineages rather than a transfer event (Fig. 2 and discussion below). In conclusion, the weak correlation between eubacterial organismal phylogeny and the distribution of *gdh* genes (Table 1) indicates that LGT must have played a significant role in the evolution of this gene family, at least among eubacteria.

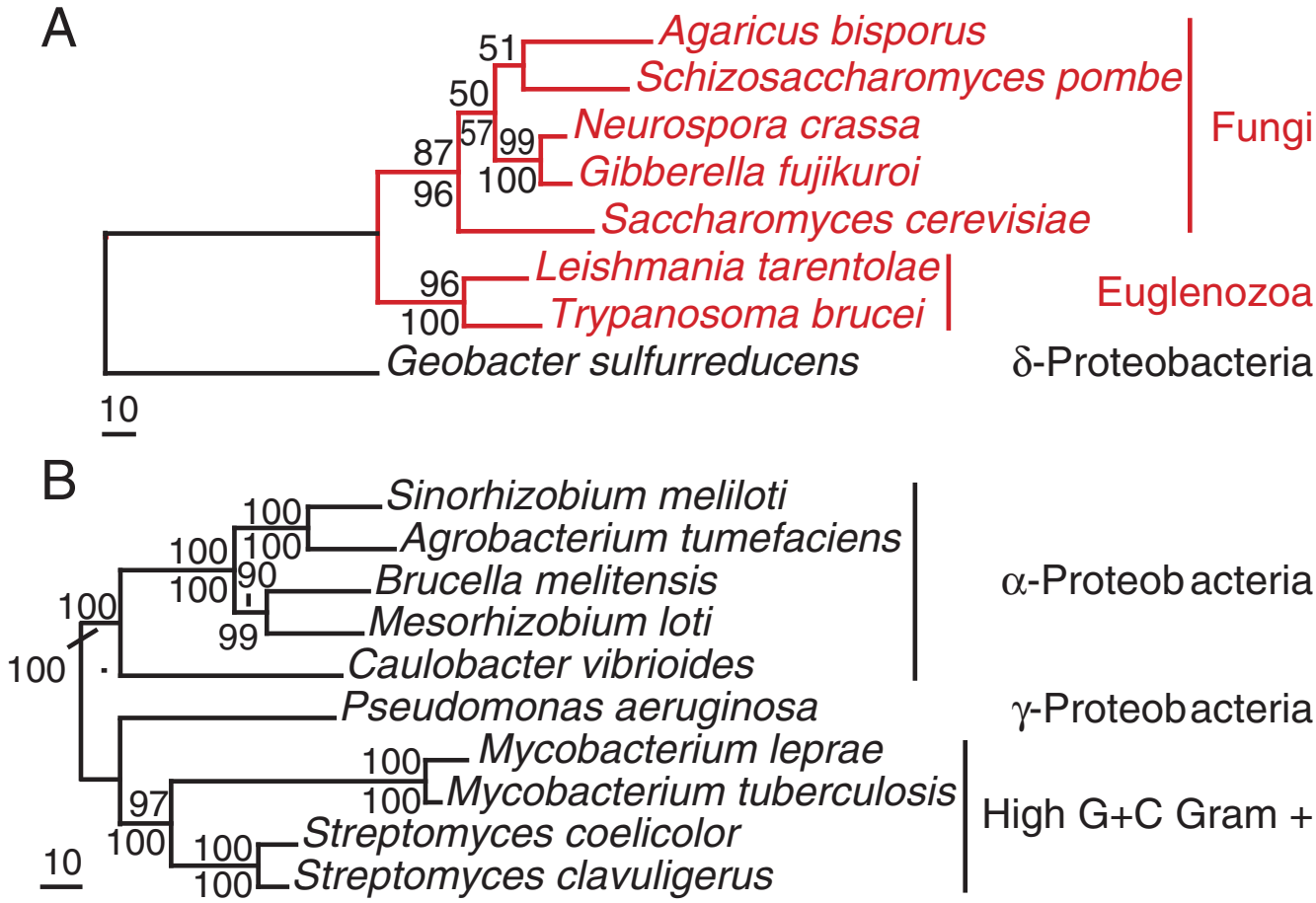

Figure 3
Maximum likelihood trees of GDH-3 and GDH-4 (A) The phylogenetic tree is based on 457 unambiguously aligned amino acid positions of GDH-3. The Γ shape parameter, α, and the fraction of invariable sites, P_{inv}, were estimated to 1.07 and 0.06, respectively. (B) The phylogenetic tree is based on 1141 unambiguously aligned amino acid positions of GDH-4. The Γ shape parameter, α, and the fraction of invariable sites, P_{inv}, were estimated to 1.14 and 0.14, respectively. The trees are arbitrarily rooted. Labeling as in Figure 1.

