J Mol Evol (2008) 66:11–20
DOI 10.1007/s00239-007-9058-0

The Early Evolution of the Phosphagen Kinases—Insights from Choanoflagellate and Poriferan Arginine Kinases

Maria Conejo · Matt Bertin · Shirley A. Pomponi · W. Ross Ellington

Received: 15 April 2007 / Accepted: 7 November 2007 / Published online: 7 December 2007

Abstract Arginine kinase (AK) is a member of a large family of phosphoryl transfer enzymes called phosphagen (guanidino) kinases. AKs are present in certain protozoans, sponges, cnidarians, and both lophotrochozoan and ecdysozoan protostomes. Another phosphagen kinase, creatine kinase (CK), is found in sponges, cnidarians, and both deuterostome and protostome groups but does not appear to be present in protozoans. To probe the early evolution of phosphagen kinases, we have amplified the cDNAs for AKs from three choanoflagellates and from the hexactinellid sponge *Aphrocallistes beatrix* and the demosponges *Suberites fuscus* and *Microciona prolifera*. Phylogenetic analysis using maximum likelihood of these choanoflagellate and sponge AKs with other AK sequences revealed that the AK from the choanoflagellate *Monosiga brevicollis* clusters with the AK from the glass sponge *Aphrocallistes* and is part of a larger cluster containing AKs from the demosponges *Suberites* and *Microciona* as well as basal and protostome invertebrates. In contrast, AKs from *Codonosiga gracilis* and *Monosiga ovata* form a distinct cluster apart from all other AK sequences. tBLASTn searches of the recently released *M. brevicollis* genome database showed that this species has three unique AK genes—one virtually identical to the *M. brevicollis* cDNA and the other two showing great similarity to *C. gracilis* and *M. ovata* AKs. Three distinct AK genes are likely present in choanoflagellates. Two of these AKs display extensive similarity to both CKs and an AK from sponges. Previous work has shown CK evolved from an AK-like ancestor prior to the divergence of sponges. The present results provide evidence suggesting that the initial gene duplication event(s) leading to the CK lineage may have occurred before the divergence of the choanoflagellate and animal lineages.

Keywords Arginine kinase · Creatine kinase · Choanoflagellates · Sponges

M. Conejo · M. Bertin · W. R. Ellington (✉)
Department of Biological Science and Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306-4370, USA
e-mail: elling@bio.fsu.edu

S. A. Pomponi
Harbor Branch Oceanographic Institution, 5600 U.S. 1 North, Ft. Pierce, FL 34946, USA

Present Address:
M. Bertin
Florida Fish and Wildlife Research Institute, 100 Eighth Avenue SE, St. Petersburg, FL 33701-5020, USA

Introduction

Phosphagen (guanidino) kinases are members of a family of phosphoryl transfer enzymes that play a central role in energy homeostasis in cells that display high and variable rates of ATP turnover (Ellington 2001). The nomenclature for this family is based on the guanidine substrate phosphorylated- arginine kinase (AK), creatine kinase (CK), lombicine kinase (LK), glycocyamine kinase (GK), and taurocyamine kinase (TK). AK is thought to be most closely related to the basal phosphagen kinase because (a) AK is the most widely distributed being present in protozoa and most invertebrate groups; (b) AKs typically are functional monomers, whereas the other phosphagen kinases exist mostly as functional oligomers; and (c) AKs utilize an unmodified amino acid substrate (arginine), while the others use specialized guanidino compounds like creatine, taurocyamine, and lombricine (Ellington 2001).

The AK gene evolved prior to the divergence of the two major eukaryotic lineages, the ophistokonts and anterokonts, and subsequently has undergone a variety of duplication/divergence and even gene fusion events (Uda et al. 2006). CKs are thought to have evolved from the AK lineage through an early gene duplication and divergence event (Watts 1975). Following this early event, multiple gene duplications occurred, leading to the extant family of CK isoforms targeted at different intracellular compartments—cytoplasmic, mitochondrial, and flagellar CKs (Suzuki et al. 2004). We know that this event occurred rather early, as the genes for two of the three known CK isoforms were present prior to the divergence of sponges from the metazoans (Sona et al. 2004; Bertin et al. 2007). It has recently been shown that the genes for LK (Suzuki et al. 1997), GK (Ellington et al. 2004), and TK (Uda et al. 2005), which are present only in certain annelid groups, evolved from a CK-like ancestor after the divergence of the annelids from the lophotrochozoan protostome lineage.

Sponges are basal metazoans with rather simple body plans. Of the three major classes in the phylum *Porifera*, hexactinellid (glass) sponges are thought to have diverged first from the metazoan lineage and are closely related to demosponges (Kruse et al. 1998). Available data support the view that calcareous sponges diverged later (e.g., see Kruse et al. 1998). Recently, considerable attention has focused on the relationship of the choanoflagellates to metazoans. A considerable base of information has been amassed indicating that choanoflagellates are basal and constitute a direct sister-group of sponges (Medina et al. 2001; Snell et al. 2001; Lang et al. 2002; Rokas et al. 2003; Philippe et al. 2004; Peterson et al. 2005). These unicellular organisms appear to be monophyletic (Medina et al. 2001; Steenkamp et al. 2005). The choanoflagellate mitochondrial genome is more protozoan-like (Burger et al. 2003) in comparison to the more metazoan-like mitochondrial genome of sponges (Lavrov et al. 2005).

Both AK and CK are present in demosponges (Robin and Guillou 1980; Ellington 2000; Perovic-Ottstadt et al. 2005; Sona et al. 2004; Bertin et al. 2007). Two CK isoform genes are present in hexactinellids, arguably the most ancient, extant group of metazoans (Bertin et al. 2007). Given the apparent absence of CK in protozoans investigated thus far (Ellington and Suzuki 2006) and the close proximity of choanoflagellates to the sponges, these two groups represent attractive systems to explore the early evolution of the phosphagen kinases. In the present effort, we have amplified the cDNAs for AKs from three choanflagellates, one hexactinellid sponge, and two demosponges. Our results show that choanoflagellates likely have at least three distinct AK genes but appear to lack CK. One of these AKs is very similar to the hexactinellid AK and, in a phylogenetic tree, is part of a large supercluster consisting of ciliate, basal metazoan, and protostome AKs. The other two choanoflagellate AKs form a unique clade and, although true AKs, display great sequence similarities to CKs. The results suggest that the early gene duplication and divergence events leading to the CK lineage may have occurred prior to the divergence of the choanoflagellate and animal lineages.

