## Supplementary Information

### Cation/$Ca^{2+}$ Exchangers

Cation/$Ca^{2+}$ exchangers, defined by the presence of two highly conserved α-repeat regions in the two transmembrane domains, comprise the NC(K)X group including the $K^+$-independent $Na^+/Ca^{2+}$ exchangers (NCXs) and the $K^+$-dependent $Na^+/Ca^{2+}$ exchangers (NCKXs), and the $Ca^{2+}/H^+$ exchanger (CAX) group (Philipson and Nicoll 2000). NC(K)Xs and CAXs were previously shown to exhibit mutually exclusive genomic distributions - NC(K)Xs in animals that mainly use $Na^+$ (and $K^+$) gradients as the driving force(s) for $Ca^{2+}$ extrusion, and CAXs in bacteria, plants, and fungi that utilize the $H^+$ gradients across the membranes (Cai and Lytton 2004; Shigaki et al. 2006). Both NCKX and CAX homologs are present in the apusozoan *T. trahens* (Cai and Clapham 2012).

*A. limacinum* possesses three classes of cation/$Ca^{2+}$ exchangers – NCKXs, NCXs and CAXs, with significant sequence homology and highly conserved signature motifs/residues (**Figs. 2 and S2**). For instance, conservation of key acidic residues in the signature motifs critical for the exchange activity and conservation of a single aspartate residue in the second α-repeat in NCKXs (Asp575 in human NCKX2) (**Fig. S2**) that is important for the $K^+$-dependent exchange activity of NCKXs (Kang et al. 2005). Therefore, the presence of all three groups of exchangers in *A. limacinum* provides evidence for the common origin of NC(K)X and CAX exchangers in ancestral protists.

### Intracellular Ion Channels

A copy of $IP_3$ receptors ($IP_3$Rs) homolog is present in *A. limacinum*. *T. trahens* contains two copies of IP3Rs (Cai and Clapham 2012). $IP_3$R is absent in plants, fungi, and many other photosynthetic eukaryotes (Verret et al. 2010). The first functionally characterized protozoan $IP_3$R was cloned from the ciliate *Paramecium tetraurelia* (Ladenburger et al. 2006). $IP_3$R homologs with varied degrees of sequence divergence have also been identified in many other protozoan species. Some protozoan $IP_3$R homologs appear to exhibit ryanodine receptor (RyR)-like features, but they are generally believed to lack critical

domains/residues conserved in animal RyRs (Wheeler and Brownlee 2008; Prole and Taylor 2011; Plattner and Verkhratsky 2013). The presence of these primitive $IP_3R$ and RyR-like homologs suggests the early evolution of intracellular $Ca^{2+}$ release channels in eukaryotes (Plattner and Verkhratsky 2013). In contrast, no RyR homologs are identified in *A. limacinum* and *T. trahens*.

Two pore channels (named for having two pore *domains*, not for having two pores; TPCs) are conserved in land plants and animals, but are not found in fungi and many algae. TPCs are present in *A. limacinum* and *T. trahens*. Plant TPCs apparently mediate slowly activating vacuolar $Ca^{2+}$ currents (Peiter et al. 2005). However, recent direct electrophysiological recording in endolysosomes demonstrates that animal TPCs are $Na^+$-selective (Wang et al. 2012; Cang et al. 2013).

**Other types of $Ca^{2+}$ signaling molecules**

The mammalian transient receptor potential (TRP) channel family comprises six subfamilies of 28 channel proteins with modest sequence homology (Wu, Sweet, and Clapham 2010). They regulate diverse cellular functions in response to various stimuli such as temperature, pH and second messengers. Most TRP channels are $Ca^{2+}$ permeable, but the degree of $Ca^{2+}$ selectivity is highly variable. Four members of the TRP ion channel family including TRPC, TRPV, TRPP, and TRPML are found in *A. limacinum* (**Fig. S1**). TRP channel homologs are also present in *T. trahens* (Cai and Clapham 2012). Even though animal-type TRP channels are not present in land plants and fungi, they are widely distributed in many algae. The physiological functions of TRP channel homologs in lower eukaryotes are largely unknown.

We also identified homologs of $Ca_V$ channels in *A. limacinum*. Animal $Ca_V$ channel homologs are found in *T. trahens* and a wide range of algae, but have been lost in land plants and fungi (Verret et al. 2010; Cai and Clapham 2012). Furthermore, cyclic nucleotide-gated ion channels (Cai 2012a), three classes of $Ca^{2+}$ ATPases, and mitochondrial $Ca^{2+}$ regulators including LETM1, MCU and MICU are also found in *A. limacinum* and *T. trahens* (**Fig. S1**).

