Sáez, A. G., Lozano, E. & Zaldívar-Riverón, A. 2009. Evolutionary history of Na,K-ATPases and their osmoregulatory role. *Genetica* 136(3):479-90.

<http://dx.doi.org/10.1007/s10709-009-9356-0>

*(Author version of the final manuscript. If you prefer to have the Genetica pdf copy,
please write to asaez AT mncn.csic.es)*

Evolutionary history of Na,K-ATPases and their osmoregulatory role

Alberto G. Sáez, Encarnación Lozano and Alejandro Zaldívar-Riverón

Department of Biodiversity and Evolutionary Biology, Museo Nacional de Ciencias
Naturales, CSIC, C/ José Gutiérrez Abascal 2, 28006 Madrid, Spain

Corresponding author: Alberto G. Sáez
Depto. Biodiversidad y Biología Evolutiva
Museo Nacional de Ciencias Naturales
C/ José Gutiérrez Abascal 2
28006 Madrid
Spain
Tel.: 34-914111328 ext. 1283
Fax: 34-915645078
E-mail: asaez@mncn.csic.es

Abstract

The Na/K pump, or Na,K-ATPase, is a key enzyme to the homeostasis of osmotic pressure, cell volume, and the maintenance of electrochemical gradients. Its α subunit, which holds most of its functions, belongs to a large family of ATPases known as P-type, and to the subfamily IIC, which also includes H,K-ATPases. In this study, we attempt to describe the evolutionary history of IIC ATPases by doing phylogenetic analysis with most of the currently available protein sequences (over 200), and pay special attention to the relationship between their diversity and their osmoregulatory role. We include proteins derived from many completed or ongoing genome projects, many of whose IIC ATPases have not been phylogenetically analyzed previously. We show that the most likely origin of IIC proteins is prokaryotic, and that many of them are present in non-metazoans, such as algae, protozoans or fungi. We also suggest that the pre-metazoan ancestor, represented by the choanoflagellate *Monosiga brevicollis*, whose genome has been sequenced, presented at least two IIC-type proteins. One of these proteins would have given rise to most current animal IIC ATPases, whereas the other apparently evolved into a lineage that, so far, has only been found in nematodes. We also propose that early deuterostomes presented a single IIC gene, from which all the extant diversity of vertebrate IIC proteins originated by gene and genome duplications.

Keywords

Na,K-ATPases, IIC-type ATPases, P-type ATPases, salt tolerance, osmoregulation, gene family, gene phylogeny.

Abbreviations

NKA: Na,K-ATPase

HKA: H,K-ATPase

α -NKA: α subunit of Na,K-ATPase

α -HKA: α subunit of H,K-ATPase

MRC: mitochondrial rich cells

Introduction

One of the most obvious examples of ecological adaptation is salt tolerance. When the salinity of an aquatic habitat is higher or lower than within the body fluids of the organisms living in it, those organisms must transport salts across their epithelia to keep osmolarity within physiological levels (alternative approaches are plant rigid walls, and matching the internal osmolarity to the external one, as osmoconformers do) (Krogh 1946).

To find out which genes are responsible for salt tolerance in animals, almost all related reports use the candidate gene approach, although see Kültz et al. (2007) for an exception. Most studies have focused on the Na,K-ATPase, but there is a list of other proteins also being investigated, such as the Na/K/2Cl cotransporter, the Na/H exchanger, the carbonic anhydrase, the V-type H-ATPase, the Cl/HCO₃ exchanger, and a few others (Towle 1997; Marshall 2002; Evans et al. 2005).

The most frequent source of evidence for a given protein to be considered relevant in osmoregulation is when its expression or activity is altered after the organism has gone from high to low salt, or vice versa, and specially when that expression is observed in osmoregulatory organs, such as gills. For example, the carbonic anhydrase, which catalyzes the conversion of CO₂ and H₂O to HCO₃⁻ and H⁺, increases its activity and mRNA expression in the gills of the euryhaline (i.e. survive wide variations of environmental salinity) crabs when exposed to low salinity water (Henry 2005; Serrano et al. 2007). Conversely, in the rainbow trout, acclimation from fresh to seawater also induces a transient increment of activity of carbonic anhydrase in the intestine (Grosell

et al. 2007). The salt-regulated proteins may not be directly responsible for the movement of ions from the medium into the organism (or vice versa). Rather, usually several of them act in collaboration with each other setting up electrochemical gradients across various membranes that, in combination, are able to keep a fitting osmotic equilibrium for the organism.

In this paper we focus on the enzyme whose involvement in osmoregulation have been most recognized, the Na,K-ATPase (NKA henceforth). First, we briefly summarize what is known about its role in such adaptation, and secondly, we build the most comprehensive phylogeny for NKAs up to date. We think that the study of the evolution of NKAs is an important step in understanding how the diversity of NKAs relates to their spectrum of functions, including osmoregulation.

The NKA was discovered in leg nerves of crabs in 1957 by Jens Skou (Skou 1957), who received the Nobel Prize for his work on this enzyme forty years later. This is a transmembrane protein that utilizes the energy from the ATP hydrolysis transporting, in the best studied cases, 3 Na⁺ from the cytoplasm to the extracellular compartment and 2 K⁺ in the opposite direction, carrying both cations against their respective gradients while maintaining also an electrical gradient. These electrochemical differences are required not only for keeping an osmotic balance, but for many other functions, such as favouring the transport of nutrients and amino acids into the cells, regulation of cell volume, and for the maintenance and restoration of the membrane potential in excitable nerve cells (Therien and Blostein 2000; Cereijido et al. 2004; Kühlbrandt 2004). The pump may contain three subunits: the largest one is the α subunit, that undergoes phosphorylation and conformational transition coupled to the ATP hydrolysis and ion

transport; the β subunit is required for targeting the α subunit to the plasma membrane; and the smallest one, the γ subunit, regulates the pumping activity in a tissue- and isoform-specific way (Garty and Karlish 2006). The three subunits exist in several isoforms depending on the organism and tissue under study. The α subunit belongs to a large superfamily of evolutionarily conserved transmembrane ATPases known as P-type, because they are primarily characterized for having a phosphorylated intermediate during the reaction cycle. There are various kinds of P-type ATPases: for instance type IIA (e.g. SERCA ATPases, or sarco(endo)plasmic reticulum Ca^{2+} -ATPase, which pump Ca^{2+} out of the cytosol into the lumen of the sarcoplasmic reticulum of muscle cells), or type IIIA (e.g. H^{+} -pumps that maintain an acidic gradient across the plasma membrane in plants and fungi). The α subunit of the Na,K-ATPase (henceforward α -NKA for short) belongs to type IIC, which also includes the α subunit of H,K-ATPases (α -HKA hereafter). Some of the last are found in stomach acid-secreting cells, transporting protons out and potassium inside of these cells (Axelsen and Palmgren 1998; Kühlbrandt 2004).

The role of NKA in osmoregulation is not fully elucidated, and varies depending on the organism or cell type under consideration (reviewed by Towle 1997; Marshall 2002 and Evans et al. 2005). Two general models are presented in figure 1: organismal hypoosmolarity (e.g. in seawater) and hyperosmolarity (e.g. in freshwater). Both scenarios place the NKA at the basolateral membrane of specialized cells, the mitochondrial rich cells (MRCs, or chloride cells), which are cells dispersed among regular epithelial cells and in osmoregulatory organs such as gills or the rectal gland of elasmobranchs. The hypoosmolarity or seawater model is more accepted and more general among cell types or organisms. It implies the active transport of Na^{+} into the

blood stream from the cytoplasm of the MRC, building a strong and local electric gradient across the body wall (positive inside), which then drives the transport of Na^+ , K^+ and Cl^- by the Na/K/2Cl cotransporter into the MRC from the blood or hemolymph, and later the passive excretion of Cl^- to the external medium through the cystic fibrosis transmembrane conductance regulator, located in the apical membrane of the MRC. The remaining Na^+ , following the built electric gradient, would leave the organism through leaky cellular junctions, despite its unfavorable chemical gradient. A K^+ channel would recycle this cation to the blood. In freshwater conditions, the NKA is also found at the basolateral membrane of MRCs, actively transporting the cytosolic Na^+ into the blood that enters the cell through a Na/H exchanger. This dual role of the NKA in hypo- and hyperosmolarity explains why its activity can increase in the acclimation from freshwater to seawater and vice versa, although this greatly depends on the organism and tissue under consideration (Holliday et al. 1990; Piermarini and Evans 2000; Madsen et al. 2007; Tsai and Lin 2007).

