

## Research Article

# The Evolutionary History of Daphniid $\alpha$ -Carbonic Anhydrase within Animalia

Billy W. Culver<sup>1,2</sup> and Philip K. Morton<sup>2,3</sup>

<sup>1</sup>Program in Ecology & Evolutionary Biology, Department of Biology, University of Oklahoma, 730 Van Vleet Oval, Norman, OK 73019, USA

<sup>2</sup>University of Oklahoma Biological Station, 15389 Station Road, Kingston, OK 73439, USA

<sup>3</sup>Murray State College, One Murray Campus, Tishomingo, OK 73460, USA

Correspondence should be addressed to Billy W. Culver; [bculv001@ou.edu](mailto:bculv001@ou.edu)

Received 5 October 2014; Revised 12 March 2015; Accepted 14 March 2015

Academic Editor: Hirohisa Kishino

Copyright © 2015 B. W. Culver and P. K. Morton. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Understanding the mechanisms that drive acid-base regulation in organisms is important, especially for organisms in aquatic habitats that experience rapidly fluctuating pH conditions. Previous studies have shown that carbonic anhydrases (CAs), a family of zinc metalloenzymes, are responsible for acid-base regulation in many organisms. Through the use of phylogenetic tools, this present study attempts to elucidate the evolutionary history of the  $\alpha$ -CA superfamily, with particular interest in the emerging model aquatic organism *Daphnia pulex*. We provide one of the most extensive phylogenies of the evolution of  $\alpha$ -CAs, with the inclusion of 261 amino acid sequences across taxa ranging from Cnidarians to *Homo sapiens*. While the phylogeny supports most of our previous understanding on the relationship of how  $\alpha$ -CAs have evolved, we find that, contrary to expectations, amino acid conservation with bacterial  $\alpha$ -CAs supports the supposition that extracellular  $\alpha$ -CAs are the ancestral state of animal  $\alpha$ -CAs. Furthermore, we show that two cytosolic and one GPI-anchored  $\alpha$ -CA in *Daphnia* genus have homologs in sister taxa that are possible candidate genes to study for acid-base regulation. In addition, we provide further support for previous findings of a high rate of gene duplication within *Daphnia* genus, as compared with other organisms.

## 1. Introduction

Organisms experience a variety of environmental stressors to which they must respond in order to survive and reproduce. Some are able to adjust to these stressors and live to produce offspring and propagate their genes, while others do not and are extirpated. There has been a plethora of work attempting to elucidate the changes in physiological and genetic mechanisms in response to human-induced stresses/impacts on aquatic habitats, including nutrient enrichment and cultural eutrophication [1–3], anthropogenically elevated carbon dioxide [4], and toxic metal contamination [5, 6]. Another important human-mediated impact to aquatic habitats, lake acidification/alkalization, has also been well studied [7–11].

Acidification and alkalization of water bodies are important ecological stressors that affect the structure of plankton communities. Although the processes of acidification

and alkalization can occur naturally through mechanisms such as bedrock leaching [12], catchment runoff [10], and chemical conversion [13, 14], increasing impacts from anthropogenic sources such as carbon dioxide emissions [11], cultural eutrophication [7], and mining activities [6, 9] are of great concern with regard to lake acidification and alkalization.

Maintaining pH homeostasis in these altered habitats is critical for organisms to survive and reproduce. Acid-base regulation in a number of aquatic organisms (e.g., fish [15–17], decapods [18, 19], and aquatic insects [20]) has been linked to the enzyme carbonic anhydrase (CA). CAs are zinc metalloenzymes that catalyze the reversible hydration/dehydration reaction:  $\text{CO}_2 + \text{H}_2\text{O} \rightleftharpoons \text{H}_2\text{CO}_3 \rightleftharpoons \text{HCO}_3^- + \text{H}^+$ , and are fundamental to many biological processes in addition to acid-base regulation, for example, photosynthesis [21], respiration [18, 22], osmoregulation [18,

22], bone resorption [23], and biomineralization [24]. CAs are classified into five evolutionarily distinct and unrelated superfamilies:  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , and  $\zeta$ ; each superfamily has different active site amino acids, primary sequences, and protein structure [25, 26]. These families are thought to be the result of convergent evolution. The  $\alpha$ -CA superfamily typically has 16 or 17 different isoforms within vertebrates, which are the primary contributors to acid-base regulation. The  $\alpha$ -CA superfamily is broken into four families: cytosolic, secretory, transmembrane/membrane-bound, and CA-related proteins (CA-RP), the latter of which have purportedly lost function due to the loss of at least one of the three active site Histidine residues [27]. In fish, decapods, and aquatic insects, the cytosolic and membrane-bound  $\alpha$ -CAs in gills have been shown to regulate internal pH (Figure 1). The  $\beta$ -CAs are typically found only in bacteria, plants, algae, and fungi; however  $\beta$ -CAs have recently been found in some animals such as *Caenorhabditis elegans* [28], *Anopheles gambiae* [29], and *Daphnia pulex* [22]. There is a lack of knowledge on the catalytic activity and expression of  $\beta$ -CAs in animals, but in plants they are catalytically similar to  $\alpha$ -CAs in animals. The  $\gamma$ -CAs have only been found in archaea and bacteria, while  $\delta$ -CAs and  $\zeta$ -CAs have only been found in marine diatoms [24]. The  $\zeta$ -CAs are unique among CAs since they replace the zinc ion with cadmium [30].

In this study, we investigated the evolutionary history of  $\alpha$ -CAs in the microcrustacean *Daphnia* genus using phylogenetic methods. Since little is understood about  $\beta$ -CAs in animals, this study focuses on  $\alpha$ -CAs. *Daphnia* spp. are keystone aquatic herbivores and an emerging model organism, whose genome has been sequenced and annotated [27]. Interestingly, the *D. pulex* genome has a high rate of gene duplication, three times as high as *Drosophila* genus and nematodes and 30% higher than humans [27]. Since *Daphnia* spp. have 30 isoforms [22] of  $\alpha$ -CAs, compared to the 15 in other organisms; this lends itself to the notion that there have been multiple duplication events within the *Daphnia*  $\alpha$ -CAs. It has been hypothesized that duplication events can be a source for evolutionary novelties and that these duplications can follow one of several evolutionary trajectories: (i) one copy may become silenced (nonfunctionalization); (ii) one copy may acquire a novel beneficial function (neofunctionalization); or (iii) both copies may experience reduced functionality (subfunctionality) [31–33].

