

RESEARCH ARTICLE

Open Access

Testis-specific glyceraldehyde-3-phosphate dehydrogenase: origin and evolution

Mikhail L Kuravsky¹, Vladimir V Aleshin^{2,3}, Dmitrij Frishman^{4*} and Vladimir I Muronetz^{1,2*}

Abstract

Background: Glyceraldehyde-3-phosphate dehydrogenase (GAPD) catalyses one of the glycolytic reactions and is also involved in a number of non-glycolytic processes, such as endocytosis, DNA excision repair, and induction of apoptosis. Mammals are known to possess two homologous GAPD isoenzymes: GAPD-1, a well-studied protein found in all somatic cells, and GAPD-2, which is expressed solely in testis. GAPD-2 supplies energy required for the movement of spermatozoa and is tightly bound to the sperm tail cytoskeleton by the additional N-terminal proline-rich domain absent in GAPD-1. In this study we investigate the evolutionary history of GAPD and gain some insights into specialization of GAPD-2 as a testis-specific protein.

Results: A dataset of GAPD sequences was assembled from public databases and used for phylogeny reconstruction by means of the Bayesian method. Since resolution in some clades of the obtained tree was too low, syntenic analysis was carried out to define the evolutionary history of GAPD more precisely. The performed selection tests showed that selective pressure varies across lineages and isoenzymes, as well as across different regions of the same sequences.

Conclusions: The obtained results suggest that GAPD-1 and GAPD-2 emerged after duplication during the early evolution of chordates. GAPD-2 was subsequently lost by most lineages except lizards, mammals, as well as cartilaginous and bony fishes. In reptilians and mammals, GAPD-2 specialized to a testis-specific protein and acquired the novel N-terminal proline-rich domain anchoring the protein in the sperm tail cytoskeleton. This domain is likely to have originated by exonization of a microsatellite genomic region. Recognition of the proline-rich domain by cytoskeletal proteins seems to be unspecific. Besides testis, GAPD-2 of lizards was also found in some regenerating tissues, but it lacks the proline-rich domain due to tissue-specific alternative splicing.

Background

Glyceraldehyde-3-phosphate dehydrogenase (GAPD, EC 1.2.1.12) is a homotetrameric glycolytic enzyme providing phosphorylation of 3-phosphoglyceraldehyde to 1,3-diphosphoglycerate coupled with reduction of NAD⁺ to NADH. Mammals are known to possess two tissue-specific GAPD isoenzymes: somatic (GAPD-1) and testis-specific (GAPD-2, GAPDS). For *Homo sapiens*, their protein sequences are 68% identical. Besides the two isoenzymes, a vast amount of GAPD pseudogenes was found in the genomes of primates and rodents [1,2].

Mammalian GAPD-1 is a well-studied protein, a high concentration of which in cells (5-15% of all cytoplasmic proteins) confirms its functional significance. Recent studies established that GAPD-1 is not simply a classical metabolic protein involved in glycolytic energy production, but rather a multifunctional protein with specific functions in numerous processes [3,4]. GAPD-1 was shown to display both cytosolic and nuclear localization participating in endocytosis [5-7], plasma membrane fusion [8], microtubule assembly [9,10], secretory vesicular transport [11,12], protein phosphotransferase/kinase reactions [13,14], translational and transcriptional controls of gene expression [15-17], regulation of telomere structure [18,19], nuclear membrane fusion [20], nuclear RNA transport [21], DNA excision-repair [22,23] and induction of apoptosis in case of oxidative stress [24-27]. Furthermore, GAPD-1 was implicated in

* Correspondence: d.frishman@wzw.tum.de; vimuronets@belozersky.msu.ru
¹Faculty of Bioengineering and Bioinformatics, M.V. Lomonosov Moscow State University, Leninskiye Gory 1-73, Moscow 119991, Russian Federation
⁴Department of Genome Oriented Bioinformatics, Technische Universität München, Wissenschaftszentrum Weißenstephan, Freising 85354, Germany
Full list of author information is available at the end of the article

Alzheimer's [28-30] and Huntington's [30-32] neurodegenerative diseases.

As opposed to soluble GAPD-1, mammalian GAPD-2 is tightly attached to the cytoskeleton, namely to the principal piece of the spermatic filament fibrous sheath [33-35]. The attachment is mediated by an additional N-terminal proline-rich domain of 74 amino acids [35,36]. GAPD-2 supplies the dynein ATPases of filament with energy, therefore playing a crucial role in the maintaining of sperm motility. Disruption of its expression generally leads to infertility [37]. Due to its strong association with cytoskeleton GAPD-2 remains within the insoluble fraction after cell breaking, significantly complicating its experimental investigation. As a result, there is only little data on GAPD-2 properties. It was recently discovered to display enhanced stability towards denaturation that may be an adaptation to the absence of protein expression in spermatozoa. Enzyme kinetics exhibited by GAPD-2 was found to differ from the one exhibited by GAPD-1 too [38]. Based on the study of short functional motives of both mammalian isoenzymes, GAPD-2 was proposed to evade involvement in most non-glycolytic processes characteristic for GAPD-1 [39].

GAPD-1 and GAPD-2 are also possessed by some other vertebrates besides Mammalia [40-42], but their expression is apparently not always tissue-specific. In the bony fish *Oplegnathus fasciatus* both GAPD mRNAs were detected ubiquitously in all of tissues examined [40], and therefore the functional specificities of the isoenzymes seem to differ from the mammalian ones. Based on the phylogenetic trees, it was hypothesized that GAPD could diverge to the isoenzymes around the origin of Bilateria, but as only vertebrates have retained GAPD-2, this scenario seems unlikely. However, some vertebrates (e.g. *Xenopus laevis*) were discovered to lack GAPD-2 [42].

Single copy genes are thought to evolve conservatively because of strong negative selective pressure. Gene duplications produce a redundant gene copy and thus release one or both copies from negative selective pressure. Thus, duplications should be an important precursor of functional divergence. The increased availability of sequences in the public databases allows the investigation of the molecular evolution of the GAPD gene family and the evaluation of selection following duplication events. In the present study we focus on the evolution of the poorly uninvestigated GAPD-2 isoenzyme. Previously GAPD-2 was discovered to be specific for vertebrates [42]. Therefore we will focus on this taxon as well as on the other groups of deuterostomes not considered in [42]. Specifically, we (1) examine the evolutionary history of GAPD-2 and other GAPD isoenzymes of deuterostomes, (2) evaluate lineage-specific

changes in selective pressure affecting GAPD isoenzymes, and (3) look into the metamorphosis of GAPD-2 to a testis-specific protein.

Results

Sequences of GAPD family members

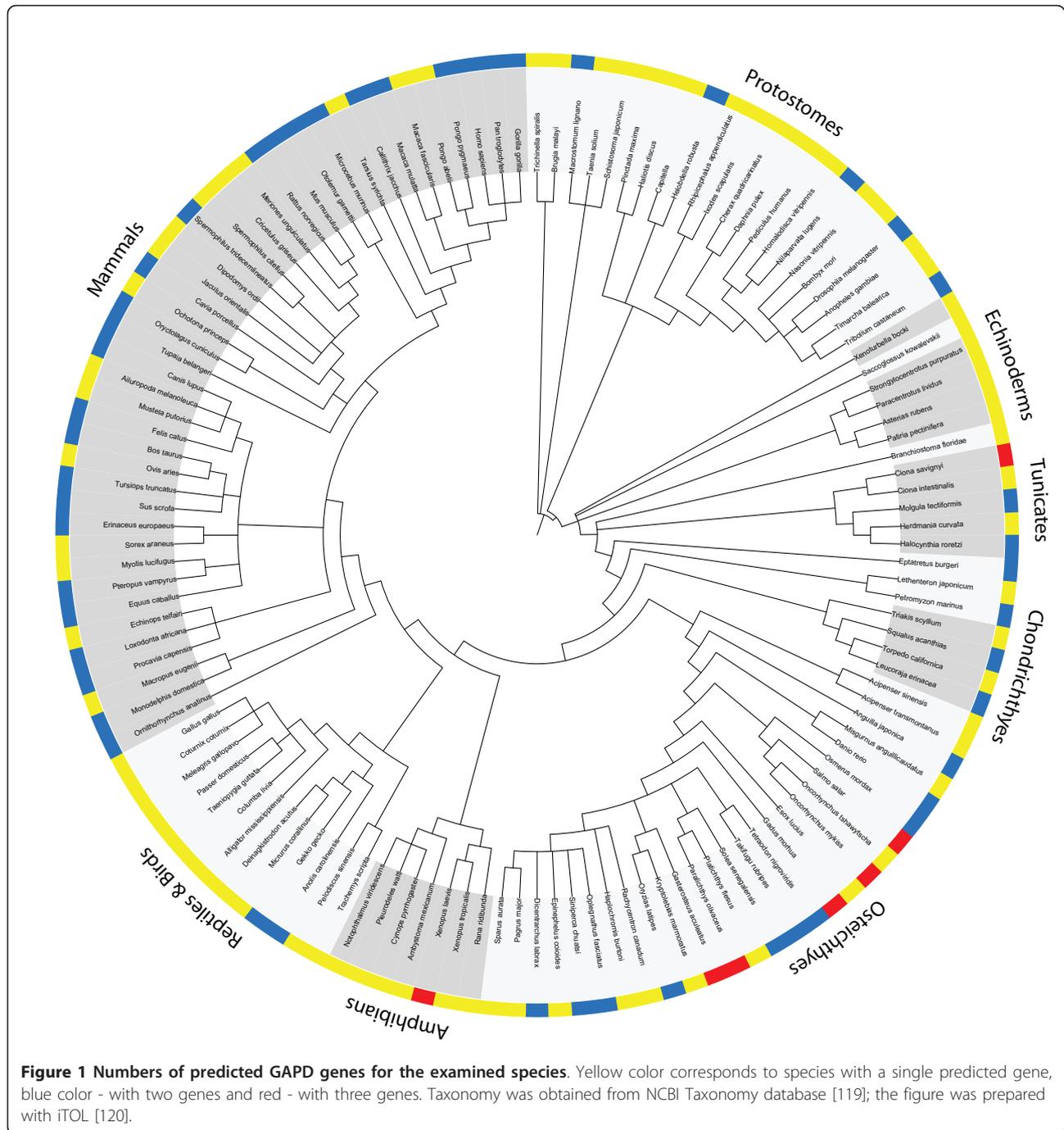
The numbers of discovered GAPD family members for all examined species are represented at Figure 1. Mammalian GAPD sequences were extracted from the Ensembl database. For most species (19 of the 25 examined) two different sequences were obtained. One of these sequences always contained an additional proline-rich domain at the N-terminus, as observed in the human GAPD-2. A single GAPD sequence was obtained for each of the 6 remaining mammalian species, either with or without the proline-rich domain. The lack of the second sequence seems to be due to incompleteness of genomes.