Our interest on the NKA and its role in osmoregulation stems from our work on the brine shrimp *Artemia*. This organism comprises several species widely distributed and highly halophilic, which are the only metazoans able to tolerate salt levels close to saturation (Abatzopoulos et al. 2002). We are generally interested in studying the molecular basis for such remarkable adaptation. In collaboration with colleagues, one of us (AGS) have previously studied the histological and temporal pattern of expression of the two known isoforms of α subunit of NKA in *A. franciscana*: $\alpha 1$ (Macías et al. 1991) and $\alpha 2$ (Baxter-Lowe et al. 1989). We showed that both isoforms are strongly expressed in osmoregulatory organs although with a different temporal and spatial pattern (Escalante et al. 1994 and 1995). We also showed that the gene coding for $\alpha 1$

shows a high degree of intraspecific variability, perhaps due to balanced polymorphisms (García-Sáez et al. 1997; Sáez et al. 2000). Interestingly, it has been recently shown that $\alpha 2$ increases its mRNA levels when the external salinity is raised, whereas $\alpha 1$'s expression apparently remains unaffected (Jorgensen and Amat 2008).

As mentioned before, in this article we focus on the diversity and evolution of α -NKAs. Some studies have already investigated the evolution of P-type ATPases (Palmgren and Axelsen 1998; Axelsen and Palmgren 1998), or even more specifically the evolution of α -NKAs (Okamura et al. 2002; Anderson et al. 2004). However, the many newly completed genomes projects, such as the one of choanoflagellate *Monosiga brevicollis* (King et al. 2008), or of the crustacean *Daphnia pulex* (a water flea species; *Daphnia* Genomics Consortium, unpublished), as well as the constant growth of protein sequence databases, recommend an update on this problem.

Materials and methods

Collection of sequences

We used UniProt at <http://beta.uniprot.org> (The UniProt Consortium 2008) to perform a blast search using the $\alpha 1$ subunit of *A. franciscana* as the query sequence (UniProt accession number P28774; Macías et al. 1991). We used the database UniRef100 (release 13.0) on 14-March-2008 using default searching options. UniRef100 combines manual and automatically annotated protein sequences from two different databases, UniProtKB and UniParc, avoiding the redundancy of identical sequences. We checked

that the clustered identical sequences belonged to the same species, and so that they only represented duplicated entries. At this point, no fragments smaller than 800 amino acids were selected (P28774 has 1004 residues). To assess the cut-off score value to use from blast, we considered the observed blast score between our query sequence and the two most basal IIC sequences according to Palmgren and Axelsen (1998): C09H5.2 and C02E7.1; scores 557 and 550, respectively. We used a lower cut-off score, 440, in order to minimize the possibilities of excluding from our data set true IIC proteins. Sequences with lower blast scores were excluded from our final data set, although we also inspected the list of excluded proteins and saw no apparent IIC proteins. A total of 229 proteins were selected from UniProtKB and 80 from UniParc by our blast on UniRef100. From this large set of 309 sequences, and in order to exclude non-informative sequences that would slow our phylogenetic analyses, we eliminated proteins for which we could find evidence that they were coded by single genes (i.e., splice variants, same genomic location); or for which we considered that possibility a very likely one, due to an equal or larger identity of 98% between proteins from the same species. We also dropped from our database several sequences with clear orthologous in closely related species, such as all *Caenorhabditis briggsae*'s sequences (we had the *C. elegans*'s homologues). After having compiled this data set we searched type IIC proteins directly into the genomes of the following model organisms: *Apis mellifera* (The Honeybee Genome Sequencing Consortium 2006), *Caenorhabditis elegans* (*C. elegans* Sequencing Consortium 1998), *Danio rerio* (Sanger Institute, unpublished), *Daphnia pulex* (Daphnia Genomics Consortium, unpublished), *Dictyostelium discoideum* (Eichinger et al. 2005), *Drosophila melanogaster* (Adams et al. 2000), *Homo sapiens* (Lander et al. 2001), *Monosiga brevicollis* (King et al. 2008), *Nematostella vectensis* (Putnam et al. 2007), *Strongylocentrotus purpuratus* (Sea

Urchin Genome Sequencing Consortium 2006), *Tribolium castaneum* (Tribolium
 Genome Sequencing Consortium 2008) and *Xenopus laevis* (US Department of Energy
 Joint Genome Institute, unpublished). After a careful search in each of these genomes,
 we only found two published sequences not present in our UniProt data set, and which
 are both among the most divergent α -NKAs (figure 2): catp-1 (*C. elegans*), and
 CG3701 (*D. melanogaster*) (Okamura et al. 2002). Thus, although we found a good
 match between both searches, in UniProt and in the genome surveys, there was the
 possibility that we could have missed similarly divergent proteins from other species.
 Therefore, we performed a second search in Uniprot using catp-1 and CG3701 as
 probes. This way we found CG3701's ortholog *Drosophila pseudoobscura*'s GA17624,
 and two additional proteins from the choanoflagellate *Monosiga brevicollis*: ID10758
 and ID28085. In June 2008, while revising this manuscript, we added to our data set six
 new sequences from a new search in an updated version of Uniprot. We also obtained
 more sequences from very recently sequenced genomes not yet available in Uniprot:
 one from the amphioxus (Cephalochordata) *Branchiostoma floridae* (Putnam et al.
 2008), two of the annelid *Capitella capitata* (US Department of Energy Joint Genome
 Institute, unpublished), and six from the crustacean *Daphnia pulex* (US Department of
 Energy Joint Genome Institute, in collaboration with the *Daphnia* Genomics
 Consortium, unpublished). Two more genomes were searched with no additional hits
 (with respect to Uniprot): *Anopheles gambiae* (Holt et al. 2002) and *Ornithorhynchus*
anatinus (Warren et al. 2008). Finally, we also included three outgroups sequences: *A.*
franciscana's sarcoplasmic reticulum Ca-ATPase (SERCA-ATPase; a IIA P-type
 ATPase), Ca-ATPase mca-1 (IIB), and Na-ATPase1 ENA1 (IID). We included three
 outgroups to be able to identify the root of the phylogenetic tree, and to have a sense of
 the genetic distances within the IIC set in relation to other P-type ATPases. Our data set

contained 224 sequences. We include as supplementary information the list of UniProt sequences used in this study.

Phylogenetic analyses

We used MAFFT v5 with default values to align the 224 sequences (Katoh et al. 2005). In order to avoid the least reliable aligned regions we used Gblocks v0.91b at the Gblocks server (Castresana 2000), with the most permissive conditions (otherwise too few sites would remain). In the final alignment only 22% of the original positions stayed, i.e. 730 out of 3315. Simulation experiments have shown that the phylogenetic trees obtained after excluding less reliable sites are closer to the true ones, despite the discard of data (Talavera and Castresana 2007). We also excluded two non-informative sequences (Uniprot codes Q6PIC6 and UPI000065FCD1) because they were almost identical to two other sequences after Gblocks; thus our final data set included 222 protein sequences. To find out the most appropriate model of amino acid substitution for our data set, we used ProtTest v1.2.7, which tested 70 models (Abascal et al. 2005). By far, the best fitting model was RtREV+I+G+F, that is, the RtREV substitution matrix considering also a percentage of invariable sites, a gamma distribution of site variation and unequal amino acid frequencies. We used PhyML v3.0 to build the phylogenetic tree, a maximum likelihood program (Guindon and Gascuel 2003), where we selected RtREV+I+G+F as the model of choice. Although we run several times PhyML, changing some optional parameters, and the results were very similar, we finally chose a BioNJ initial tree and a tree topology search based in “best of NNIs and SPRs”, which gave us the highest likelihood value for the selected model. To assess the reliability of the outcoming tree, we run PhyML on 1000 bootstrap replicates, and built

from them a majority-rule consensus tree using PAUP* (Swofford 2000), from where we obtained the non-parametric bootstrap values. In order to speed up computing time, during bootstrap we used the gamma shape parameter and the proportion of invariable sites that were estimated by PhyML from the full data set (0.901 and 0.016 respectively).