Materials and Methods

Animals

Cultures of three choanoflagellate species were obtained from the American Type Culture Center (ATCC; Manassas, VA): *Monosiga brevicollis* (ATCC-50154), *Codonosiga gracilis* (listed as *Monosiga gracilis* by ATCC; ATCC-50964), and *Monosiga ovata* (ATCC-50635). Choanoflagellates were cultured in 15-cm-diameter, untreated, plastic cell culture dishes at 25°C using ATCC medium 1525 for *M. brevicollis* and *C. gracilis* and ATTC medium 802 for *M. ovata*. Cells were dislodged from the culture dishes using disposable, sterile cell scrapers (Fisher Scientific, Atlanta, GA) every 4 days (*M. brevicollis* and *M. gracilis*) or 14 days (*M. ovata*), and 6 ml of the suspension was transferred to 60 ml of new medium. Ongoing, stable cultures of all three species were established.

A specimen of a hexactinellid sponge (subsequently identified as *Aphrocallistes beatrix* by Dr. Henry Reiswig, University of Victoria, Victoria, British Columbia, Canada) was collected via submersible by Harbor Branch Oceanographic Institution, Fort Pierce, Florida, USA. Hexactinellid tissue was preserved in RNAlater (Ambion, Austin, TX). Specimens of demosponges *Suberites fuscus* and *Microciona prolifera* were purchased from the Marine Biological Laboratory (Woods Hole, MA).

RNA Isolation

The contents of 8 to 15 culture dishes of each choanoflagellate species were pooled and pelleted by gentle centrifugation in 250-ml centrifuge bottles. The resulting pellets were homogenized in 10–20 ml TRIZOL (Invitrogen, Carlsbad, CA) using a Brinkmann Polytron (Westbury, NY). Hexactinellid tissue preserved in RNA later was scooped directly from the cryostorage vial. Fresh demosponge tissue was diced into approximately 2-mm cubes using a razor blade. Approximately 0.5–2 g of sponge tissue was homogenized in 15 to 30 ml of TRIZOL reagent as above. The homogenates were then centrifuged at 7650g for 30 min and total RNA was isolated according

to the manufacturer's protocol. Final total RNA pellets were resuspended in 75% ethanol or dissolved in approximately 0.20 ml of nuclease-free water. mRNA was selected from the total RNA of *M. brevicollis*, *C. gracilis*, and *M. ovata* using Qiagen (Valencia, CA) Oligotex kits according to the manufacturer's protocol. All RNA samples were stored at −70°C.

RTPCR Amplification of AK cDNAs

M. brevicollis AK

The 3′ end of the cDNA was amplified with an Invitrogen 3′-RACE (rapid amplification of cDNA ends) (Frohman et al. 1988) kit using total RNA, a lock-docking oligo-dT (LD/dT) primer (Borson et al. 1992), and a forward primer (5′-GGTACGGTGCTGAACGAGGTGTTTGAC-3′) based on unique EST sequence data provided by Dr. Nicole King (Departments of Molecular & Cell Biology and Integrative Biology, University of California, Berkeley). PCR amplifications were conducted in a Hybaid (Middlesex, UK) PCR Sprint thermocycler using Takara (Santa Ana, CA) ExTaq HS DNA polymerase in a touchdown protocol (annealing temperatures of 65, 62, 59, and 56°C, 4 × each, followed by 61°C, 16×). The resulting 700-bp product was subcloned and sequenced. Due to some ambiguity in the 3′-end of the cDNA sequence, a second round 3′-RACE was conducted using mRNA and a BD Clontech (Palo Alto, CA) SMART RACE kit. The cDNA was amplified using a new gene specific, forward primer (5′-TCCTTTGCCATGTCACCC ACCCAGATG-3′) and Clontech DNA polymerase in a touchdown protocol as above. The resulting 500-bp product yielded high-quality sequence.

To obtain the 5′-end sequence for *M. brevicollis* AK, the cDNA was amplified using an Invitrogen 5′-RACE kit with a reverse primer based on the second-round 3′-RACE results (5′-CGCATGCCCGTGCCCAGGTTGGTCGG-3′) and ExTaq HS DNA polymerase. Amplification (65°C, 33×) yielded an 800-bp product which contained most of the AK open reading frame (ORF). To obtain the complete 5′-end sequence, it was necessary to do two more rounds of amplifications. First-strand cDNA was prepared using a MonsterScript kit (Epicentre Biotechnologies, Madison, WI) with a reverse primer (5′-CCGGGTGCTCTGGCGT-CAGCGAGT-3′) derived from the 3′-RACE results. The cDNA was then amplified with an Invitrogen 5′-RACE kit using the supplied abridged anchor primer, a gene specific primer (5′-AGGTCCAGGCAGCCGTCCGCAGGC-3′), ExTaq HS DNA polymerase, and a touchdown protocol (68, 65, 62, and 59°C, 4 × each; 65°C, 16×). The 1:100 dilution of the above reaction mixture was reamplified under the same conditions, and a 400-bp product from this reaction was purified using a Qiagen gel purification kit and reamplified as above. The final product was subcloned and sequenced, yielding the complete 5′-end sequence.

C. gracilis AK

Both 3′- and 5′-RACE amplifications were conducted using the BD Clontech SMART RACE kit with. For 3′-RACE, single-stranded cDNA was produced with the LD/dT primer and then amplified using a universal, redundant primer (5′-GTNTGGRTNAAYGARGARGAYCA-3′), which corresponds to the region of the highly conserved NEEDH sequence found in all phosphagen kinases (Suzuki and Furukohri 1994), and the LD/dT primer. Amplification was conducted using ExTaq HS DNA polymerase (55°C, 33×). The resulting 500-bp product was subcloned, sequenced, and yielded an AK sequence through the 3′-end. The 5′-end of this AK was amplified with the BD Clontech universal primer mix, a gene specific reverse primer (5′-GGCAGCAATTCCACGGCTCAAACGCTC-3′) derived from the 3′-end amplification results, and ExTaq HS DNA polymerase using a touchdown protocol (65, 62, 59, and 56°C, 4 × each; 62°C, 16×). The resulting ∼1000-bp product was gel extracted, reamplified as above, subcloned, and sequenced.