GAPD sequences of teleosts were obtained using both Ensembl (5 species present in this database) and BLAST searches against RefSeq transcripts and the EST division of GenBank (species not covered by Ensembl). Three different sequences were discovered for 4 species, two sequences - for 6 species and a single sequence - for 3 species. The differences between the numbers of obtained GAPD sequences are not necessarily a result of data incompleteness and may be biologically relevant. For example, only two sequences were identified within a complete genome of zebrafish.

Identification of GAPD sequences of all other species was performed by conducting BLAST searches against RefSeq transcripts and the EST division of GenBank. Two different GAPD sequences were discovered for lizards, some cartilaginous fishes, some jawless vertebrates, some tunicates and a few non-deuterostomes (3 of 10 insects, a leech and a flatworm). Single GAPD sequences were discovered for all examined birds, reptiles except lizards, amphibians, lancelet, echinoderms, acorn worm, *Xenoturbella bocki*, as well as for the remaining cartilaginous fishes, jawless vertebrates, tunicates and most examined non-deuterostomes. Two species (*Xenopus laevis* and *Ciona savignyi*) were revealed to possess even three, but slightly different GAPD family members.

Tissue-specific translation of the proline-rich domain in lizard GAPD

Besides mammalian GAPD-2, proline-rich domains were detected only in one of the GAPD isoenzymes of lizard species: *Anolis carolinensis* and *Gekko gekko*. ESTs of *A. carolinensis* encoding this isoenzyme originated from testis [GenBank:FG786985, GenBank:FG793471, GenBank:FG801901, GenBank:FG802958], regenerating tail [GenBank:FG771974, GenBank:FG779496] and the whole embryo [GenBank:FG720854]. It is remarkable that ESTs



from regenerating tail and embryo lack a fragment of 103 nucleotides (shortened variant), which is present in ESTs from testis (full-length variant; see Figure 2 and additional file 1: Alignment of the two forms of *Anolis carolinensis* GAPD-2 mRNA). This fragment is situated near the 5'-terminus and encodes the beginning of the proline-rich domain including an ATG start codon. The next possible start codon, which is present in both EST

variants, is located right after the proline-rich domain. So the protein with the proline-rich domain should be translated only from the full-length EST variant. Translation of the shortened EST variant should begin from the second start codon such that the product will not possess the proline-rich domain.

Availability of the two EST variants must be a result of tissue-specific alternative splicing: the exon of 103

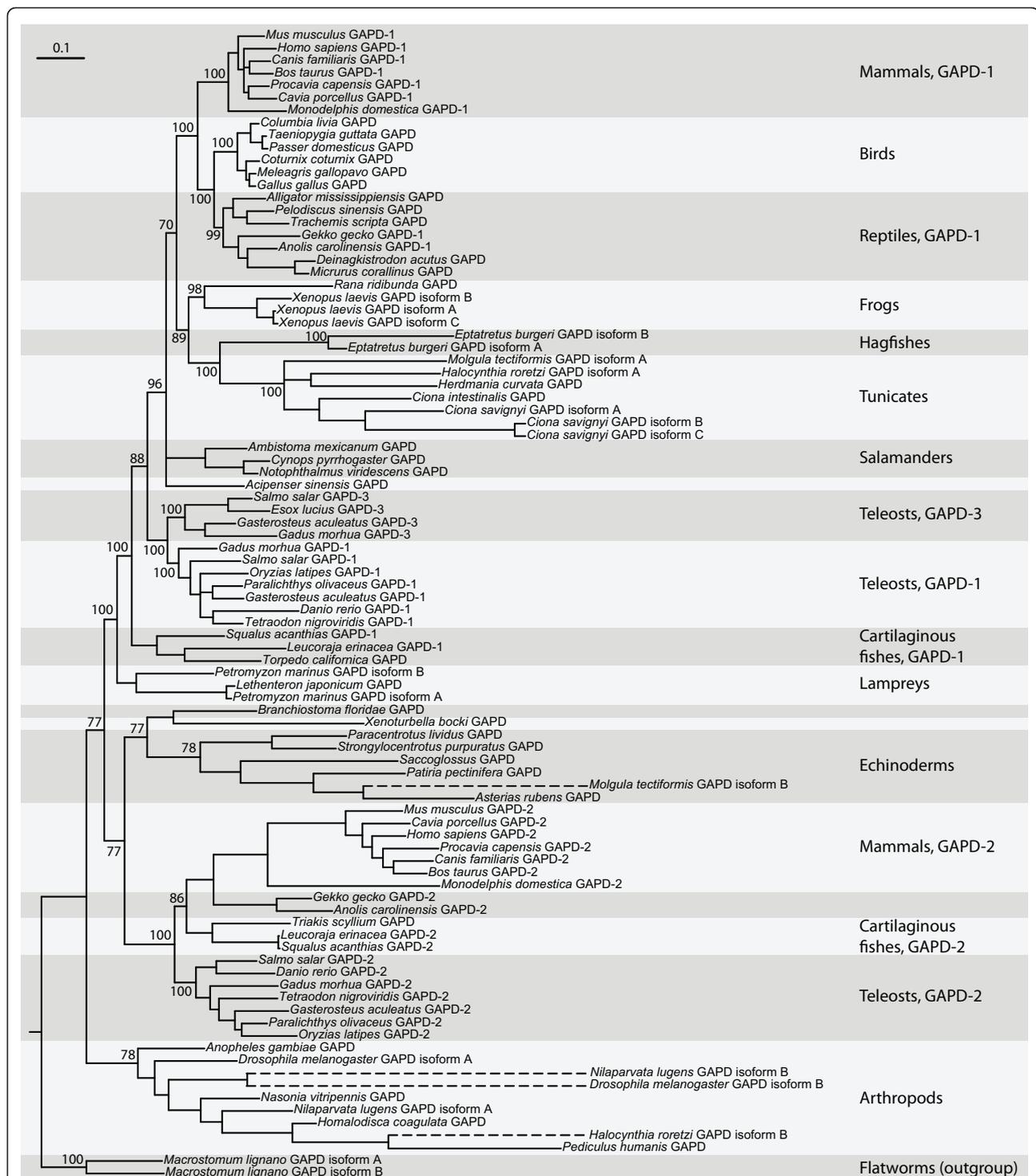


Figure 3 Phylogenetic tree of 92 GAPD isoenzymes. Phylogenetic tree constructed on nucleotide sequences using the Bayesian algorithm. Numbers at nodes are the obtained posterior probabilities. Discontinuous lines mark the branches of enormous high length (more than 0.75), which can correspond to pseudogenes or contaminated samples. The tree does not fit the common knowledge about the evolution in details; nevertheless it provides some useful information. For more accurate definition of GAPD evolution, syntenic analysis was also used.

the group including mammalian GAPD-1 and the group including mammalian GAPD-2. GAPD of insects separates before these two groups diverge, which means that the duplication into GAPD-1 and GAPD-2 took place after the divergence of protostomes and deuterostomes. The orthologs of mammalian GAPD-1 and GAPD-2 are further referred to as GAPD-1 and GAPD-2, correspondingly.

The clade including mammalian GAPD-1 is supported by a high posterior probability (100%). Inside this clade a number of additional duplications were detected. One of them apparently happened near the origin of teleosts and produced a third GAPD isoenzyme hereinafter referred to as GAPD-3. Other independent duplications produced additional GAPD isoenzymes in lamprey, hagfish, sea squirt and *Xenopus laevis*.

The clade including mammalian GAPD-2 is based on a less robust branch with a posterior probability of 77%. It splits into a clade of vertebrates (100% posterior probability) and a clade including the only GAPD isoenzymes of echinoderms, lancelet, hemichordates and *Xenoturbella bocki*, as well as the second GAPD isoenzyme of some tunicates (77% posterior probability). On account of lower support value, merging of these two clades into one is questionable and needs confirmation.

The syntenic analysis showed that GAPD family members of the examined species can be linked to either of two loci: the locus syntenic to human GAPD-1 contains GAPD-1 of zebrafish, GAPD-1 and GAPD-3 of stickleback, the only GAPD of lancelet and the only GAPD of sea squirt; the locus syntenic to human GAPD-2 contains GAPD-2 of both zebrafish and stickleback (Figure 4). The similarity between gene layouts within both loci is rather low, multiple genome micro-rearrangement events such as deletions and inversions were detected. The surroundings of GAPD genes in sea urchin and acorn worm genomes do not contain any common genes with both each other and the revealed two syntenic loci. The genes in these surroundings do not form any clusters in the genomes of other examined species as well. This can be accounted for distant relationships between the species.

BLAST searches of genes which are syntenic to human and fish GAPD-2 were carried out in the genomes of lancelet, sea squirt, sea urchin and acorn worm. They showed that these genes are dispersed in the genomes rather than combined together in a single locus.