To specifically focus on the evolution of vertebrate α -NKAs and α -HKAs, we ran a second phylogenetic analyses using a subset of these proteins, following again all the steps previously mentioned: alignment (MAFFT), selection of conserved blocks (Gblocks), model selection (ProtTest), and phylogenetic search with 1000 replicates bootstrap (PhyML). This data set had 41 proteins, including echinoderm and cephalochordate outgroups, and excluding several proteins with clear orthologs already present in the database. Here the best-fitting model found was WAG+I+G+F, and the final number of sites was 979 (82% of the original 1183 positions).

Results and discussion

In this study we reconstruct the most comprehensive phylogenetic tree produced so far for α -NKAs and the highly related α -HKAs, both known as IIC P-type ATPases. As explained in Materials and methods, it is made of a selection of 219 proteins sequences plus three outgroup sequences belonging to other P-type ATPase subfamilies (i.e. IIA, IIB, IID). Sequences were chosen on the basis of their similarity to a few α -NKAs, whilst excluding other well-characterized P-type ATPases (except for the outgroup), and therefore some of them may not be Na/K (or H/K) pumps, as the vast majority of

the correspondent enzymes have never been biochemically tested. Keeping this consideration in mind, the first thing we can say about the tree in figure 2 is that the origin of IIC proteins is probably prokaryotic, as the three outgroup sequences are found within the prokaryotic clade. This prokaryotic clade is made of a combination of bacterial and archaeal sequences, and is well supported (87% bootstrap; $\geq 70\%$ bootstrap is usually indicative of good support (Hillis and Bull 1993)). However, to our knowledge, no prokaryotic NKA activity has been described yet.

Most fungal sequences cluster in a very well supported clade (100% bootstrap). The proteins that have been studied at the functional level in this group are known as ACU (i.e., alkali cation uptake) ATPases. They have been shown to be transporters that mediate the uptake of K^+ and Na^+ (the later with less affinity), although apparently not in exchange of one another (Benito et al. 2004). A new subfamily of P-type ATPases, IIE, has been proposed to account for ACU proteins, especially taking into consideration their distinct phylogenetic position (Benito et al. 2004; figure 2).

Three additional proteins from fungi, two of them known as PAT1 and PAT2 (i.e. P-type ATPases 1 and 2), are placed in a second fungal clade, also supported with 100% bootstrap, and not closely related to the first clade. Phosphoenzyme formation assays suggest that PAT activity corresponds to a Na,K-ATPase, which if confirmed would imply that the ancestral NKA preceded the split between fungi and metazoa (Fietto et al. 2002). In fact, many other non-metazoan sequences appear on the tree of figure 2. Functional studies have suggested that at least one of them, rhodophyta *Porphyra yezoensis*'s kpa-1 (i.e., K^+ P-type ATPase), is a NKA (Barrero-Gil et al. 2005). However, it is possible that important differences exist with respect to typical NKAs. In

the case of stramenopile *Heterosigma akashiwo*'s HANA (*H. akashiwo* Na⁺-ATPase), it has been suggested that this protein is unlikely to interact with a β subunit and that it is resistant to the glycoside ouabain, being both features contrary to those of typical NKAs (Shono et al. 2001). The phylogenetic relationship among the non-metazoan putative NKAs presents many uncertainties (figure 2). Protozoan sequences form a distinct clade corresponding to nine genes from *Paramecium tetraurelia* and 14 from *Tetrahymena thermophila*. With a 99% bootstrap value supporting this protozoan clade, it appears that the evolution of these proteins have emerged through multiple gene duplications, which is no exception in the genome of those two species (Aury et al. 2006; Eisen et al. 2006). In *T. thermophila*, we found that most of the correspondent genes are physically linked, which suggests their relatively recent origin. It is quite interesting to note the high support (88%) of a clade made of three isoforms from nematodes (*Caenorhabditis*), two from the choanoflagellate *Monosiga brevicollis*, and one from the oomycete *Pythium aphanidermatum*. One of the *Caenorhabditis* proteins, catp-1, has been assigned a role in developmental timing that is independent of its predicted pumping function (Ruaud and Bessereau 2007). On the other hand, the proximity of these nematode and *M. brevicollis* ATPases is actually expected, as choanoflagellates are the sister group of metazoa (King et al. 2008). However, that the sequence of an oomycete turns out to be closely related to animal and choanoflagellate sequences is somewhat surprising. It is known that fungi and animals are relatively close phylogenetically, being amoebozoa (mycetozoa, *Dictyostelium discoideum*) next to them in the tree of life (all represented in figure 2). Oomycetes, however, are not fungi but stramenopiles, a more distant group also including brown algae, diatoms, and others (see <http://tolweb.org/Eukaryotes/3>).

The lower part of the tree of figure 2, in grey, is shown expanded in figure 3. This part includes only animal IIC ATPases that, generally, may be considered bona fide α -HKAs - and α -NKAs (Mense et al. 2002; Jorgensen et al. 2003). There is no bootstrap support for the separation of both sub-trees (i.e., below 50%). However, we noticed that this instability was caused by two very fast evolving sequences in our tree: *Drosophila melanogaster*'s CG3701 and its *D. pseudoobscura*'s ortholog (whose clade is indicated by an asterisk in figures 2 and 3). If we ignore the existence of these two proteins in our tree, the bootstrap support for the separation of the sub-tree of figure 3 is actually very high (98%). Moreover, four more internal branches rise above the 50% bootstrap value in that case (all indicated by asterisks in figures 2 and 3). Nevertheless, even without CG3701 proteins, the phylogenetic relationships among the differently coloured clades in figure 2 (only online version of this article) remain poorly supported.

It is important to focus on the position of *M. brevicollis*'s sequences in our tree. As previously mentioned, this species represents, as a choanoflagellate, a resembling living relative of the pre-metazoan ancestor (King et al. 2008), and so its genome can shed light on the origins of animal α -NKAs. We have found three putative α -NKAs in *M. brevicollis*. In figure 2 we can find two of them next to the only animal (nematode) sequences indicated in this figure. One of the proteins, annotated as ID10758, has only 284 amino acids; the other, ID28085, is also shorter than all the rest of α -NKAs in our alignment: 666 amino acids. Although these two proteins present little similarity to each other and very short overlapping, their close phylogenetic position suggest that they may have a common origin perhaps subsequent to the split between choanoflagellates and metazoans. The third *M. brevicollis* IIC ATPase is located at the base of the tree of "core" animal α -NKAs (figure 3), with ID35405 (1042 amino acids).

Therefore, it appears as if the ancestral pre-metazoan genome coded for at least two α -NKA-like proteins. One would have given origin to those putative α -NKAs today present in nematodes (catp-1, C09H5.2 and C02E7.1 in *Caenorhabditis elegans*; with identified *C. briggsae* orthologs, see Material and methods), but without known representatives in other animal lineages. In contrast, a second pre-metazoan α -NKA, similar to *M. brevicollis* ID35405, would have evolved into the vast majority of current animal α -NKAs and α -HKAs (figure 3).