In addition, we used the phylogenetic analysis of the superfamily of  $\alpha$ -CAs to clarify which *Daphnia*  $\alpha$ -CAs may be investigated further for their role in acid-base regulation. The criteria for this analysis involved examining *Daphnia*  $\alpha$ -CA genes with functioning  $\alpha$ -CA homologs in other crustaceans [18, 19], aquatic insects [20], and fish [15–17]. Further, we investigated the evolutionary history of  $\alpha$ -CAs in *Daphnia* to elucidate the functionality of duplicate  $\alpha$ -CA genes, if they indeed exist.

## 2. Materials and Methods

**2.1. Sequence Retrieval.** All sequences, except *Daphnia* sequences, were obtained from the National Center for

Biotechnology Information (NCBI). A key word search for “*Homo sapiens* Carbonic Anhydrase” was performed for each of the 16 human isoforms of  $\alpha$ -CA and the amino acid sequences were obtained. For each human isoform, a BLAST search was performed using the BLASTP algorithm with default settings from the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>). Only protein sequences from all taxa with an *E* value lower than  $e^{-75}$  were selected for analysis. The list of sequences was screened to ensure there were no duplicate sequences, based upon 100% sequence conservation in the gene within a given species. Partial sequences were discarded in the final analysis. Twelve  $\alpha$ -CA amino acid sequences were retrieved from bacteria to use as an outgroup. This search resulted in 213 amino acid sequences from taxa ranging from cnidarians to mammals (Table S1 in Supplementary Material available online at <http://dx.doi.org/10.1155/2015/538918>). Our final list of taxa included more vertebrates than invertebrates; this bias is the result of the lack of whole genomes or CA loci sequences within invertebrates. While the bias towards more vertebrates does not affect the overall topology of the phylogeny, more invertebrates could have enhanced the resolution and support of some of the invertebrate clades.

*Daphnia* sequences were obtained from the *Daphnia* Genomics Consortium (DGC) (<http://wflbase.org/>). The *D. pulex* sequences were retrieved using the search function by entering the gene name. The nucleotide sequences were converted to amino acid sequences using MEGA 5.0 [34]. The *D. galeata* sequences were found by blasting the *D. pulex* CAs against the *D. galeata* database from the DGC using the TBLASTN algorithm with default settings [35]. *D. pulicaria* were sequenced (Culver & Morton unpublished data) and converted to amino acid sequences using MEGA 5.0. This search resulted in 30 *D. pulex*, 25 *D. galeata*, and 3 *D. pulicaria* amino acid sequences (Table S1). In addition, each *D. pulex*  $\alpha$ -CA was mapped to their respective chromosome to infer duplication history and duplication events. We mapped each *D. pulex*  $\alpha$ -CA isoform using known scaffold positions on their respective chromosomes [31].

**2.2. Sequence Alignment and Phylogenetic Analysis.** The 271 amino acid sequences were uploaded into CLUSTALX 2.1 [36] and a multisequence alignment was run with iterations after each alignment step. The aligned sequences were then uploaded into MEGA 5.0 and a best model fit was performed. The results from the best model fit indicated a Whelan and Goldman (WAG) model with gamma distribution and invariant sites [37]. Aligned amino acid sequences were then uploaded into the CIPRES web portal [38] and a Bayesian maximum likelihood phylogeny was created using MrBayes 3.1.2 [39] with the following parameters: 1 million iterations, 250 K burn-in, and 2 runs with 8 chains each. In addition, a bootstrapped maximum likelihood RaxML version 8.0 tree was constructed with 1000 iterations [40]. The resultant consensus tree was visualized using FigTree version 1.3.1 [41]; branches were collapsed for ease in reading the rather large phylogeny. Species composition of the collapsed branches can