The constructed synteny maps provide support for orthology between GAPD-1 of human, either GAPD-1 or GAPD-3 of stickleback, GAPD of lancelet and sea squirt, as well as between GAPD-2 of human and both fishes. These results generally agree with the phylogenetic trees, indicating orthology between appropriate isoenzymes of human and fishes. Syntenic analysis

helped to identify the origin of lancelet and sea squirt GAPDs, which was not determined with confidence by phylogenetic trees construction because of low branch support values. The evidence is also given for the origination of GAPD-1 and GAPD-3 of stickleback and probably some other bony fishes as a result of teleost-specific whole genome duplication.

Selective pressure estimation

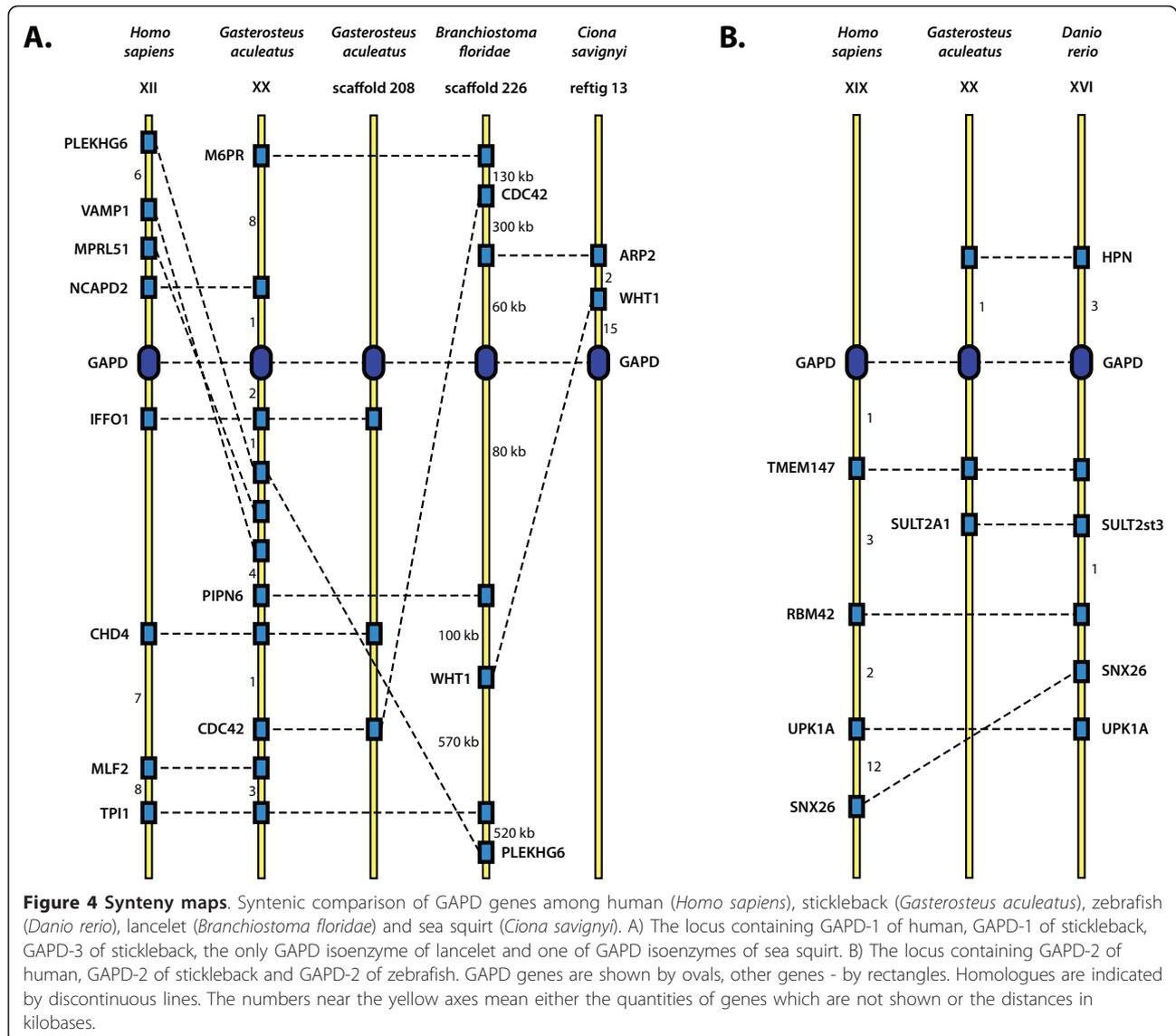
K_a/K_s profiles were compared in four clades: mammalian GAPD-1 and GAPD-2, teleost GAPD-1 and GAPD-2, while GAPD of insects was used as an outgroup. To avoid saturation in synonymous substitutions which can significantly affect the results, pairs of closely related sequences were considered (Table 1).

Results of K_a/K_s profile calculation show that selective pressure varies for different regions of GAPD sequences (Figure 5). Most regions of all examined sequences are suggested to be under strong purifying selection ($K_a/K_s < 0.1$). However, a part of the proline-rich domain of mammalian GAPD-2 is not restrained by purifying selection with K_a/K_s up to 1.1.

In mammalian GAPD-1 and GAPD-2, teleost GAPD-2 and insect GAPD, the purifying selection is impaired approximately between the 85th and 105th positions of protein sequences (Section 1; from here on the numbering of amino acid positions as in mammalian GAPD-1). In mammalian GAPD-2 purifying selection is also weakened between the 265th and 285th positions (Section 2). In teleost GAPD-1 purifying selection is weakened between the 55th and 75th positions (Section 3).

The regions under impaired purifying selection were mapped on the 3D-structure of human GAPD-1 (PDB ID 1u8f). Section 1 corresponds to a buried β -strand and adjacent loops near the NAD-binding site. Sections 2 and 3 are solvent-exposed regions of the polypeptide chain also composed of both β -strands and loops.

Selective pressure affecting GAPD family members was also investigated by means of branch-specific models as implemented in PAML. Six datasets were examined: mammalian GAPD-1 (17 sequences) and GAPD-2 (12 sequences), teleost GAPD-1 (10 sequences), GAPD-2 (7 sequences) and GAPD-3 (4 sequences) as well as insect GAPD (8 sequences). To determine whether the selective constrains vary for different isoenzymes and lineages, two models were compared: one-ratio (R1) and six-ratios (R6). R1 assumed constant K_a/K_s ratio for all examined GAPD datasets, whereas R6 assumed different ratios for each dataset. The obtained K_a/K_s ratios and the likelihoods of the models are represented in Table 2. The likelihood ratio test (LRT) indicated a significant difference between the likelihoods of R1 and R6 ($2d = 148.57$, $df = 5$, p -value = 0.00), implying variation of selective constrains at least for some datasets.



Following the results obtained for R6 model, the K_a/K_s ratios of mammalian GAPD-2 and insect GAPD differ from the mean value above all (Figure 6). Therefore the hypotheses stating that the selective constrains differ between these two and the other datasets were tested. Three models were compared with R6: R2m model assuming constant K_a/K_s ratio for all datasets except

mammalian GAPD-2, R2i model assuming constant K_a/K_s ratio for all datasets except insect GAPD and R3 model assuming constant K_a/K_s ratio for all datasets except both mammalian GAPD-2 and insect GAPD (see Table 2 for the obtained ω -values and likelihoods). LRT revealed that the likelihoods of R3 and R6 are not significantly different ($2d = 6.73$, $df = 3$, p -value = 0.08), while

Table 1 Pairs of sequences used for K_a/K_s calculation

Taxon	Species	Isoenzyme	Sequence identity, %
Mammals	<i>Homo sapiens</i> - <i>Microcebus murinus</i>	GAPD-1	91
	<i>Homo sapiens</i> - <i>Callitrix jacchus</i>	GAPD-2	94
Teleosts	<i>Tetraodon nigroviridis</i> - <i>Takifugu rubripes</i>	GAPD-1	94
	<i>Tetraodon nigroviridis</i> - <i>Takifugu rubripes</i>	GAPD-2	92
Insects	<i>Drosophila ananassae</i> - <i>Drosophila virilis</i>	GAPD	87

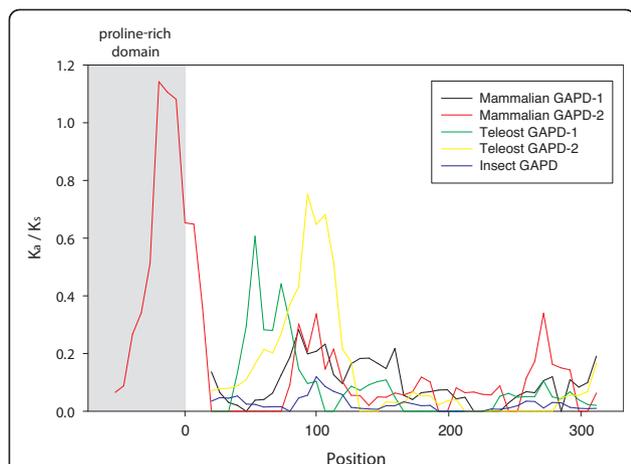


Figure 5 K_a/K_s profiles. Five GAPD isoenzymes are compared: GAPD-1 and GAPD-2 of mammals, GAPD-1 and GAPD-2 of teleosts and GAPD of insects. Numbering of the X axis starts immediately after the proline-rich domain of mammalian GAPD-2 and corresponds to the positions in protein sequences.