There has been speculation on the properties of the ancestral animal α -NKAs (Okamura et al. 2002), and the *M. brevicollis* ID35405 sequence gives us the opportunity to enter this discussion. So, does this α -NKA share two of the unique features of most α -NKAs, namely inhibition by the cardiac glycoside ouabain and assembling with a β subunit? There are many residues key to ouabain binding that have been identified (Croyle et al. 1997). One of the most relevant is Asp121 (numbered according to sheep α 1), because when substituted by Glu, using site-directed mutagenesis, it confers great resistance to the drug. The equivalent position in ID35405 is precisely occupied by Glu, and thus, it is consistent with the idea that the ancestral pre-metazoan protein was resistant to ouabain. What about the α/β assemble? The main amino acids responsible for this feature, a SYGQ tract between the seventh and eighth transmembrane domains (Okamura et al. 2002), are present in ID35405, and therefore it could be expected that *M. brevicollis* had a β -NKA subunit. However, searching it through the current version of its genome (1.0), we could not find a single hit for such protein, despite numerous searches relaxing blast parameters. We also searched ouabain binding sites and α/β assembling amino acids in *M. brevicollis*' ID28085 and ID10758, but we did not find them.

Although to less extent than in figure 2, in figure 3 we also see few supporting bootstrap values (i.e. >70%) for relationships between major clades. One of them is for the separation between vertebrate α -NKAs and α -HKAs from the rest of the sequences (85%). In a separate analysis focusing on the vertebrate sequences alone this value is even higher (100%; figure 4). Interestingly, at the base of the vertebrate clade we find two sequences, one from the cephalochordate *Branchiostoma floridae* (amphioxus) and another from the echinoderm *Strongylocentrotus purpuratus* (sea urchin). We only found one α -NKA isoform in each of these two sequenced genomes. Given that cephalochordates represent the most basal extant chordate lineage, whereas echinoderms (together with hemichordates) are the closest living relatives of chordates (Putnam et al. 2008), our tree suggests that all vertebrate α -NKAs and α -HKAs probably evolved from a single deuterostome α -NKA.

Judging by bootstrap support (85%), protostome α -NKAs in figure 3 could perhaps fall into two major groups: one made of seemingly more divergent insect and nematode sequences (upper part of figure 3), and a second more inclusive group having α -NKAs not only from insects and nematodes, but from the remaining taxa (annelids, molluscs, crustaceans, etc; lower part of figure 3). In relation to the first class of α -NKAs, the tree suggests that an old lineage of them would have diversified in insects, persisted in nematodes, and disappeared from the remaining animal groups represented in figure 3. It is striking to notice the length of the clade made by *Drosophila melanogaster*'s CG3701 and its *D. pseudoobscura*'s ortholog (indicated by an asterisk; figures 2). As previously explained, the position of this clade is particularly uncertain. In other maximum likelihood trees (under a different model of evolution, or in bootstrap

replicates) it branches off early from the main axis of the tree in figure 2, which could be due to the long-branch attraction artefact (Felsenstein 1978). In those cases, again it lacks bootstrap support and it appears unrelated to the animal sequences indicated in figure 2. Pairwise comparisons between *D. melanogaster* and *D. pseudoobscura* for the three *Drosophila* α -NKAs in our tree show that the rate of evolution of CG3701-like proteins is much faster than for the other two (i.e., 42.2% of identity versus 79.6% and 96.9%, respectively). Consistent with this observation, we only found CG3701-like annotated proteins within the *Drosophila* genus, albeit in all twelve *Drosophila*'s whose genomes have been sequenced (Clark et al. 2007; <http://flybase.org/blast/>).

Among the rest of invertebrate α -NKAs, the crustacean *Daphnia pulex* (water flea) presented six sequences in the same clade (figure 3). Three of the corresponding paralogous genes are found contiguous in the same genome project scaffold (No. 91) and forming a well supported clade (100%), and thus, they are expected to be the result of recent tandem duplications. Perhaps this α -NKA variation could represent part of the adaptation of *D. pulex* to resist different salinities (Weider and Hebert 1987). The two known α -NKAs from the brine shrimp *Artemia franciscana*, α 1 and α 2, appear in the same clade than *D. pulex*'s α -NKAs, and we have preliminary data about two more isoforms. Thus, it may be that the adaptation of invertebrates to osmotic fluctuating environments could have forced in them the evolution of an array of α -NKA isoforms. In contrast, marine organisms represented in figure 3, such as sea urchins or sea anemones, present only one or two α -NKA isoforms. Consistent with the idea that osmoregulatory fluctuations drive the diversity of paralogous α -NKAs is the observation that two highly conserved amino acids, in transmembrane domains fourth and fifth, vary in their electrical charge among the different copies of *D. pulex* and *A.*

franciscana, but not in other invertebrates. These two amino acids are Asn324 and Asn776 (numbers according to pig $\alpha 1$). Jorgensen and Amat (2008) recently pointed out that these two positions, when occupied by a positively charged amino acid, like in *A. franciscana*'s $\alpha 2$ (Lys in both), reduce the number of ions transported per catalytic cycle, enabling the enzyme to work under stronger osmotic pressure. Two of the six *D. pulex*'s isoforms also have Lys at one or two of these sites (GNO_342614 shows Tyr and Lys, respectively, and GNO_290684 Lys in both). In marked contrast, none but six other sequences in our 222 proteins alignment showed a basic charged residue in any of these two sites. Four of these changes are in metazoans: zebrafish (*Danio rerio*) closely related isoforms $\alpha 1a.2$, $\alpha 1a.3$ and $\alpha 1a.5$, plus one trout ortholog (figure 4), all with an Asn776Lys substitution.

Colina et al. (2007) has studied four α -NKA sites that are more basic in the squid (also in our tree in figure 3, i.e. Mollusc) than in the human $\alpha 1$ isoform. They claim that these four substitutions reduce the sensitivity to external Na^+ , helping osmoconformers to adapt to marine environments. Looking at these positions in our alignment, we do not see a clear pattern of substitutions related to the environmental conditions in which the organisms live in. However, three of these sites display striking substitutions. Two of them turn to the even more basic Lys (compared to the squid): Tyr127Lys in the sea urchin, and Gly319Lys in one of the two sea anemone isoforms. The third site, Lys914 in the squid (and Ile at the human $\alpha 1$), becomes an acidic Glu in *Hydra vulgaris* and in one of the sequences from *D. pulex*, perhaps relating to their freshwater style.

In relation to vertebrate α -NKAs and α -HKAs, we have conducted a separate analysis only for them (figure 4), apart from the one performed for our whole database (figure

3). We did that in order to avoid very divergent sequences, and thus, obtaining a more reliable alignment (and hence tree). This way, after we run Gblocks, the software to select the most unquestionable parts of an alignment (Castresana 2000), we obtained more positions from the vertebrate matrix than for the whole data set (979 vs. 730, respectively). Therefore, we think that the tree shown in figure 4 is better founded than the vertebrate part of the tree in figure 3, although both results can be considered. As explained in Material and methods, in the vertebrate separate analysis we also present a more readable tree (figure 4) by excluding some vertebrate sequences, especially from mammals, already represented in our analysis by unambiguous orthologs.