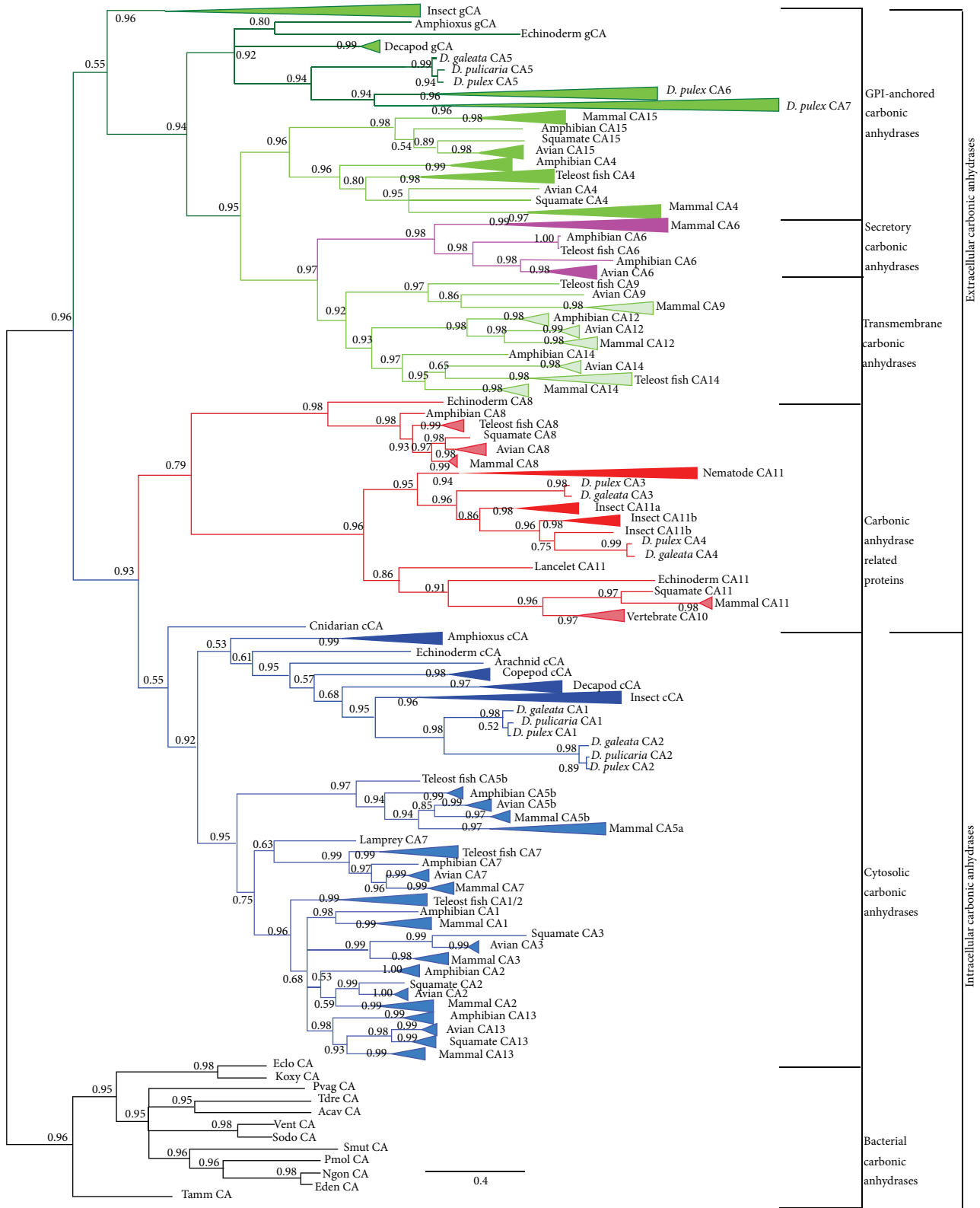


FIGURE 1: Phylogeny of  $\alpha$ -CAs inferred from a maximum likelihood analysis performed with MrBayes; posterior probabilities of branches are indicated at the nodes. Species are collapsed within a larger taxonomical grouping. Branches are colored according to alpha-carbonic anhydrase families: GPI-anchored (dark green: invertebrate; medium green: vertebrate), membrane-bound (light green), secretory (purple), CA-RP (red), and cytosolic (dark blue: invertebrate; light blue: vertebrate). Black branches represent the bacterial outgroup  $\alpha$ -CA families.

be found in Table S1. The aligned sequences were also used to determine residues that were conserved within each group of  $\alpha$ -CAs (Table S2). A cutoff of 80% was used to determine if residues were conserved within an  $\alpha$ -CA group across the entire phylogeny; however, if less than three species were in a group, then 100% conservation was used.

**2.3. Determination of Ancestral States.** Ancestral states of amino acids were inferred using a maximum likelihood approach within MEGA 5.0. Parameterization for the analysis employed a WAG model with gamma distribution and invariant sites and very strong branch swap filters. Criteria to elucidate the ancestral state of amino acids residues were determined by using those residues that are 80% conserved in the bacteria outgroup. This resulted in a reference sequence template that could be used to compare the other isoforms. Residues that were 90% conserved among all the isoforms in the phylogeny were excluded because they were not informative. Residues that were not shared among 50% of the isoforms in each  $\alpha$ -CA group were also excluded to reduce noise. Twenty-seven residues remained for ancestral state analysis. As it is cumbersome to view the changes in ancestral states on the phylogeny, a table (Table S3) was created to facilitate a summary of amino acid residue evolution through the phylogeny. The table includes the predicted ancestral sequence at all nodes (the most recent common ancestor) and the number of amino acid changes from the most recent common ancestor (including homoplasies).

**2.4. N-Terminus, GPI-Anchored, and Transmembrane Prediction of *Daphnia* CA6s and CA7s.** To predict the transmembrane domains in the *Daphnia* CA6s and CA7s, the TMHMM Server v. 2.0 [42] on the Center for Biological Sequence Analysis (CBS) Prediction Server (<http://www.cbs.dtu.dk/>) was utilized. TMHMM uses a hidden Markov model to predict the location and likelihood of transmembrane helices. First, the amino acid sequences from the vertebrate extracellular CAs, determined from the phylogenetic analysis, were uploaded to the TMHMM server to determine if the software could successfully predict the known transmembrane CAs from the secretory and glycosylphosphatidylinositol (GPI-) anchored  $\alpha$ -CAs. The *Daphnia* CAs amino acid sequences were then uploaded into the TMHMM server. Those sequences that had a posterior probability greater than 0.80 and no N-terminus signal peptides were predicted to be transmembrane CAs.

To determine N-terminal sequences and cleavage sites in the *Daphnia* CA6s and CA7s, we used the TargetP 1.1 Server [43] on the CBS Prediction Server. As with predicting transmembrane domains, amino acid sequences from known vertebrate extracellular CAs, as determined by the phylogenetic analysis, were uploaded into the TargetP 1.1 Server with the following parameters: (i) nonplant organisms; (ii) cleavage sites predicted; and (iii) a specificity cutoff of greater than 0.7. This run was used to determine if the software could successfully predict the  $\alpha$ -CAs with known N-terminus sequences. *Daphnia* CAs amino acids from the CA6s and

CA7s were then uploaded in the TargetP 1.1 Server with the same parameters to predict N-terminus signal peptides.

GPI-anchored proteins in *Daphnia* genus were predicted using the online based software, GPI-SOM (<http://gpi.unibe.ch/>) [44]. GPI-SOM uses a Kohonen self-organizing mapping approach to predict C-terminus anchoring signal and anchoring site. GPI-anchoring sites are only found in the C-terminus of a protein. GPI-anchoring proteins also contain N-terminus signaling peptides. Known vertebrate extracellular amino acid sequences, as determined from the phylogenetic analysis, were uploaded into GPI-SOM and were run with default parameters. The results were used to elucidate whether GPI-SOM could successfully predict the known GPI-anchored  $\alpha$ -CAs from the secretory and transmembrane  $\alpha$ -CAs. Amino acids from *Daphnia* CA6s and CA7s were uploaded into GPI-SOM to predict *Daphnia* GPI-anchored  $\alpha$ -CAs.

The prediction software was able to successfully place the vertebrate  $\alpha$ -CA subfamilies into their respective subfamilies, for instance transmembrane, secretory, and GPI-anchored  $\alpha$ -CAs.

