Phylogenetic analysis of the diacylglycerol kinase family of proteins and identification of multiple highly-specific conserved inserts and deletions within the catalytic domain that are distinctive characteristics of different classes of DGK homologs

Radhey S. Gupta*, Richard M. Epand

Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario, Canada

* gupta@mcmaster.ca

Abstract

Diacylglycerol kinase (DGK) family of proteins, which phosphorylates diacylglycerol into phosphatidic acid, play important role in controlling diverse cellular processes in eukaryotic organisms. Most vertebrate species contain 10 different DGK isozymes, which are grouped into 5 different classes based on the presence or absence of specific functional domains. However, the relationships among different DGK isozymes or how they have evolved from a common ancestor is unclear. The catalytic domain constitutes the single largest sequence element within the DGK proteins that is commonly and uniquely shared by all family members, but there is limited understanding of the overall function of this domain. In this work, we have used the catalytic domain sequences to construct a phylogenetic tree for the DGK family members from representatives of the main vertebrate classes and have also examined the distributions of various DGK isozymes in eukaryotic phyla. In a tree based on catalytic domain sequences, the DGK homologs belonging to different classes formed strongly supported clusters which were separated by long branches, and the different isozymes within each class also generally formed monophyletic groupings. Further, our analysis of the sequence alignments of catalytic domains has identified >10 novel sequence signatures consisting of conserved signature indels (inserts or deletions, CSIs) that are distinctive characteristics of either particular classes of DGK isozymes, or are commonly shared by members of two or more classes of DGK isozymes. The conserved indels in protein sequences are known to play important functional roles in the proteins/organisms where they are found. Thus, our identification of multiple highly specific CSIs that are distinguishing characteristics of different classes of DGK homologs points to the existence of important differences in the catalytic domain function among the DGK isozymes. The identified CSIs in conjunction with the results of blast searches on species distribution of DGK isozymes also provide useful insights into the evolutionary relationships among the DGK family of proteins.

Introduction

Diacylglycerol (DAG) and phosphatidic acid (PA) are two main signalling molecules within eukaryotic cells which, through their interactions with different effector proteins, play central roles in regulating diverse cellular processes [1–6]. The main source of DAG in cells is via the hydrolysis of phosphoinositides by the enzyme phospholipase C in response to a variety of extracellular stimuli including growth factors and hormones. The DAG produced serves as a substrate for the enzyme DAG kinase (DGK), which phosphorylates it into PA. Due to the ability of DGK to convert one important signalling molecule (DAG) into another (PA), the activity of DGK in different cells is tightly controlled for maintenance of normal physiological conditions [1–6]. Because DGK plays an important role in controlling diverse cellular processes including development, cell division and proliferation, neuronal and immune responses, vascular traffic, apoptosis, cytoskeletal reorganization, etc. multiples forms of DGK are generally found in most eukaryotic organisms [1–10].

In mammalian species, where the DGK family of proteins has been best studied, 10 different isozymes of DGK, designated as α, β, γ, δ, ε, ζ, η, θ, ι and κ, differing in their biochemical properties, tissue distributions, as well as their lengths (ranging from 567 aa to >1150 aa) have been identified [1–6,11]. Based on sequence similarities between these isozymes and the presence or absence of specific functional domains, the known DGK members have been grouped into 5 different classes or Types (Fig 1) [1–6,11,12]. All of these isozymes share in common a large catalytic domain, which is sometimes divided into two parts–catalytic and accessory domains, and 2 or 3 cysteine-rich domains, referred to as the C1 domains [reviewed in[1–6,12, 13]]. The simplest and shortest (567 aa) of these isozymes is DGK-ε, sole member of the class III DGK, which contains only the commonly shared catalytic domain and the two C1 domains. In addition, DGK-ε also contains a conserved helical segment near its N-terminal end that is indicated to play an important role in its membrane interaction [12,13]. The class I DGKs (α, β and γ isozymes), in addition to containing the commonly shared domains, are characterized by the presence of two EF-hand motifs and a conserved domain of unknown function near the N-terminal end [1–6,11,12]. Novel characteristics of the class II DGK isozymes (δ, η and κ isoforms) include the presence of a plecstrin homology (PH) domain, a sterile α motif (SAM) domain and a large insert within the catalytic domain separating it into two parts. The class IV DGKs (ζ and ι isoforms) are distinguished from the others due to their containing a sequence homologous to the MARCKS phosphorylation site domain and four ankyrin repeats near the C-terminal end. Lastly, DGK-θ is the sole member of the class V DGK and it contains three C1 domains, a Gly-Pro rich domain and a PH-domain-like region with overlapping RAS-associating domain [1–6,11]. The unique domains present in DGK isozymes, through their interactions with various regulatory proteins and other molecules present in different cells and tissues, play key roles in the diverse physiological functions exhibited by different DGK isozymes [1–6].

Although the differences in the structural characteristics of DGK isozymes and their involvement in diverse cellular functions are widely known [1–6], there is limited (or no) information available at present as to how different DGK isozymes, or the different classes of DGKs, are related to each other, and have possibly evolved from a common ancestor. There is also limited understanding of the distributions of DGK isozymes in eukaryotic organisms other than mammals [3,10,14–16]. Knowledge of the species distribution and evolutionary relationships among DGK isozymes can provide useful insights into their physiological functions. Within the DGK family of proteins, the catalytic domain constitutes the single largest sequence element that is commonly and uniquely shared by all of the family members [1–6,11]. In the present work, we have used the catalytic domain sequences to construct a
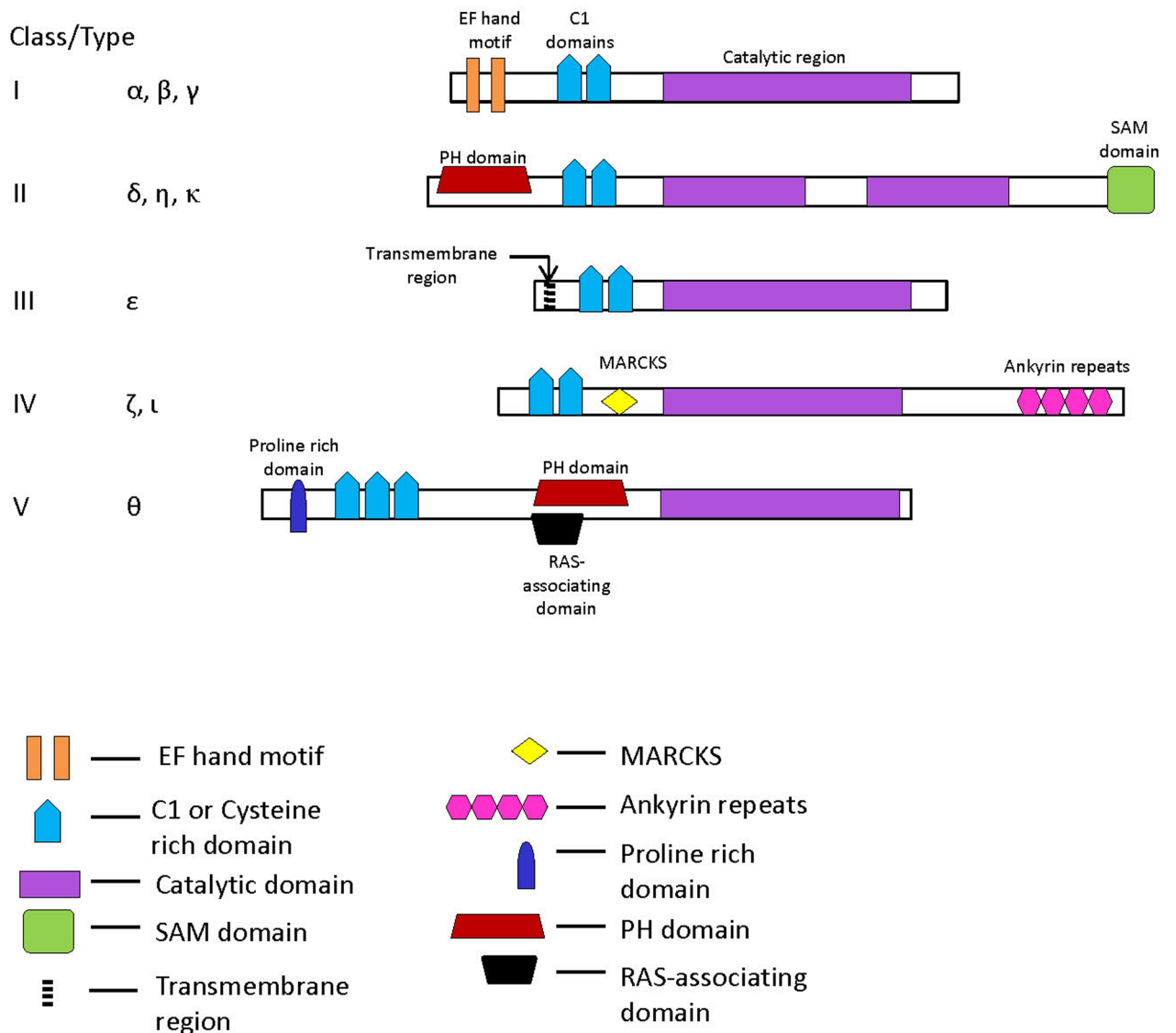
Fig 1. Sequence characteristics of different forms of DGK isozymes that are found in mammalian/vertebrate species. The ten known isozymes are grouped into five Classes (or Types) and a schematic of the known main functional domains in these isozymes is shown. Abbreviations; PH, plecstrin homology; MARCKS, sequence similar to the MARCKS phosphorylation site; SAM, sterile α motif. The information for the presence of different domains in the DGK isozymesis based on the following reviews [1–6,11,12].

https://doi.org/10.1371/journal.pone.0182758.g001

phylogenetic tree for the DGK family members. Additionally, and more importantly, our analysis of the sequence alignment of the catalytic domain has identified large numbers of novel sequence signatures in the forms of conserved inserts and deletions (indels) that are distinctive of either particular classes of DGK isozymes, or are commonly shared by members of two or more classes of DGK isozymes. These results point to the existence of important differences in the function of catalytic domain for different classes of DGK isozymes and are also useful in understanding the evolutionary relationships among the DGK family of proteins.

Methods

Phylogenetic tree construction and identification of conserved indels in the catalytic domain sequences

The protein sequences of all ten DGK isozymes from representatives of the main vertebrate groups were retrieved from the NCBI nr database. In most cases, sequences were obtained from the following species covering the diversity of the vertebrate phylum: Mammals (*Homo sapiens*, *Rattus novergicus*), Birds (*Serinus canaria*, *Sturnus vulgaris*), Reptiles (*Protobothrops mucrosquamatus*, *Python bivittatus*), Amphibia (*Xenopus tropicalis*) and Fishes (*Pundamilia nyererei*, *Maylandia zebra*). The information for accession numbers and sequence lengths of different DGK homologs that were used for phylogenetic studies is provided in the S1 Table. Multiple sequence alignments of different DGK isozymes were created using the Clustal X ver. 2.0 program [17], and based on these alignments, sequences corresponding to the commonly shared catalytic domain were identified. The large insert present within the catalytic domain (CD) of Type II DGK homologs was excluded from these alignments. The resulting sequences for the DGK CDs were then realigned and, based on these alignments, a segment of 364 aa was identified that showed good conservation amongst different DGK classes. This sequence alignment, which was used for phylogenetic analysis, is provided in S1 Fig. All of the columns with sequence gaps were not considered during tree construction leaving a total of 260 aligned positions that were used for phylogenetic analysis. The evolutionary history based on this sequence alignment was inferred using the Maximum likelihood (ML) method, with 100 bootstrap replication, using the MEGA 6 [18] based on the Jones–Taylor–Thornton (JTT) model [19]. The tree with the highest log likelihood (-7845.5131) is shown. The numbers shown on the branches identifies clades in which the associated taxa clustered together >50% time in bootstrap replicates.

Sequence alignments of the DGK homologs were also examined for the presence of conserved indels (insertions or deletions) which were flanked on both sides by at least 5–6 conserved residues in the neighboring 30–40 amino acids and which appeared to be specifically present in either a particular class of DGK isozymes, or which were shared by the DGK homologs from more one than classes [20–22]. For conserved indels meeting these criteria, additional Blastp searches were carried out using the NCBI NR (non-redundant) database to determine the specificity of the indels for the DGK classes of isozymes. These blast results were examined for the presence/absence of the identified CSIs in top 500 blast hits. SIG_CREATE and SIG_STYLE programs (available on Gleans.net) were used to format and create the signature CSIs files shown here [22,23]. Due to space limitations, sequence information for only 1–2 representative species from each vertebrate group (viz. generally the same as those employed for phylogenetic studies for mammals, birds, fishes, reptiles and amphibians) for different DGK isozymes is shown in the alignment files. However, unless otherwise noted, all of the described CSIs are specifically present in the indicated groups/classes of DGK isozymes, in different vertebrate species, where the distribution of these CSIs was studied in detail.

The presence or absence of different DGK isozymes in members from different kingdoms and main phyla of eukaryotes was assessed by carrying out Blastp searches on sequences of different DGK isozymes from human and mouse species against annotated members of specific kingdoms or phyla present in the NCBI nr database. These searches were carried out using the entire sequences of the DGK isozymes as well as utilizing sequences fonly the catalytic domain sequence. Based on significant hits observed in these Blast searches, and to which DGK isozymes the observed hits exhibited maximal similarity, tentative inferences regarding the presence or absence of specific DGK isozymes in deeper branching eukaryotic taxa were made.