Within the monophyletic (100% bootstrap; figure 4) vertebrate clade, we observed two major clades, α -NKAs and α -HKAs. The evolution of α -HKAs appears more reliable, because of higher bootstrap supports, than the evolution of α -NKAs. We also noticed a higher rate of amino acid substitution in α -HKAs than in α -NKAs, indicated by the longer branches of the first, which is probably related to their functional shift, from transporting sodium to proton. There are two kinds of α -HKAs: gastric and nongastric. The first are responsible for acidifying the stomach and secreting acid in kidneys. Nongastric α -HKAs are less known, and they are expressed in the mammalian colon and prostate (Silver and Soleimani 1999). Seemingly, they both are missing from most living fishes, although the gastric one has been isolated from the Atlantic stingray, an elasmobranch (cartilaginous fishes; Choe et al. 2004). We have searched for α -HKAs in the ongoing genome project of the zebrafish (*Danio rerio*) but to no avail. Thus, it appears that α -HKAs were already present in our most recent common ancestor with cartilaginous fishes, because of their presence in elasmobranchs, probably disappearing later from teleosts (boney fishes). We also noticed the apparent loss of gastric α -HKAs

in the avian and monotreme lineages, with the chicken and the platypus genomes already available (International Chicken Genome Sequencing Consortium 2004; Warren et al. 2008). There is an outlier sequence from *Xenopus tropicalis* at the base of the gastric α -HKA clade. This sequence is described as a nongastric α -HKA in UniProt (Q5BL65), although we have noticed that it presents a specific Glu residue characteristic of gastric α -HKAs in M6 (transmembrane domain number six; Okamura et al. 2002). There are no representative sequences from *X. tropicalis* in the nongastric clade, although we found two from *X. laevis* and one from *Bufo marinus*, and so one possibility is that the outlier *X. tropicalis* protein derived in fact from a nongastric α -HKA, suffering later a convergent process of evolution towards gastric α -HKAs. We also noticed that *Monodelphis domestica*, the gray short-tailed opossum, the first marsupial whose genome has been sequenced (Mikkelsen et al. 2007), has three contiguous gene copies coding for nongastric α -HKAs instead of the single one found in eutherians (e.g. rat, human, mouse, dog) and monotremes (i.e. platypus), and therefore this triplication appears to be specific to the opossums or perhaps marsupials (metatherians) at large.

Vertebrate α -NKAs are the most studied α -NKAs, especially with regard to mammalian isoforms, α 1, α 2, α 3 and α 4, which present different expression patterns and functions (Blanco and Mercer 1998). Roughly speaking, our trees support four vertebrate clades corresponding to these isoforms (figures 3 and 4). However, the evolutionary relationships among these four clades are uncertain. Besides, some of the proteins do not easily fall in one of these four groups (figure 4). One of these proteins is a platypus sequence at the base of the unsupported (i.e., bootstrap <50%) α 1 clade. Actually, if we consider this protein as part of the α 1 clade, then the clade has a support of 70%.

However, its phylogenetic position is awkward, emerging from a very old node, ancestral to birds, amphibious and fishes. Nevertheless, the low bootstrap support for the nodes between the two *O. anatinus* sequences in the $\alpha 1$ clade, together with the 70% support for that $\alpha 1$ clade, suggest to us that the outlier platypus sequence is a fast evolving $\alpha 1$. The second protein whose position is unclear is *D. rerio*'s $\alpha 2$ (figure 4). Despite of its name, we do not find it in the $\alpha 2$ clade. However, there is no bootstrap support for its current position in our tree, and so we cannot contradict its current classification. If we take that classification as correct, then we share with fishes all mammalian α isoforms except $\alpha 4$, which has only been found in therian mammals (marsupials and eutherians). It is also interesting to observe the long lengths of $\alpha 4$ branches in our tree (figure 4), which are probably due to the functional specialization of $\alpha 4$'s, which are expressed mainly in testis (Blanco et al. 2000). It is also of interest to note that boney fishes, or at least their model representative, *Danio rerio*, whose genome is in an advanced stage of completion, has six paralogous $\alpha 1$ genes ($\alpha 1a.1$, $\alpha 1a.2$, $\alpha 1a.3$, $\alpha 1a.4$, $\alpha 1a.5$ and $\alpha 1b$), two $\alpha 3$'s ($\alpha 3a$, $\alpha 3b$), and as discussed previously, probably one $\alpha 2$. It has been argued that this rich diversification has arisen through a combination of gene and genome duplications (Serluca et al. 2001). The functional role of these multiple isoforms is far from being understood, but it is thought that, at least in part, may have to do with the adaptive response to osmoregulatory changes (Richards et al. 2003; Bystriansky et al. 2006).

Acknowledgements

We thank Gernot Segelbacher and Jacob Höglund for the opportunity to make this

contribution and for organizing the Ecological Genomics meeting. We are supported by programs I3P-Postdoc (EL), Juan de la Cierva (AZR) and Ramón y Cajal (AGS), whereas our research is funded by Plan Nacional I+D (CGL2005-00307), all from the Ministry of Education and Science, Spain. We also thank Bioportal (University of Oslo) and Centro Técnico de Informática (CSIC) for their computing resources, and all the genomic consortiums for making their databases available.

References

Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104-2105

Abatzopoulos TJ, Beardmore JA, Clegg JS, Sorgeloos P (2002) *Artemia*: basic and applied biology. Kluwer Academic Publisher, Dordrecht

Adams MD, Celniker SE, Holt RA et al (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287:2185-2195

Axelsen KB, Palmgren MG (1998) Evolution of substrate specificities in the P-type ATPase superfamily. *J Mol Evol* 46:84-101

Anderson FE, Cordoba AJ, Thollesson M (2004) Bilaterian phylogeny based on analyses of a region of the sodium-potassium ATPase beta-subunit gene. *J Mol Evol* 58:252-268

Aury JM, Jaillon O, Duret L et al (2006) Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*.

Nature 444:171-178

Baxter-Lowe LA, Guo JZ, Bergstrom EE, Hokin LE (1989) Molecular cloning of the Na,K-ATPase alpha-subunit in developing brine shrimp and sequence comparison with higher organisms. FEBS Lett 257:181–187

Barrero-Gil J, Garcíadeblás B, Benito B (2005) Sodium, potassium-atpases in algae and oomycetes. J Bioenerg Biomembr 37:269-278

Benito B, Garcíadeblás B, Schreier P, Rodríguez-Navarro A (2004) Novel P-type ATPases mediate high-affinity potassium or sodium uptake in fungi. Eukaryot Cell 3:359-368

Blanco G, Mercer RW (1998) Isozymes of the Na-K-ATPase: heterogeneity in structure, diversity in function. Am J Physiol 275:F633-50

Blanco G, Sánchez G, Melton RJ, Tourtellotte WG, Mercer RW (2000) The alpha-4 isoform of the Na,K-ATPase is expressed in the germ cells of the testes. J Histochem Cytochem 48:1023-1032

Bystriansky JS, Richards JG, Schulte PM, Ballantyne JS (2006) Reciprocal expression of gill Na⁺/K⁺-ATPase alpha-subunit isoforms α 1a and α 1b during seawater

acclimation of three salmonid fishes that vary in their salinity tolerance. J Exp Biol 209:1848-1858

C. elegans Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 282:2012-2018

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540-552

Cereijido M, Contreras RG, Shoshani L (2004) Cell adhesion, polarity, and epithelia in the dawn of metazoans. Physiol Rev 84:1229-1262

Choe KP, Verlander JW, Wingo CS, Evans DH (2004) A putative H⁺-K⁺-ATPase in the Atlantic stingray, *Dasyatis sabina*: primary sequence and expression in gills. Am J Physiol Regul Integr Comp Physiol 287:R981-991

Clark AG, Eisen MB, Smith DR et al (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450:203-218

Colina C, Rosenthal JJ, DeGiorgis JA et al (2007) Structural basis of Na⁽⁺⁾/K⁽⁺⁾-ATPase adaptation to marine environments. Nat Struct Mol Biol 14:427-431

Croyle ML, Woo AL, Lingrel JB (1997) Extensive random mutagenesis analysis of the Na⁺/K⁺-ATPase alpha subunit identifies known and previously unidentified amino acid

residues that alter ouabain sensitivity--implications for ouabain binding. *Eur J Biochem* 248:488-95

Eichinger L, Pachebat JA, Glockner G et al (2005) The genome of the social amoeba *Dictyostelium discoideum*. *Nature* 435:43-57

Eisen JA, Coyne RS, Wu M et al (2006) Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol* 4:e286

Escalante R, García-Sáez A, Ortega MA, Sastre L (1994) Gene expression after resumption of development of *Artemia franciscana* cryptobiotic embryos. *Biochem Cell Biol* 72:78-83

Escalante R, García-Sáez A, Sastre L (1995) In situ hybridization analyses of Na,K-ATPase α -subunit expression during early larval development of *Artemia franciscana*. *J Histochem Cytochem* 43:391-399