Piezo ion channels are $Ca^{2+}$-permeable mechanically-activated channels that are found in animals, plants and some protists (Coste et al. 2010). Piezo channels seem to be absent in fungi, bacteria and many protists. Piezo homologs are identified in *T. trahens* and *M. brevicollis*, but not in *A. limacinum*.

The ligand-gated ion channel superfamily in animals consists of P2X receptors, ionotropic glutamate receptors (iGluRs), and cysteine-loop (Cys-loop) receptors. P2X receptors exist in all vertebrates, many invertebrates, basal fungi, and certain protists. Modern fungi, land plants and some invertebrate species do not contain P2X receptor homologs (Cai 2012b). P2X receptor homologs are identified in both *A. limacinum* and *T. trahens*. Both animals and land plants possess iGluRs that are $Ca^{2+}$-permeable. iGluR homologs can also be found in certain protist and bacterial species in which iGluRs appear to be $K^{+}$-selective (Chen et al. 1999). We found in *T. trahens* three putative iGluRs homologs that are closer to their bacterial counterparts, while no iGluR homolog is present in *A. limacinum*. Cys-loop receptors are pentameric ligand-gated ion channels that include $Ca^{2+}$-permeable nicotinic acetylcholine and 5-HT3 receptors, and anion-selective GABA and glycine receptors. Cys-loop receptors are widespread in animals and are also present in several bacterial and protist species (Verret et al. 2010). Cys-loop receptor homologs are found in *A. limacinum* but are absent in *T. trahens*.

Calmodulin (CaM) is a small, soluble $Ca^{2+}$-binding protein that regulates a wide variety of cellular processes (Toutenhoofd and Strehler 2000). CaM is ubiquitous throughout eukaryotic evolution and its ancestral form, with four EF-hand $Ca^{2+}$ binding motifs, is present in bacteria (Michiels et al. 2002). CaM is present in both *A. limacinum* and *T. trahens*.

## References


Burki F, Pawlowski J. 2006. Monophyly of Rhizaria and multigene phylogeny of unicellular bikonts. **Mol Biol Evol**. 23:1922-1930.

Cai X. 2012a. Evolutionary genomics reveals the premetazoan origin of opposite gating polarity in animal-type voltage-gated ion channels. **Genomics**. 99:241-245.

Cai X. 2012b. P2X receptor homologs in basal fungi. **Purinergic Signal**. 8:11-13.

Cai X, Clapham DE. 2012. Ancestral $Ca^{2+}$ signaling machinery in early animal and fungal evolution. **Mol Biol Evol**. 29:91-100.

Cai X, Lytton J. 2004. The cation/$Ca^{2+}$ exchanger superfamily: phylogenetic analysis and structural implications. **Mol Biol Evol**. 21:1692-1703.

Cang C, Zhou Y, Navarro B, Seo YJ, Aranda K, Shi L, Battaglia-Hsu S, Nissim I, Clapham DE, Ren D. 2013. mTOR regulates lysosomal ATP-sensitive two-pore Na(+) channels to adapt to metabolic state. **Cell**. 152:778-790.

Chen GQ, Cui C, Mayer ML, Gouaux E. 1999. Functional characterization of a potassium-selective prokaryotic glutamate receptor. **Nature**. 402:817-821.

Coste B, Mathur J, Schmidt M, Earley TJ, Ranade S, Petrus MJ, Dubin AE, Patapoutian A. 2010. Piezo1 and Piezo2 are essential components of distinct mechanically activated cation channels. **Science**. 330:55-60.

Derelle R, Lang BF. 2012. Rooting the eukaryotic tree with mitochondrial and bacterial proteins. **Mol Biol Evol**. 29:1277-1289.

Kang KJ, Shibukawa Y, Szerencsei RT, Schnetkamp PP. 2005. Substitution of a single residue, Asp575, renders the NCKX2 $K^+$-dependent $Na^+/Ca^{2+}$ exchanger independent of $K^+$. **J Biol Chem**. 280:6834-6839.

Ladenburger EM, Korn I, Kasielke N, Wassmer T, Plattner H. 2006. An Ins(1,4,5)P3 receptor in Paramecium is associated with the osmoregulatory system. **J Cell Sci**. 119:3705-3717.

Michiels J, Xi C, Verhaert J, Vanderleyden J. 2002. The functions of Ca(2+) in bacteria: a role for EF-hand proteins? **Trends Microbiol**. 10:87-93.

Peiter E, Maathuis FJ, Mills LN, Knight H, Pelloux J, Hetherington AM, Sanders D. 2005. The vacuolar $Ca^{2+}$-activated channel TPC1 regulates germination and stomatal movement. **Nature**. 434:404-408.

Philipson KD, Nicoll DA. 2000. Sodium-calcium exchange: a molecular perspective. **Annu Rev Physiol**. 62:111-133.