Results

Phylogenetic analysis of the DGK isozymes based on catalytic domain sequences

The DGK isozymes differ considerably from each other in their sequence lengths and different domains present in them (see introduction and Fig 1). Pairwise sequence comparison studies (unpublished) indicate that the overall sequence similarity between members of different classes is in the range of 15–30% and much of the observed similarity is seen within the commonly shared catalytic domain. Based on sequence alignments of DGK homologs from representatives of different groups of vertebrates, conserved regions from the catalytic domain which could be properly aligned amongst different DGK classes were identified and these were subjected to phylogenetic analyses. A maximum-likelihood tree based on 100 bootstrap samples of this sequence region was constructed using the JTT model (see Methods) to infer the evolutionary relationships amongst different DGK homologs. The tree shown in Fig 2 is based on 260 aligned positions that were left after removal of all insertions and deletions. In this phylogenetic tree (Fig 2), DGK homologs belonging to all five classes or Types formed strongly supported clusters. These clusters were separated by long branches indicating that the DGK sequences for different classes have diverged considerably from each other even within the conserved catalytic domain. Within each DGK class, homologs belonging to different DGK isozymes also generally formed separate clusters, except for the homologs of kappa and zeta isozymes which exhibited polyphyletic branching and were separated into two clusters. The branches separating different DGK isozymes within any given class were generally short indicating that the isozymes within different classes are very similar to each other. Based on the phylogenetic tree, some information regarding relationships among different isozymes within different classes can also be gleaned. In this tree, DGK homologs belonging to the classes I and IV grouped together but there was only weak support (clustered together in 51% of bootstrap samples) for this grouping. Within the class I isozymes, the DGK-α homologs formed outgroups relative to a cluster consisting of the DGK-β and DGK-γ homologs, indicating that the sequences from the latter two groups are more closely related. Similarly, within the class II isozymes, a grouping of DGK-δ and DGK-η homologs was supported by the observed bootstrap score, with the DGK-κ sequences forming an outgroup of this cluster. Lastly, in the tree shown in Fig 2, which was rooted at midpoint, the DGK-ε and DGK-θ homologs formed the most outlying clades.

Conserved signatures indels in the catalytic domain shared by different classes of DGK isozymes

In the tree shown in Fig 2, while the distinction between different DGK classes was strongly supported, the nodes connecting these classes had low bootstrap scores indicating that the evolutionary relationships among different DGK classes were not reliably resolved. Thus, other molecular sequence based approaches which can provide insights are needed. Conserved signature indels (CSIs) represent an important class of molecular markers that have been used in the past to resolve a number of important evolutionary questions [21,24–27]. The CSIs that are useful for evolutionary studies are generally of fixed lengths, present at specific positions iparticular genes/proteins, and they are flanked on both sides by conserved regions to ensure that they constitute reliable characteristics [21,25,27,28]. The CSIs in genes/proteins sequences generally result from rare genetic changes and the most parsimonious explanation to account for their shared presence in a given gene or protein from a specific group of species is that the genetic change giving rise to the CSI occurred in a common ancestor of the indicated group

Fig 2. A maximum-likelihood tree for the DGK isozymes from representative vertebrate species based on the conserved regions in catalytic domain. The accession numbers of different sequences which were used for tree construction are shown after the species names. The tree was constructed as described in the

Methods section and the branches which were supported by >50% bootstrap scores are marked. The tree shown was rooted at midpoint and drawn to scale, with branch lengths measured in the number of substitutions per site. Different DGK isozymes as well as the five main classes are marked.

https://doi.org/10.1371/journal.pone.0182758.g002

and then it was vertically inherited by the other group members [21,22,25,29]. Due to the discrete natures of the CSIs and the fact that they are located within conserved regions,their presence or absence in different lineages or proteins is generally not affected by factors such as differences in evolutionary rates among different species, or proteins, and long-branch attraction artefacts [21,24–26]. We have examined the sequence alignment of catalytic domains from DGK homologs for the presence of CSIs and these analyses have identified a number of useful CSIs that are specifically shared by particular groups of DGK isozymes.

When a conserved indel is present in a given protein, to infer whether the observed indel is an insert or a deletion, information for the presence or absence of the indel in the ancestral state of the protein is needed [22]. In most cases, such information could be obtained by determining the presence or absence of the indel in species which are ancestral to those containing the indel [25,27,28,30]. In some instances, if two proteins have evolved by an ancient gene duplication event, then the presence or absence of the indel in the other protein has also proven useful to infer whether the observed indel is an insert or a deletion [25,30]. However, due to the uniqueness of the DGK catalytic domain sequence for the eukaryotic homologs, no suitable outgroup is available for determining the ancestral state of the protein or reliably inferring whether a given CSI represents an insertion within a certain class of DGK homologs or a deletion in the rest of the DGK homologs. Hence, our use of the term insert(s) or deletion(s) to describe the identified CSIs is only in reference to the presence or absence of the same or similar sequence characteristics in other classes of DGK homologs. Additionally, it should also be acknowledged that while the presence or absence of a CSI within a given conserved region can be reliably identified (due to the sequence conservation of flanking regions), it is often difficult to specify the exact location of the CSI within a conserved region. This is especially true when the sequences of different taxa or homologs that are being compared differ considerably from each other and there is no structural information available for the proteins to guide the sequence alignments for correct placement of the CSI. Due to this limitation, the positions of various CSIs in the sequence alignment shown here represent their most likely positions based on our judgement of the sequence conservation of the flanking regions. However, we cannot exclude the possibility that in the optimal alignment that could be supported by other data the actual positions of some of these CSIs might be shifted by a few residues either to the left or on the right. Information regarding the characteristics of various CSIs that have been identified by our analysis is provided below.

The first of the identified CSIs within the catalytic domain is a 16–24 aa insert (marked signature ❶ in Fig 3) that is uniquely found in all of the Class I DGK homologs from different species. The observed CSI is present in a conserved region near the C-terminal end of the catalytic domain and its shared presence by all DGK -α, -β and -γ homologs provides further evidence of the close and specific relationship among these isozymes belonging to class I DGKs. For DGK-α and DGK-γ, interesting differences are also seen in the length of this CSI amongst the vertebrate species. For DGK-α, in contrast to the fishes (e.g. *Pundamila nyererei* and *Maylinda zebra*), homologs found in mammals (e.g. humans, *Rattus novergicus*), birds (*Serinus canaria*) and amphibians (*Xenopus tropocalis*) are 6 aa longer, suggesting that an additional insert within the DGK-α likely occurred in a common ancestor of the latter groups of vertebrate species. Similarly, for DGK-γ, the birds and reptiles harbor a longer insert in contrast to that present in mammals. Interestingly, the homologs for DGK-γ were not detected in fishes

```
                                                                ❶
                                            657(446)                                      709(481)
                                            ILNIPSMYGGTNLWG  ENKKNRAVIR      ESRKGV      TDPKELKFCVQDLSDQLLEVVG
γ   Homo sapiens γ          NP_001337.2    ---------------  -T--------      ----S-      -------C--------------
    Rattus novergicus γ     NP_037258      ---------------  -T--------      ----S-      -------C--------------
    Serinus canaria γ       XP_009087565   V---------S----  -T--Q-GYN-LGKKVP-KVP-TVV    --A-----------H------
    Python bivittatus γ     XP_015743514   V----------S----  -T--Q-SYN-MSKKIP-KLQSSVV   --A------------------
β   Homo sapiens β          XP_016867279   -------H--S----  -S--R-SHR-IEKKGSDK-TT-      --A-----AS-----------
    Rattus novergicus β     NP_062177      -------H--S----  -S--K-SHR-IEKKGSDK-PTL      --A-----AS-----------
    Serinus canaria β       XP_018765012   -------H--S----  -T--R-SHR-TEKKRSDK-TT-      --A-----VC-------M----
    Pro. mucrosquamatus β   XP_015672258   -------H--S----  -T--R-SHR-FEKNRPDK-IT-      --A-----AC-------V----
    Xenopus tropicalis β    NP_001096214   -------H--S----  -T--R-SNR-TDKKNSDK-TT-      --A-----AA-----------
    Maylandia zebra β       XP_004560243   -------H--S----  -S--R-GYRKGGKKGQDK-TP-      L-----M-A------------
α   Homo sapiens α          XP_005268745   V------H--S----  DTRRPHGD-YGINQALGATAK-I     ---DI--T--P----KR-----
    Rattus novergicus α     XP_008763241   V------H--S----  DT-RPHGDTCGINQALG-VAKII     ---DI--T--P-M--KR-----
    Xenopus tropicalis α    XP_002934577   VI-----H--S----  -T--HWGEP-   NASCHNAPEAI     ---DA--S-------KR-----
    Pundamilia nyererei α   XP_005749046   V------H--S----  -S--SDG MAGLEQ      SE-I     ---EA--IVS--I--KR-----
    Maylandia zebra α       XP_004560486   V------H--S----  -S--SDG VAGLEQ      SE-I     ---EA--IVS--I--KR-----
ι   Homo sapiens ι          XP_016868274   F----RYCA--MP--                              NPGDHHD-EP-RHD-GYI--I-
    Rattus novergicus ι     NP_942077      F----RYCA--MP--                              NPGDHHD-EP-RHD-GYI--I-
    Serinus canaria ι       XP_009092397   F----RYCA--MP--                              NPGDHRD-EP-RHD-GYI--I-
    Python bivittatus ι     XP_015678872   F----RYCA--MP--                              NPGDHRE-EP-RHD-GYI--I-
    Pundamilia nyererei ι   XP_005747329   F----RYCA--MP--                              NTGDHRD-EP-RHD-GCI--I-
    Maylandia zebra ι       XP_004573783   F----RYCA--MP--                              NTGDHRD-EP-RHD-GCI--I-
ζ   Homo sapiens ζ          NP_001186196   F----RYCA--MP--                              HPGEHHD-EP-RHD-GY---I-
    Rattus novergicus ζ     NP_112405      F----RYCA--MP--                              HPGEHHD-EP-RHD-GY---I-
    Serinus canaria ζ       XP_009100702   F----RYCA--MP--                              NPGEHHD-EP-RHD-GCI--I-
    Python bivittatus ζ     XP_007429871   F----RYCA--MP--                              NPGEHHD-EP-RHD-GC---I-
    Xenopus tropicalis ζ    XP_002934142   F----RYCA--MP--                              NPGEHHD-EP-RHD-GCI--I-
    Pundamilia nyererei ζ   XP_013770097   F----RYCA--TP--                              NPSEHQD-EP-RHD-GCI--I-
    Maylandia zebra ζ       XP_014263169   F----RYCA--TP--                              NPSEHQD-EP-RHD-GCI--I-
η   Homo sapiens η          NP_001284358   V-----YA----F--                              GTKEDDI-AAPSFD-KI----A
    Rattus novergicus η     XP_006222097   V-----YA----F--                              GTKEDDI-AAPSFD-KI----A
    Serinus canaria η       XP_009101226   V-----YA----F--                              GTKEDDI-GAPSFD-KI----A
    Python bivittatus η     XP_007425896   V-----YA----F--                              GTKEDDI-GAPSFD-KI----A
    Xenopus tropicalis η    XP_012813202   V-----YA----F--                              GTKEDDI-GAPSFD-KI----A
    Maylandia zebra η       XP_004573039   V-----YA----F--                              GTKEDDI--APSFD-KI----A
δ   Homo sapiens δ          XP_011510341   V-----YA----F--                              GTKEDDT-AAPSFD-KI----A
    Rattus novergicus δ     XP_006226999   V-----YA----F--                              GTKEDDT-AAPSFD-KI----A
    Serinus canaria δ       XP_009087492   V-----YA----F--                              GTKEDDT-TAPSFD-KI----A
    Python bivittatus δ     XP_015745911   V-----YA----F--                              GTKEDDT-TAPSFD-KI----A
    Xenopus tropicalis δ    XP_012826605   V-----YA----F--                              GTKEDDI-GAPSFD-KI----A
    Maylandia zebra δ       XP_012780570   V-----YA----F--                              GTKEDDT-TAPSFD-KI----A
κ   Homo sapiens κ          XP_016884757   V---T-YA--I-F--                              SNTATTEYEAPAID-GK----A
    Rattus novergicus κ     XP_008771334   V---T-YA--V-F--                              SNTATTEYEAPAID-GK----A
    Sturnus vulgaris κ      XP_014746482   V-----YA--I-F--                              GTKEDNN-GAPSFD-KK----A
    Proto. mucrosquamatus κ XP_015687893   V-----YA--I-F--                              GTKEDSN-GAPSFD-KK----A
ε   Homo sapiens ε          NP_003638      V---GYWG--CR--E                              GMGDE TYPLARHD-G------
    Rattus novergicus ε     NP_001034430   V---GYWG--CR--E                              GVGDE TYPLARHD-G------
    Serinus canaria ε       XP_009092615   V---GYWG--CR--E                              GMGDE PYPLARHD-G------
    Python bivittatus ε     XP_007428710   V---AYWG--CR--E                              GMGDE LYPLARHD-G------
    Xenopus tropicalis ε    XP_002937855   V---GYWG--CR--E                              GMGDE PYPLSRHD-G------
    Pundamilia nyererei ε   XP_005737401   VC--GYWG--CR--E                              GMGDE PCPPTR-D-G------
    Maylandia zebra ε       XP_004538850   VC--GYWG--CR--E                              GMGDE PCPPTR-D-G------
θ   Homo sapiens θ          NP_001338      FI----WGS-AD---                              SDSDT R-EKPRMD-G------
    Rattus novergicus θ     NP_001185733   FI----WGS-AD---                              SDSDS R-EKPRID-G------
    Serinus canaria θ       XP_009094351   FI----WGS-AD---                              SESDN R-EKPRID-G------
    Python bivittatus θ     XP_007441613   FI----WGS-AD---                              SDNDS R-EKPKID-G------
    Xenopus tropicalis θ    XP_012808691   F-----WGS-AD---                              SDNDN RYGKPRID-G------
    Pundamilia nyererei θ   XP_005737796   F-----WGS-AD---                              SEVDS RYEKPSID-G------
    Maylandia zebra θ       XP_004552088   F-----WGS-AD---                              SEVDS RYEKPSID-G------
```