The scarceness of complete genome sequences for eukaryotes makes it problematic to use phylogenetic distribution patterns as evidence for, or against, gene transfer events affecting eukaryotes. The apparent lack of a gene family from an organism may simply reflect the incompleteness of our knowledge of its genome. Nevertheless, even given this incomplete knowledge, the distribution pattern of *gdh* genes is difficult to reconcile with current accounts of eukaryotic phylogeny. For example, metazoa and fungi share a eukaryotic ancestor to the exclusion of many other eukaryotic groups [18], but only *gdh2* genes are found in the complete metazoan genome sequences, while *gdh1* and *gdh3* genes are found in fungi (Table 1 and Figs. 1,2,3). One possible explanation for this pattern is that the ancestor of fungi and metazoa encoded all three classes with subsequent losses of *gdh2* in the fungal lineage and *gdh1* and *gdh3* in the metazoan lineage. Alternatively, at least one gene transfer to the metazoan or the fungal lineage could have occurred after their divergence creating the observed distribution of the *gdh* gene families. We favor the gene transfer scenario since no eukaryote has yet been found to encode three *gdh* genes. Improving the taxon sampling of *gdh* genes from the two groups and their closest relatives combined with phylogenetic analyses should clarify this issue.

Phylogenetic analyses of GDH sequences

The observed distribution of *gdh* genes could, in principle, *always* be explained by the presence of all families in a common ancestor, followed by massive differential losses

Table 1: Distribution of GDH among the organismal groups for which at least one complete genome sequence is available

Domain[a]	Group	#species[b]	GDH-1	GDH-2	GDH-3	GDH-4
A	Crenarchaeota	4		+		
A	Euryarchaeota	5		+		
A	Euryarchaeota	4				
B	Aquificales	1				
B	Chlamydiales	3				
B	Cyanobacteria	2		+		
B	High G+C Gram +	1	+			
B	High G+C Gram +	2				+
B	Low G+C Gram +	1	+	+		
B	Low G+C Gram +	5	+			
B	Low G+C Gram +	3		+		
B	Low G+C Gram +	6				
B	α-proteobacteria	1	+			+
B	α-proteobacteria	1		+		+
B	α-proteobacteria	5				+
B	β-proteobacteria	1	+	+		
B	β-proteobacteria	1		+		
B	γ-proteobacteria	4	+			
B	γ-proteobacteria	1	+	+		
B	γ-proteobacteria	1	+			+
B	γ-proteobacteria	2				+
B	γ-proteobacteria	1				
B	ε-proteobacteria	1	+			
B	ε-proteobacteria	1				
B	Spirochaetales	2				
B	Thermotogales	1		+		
B	Thermus/Deinoc-occus	1	+	+		
E	Fungi	2	+		+	
E	Metazoa	3		+		
E	Microsporidia	1				
E	Viridiplantae	1	+	+		

[a] A, Archaea; B, Eubacteria; E, Eukaryotes. [b] The number of species in the group that show the distribution pattern.

of paralogs. Phylogenetic reconstructions are therefore necessary to distinguish between this scenario and a situation whereby LGT created the scrambled distribution pattern of *gdh* genes. Differential gene loss of ancient paralogs is expected to produce phylogenetic trees for each family that broadly resemble the organismal phylogeny – i.e. accepted monophyletic organismal groups, such as, for example, proteobacteria and eukaryotes, should be recovered. LGT events, on the other hand, are expected to produce trees that are at odds with the expected organismal phylogenies.