### 3. Results and Discussion

**3.1. General Phylogenetic Distribution of CA Types in Animals.** Typically  $\alpha$ -CAs are characterized by (i) four active site residues: Histidine- (His-) 316 (His-64 using nomenclature of vertebrate  $\alpha$ -CAs), Glutamine- (Gln-) 353, Glutamic Acid- (Glu-) 372, and Threonine- (Thr-) 498; (ii) three zinc-binding site residues: His-355, His-357, and His-385; and (iii) two substrate-binding site residues: Thr-498 and Thr-499. The residue His-316 acts as a proton shuttle from the zinc ion and is considered a rate-limiting step in the catalytic process [45]. So the inclusion of the His at 316 is important in determining the activity level of the enzyme. Further, the residues Thr-498 and Thr-499 result in a threonine loop, which coordinates the zinc ion and is important in the overall activity of the enzyme [46]. The amino acid alignment shows that these residues are highly conserved throughout the phylogeny (Table S2). Also the residues surrounding these highly conserved residues have recognizable motifs that are also highly conserved. There are also three motifs that are highly conserved that are not associated sequentially with any of the active, zinc-binding, or substrate-binding sites: the motif QSPINI found at residues 219–224, GLAVLG found at residues 408–413, and N-RP-QPL at residues 570–577.

The phylogenetic results of the MrBayes (Figure 1) and RAXML (Figure S1) analyses produced similar topologies. The phylogenies indicate that the first divergence in  $\alpha$ -CAs resulted in two sister clades representing extracellular and intracellular  $\alpha$ -CAs and appeared after the split of animals, plants, and fungi from bacteria. Before this early divergence, the most likely ancestral state of the  $\alpha$ -CAs was extracellular (which include the GPI-anchored, transmembrane/membrane-bound, and secretory  $\alpha$ -CAs), as is evident by the bacterial  $\alpha$ -CAs having similarly conserved residues as the extracellular  $\alpha$ -CAs in animals (Table S3). Another line of evidence suggests that bacterial  $\alpha$ -CAs are

formed near to or on the cytoplasmic membrane [47]. In addition, Le Roy et al. [24] found that Porifera  $\alpha$ -CAs were more similar to extracellular  $\alpha$ -CAs and were more basal phylogenetically than intracellular  $\alpha$ -CAs. In particular, both bacterial  $\alpha$ -CAs and extracellular animal  $\alpha$ -CAs share the same active site residues, zinc-binding site residues, and substrate-binding site residues. In addition, they have disulfide bonding sites at residues Cysteine- (Cys-) 214 and Cys-502 that are not found in cytosolic  $\alpha$ -CAs (Tables S2 and S3). Further, extracellular  $\alpha$ -CAs and bacterial  $\alpha$ -CAs share the following conserved residues that are not found in intracellular  $\alpha$ -CAs (however, they are found in some of the CA-RPs): Asparagine- (Asn-) 314, Asn-434, Tyrosine- (Tyr-) 491, Arginine- (Arg-) 492, and Arg-578 (Table S3). These results are contrary to the commonly held notion that the intracellular  $\alpha$ -CAs are the ancestral state [25].

The GPI-anchored  $\alpha$ -CAs, found within the extracellular  $\alpha$ -CAs, form three monophyletic clades: these clades consist of an invertebrate clade (including the chordate amphioxus), vertebrate clade, and an insect clade. Of note is the fact that insects did not fall within the invertebrate clade; however, the insect GPI-anchored clade has weak support (posterior probability = 0.55, Figure 1). The vertebrate subclade of GPI-anchored  $\alpha$ -CAs is characterized by CA4 and CA15. In vertebrates, CA4 is localized in the kidneys, gastrointestinal tract, and endothelium, while CA15 is localized in the kidneys and is not expressed in humans [45, 47]. One of the weaknesses of constructing robust phylogenies of metazoans  $\alpha$ -CAs is that there is a lack of depth of taxon sequence coverage in invertebrate organisms. Some researchers choose to limit their analysis to organisms that have whole genome sequences, in order to increase the likelihood of capturing all isoforms; however, this limits the number of taxa that can be used. We chose to use both whole genomes and individually sequenced  $\alpha$ -CA isoforms to increase coverage of both isoforms and taxa [24]. However, even taking this approach, there is a severe lack of data on sequenced  $\alpha$ -CAs within invertebrates, thus weakening support for some relationships within the phylogeny. This may also cause some sampling bias when trying to deduce the rate of duplication events between invertebrate taxa. With the plethora of next-generation studies taking place, perhaps this lack in data will be resolved in the near future.

The GPI-anchored  $\alpha$ -CAs further diverged into the secretory type  $\alpha$ -CAs due to the loss of the C-terminus cleavage and anchoring site [22, 46], which occurred after the appearance of amphioxus. This can be deduced, since the secretory  $\alpha$ -CA appears in all vertebrates. The secretory  $\alpha$ -CA is characterized by CA6s, which is localized in the saliva of vertebrates. Membrane-bound  $\alpha$ -CAs diverged from a common ancestor with the secretory  $\alpha$ -CAs based on phylogenetic support that shows the divergence occurring after amphioxus, but before the amphibian/fish divergence. The transmembrane  $\alpha$ -CAs are characterized by the further loss of the N-terminus signal peptides and the development of helices that are embedded in the cell membrane and are represented by CA9, CA12, and CA14.

After the split of animal phyla, extracellular  $\alpha$ -CAs diverged from intracellular  $\alpha$ -CAs. The intracellular  $\alpha$ -CAs

are characterized by an amino acid change from the ancestral state at the following residues: 233 from Isoleucine (Ile) to Proline (Pro), 314 from Asn to Thr, 318 from Ile to Serine (Ser), 319 from Gln to Phenylalanine (Phe), 448 from Ile to Thr, 491 from Tyr to Tryptophan (Trp), 492 from Arg to Thr, and 505 from Glycine (Gly) to Ser (Table S3). During the evolution of intracellular  $\alpha$ -CAs, a duplication event likely occurred, which split intracellular  $\alpha$ -CAs into two clades: CA-related proteins (CA-RPs) and cytosolic  $\alpha$ -CAs. The CA-RPs are characterized by an amino acid change at the active site residue 353 from Gln to Glu in all the CA-RPs, an amino acid change at the active site residue 316 from His to Ser in the CA11s, and an additional amino acid change at the zinc-binding site residue 385 from His to Gln in the CA10s and CA11s, which resulted in the complete loss of function (nonfunctionalization) or a different function (neofunctionalization) in these enzymes [45]. According to the phylogeny, this duplication must have occurred before the emergence of cnidarians.