M. ovata AK

The cDNA for *M. ovata* AK was amplified according to the above protocol for the *C. gracilis* AK. The 5′-RACE amplification used an *M. ovata* specific primer (5′-TCTTGACGACGCCCTCGACGGCATTGAT-3′) derived from the 3′-end amplification results, which yielded an ∼1100-bp product that contained the remaining portions of the ORF as well as the 5′-UTR.

A. beatrix, S. fuscus, and M. prolifera AKs

Reverse transcription was carried out on total RNA using the BD Clontech SMART RACE kit. The 3′ end of the AK cDNAs was amplified (55°C, 33×) using the LD/dT primer, the redundant NEEDH region primer used in the *C. gracilis* and *M. ovata* 3′-RACE amplifications and ExTaq HS DNA polymerase. Products of ∼800 bp representing the 3′ ends of AK were subcloned and sequenced. Reverse, gene-specific primers were designed for 5′-RACE amplifications from the previously sequenced 3′-RACE products (*A. beatrix*, 5′-AAATTCTTTCGAATGGGCAAATCCCCC-3′; *S. fuscus*, 5′-TTCCATGGAGATGATACGAAGATGATC-3′; *M. prolifera*, 5′-ACCAATGTCAGCTCCCTTCTGCAT

GGA-3′). Amplification was carried out using 5′-RACE-ready cDNA with the BD SMART RACE kit and the universal primer mix and ExTaq HS DNA polymerase. PCR amplifications were carried out under the same conditions as for the 3′-RACE, with annealing temperatures 10°C lower than the respective gene-specific primer melting temperatures. Products in the size range of 400 to 800 bp were subcloned and sequenced, yielding the 5′-end sequences for all three AKs including the 5′-UTR sequences.

Sequence and Phylogenetic Analysis

AK sequences were assembled using the Sequencher software package (Gene Codes, Ann Arbor, MI). Multiple sequence alignments were carried out in GCG's Seqlab (Accelrys, San Diego, CA) using the Pileup function and manual refinement. A consensus mask was applied and areas of less than 15% consensus were excluded. Multiple sequence files were imported into the PHYLIP package 3.6. A total of 26 AK sequences were used; 3 sponge CK sequences were used as an outgroup. Maximum Likelihood (ML) tree construction was carried out using the ProML program (Amino Acid Sequence Maximum Likelihood method, version 3.6a2.1) in the PHYLIP package. The sequence alignment was run in the Seqboot program (Bootstrapping algorithm, version 3.6a2.1), generating 100 replications of the alignment. This alignment was then run in 10 independent ProML programs, generating 100 ProML trees. The program was set to the Jones-Taylor-Thornton amino acid model, gamma distribution, globally rearranged 10 times with a random input order. A consensus tree was then created using PHYLIP's Consense (a consensus tree program, version 3.6a2.1) to provide bootstrap values for the ML tree. Bootstrap values indicate the number of times a node appeared, of 100 ML trees.

Results and Discussion

Amplification and Analysis of Choanoflagellate and Sponge AK cDNAs

Full-length AK cDNAs sequences, containing the 5′- and 3′-UTRs and the ORFs, were obtained for the choanoflagellates *M. brevicollis* and *M. ovata*. These cDNAs code for 394- and 389-residue proteins with estimated relative molecular masses (M_r) of 42.57 and 43.31 kDa, respectively. Attempts to amplify the entire 5′-end of the cDNA for the AK from *C. gracilis* consistently failed. However, the resulting cDNA contained what appeared to be all of the ORF minus the start codon. This cDNA codes for a 382-residue protein with an estimated M_r of 42.25 kDa. Full-length AK cDNAs sequences, containing the 5′- and 3′-UTRs and the ORFs, were obtained for the hexactinellid sponge *A. beatrix* and the demosponges *S. fuscus* and *M. prolifera*. These cDNAs code for 367-, 382-, and 361-residue proteins with estimated M_r values of 41.03, 43.12, and 40.3, respectively. All six cDNA sequences have been deposited in GenBank under the following accession numbers: *M. brevicollis*, AK-EF394320; *M. ovata*, AK-EF394322; *C. gracilis*, AK-EF394321; *A. beatrix*, AK-DQ786177; *S. fuscus*, AK-DQ786183; and *M. prolifera*, AK-DQ786180.

The deduced amino acid sequences based on these cDNAs as well as deduced amino acid sequences for mitochondrial and protoflagellar CKs from the sponges *A. beatrix* and *S. fuscus* are shown in Fig. 1. The cDNAs amplified and sequenced in the present effort clearly code for true arginine kinases, as they contain the key, diagnostic AK catalytic residues as shown in the available x-ray crystal structures (Zhou et al. 1998; Yousef et al. 2003). AKs have a shortened "specificity loop" (Uda et al. 2006) located in the N-terminal region (residue positions 98–108 in Fig. 1) as well as a key glutamate residue (E_{382} in Fig. 1) that interacts with the arginine substrate (Zhou et al. 1998. In contrast, CKs have an extended "specificity loop" (see residues 98–108 in Fig. 1) and a key isoleucine residue in this loop (Ile_{104} in Fig. 1), which, in concert with a valine residue substituted for the E_{382} equivalent residue (V_{382} in Fig. 1), form a binding "pocket" for the methyl group of creatine (Lahiri et al. 2002; Novak et al. 2004).

To further validate whether the three choanoflagellates express AKs, we prepared crude lysates from pelleted cells and looked for AK and CK activities using spectrophotometric enzyme assays (Strong and Ellington 1996); only AK activity was observed in these lysates (M. Conejo, unpublished observations). Furthermore, a full-length cDNA ORF for *M. brevicollis* AK was inserted into a pETBlue1 expression vector (EMD Novagen, San Diego, CA), which was then used to transform *E. coli* expression hosts. Pilot expression experiments yielded large amounts of recombinant protein displaying AK enzymatic activity (M. Conejo, unpublished observations). Assays conducted on cell-free extracts of the choanoflagellate *Salpingoeca infusorium* (ATCC-50559) also yielded AK activity only (M. Conejo, unpublished observations); it seems likely that CK is absent in choanoflagellates. Expression constructs for the AKs from *A. beatrix* and *M. prolifera* were prepared, which also yielded soluble AK activity (M. Conejo, unpublished observations).