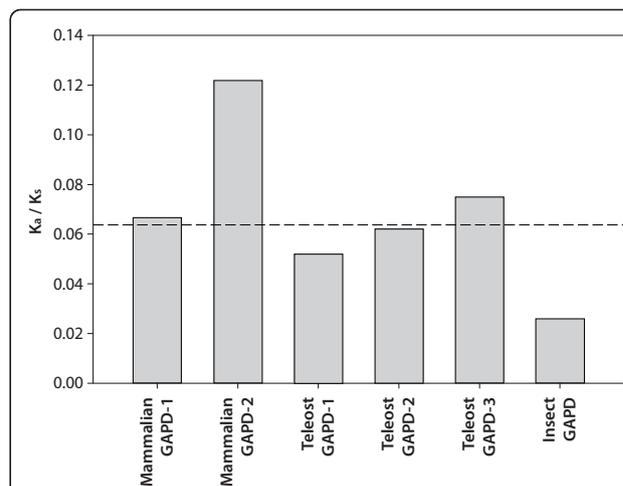


Figure 6 K_a/K_s values obtained with the aid of branch-specific models. The shown K_a/K_s values were calculated using the six-ratio (R6) model, which implies different selection constrains for GAPD-1 and GAPD-2 of mammals, GAPD-1, GAPD-2 and GAPD-3 of teleosts, as well as GAPD of insects. The values for all isoenzymes except mammalian GAPD-2 and insect GAPD were found not to differ significantly. Discontinuous horizontal line is the K_a/K_s value obtained using the one-ratio (R1) model, which implies the same selection constrains for all isoenzymes.

the likelihoods of both R2m and R2i are significantly lower ($2d = 64.84$, $df = 4$, $p\text{-value} = 0$ and $2d = 60.5$, $df = 4$, $p\text{-value} = 0$, respectively). It means that the selective constrains are more or less similar for all three teleost GAPD isoenzymes and mammalian GAPD-1, greater for insect GAPD and weaker for mammalian GAPD-2.

Discussion

Evolutionary relationships between GAPD isoenzymes

In this study we sought to expand the previous phylogenetic investigations of GAPD [42-50] by concentrating on deuterostomes. As compared to the study in reference [42], which is also focused on deuterostomes, we introduced a number of new sequences especially from non-mammalian and non-teleost species and carried out the syntenic analysis. This allowed more accurate

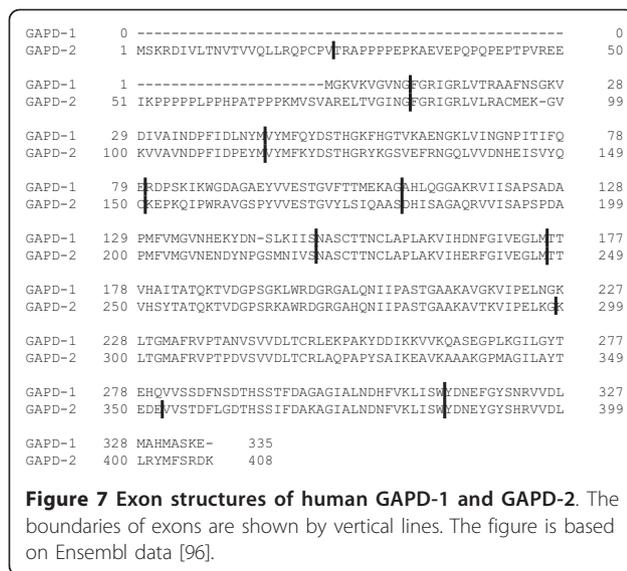
Table 2 The K_a/K_s ratio estimates for GAPD isoenzymes under various branch-specific models

Model	K_a/K_s ratio	Log-likelihood
R1	0.06369	-22221.41
R2m	0.05363 (all except mammalian GAPD-2) 0.12179 (mammalian GAPD-2)	-22179.54
R2i	0.07405 (all except insect GAPD) 0.02642 (insect GAPD)	-22177.37
R3	0.06263 (all except listed below) 0.12186 (mammalian GAPD-2) 0.02567 (insect GAPD)	-22150.49
R6	0.06672 (mammalian GAPD-1) 0.05218 (fish GAPD-1) 0.07492 (fish GAPD-3) 0.12187 (mammalian GAPD-2) 0.06218 (fish GAPD-2) 0.02593 (insect GAPD)	-22147.12

determination of phylogeny, as well as the identification of some novel GAPD isoenzymes, for example the third isoenzyme of teleosts.

The constructed phylogenetic trees provide evidence for duplication in the early evolution of chordates which gave rise to GAPD-1 and GAPD-2 isoenzymes. It presumably took place even before the first whole-genome duplication of vertebrates [51-53]. The loci of GAPD-1 and GAPD-2 were found not to be syntenic to each other. It can be explained either by a single-gene duplication, which produced a copy of the ancestral GAPD gene, or by loss of synteny after a duplication of longer genome segment. However, the emergence of GAPD-1 and GAPD-2 is surely not a result of a retroposition, as it was concluded in early studies [54,55], documented by similar exon structures of the isoenzymes (Figure 7). It should be noted that GAPD is one of the few glycolytic enzymes that did not acquire any additional isoenzymes during the vertebrate-specific whole-genome duplication events; neither did phosphoglucose isomerase, triosephosphate isomerase and phosphoglycerate kinase. The other glycolytic enzymes gained from one to three extra copies that evolved to the tissue-specific proteins [42,56-61].

GAPD-2 was lost in most lineages and retained only by mammals, lizards, teleosts and cartilaginous fishes. The presence of both isoenzymes in these organisms raises the question of a functional difference between them. It is assumed that if two isoenzymes perform the



same function in the same set of tissues, one of them is free from functional constraints and its gene will eventually turn into a non-functional pseudogene or will be deleted [62-64]. In mammals and lizards GAPD-1 and GAPD-2 specialized to tissue-specific proteins and this is probably the reason why one of them avoided the lost. Generally, specialization towards tissue-specificity is a trend among glycolytic enzymes that have acquired additional copies. In vertebrates, they usually have distinctive isoenzymes in liver, muscle and brain, sometimes in erythrocytes and other tissues [42,60]. The situation with GAPD of teleosts and cartilaginous fishes is more complex. According to EST data, GAPD-1 and GAPD-2 of fishes are expressed in the same tissues. The results of branch-specific tests indicate that the evolutionary rates of both isoenzymes are accelerated as compared to the ancestral GAPD (GAPD of insects, which separated before the emergence of GAPD-1 and GAPD-2, was considered to evolve with the similar rate as the ancestral protein). This is in line with the model of gene duplications proposed by Hughes [65,66]. It suggests that the original gene was performing two or more functions. After duplication each copy specialized on performing a part of them. GAPD is known to be a multifunctional protein participating in many processes beyond glycolysis. As the catalytic center is conserved in both isoenzymes, GAPD-1 and GAPD-2 of teleosts and cartilaginous fishes may specialize on performing different non-glycolytic functions, as also evidenced by K_a/K_s profiling. Different regions of teleost GAPD-1 and GAPD-2 are under impaired purifying selection. These regions can correspond to the parts of proteins which are responsible for performing isoenzyme-specific non-glycolytic functions.

A number of additional duplications of GAPD genes occurred independently in certain lineages. For example, some teleosts possess the third GAPD isoenzyme (GAPD-3) in addition to GAPD-1 and GAPD-2. Taking into account both the constructed phylogenetic trees and the obtained data on syntenies, it can be concluded that GAPD-3 originated from GAPD-1 during the teleost-specific whole-genome duplication [67]. However, GAPD-3 was not found in complete genomes of zebrafish, tetraodon and fugu, which means that it was lost.

The retention of GAPD-3 by certain species of teleosts agrees with the model of dosage balance proposed by Papp with colleagues [65,68]. It states that genes having optimal dosages that are dependent on each other may be lost only synchronously after whole-genome duplications. Therefore they are preferentially kept. In the study by [42] most of the other glycolytic enzymes were shown to have extra copies in teleosts, which also originated during the whole-genome duplication. Therefore, GAPD-3 as well as the other additional glycolytic isoenzymes in teleosts may be retained to prevent dosage imbalance leading to glycolysis malfunction.

The model of dosage balance also provides an explanation for *Xenopus laevis* possessing three slightly different GAPD isoenzymes [Swiss-Prot:P51469, GenBank:BC043972, GenBank:BC048770]. Following the results of phylogenetic analysis (Figure 3), the duplications of GAPD-1 gene giving rise to these isoenzymes seem to have taken place after the divergence of *Xenopus* and *Rana* genera of frogs. *X. laevis* is known to have undergone a whole-genome duplication event about 40 million years ago [69,70] and most of its genes have two copies [71]. Furthermore, the GAPD genes in reference [GenBank:BC043972, GenBank:BC048770] might be the allelic variants of a single gene since they are 99% identical and their evidence is only at transcript level. If so, *X. laevis* would have only two GAPD genes, in line with the dosage balance model.

Sea squirt *Ciona savignyi* was discovered to possess three different GAPD isoenzymes as well. All of them seem to have originated from GAPD-1 after the emergence of tunicates (Figure 3). To check whether these isoenzymes are encoded by distinct genes or allelic variants of a single gene, we turned to the *C. savignyi* genome assembly version 2.0 (Broad Institute) with removed redundant alleles [72] available via Ensembl. There were only two GAPD genes [Ensembl:ENSCSAVG00000004357, Ensembl:ENSCSAVG00000007442] corresponding to the two isoenzymes. The remaining isoenzyme was 97% identical to one of the others. Perhaps, it is nothing but an allelic variant since *C. savignyi* displays extremely high allelic polymorphism [73].

It looks like the duplication giving rise to the GAPD-1 copies in *C. savignyi* is advantageous itself by increasing

GAPD dosage. Otherwise fixation of even two consecutive duplications seems to be unlikely. The model of gene duplications assuming beneficial increase in gene dosage has been extensively studied and shown to be applicable in a number of cases [74-76]. The duplications of GAPD in the considered species may be explained by the emerged necessity of enhancing of some non-glycolytic functions of GAPD, as it is hard to imagine that such a conserved process as glycolysis needs an increase of a dose of one of its enzymes.