Fig 3. Excerpts from the sequence alignment of DGK isozymes showing a conserved insert (marked as Sig ❶) within the catalytic domain that is a unique characteristic of the Class I DGK homologs. The insert shown is present in a conserved region as the dashes (-) in the alignment denote identity with the amino acid shown on the top line. Sequence information is shown for only 1–2 representative species from different main groups within the vertebrates. However, this insert is specific for the Class I DGK homologs from all other vertebrates species examined. The second column shows the Genbank ID numbers of different sequences. The numbers on top of the alignment show the location of the depicted sequence in two reference DGK sequences. The first number outside of the parenthesis indicates the sequence position in the human DGK-γ homolog (accession number NP_001337.2) shown on the first line of the alignment, whereas the numbers within the parenthesis correspond to the location of this sequence in athe reference human DGK-ε homolog (accession number NP_003638). The vertebrate species for which the sequences are shown include the following groups: Mammals—*Homo sapiens* and *Rattus novergicus*; Birds—*Serinus canaria* and *Sturnus vulgaris*; Reptiles—*Protobothrops mucrosquamatus* and *Python bivittatus*; Amphibians—*Xenopus tropicalis* and Fish—*Pundamilia nyererei* and *Maylandia zebra*.

and amphibian species. In addition to the CSI shown in Fig 3, DGK homologs contain one additional 2 aa conserved insert (marked Sig ❷A in Fig 4), which is also commonly shared by the class I DGK homologs. In the same position, where this 2 aa CSI is found in class I homologs, a 4 aa insert (marked as Sig ❷B) is present in the DGK-i and DGK-ζ homologs (class IV isozymes). These signatures serve to distinguish the Class I and Class IV DGK homologs from all others and the presence of the inserts in the same position only in these two classes of homologs suggests that they may be more closely related to each other than to the other DGK classes. In close proximity of signatures ❷A,B, another CSI consisting of 2 aa deletion (denoted by ❸ and highlighted in light blue color) is uniquely present in the DGK-θ homologs.

Our analysis has also identified another CSI consisting of a 1 aa deletion in a highly conserved region (denoted as Sig ❹ in Fig 5), which is commonly shared by all of the Class I and Class IV DGK homologs. This CSI, in conjunction with the signatures ❷A,B (Fig 4) and the grouping together of the members of these two classes in the phylogenetic tree (Fig 2), provides evidence that the homologs from these two classes are specifically related to each other. In the proximity of Sig ❹, DGK-θ homologs also contain a 1 aa deletion (marked as Sig ❺, shown in light green color). Although, the CSIs ❹ and ❺ are placed in different positions in Fig 5, due to limited sequence similarity observed for the DGK-θ homologs in the vicinity of this indel, it is possible that the indel in DGK-θ (marked as Sig ❺) also occurred in the same position as in the Class I and Class IV DGK isozymes. In that case, the indel marked as ❹ would be a shared characteristic of the Class I, Class IV and Class V DGK isozymes.

The class II DGK homologs are known to contain a large insert in the middle of the catalytic domain. In the same location, where this large insert is present in the class II homologs, the DGK homologs belonging to other classes are also found to contain smaller CSIs of specific lengths that serve to distinguish them from each other (Fig 6). The Class I DGK homologs contain two different CSIs (a 5 aa deletion and a 2 aa deletion, denoted as Sig ❻) in this region whereas the DGK-ε homologs (class III) all have an 8 aa deletion in this position (marked as Sig ❽). The location of the large insert in the Class II DGK homologs is marked by Sig ❼. In addition, the DGK-θ and DGK-i isozymes can also be distinguished from others by the presence of 1 aa CSIs (deletions) that appear specific for these isoforms. Lastly, two additional CSIs are present in the catalytic domain in the proximity of the predicted ATP-binding site [2,31,32] shown in the sequence alignment in Fig 7. One of these CSIs, a 2 aa insert, is specific for the DGK-ε homologs (marked as Sig ❾), whereas in a nearby region, both DGK-θ as well as DGK-ε are found to contain CSIs of different lengths in the same position (marked as Sig ❿A, B). However, the sequence of the ATP binding motif (G-X-G-X-X-G) is nearly perfectly conserved in all DGK-isozymes in vertebrates, with the exception of DGK-κ isoform, where a Ser is present in place of the last Gly residue in the mammalian homologs. A number of other residues flanking the ATP-binding motif are also either completely or highly conserved in all DGK isozymes.

Distribution of DGK isozymes within the domain Eukarya

Much of the work on DGK isozymes has been carried out in mammalian systems and is limited to humans and rodent species. In addition, somework has been carried out on documenting and studying the functions of DGK isozymes in other eukaryotic organisms such as Drosophila, nematodes, slime molds and plants [3,6,10,14–16,33–35]. However, it is unclear whether all of the different DGK isozymes are present in different vertebrate classes and other deeper branching eukaryotic phyla. To gain insights in these regards, we have examined the species distributions of various DGK isozymes in different eukaryotic kingdoms and phyla. These studies were carried out as described in the Methods section and their results are
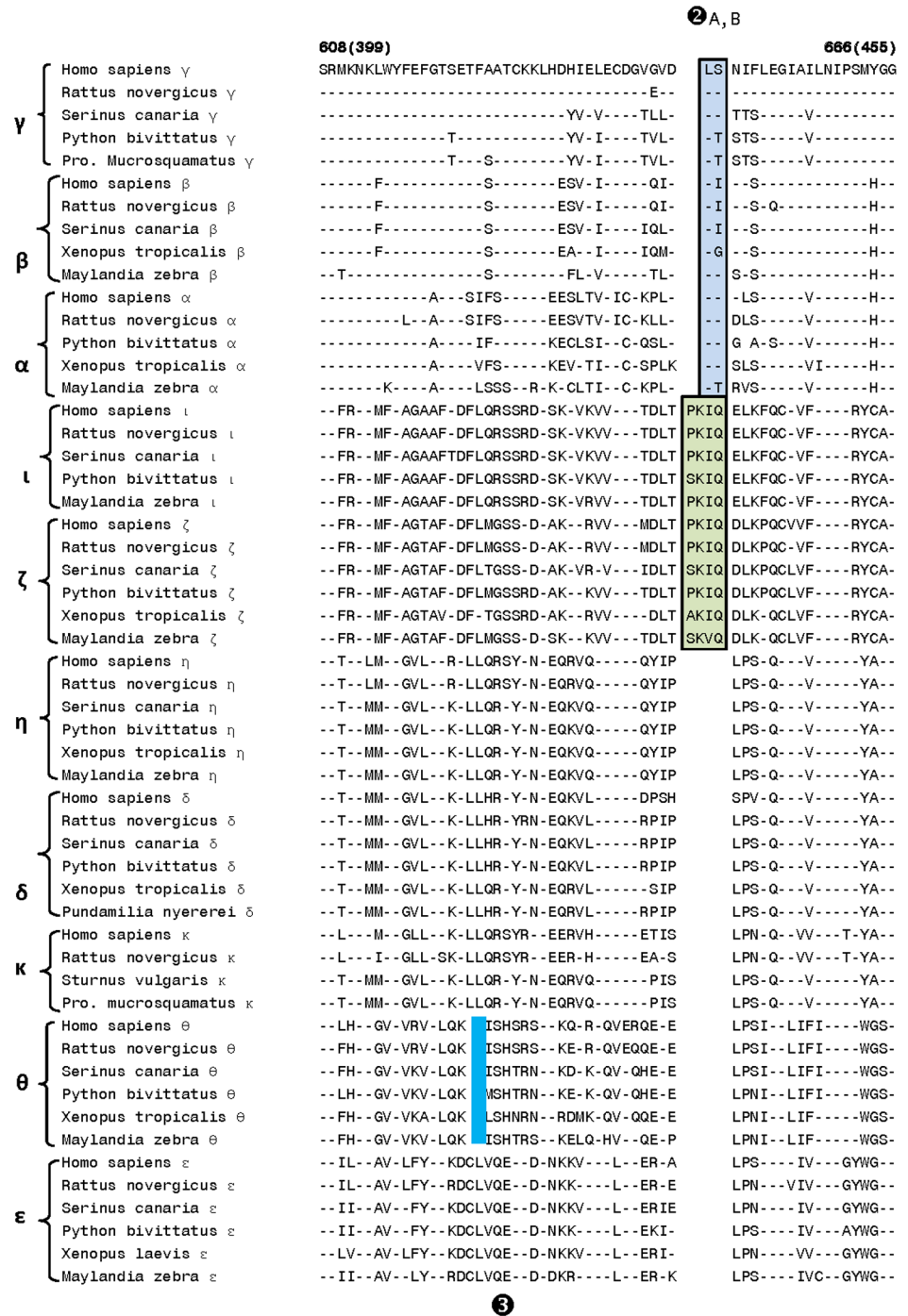
Fig 4. Partial sequence alignment of the DGK catalytic domain showing a number of signatures that are specific for different DGK isozymes. In the position marked Sig❷A,B, a 2 aa insert is specifically present in all class I DGK homologs, whereas in the same position a 4 aa insert is found in the class II DGK isozymes. Another signature marked ❸ consists of a 2 aa deletion that is uniquely present in the DGK-θ homologs. The signatures shown are specific for the indicated DGK isoforms in sequenced vertebrate species. The dashes (-) in the alignment denote identity with the amino acid on the top and the Genbank ID numbers for these sequences are provided in Fig 3.