Phylogenetic and Sequence Analyses

The relationship of the choanoflagellate and sponge AKs to AKs from the ciliate *Tetrahymena*, the protozoan

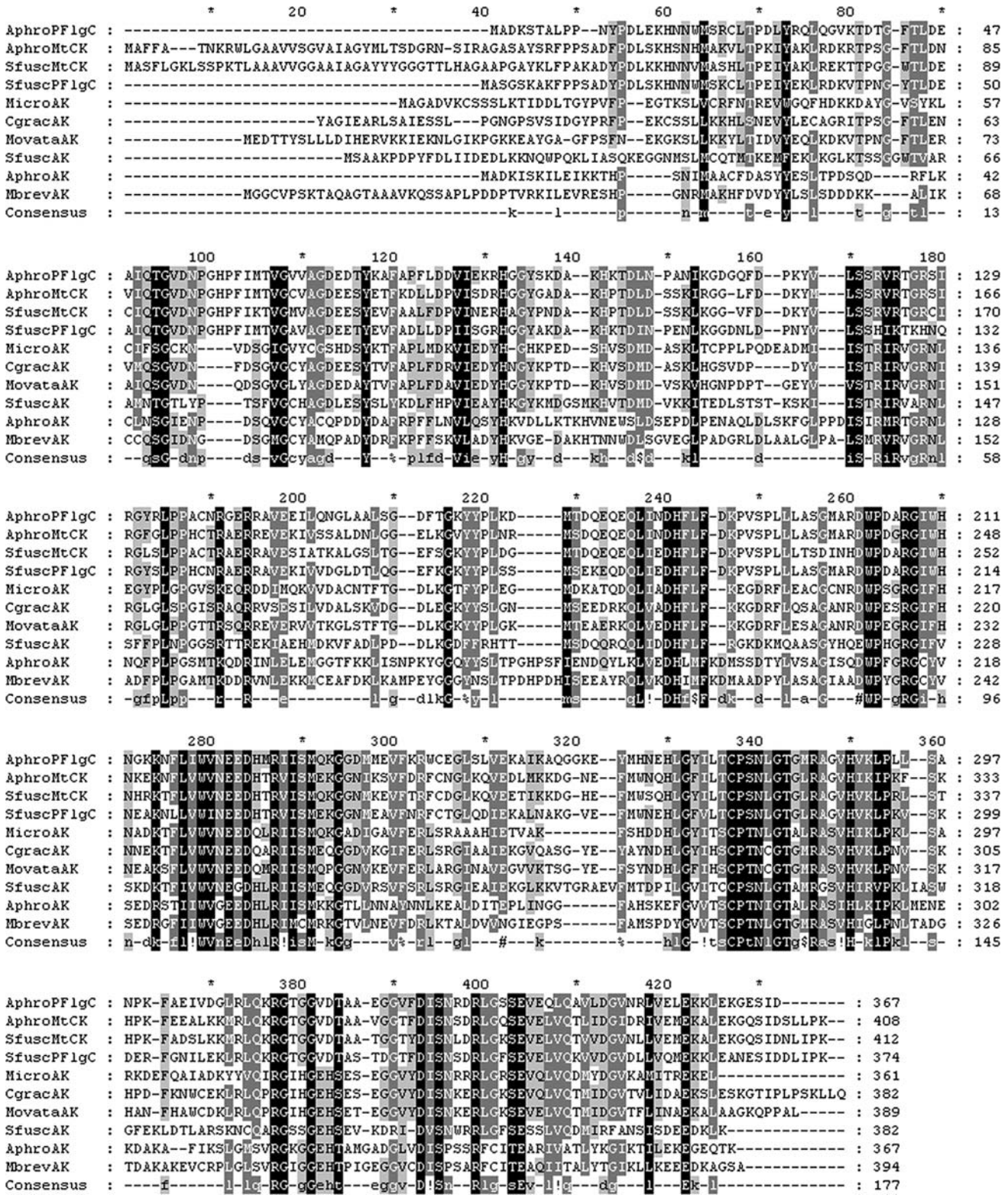

Fig. 1 Multalin (http://www.bioinfo.genopole-toulouse.prd.fr/multalin/) alignment of the deduced amino acid sequences of AKs from *S. fuscus* (SfuscAK), *M. prolifera* (MicroAK), *C. gracilis* (CgracAK), *M. ovata* (MovataAK), *A. beatrix* (AphroAK), and *M. brevicollis* (MbrevAK) as well as mitochondrial CKs from the sponges *A. beatrix* (AphroMtCK) and *S. fuscus* (SfuscMtCK) and protoflagellar CKs from *A. beatrix* (AphroPFlgCK) and *S. fuscus* (SfuscPFlgCK) previously reported by Bertin et al. (2007). Black, medium-gray, and light-gray shading corresponds to 100%, 80%, and 60% conserved, respectively. Alignment is viewed in GeneDoc obtained from the National Resource for Biomedical Computing (http://www.nrbsc.org/). Based on the crystal structure of horseshoe crab AK (Zhou et al. 1998), the juncture of the small N-terminal domain and the large C-terminal domain occurs at residue 156 in the alignment

Trypanosoma, the two domains of a contiguous dimeric AK from a cnidarians, and a broad range of ecdysozoan and lophotrochozoan AKs was analyzed. Several sponge CKs were used as an outgroup. The most striking feature of the resulting ML tree (Fig. 2) is the large assemblage of five distinct and well-supported AK clades. One of these consists of AKs from ecdysozoan and lophotrochozoan protostomes, each grouped into distinct lineages (Fig. 2). Note that the AK from *Trypanosoma* is within the ecdysozoans, a linkage which has been observed on many occasions (Uda et al. 2006) and is thought to be due to the fact that AK in this group may have been acquired by horizontal gene transfer from early arthropod hosts (Pereira et al. 2000).

The remaining clades within the large AK supercluster consist of the following: (a) the two domains of the contiguous dimeric AK from the sea anemone *Anthopleura japonica*, (b) the choanoflagellate *M. brevicollis* AK and the AK from the hextactinellid sponge *A. beatrix*, (c) the AKs from the sponges *Suberites fuscus* and *S. domuncula*, and finally, (d) a clade consisting of the AK from the sponge *M. prolifera* and the AKs from the ciliate *Tetrahymena thermophila* (Fig. 2).