The CA-RPs form a large monophyletic group made up of three subclades. One subclade consists of CA8 and contains only deuterostomes. An interesting feature of the CA8 subclade is that there are relatively short branch lengths across a diverse group of taxa, suggesting high conservation within this subclade despite these isoforms being noncatalytic. These results suggest that CA8 may have an important biological function within deuterostomes [27]. Further, the CA8 subclade is sister to both CA11s and CA10s. The CA11s (including *Daphnia* CA3 and CA4, nomenclature for the  $\alpha$ -CAs of *Daphnia* genus and many invertebrates are not consistent with the nomenclature of  $\alpha$ -CAs used for mammals) form two distinct groupings: protostomes and deuterostomes.

The cytosolic  $\alpha$ -CAs make up a monophyletic group and are characterized by the loss of the disulfide bond at residue 214 due to the Cys converting to different amino acids that do not facilitate disulfide bonds. The loss of the disulfide bond at residue 214 in the cytosolic  $\alpha$ -CAs suggests relaxed selection, since these enzymes do not need the extra structural integrity provided by the disulfide bond to deal with the environment outside the cell [24, 46]. Also of interest within the cytosolic  $\alpha$ -CAs is the fact that there was an amino acid change at the active site residue 316 from His to Asn in vertebrates and to a Thr in invertebrates, with the subsequent reemergence of the His at residue 316 in vertebrate CA1, 2, 7, and 13, and fish CA1/2 (Table S3). The His-316 residue is important in the activity of the enzyme in that it acts as a proton shuttle from the zinc ion and is considered a rate-limiting reaction [46]. In addition, vertebrate CA1 and CA13 have a conversion of Thr-499, an important residue in the coordination of the zinc ion and is important in catalytic activity, to His in CA1 and Valine (Val) in CA7. This supports why CA2 has been determined to be the most active of the vertebrate cytosolic  $\alpha$ -CAs, while the others have varying degrees of lower activity [46]. Within the cytosolic  $\alpha$ -CAs, after the (weakly supported: posterior probability = 0.55) divergence of cnidarians, there is a split resulting in an exclusively vertebrate clade and a clade containing all the invertebrates (including amphioxus). The most basal group of the vertebrate clade consists exclusively of CA5, which is associated with mitochondria [46], with

the next divergence from the CA5/7 common ancestor being CA7 followed by the teleost fish CA1/2. The teleost fish CA1/2 clade shows evidence of a duplication event [48, 49]; however, it is not universal to all teleosts. This is represented within the collapsed clade of teleost fish CA1/2 of the cytosolic  $\alpha$ -CAs (Figure 1). In addition, after the divergence of teleost fish CA1/2, a polytomy is formed and the relationship among CA1, CA2, CA3, and CA13 type cytosolic  $\alpha$ -CAs cannot be resolved. Here, the RAxML tree resolves the polytomy but has relatively weak support (bootstrap values = 19–44, Figure S1). The sister group to the exclusively vertebrate subclade contains all the invertebrates and amphioxus.

**3.2. *Daphnia* CA Isoforms.** Of the *Daphnia*  $\alpha$ -CAs, two fall within the cytosolic family (CA1 and CA2), clustering with other arthropods, echinoderms, and cnidarians (Figure 1). *Daphnia* CA5 clusters with GPI-anchoring  $\alpha$ -CAs of other arthropods (Figure 1). Two other  $\alpha$ -CAs (CA3 and CA4) cluster within the CA-RP clade and are sisters to hexapod CA-RPs (Figure 1). Specifically, CA3 is closely associated with hexapod CA11a, while CA4 is sister to hexapod CA11b.

The remaining 25  $\alpha$ -CAs form two sister clades, CA 6B-G and CA7A-Q (including CA6A and CA6H), that diverged from CA5. In previous work [22], CA6H was the first branch in the CA6s clade, while CA6A was excluded from the phylogeny. Since in this study CA6A and CA6H cluster with the CA7s, we would propose to rename these genes as CA7R and CA7S, respectively, since each of the nodes have good posterior probability support (0.96) (Figure 2). Weber and Pirow [22] suggest that CA6s and CA7s are secretory CAs due to fact that they have N-terminus signaling peptides; however, our analysis does not support that all the CA6s and CA7s are secretory. Using transmembrane, N-terminus, and GPI-anchoring software we found that, like Weber and Pirow, none of the CA6s or CA7s are transmembrane  $\alpha$ -CAs using a posterior probability cutoff of >0.8. We did find evidence to support that *Daphnia* CA6F, CA7H, CA7K, and CA7O are GPI-anchored  $\alpha$ -CAs in that they all had N-terminus signaling peptides and C-terminus cleavage and anchoring sites (Table 1 and Figure 2). All the remaining *Daphnia* CA6s and CA7s, except CA6E and CA7Q, had N-terminus signaling peptides (specificity > 0.7) without the C-terminus cleavage and anchoring sites, suggesting that these  $\alpha$ -CAs are secretory (Table 2, Figure 2). The two remaining *Daphnia*  $\alpha$ -CAs, CA6E and CA7C, were not predicted to be transmembrane, secretory, or GPI-anchoring proteins and may have some cytosolic function (Table 1, Figure 2). Le Roy et al. [24] also found cytosolic-like CAs in poriferans and mollusks that are involved in biocalcification in  $\alpha$ -CAs within their respective extracellular clade. They suggest that this may be an internalization of a formally secreted  $\alpha$ -CA or they may be secreted proteins that are shuttled out of the cell in a novel manner. Further research is warranted to verify the function and localization of these  $\alpha$ -CAs.