https://doi.org/10.1371/journal.pone.0182758.g004

```
                                   518(309)             ❹              559(351)
                              PLGTGNDLARCLRWGGGYE    GGSLT K ILKDIEQSPLVMLDRWH
       ┌ Homo sapiens γ       -------------------    ----- - ---E-----------Y
       │ Rattus norvegicus γ  -------------------    --N-M - V-----H-TE-------
  γ    │ Serinus canaria γ    -------------------    --N-I - ------H-TE------Q
       └ Python bivittatus γ  -------------------    -EN-M - ------N-TEI-----K
       ┌ Homo sapiens β       -------------------    -EN-M - ------S-TEI-----K
       │ Rattus norvegicus β  -------------------    -EN-M - ------N-SEIL----K
  β    │ Serinus canaria β    -------------------    -EN-L - ------N-TEIL----K
       │ Python bivittatus β  -------------------    -EN-M - F-----IATV-L----K
       │ Xenopus tropicalis β -------------------    ----V - V-R---H-TE-V----
       └ Maylandia zebra β    -------------------
       ┌ Homo sapiens α       -------------------    -QN-A - ----L-M-KV-HM---S
       │ Rattus norvegicus α  ---------------R---    -EN-R - ------I-KV-Y----
  α    │ Python bivittatus α  -----------------D     -EN-V - ------A-SILQM---S
       │ Xenopus tropicalis α ------------S-----D     -EN-M - F-----K-VPIKM---N
       └ Maylandia zebra α    -----------------D     -SD-K E -M-E--G-E--PM---S
       ┌ Homo sapiens ι       ----------T-N-----T    DEPVS - --CQV-DGTV-Q----N
       │ Rattus norvegicus ι  ----------T-N-----T    DEPVS - --CQV-DGTI-Q----N
  ι    │ Serinus canaria ι    --------T-N-----T      DEPVS - --CHV-DGTI-Q----N
       │ Python bivittatus ι  ----------T-N-----T    DEPVA - --CHV-DGAI-Q----N
       │ Xenopus tropicalis ι ----------T-N-----T    DEPVS - --CHV-DGTN-Q----N
       └ Maylandia zebra ι    ----------T-N-----T    DEPVS - V-CHV-DGSV-Q----N
       ┌ Homo sapiens ζ       ----------T-N-----T    DEPVS - --SHV-EGNV-Q----D
       │ Rattus norvegicus ζ  ----------T-N-----T    DEPVS - --SHV-EGNV-Q----D
  ζ    │ Serinus canaria ζ    ----------T-N-----T    DEP-S - --SHV-DGNI-Q----N
       │ Python bivittatus ζ  ----------T-N-----T    DEP-S - --SHV-EGEI-Q----N
       │ Xenopus tropicalis ζ ----------T-N-----T    DEP-S - --SHV-EGIV-Q----N
       └ Maylandia zebra ζ    ----------T-N-----T    DEPVS - V-CHV-DGSV-Q----N
       ┌ Homo sapiens η       ----------V-G---S-D D DTQ-P Q --EKL-RASTK-----S
       │ Rattus norvegicus η  ----------V-G---S-D D DTQ-P Q --EKL-RASTK-----S
  η    │ Serinus canaria η    ----------V-G---SCD D DTQ-P Q --EKL-RASTK-----S
       │ Python bivittatus η  ----------V-G---SCD D DTQ-P Q --EKL-RASTK-----S
       │ Xenopus tropicalis η ----------V-G--ASCD D DTQ-P Q --EKL-RASTK-----S
       └ Maylandia zebra η    ----------V-G--PSCD D DTQ-P Q --EKL-RASTK-----S
       ┌ Homo sapiens δ       ----------V-G--SACD D DTQ-P Q --EKL-RASTK-----S
       │ Rattus norvegicus δ  ----------V-G--SACD D DTQ-P Q --AKL-RASTK-----S
  δ    │ Serinus canaria δ    ----------V-G--SACD D DTQ-P Q --EKL-RASTK-----S
       │ Python bivittatus δ  ----------V-G--SACD D DTQ-P Q --EKL-RASTK-----S
       │ Xenopus tropicalis δ ----------V-G--SACD D DTQ-P Q --EKL-RAGTK-----S
       └ Maylandia zebra δ    ----------V-G--SACD D DTQ-P Q --EKL-RASTK-----S
       ┌ Homo sapiens κ       ----------V-G--AFWN K SK-PL D --NRV--ASVRI----S
  κ    │ Rattus norvegicus κ  ----------V-G--AFWS K -K-PL E --SRV--AHVRI----S
       │ Sturnus vulgaris κ   ----------V-G--SLCD D DTQ-L Q --EKL-RATTK-----S
       └ Pro. mucrosquamatus κ ----------V-G--SLCD D DTQ-L Q V-EKL-RATTK-----S
       ┌ Homo sapiens ε       --------G-V----A--S G EDPFS   V-LSVDEADA-LM---T
       │ Rattus norvegicus ε  --------G-V----A--S G EDPFS   V-VSVDEADA-LM---T
       │ Serinus canaria ε    --------G-V----A--S G EDPFS   V-LSVDEADA-LV---T
  θ    │ Python bivittatus ε  --------G-V----A--S G EDPYS   --VSVDEADD-LM---T
       │ Xenopus tropicalis ε --------G-V----A--S G EDPYS   --ISVNEAEN-LM---T
       │ Pundamilia nyererei ε ----------I----P--S G EDPYH   --VSVYEADEILM---T
       └ Maylandia zebra ε    ----------I----P--S G EDPYH   --VSVYEADEILM---T
       ┌ Homo sapiens θ       --------SNT-G--T--A G EIPVA Q V-RNVMEADGIK----K
       │ Rattus norvegicus θ  --------SNT-G--T--A G EIPVA Q V-RNVMEADGIK----K
  ε    │ Serinus canaria θ    --------SNT-G--A--A G EVPVE Q --RNVMEADGIK----K
       │ Python bivittatus θ  --------SNT-G--A--A G EVPVE Q --HN-MDADGIK----K
       │ Xenopus tropicalis θ ---------NT-G--A--A G DVPVE Q --RN-MDADSIK----K
       └ Maylandia zebra θ    --------SNT-G--A--A G EIPVE Q V-RN-LDAEV-KM---K
```
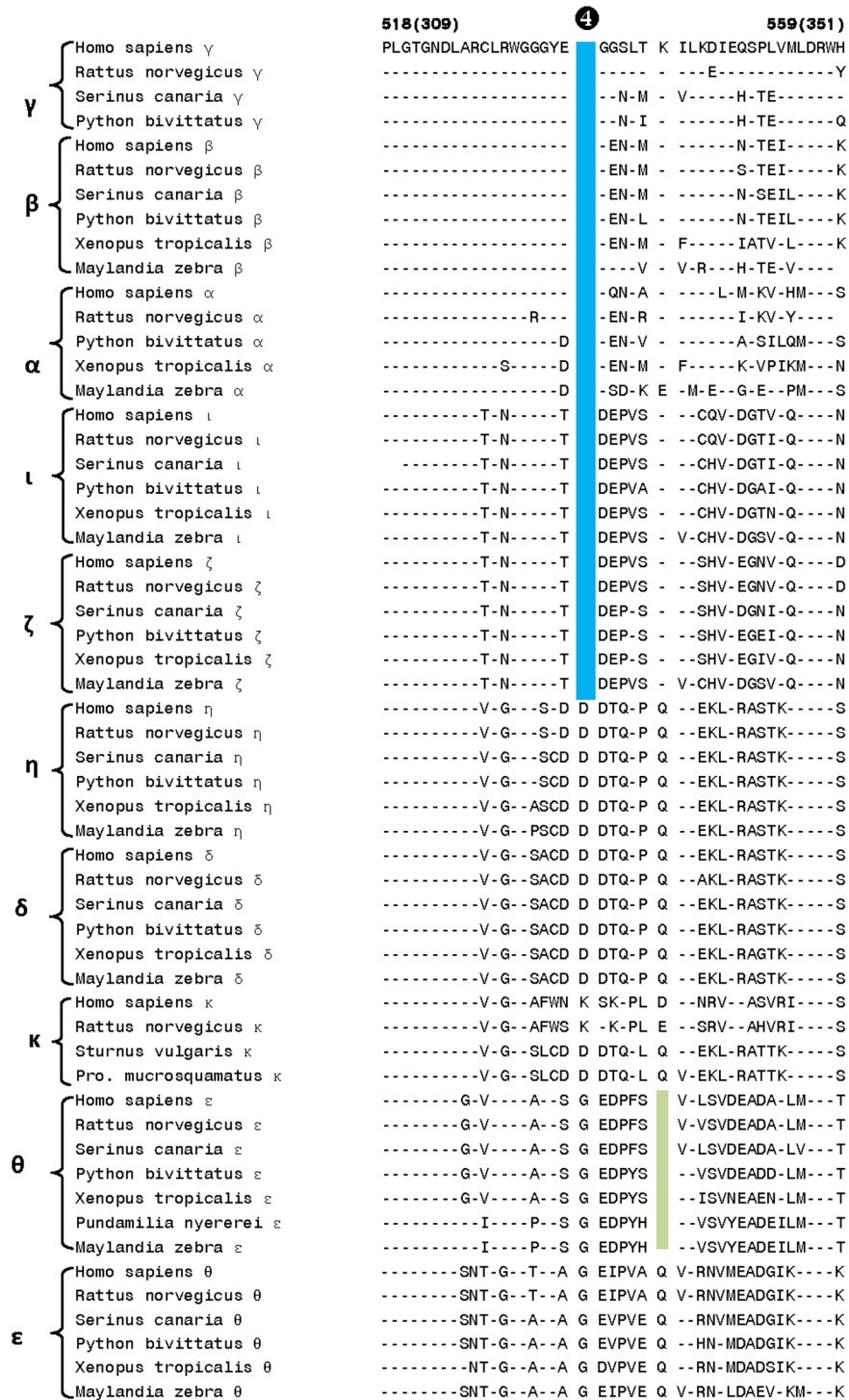
Fig 5. Excerpts from the multiple sequence alignment of catalytic domain from DGK isozymes showing a 1 aa conserved indel (deletion) (marked ❹) that is commonly shared by the class I and class IV isozymes. In the proximity of Sig ❹, a 1 aa deletion is also present in the DGK-θ homologs (marked as signature ❺). Although these two indels are placed in different positions, the possibility that they have occurred in the same position as ❹ cannot be excluded.

```
                              543(335)              ⑥           ⑥                        609(400)
    ┌ Homo sapiens      γ    ILKDIEQSPLVMLDRWHLEVIPRE   EVENGDG   VPYSIMNNYFSIGVDASIAHRFHVMREKHPEKFNSR
    │ Rattus norvegicus γ    ---E-----------Y---M---    -------   ---N--------------------------
  γ ┤ Serinus canaria   γ    V-----H-TE-------IDI--TD   REA---F   --ST-I----------------I----------
    │ Python bivittatus γ    ------H-TE------QID---ND   KE----F   -----I----------------L----------
    └ Pro. Mucrosquamatus γ  ------H-TE------QID---ND   KE----F   -----I----------------L----------
    ┌ Homo sapiens      β    ------N-TEI-----KF----ND   KD-K--F   -----I----------------I----------
    │ Rattus norvegicus β    ------S-TEI-----KF--T-ND   KD-K--F   -----I----------------I----------
  β ┤ Serinus canaria   β    ------N-SEIL----KF----ND   KD-K--F   ---N-I----------------I----------
    │ Python bivittatus β    ------N-TEIL----KF----ND   KD-K--F   -----I----------------I----------
    │ Xenopus tropicalis β   -----IATV-L----KID---ND    KD-K--F   -----I----------------L----------
    └ Maylandia zebra   β    --R---N-TE------KID-T-TD   KE-R--F   -----I----------------
    ┌ Homo sapiens      α    ----L-M-KV-HM---SV----QG   TE-KS-F   --FQ-I----------------I----Y------
    │ Rattus norvegicus α    ------I-KV-Y----L-----QG   NG-KS-F   --SQ-I----------------L----Y------
  α ┤ Python bivittatus α    ------A-SILQM---SVQ-M-EN   PD-K--F   ---E-I----------------Y------
    │ Xenopus tropicalis α   -----K-VPIKM---NI----EN    PD-K--F   ---E-I----------------N----Y------
    └ Maylandia zebra   α    -M-E--G-E--PM---SIQ---ND   PH-A--F   --NE-I----------------S------QR----
    ┌ Homo sapiens      ι    --CQV-DGTV-Q----N-H-ERNPDLPPE-L-D-VCK   L-LNVF-----L-F--HVTLE--ES--AN-------
    │ Rattus norvegicus ι    --CQV-DGTI-Q----N-H-ERNPDLPPE-LE-GVCK   L-LNVF-----L-F--HVTLE--ES--AN-------
  ι ┤ Serinus canaria   ι    --CHV-DGTI-Q----N-H-ERNPDLPQD-L-D-ARK   L-LNVF-----L-F--HVTLE--ES--AN-------
    │ Python bivittatus ι    --CHV-DGAI-Q----N-Q-ERNPDLPPN-L-D-A-K   L-L-VF-----L-F--HVTLE--ES--AN-------
    │ Xenopus tropicalis ι   --CHV-DGTN-Q----N-H-ERNPDLLHE-LDD-THK   L-LNVF-----L-F--HVTLE--ES-
    └ Maylandia zebra   ι    V-CHV-DGSV-Q----N-L-EKSS       AQP-E-T-K   L-LNVF-----L-F--HVTLE--ES--AN-------
    ┌ Homo sapiens      ζ    --SHV-EGNV-Q----D-HAE-NPEAGPEDRDE-ATDRL-LDVF-----L-F--HVTLE--ES--AN-------
    │ Rattus norvegicus ζ    --SHV-EGNV-Q----D-RAE-NPEAGPE-RD-GATDRL-LDVF-----L-F--HVTLE--ES--AN-------
  ζ ┤ Serinus canaria   ζ    --SHV-DGNI-Q----N-H-E-NPDTNPE-KDEAAADKL-LDVF-----L-F--RVTLE--ES--AN-------
    │ Python bivittatus ζ    --SHV-EGEI-Q----N-Q-D-TPEGNPE-KDETATDKL-LDVF-----L-F--RVTLE--ES--AN-------
    │ Xenopus tropicalis ζ   --SHV-EGIV-Q----N-V-ERNPEAWEDDKGD-ATDKL-LDVF-----L-F--HVTLE--ES--AN-------
    └ Maylandia zebra   ζ    --SHV-DGNI-Q----N-N-EANPEARPEDQDEHQTDKL-IDVF-----L-F--HVTLG--ES--AN-------
    ┌ Homo sapiens      η    --EKL-RASTK-----SI-K                                 CV-----G--L--K-SLE-NNK--E----CR--
    │ Rattus norvegicus η    --EKL-RASTK-----SI-K                                 CV-----G--L--K-SLE-NNK--E----CR--
  η ┤ Serinus canaria   η    --EKL-RASTK-----SI-K                                 CV-----G--L--K-SLE-NNK--E----CR--
    │ Python bivittatus η    --EKL-RASTK-----SI-K          ⑦                      CV-----G--L--K-SLE-NNK--E----CR--
    │ Xenopus tropicalis η   --EKL-RASTK-----SI-K    Large insert in              CV-----G--L--K-SLE-NNK--E----CR--
    └ Maylandia zebra   η    --EKL-RASTK-----SI-K      Class II DGK               CV-----G--L--K-SLE-NNK--E----CR--
    ┌ Homo sapiens      δ    --EKL-RASTK-----SV-K        homologs                 CV-----G--L--K-SLD-NNK-DE----CR--
    │ Rattus norvegicus δ    --AKL-RASTK-----SV-K                                 CV-----G--L--K-SLD-NNK-DE----CR--
  δ ┤ Serinus canaria   δ    --EKL-RASTK-----SI-K                                 CV-----G--L--K-SLD-NNK-DE----CR--
    │ Python bivittatus δ    --EKL-RASTK-----SI-K                                 CV-----G--L--K-SLD-NNK-DE----CR--
    │ Xenopus tropicalis δ   --EKL-RAGTK-----SI-K                                 CV-----G--L--K-SLD-NNK-DE--K-CR--
    └ Maylandia zebra   δ    --EKL-RASTK-----SI-K                                 CV-----G--L--K-SLD-NNK-DE----CR--
    ┌ Homo sapiens      κ    --NRV--ASVRI----SV-K                                 CV-----G--L--K-SLD-NTR-DE--GQY---
    │ Rattus norvegicus κ    --SRV--AHVRI----SV-R                                 CV-----G--L--K-SLE-NAR--E---QY---
  κ ┤ Serinus canaria   κ    --EKL-RASTK-----SI-K                                 CV-----G--L--K-SLD-NNK-DE----CR--
    │ Sturnus vulgaris   κ    --EKL-RATTK-----SV-K                                 CV-----G--L--K-SLE-NNK-DE--K-CS--
    └ Pro. mucrosquamatus κ  V-EKL-RATTK-----SV-K                                 CV-----G--L--K-SLE-NNK-DE--K-CS--
    ┌ Homo sapiens      ε    V-RNVMEADGIK----KVQ-TNKG                   YYNLRKPKEFT------V-P--LM-LN--AH---A-SL-S--
    │ Rattus norvegicus ε    V-RNVMEADGIK----KVQ-TNKG                   YYSLRKPKEFT--------P--LM-LN--AH---A-SL-S--
  ε ┤ Serinus canaria   ε    --RNVMEADGIK----KVQ-TNKG      ⑧            YYNLRKPKVFT------V-P--LM-LN--AH---T-SL-S--
    │ Python bivittatus ε    --HN-MDADGIK----KVQ-TNKG                   YYNLRKLKVFT------V-P--LM-LN--TH---T-SL-T--
    │ Xenopus tropicalis ε   --RN-MDADGIK----KVQ-TNKG                   Y SLRKPKVLS------V-P--LM-LN--TH---T-SL-S--
    └ Maylandia zebra   ε    V-RN-LDAEV-KM---KVQ-ASKG                   VY FRKPKVLS------V-P--LM-LS--AH---T-SF-S--
    ┌ Homo sapiens      θ    V-LSVDEADA-LM---TILLDAH-AGSAENDTADAEF   PKIVQ-S--CG--I--ELSLD--QA--EE-G--T--
  θ ┤ Rattus norvegicus θ    V-VSVDEADA-LM---TILLDAH-IDSTENNVVETEF   PKIVQ----CG--I--ELSLD--QA--EE-G--T--
    │ Pundamilia nyererei θ  --VSVYEADEILM---TILLDAQDVSEDGKENEFLEF   PKIVQ-----GL-I--ELSLD--LA--DE-D--T--
    └ Maylandia zebra   θ    --VSVYEADEILM---TILLDAQDVSEDGKENDFLEF   PKIVQ-----GL-I--ELSLD--LA--DE-D--T--
```
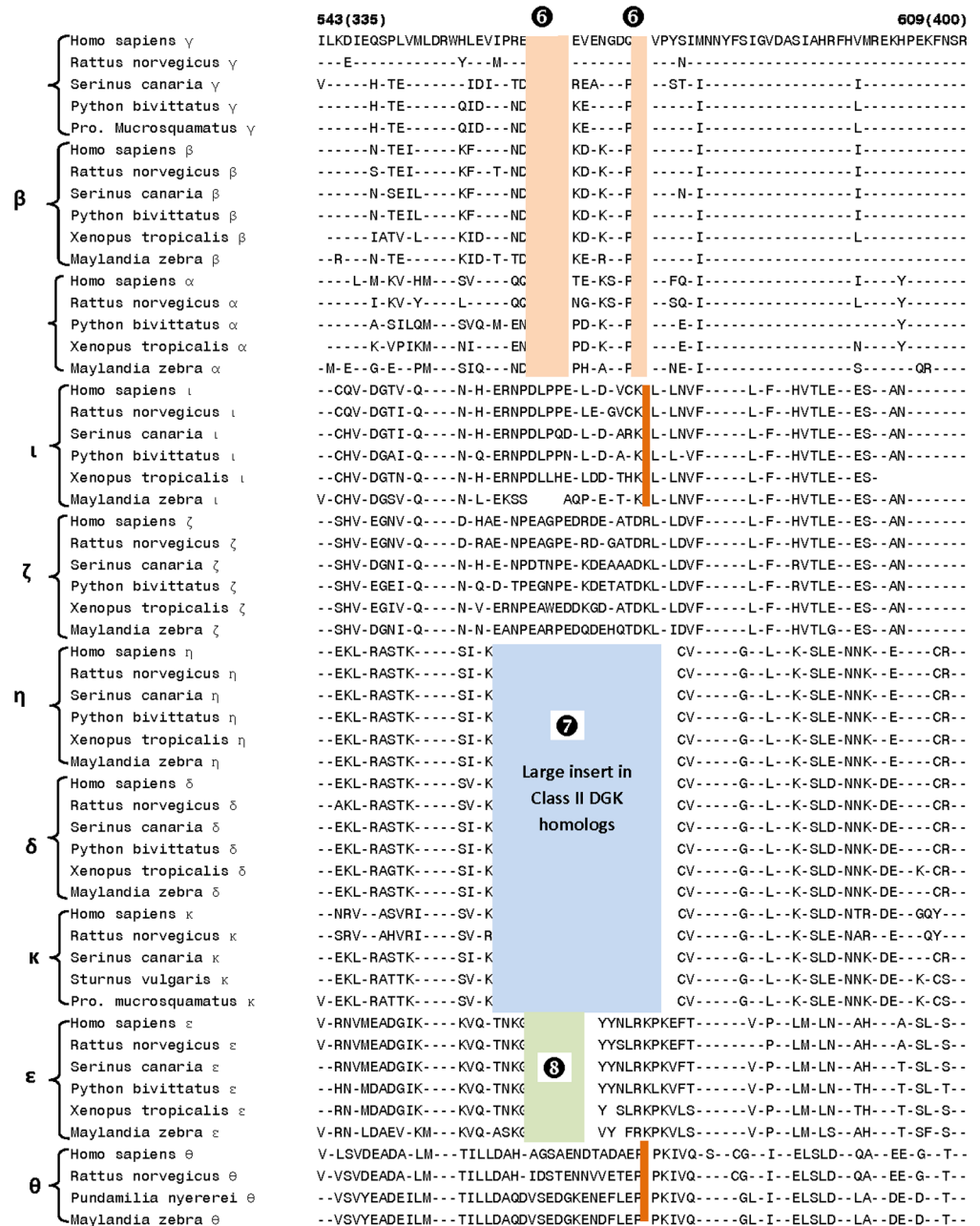
Fig 6. Partial sequence alignment of DGK isozymes showing a region of the catalytic domain where indels of different lengths are present in different DGK isozymes. Signatures marked ⑥ are specific for class I DGKs, the large deletion marked ⑦ is a shared characteristic of class II DGKs and the CSI marked as ⑧ is present in different DGK-θ homologs.