Plattner H, Verkhratsky A. 2013. Ca2+ signalling early in evolution--all but primitive. **J Cell Sci**. 126:2141-2150.

Prole DL, Taylor CW. 2011. Identification of intracellular and plasma membrane calcium channel homologues in pathogenic parasites. **PLoS One**. 6:e26218.

Shigaki T, Rees I, Nakhleh L, Hirschi KD. 2006. Identification of three distinct phylogenetic groups of CAX cation/proton antiporters. **J Mol Evol**. 63:815-825.

Toutenhoofd SL, Strehler EE. 2000. The calmodulin multigene family as a unique case of genetic redundancy: multiple levels of regulation to provide spatial and temporal control of calmodulin pools? **Cell Calcium**. 28:83-96.

Verret F, Wheeler G, Taylor AR, Farnham G, Brownlee C. 2010. Calcium channels in photosynthetic eukaryotes: implications for evolution of calcium-based signalling. **New Phytol**. 187:23-43.

Wang X, Zhang X, Dong XP et al. 2012. TPC proteins are phosphoinositide- activated sodium-selective ion channels in endosomes and lysosomes. **Cell**. 151:372-383.

Wheeler GL, Brownlee C. 2008. $Ca^{2+}$ signalling in plants and green algae--changing channels. **Trends Plant Sci**. 13:506-514.

Wu LJ, Sweet TB, Clapham DE. 2010. International Union of Basic and Clinical Pharmacology. LXXVI. Current progress in the mammalian TRP ion channel family. **Pharmacol Rev**. 62:381-404.

## Legends for Supplementary Figures

**Figure S1. $Ca^{2+}$ signaling machineries in select species of protists, plants, fungi and animals across Unikonta and Bikonta.** The evolutionary relationship is inferred from the Tree of Life project (http://www.tolweb.org/) and recent references on the eukaryotic tree (Burki and Pawlowski 2006; Derelle and Lang 2012). The number of protein homologs derived from currently available genomic databases is shown. In *C. paradoxa* whose complete genome sequence is not available and in few other instances, a *black dot* indicates the presence of protein homolog (s). Absence of a number and a black dot indicates that no homolog was identified. *Asterisk* (*) indicates that diatom TRP-like sequences do not group with known mammalian TRP subfamilies. Similar comparative genomics analyses of $Ca^{2+}$ signaling molecules have been shown (Wheeler and Brownlee 2008; Verret et al. 2010). Abbreviations: *CAX*, $Ca^{2+}/H^{+}$ exchanger; *CatSper*, sperm-associated cation channel; *CaV*, voltage-gated $Ca^{2+}$ channel; ; *Cch1*, $Ca^{2+}$ channel protein 1; *CNG*, cyclic nucleotide-gated channel; Cys-loopR, cysteine-loop ligand-gated receptor; iGluR, ionotropic glutamate receptor; *$IP_3R$*, inositol 1,4,5-trisphosphate receptor; *Letm1*, leucine zipper-EF-hand containing transmembrane protein 1; *MCU*, mitochondrial $Ca^{2+}$ uniporter; *MICU*, mitochondrial EF hand $Ca^{2+}$ uniporter regulator; *Mid1*, putative stretch-activated cation channel; *NC(K)X*, $Na^{+}/Ca^{2+}$ ($K^{+}$-dependent) exchanger; *Orai*, Orai $Ca^{2+}$ release-activated $Ca^{2+}$ channel; *P2XR*, P2X purinergic receptor; Piezo, Piezo mechanically activated cation channel; *PMC*, vacuole $Ca^{2+}$ ATPase; *PMCA*, plasma membrane $Ca^{2+}$ ATPase; *PMR*, yeast P-type $Ca^{2+}$ ATPase; *RyR*, ryanodine receptor; *SERCA*, sarco/endoplasmic reticulum $Ca^{2+}$ ATPase; *SPCA*, secretory pathway $Ca^{2+}$ ATPase; *TPC*, two-pore channel; *TRP*, transient receptor potential channel; *TRPY1*, yeast vacuolar channel protein.

**Figure S2. Identification of NC(K)X and CAX exchangers in *A. limacinum*.** Sequence alignment of the second α regions of select NCX, NCKX, and CAX exchangers from *A. limacinum* (*Ali*), *Homo sapiens* (*Hsa*), and the land plant *A. thaliana* (*Ath*). The location of signature motifs in the α2 repeats for each class of exchanger (Cai and Lytton 2004) is underlined. The location of the critical acidic residue

essential for exchanger activity is indicated by the asterisk symbol – an aspartate for NC(K)X and a glutamate for CAX. The filled circle indicates the location of the conserved aspartate residue critical for $K^+$-dependence of NCKX exchangers. *Abbreviations*: *CAX*, $Ca^{2+}/H^+$ exchangers; *NCKX*, $K^+$-dependent $Na^+/Ca^{2+}$ exchangers; *NCX*, $K^+$-independent $Na^+/Ca^{2+}$ exchangers.