GAPD-2 specialization to a testis-specific protein

Mammalian GAPD-2 is known to be a highly specialized isoenzyme, which is present solely in testis (microarray data are available in the ArrayExpress database at <http://www.ebi.ac.uk/arrayexpress> under accession numbers E-GEOD-7307, E-GEOD-3526, E-TABM-969 and E-GEOD-2361) [77-79]. We have found that GAPD-2 is expressed in a testis-specific way also by two lizard species. Lizards are also the only lineage besides mammals in which GAPD-2 possesses the proline-rich domain. Taking into account that this domain serves as an anchor to spermatid filament cytoskeleton, correlation between its presence and testis-specific expression seems to be evident.

As GAPD-2 is a testis-specific protein only in mammals and lizards, it is likely to have specialized in this way during the early evolution of amniotes. However, birds have completely lost GAPD-2. We could not detect it in any of the examined bird species including *Gallus gallus* and *Taeniopygia guttata* with complete genomes. So, the same GAPD isoenzyme should act in both somatic tissues and testis. It remains unclear what changes in bird spermatozoa rendered testis-specific GAPD-2 unnecessary.

GAPD-2 is not the only testis-specific glycolytic isoenzyme. There are also testis-specific isoenzymes of phosphoglycerate kinase (PGK-2) [80,81] and lactate dehydrogenase (LDHC) [82-84]. It is remarkable that both are possessed only by mammals, thus resembling GAPD-2. PGK-2 originated from PGK-1 isoenzyme by retrotransposition [85,86], while LDHC stems from the LDHA isoenzyme [60]. These events are supposed to have taken place during the early evolution of mammals. Perhaps, the gain of three testis-specific glycolytic isoenzymes is a consequence of an alteration of spermatozoa structure. Mammalian spermatozoa are known to have a relatively long and thin tail, complicating ATP diffusion from mitochondria along it [87]. Therefore, energy is generated mostly by glycolytic enzymes located in the tail cytoplasm [34,88]. Such reorganization of metabolism may require special isoenzymes with distinctive catalytic properties.

As mentioned before, a unique feature of testis-specific GAPD-2 is the additional N-terminal proline-rich domain, which is absent in all other GAPD isoenzymes. Moreover, there are no additional fragments in PGK-2 and LDHC. The spatial structure of the proline-rich domain is still unsolved. We have found that for the majority of mammals it is encoded by two exons. The first exon encodes a conservative segment of 22 amino acids. The second exon encodes a segment with a high content of proline residues, highly variable in both length (58-97 amino acids) and composition (see additional file 2: The proline-rich N-terminal domains of mammalian GAPD-2). The layout of proline residues has a strikingly repetitive character. They form P_n and $(XP)_n$ motifs, where X is any amino acid (often cysteine, glutamic acid or glutamine). Generally, polyproline repetitive motifs are known to participate in strong but unspecific protein-protein interactions [89]. Apparently they play the same role in the proline-rich domain of GAPD-2 mediating the binding to spermatid filament cytoskeleton. The presence of two different kinds of polyproline motifs suggests GAPD-2 being bound to more than one protein of cytoskeleton.

An evidence for unspecific proline-rich domain recognition by cytoskeletal proteins is also furnished by the results of K_a/K_s calculation. K_a/K_s value estimated for the variable segment of the proline-rich domain of mammalian GAPD-2 was close to unity, which means that this domain is subjected to neither purifying nor positive selection and therefore its specific sequence is not important for functioning.

The proline-rich domain is likely to be relatively young since it is absent in all other GAPD isoenzymes and no similar sequences have been revealed in other proteins by means of BLAST searches. So-called exonization of non-coding sequences is now assumed to be the source of new protein domains [90-93]. The repetitive character of the proline-rich domain sequence implies that it could have emerged from a microsatellite region. This way of new domain origination was proposed to be a general mechanism for the repetitive protein sequences [90,94,95].

Tissue-specific alternative splicing was discovered to govern the presence of proline-rich domain in GAPD-2 of a lizard *Anolis carolinensis*: it depends on a cassette exon being either spliced or retained. Unfortunately, no conclusion can be made as to whether this mechanism preceded GAPD-2 specialization to a testis-specific protein or appeared after it. It may be that GAPD-2 first incorporated the proline-rich domain as a rare optional splice variant in some tissues and only then specialized towards testis-specificity.

Conclusions

The results of our study substantially expand the current knowledge on evolution of GAPD family members. We show that GAPD-1 and GAPD-2 isoenzymes of mammals are also present in other lineages. We speculate that they emerged after duplication of the ancestral GAPD gene during the early evolution of chordates. GAPD-1 then underwent a number of additional independent duplications in different species, while GAPD-2 was lost in most lineages and is now found only in mammals and lizards, as well as cartilaginous and bony fishes.

We have demonstrated that GAPD-2 of mammals and lizards is specialized to a testis-specific protein. Accordingly, in these lineages GAPD-2 has acquired the novel N-terminal proline-rich domain anchoring the protein to the sperm tail cytoskeleton. This domain is likely to have originated by exonization of a microsatellite genomic region in a common ancestor of amniots. Estimates of selective pressure suggest unspecific recognition of the proline-rich domain by cytoskeletal proteins. Besides testis, GAPD-2 of lizards was also found in some regenerating tissues, but lacking the proline-rich domain due to tissue-specific alternative splicing.

Methods

Sequence data

In the previous study [42], GAPD-2 was shown to be specific for vertebrates. Therefore we decided to limit the consideration of GAPD isoenzymes and focused only on those belonging to vertebrates and also to the other groups of deuterostomes since they were not examined in [42]. In order to find all GAPD sequences of deuterostomes, we first turned to the Ensembl database [96]. 69 sequences of mammals and bony fishes were obtained from it as belonging to glyceraldehyde-3-phosphate dehydrogenase protein family [Ensembl: ENSFM0025000000211]. Second, a PSI-BLAST [97] search using the human GAPD-1 [SwissProt:P04406] as query (which was selected to be a typical example of GAPD) was conducted against UniProt [98]. Since GAPD is known to be a well-conserved protein, a strict e-value threshold of 10^{-6} was chosen. The search converged in 6 steps returning 8957 hits, all of which showed more than 30% of identity to the query sequence. All in all 60 sequences of deuterostomes were picked out (excluding fragments and those previously obtained from the Ensembl database). We also selectively picked out 13 sequences of the major protostome phyla (arthropods, mollusks, annelid worms, roundworms and flatworms). Third, additional 55 sequences were obtained by employing TBLASTN algorithm with default parameters [97] to search with human GAPD-1

[SwissProt:P04406] and GAPD-2 [SwissProt:O14556] as queries in the EST division of GenBank [99]. EST hits, which usually represent fragments of complete mRNAs, were manually scanned for extensive overlapping regions and then joined into larger sequences. Further inspection revealed some cases of contamination, which were excluded from the analysis. Specifically, we identified a chicken EST [GenBank:AM067846] actually belonging to *Aspergillus flavus* and three lancelet ESTs [GenBank:FE567488, GenBank:FE567489, GenBank:BW781185] belonging to some diatoms. As a result of this three step procedure the total of 197 GAPD sequences were identified for 131 species (109 deuterostomes and 22 other animals, see additional file 3: Accession codes of GAPD sequences used in the analysis).

Multiple alignment and phylogeny reconstruction

Since phylogenetic tree reconstruction is a computationally expensive process, only a part of the obtained sequences was subjected to the analysis. No more than 7 species from each class of deuterostomes were considered, as well as 6 species of insects as the representatives of protostomes (for more details see additional file 3: Accession codes of GAPD sequences used in the analysis). Two slightly different GAPD sequences from the flatworm *Macrostomum lignano*, both derived from several ESTs [GenBank:EG952499, GenBank:EG951174, GenBank:EG952414, GenBank:EG952720, GenBank:EG953822], were used as an outgroup. The total dataset for phylogenetic analysis comprised 92 GAPD sequences. Multiple alignment of protein sequences was performed by MUSCLE [100] and then manually edited. The alignment of nucleic sequences was constructed by means of RevTrans 1.4 Server [101] based on the protein alignment (see additional file 4: Raw alignment of GAPD nucleic sequences used in the phylogenetic analysis). Columns with gaps were eliminated before phylogenetic analysis.

The phylogenetic relationships between GAPD family members were reconstructed using both protein and nucleic sequences. The Bayesian method of tree reconstruction as implemented in MrBayes 3.1.2 [102,103] software was applied. The JTT model of amino-acid change [104], as well as the GTR model of nucleotide substitutions [105] were used. Preliminary analyses indicated that variation at the third position was saturated and confounded resolution at deep internal nodes. Therefore, trees based on nucleotide data were reconstructed in MrBayes by partitioning the data into the first, second and third codon positions, and allowing each partition to evolve at its own rate with its own shape parameter of gamma distribution.

For the Bayesian analyses, two independent runs were performed, each with four simultaneous chains that

sampled every 100 generation. Trees sampled before the cold chain reached stationarity based on plots of the maximum likelihood scores were discarded. Sampling continued until convergence was achieved based on the average standard deviation of the split frequencies as given in MrBayes. Node support was assessed as Bayesian posterior probabilities.

Synteny

Syntenic analysis is a reliable approach for establishing orthology. It is based on the assumption that local surroundings of genes are rarely affected by genomic rearrangements. Therefore, if the two genes have homologous neighbors, they are likely to have originated by vertical descent from a single ancestor and, in other words, be orthologous.

A syntenic analysis of the relationship between GAPD family members was performed by identification of positions of up to 20 genes both upstream and downstream of GAPD genes in human (*Homo sapiens*), stickleback (*Gasterosteus aculeatus*), zebrafish (*Danio rerio*), lancelet (*Branchiostoma floridae*), sea squirt (*Ciona savignyi*), sea urchin (*Strongylocentrotus purpuratus*) and acorn worm (*Saccoglossus kowalevskii*). Syntenic maps were constructed based on the information regarding gene location either available from Ensembl (human, both fishes and sea squirt) or obtained by conducting BLASTX searches of adjacent genomic regions against non-redundant protein databases. In the latter case, the homology between genes was decided if the identities of their protein product sequences were greater than 30%. The following genomes were used: *B. floridae* version 2.0 (Joint Genome Institute) [51], *S. purpuratus* version 2.1 (Human Genome Sequencing Center) [106] and *S. kowalevskii* version 1.0 (Human Genome Sequencing Center). Since genomic micro-rearrangements might occur, the matches between the local surroundings of GAPD genes were not required to be co-linear for establishing orthology. Gene losses and insertions were allowed as well.