summarized in Table 1. The subphylum vertebrata, which encompasses all of the large animal species, contains five main groups: fishes, amphibians, birds, reptiles and mammals. Of the known DGK isozymes, homologs of DGK- α, -β, -δ, -η, -ζ, -θ and -ε were broadly distributed in all major groups within the vertebrates. While homologs of DGK-κ were mainly found in mammals, birds and some fishes and reptiles, homologs of DGK-γ were not detected in fishes and amphibians. In the invertebrate phyla that were examined (viz. Arthropoda, Nematoda,
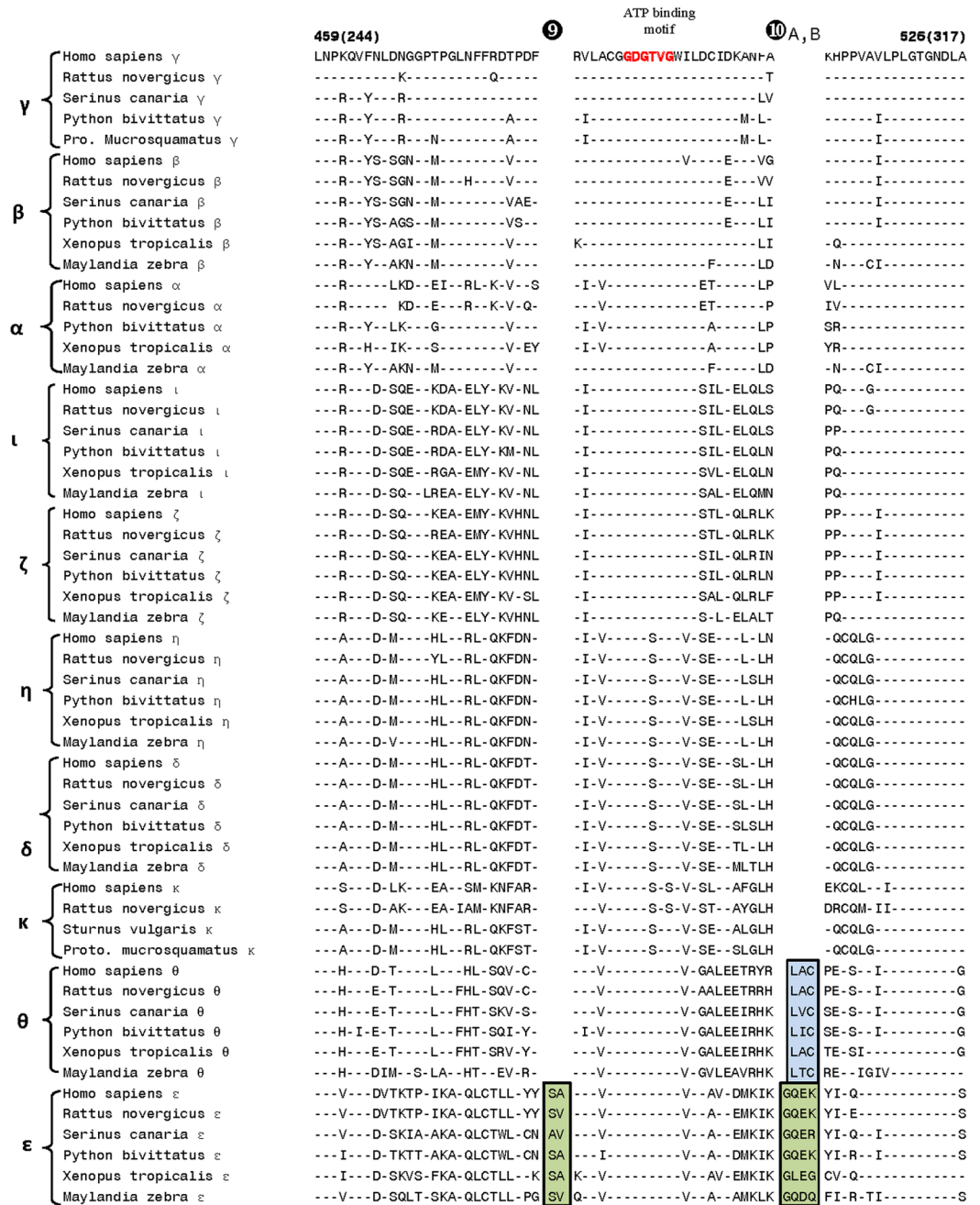
```
                              459(244)              ⑨       ATP binding      ⑩A,B        526(317)
                                                            motif
         Homo sapiens      γ   LNPKQVFNLDNGGPTPGLNFFRDTPDF        RVLACGGDGTVGWILDCIDKANFA        KHPPVAVLPLGTGNDLA
         Rattus novergicus γ   ----------K---------Q-----        ----------------------T        -----------------
  γ      Serinus canaria   γ   ---R--Y---R---------------        ----------------------LV       -----------------
         Python bivittatus γ   ---R--Y---R-----------A---        -I------------------M-L-       ------I----------
         Pro. Mucrosquamatus γ ---R--Y---R--N--------A---        -I------------------M-L-       ------I----------
         Homo sapiens      β   ---R--YS-SGN--M--------V---        ------------V----E---VG        ------I----------
         Rattus novergicus β   ---R--YS-SGN--M---H----V---        ------------------E---VV       ------I----------
  β      Serinus canaria   β   ---R--YS-SGN--M-------VAE-        ------------------E---LI       ------I----------
         Python bivittatus β   ---R--YS-AGS--M--------VS--        ------------------E---LI       ------I----------
         Xenopus tropicalis β  ---R--YS-AGI--M--------V---        K-------------------LI         -Q---------------
         Maylandia zebra   β   ---R--Y--AKN--M--------V---        ------------------F-----LD     -N---CI----------
         Homo sapiens      α   ---R-----LKD--EI--RL-K-V--S       -I-V-----------ET-----LP       VL---------------
         Rattus novergicus α   ---R----- KD--E---R--K-V-Q-       ---V-----------ET-----P        IV---------------
  α      Python bivittatus α   ---R--Y--LK---G--------V---        -I-V-----------A-----LP        SR---------------
         Xenopus tropicalis α  ---R--H--IK---S--------V-EY       -I-V-----------A-----LP        YR---------------
         Maylandia zebra   α   ---R--Y--AKN--M--------V---        ---------------F-----LD        -N---CI----------
         Homo sapiens      ι   ---R---D-SQE--KDA-ELY-KV-NL       -I------------SIL-ELQLS        PQ---G-----------
         Rattus novergicus ι   ---R---D-SQE--KDA-ELY-KV-NL       -I------------SIL-ELQLS        PQ---G-----------
  ι      Serinus canaria   ι   ---R---D-SQE--RDA-ELY-KV-NL       -I------------SIL-ELQLS        PP---------------
         Python bivittatus ι   ---R---D-SQE--RDA-ELY-KM-NL       -I------------SIL-ELQLN        PQ---------------
         Xenopus tropicalis ι  ---R---D-SQE--RGA-EMY-KV-NL       -I------------SVL-ELQLN        PQ---------------
         Maylandia zebra   ι   ---R---D-SQ--LREA-ELY-KV-NL       -I------------SAL-ELQMN        PQ---------------
         Homo sapiens      ζ   ---R---D-SQ---KEA-EMY-KVHNL       -I------------STL-QLRLK        PP----I----------
         Rattus novergicus ζ   ---R---D-SQ---REA-EMY-KVHNL       -I------------STL-QLRLK        PP----I----------
  ζ      Serinus canaria   ζ   ---R---D-SQ---KEA-ELY-KVHNL       -I------------SIL-QLRIN        PP----I----------
         Python bivittatus ζ   ---R---D-SQ---KEA-ELY-KVHNL       -I------------SIL-QLRLN        PP----I----------
         Xenopus tropicalis ζ  ---R---D-SQ---KEA-EMY-KV-SL       -I------------SAL-QLRLF        PP----I----------
         Maylandia zebra   ζ   ---R---D-SQ---KE--ELY-KVHNL       -I------------S-L-ELALT        PQ---------------
         Homo sapiens      η   ---A---D-M----HL--RL-QKFDN-       -I-V-----S---V-SE---L-LN       -QCQLG-----------
         Rattus novergicus η   ---A---D-M----YL--RL-QKFDN-       -I-V-----S---V-SE---L-LH       -QCQLG-----------
  η      Serinus canaria   η   ---A---D-M----HL--RL-QKFDN-       -I-V-----S---V-SE---LSLH       -QCQLG-----------
         Python bivittatus η   ---A---D-M----HL--RL-QKFDN-       -I-V-----S---V-SE---L-LH       -QCHLG-----------
         Xenopus tropicalis η  ---A---D-M----HL--RL-QKFDN-       -I-V-----S---V-SE---LSLH       -QCQLG-----------
         Maylandia zebra   η   ---A---D-V----HL--RL-QKFDN-       -I-V-----S---V-SE---L-LH       -QCQLG-----------
         Homo sapiens      δ   ---A---D-M----HL--RL-QKFDT-       -I-V-----S---V-SE--SL-LH       -QCQLG-----------
         Rattus novergicus δ   ---A---D-M----HL--RL-QKFDT-       -I-V-----S---V-SE--SL-LH       -QCQLG-----------
  δ      Serinus canaria   δ   ---A---D-M----HL--RL-QKFDT-       -I-V-----S---V-SE--SL-LH       -QCQLG-----------
         Python bivittatus δ   ---A---D-M----HL--RL-QKFDT-       -I-V-----S---V-SE--SLSLH       -QCQLG-----------
         Xenopus tropicalis δ  ---A---D-M----HL--RL-QKFDT-       -I-V-----S---V-SE--TL-LH       -QCQLG-----------
         Maylandia zebra   δ   ---A---D-M----HL--RL-QKFDT-       -I-V-----S---V-SE--MLTLH       -QCQLG-----------
         Homo sapiens      κ   ---S---D-LK---EA--SM-KNFAR-       -I-V-----S-S-V-SL--AFGLH       EKCQL--I---------
  κ      Rattus novergicus κ   ---S---D-AK---EA-IAM-KNFAR-       ---V-----S-S-V-ST--AYGLH       DRCQM-II---------
         Sturnus vulgaris  κ   ---A---D-M----HL--RL-QKFST-       -I-V-----S---V-SE-ALGLH        -QCQLG-----------
         Proto. mucrosquamatus κ ---A---D-M----HL--RL-QKFST-     -I-V-----S---V-SE-SLGLH        -QCQLG-----------
         Homo sapiens      θ   ---H---D-T----L---HL-SQV-C-       ---V---------V-GALEETRYR  [LAC] PE-S--I---------G
         Rattus novergicus θ   ---H---E-T----L--FHL-SQV-C-       ---V---------V-AALEETRRH  [LAC] PE-S--I---------G
  θ      Serinus canaria   θ   ---H---E-T----L--FHT-SKV-S-       ---V---------V-GALEEIRHK  [LVC] SE-S--I---------G
         Python bivittatus θ   ---H-I-E-T----L--FHT-SQI-Y-       -I-V---------V-GALEEIRHK  [LIC] SE-S--I---------G
         Xenopus tropicalis θ  ---H---E-T----L--FHT-SRV-Y-       ---V---------V-GALEEIRHK  [LAC] TE-SI----------G
         Maylandia zebra   θ   ---H--DIM--S-LA--HT--EV-R-        ---V---------V-GVLEAVRHK  [LTC] RE--IGIV---------
         Homo sapiens      ε   ---V---DVTKTP-IKA-QLCTLL-YY [SA]  ---V---------V--AV-DMKIK  [GQEK] YI-Q-----------S
         Rattus novergicus ε   ---V---DVTKTP-IKA-QLCTLL-YY [SV]  ---V---------V-A--EMKIK   [GQEK] YI-E-----------S
  ε      Serinus canaria   ε   ---V---D-SKIA-AKA-QLCTWL-CN [AV]  ---V---------V-A--EMKIK   [GQER] YI-Q--I--------S
         Python bivittatus ε   ---I---D-TKTT-AKA-QLCTWL-CN [SA]  ---I---------V--A-DMKIK   [GQEK] YI-R--I--------S
         Xenopus tropicalis ε  ---I---D-SKVS-FKA-QLCTLL--K [SA]  K--V---------V--AV-EMKIK  [GLEG] CV-Q-----------
         Maylandia zebra   ε   ---V---D-SQLT-SKA-QLCTLL-PG [SV]  Q--V---------V--A--AMKLK  [GQDQ] FI-R-TI--------S
```