## Supplementary Information

### Extended Material and Methods

### Database Searches

BlastP and TBlastN searches (Altschul et al. 1997) were performed using protein sequences of $Ca^{2+}$ signaling molecules from *Homo sapiens*, *T. trahens* and *Saccharomyces cerevisiae* as queries against available genomic sequences of bikont organisms at the National Center for Biotechnology Information Genome Database, the genome portal of the Department of Energy Joint Genome Institute, and the *C. paradoxa* genome website (Price et al. 2012). $Ca^{2+}$ signaling molecules such as $Ca^{2+}$ ATPases and mitochondrial $Ca^{2+}$ regulators that are widely distributed in lower eukaryotes were not used in initial screenings. The initial query proteins included CatSper subunits, $Ca^{2+}/H^{+}$ (CAX) exchangers, and $Na^{+}/Ca^{2+}$ exchangers , all of which are known to show distinct evolution patterns between animals and fungi, and are all present in *T. trahens* (Cai and Clapham 2012). Repeated BLAST searches using hit sequences from the first round of searches were also performed to identify potential distantly related homologs that might not be detected by using query proteins. In addition, PHMMER searches (HMMER 3.0, http://hmmer.janelia.org/) were also performed against select protein datasets.

Gene prediction programs GeneMark.hmm (Lomsadze et al. 2005) and FGENESH (http://www.softberry.com/berry.phtml) were employed when putative genes of interest identified by TBlastN searches had not been annotated or currently annotated genes in the database were believed to be incomplete.

Conserved protein domain/motif searches were performed using the SMART (Letunic, Doerks, and Bork 2009) and Pfam (Finn et al. 2010) web servers. The locations of conserved protein motifs/amino acid residues critical for ion channel function were obtained by literature searches.

**Multiple Sequence Alignments and Phylogeny Analyses**

Non-redundant protein sequences were aligned using MAFFT (Katoh and Toh 2008) and the alignments displayed with the Blosum62 Similarity Scoring Table were manually edited for improvement in GeneDoc (Nicholas, Nicholas, and Deerfield 1997). Unreliably aligned sites in the multiple sequence alignments were eliminated by using the GUIDANCE web server (Penn et al. 2010) and aligned columns containing more than 50% gaps were then removed by Gap Strip/Squeeze (v 2.1.0). Unambiguous sequence alignments were subsequently converted to files in PHYLIP format. Next, the best-fit evolution model and parameter estimates for the phylogeny reconstruction were selected by ProtTest 3.4 (Darriba et al. 2011). Maximum likelihood phylogenies with 100 bootstrap replicates were performed at the GARLI web server (http://molecularevolution.org/) (Bazinet, Zwickl, and Cummings 2014) by using GARLI 2.0 (Genetic Algorithm for Rapid Likelihood Inference) (Zwickl 2006). Consensus trees were computed by using the CONSENSE program (PHYLIP package, V 3.69) (Felsenstein 1996).

## References


Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Res**. 25:3389-3402.

Bazinet AL, Zwickl DJ, Cummings MP. 2014. A Gateway for Phylogenetic Analysis Powered by Grid Computing Featuring GARLI 2.0. **Syst Biol**.

Cai X, Clapham DE. 2012. Ancestral $Ca^{2+}$ signaling machinery in early animal and fungal evolution. **Mol Biol Evol**. 29:91-100.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. **Bioinformatics**. 27:1164-1165.

Felsenstein J. 1996. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. **Methods Enzymol**. 266:418-427.

Finn RD, Mistry J, Tate J et al. 2010. The Pfam protein families database. **Nucleic Acids Res**. 38:D211-222.

Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. **Brief Bioinform**. 9:286-298.

Letunic I, Doerks T, Bork P. 2009. SMART 6: recent updates and new developments. **Nucleic Acids Res**. 37:D229-232.

Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. **Nucleic Acids Res**. 33:6494-6506.

Nicholas K, Nicholas H, Deerfield D. 1997. Genedoc: analysis and visualization of genetic variation. **EMBNET News**. 4:1-4.

Penn O, Privman E, Ashkenazy H, Landan G, Graur D, Pupko T. 2010. GUIDANCE: a web server for assessing alignment confidence scores. **Nucleic Acids Res**. 38:W23-28.

Price DC, Chan CX, Yoon HS et al. 2012. Cyanophora paradoxa genome elucidates origin of photosynthesis in algae and plants. **Science**. 335:843-847.

Zwickl DJ. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. The University of Texas at Austin.