Synonymous and non-synonymous substitution rates

To examine whether the GAPD family members are subjected to adaptive evolution, an analysis of variation under selective pressure was performed. Usually selective pressure is estimated by comparing the rates of synonymous (K_s) and non-synonymous substitutions (K_a) for the entire sequence. If K_a/K_s value is greater than unity, the whole sequence is supposed to be under positive selection, otherwise under purifying selection [107-109]. However, since each amino acid has a different function, the type and strength of natural selection

may be different for each amino acid. To detect the variation in K_a/K_s values across the sequence a sliding-window approach is often used [110,111].

Alignments of nucleotide sequences were constructed by PAL2NAL [112] based on protein alignments. K_a/K_s profiles were generated using a window of 120 base pairs and a step of 20 base pairs. Such a wide window was used because of high conservation of the analyzed sequences. Calculations of K_a and K_s for each window position were carried out with the aid of DnaSP 5.10 software [113].

Branch-specific selection tests

The differences in selective pressure between GAPD isoenzymes were also examined by means of more sophisticated branch-specific models as implemented in codeml program from PAML software [114]. Such kind of models assumes separate K_a/K_s values for different branches of the phylogenetic tree. They are often used for detecting selection changes after gene duplications, where one copy might evolve at a different rate due to acquisition of a new function or the loss of an old one [115-118].

First, GAPD sequences were divided into groups according to the results of phylogenetic analysis. Then a number of branch-specific models assuming separate K_a/K_s ratios for different combinations of groups were assayed. Likelihood ratio test (LRT) was used to determine whether the likelihoods of a pair of alternative branch-specific models are significantly different.

Additional material

Additional file 1: Alignment of the two forms of *Anolis carolinensis* GAPD-2 mRNA. The sequence of protein product is represented; the possible start codons are in bold. Translation of the full-length mRNA seems to begin from the first start codon and the protein product will possess the proline-rich domain (shown in italic). The shortened mRNA lacks the first start codon. Therefore translation should begin from the second one resulting in a protein product without the proline-rich domain.

Additional file 2: The proline-rich N-terminal domains of mammalian GAPD-2.

Additional file 3: Accession codes of GAPD sequences used in the analysis.

Additional file 4: Raw alignment of GAPD nucleic sequences used in the phylogenetic analysis.

Acknowledgements

We thank Dr. Dmitry Ivankov for providing helpful comments. The work was supported by the Russian Foundation for Basic Researches (grants 09-04-01122-a, 09-04-01150, 09-04-92740-NNIOM_a and 09-04-92741-NNIOM_a), the Ministry of Education and Science of Russian Federation (Federal Target Program "Scientific and scientific-pedagogical personnel of the innovative Russia for 2009-2013"), and the DFG International Research Training Group "Regulation and Evolution of Cellular Systems" (GRK 1563).

Author details

¹Faculty of Bioengineering and Bioinformatics, M.V. Lomonosov Moscow State University, Leninskiye Gory 1-73, Moscow 119991, Russian Federation. ²A.N. Belozersky Institute for Physical and Chemical Biology, M.V. Lomonosov Moscow State University, Leninskiye Gory 1-40, Moscow 119991, Russian Federation. ³Kharkevich Institute for Information Transmission Problems of the Russian Academy of Sciences, Moscow, 127994, Russian Federation. ⁴Department of Genome Oriented Bioinformatics, Technische Universität München, Wissenschaftszentrum Weihenstephan, Freising 85354, Germany.

Authors' contributions

MLK participated in collection of sequences, construction of phylogenetic trees, syntenic analyses and performing selection tests, and drafted the manuscript. WA participated in collection of sequences, construction of phylogenetic trees, syntenic analyses, and performing selection tests. DF helped to collect sequences, participated in performing selection tests, and helped write the manuscript. All authors read and approved the final manuscript. VIM conceived the study, participated in its design and coordination, gained funding and helped to draft the manuscript.

Received: 24 November 2010 Accepted: 10 June 2011

Published: 10 June 2011

References

- Liu YJ, Zheng D, Balasubramanian S, Carriero N, Khurana E, Robilotto R, Gerstein MB: **Comprehensive analysis of the pseudogenes of glycolytic enzymes in vertebrates: the anomalously high number of GAPDH pseudogenes highlights a recent burst of retrotrans-positional activity.** *BMC Genomics* 2009, **10**:480.
- Li Y, Nowotny P, Holmans P, Smemo S, Kauwe JSK, Hinrichs AL, Tacey K, Doil L, van Luchene R, Garcia V, et al: **Association of late-onset Alzheimer's disease with genetic variation in multiple members of the GAPD gene family.** *Proc Natl Acad Sci USA* 2004, **101**:15688-15693.
- Sirover MA: **New nuclear functions of the glycolytic protein, glyceraldehyde-3-phosphate dehydrogenase, in mammalian cells.** *J Cell Biochem* 2005, **95**:45-52.
- Sirover MA: **New insights into an old protein: the functional diversity of mammalian glyceraldehyde-3-phosphate dehydrogenase.** *Biochim Biophys Acta* 1999, **1432**:159-184.
- Raje CI, Kumar S, Harle A, Nanda JS, Raje M: **The macrophage cell surface glyceraldehyde-3-phosphate dehydrogenase is a novel transferrin receptor.** *J Biol Chem* 2007, **282**:3252-3261.
- Glaser PE, Gross RW: **Rapid plasmenylethanolamine-selective fusion of membrane bilayers catalyzed by an isoform of glyceraldehyde-3-phosphate dehydrogenase: discrimination between glycolytic and fusogenic roles of individual isoforms.** *Biochemistry* 1995, **34**:12193-12203.
- Robbins AR, Ward RD, Oliver C: **A mutation in glyceraldehyde 3-phosphate dehydrogenase alters endocytosis in CHO cells.** *J Cell Biol* 1995, **130**:1093-1104.
- Hessler RJ, Blackwood RA, Brock TG, Francis JW, Harsh DM, Smolen JE: **Identification of glyceraldehyde-3-phosphate dehydrogenase as a Ca²⁺-dependent fusogen in human neutrophil cytosol.** *J Leukoc Biol* 1998, **63**:331-336.
- Cueille N, Blanc CT, Riederer IM, Riederer BM: **Microtubule-associated protein 1B binds glyceraldehyde-3-phosphate dehydrogenase.** *J Proteome Res* 2007, **6**:2640-2647.
- Volker KW, Reinitz CA, Knull HR: **Glycolytic enzymes and assembly of microtubule networks.** *Comp Biochem Physiol B Biochem Mol Biol* 1995, **112**:503-514.
- Tisdale EJ, Azizi F, Artalejo CR: **Rab2 utilizes glyceraldehyde-3-phosphate dehydrogenase and protein kinase C[α] to associate with microtubules and to recruit dynein.** *The Journal of biological chemistry* 2009, **284**:5876-5884.
- Bryksin AV, Laktionov PP: **Role of glyceraldehyde-3-phosphate dehydrogenase in vesicular transport from golgi apparatus to endoplasmic reticulum.** *Biochemistry Biokhimiia* 2008, **73**:619-625.
- Duclos-Vallee JC, Capel F, Mabit H, Petit MA: **Phosphorylation of the hepatitis B virus core protein by glyceraldehyde-3-phosphate dehydrogenase protein kinase activity.** *J Gen Virol* 1998, **79**(Pt 7):1665-1670.
- Engel M, Seifert M, Theisinger B, Seyfert U, Welter C: **Glyceraldehyde-3-phosphate dehydrogenase and Nm23-H1/nucleoside diphosphate kinase A. Two old enzymes combine for the novel Nm23 protein phosphotransferase function.** *J Biol Chem* 1998, **273**:20058-20065.
- Kondo S, Kubota S, Mukudai Y, Nishida T, Yoshihama Y, Shiota T, Shintani S, Takigawa M: **Binding of glyceraldehyde-3-phosphate dehydrogenase to the cis-acting element of structure-anchored repression in ccn2 mRNA.** *Biochemical and biophysical research communications* 2011, **405**:382-387.
- Li Y, Huang T, Zhang X, Wan T, Hu J, Huang A, Tang H: **Role of glyceraldehyde-3-phosphate dehydrogenase binding to hepatitis B virus posttranscriptional regulatory element in regulating expression of HBV surface antigen.** *Arch Virol* 2009, **154**:519-524.
- Dai RP, Yu FX, Goh SR, Chng HW, Tan YL, Fu JL, Zheng L, Luo Y: **Histone 2B (H2B) expression is confined to a proper NAD⁺/NADH redox status.** *J Biol Chem* 2008, **283**:26894-26901.
- Demarse NA, Ponnusamy S, Spicer EK, Apohan E, Baatz JE, Ogretmen B, Davies C: **Direct binding of glyceraldehyde 3-phosphate dehydrogenase to telomeric DNA protects telomeres against chemotherapy-induced rapid degradation.** *Journal of molecular biology* 2009, **394**:789-803.
- Sundararaj KP, Wood RE, Ponnusamy S, Salas AM, Szulc Z, Bielawska A, Obeid LM, Hannun YA, Ogretmen B: **Rapid shortening of telomere length in response to ceramide involves the inhibition of telomere binding activity of nuclear glyceraldehyde-3-phosphate dehydrogenase.** *J Biol Chem* 2004, **279**:6152-6162.
- Nakagawa T, Hirano Y, Inomata A, Yokota S, Miyachi K, Kaneda M, Umeda M, Furukawa K, Omata S, Horigome T: **Participation of a fusogenic protein, glyceraldehyde-3-phosphate dehydrogenase, in nuclear membrane assembly.** *J Biol Chem* 2003, **278**:20395-20404.
- Singh R, Green MR: **Sequence-specific binding of transfer RNA by glyceraldehyde-3-phosphate dehydrogenase.** *Science* 1993, **259**:365-368.
- Azam S, Jouvett N, Jilani A, Vongsamphanh R, Yang X, Yang S, Ramotar D: **Human glyceraldehyde-3-phosphate dehydrogenase plays a direct role in reactivating oxidized forms of the DNA repair enzyme APE1.** *J Biol Chem* 2008, **283**:30632-30641.
- Meyer-Siegler K, Mauro DJ, Seal G, Wurzer J, deRiel JK, Sirover MA: **A human nuclear uracil DNA glycosylase is the 37-kDa subunit of glyceraldehyde-3-phosphate dehydrogenase.** *Proc Natl Acad Sci USA* 1991, **88**:8460-8464.
- Hwang NR, Yim S-H, Kim YM, Jeong J, Song EJ, Lee Y, Lee JH, Choi S, Lee K-J: **Oxidative modifications of glyceraldehyde-3-phosphate dehydrogenase play a key role in its multiple cellular functions.** *Biochem J* 2009, **423**:253-264.
- Sen N, Hara MR, Kornberg MD, Cascio MB, Bae B-I, Shahani N, Thomas B, Dawson TM, Dawson VL, Snyder SH, et al: **Nitric oxide-induced nuclear GAPDH activates p300/CBP and mediates apoptosis.** *Nat Cell Biol* 2008, **10**:866-873.
- Hara MR, Cascio MB, Sawa A: **GAPDH as a sensor of NO stress.** *Biochim Biophys Acta* 2006, **1762**:502-509.
- Arutyunova EI, Danshina PV, Domnina LV, Pleten AP, Muronetz VI: **Oxidation of glyceraldehyde-3-phosphate dehydrogenase enhances its binding to nucleic acids.** *Biochem Biophys Res Commun* 2003, **307**:547-552.
- Butterfield DA, Hardas SS, Lange ML: **Oxidatively modified glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and Alzheimer's disease: many pathways to neurodegeneration.** *J Alzheimers Dis* 2010, **20**:369-393.
- Naletova I, Schmalhausen E, Kharitonov A, Katrukha A, Saso L, Caprioli A, Muronetz V: **Non-native glyceraldehyde-3-phosphate dehydrogenase can be an intrinsic component of amyloid structures.** *Biochim Biophys Acta* 2008, **1784**:2052-2058.
- Mazzola JL, Sirover MA: **Reduction of glyceraldehyde-3-phosphate dehydrogenase activity in Alzheimer's disease and in Huntington's disease fibroblasts.** *J Neurochem* 2001, **76**:442-449.
- Bae BI, Hara MR, Cascio MB, Wellington CL, Hayden MR, Ross CA, Ha HC, Li XJ, Snyder SH, Sawa A: **Mutant huntingtin: nuclear translocation and cytotoxicity mediated by GAPDH.** *Proc Natl Acad Sci USA* 2006, **103**:3405-3409.
- Mazzola JL, Sirover MA: **Alteration of nuclear glyceraldehyde-3-phosphate dehydrogenase structure in Huntington's disease fibroblasts.** *Brain Res Mol Brain Res* 2002, **100**:95-101.
- Shchutskaya YY, Elkina YL, Kuravsky ML, Bragina EE, Schmalhausen EV: **Investigation of glyceraldehyde-3-phosphate dehydrogenase from human sperms.** *Biochemistry (Mosc)* 2008, **73**:185-191.