Fig 7. Excerpts from the sequence alignment of DGK isozymes showing a number of CSIs that are specific for the DGK-θ and DGK-ε isozymes. The 2 aa insert marked as Sig❾ is only found in DGK-ε homologs, whereas in a nearby region, both DGK-θ as well as DGK-ε are found to contain CSIs of different lengths in the same position (marked as Sig ❿A, B). These CSIs are located in the proximity of ATP-binding motif (shown in red), which is conserved in all isozymes.

Mollusca and Cnidarians), homologs that were most similar to DGK-ε, DGK-θ, DGK- ζ, DGK-β and DGK-δ were detected in either all or most species from phyla, whereas homologs for the other DGK isozymes were generally not detected or their presence could not be reliably ascertained. In accordance with earlier studies, homologs showing significant similarity to the DGK isozymes were not detected in Fungi [3,4,6], but in plants, homologs that were most

Table 1. Distribution of different DGK isozymes in the main phyla of vertebrates and other eukaryotes.

Eukaryotic Groups	DAG kinase isoforms									
	α	β	γ	δ	η	κ	ι	ζ	ε	θ
Mammals (144)	+	+	+	+	+	+	+	+	+	+
Amphibians (1)	+	+	-	+	+	-	+	+	+	+
Birds (61)	-[3]	+	+	+	+	+	+	+	+	+
Reptiles (11)	+	+	+	+	+/-	-[3]	+/-	+	+	+
Teleost Fishes (35)	+	+	-	+	+	+/-	+	+	+	+
Plants (81)	-[1]	+	+	-[1]	-[1]	-[1]	-[1]	-[1]	-[1]	-[1]
Fungi (704)	-	-[3]	-[3]	-	-	-	-	-	-	-
Plasmodium (45)	-[1]	-	+/-[2]	+/-[2]	-[1]	-[1]	-[1]	-[1]	-[1]	-[1]
Amoebozoa (16)	-[1]	-	+/-[1]	-[1]	-[1]	-[1]	-[1]	-[1]	+/-[1]	-[1]
Nematoda (47)	-	+	-	+/-[2]	-[1]	-[1]	+/-[2]	+/-[2]	+	+
Mollusca (5)	-	+	-	+	-	-	-	+	+	+
Cnidarians (5)	-	-	-[3]	-[3]	+/-	-	-	+	+	+
Arthropoda (107)	-[1]	+/-[2]	-[1]	-	+	-	+/-[1]	+/-[2]	+	+
Platyhelminthes (7)	-	-	-	-	-	-	+	+/-	-	+
Euglenozoa (47)	+/-[2]	-[1]	-[1]	-[1]	-[1]	-[1]	-[1]	-[1]	+/-[1]	-[1]

The presence or absence of different DGK isoforms in the indicated groups was determined based on blastp searches on the annotated genomes in the NCBI nr database as described in the Methods section. The numbers in parenthesis after the group names indicatedthe numbers of annotated genomes that were present in the NCBI nr database from the given group.

(+/-) indicates that only some species may contain the isoform.

[1]No isoform was specified however the top 5 hits show limited similarity to the indicated isoform.

[2]No isoform was specified however the top 5 hits are most similar to the indicated isoform.

[3]Limited similarity (less than 3 hits for the isoform of interest).

https://doi.org/10.1371/journal.pone.0182758.t001

similar to the β- and γ- isozymes were detected [3,10]. Blast searches were also carried out against members of a number of protists phyla (viz. Plasmodium, Euglenezoa, Amoebozoa), which form the deepest branching lineages within the eukaryotes. Except for an isolated hit showing limited similarity to these proteins, the presence of DGK-isozymes in these species could not be reliably ascertained.

Discussion

Most vertebrate species contain 10 different DGK isozymes which are grouped into 5 different classes or types (Fig 1). The DGKs from different classes are distinguished from each other by the presence or absence of specific functional domains; however, all of them share a large catalytic domain (CD), as well as two cysteine-rich domains. In the simplest of the DGK isozymes, DGK-ε (sole member of class III DGK), the catalytic domain (> 300 aa) makes up nearly 60% of the entire protein and thus it comprises the single largest sequence element specific for the DGK family of proteins. In contrast, the cysteine-rich C1 domains are of short lengths (~ 50 aa), show considerable sequence variation, and they are also present in numerous other proteins including the protein kinase C family of proteins [2–4,12,36]. Due to the prominence and uniqueness of the catalytic domain for the DGK family of proteins, analysis of its sequence provides an important resource for gaining an understanding of the evolution of this protein family as well as functional differences among its members. Our analyses of the catalytic domain sequences presented here have identified many novel sequence features that are specific for different types of DGKs and provide useful insights into the evolution of this protein family.
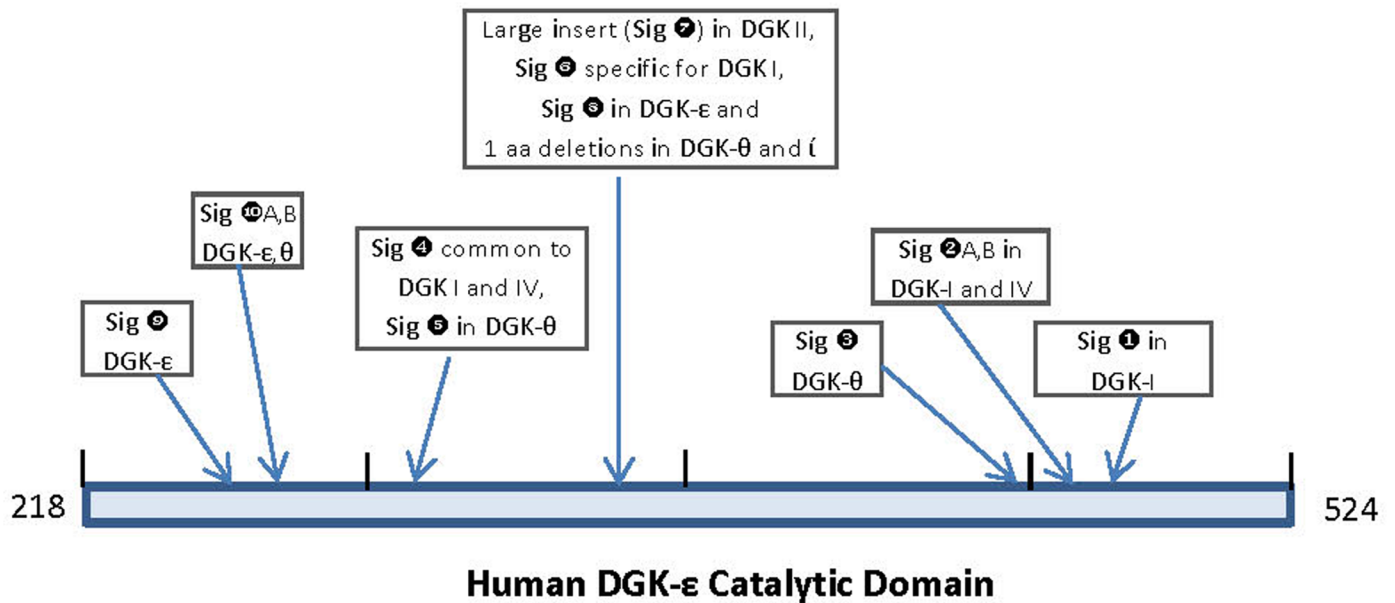
Fig 8. A summary diagram indicating the locations of the CSIs within the catalytic domain that are specific for different DGK classes. The locations of the CSIs are marked using the catalytic domain sequence for the human DGK-ε isoform (accession number NP_003638) as reference.