**3.3. Duplication Events in CA Isoforms in *Daphnia* Genus.** Phylogenetic results also support the hypothesis of multiple

duplication events in *Daphnia* genus. The first duplication event seems to be the result of gene-level duplication in an ancestral species that resulted in the divergence of cytosolic  $\alpha$ -CAs from extracellular  $\alpha$ -CAs and CA-RPs. This is supported by the fact that extracellular  $\alpha$ -CAs and CA-RPs (the predicted ancestral state of  $\alpha$ -CAs), as a group, are found in tandem on chromosome 7 (Table 2). The second duplication is the result of a genome-level duplication event in an ancestral species, which led to the divergence of the CA-RPs from the cytosolic  $\alpha$ -CAs. Evidence in support of this is that the CA-RPs (and extracellular  $\alpha$ -CAs) are found on chromosome 7, while the cytosolic  $\alpha$ -CAs are found on chromosome 4. Since, *Daphnia* spp. are known to have a high level transposable elements [27], this could be a potential mechanism through which the gene was able to move within the genome. Another potential mechanism could be chromosome duplication. Further investigation is needed to determine which mechanism is supported. As with the already discussed isoforms, the remaining isoforms (CA1 and CA2, CA3 and CA4, and CA6s and CA7s) also appear to be the result of duplication events. One possibility is that these duplications are the result of tandem duplications, as many of the genes are in synteny (Table 2). Although there is only the one GPI-anchored  $\alpha$ -CA in *Daphnia* genus (CA5), there is a radiation of 25  $\alpha$ -CAs (CA6A-H and CA7A-Q), which diverged from CA5. When the CA6s and CA7s diverged from CA5, they lost their GPI-anchoring site but retained the N-terminus signaling peptide sequence allowing for neofunctionalization as secretory  $\alpha$ -CAs (Table 1 and Figure 2). Four isoforms, CA6F, CA7H, CA7K, and CA7O, later reverted to GPI-anchored  $\alpha$ -CAs through convergent evolution. Additionally, two isoforms, CA6E and CA7Q, lost both the N-terminus signaling peptide and C-terminus cleavage sequence. This suggests that they either became cytosolic or developed a novel secretory pathway [24]. Several studies of *Daphnia* genus and other invertebrate genes and genomes have unveiled duplicated genes that have led to neofunctionalization, such as the spooky genes in arthropods [50]. If these duplications prove to be neofunctional, then *Daphnia*  $\alpha$ -CAs would have a larger than expected number of neofunctional isoforms. Kondrashov [51], in his review, explains how it is possible for duplicated genes to persist in the genome long enough to eventually evolve into neofunctional genes through the redundancy hypothesis, which postulates that duplicate genes are not deleterious but are maintained through neutral processes and can evolve into neofunctional genes if they lead to a fitness advantage. *Daphnia's* two CA-RPs do not appear to be duplicated within the genus but belong to a larger duplication within the phylum Arthropoda that occurred after the divergence of arthropods and nematodes.

The fact that *Daphnia* genus has 30 isoforms of  $\alpha$ -CA, while most vertebrates only have 15 or 16, lends support to previous work on the *Daphnia* genome, which found that *Daphnia* genus has a relatively high rate of gene duplication, at least within *D. pulex*. These high duplication rates are not novel to *Daphnia* genus and have been shown in another cyclically parthenogenetic organism, the pea aphid, *Acyrtosiphon pisum* [52, 53].

TABLE 1: Results of prediction software to determine whether proteins are transmembrane, secretory, or GPI-anchored. Transmembrane proteins were determined using the TMHMM server on the CBS Prediction Server with a posterior probability  $>0.8$  and no N-terminus prediction on a transmembrane protein. N-terminus signaling peptides were elucidated using TargetP on the CBS Prediction Server, with a specificity  $>0.7$  indicating a high probability of a N-terminus signaling peptide. GPI-SOM was used to predict C-terminus cleavage and anchoring sites. If a protein was not transmembrane and has both N-terminus and C-terminus it was predicted to be a GPI-anchored protein. If it had only a N-terminus prediction, it was classified as secretory protein. If it did not fit any category it was classified as a cytosolic-like protein.

<i>Daphnia</i> CA isoform criteria ->	Transmembrane $>0.8$ and no N-terminus	N-Terminus $>0.7$ specificity	C-Terminus Most probable	Prediction
CA5	No	Yes	Yes	GPI-anchored
CA6A	No	Yes	Not	Secretory
CA6B	No	Yes	Not	Secretory
CA6C	No	Yes	Not	Secretory
CA6D	No	Yes	Not	Secretory
CA6E	No	0.178	Not	Cytosolic-like
CA6F	No	Yes	Yes	GPI-anchored
CA6G	No	Yes	Not	Secretory
CA6H	No	Yes	Not	Secretory
CA7A	No	Yes	Not	Secretory
CA7B	No	Yes	Not	Secretory
CA7C	No	Yes	Not	Secretory
CA7D	No	Yes	Not	Secretory
CA7E	No	Yes	Not	Secretory
CA7F	No	Yes	Not	Secretory
CA7G	No	Yes	Not	Secretory
CA7H	No	Yes	Yes	GPI-anchored
CA7I	No	Yes	Not	Secretory
CA7J	No	Yes	Not	Secretory
CA7K	No	Yes	Yes	GPI-anchored
CA7L	No	Yes	Not	Secretory
CA7M	No	Yes	Not	Secretory
CA7N	No	Yes	Not	Secretory
CA7O	No	Yes	Yes	GPI-anchored
CA7P	No	Yes	Not	Secretory
CA7Q	No	0.288	Not	Cytosolic-like

The phylogeny presented here shows that this may be a genus-wide phenomenon, as *D. galeata* also shares this radiation event within the CA6s and CA7s; however, the *D. galeata* radiation is not as extensive: 20 isoforms of CA6s and CA7s, as compared to 25 in *D. pulex*. As the genomes of two additional daphniid species (i.e., *D. magna* and *D. pulicaria*) are completed, it will be of interest to determine if these genomes support the finding of a large radiation of CA6s and CA7s within the genus *Daphnia*. It also appears that CA1 and CA2 are the products of duplication within

the genus, whereas the duplication of CA3 and CA4 appears to be within the whole arthropod phylum; however, as more arthropod genomes are sequenced, this may fill in gaps in the phylogeny.