Evans DH, Piermarini PM, Choe KP (2005) The multifunctional fish gill: dominant site of gas exchange, osmoregulation, acid-base regulation, and excretion of nitrogenous waste. *Physiol Rev* 85:97-177

Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. *Syst Zool* 27:401-410

- Fietto LG, Pugliese L, Gomes SL (2002) Characterization and expression of two genes encoding isoforms of a putative Na, K-ATPase in the chytridiomycete *Blastocladiella emersonii*. *Biochim Biophys Acta* 1576:59-69
- García-Sáez A, Perona R, Sastre L (1997) Polymorphism and structure of the gene coding for the $\alpha 1$ subunit of the *Artemia franciscana* Na/K-ATPase. *Biochem J* 321:509-518
- Garty, H, Karlish, SJ (2006) Role of FXYD proteins in ion transport. *Annu Rev Physiol* 68:431–459
- Grosell M, Gilmour KM, Perry SF (2007) Intestinal carbonic anhydrase, bicarbonate, and proton carriers play a role in the acclimation of rainbow trout to seawater. *Am J Physiol Regul Integr Comp Physiol* 293:R2099-111
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696-704
- Henry RP (2005) Critical salinity, sensitivity, and commitment of salinity-mediated carbonic anhydrase induction in the gills of two euryhaline species of decapod crustaceans. *J Exp Zool A* 303:45-56
- Hillis DM, Bull JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analyses. *Syst Biol* 42:182–192

Holliday CW, Roye DB, Roer RD (1990) Salinity-induced changes in branchial Na⁺/K⁺-ATPase activity and transepithelial potential difference in the brine shrimp *Artemia salina*. J Exp Biol 151:279-296

Holt RA, Subramanian GM, Halpern A et al (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. Science 298:129-149

International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695-716

Jorgensen PL, Amat F (2008) Regulation and function of lysine-substituted Na,K pumps in salt adaptation of *Artemia franciscana*. J Membr Biol 221:39-49

Jorgensen PL, Hakansson KO, Karlsh SJ (2003) Structure and mechanism of Na,K-ATPase: functional sites and their interactions. Annu Rev Physiol 65:817-849

Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: Improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 33:511–518

King N, Westbrook MJ, Young SL et al (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. Nature 451:783-788

Krogh A (1946) The active and passive exchanges of inorganic ions through the surfaces of living cells and through living membranes generally. Proc R Soc Lond B

133:140-200

Kühlbrandt W (2004) Biology, structure and mechanism of P-type ATPases. *Nat Rev Mol Cell Biol* 5:282-295

Kültz D, Fiol D, Valkova N et al (2007) Functional genomics and proteomics of the cellular osmotic stress response in “non-model” organisms. *J Exp Biol* 210:1593-1601

Lander ES, Linton LM, Birren B et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860-921

Macías MT, Palmero I, Sastre L (1991) Cloning of a cDNA encoding an *Artemia franciscana* Na/K ATPase α -subunit. *Gene* 105:197–204

Madsen SS, Jensen LN, Tipsmark CK et al (2007) Differential regulation of cystic fibrosis transmembrane conductance regulator and Na⁺,K⁺ -ATPase in gills of striped bass, *Morone saxatilis*: effect of salinity and hormones. *J Endocrinol* 192:249-260

Marshall WS (2002) Na(+), Cl(-), Ca(2+) and Zn(2+) transport by fish gills: retrospective review and prospective synthesis. *J Exp Zool* 293:264-283

Mense M, Rajendran V, Blostein R, Caplan MJ (2002) Extracellular domains, transmembrane segments, and intracellular domains interact to determine the cation selectivity of Na,K- and gastric H,K-ATPase. *Biochemistry* 41:9803-9812

- Mikkelsen TS, Wakefield MJ, Aken B et al (2007) Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* 447:167-77
- Okamura H, Yasuhara JC, Fambrough DM, Takeyasu K (2002) P-type ATPases in *Caenorhabditis* and *Drosophila*: implications for evolution of the P-type ATPase subunit families with special reference to the Na,K-ATPase and H,K-ATPase subgroup. *J Membr Biol* 191:13-24
- Palmgren MG, Axelsen KB (1998) Evolution of P-type ATPases. *Biochim Biophys Acta* 1365:37-45
- Piermarini PM, Evans DH (2000) Effects of environmental salinity on Na(+)/K(+)-ATPase in the gills and rectal gland of a euryhaline elasmobranch (*Dasyatis sabina*). *J Exp Biol* 203:2957-2966
- Putnam NH, Srivastava M, Hellsten U et al (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317:86-94
- Putnam NH, Butts T, Ferrier DE et al (2008) The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064-1071
- Richards JG, Semple JW, Bystriansky JS, Schulte PM (2003) Na⁺/K⁺-ATPase α -isoform switching in gills of rainbow trout (*Oncorhynchus mykiss*) during salinity transfer. *J Exp Biol* 206:4475-4486

Ruaud AF, Bessereau JL (2007) The P-type ATPase CATP-1 is a novel regulator of *C. elegans* developmental timing that acts independently of its predicted pump function. *Development* 134:867-879

Sáez AG, Escalante R, Sastre L (2000) High DNA sequence variability at the α 1 Na/K-ATPase locus of *Artemia franciscana* (brine shrimp): polymorphism in a gene for salt-resistance in a salt-resistant organism. *Mol Biol Evol* 17:235-250

Sea Urchin Genome Sequencing Consortium (2006) The Genome of the Sea Urchin *Strongylocentrotus purpuratus*. *Science* 314:941-952

Serluca FC, Sidow A, Mably JD, Fishman MC (2001) Partitioning of tissue expression accompanies multiple duplications of the Na⁺/K⁺ ATPase alpha subunit gene. *Genome Res* 11:1625-31

Serrano L, Halanych KM, Henry RP (2007) Salinity-stimulated changes in expression and activity of two carbonic anhydrase isoforms in the blue crab *Callinectes sapidus*. *J Exp Biol* 210:2320-2332

Shono M, Wada M, Hara Y, Fujii T (2001) Molecular cloning of Na⁺-ATPase cDNA from a marine alga, *Heterosigma akashiwo*. *Biochim Biophys Acta* 1511:193-199

Silver RB, Soleimani M (1999) H⁺-K⁺-ATPases: regulation and role in pathophysiological states. *Am J Physiol* 276:F799-811

Skou JC (1957) The influence of some cations on an adenosine triphosphatase from peripheral nerves. *Biochim Biophys Acta* 23:394-401

Swofford DL (2000) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts

Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 56:564-577

The Honeybee Genome Sequencing Consortium (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443:931-949

The UniProt Consortium (2008) The Universal Protein Resource (UniProt). *Nucleic Acids Res* 36:D190-D195

Therien AG, Blostein R (2000) Mechanisms of sodium pump regulation. *Am J Physiology Cell Physiol* 279:C541-C566

Towle DW (1997) Molecular approaches to understanding salinity adaptation of estuarine animals. *Am Zool* 37:575-584

Tribolium Genome Sequencing Consortium (2008) The genome of the model beetle and pest *Tribolium castaneum*. *Nature* 452:949-955

Tsai JR, Lin HC (2007) V-type H⁺-ATPase and Na⁺,K⁺-ATPase in the gills of 13 euryhaline crabs during salinity acclimation. *J Exp Biol* 210:620-627

Warren WC, Hillier LW, Marshall Graves JA et al (2008) Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453:175-183

Weider LJ, Hebert PDN (1987) Ecological and physiological differentiation among low-arctic clones of *Daphnia pulex*. *Ecology* 68:188-198

Figure legends

Fig. 1

Models for the osmoregulatory role of Na,K-ATPases in hypersaline (e.g. seawater) and hyposaline environments (e.g. freshwater). See text for a description.