Figure 2 shows that the *C. gracilis* and *M. ovata* AKs form a well-supported clade which has an intermediate position and is distinct from the CK "outgroup" and the large assemblage of protozoan and animal AKs. A cursory inspection of the AK-CK sequence alignment in Fig. 1 shows that both *C. gracilis* and *M. ovata* AKs, although true arginine kinases, resemble sponge CKs in a number of respects. This similarity is further reinforced when these sequences are subjected to pairwise amino acid comparisons by MatGat (Campanella et al. 2003) for percentage identity and similarity as shown in Table 1. The shaded

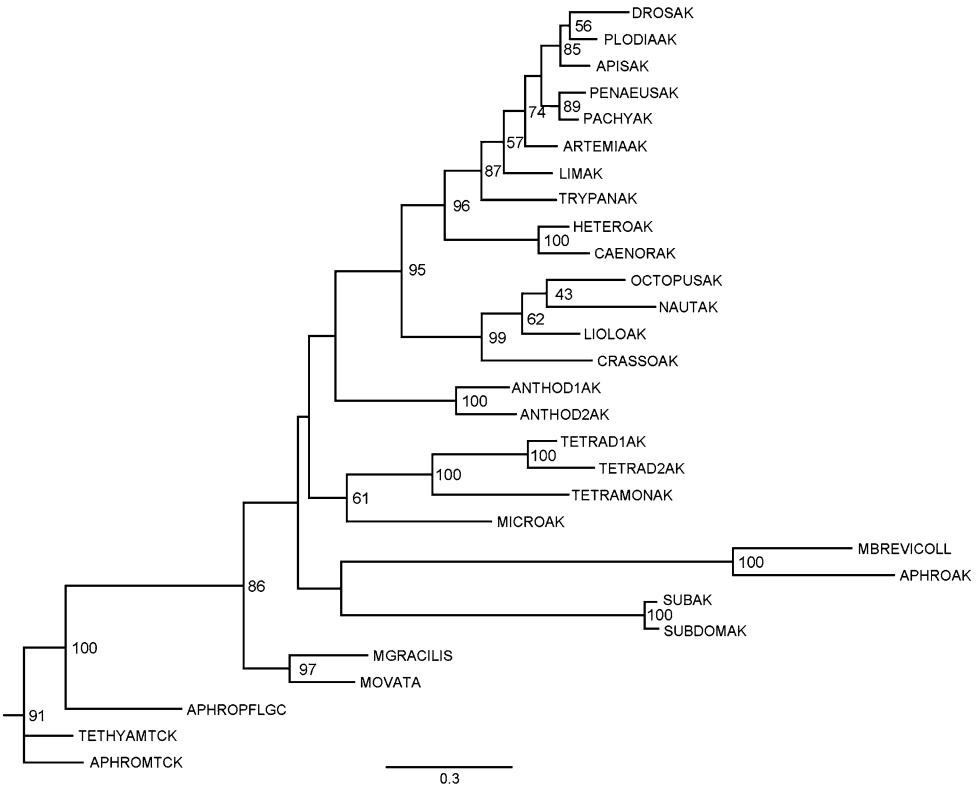

Fig. 2 Maximum likelihood tree of AK sequences using three sponge CK sequences as an outgroup. Numbers at nodes are bootstrap values. The order of the nodes resulting from the consensus tree at bootstrap values < 50 is not consistent with the node order that resulted from the ML tree, indicating that the order is not supported by the bootstrap values and are thus not shown. The horizontal bar corresponds to evolutionary distance (0.3 amino acid substitution per site). Arginine kinases: CGRACILIS, *C. gracilis* AK; MOVATA, *M. ovata* AK; APHROAK, *A. beatrix* AK; MBREVICOLI, *M. brevicollis* AK; SUBAK, *S. fuscus* AK; SUBDOAK, *Suberites domuncula* AK; MICROAK, *M. prolifera* AK; TETRA D1AK and D2AK, domains 1 and 2 of the contiguous dimeric AK from the ciliate *Tetrahymena*; TETRAMONAK, monomeric AK from *Tetrahymena*; ANTHO D1AK and D2AK, domains 1 and 2 of the contiguous dimeric AK from the sea anemone *Anthopleura*; CRASSOAK, oyster *Crassostrea* AK; LIOLOAK, gastropod *Liolophora* AK; NAUTAK, *Nautilus* AK; OCTOPUSAK, *Octopus* AK; HETEROAK, nematode *Heterodera* AK; CAENORAK, nematode *Caenorhabditis* AK; TRYPANAK, *Trypanosoma* AK; LIMAK, horseshoe crab *Limulus* AK; ARTEMIAAK, brineshrimp *Artemia* AK; APISAK, insect *Apis* AK; PLODIAAK, insect *Plodia* AK; DROSAK, insect *Drosophila* AK; PENAEUSAK, shrimp *Penaeus* AK; PACHYAK crab, *Pachygrapsus* AK. Creatine kinases: APHROMTCK, sponge *Aphrocallistes* mitochondrial CK; APHROPFCK, sponge *Aphrocallistes* protoflagellar CK; TETHYAMTCK, sponge *Tethya* mitochondrial CK. All sequences were obtained from available databases

Table 1 MatGat (Campanella et al. 2003) pairwise comparisons of the deduced amino acid sequences of AKs from *S. fuscus* (SfuscAK), *M. prolifera* (MicroAK), *C. gracilis* (CgracAK), *M. ovata* (MovataAK), *A. beatrix* (AphroAK), and *M. brevicollis* (MbrevAK) as well as mitochondrial CKs from the sponges *A. beatrix* (AphroMtCK) and *S. fuscus* (SfuscMtCK) and protoflagellar CKs from *A. beatrix* (AphroPFlgCK) and *S. fuscus* (SfuscPFlgCK) previously reported by Bertin et al. (2007)