The inferred GDH amino acid sequences were aligned for each class individually and unambiguously aligned regions were identified. Closely related sequences from different strains of the same species and closely related species were excluded to decrease the computational time for the analyses. Sequences with deviant amino acid composition were excluded to reduce the noise relative to the phylogenetic signal in the dataset. In previous studies, the different families of GDH have been aligned and combined phylogenetic analyses have been performed [12–14]. However, only two of the families, GDH-1 and GDH-2, can be unambiguously aligned over a significant part of the molecules. Phylogenetic reconstructions that include both families show a very long internal branch whose placement within each subtree is dependent on the taxon sampling within each family (data not shown). Therefore, separate maximum likelihood phylogenetic analyses for each GDH family were performed (Figs. 1,2,3).

Frequent eubacterial intra-domain LGTs

The GDH-1 and GDH-2 phylogenetic analyses strongly indicate that LGT has played an important role in the evolution of these gene families. For example, proteobacterial sequences are found in five groups in GDH-1, three of which are separated with statistical support values >80% in both bootstrap analyses, and five groups in GDH-2,

four of which separated with >85% bootstrap support in both analyses (Figs. 1 & 2). In the GDH-1 phylogenetic reconstruction the α-proteobacteria *Sinorhizobium meliloti* groups with the high G+C gram positive *Corynebacterium glutamicum*, and the γ-proteobacteria *Pasteurella multicoda* and *Salmonella typhimurium* form a strong group with *Deinococcus radiodurans* and the unicellular eukaryote *Trypanosoma cruzi*, to the exclusion of the other proteobacterial sequences in the tree (Fig. 1). Similarly, in the GDH-2 reconstruction the α-proteobacteria *Brucella melitensis* groups with two cyanobacterial sequences, while the two other α-proteobacterial sequences group with a large eukaryotic cluster. Also, the two β-proteobacterial sequences fail to group together in the GDH-2 tree – one is an immediate outgroup to a group with low G+C gram positive sequences, while the other is in a strongly supported cluster with a γ-proteobacterial sequence and the *Deinococcus* sequence (Fig. 2). Thus, several LGT events involving other eubacterial groups as well as eukaryotes have to be inferred to explain the distribution of the proteobacterial GDH sequences. The polyphyletic pattern is not unique to proteobacteria within the eubacterial domain; the three cyanobacterial GDH-2 sequences are separated into two distinct clusters with bootstrap support values of 99% and 95%, respectively (Fig. 2), and the low G+C gram positives sequences are split into at least two groups each for GDH-1 and GDH-2, albeit with slightly weaker support from the bootstrap analyses (Figs. 1 & 2). Taken together, the phylogenetic reconstructions strongly support our predictions based on the distribution pattern (Table 1) that there has been frequent LGT in the evolution of eubacterial GDH-1 and GDH-2.

LGT between the two prokaryotic domains

Only the GDH-2 gene family is found in Archaea (Table 1). At first glance, this might be taken as evidence that an ancestral archaeon encoded this class of the gene and that it was passed on to extant Archaea by vertical inheritance. However, the phylogenetic analysis of GDH-2 argues against this simple interpretation. The archaeal sequences are split into three distinct groups (Fig. 2); the *Thermoplasma* and *Thermococcus* sequences form a cluster with a cyanobacterial/α-proteobacterial group which is nested within a cluster of low G+C gram positive eubacteria with a bootstrap support value of 71% in the maximum likelihood analysis (Fig. 2), the crenarchaeote sequences group with another cluster of low G+C gram positives with 74% bootstrap support in the same analysis, and the *Halobacterium* and the *Haloferax* sequences form a group that is excluded from the two other archaeal clusters. The support values for these bipartitions are much weaker in the distance analysis, 18% and 24%, respectively. However, the archaeal sequences were never recovered as a monophyletic cluster in any of the 500 bootstrap replicates with either of the two methods. In fact, a cluster of the *Thermococcus*/*Thermoplasma* and crenarchaeote sequences was the only pairwise combination of the three archaeal groups obtained in the optimal maximum likelihood tree (Fig. 2) that was recovered in any of the bootstrap analyses; this grouping was found in 0,6% and 0,4% of the replicates in the maximum likelihood and distance analyses, respectively. Thus, the phylogenetic analysis fail to support the indications from the distribution analysis that the archaeal sequences are unaffected by gene transfer events. Rather, the recovered tree suggests two independent LGT events between eubacteria and archaea: one transfer between the low G+C gram positive eubacterial lineage and the crenarchaeotes, and another transfer to the *Thermococcus*/*Thermoplasma* lineage (Fig. 2).