34. Krisfalusi M, Miki K, Magyar PL, O'Brien DA: **Multiple glycolytic enzymes are tightly bound to the fibrous sheath of mouse spermatozoa.** *Biol Reprod* 2006, **75**:270-278.
35. Bunch DO, Welch JE, Magyar PL, Eddy EM, O'Brien DA: **Glyceraldehyde 3-phosphate dehydrogenase-S protein distribution during mouse spermatogenesis.** *Biol Reprod* 1998, **58**:834-841.
36. Westhoff D, Kamp G: **Glyceraldehyde 3-phosphate dehydrogenase is bound to the fibrous sheath of mammalian spermatozoa.** *J Cell Sci* 1997, **110**(Pt 15):1821-1829.
37. Miki K, Qu W, Goulding EH, Willis WD, Bunch DO, Strader LF, Perreault SD, Eddy EM, O'Brien DA: **Glyceraldehyde 3-phosphate dehydrogenase-S, a sperm-specific glycolytic enzyme, is required for sperm motility and male fertility.** *Proc Natl Acad Sci USA* 2004, **101**:16501-16506.
38. Elkina YL, Kuravsky ML, El'darov MA, Stogov SV, Muronetz VI, Schmalhausen EV: **Recombinant human sperm-specific glyceraldehyde-3-phosphate dehydrogenase: structural basis for enhanced stability.** *Biochim Biophys Acta* 2010, **1804**:2207-2212.
39. Kuravsky ML, Muronetz VI: **Somatic and sperm-specific isoenzymes of glyceraldehyde-3-phosphate dehydrogenase: comparative analysis of primary structures and functional features.** *Biochemistry (Mosc)* 2007, **72**:744-749.
40. Cho YS, Lee SY, Kim KH, Nam YK: **Differential modulations of two glyceraldehyde 3-phosphate dehydrogenase mRNAs in response to bacterial and viral challenges in a marine teleost *Oplegnathus fasciatus* (Perciformes).** *Fish Shellfish Immunol* 2008, **25**:472-476.
41. Manchado M, Infante C, Asensio E, Canavate JP: **Differential gene expression and dependence on thyroid hormones of two glyceraldehyde-3-phosphate dehydrogenases in the flatfish *Senegalese sole* (*Solea senegalensis* Kaup).** *Gene* 2007, **400**:1-8.
42. Steinke D, Hoegg S, Brinkmann H, Meyer A: **Three rounds (1R/2R/3R) of genome duplications and the evolution of the glycolytic pathway in vertebrates.** *BMC Biol* 2006, **4**:16.
43. Baibai T, Oukhattar L, Mountassif D, Assobhei O, Serrano A, Soukri A: **Comparative molecular analysis of evolutionarily distant glyceraldehyde-3-phosphate dehydrogenase from *Sardina pilchardus* and *Octopus vulgaris*.** *Acta Biochim Biophys Sin (Shanghai)* 2010, **42**:863-872.
44. A M, Rangarajan L, Bhat S: **Computational approach towards finding evolutionary distance and gene order using promoter sequences of central metabolic pathway.** *Interdiscip Sci* 2009, **1**:128-132.
45. Takishita K, Inagaki Y: **Eukaryotic origin of glyceraldehyde-3-phosphate dehydrogenase genes in *Clostridium thermocellum* and *Clostridium cellulolyticum* genomes and putative fates of the exogenous gene in the subsequent genome evolution.** *Gene* 2009, **441**:22-27.
46. Akinyi S, Gaona J, Meyer EVS, Barnwell JW, Galinski MR, Corredor V: **Phylogenetic and structural information on glyceraldehyde-3-phosphate dehydrogenase (G3PDH) in *Plasmodium* provides functional insights.** *Infect Genet Evol* 2008, **8**:205-212.
47. Stechmann A, Baumgartner M, Silberman JD, Roger AJ: **The glycolytic pathway of *Trimastix pyriformis* is an evolutionary mosaic.** *BMC Evol Biol* 2006, **6**:101.
48. Oslancova A, Janecek S: **Evolutionary relatedness between glycolytic enzymes most frequently occurring in genomes.** *Folia Microbiol (Praha)* 2004, **49**:247-258.
49. Petersen J, Brinkmann H, Cerff R: **Origin, evolution, and metabolic role of a novel glycolytic GAPDH enzyme recruited by land plant plastids.** *J Mol Evol* 2003, **57**:16-26.
50. Canback B, Andersson SG, Kurland CG: **The global phylogeny of glycolytic enzymes.** *Proc Natl Acad Sci USA* 2002, **99**:6097-6102.
51. Putnam NH, Butts T, Ferrig DEK, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu J-K, et al: **The amphioxus genome and the evolution of the chordate karyotype.** *Nature* 2008, **453**:1064-1071.
52. Panopoulou G, Poustka AJ: **Timing and mechanism of ancient vertebrate genome duplications – the adventure of a hypothesis.** *Trends Genet* 2005, **21**:559-567.
53. Hokamp K, McLysaght A, Wolfe KH: **The 2R hypothesis and the human genome sequence.** *J Struct Funct Genomics* 2003, **3**:95-110.
54. Hanauer A, Mandel JL: **The glyceraldehyde 3 phosphate dehydrogenase gene family: structure of a human cDNA and of an X chromosome linked pseudogene; amazing complexity of the gene family in mouse.** *EMBO J* 1984, **3**:2627-2633.
55. Piechaczyk M, Blanchard JM, Riaad-El Sabouty S, Dani C, Marty L, Jeanteur P: **Unusual abundance of vertebrate 3-phosphate dehydrogenase pseudogenes.** *Nature* 1984, **312**:469-471.
56. Irwin DM, Tan H: **Molecular evolution of the vertebrate hexokinase gene family: Identification of a conserved fifth vertebrate hexokinase gene.** *Comp Biochem Physiol Part D Genomics Proteomics* 2008, **3**:96-107.
57. Sato Y, Nishida M: **Post-duplication charge evolution of phosphoglucose isomerases in teleost fishes through weak selection on many amino acid sites.** *BMC Evol Biol* 2007, **7**:204.
58. Piast M, Kustrzeba-Wojcicka I, Matusiewicz M, Banas T: **Molecular evolution of enolase.** *Acta Biochim Pol* 2005, **52**:507-513.
59. Kao HW, Lee SC: **Phosphoglucose isomerases of hagfish, zebrafish, gray mullet, toad, and snake, with reference to the evolution of the genes in vertebrates.** *Mol Biol Evol* 2002, **19**:367-374.
60. Li YJ, Tsoi SC, Mannen H, Shoei-lung Li S: **Phylogenetic analysis of vertebrate lactate dehydrogenase (LDH) multigene families.** *J Mol Evol* 2002, **54**:614-624.
61. Tracy MR, Hedges SB: **Evolutionary history of the enolase gene family.** *Gene* 2000, **259**:129-138.
62. Lynch M, Force A: **The probability of duplicate gene preservation by subfunctionalization.** *Genetics* 2000, **154**:459-473.
63. Wagner A: **The fate of duplicated genes: loss or new function?** *Bioessays* 1998, **20**:785-788.
64. Walsh JB: **How often do duplicated genes evolve new functions?** *Genetics* 1995, **139**:421-428.
65. Innan H, Kondrashov F: **The evolution of gene duplications: classifying and distinguishing between models.** *Nat Rev Genet* 2010, **11**:97-108.
66. Hughes AL: **The evolution of functionally novel proteins after gene duplication.** *Proc Biol Sci* 1994, **256**:119-124.
67. Donoghue PCJ, Purnell MA: **Genome duplication, extinction and vertebrate evolution.** *Trends Ecol Evol* 2005, **20**:312-319.
68. Papp B, Pal C, Hurst LD: **Dosage sensitivity and the evolution of gene families in yeast.** *Nature* 2003, **424**:194-197.
69. Bisbee CA, Baker MA, Wilson AC, Haji-Azimi I, Fischberg M: **Albumin phylogeny for clawed frogs (*Xenopus*).** *Science* 1977, **195**:785-787.
70. Evans BJ, Kelley DB, Tinsley RC, Melnick DJ, Cannatella DC: **A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution.** *Mol Phylogenet Evol* 2004, **33**:197-213.
71. Hellsten U, Khokha MK, Grammer TC, Harland RM, Richardson P, Rokhsar DS: **Accelerated gene evolution and subfunctionalization in the pseudotetraploid frog *Xenopus laevis*.** *BMC Biol* 2007, **5**:31.
72. Vinson JP, Jaffe DB, O'Neill K, Karlsson EK, Stange-Thomann N, Anderson S, Mesirov JP, Satoh N, Satou Y, Nusbaum C, Birren B, Galagan JE, Lander ES: **Assembly of polymorphic genomes: algorithms and application to *Ciona savignyi*.** *Genome Res* 2005, **15**:1127-1135.
73. Small KS, Bruno M, Hill MM, Sidow A: **Extreme genomic variation in a natural population.** *Proc Natl Acad Sci USA* 2007, **104**:5698-5703.
74. Veitia RA: **Gene dosage balance: deletions, duplications and dominance.** *Trends Genet* 2005, **21**:33-35.
75. Kondrashov FA, Koonin EV: **A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications.** *Trends Genet* 2004, **20**:287-290.
76. Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV: **Selection in the evolution of gene duplications.** *Genome Biol* 2002, **3**:RESEARCH0008.
77. Welch JE, Barbee RR, Magyar PL, Bunch DO, O'Brien DA: **Expression of the spermatogenic cell-specific glyceraldehyde 3-phosphate dehydrogenase (GAPDS) in rat testis.** *Mol Reprod Dev* 2006, **73**:1052-1060.
78. Welch JE, Brown PL, O'Brien DA, Magyar PL, Bunch DO, Mori C, Eddy EM: **Human glyceraldehyde 3-phosphate dehydrogenase-2 gene is expressed specifically in spermatogenic cells.** *J Androl* 2000, **21**:328-338.
79. Fenderson BA, Toshimori K, Muller CH, Lane TF, Eddy EM: **Identification of a protein in the fibrous sheath of the sperm flagellum.** *Biol Reprod* 1988, **38**:345-357.
80. Danshina PV, Geyer CB, Dai Q, Goulding EH, Willis WD, Kitto GB, McCarrey JR, Eddy EM, O'Brien DA: **Phosphoglycerate kinase 2 (PGK2) is essential for sperm function and male fertility in mice.** *Biol Reprod* 2010, **82**:136-145.
81. VandeBerg JL, Cooper DW, Close PJ: **Testis specific phosphoglycerate kinase B in mouse.** *J Exp Zool* 1976, **198**:231-240.
82. Goldberg E, Eddy EM, Duan C, Odet F: **LDHC: the ultimate testis-specific gene.** *J Androl* 2010, **31**:86-94.