	1	2	3	4	5	6	7	8	9	10
1. AphroPFlgCK	–	59.3	66.7	60.9	34.4	29.6	38.7	44.1	45.6	32.4
2. AphroMtCK	73.0	–	64.1	71.4	34.6	28.3	36.4	45.4	46.2	31.4
3. SfuscFlgCK	80.7	76.7	–	62.2	32.8	28.1	37.3	44.5	45.3	28.3
4. SfuscMtCK	72.1	83.3	72.8	–	31.7	26.4	35.8	45.1	46.0	29.4
5. SfuscAK	54.7	53.7	55.0	51.7	–	28.1	39.7	38.1	37.7	28.2
6. AphroAK	53.7	49.0	51.3	47.6	52.4	–	31.5	30.1	30.4	51.3
7. MicroAK	59.1	55.6	56.4	51.5	55.0	52.6	–	50.6	51.3	32.8
8. CgracAK	66.2	65.4	66.0	64.8	57.6	53.4	64.9	–	67.8	33.4
9. MovataAK	63.5	64.5	63.2	63.3	59.6	51.2	65.3	80.2	–	32.5
10. MbrevAK	50.3	49.8	48.2	47.3	48.0	66.2	50.8	52.2	50.5	–

Note. Comparisons used the default parameters and Blosum 62. The upper triangle contains the percentage identity values, while the lower triangle shows the percentage similarity values. Significance of the shaded cells is discussed in the text.

portions of the table correspond to comparisons of *C. gracilis* and *M. ovata* AKs with the four sponge CKs as well as the AK from the sponge *M. prolifera*. All show comparable degrees of percentage identity and similarity (Table 1). In contrast, *M. brevicollis* AK displays a high degree of similarity and identity to only the AK from the hexactinellid sponge *A. beatrix* (Table 1).

The positioning of *C. gracilis* and *M. ovata* AKs outside of the large assemblage of protozoan, basal metazoan, and eumetazoan AKs is, indeed, a very significant observation. All phosphagen kinases consist of a small, ~100-residue, N-terminal domain and a larger, 250+-residue, C-terminal domain (Zhou et al. 1998). The N-terminal domain undergoes significant conformational movements during catalysis, closing down on the catalytic pocket. Given these two distinct domains, we conducted phylogenetic analyses by ML of the small domains only and the large domains only of the sequences used in construction of the ML tree in Fig. 2 (domain juncture indicated in Fig. 1). The results for the former were not conclusive due to the short length of the sequences but the latter analysis positioned the *C. gracilis* and *M. ovata* AKs outside of the AK assemblage (data not shown).

At Least Three AK Genes Are Present in Choanoflagellates

The recent release of the *Monosiga brevicollis* genome dataset (http://www.genome.jgi-psf.org/Monbr1/Monbr1.home.html) allowed us to conduct tBLASTn searches of the genome scaffolds using the *M. brevicollis*, *C. gracilis*, and *M. ovata* deduced amino acid sequences as queries. Our searches were quite fruitful, yielding three AK genes—scaffold 2:116335, scaffold 2:733464, and scaffold 7:781101. The deduced amino acid sequences from the exons of each of these genes were assembled into AK sequences denoted S2:116335, S2:733464, and S7:781101, respectively. These sequences are aligned with the *M. brevicollis*, *C. gracilis*, and *M. ovata* deduced amino acid sequences in Fig. 3. The S2:116335 sequence is virtually identical to the sequence derived from the *M. brevicollis* cDNA except for two amino acid substitutions (Fig. 3). In contrast, S2:733464 and S7:781101 are very different from S2:116335 and *M. brevicollis* AKs and are very similar, but not identical, to each other. Of great interest is the fact that both S2:733464 and S7:781101 are much more similar to the cDNA derived amino acid sequences of *M. gracilis* and *M. ovata* AKs (Fig. 3).

There are three unique AK genes in the choanoflagellate *M. brevicollis*. The cDNAs that we have amplified from *C. gracilis* and *M. ovata* code for AKs that are very similar to two of the three *M. brevicolis* AK genes. While not absolutely definitive, a reasonable conclusion from our results is that at least three different AK genes are present in choanoflagellates. Our RTPCR amplifications seem to have serendipitously generated evidence for all three.

Implications with Respect to Phosphagen Kinase Evolution

As indicated previously, AKs are widespread in protozoans and basal and protostome invertebrates. CKs are thought to