Inter-domain LGT involving eukaryotes

The phylogenetic analysis of GDH-1 suggests that inter-domain transfer may not be limited to prokaryotes – the *T. cruzi* and the *Entodinium caudatum* sequences are phylogenetically distant from the other eukaryotic clusters (Fig. 1). The *Trypanosoma* sequence forms a group with two proteobacterial sequences and a *Deinococcus* sequence, with a bootstrap support values of >90% for the bipartition, indicative of a inter-domain gene transfer to the kinetoplastid lineage from a eubacterial lineage, possibly a γ-proteobacterium (Fig. 1). A second LGT event has to be inferred to explain the presence of a *gdh1* gene sequence in the ciliate *E. caudatum* which groups with sequences from *Bacteroides* and *Porphyromonas* with a bootstrap support values of >75% (Fig. 1). The phylogenetic analysis of GDH-1 suggests additional gene transfer events; eukaryotes emerge in five different places in the tree (Fig. 1). Unfortunately, the additional LGT events implied by this branching pattern can neither be proved nor disproved, since the backbone of the GDH-1 phylogenetic tree is poorly resolved. Three of the eukaryotic groups – the plant/oomycete cluster, the large protist cluster and the fungi/red algal cluster – could indeed represent a large eukaryotic GDH-1 group (Fig. 1).

Two eukaryotic groups are found in the GDH-2 phylogenetic tree, one cluster with two green algal sequences and an *Arabidopsis* sequence, and a second larger cluster. As mentioned, two α-proteobacterial sequences form a strongly supported group with the larger eukaryotic cluster (Fig. 2). This α-proteobacterial/eukaryotic cluster is a sister to the green algal/land plant cluster. Several different evolutionary scenarios could have produced this pattern. A gene transfer may have occurred from the common ancestor of metazoan, slime mold and ciliate sequences to the α-proteobacterial lineage. In this case, the eukaryote lineage would be one large clade with the eubacterial grouping arising from within them. Alternatively, this transfer event could have happened in the opposite direction and the eukaryotic groups origi-

nated via one, or maybe two, gene transfer events from eubacteria. If so, the transfer to the larger eukaryotic group could, in principle, represent an endosymbiotic gene transfer, since the ancestor of the mitochondria was an α-proteobacterium. However, this scenario is problematic for two reasons; α-proteobacterial sequences are also found in another part of the GDH-2 tree, as well as in the GDH-1 and GDH-4 trees (Figs. 1,2,3), and multiple independent losses in eukaryotic lineages have to be inferred.

Phylogeny of GDH-3 and GDH-4

The phylogenetic reconstructions of GDH-1 and GDH-2 in combination with the phylogenetic distribution analyses (Table 1 and Figs. 1 and 2) provide strong support for multiple inter- and intra-domain gene transfer events. The phylogeny of GDH-3 and GDH-4, on the other hand, fail to indicate transfer events – all recovered clusters represent expected organismal groups (Fig. 3). However, this should not be taken as evidence that these two classes have never suffered LGT – the recovered organismal groups are only distantly related. For example, the presence of *gdh3* genes in eukaryotes, and δ-proteobacteria among prokaryotes, indicates at least one transfer between eubacteria and eukaryotes, unless an enormous number of parallel losses of the gene are to be invoked amongst eubacteria. Also, since fungi and Euglenozoa are rather distantly related eukaryotic groups [18], either several independent losses of the *gdh3* gene in other eukaryotic lineages must have occurred, or a gene transfer event must be invoked (Fig. 3A). Similarly, the *gdh4* gene is only present in a few species that do not form one coherent organismal group, which is indicative of intra-domain LGT (Fig. 3B).