In a phylogenetic tree based on CD sequences, DGK homologs belonging to different classes formed strongly supported clusters, which were separated by long branches. The observed results indicate strongly that apart from the other known differences in different types of DGK homologs (see [1–6]) their catalytic domain sequences have also diverged considerably from each other and provide clear distinction among different types of DGK homologs. In contrast, the DGK isozymes within each DGK class were tightly clustered in the phylogenetic tree indicating that their catalytic domain sequences are very similar. A major focus of the present work was on identifying novel sequence features within the CD, which are specific for either a particular type of DGK homologs or which are commonly shared by members of different DGK classes. A large number of sequence motifs or signatures, in the forms of CSIs, have been identified in the present work and a summary of the identified CSIs, their DGK class specificity and location within the catalytic domain is provided in Fig 8. Locations of the different CSIs in the CD are indicated using the sequence of the human DGK-ε isoform as reference (accession number NP_003638).

As seen from Fig 8, DGK homologs of all five classes can be clearly distinguished from each other by the presence of specific CSIs that are uniquely found in them. For class I, class III, class IV and class V DGK homologs, multiple CSIs specific for these groups were identified. The described signatures are uniquely present in all members of the indicated classes from vertebrate species (the main focus of the present work) indicating that the genetic changes responsible for these CSIs first occurred in the ancestors of specific DGK isoforms, and they were then retained in the descendent species. Further, according to the definition of the CSI(s) used in our work, the large insert (>300 aa) within the class II DGK members [2–6,12], also represents a CSI (marked as Sig ❼) specific for the class II DGK homologs. The sequence region in the CD where this large CSI is found in class II homologs, it is flanked on both sides by several residues that are conserved in all DGK homologs. Within this conserved region, CSIs of specific lengths (Sig ❻, ❽ and ❽A) which are distinctive characteristics of the other classes of DGK homologs are also found. Based on the differences in the lengths of the CSIs in different

classes of DGK homologs, this region is predicted to specify important functional differences in the properties of different classes of DGK homologs. Another noteworthy CSI within the CD is a 16–24 aa insert (Sig ❶ in Fig 3) that is uniquely found in all of the Class I DGK homologs. Some of the identified CSIs were also commonly shared by members from two different DGK types. For example, the CSI in Sig ❹ (Fig 5) is commonly shared by class I and IV, and also possibly class V (Sig ❺), DGK homologs. A closer relationship between class I and IV homologs is also suggested by Sig❷A, B, where inserts that are either 2 aa or 4 aa long are specifically present in these two classes of DGK homologs and by the observed branching pattern of these homologs in the phylogenetic tree. In another instance (Sig❿A, B), CSIs of different lengths (3 aa and 4 aa) in the same position are a uniquely present in theε- and θ- classes of DGK-homologs. Different lengths of CSIs in the same position in specific classes of DGK homologs can results from several possibilities. First, it is possible that independent genetic changes have occurred in the ancestor(s) of specific DGK classes, giving rise to CSIs of different lengths. However, in some cases such as Sig❷A, B (and also Sig❿A, B), it is also possible that initially a genetic change leading to either 2 aa or 4 aa insert occurred in a common ancestor of the class I and IV DGK homologs, and then another insertion or deletion event occurred in the same position in the ancestor of one of these two groups leading to the observed differences in the lengths of the CSIs in these two classes. Although it is difficult to distinguish between these two possibilities, the fact that the class I and class IV DGK homologs also contain another uniquely shared CSI (viz. Sig❹) and they branch together in the phylogenetic tree corroborate the view that the homologs from these two classes are closely related and that the CSI ❷A,B initially occurred in a common ancestor of these two classes.

There is limited understanding at present concerning the overall function of the catalytic domain for eukaryotic DGKs [1–5,31]. Some differences have been recently reported in the activities of different classes of DGKs towards 1-monoacylglycerol and 2-monoacylglycerol [11]. The catalytic domain contains a highly-conserved sequence motif (G-X-G-X-X-G) similar to that present in the ATP-binding sites of other protein kinases [2,31,32]. The importance of this motif for the functioning of DGKs has been established by mutational studies [2,31,37]. The CDs from DGK-ε, -ζ, and -θ proteins exhibit very little DGK activity when expressed as isolated subunits [2]. This and other observations have led to the suggestion that the proper functioning of the CD for mammalian DGKs requires interactions with other motifs (proteins) [1–5,31]. However, our knowledge of the sequence motifs in the CD that may facilitate its interactions with other proteins/effectors molecules, and how different DGK isozymes, or classes of isozymes, differ in this regard is very limited. The lack of structural information for the CD for eukaryotic DGKs is also a major hindrance in understanding its function. In this context, our identification of multiples CSIs within the CD that are specific for different classes of DGK homologs is of much interest and significance. Based on earlier work, the CSIs in protein sequences are known to play important functional roles in the organisms in which they are found [38–42]. The mutational changes in the CSIs affect proper functioning of the proteins as well as cell growth, and both large and small CSIs are indicated to be significant in this regard [38]. Further, extensive work on CSIs show that they are generally located in protein structures in surface loops [39,42–45], which are indicated to play important roles in mediating protein-protein interactions [39,44,46–49]. Based on these observations, it can be hypothesized (predicted) that the CSIs in the CD that are specific for different classes of DGK homologs should also be playing important roles in the interactions of different DGK isozymes with specific proteins (or other target molecules) that may be responsible for the differences in the cellular functions of the different DGK isozymes. Thus, further studies on understanding the cellular functions of the identified CSIs, and the specific interactions that they may be involved in, could provide important insights regarding the biochemical basis of the differences in the

functions of different DGK isozymes. It should also be noted that apart from the CSIs that are specific for different classes of DGKs, sequence alignments shown in Figs 2–7 and S1 Fig also contain large numbers of specific amino acid substitutions (or polymorphisms) that are specific for particular classes of DGK homologs. These sequence polymorphisms constitute another type of molecular signature, which are also expected to be of functional significance [12,38]. The DGK family of proteins, due to their key roles in controlling the cellular concentrations of two important signalling molecules and thereby controlling diverse cellular processes, are targets for development of novel therapeutics for treatment of cancer, neuronal, metabolic and inflammatory diseases [9,50–57]. In this context, the CSIs identified in the present work, which are specific for different classes of DGK homologs and predicted to play important roles in the cellular functions of these isoforms, thus provide potentially useful drug targets [58].

The DGK isozymes are primarily found in eukaryotic organisms. Although a protein exhibiting DGK activity is present in many bacteria [4,5,59,60], it shows minimal similarity (15–18%) over a short length, and with multiple intervening gaps, to the eukaryotic DGKs [13] [61]. The bacterial DGK homologs are short in comparison to the eukaryotic proteins and they do not contain any of the canonical sequence features (viz. catalytic domain and cysteine-rich C1 domains), which are shared characteristics of all eukaryotic DGK homologs [2,4,5,12,59]. These observations strongly suggest that the bacterial DGK homologs are evolutionarily unrelated to the eukaryotic counterparts, and that the proteins exhibiting DGK activity have evolved independently in bacteria. In view of the importance of DGK isozymes in cellular signalling processes and their involvement in a wide variety of cellular functions, it is of much interest to know how this large family of proteins has evolved. The phylogenetic approach employed in this work was able to clearly distinguish the five different classes of DGKs; however, as the canonical DGK homologs are only found in eukaryotic organisms, it was difficult to root the tree and determine how members of different DGK classes are related to each other and which of these DGK isozymes is ancestral. The CSIs in protein sequences provide another useful means for understanding the evolutionary relationships among distantly related species; however, in the present instance, due to the absence of a suitable outgroup, it was difficult to establish the ancestral state of the DGK protein sequence and reliably infer whether the observed CSIs are insertions or deletions in the indicated classes of DGK homologs.

Although the present work does not resolve how different classes of DGK proteins have originated from a common ancestor and which of these isoforms is most ancestral, the results presented here show that of the different classes of DGKs, the DGK-ε, -θ and -ζ isoforms are most widely distributed in eukaryotic organisms. In addition to their nearly ubiquitous presence in vertebrate species, the homologs for these isoforms are also indicated to be present in most invertebrate species. In contrast, homologs for the other DGK isozymes were not detected in a number of invertebrate phyla (Table 1). Although more detailed studies are needed to establish with certainty the identity of the DGK isoforms that are present in deep branching eukaryotic phyla, the observed broader distribution of the DGK-ε, DGK-θ and DGK-ζ isoforms within eukaryotic organisms suggests that one of these isoforms likely represent the ancestral form of the protein. Of these isoforms, DGK-ε represents the most rudimentary form of the protein containing all of the main sequence features that are commonly shared by all other DGK family members. Further, DGK-ε is the only isoform which shows specificity for a particular substrate i.e. arachidonoyl-containing diacylglycerol [2,3,6,12,62]. Thus, it is possible that the DGK-ε originated first, and all other DGK isoforms originated from it by acquisition of different domains and other identified signatures. However, the present study is based only on the analyses of the catalytic domain sequences and further work on the analyses of other sequence regions of the DGK protein family could provide additional useful information to clarify the relationships among different DGK isoforms.

Supporting information

S1 Fig. Edited multiple sequence alignment of the catalytic domain for different DGK isozymes that was used for phylogenetic tree construction.
(PDF)

S1 Table. Sequence information for different DGK homologs used in phylogenetic studies.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: Radhey S. Gupta, Richard M. Epand.

Data curation: Radhey S. Gupta.

Formal analysis: Radhey S. Gupta.

Funding acquisition: Radhey S. Gupta.

Investigation: Radhey S. Gupta.

Methodology: Radhey S. Gupta.

Project administration: Radhey S. Gupta.

Resources: Radhey S. Gupta.

Software: Radhey S. Gupta.

Supervision: Radhey S. Gupta.

Validation: Radhey S. Gupta.

Visualization: Radhey S. Gupta.

Writing – original draft: Radhey S. Gupta.

Writing – review & editing: Radhey S. Gupta, Richard M. Epand.

References

1. Shulga YV, Topham MK, Epand RM. Regulation and functions of diacylglycerol kinases. Chem Rev.2011; 111: 6186–6208. https://doi.org/10.1021/cr1004106 PMID: 21800853

2. Topham MK, Epand RM. Mammalian diacylglycerol kinases: molecular interactions and biological functions of selected isoforms. Biochim Biophys Acta.2009; 1790: 416–424. https://doi.org/10.1016/j.bbagen.2009.01.010 PMID: 19364481

3. Merida I, Avila-Flores A, Merino E. Diacylglycerol kinases: at the hub of cell signalling. Biochem J.2008; 409: 1–18. https://doi.org/10.1042/BJ20071040 PMID: 18062770

4. Sakane F, Imai S, Kai M, Yasuda S, Kanoh H. Diacylglycerol kinases: why so many of them? Biochim Biophys Acta.2007; 1771: 793–806. https://doi.org/10.1016/j.bbalip.2007.04.006 PMID: 17512245

5. van Blitterswijk WJ, Houssa B. Properties and functions of diacylglycerol kinases. Cell Signal.2000; 12: 595–605. PMID: 11080611

6. Goto K, Hozumi Y, Nakano T, Saino-Saito S, Martelli AM. Lipid messenger, diacylglycerol, and its regulator, diacylglycerol kinase, in cells, organs, and animals: history and perspective. Tohoku J Exp Med.2008; 214: 199–212. PMID: 18323690

7. Diao J, Wu C, Zhang J, Liu J, Zhang X, Hao P, et al. Loss of Diacylglycerol Kinase-Zeta Inhibits Cell Proliferation and Survival in Human Gliomas. Mol Neurobiol.2016; 53: 5425–5435. https://doi.org/10.1007/s12035-015-9419-4 PMID: 26452358

8. Singh BK, Kambayashi T. The Immunomodulatory Functions of Diacylglycerol Kinase zeta. Front Cell Dev Biol.2016; 4: 96. https://doi.org/10.3389/fcell.2016.00096 PMID: 27656643

9. Purow B. Molecular Pathways: Targeting Diacylglycerol Kinase Alpha in Cancer. Clin Cancer Res.2015; 21: 5008–5012. https://doi.org/10.1158/1078-0432.CCR-15-0413 PMID: 26420856

10. Arisz SA, Testerink C, Munnik T. Plant PA signaling via diacylglycerol kinase. Biochim Biophys Acta.2009; 1791: 869–875. https://doi.org/10.1016/j.bbalip.2009.04.006 PMID: 19394438