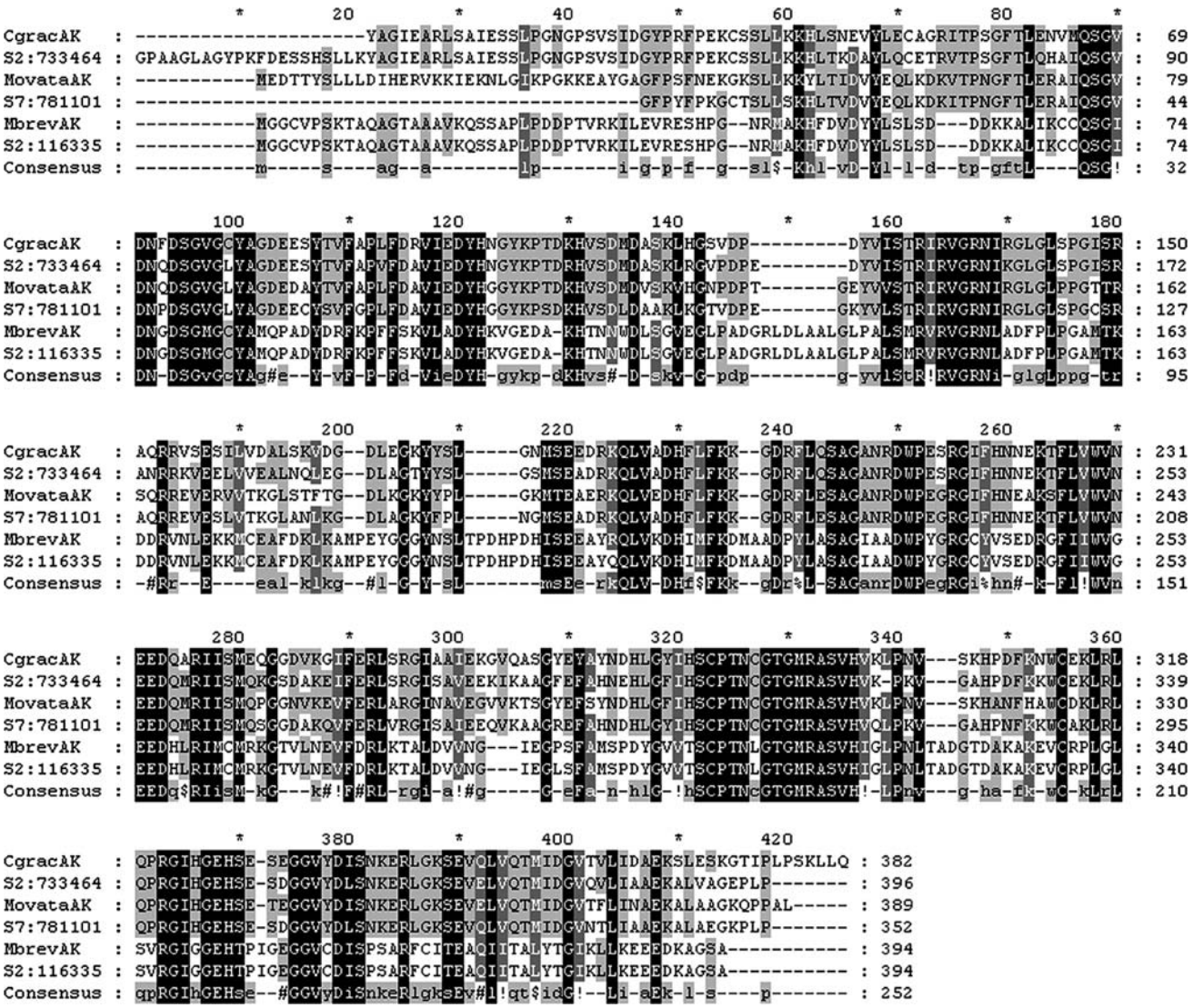

Fig. 3 Multalin (http://www.bioinfo.genopole-toulouse.prd.fr/multalin/) alignment of the sequences for *C. gracilis* (CgracAK), *M. ovata* (MovataAK), *M. brevicollis* (MbrevAK) AK scaffolds S7:781101, S2:733464, and S2:116335. Black, medium-gray, and light-gray shading corresponds to 100%, 80%, and 60% conserved, respectively. Alignment is viewed in GeneDoc obtained from the National Resource for Biomedical Computing (http://www.nrbsc.org/). Based on the crystal structure of horseshoe crab AK (Zhou et al. 1998), the juncture of the small N-terminal domain and the larger C-terminal domain occurs at residue 148 in the alignment

have evolved from an AK-like ancestor via an initial gene duplication event followed by acquisition of the structural correlates of creatine specificity and dimerization (Ellington 2001). A further series of duplication/divergence events led to extant family of CK isoforms (Suzuki et al. 2004). These early duplication/divergence events as well as the one leading to the formation of the mitochondrial and protoflagellar CK isoform genes occurred prior to the divergence of the hexactinellid sponges (Bertin et al. 2007). There is no evidence for any CK genes in the protozoan groups (Ellington and Suzuki 2006), yet not one, but two, CK genes are present in the oldest, extant class of sponges, the hexactinellids (Bertin et al. 2007). The origin of the CK gene(s) can thus likely be traced to earliest metazoans or their protozoan ancestors (Sona et al. 2004). Due to their pivotal position as a sister-group to the basal metazoans, the protozoan choanoflagellates have received considerable attention, especially with respect to their possession of some elements of the molecular infrastructure required for multicellularity (King et al. 2003; King 2004; King 2005).

Our present results demonstrate that at least three AK genes are likely present in choanoflagellates. Our ML tree (Fig. 2) shows that one of these AKs (from *M. brevicollis*) forms a clade with the AK from the hexactinellid sponge *A. beatrix* and is part of larger assemblage of protozoan, basal

metazoan, and eumetazoan AKs. In contrast, the AKs from *C. gracilis* and *M. ovata*, which are very similar to two of the AK genes mined from the *M. brevicollis* genomic database (Fig. 3), form a well-supported clade outside of the large assemblage of protozoan and animal AKs (Fig. 2). These AKs show extensive sequence similarities to sponge CKs and to at least one sponge AK. Given the topology of the ML tree and the sequence similarities, it is tempting to speculate that the multiple AK genes present in the chonaoflagellates constitute the molecular precursors for the divergence events that led to the creatine kinase lineage in animals.

The Physiological Roles of AK and CK in Sponges and Choanoflagellates

Phosphagen kinase reactions, like those catalyzed by AK and CK, mitigate spatial and temporal mismatches of ATP supply and demand in cells (Ellington 2001). The present data and prior work show that CK and AK are widespread in sponges. AK, but not CK, seems to be widespread in choanoflagellates. Ellington (2001) argued that the early physiological role of AK and CK reactions was to facilitate energy transport between ATP sources (mitochondria) and ATP sinks (peripheral ATPases). Perovic-Ottstadt et al. (2005) showed that AK expression was up-regulated in response to addition of silicate in the sponge *Suberites domuncula* suggesting a role for AK in energy homeostasis in sclerocytes of sponges. Sponges contain flagellated choanocytes which generate water currents. It has been suggested that CK could play a role in energy transport within these highly polarized cells (Bertin et al. 2007); a similar role could also be played by AK. Finally, flagella in choanoflagellates generate motility and water currents for feeding. The AK system present in these protozoans could mitigate reaction diffusion constraints with respect to energy transport between mitochondria in the cell body and the dynein ATPases along the length of the flagellum.

Acknowledgments This research was supported by National Science Foundation Grants IOB-0130024 and IOB-0542236 to W.R.E. We thank Steve Thompson of the School of Computational Science at Florida State University for assistance with the phylogenetic analyses. We thank Dr. Nicole King, University of California at Berkeley, for providing unpublished EST data as well as for giving advice on choanoflagellate culture procedures.