The relative rates of LGT may be comparable in prokaryotes and microbial eukaryotes

The strongest evidence for a gene transfer is a close phylogenetic relationship between gene sequences of distantly related organisms to the exclusion of gene sequences of more closely related organisms. Above we have described this kind of evidence for LGT within the *gdh* gene families – many relationships are at odds with accepted views of organismal relationships in the phylogenetic reconstructions of GDH-1 and GDH-2 (Figs. 1 & 2). Among the potential transfers with maximum likelihood bootstrap support values above 50% (green and red ovals in Figs. 1 and 2), there are seven transfers suggested between different lineages of eubacteria and two transfers between eubacteria and archaea (Figs. 1 & 2). Intriguingly, the phylogenetic reconstructions also indicate three cases of gene transfers involving eukaryotes; a ciliate and a kinetoplastid probably acquired *gdh1* genes from two different eubacterial lineages (Fig. 1), and the large eukaryotic group most likely exchanged a *gdh2* gene with the α-proteobacterial lineage (Fig. 2). The strength of the support for the putative transfers differs between the individual cases and between the two phylogenetic methods. While all three putative transfers involving eukaryotes show strong support from both analyses and most likely represent true cases of LGT, two of the suggested intra-domain eubacterial transfers and both the prokaryotic inter-domain transfers show only weak support from the maximum likelihood distance analysis and therefore should be viewed as more tentative cases (Figs. 1 & 2). All potential transfers affecting eukaryotes seem to have involved microbes – the two transfers of *gdh1* genes involve protists and the common ancestor of the large eukaryotic GDH-2 group was most likely unicellular. Unfortunately, the relative rates of transfers are extremely difficult to estimate due to the limited number of events, poorly resolved phylogenies, a highly non-random taxon sampling and our incomplete knowledge of organismal relationships within the three domains of life. Nevertheless, the observed possible transfers suggest that the rate at which LGT occurs within the *gdh* gene families in microbial eukaryotes is comparable to the rate in prokaryotes.

Conclusions

This work clearly demonstrates that analyses of distribution patterns of genes should be complemented with phylogenetic reconstructions of the genes themselves in order to distinguish between differential gene loss and gene transfer [19]. The combination of phylogenetic reconstructions and analyses of phylogenetic distribution patterns of the four *gdh* gene families provide strong support for numerous gene transfers involving prokaryotes, as well as microbial eukaryotes. Differential gene loss, on the other hand, does not seem to have played an important role in the evolution of *gdh* genes in any of the three domains of life. The rates at which lateral gene transfer occurs in prokaryotes versus microbial eukaryotes may be similar. We predict that systematic analyses, such as this, of a much wider array of gene families will show that LGT is an important evolutionary mechanism in genome evolution among protists.

Methods

PCR and sequencing of eukaryotic gdh genes

To extend the sampling of *gdh* genes from diverse eukaryotes we PCR amplified and sequenced *gdh*1 genes from the diplomonads *S. vortens* (strain ATCC 50386) and *H. inflata* (strain AZ-4) and the parabasalid *Monocercomonas* sp. (strain NS-1PRR ATCC 50210). The degenerate primers GDH1f1 (GCTCTCGGNCCNTAYAARGG), GDH1f2 (CCGGAGGCNACNGGNTAYGG), GDH1r1 (TCGTTCTGNGTNGCRCANGG) and GDH1r2 (AACCCGGCDATRTTNGCNCC) designed against conserved regions of the *gdh*1 gene were used with genomic DNA of the three species in PCR reactions. Samples of genomic DNAs were obtained as gifts: *H. inflata* was a gift from H.

van Keulen, *S. vortens* was a gift from P.J. Keeling and *Monocercomonas* sp. was a gift from M. Müller. The resulting PCR products were purified using the Qiaquick PCR Purification Kit (Qiagen) and directly sequenced using the ABI PRISM BigDye Termination Cycle Sequencing Kit (Applied Biosystems) using the primers used in the amplification as well as internal exact-match primers. cDNA clones encoding GDH-1 were retrieved from EST projects for the parabasalid *T. vaginalis*, the diplomonad *S. barkhanus* and the oomycete *P. infestans* [15]. cDNAs clones encoding both GDH-1 and GDH-2 were retrieved from the red alga *P. yezoensis* [16] and a cDNA clone encoding GDH-2 was retrieved from the EST project for the green alga *C. reinhardtii* [17]. These clones were completely sequenced by primer walking.