11. Sato Y, Murakami C, Yamaki A, Mizuno S, Sakai H, Sakane F. Distinct 1-monoacylglycerol and 2-monoacylglycerol kinase activities of diacylglycerol kinase isozymes. Biochim Biophys Acta.2016; 1864: 1170–1176. https://doi.org/10.1016/j.bbapap.2016.06.012 PMID: 27346717

12. Epand RM, So V, Jennings W, Khadka B, Gupta RS, Lemaire M. Diacylglycerol Kinase-epsilon: Properties and Biological Roles. Front Cell Dev Biol.2016; 4: 112. https://doi.org/10.3389/fcell.2016.00112 PMID: 27803897

13. Jennings W, Doshi S, D'Souza K, Epand RM. Molecular properties of diacylglycerol kinase-epsilon in relation to function. Chem Phys Lipids.2015; 192: 100–108. https://doi.org/10.1016/j.chemphyslip.2015.06.003 PMID: 26134136

14. De La Roche MA, Smith JL, Rico M, Carrasco S, Merida I, Licate L, et al. Dictyostelium discoideum has a single diacylglycerol kinase gene with similarity to mammalian theta isoforms. Biochem J.2002; 368: 809–815. https://doi.org/10.1042/BJ20021027 PMID: 12296770

15. Biron D, Shibuya M, Gabel C, Wasserman SM, Clark DA, Brown A, et al. A diacylglycerol kinase modulates long-term thermotactic behavioral plasticity in C. elegans. Nat Neurosci.2006; 9: 1499–1505. https://doi.org/10.1038/nn1796 PMID: 17086178

16. Pilz A, Schaap D, Hunt D, Fitzgibbon J. Chromosomal localization of three mouse diacylglycerol kinase (DAGK) genes: genes sharing sequence homology to the Drosophila retinal degeneration A (rdgA) gene. Genomics.1995; 26: 599–601. PMID: 7607687

17. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics.2007; 23: 2947–2948. https://doi.org/10.1093/bioinformatics/btm404 PMID: 17846036

18. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol.2013; 30: 2725–2729. https://doi.org/10.1093/molbev/mst197 PMID: 24132122

19. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Computer applications in the biosciences: CABIOS.1992; 8: 275–282. PMID: 1633570

20. Gupta RS. Molecular signatures that are distinctive characteristics of the vertebrates and chordates and supporting a grouping of vertebrates with the tunicates. Mol Phylogenet Evol.2016; 94: 383–391. https://doi.org/10.1016/j.ympev.2015.09.019 PMID: 26419477

21. Gupta RS. Impact of genomics on the understanding of microbial evolution and classification: the importance of Darwin's views on classification. FEMS Microbiol Rev.2016; 40: 520–553. https://doi.org/10.1093/femsre/fuw011 PMID: 27279642

22. Gupta RS: Identification of Conserved Indels that are Useful for Classification and Evolutionary Studies. In Bacterial Taxonomy, Methods in Microbiology Volume 41. Edited by Goodfellow M, Sutcliffe IC, Chun J. London: Elsevier; 2014:153–182.

23. Zhang G, Gao B, Adeolu M, Khadka B, Gupta RS. Phylogenomic Analyses and Comparative Studies on Genomes of the Bifidobacteriales: Identification of Molecular Signatures Specific for the Order Bifidobacteriales and Its Different Subclades. Front Microbiol.2016; 7: 978. https://doi.org/10.3389/fmicb.2016.00978 PMID: 27446019

24. Rokas A, Holland PW. Rare genomic changes as a tool for phylogenetics. Trends Ecol Evol.2000; 15: 454–459. PMID: 11050348

25. Gupta RS. Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. Microbiol Mol Biol Rev.1998; 62: 1435–1491. PMID: 9841678

26. Springer MS, Stanhope MJ, Madsen O, de Jong WW. Molecules consolidate the placental mammal tree. Trends Ecol Evol.2004; 19: 430–438. https://doi.org/10.1016/j.tree.2004.05.006 PMID: 16701301

27. Baldauf SL, Palmer JD. Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. Proc Natl Acad Sci USA.1993; 90: 11558–11562. PMID: 8265589

28. Rivera MC, Lake JA. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. Science.1992; 257: 74–76. PMID: 1621096

29. Gupta RS. Molecular signatures that are distinctive characteristics of the vertebrates and chordates and supporting a grouping of vertebrates with the tunicates. Mol Phylogenet Evol.2015; 94: 383–391. https://doi.org/10.1016/j.ympev.2015.09.019 PMID: 26419477

30. Valas RE, Bourne PE. Structural analysis of polarizing indels: an emerging consensus on the root of the tree of life. Biol Direct.2009; 4: 30. https://doi.org/10.1186/1745-6150-4-30 PMID: 19706177

31. Goto K, Kondo H. Functional implications of the diacylglycerol kinase family. Adv Enzyme Regul.2004; 44: 187–199. https://doi.org/10.1016/j.advenzreg.2003.11.007 PMID: 15581490

32. Hanks SK, Quinn AM, Hunter T. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science.1988; 241: 42–52. PMID: 3291115

33. Feng H, Ren M, Chen L, Rubin CS. Properties, regulation, and in vivo functions of a novel protein kinase D: Caenorhabditis elegans DKF-2 links diacylglycerol second messenger to the regulation of stress responses and life span. J Biol Chem.2007; 282: 31273–31288. https://doi.org/10.1074/jbc.M701532200 PMID: 17728253

34. Ostroski M, Tu-Sekine B, Raben DM. Analysis of a novel diacylglycerol kinase from Dictyostelium discoideum: DGKA. Biochemistry.2005; 44: 10199–10207. https://doi.org/10.1021/bi0507276 PMID: 16042397

35. Frolov MV, Benevolenskaya EV, Birchler JA. Molecular analysis of a novel Drosophila diacylglycerol kinase, DGKepsilon. Biochim Biophys Acta.2001; 1538: 339–352. PMID: 11336805

36. Hurley JH, Newton AC, Parker PJ, Blumberg PM, Nishizuka Y. Taxonomy and function of C1 protein kinase C homology domains. Protein Sci.1997; 6: 477–480. https://doi.org/10.1002/pro.5560060228 PMID: 9041654

37. Sanjuan MA, Jones DR, Izquierdo M, Merida I. Role of diacylglycerol kinase alpha in the attenuation of receptor signaling. J Cell Biol.2001; 153: 207–220. PMID: 11285286

38. Singh B, Gupta RS. Conserved inserts in the Hsp60 (GroEL) and Hsp70 (DnaK) proteins are essential for cellular growth. Mol Genet Genomics.2009; 281: 361–373. https://doi.org/10.1007/s00438-008-0417-3 PMID: 19127371

39. Gupta RS, Nanda A, Khadka B. Novel molecular, structural and evolutionary characteristics of the phosphoketolases from bifidobacteria and Coriobacteriales. PLoS ONE.2017; 12: e0172176. https://doi.org/10.1371/journal.pone.0172176 PMID: 28212383

40. Clarke JH, Irvine RF. Evolutionarily conserved structural changes in phosphatidylinositol 5-phosphate 4-kinase (PI5P4K) isoforms are responsible for differences in enzyme activity and localization. Biochem J.2013; 454: 49–57. https://doi.org/10.1042/BJ20130488 PMID: 23758345

41. Geszvain K, Gruber TM, Mooney RA, Gross CA, Landick R. A hydrophobic patch on the flap-tip helix of E.coli RNA polymerase mediates sigma(70) region 4 function. J Mol Biol.2004; 343: 569–587. https://doi.org/10.1016/j.jmb.2004.08.063 PMID: 15465046

42. Alnajar S, Khadka B, Gupta RS. Ribonucleotide reductases from Bifidobacteria contain multiple conserved indels distinguishing them from all other organisms: In silico analysis of the possible role of a 43 aa bifidobacteria-specific insert in the Class III RNR homolog. Front Microbiol.2017; 8: Article 1409.

43. Gupta RS. Moelcular Markers for Photosynthetic Bacteria and Insights into the Origin and Spread of Photosynthesis. Advances in Botanical Research.2013; 66: 37–66.

44. Akiva E, Itzhaki Z, Margalit H. Built-in loops allow versatility in domain-domain interactions: lessons from self-interacting domains. Proc Natl Acad Sci U S A.2008; 105: 13292–13297. https://doi.org/10.1073/pnas.0801207105 PMID: 18757736

45. Khadka B, Gupta RS. Identification of a Conserved 8 aa Insert in the PIP5K Protein in the Saccharomycetaceae family of Fungi and the Molecular Dynamics Simulations and Structural Analysis to investigate its Potential Functional Role. Proteins.2017.

46. Hashimoto K, Panchenko AR. Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. Proc Natl Acad Sci U S A.2010; 107: 20352–20357. https://doi.org/10.1073/pnas.1012999107 PMID: 21048085

47. Chatterji M, Unniraman S, Maxwell A, Nagaraja V. The additional 165 amino acids in the B protein of Escherichia coli DNA gyrase have an important role in DNA binding. J Biol Chem.2000; 275: 22888–22894. https://doi.org/10.1074/jbc.M001047200 PMID: 10764756

48. Hormozdiari F, Salari R, Hsing M, Schonhuth A, Chan SK, Sahinalp SC, et al. The effect of insertions and deletions on wirings in protein-protein interaction networks: a large-scale study. J Comput Biol.2009; 16: 159–167. https://doi.org/10.1089/cmb.2008.03TT PMID: 19193143

49. Cherkasov A, Nandan D, Reiner NE. Selective targeting of indel-inferred differences in spatial structures of highly homologous proteins. Proteins.2005; 58: 950–954. https://doi.org/10.1002/prot.20391 PMID: 15657927

50. Yang E, Singh BK, Paustian AM, Kambayashi T. Diacylglycerol Kinase zeta Is a Target To Enhance NK Cell Function. J Immunol.2016; 197: 934–941. https://doi.org/10.4049/jimmunol.1600581 PMID: 27342844

51. Zamani A, Decker C, Cremasco V, Hughes L, Novack DV, Faccio R. Diacylglycerol Kinase zeta (DGKzeta) Is a Critical Regulator of Bone Homeostasis Via Modulation of c-Fos Levels in Osteoclasts. J Bone Miner Res.2015; 30: 1852–1863. https://doi.org/10.1002/jbmr.2533 PMID: 25891971

52. Kaneko YK, Ishikawa T. Diacylglycerol Signaling Pathway in Pancreatic beta-Cells: An Essential Role of Diacylglycerol Kinase in the Regulation of Insulin Secretion. Biol Pharm Bull.2015; 38: 669–673. https://doi.org/10.1248/bpb.b15-00060 PMID: 25947912

53. Shirai Y, Saito N. Diacylglycerol kinase as a possible therapeutic target for neuronal diseases. J Biomed Sci.2014; 21: 28. https://doi.org/10.1186/1423-0127-21-28 PMID: 24708409

54. Prinz PU, Mendler AN, Brech D, Masouris I, Oberneder R, Noessner E. NK-cell dysfunction in human renal carcinoma reveals diacylglycerol kinase as key regulator and target for therapeutic intervention. Int J Cancer.2014; 135: 1832–1841. https://doi.org/10.1002/ijc.28837 PMID: 24615391

55. Sakane F, Imai S, Kai M, Yasuda S, Kanoh H. Diacylglycerol kinases as emerging potential drug targets for a variety of diseases. Curr Drug Targets.2008; 9: 626–640. PMID: 18691010

56. Brown JR, Auger KR. Phylogenomics of phosphoinositide lipid kinases: perspectives on the evolution of second messenger signaling and drug discovery. BMC Evol Biol.2011; 11: 4. https://doi.org/10.1186/1471-2148-11-4 PMID: 21208444

57. Merida I, Torres-Ayuso P, Avila-Flores A, Arranz-Nicolas J, Andrada E, Tello-Lafoz M, et al. Diacylglycerol kinases in cancer. Adv Biol Regul.2017; 63: 22–31. https://doi.org/10.1016/j.jbior.2016.09.005 PMID: 27697466

58. Nandan D, Lopez M, Ban F, Huang M, Li Y, Reiner NE, et al. Indel-based targeting of essential proteins in human pathogens that have close host orthologue(s): discovery of selective inhibitors for Leishmania donovani elongation factor-1alpha. Proteins.2007; 67: 53–64. https://doi.org/10.1002/prot.21278 PMID: 17243179

59. Lightner VA, Bell RM, Modrich P. The DNA sequences encoding plsB and dgk loci of Escherichia coli. J Biol Chem.1983; 258: 10856–10861. PMID: 6309817

60. Jerga A, Lu YJ, Schujman GE, de Mendoza D, Rock CO. Identification of a soluble diacylglycerol kinase required for lipoteichoic acid production in Bacillus subtilis. J Biol Chem.2007; 282: 21738–21745. https://doi.org/10.1074/jbc.M703536200 PMID: 17535816

61. Miller DJ, Jerga A, Rock CO, White SW. Analysis of the Staphylococcus aureus DgkB structure reveals a common catalytic mechanism for the soluble diacylglycerol kinases. Structure.2008; 16: 1036–1046. https://doi.org/10.1016/j.str.2008.03.019 PMID: 18611377

62. Epand RM, Topham MK. Measurement of mammalian diacylglycerol kinase activity in vitro and in cells. Methods Enzymol.2007; 434: 293–304. https://doi.org/10.1016/S0076-6879(07)34016-0 PMID: 17954254