References

Bertin M, Pomponi SM, Kokuhuta C, Iwasaki N, Suzuki T, Ellington WR (2007) Origin of the genes for the isoforms of creatine kinase. Gene 392:273–282

Borson ND, Salo W, Drewes LR (1992) A lock-docking oligo (DT) primer for 5′- and 3′-RACE PCR. PCR Meth Appl 2:144–148

Burger G, Forget L, Zhu Y, Gray MW, Lang BF (2003) Unique mitochondrial genome architecture in unicellular relatives of animals. Proc Natl Acad Sci USA 100:892–897

Campanella JJ, Bitincka L, Smalley J (2003) MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinform 4:29

Ellington WR (2000) A dimeric creatine kinase from a sponge: implications in terms of phosphagen kinase evolution. Comp Biochem Physiol B 126:1–7

Ellington WR (2001) Evolution and physiological roles of phosphagen systems. Annu Rev Physiol 63:289–325

Ellington WR, Suzuki T (2006) Evolution and divergence of creatine kinase genes, In: Vial C (ed) Molecular anatomy and physiology of proteins: creatine kinase. Nova Science, New York, pp 1–27

Ellington WR, Yamashita D, Suzuki T (2004) Alternative splicing produces transcripts coding for alpha and beta chains of a hertero-dimeric phosphagen kinase. Gene 334:167–174

Frohman MA, Dus MK, Martin GR (1988) Rapid production of full length cDNAs from rare transcripts: amplification using a single gene specific primer. Proc Natl Acad Sci USA 85:8998–9002

King N (2004) The unicellular ancestry of animal development. Dev Cell 7:313–325

King N (2005) Choanoflagellates. Curr Biol 15:R113–R114

King N, Hittinger CT, Carroll SB (2003) Evolution of key cell signaling and adhesion protein families predates the origin of animals. Science 301:361–363

Kruse M, Leys SP, Müller IM, Müller WE (1998) Phylogenetic position of the *Hexactinellida* within the phylum *Porifera* based on the amino acid sequence of the protein kinase C from *Rhabdocalyptus dawsoni*. J Mol Evol 46:721–728

Lahiri SD, Wang PF, Babbit PC, McLeish MJ, Kenyon GL, Allen KN (2002) The 2.1 Å structure of *Torpedo californica* creatine kinase complexed with the ADP-Mg^{2+}-NO^{-} creatine transition-state analog complex. Biochem 41:13861–134867

Lang BF, Okelly C, Nerad T, Gray MW, Burger G (2002) The closest unicellular ancestors of animals. Curr Biol 12:1773–1778

Lavrov DV, Forget L, Kelly M, Lang BF (2005) Mitochondrial genomes of two demosponges provide insights into an early stage of animal evolution. Mol Biol Evol 22:1231–1239

Medina M, Collins AG, Silberman JD, Sogin ML (2001) Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit RNA. Proc Natl Acad Sci USA 98:9707–9712

Novak WRP, Wang P-F, McLeish MJ, Kenyon GL, Babbitt PC (2004) Isoleucine 69 and valine 325 form a specificity pocket in human muscle creatine kinase. Biochemistry 43:13766–13774

Pereira CA, Alkonso GD, Paveto MC, Iribarren A, Cabanas ML, Torres HN, Flawia MM (2000) *Trypanosoma* cruzi arginine kinase characterization and cloning. A novel energetic pathway in protozoan parasites. J Biol Chem 275:1495–1501

Perovic-Ottstadt S, Wiens M, Schröder HC, Batel R, Giovine M, Krasko A, Müller IM, Müller WE (2005) Arginine kinase in the demosponge *Suberites domuncula*: regulation of its expression and catalytic activity by silicic acid. J Exp Biol 208:637–646

Peterson FJ, McPeek MA, Evans DAD (2005) Tempo and mode of early animal evolution:inferences from rocks, Hox and molecular clocks. Paleobiology 31:36–55

Philippe H, Snell EA, Bapteste E, Lopez P, Holland PWH, Casane D (2004) Phylogenomics of eukaryotes: impact of missing data on large alignments. Mol Biol Evol 21:1740–1752

Robin Y, Guillou Y (1980) Quelque aspects du metabolisme enegrgetique les esponges. Comptes Rendus 148:121–126

Rokas A, King N, Finnerty J, Carroll SB (2003) Conflicting signals at the base of the metazoan tree. Evol Dev 5:346–359

Snell EA, Furlong RF, Holland PWH (2001) Hsp 70 sequences indicate that choanoflagellates are closely related to animals. Curr Biol 11:967–970

Sona S, Suzuki T, Ellington WR (2004) Cloning and expression of mitochondrial and protoflagellar creatine kinases from a marine sponge: implications for the origin of intracellular energy transport systems. Biochem Biophys Res Comm 317:1207–1214

Steenkamp ET, Wright J, Baldauf SL (2005) The protistan origin of animals and fungi. Mol Biol Evol 23:93–106

Strong S, Ellington WR (1996) Expression of horseshoe crab arginine kinase in *Escherichia coli* and site directed mutations of the reactive cysteine peptide. Comp Biochem Physiol 113B:809–816

Suzuki T, Furukohri T (1994) Evolution of phosphagen kinase. Primary structure of glycocyamine kinase and arginine kinase from invertebrates. J Mol Biol 237:353–357

Suzuki T, Kawasaki Y, Furukohri T, Ellington WR (1997) Evolution of phosphagen kinase. VI. Isolation, characterization and cDNA-derived amino acid sequence of lombricine kinase from the earthworm *Eisenia foetida*, and identification of a possible candidate for the guanidine substrate recognition site. Biochim Biophys Acta 1343:152–159

Suzuki T, Mizuta C, Uda K, Ishida K, Sona S, Compaan DM, Ellington WR (2004) Evolution and divergence of the genes for cytoplasmic, mitochondrial and flagellar creatine kinases. J Mol Evol 59:218–226

Uda K, Saishoji N, Ichinari S, Ellington WR, Suzuki T (2005) Origin and properties of cytoplasmic and mitochondrial isoforms of taurocyamine kinase. FEBS J 272:3521–3530

Uda K, Fujimoto N, Akiyama Y, Mizuta K, Tanaka K, Ellington WR, Suzuki T (2006) Evolution of the arginine kinase gene family. Comp Biochem Physiol D 1:209–218

Watts DC (1975) Evolution of phosphagen kinases in the chordate line. Symp Zool Soc London 36:105–127

Yousef MS, Clark SA, Pruett PK, Somasundaram T, Ellington WR, Chapman MS (2003) Induced fit in guanidino kinases—comparison of substrate-free and transition state analog structures of arginine kinase. Prot Sci 12:103–111

Zhou G, Somasundaram T, Blanc E, Parthsarathy G, Ellington WR, Chapman MS (1998) Transition state structure of arginine kinase:implications for catalysis of bimolecular reactions. Proc Natl Acad Sci USA 95:8449–8454