Assembly of GDH datasets

GDH sequences were identified using BLAST searches against a variety of databases with representatives from the four different classes of the enzyme as 'probes' (queries). This work was performed in April 2002. All published homologs were retrieved from the National Center for Biotechnology Information (NCBI) [20]. Unpublished eukaryotic sequences were retrieved from NCBI using the "other eukaryotes" BLAST service in their genomic BLAST pages. In addition, the *Dictyostelium discoideum* Genome Project database [21] and the "microbial genomes" BLAST service at NCBI were searched for homologs. Two unpublished GDH-3 sequences were retrieved. Unpublished α-proteobacterial and cyanobacterial sequences were retrieved for GDH-1 and GDH-2, in order to explore the possibility that the eukaryotic groups originated via endosymbiotic gene transfer from the mitochondria and chloroplast, respectively. All other unpublished prokaryotic GDH-1, GDH-2 and GDH-4 sequences were excluded to reduce the computational burden of the phylogenetic reconstructions. After inclusion of the newly generated GDH sequences and removal of sequences from different strains of the same species, the sizes of the datasets were 66, 73, 12 and 15 taxa for GDH-1, GDH-2, GDH-3 and GDH-4, respectively.

Phylogenetic analyses

The amino acid sequences within each dataset were aligned using CLUSTALW [22] and unambiguously aligned regions were identified and removed by visual inspection. Sequences with >85% amino acid sequence identity within the unambiguously aligned regions were excluded from the dataset to reduce the computational time. The χ^2 tests for deviation of amino acid frequencies implemented in TREE-PUZZLE, version 4.02 [23], were applied to the datasets and sequences that failed the test were excluded from further phylogenetic analysis since the currently available phylogenetic methods cannot deal with strong amino acid compositional heterogeneity in the data [24].

Protein maximum likelihood phylogenies were inferred using PMBML [25], a modified version of the of PROML within the PHYLIP package, version 3.6a2 [26]. The rationale for using PMBML rather than PROML was that the available version of PROML did not support the use of a Jones-Taylor-Thornton (JTT) substitution model at the time of analysis. A mixed four-category discrete-gamma model of among-site rate variation plus invariable sites (JTT + Γ + Inv) and 10 random additions (jumbles) with global rearrangements were used to find the optimal trees. The Γ shape parameters of the gamma distribution, α, the resulting rate categories and the fraction of invariable sites, P_{inv}, were estimated using TREE-PUZZLE, version 4.02 [23]. Protein maximum likelihood bootstrap values were calculated by analysis of 500 resampled datasets using the same parameters, except that only one round of random addition (jumble) followed by global rearrangements were performed for each replicate. Protein maximum likelihood distance bootstrap values for bipartitions were calculated by analysis of 500 resampled datasets using PUZZLEBOOT [27] with a mixed eight-category discrete-gamma model of among-site rate variation plus invariable sites (JTT + Γ + Inv).

Accession numbers

All alignments, and the complete list of accession numbers for the sequences used in the analyses, are available on request from J.O.A. (Jan.Andersson@icm.uu.se). The sequences reported here were deposited in GenBank under the accession numbers AF533881-AF533889.

Authors' contributions

JOA carried out the molecular biology studies, bioinformatic and phylogenetic analyses and drafted the manuscript. AJR provided advice on analyses and edited the manuscript. Both authors read and approved the final manuscript.

Acknowledgements

We thank T. Martin Embley, Robert P. Hirt and David S. Horner (Natural History Museum, London, UK), Nobumi Kusuhara (Kazusa DNA Research Institute, Japan) and Francine Govers (Wageningen University, the Netherlands) for generous gifts of cDNA clones. We also thank Brian Hoyt and Milan Horacek for help with computer facilities. Preliminary sequence data were obtained from the Institute for Genomic Research, the DOE Joint Genome Institute, the Stanford Genome Technology Center, the Genome Sequencing Centre Jena, and the Pennsylvania State University. A.J.R. is supported by the Canadian Institute for Advanced Research, Program in Evolutionary Biology. This work was supported by a Natural Sciences and Engineering Research Council (NSERC) Genomics Project Grant 228263-99 and a Canadian Institutes of Health Research (CIHR) Grant IG1-60759 awarded to A.J.R.