Fig. 2

Maximum likelihood best tree for a representation of all known P-type IIC ATPases (222 sequences). Various colours (online version only) are used to distinguish branches referring to different groups of organisms. Three ATPases (types IIA, IIB, IID) are used as outgroup sequences to indicate the root of the tree. Only the black branches of the ingroup were considered for presenting bootstrap values (1000 replicates), shown when equal or above 50%. Names of genes or proteins, if available, follow the name of organism(s); otherwise “undeterm.” (i.e. undetermined) is written. The gray part (see the online version of this article) of the tree, where only metazoan α subunits from H,K-

and Na,K-ATPases are found, is shown expanded in figure 3. The asterisk by one of the branches indicates the *Drosophila* clade with protein CG3701, whereas bootstrap values followed by an asterisk correspond to those values in the bootstrap consensus tree, whilst ignoring the presence of *Drosophila*'s CG3701 clade (see text). A scale bar shows the correspondence between branch length and rate of amino acidic substitution per site.

Fig. 3

Sub-tree of figure 2, there in gray (see online version), shown here in larger detail, and with the same presentation features as in figure 2.

Fig. 4

Maximum likelihood tree for representative α subunits of H,K- and Na,K-ATPases from vertebrates. Outgroup sequences correspond to the echinoderm *S. purpuratus* and the cephalochordate *B. floridae*. Bootstrap values (1000 replicates) equal or above 50% are shown at each node. Vertical lines group α subunit isoforms of Na,K-ATPases ($\alpha 1$ to $\alpha 4$), and of H,K-ATPases (α -HKAs); discontinuous lines suggest unsure associations.

Figure 1

Figure 2

Figure 3

Fig. 4

Accession	Entry Name	Status	Protein Names	Genes	Organism	Length
AA86C3	AA86NC3_AKAC5C	unreviewed	Sodium/potassium-ATPase alpha subunit		Acanthopagrus schlegelii (Black porgy)	1025
Q16H74	Q16H74_ADAE	unreviewed	Na+/K+ ATPase alpha subunit	NaK-ATPase	Aedes aegypti (Yellowfever mosquito)	1001
Q16H75	Q16H75_ADAE	unreviewed	Na+/K+ ATPase alpha subunit	NaK-ATPase	Aedes aegypti (Yellowfever mosquito)	1001
Q16H76	Q16H76_ADAE	unreviewed	Na+/K+ ATPase alpha subunit	NaK-ATPase	Aedes aegypti (Yellowfever mosquito)	1001
Q3HSP9	Q3HSP9_ANAN	unreviewed	ATPase, E1-E2 type (EC 3.6.3.9)	Axa_4083	Anabenna variabilis (strain ATCC 29413 / PCC 7937)	971
Q92930	AT1A1_ANO4	reviewed	Sodium/potassium-transporting ATPase subunit alpha-1 precursor (EC 3.6.3.9) (Sodium pump subunit alpha-1) (Na(+)/K(+) ATPase alpha-1 subunit)	atp1a1	Anguilla anguilla (European freshwater eel)	1022
Q92931	Q92931_ANO4	reviewed	ATPase subunit alpha-1 precursor (EC 3.6.3.9) (Sodium pump subunit alpha-1) (Na(+)/K(+) ATPase alpha-1 subunit)	atp1a1	Anguilla gangetica sp. PE87	1060
P71226	AT1A_ARTISF	reviewed	Sodium/potassium-transporting ATPase subunit alpha-A (EC 3.6.3.9) (Sodium pump subunit alpha-A) (Na(+)/K(+) ATPase alpha subunit A)	ATP1A	Artemia franciscana (Brine shrimp) (Artemia franciscana)	996
T28774	AT1B_T28774	reviewed	Sodium/potassium-transporting ATPase subunit alpha-B (EC 3.6.3.9) (Sodium pump subunit alpha-B) (Na(+)/K(+) ATPase alpha subunit B)	ATP1B	Artemia franciscana (Brine shrimp) (Artemia franciscana)	1000
A1C554	A1C554_ASPCL	unreviewed	H/K ATPase alpha subunit, putative	ACLA_004560	Aspergillus clavatus	1096
A1CDE2	A1CDE2_ASPG	unreviewed	Na/K ATPase alpha 1 subunit, putative	ACLA_005260	Aspergillus clavatus	1101
Q4W039	Q4W039_ASPG	unreviewed	H/K ATPase alpha subunit, putative (EC 3.6.3.10)	AFUA_500730	Aspergillus fumigatus (Sartoryia fumigata)	1099
Q4W0C6	Q4W0C6_ASPU	unreviewed	Na/K ATPase alpha 1 subunit, putative (EC 3.6.3.9)	AFUA_7004570	Aspergillus fumigatus (Sartoryia fumigata)	1105
AQ2761	AQ2761_ASPG	unreviewed	Putative frameshift	AN950930	Aspergillus niger	1108
A2R2X5	A2R2X5_ASPG	unreviewed	Induction: the mRNA for the alpha 1 isoform of the NKA1 gene from R. norvegicus (EC 3.6.3.9)	AI4010230	Aspergillus niger	1099
Q2U302	Q2U302_ASPOR	unreviewed	Na+/K+ ATPase	AO909030008	Aspergillus oryzae	1049
Q2U764	Q2U764_ASPOR	unreviewed	Na+/K+ ATPase	AO90903000141	Aspergillus oryzae	1032
Q4CPH1	Q4CPH1_ASPH1	unreviewed	Putative uncharacterized protein	AP02_04343	Aspergillus terreus (strain NH 2624)	1087
Q43134	Q43134_BLAEM	unreviewed	P-type cation-transporting ATPase	PA12	Blattodactylla emersonii (Aquatic fungus)	1080
QRV111	QRV111_BLAEM	unreviewed	P-type cation-transporting ATPase	PA12	Blattodactylla emersonii (Aquatic fungus)	1078
Q68424	AT1A1_Q68424	reviewed	Sodium/potassium-transporting ATPase subunit alpha-1 precursor (EC 3.6.3.9) (Sodium pump subunit alpha-1) (Na(+)/K(+) ATPase alpha-1 subunit)	ATP1A1	Bos taurus (Bovine)	1021
A2V016	AT1A2_BOVIN	reviewed	Sodium/potassium-transporting ATPase subunit alpha-2 precursor (EC 3.6.3.9) (Sodium pump subunit alpha-2) (Na(+)/K(+) ATPase alpha-2 subunit)	ATP1A2	Bos taurus (Bovine)	1020
AR6V11	AR6V11_BO7FB	unreviewed	Putative uncharacterized protein	BC1G_04401	Botryotinia fuckeliana (strain B05.10) (Noble rot fungus) (Botrytis cinerea)	1102
A6S135	A6S135_BO7FB	unreviewed	Putative uncharacterized protein	BC1G_06532	Botryotinia fuckeliana (strain B05.10) (Noble rot fungus) (Botrytis cinerea)	1131
AR6E11	AR6E11_BO7FB	unreviewed	Na-K-ATPase alpha subunit, putative	Bml_1_3140	Bruguiera malay (Filarial nematode worm)	1047
Q92936	AT12A_BUFMA	reviewed	Potassium-transporting ATPase alpha chain 2 (EC 3.6.3.10) (Proton pump) (Non-gastric H(+)/K(+) ATPase subunit alpha)	ATP12A	Bufo marinus (Giant toad) (Cane toad)	1042
P73074	AT1A1_BUFMA	reviewed	Sodium/potassium-transporting ATPase subunit alpha-1 precursor (EC 3.6.3.9) (Sodium pump subunit alpha-1) (Na(+)/K(+) ATPase alpha-1 subunit)	ATP1A1	Bufo marinus (Giant toad) (Cane toad)	1023
AR6X13	AR6X13_CABER	unreviewed	Putative uncharacterized protein	CBN11219	Cenorchabitis briggsae	1054
AR6X41	AR6X41_CABER	unreviewed	Putative uncharacterized protein	CBN11219	Cenorchabitis briggsae	996
AR6X63	AR6X63_CABER	unreviewed	Putative uncharacterized protein (Fragment)	CBN11219	Cenorchabitis briggsae	1056
AR6X78	AR6X78_CABER	unreviewed	Putative uncharacterized protein	CBN11219	Cenorchabitis briggsae	1054
O16331	O16331_CAEL	unreviewed	Putative uncharacterized protein	COH95.2	Cenorchabitis elegans	1054
O16436	O16436_CAEL	unreviewed	Putative uncharacterized protein	CO2E7.1	Cenorchabitis elegans	1050
Q4S240	Q4S240_CAEL	unreviewed	Putative uncharacterized protein	CO2E7.1	Cenorchabitis elegans	1048
P90735	P90735_CAEL	unreviewed	Putative uncharacterized protein est-6	est-6 (B0035.3)	Cenorchabitis elegans	996
Q27461	Q27461_CAEL	unreviewed	Na-K-ATPase alpha subunit	NaK	Cenorchabitis elegans	996
Q27462	Q27462_CAEL	unreviewed	Putative uncharacterized protein	NaK	Cenorchabitis elegans	996
Q9SG02	Q9SG02_CALS	unreviewed	Na+/K+ ATPase alpha subunit	NaK	Callinectes sapidus (Blue crab)	1039
P50997	AT1A1_CANFA	reviewed	Sodium/potassium-transporting ATPase subunit alpha-1 precursor (EC 3.6.3.9) (Sodium pump subunit alpha-1) (Na(+)/K(+) ATPase alpha-1 subunit)	ATP1A1	Canis familiaris (Dog)	1021
P50998	ATPAA_CANFA	reviewed	Potassium-transporting ATPase alpha chain 1 (EC 3.6.3.10) (Proton pump) (Gastric H(+)/K(+) ATPase subunit alpha)	ATPAA	Canis familiaris (Dog)	1047
Q90W67	Q90W67_CARAU	unreviewed	Na-K-ATPase alpha subunit 3	ATP1A3	Carassius auratus (Goldfish)	1022
Q50544	Q50544_CARRA	unreviewed	Na+/K+ ATPase alpha subunit (Fragment)	ATP1A3	Carassius auratus (Goldfish)	831
P25489	AT1A1_CATCO	reviewed	Sodium/potassium-transporting ATPase subunit alpha-1 precursor (EC 3.6.3.9) (Sodium pump subunit alpha-1) (Na(+)/K(+) ATPase alpha-1 subunit)	atp1a1	Carotus marinus (Common sole) (Sole)	1022
Q68499	AT1A1_CATCO	reviewed	Sodium/potassium-transporting ATPase subunit alpha-1 precursor (EC 3.6.3.9) (Sodium pump subunit alpha-1) (Na(+)/K(+) ATPase alpha-1 subunit)	atp1a1	Carotus marinus (Common sole) (Sole)	1022
AX9SL2	AX9SL2_CHACN	unreviewed	Sodium/potassium-transporting ATPase alpha-1 subunit	CHLEDRRAFT_187139	Chanos chanos (Milkfish)	1024
ABXN15	ABXN15_CHLRE	unreviewed	Sodium/potassium-transporting ATPase alpha subunit	CHLEDRRAFT_187139	Chlamydomonas reinhardtii	1112
Q27742	Q27742_DANRE	unreviewed	Sodium/potassium-transporting ATPase alpha subunit	atp1a3	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q4KMK4	Q4KMK4_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3b polypeptide	atp1a3b	Danio rerio (Zebrafish) (Brachydanio rerio)	1023
Q92934	Q92934_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1023
Q92771	Q92771_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a (atp1a3a-2) (CH21-1)	Danio rerio (Zebrafish) (Brachydanio rerio)	1023
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q92726_DANRE	unreviewed	ATPase, Na+/K+ transporting, alpha 3a polypeptide	atp1a3a	Danio rerio (Zebrafish) (Brachydanio rerio)	1024
Q92726	Q9					