83. Blanco A, Zinkham WH: Lactate Dehydrogenases in Human Testes. *Science* 1963, **139**:601-602.
84. Goldberg E: Lactic and Malic Dehydrogenases in Human Spermatozoa. *Science* 1963, **139**:602-603.
85. Boer PH, Adra CN, Lau YF, McBurney MW: The testis-specific phosphoglycerate kinase gene pgk-2 is a recruited retroposon. *Mol Cell Biol* 1987, **7**:3107-3112.
86. McCarrey JR, Thomas K: Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. *Nature* 1987, **326**:501-505.
87. Turner RM: Tales from the tail: what do we really know about sperm motility? *J Androl* 2003, **24**:790-803.
88. Mukai C, Okuno M: Glycolysis plays a major role for adenosine triphosphate supplementation in mouse sperm flagellar movement. *Biol Reprod* 2004, **71**:540-547.
89. Williamson MP: The structure and function of proline-rich regions in proteins. *Biochem J* 1994, **297**(Pt 2):249-260.
90. Schmidt EE, Davies CJ: The origins of polypeptide domains. *Bioessays* 2007, **29**:262-270.
91. Zhang XHF, Chasin LA: Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. *Proc Natl Acad Sci USA* 2006, **103**:13427-13432.
92. Wang W, Zheng H, Yang S, Yu H, Li J, Jiang H, Su J, Yang L, Zhang J, McDermott J, et al: Origin and evolution of new exons in rodents. *Genome Res* 2005, **15**:1258-1264.
93. Kondrashov FA, Koonin EV: Evolution of alternative splicing: deletions, insertions and origin of functional parts of proteins from intron sequences. *Trends Genet* 2003, **19**:115-119.
94. Tompa P: Intrinsically unstructured proteins evolve by repeat expansion. *Bioessays* 2003, **25**:847-855.
95. Bondareva AA, Schmidt EE: Early vertebrate evolution of the TATA-binding protein, TBP. *Mol Biol Evol* 2003, **20**:1932-1939.
96. Flicek P, Aken BL, Ballester B, Beal K, Bragin E, Brent S, Chen Y, Clapham P, Coates G, Fairley S, et al: Ensembl's 10th year. *Nucleic Acids Res* 2010, **38**:557-562.
97. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997, **25**:3389-3402.
98. The Universal Protein Resource (UniProt) in 2010. *Nucleic acids research* 2010, **38**:D142-148.
99. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: GenBank. *Nucleic Acids Res* 2008, **36**:25-30.
100. Edgar RC: MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 2004, **5**:113.
101. Wernersson R, Pedersen AG: RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res* 2003, **31**:3537-3539.
102. Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F: Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* 2004, **20**:407-415.
103. Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP: Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 2001, **294**:2310-2314.
104. Jones DT, Taylor WR, Thornton JM: The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 1992, **8**:275-282.
105. Tavaré S: Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. In *American Mathematical Society: Lectures on Mathematics in the Life Sciences. Volume 17*. Edited by: Miura RM. Providence, RI: Amer Mathematical Society; 1986:57-86.
106. Sodergren E, Shen Y, Song X, Zhang L, Gibbs RA, Weinstock GM: Shedding genomic light on Aristotle's lantern. *Dev Biol* 2006, **300**:2-8.
107. Nielsen R, Yang Z: Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 1998, **148**:929-936.
108. Goldman N, Yang Z: A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* 1994, **11**:725-736.
109. Nei M, Gojobori T: Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 1986, **3**:418-426.
110. Lee YS, Kim T-H, Kang T-W, Chung W-H, Shin G-S: WSPMaker: a web tool for calculating selection pressure in proteins and domains using window-sliding. *BMC Bioinformatics* 2008, **9**(Suppl 12):S13.
111. Xia X, Kumar S: Codon-based detection of positive selection can be biased by heterogeneous distribution of polar amino acids along protein sequences. *Comput Syst Bioinformatics Conf* 2006, **4**:335-340.
112. Suyama M, Torrents D, Bork P: PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 2006, **34**:609-612.
113. Librado P, Rozas J: DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009, **25**:1451-1452.
114. Yang Z: PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 2007, **24**:1586-1591.
115. Hughes J, Criscuolo F: Evolutionary history of the UCP gene family: gene duplication and selection. *BMC Evol Biol* 2008, **8**:306.
116. Karn RC, Clark NL, Nguyen ED, Swanson WJ: Adaptive evolution in rodent seminal vesicle secretion proteins. *Mol Biol Evol* 2008, **25**:2301-2310.
117. Lynch VJ, Roth JJ, Wagner GP: Adaptive evolution of Hox-gene homeodomains after cluster duplications. *BMC Evol Biol* 2006, **6**:86.
118. Soyer Y, Orsi RH, Rodriguez-Rivera LD, Sun Q, Wiedmann M: Genome wide evolutionary analyses reveal serotype specific patterns of positive selection in selected *Salmonella* serotypes. *BMC Evol Biol* 2009, **9**:264.
119. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Federhen S, et al: Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2010, **38**:5-16.
120. Letunic I, Bork P: Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 2007, **23**:127-128.

doi:10.1186/1471-2148-11-160

Cite this article as: Kuravsky et al: Testis-specific glyceraldehyde-3-phosphate dehydrogenase: origin and evolution. *BMC Evolutionary Biology* 2011 **11**:160.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