References

1. Doolittle WF: **Phylogenetic classification and the universal tree** *Science* 1999, **284:**2124-2129.
2. Ochman H, Lawrence JG and Groisman EA: **Lateral gene transfer and the nature of bacterial innovation** *Nature* 2000, **405:**299-304.
3. Nesbø CL, Boucher Y and Doolittle WF: **Defining the core of non-transferable prokaryotic genes: the euryarchaeal core** *J Mol Evol* 2001, **53:**340-350.
4. Wolf YI, Aravind L, Grishin NV and Koonin EV: **Evolution of aminoacyl-tRNA synthetases-analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events** *Genome Res* 1999, **9:**689-710.
5. Boucher Y and Doolittle WF: **The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways** *Mol Microbiol* 2000, **37:**703-716.
6. Qian Q and Keeling PJ: **Diplonemid glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and prokaryote-to-eukaryote lateral gene transfer** *Protist* 2001, **152:**193-201.
7. Striepen B, White MW, Li C, Guerini MN, Malik S-B, Logsdon Jr JM, Liu C and Abrahamsen MS: **Genetic complementation in apicomplexan parasites** *Proc Natl Acad Sci U S A* 2002, **99:**6304-6309.
8. Andersson JO and Roger AJ: **Evolutionary analyses of the small subunit of glutamate synthase: gene order conservation, gene fusions and prokaryote-to-eukaryote lateral gene transfers** *Eukaryot Cell* 2002, **1:**304-310.
9. Andersson JO, Sjögren ÅM, Davis LAM, Embley TM and Roger AJ: **Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes** *Curr Biol* 2003, **13:**94-104.
10. Katz LA: **Lateral gene transfers and the evolution of eukaryotes: theories and data** *Int J Syst Evol Microbiol* 2002, **52:**1893-1900.
11. Gogarten JP: **Gene transfer: gene swapping craze reaches eukaryotes** *Curr Biol* 2003, **13:**R53-R54.
12. Benachenhou-Lahfa N, Forterre P and Labedan B: **Evolution of glutamate dehydrogenase genes: evidence for two paralogous protein families and unusual branching patterns of the archaebacteria in the universal tree of life** *J Mol Evol* 1993, **36:**335-346.
13. Brown JR and Doolittle WF: **Archaea and the prokaryote-to-eukaryote transition** *Microbiol Mol Biol Rev* 1997, **61:**456-502.
14. Minambres B, Olivera ER, Jensen RA and Luengo JM: **A new class of glutamate dehydrogenases (GDH). Biochemical and genetic characterization of the first member, the AMP-requiring NAD-specific GDH of *Streptomyces clavuligerus*** *J Biol Chem* 2000, **275:**39529-39542.
15. Kamoun S, Hraber P, Sobral B, Nuss D and Govers F: **Initial assessment of gene diversity for the oomycete pathogen *Phytophthora infestans* based on expressed sequences** *Fungal Genet Biol* 1999, **28:**94-106.
16. Nikaido I, Asamizu E, Nakajima M, Nakamura Y, Saga N and Tabata S: **Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, *Porphyra yezoensis*** *DNA Res* 2000, **7:**223-227.
17. Asamizu E, Miura K, Kucho K, Inoue Y, Fukuzawa H, Ohyama K, Nakamura Y and Tabata S: **Generation of expressed sequence tags from low-CO_2 and high-CO_2 adapted cells of *Chlamydomonas reinhardtii*** *DNA Res* 2000, **7:**305-307.
18. Baldauf SL, Roger AJ, Wenk-Siefert I and Doolittle WF: **A kingdom-level phylogeny of eukaryotes based on combined protein data** *Science* 2000, **290:**972-977.
19. Andersson JO, Doolittle WF and Nesbø CL: **Are there bugs in our genome?** *Science* 2001, **292:**1848-1850.
20. **National Center for Biotechnology Information** [http://www.ncbi.nlm.nih.gov/]
21. ***Dictyostelium discoideum* Genome Project** [http://www.uni-koeln.de/dictyostelium/]
22. Thompson JD, Higgins DG and 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.
23. Strimmer K and von Haeseler A: **Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies** *Mol Biol Evol* 1996, **13:**964-969.
24. Foster PG and Hickey DA: **Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions** *J Mol Evol* 1999, **48:**284-290.
25. Veerassamy S, Smith A and Tillier ERM: **A transition probability model for amino acid substitutions from BLOCKS** *J Comput Biol* .
26. Felsenstein J: **Phylogeny Inference Package (Version 3.2)** *Cladistics* 1989, **5:**166.
27. **Roger Laboratory Home Page** [http://hades.biochem.dal.ca/Rogerlab/]

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

BioMedcentral