#### 4. Conclusions

The results of this phylogenetic study support the previously held organization of the  $\alpha$ -CA superfamily of genes, namely, the fact that  $\alpha$ -CAs are clustered into the following

TABLE 2: Results of chromosome mapping which reflect the *D. pulex*  $\alpha$ -CA isoforms and their scaffold designation along with their start and end positions on the scaffold. Three isoforms could not be mapped to a chromosome because their scaffolds have not been mapped to their respective chromosome.

<i>D. pulex</i> CA	Scaffold	Start position	End position	Chromosome	Dappu ID
CA1	8	293280	297489	4	442498
CA2	8	1005314	1007373	4	442497
CA3	74	63490	73363	NA	442499
CA4	4	1033301	1039412	7	442496
CA5	20	1028754	1037862	NA	442477
CA6A	4	1676667	1677698	7	442779
CA6B	4	1678702	1680800	7	442471
CA6C	4	1682181	1683985	7	442472
CA6D	4	1687613	1689716	7	442467
CA6E	4	1692426	1694512	7	442475
CA6F	4	1699762	1703139	7	442468
CA6G	4	1707093	1708695	7	442476
CA6H	4	2922515	2924220	7	442478
CA7A	4	2427959	2429626	7	442480
CA7B	4	2430816	2432334	7	442481
CA7C	4	2435092	2435571	7	442482
CA7D	4	2436638	2438161	7	442483
CA7E	4	2438986	2440394	7	442484
CA7G	4	1707093	1708695	7	442494
CA7H	4	2463490	2465025	7	442485
CA7I	4	2466064	2467469	7	442486
CA7J	4	2468464	2470065	7	442487
CA7K	4	2470727	2472139	7	442488
CA7L	4	2474751	2475358	7	442489
CA7M	4	2477904	2479557	7	442491
CA7N	4	2480236	2482046	7	442490
CA7O	4	2482392	2383774	7	442492
CA7P	4	2486891	24888402	7	442493
CA7Q	40	788747	790739	NA	442495

families: cytosolic, CA-RP, GPI-anchored, secretory, and membrane-bound [25]. Previous thought, however, was that intracellular  $\alpha$ -CAs were the most likely ancestral state. In contrast, our results provide support that extracellular  $\alpha$ -CAs are the likely ancestral state. The added knowledge from this extensive phylogeny elucidates the relationship among invertebrates and vertebrates. For instance, the GPI-anchored and cytosolic  $\alpha$ -CAs are divided into invertebrate and vertebrate groups. The nomenclature that is used for the vertebrate CAs does not hold up when looking at invertebrate groups. For instance, some of the cytosolic invertebrate CAs are named CA1, CA2, or cCA but do not have any phylogenetic relationship to vertebrate CA1 or CA2. In fact the invertebrate  $\alpha$ -CAs are more closely related to the more basal vertebrate  $\alpha$ -CAs, CA7 and CA5 (Table S3). Since the vertebrate CA5 is associated with mitochondria, invertebrate cytosolic  $\alpha$ -CAs are, therefore, more similar to vertebrate CA7. This is also true for the extracellular  $\alpha$ -CAs; however, since there are many homoplasies occurring within each invertebrate taxonomical clade, it is difficult to determine their relationship

to the established nomenclature of vertebrate extracellular  $\alpha$ -CAs. Further, invertebrate groups have reduced diversity of cytosolic and extracellular  $\alpha$ -CAs, when compared to vertebrates. However, this may be an artifact of the fact that invertebrate  $\alpha$ -CAs have not been well investigated. A study of the purple sea urchin (*Strongylocentrotus purpuratus*) has uncovered 19 isoforms of  $\alpha$ -CA, most of which are involved in acid-base regulation [54]. Also, *D. pulex* has 30 different isoforms [22]. Most of these isoforms are the result of a radiation of CA6s and CA7s that diverged from *Daphnia* CA5, which is a GPI-anchored  $\alpha$ -CA. Further investigation of invertebrate  $\alpha$ -CAs may uncover a greater diversity of  $\alpha$ -CAs within these families.

We had several overarching goals in performing this study with regard to the *Daphnia* genus. First, we were interested in the evolution of *Daphnia*  $\alpha$ -CAs. Homologs of acid-base regulating  $\alpha$ -CAs in organisms, such as crustaceans, aquatic insects, and fish were used to provide evidence in support of potential acid-base regulating  $\alpha$ -CAs in *Daphnia* genus. Second, we were interested in gene duplication events