[illegible]

Q6P941	Q6P941_XENLA	unreviewed	Atp1a1a-1 protein	MGCE8460	Xenopus laevis (African clawed frog)	1024
Q6PAG0	Q6PAG0_XENLA	unreviewed	MGCE8460 protein	MGCE8460	Xenopus laevis (African clawed frog)	1020
Q7ZXD3	Q7ZXD3_XENLA	unreviewed	MGCE8460 protein	MGCE8460	Xenopus laevis (African clawed frog)	1024
Q7ZXD3	Q7ZXD3_XENLA	unreviewed	MGCE8460 protein	MGCE8460	Xenopus laevis (African clawed frog)	1024
Q7ZYK8	Q7ZYK8_XENLA	unreviewed	Atp1a3-prov protein		Xenopus laevis (African clawed frog)	1025
Q801M8	Q801M8_XENLA	unreviewed	Atp12a-prov protein		Xenopus laevis (African clawed frog)	1042
A4IIS7	A4IIS7_XENTR	unreviewed	LOC100125050 protein	LOC100125050	Xenopus tropicalis (Western clawed frog) (Silurana tropicalis)	1020
Q5BL65	Q5BL65_XENTR	unreviewed	ATPase, H+/K+ transporting, nongastric, alpha polypeptide, locus 1	atp4a	Xenopus tropicalis (Western clawed frog) (Silurana tropicalis)	1033
Q6P157	Q6P157_XENTR	unreviewed	ATPase, Na+/K+ transporting, alpha 1 polypeptide	atp1a1 (TTPA007m23.1-001)	Xenopus tropicalis (Western clawed frog) (Silurana tropicalis)	1023

In green: removed series from database, either same gene, or one protein of any size with identity equal or larger than 98% within the same species, plus several sequences with unambiguous homologous sequences already present in the data base.

All the above samples are from uniprot.org (UniRef_100) on 14 March 2008. Cut off at Score (max) > 440. Release 13.0 - Feb 26, 2008. No proteins smaller than 800 amino acids were selected.

Q232A9 was removed beforehand because it had a long string of Xs, which could not aligned with anything well.

THE FOLLOWING UNIPROT PROTEINS WERE ADDED TO THE DATASET:

Q9W248		CG3701-PA	CG3701	Drosophila melanogaster	1030
Q8WQ81		catp-1 (same as A1KR00) (Y105EBA.12)	catp-1	Caenorhabditis elegans	1121
Q28ZL5	Q28ZL5_DROPS	GA17624-PA	GA17624	Drosophila pseudoobscura	974
A9V754		ID: 10758	ORF Names: 10758	Monosiga brevicollis	
A9V756		ID: 28085	ORF Names: 28085	Monosiga brevicollis	

Sequences obtained from other databases:

Branchiostoma floridae (http://www.jgi.doe.gov/):		Location:
Protein ID:57104	>jgi Brafl1 57104 lgemesh2_pm.scaffold_8000049	Brafl1/scaffold_8:4073228-4096652
Capitella capitata (http://www.jgi.doe.gov/):		
Protein ID:168025	>jgi Capca1 168025 estExt_Genewise1Plus.C_1710030	Capca1/scaffold_171:102025-109032
Protein ID:182016	>jgi Capca1 182016 estExt_Genewise1Plus.C_120047	Capca1/scaffold_12:226996-232954
Daphnia pulex (http://wlfesbase.org):		
NCBI_GNO_552044		
NCBI_GNO_864044		
NCBI_GNO_376613		
NCBI_GNO_342614		
NCBI_GNO_338614		
NCBI_GNO_290684		

Outgroup:

P35316-1	ATC_ARTSF	Calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type (isoform 1) (SERCA-ATPase; Palmero and Sastre 1989) (P-type IIA)	Artemia franciscana	1003
P13587	ATN1_YEAST	Sodium transport ATPase 1 (ENA1) (P-type IIB)	Saccharomyces cerevisiae	
O45215	O45215_CAEL	Calcium ATPase (Protein W09C2.3b, partially confirmed by transcript evidence) (mca-1) (P-type IIB)	Caenorhabditis elegans	