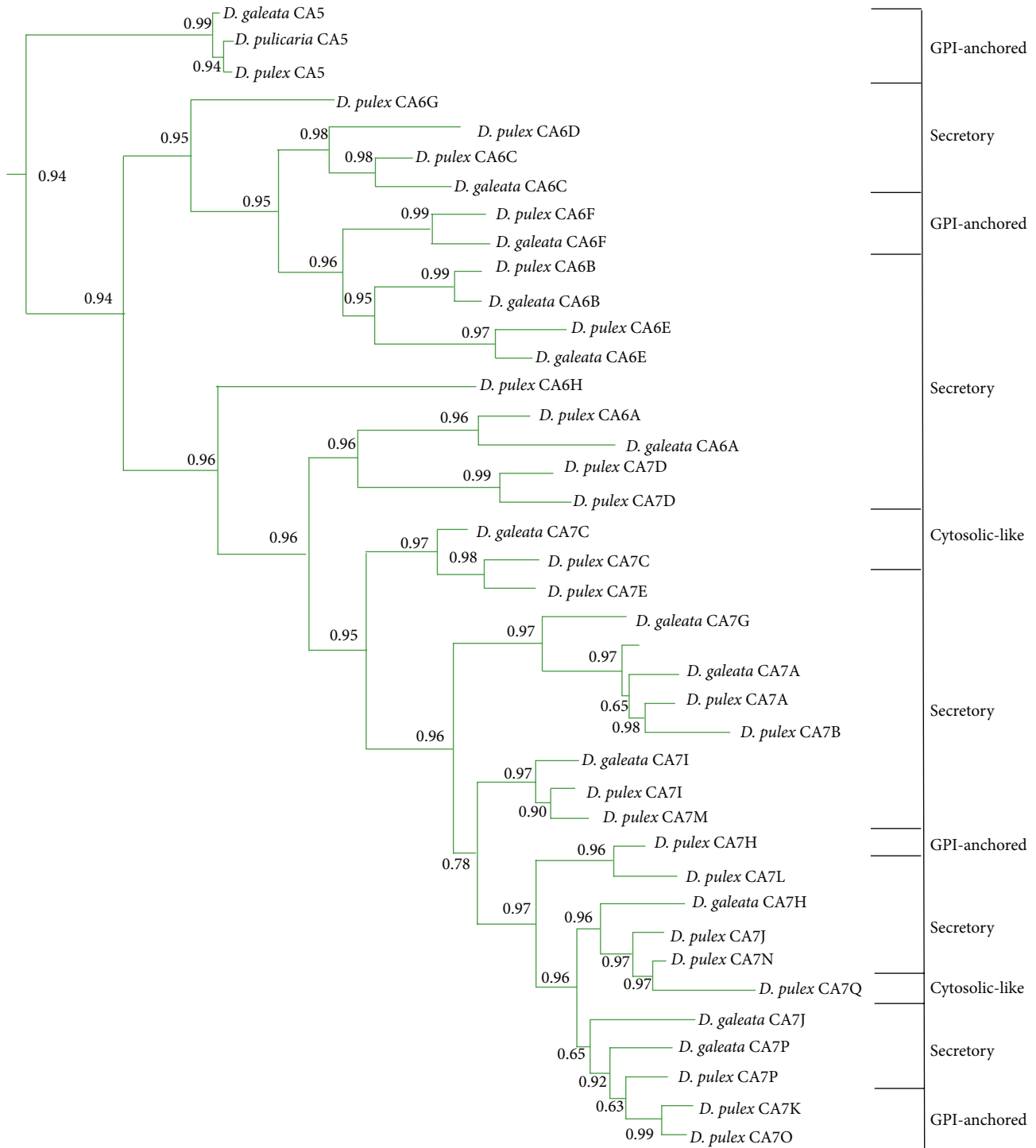


FIGURE 2: Isolated view of *Daphnia* CA5, CA6s, and CA7s based on the phylogeny represented in Figure 1. Posterior probabilities of the branches are indicated at the nodes. On the right side of the phylogeny are the predicted states for *Daphnia* CA5, CA6s, and CA7s.

in the  $\alpha$ -CA superfamily, specifically within the *Daphnia* genus, and the fate of these duplicated genes evolutionarily.

To address the goal of identifying potential genes involved in acid-base regulation in *Daphnia* genus, several candidate genes, including CA1, CA2, and CA5, may be implicated as a starting point for investigation. Since these three genes have homologs in other arthropods that have been

previously determined physiologically to be active in acid-base regulation [18], these *Daphnia* CA genes warrant further study (e.g., physiological fitness assays across a range of pH conditions) to determine their functionality. To date, it is uncertain what the expression levels of  $\alpha$ -CAs are in *Daphnia* genus. In other organisms, however, some experiments have shown differential expression of  $\alpha$ -CAs across a pH gradient.

For instance, Evans et al. [54] found that  $\alpha$ -CA12 was differentially expressed in larval *S. purpuratus* at low pH conditions. Also, Lin et al. [16] found differential expression of  $\alpha$ -CA2 and CA15 in the gills of zebrafish under differing pH conditions. Current work in our lab is trying to characterize these genes and elucidate the differential expression of  $\alpha$ -CAs across pH gradients. Furthermore, this study can be useful as a reference for any future acid-base regulation work in other arthropods, particularly crustaceans.

Within *Daphnia* CAs, there is a major radiation within the CA6 and CA7s of 25 CA isoforms, which diverge from CA5. In addition, CA1/CA2 and CA3/CA4 represent additional, independent duplications compared to other arthropods. This recurrent observation of multiple duplication events in *Daphnia* genus lends support to the hypothesis that the ecoresponsive nature of this organism may be due to possible neofunctionalization resulting from the high levels of gene duplication. Thus, the genome duplications in *Daphnia* genus may allow this organism to withstand an extensive range of environmental (e.g., pH) conditions that are encountered in aquatic habitat [27].

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This study was supported by the U.S. National Science Foundation (NSF-IOS-OEI) grant to L. J. Weider (no. 0924289), the University of Oklahoma Department of Biology Adams Scholarship to Billy W. Culver, and the University of Oklahoma Biological Station Summer Research Grant to Billy W. Culver. The authors thank the University of Minnesota LacCORE staff (in particular R. O'Grady), for assistance with coring and providing *D. pulicaria* samples from Madison Lake, MN. A. Harris and J. Medders facilitated sequencing of *D. pulicaria*. The authors also thank L. J. Weider for his advice, mentorship, and comments on this paper. In addition, the authors also thank R. Broughton, D. Hambright, I. Schlupp, and R. Cichewicz for their input. The authors thank J. Cureton and R. Hartnett who reviewed and provided invaluable comments on earlier drafts of this paper. The authors would also like to thank the two anonymous reviewers that greatly improved the quality of this paper. This paper constitutes a portion of BWC's doctoral dissertation at the University of Oklahoma.

## References

- [1] P. D. Tortell, "Evolutionary and ecological perspectives on carbon acquisition in phytoplankton," *Limnology and Oceanography*, vol. 45, no. 3, pp. 744–750, 2000.
- [2] N. G. Hairston Jr., C. L. Holtmeier, W. Lampert et al., "Natural selection for grazer resistance to toxic cyanobacteria: evolution of phenotypic plasticity?" *Evolution*, vol. 55, no. 11, pp. 2203–2214, 2001.
- [3] P. D. Jeyasingh, A. Ragavendran, S. Paland, J. A. Lopez, R. W. Sterner, and J. K. Colbourne, "How do consumers deal with stoichiometric constraints? Lessons from functional genomics using *Daphnia pulex*," *Molecular Ecology*, vol. 20, no. 11, pp. 2341–2352, 2011.
- [4] J. Urabe, J. Togari, and J. J. Elser, "Stoichiometric impacts of increased carbon dioxide on a planktonic herbivore," *Global Change Biology*, vol. 9, no. 6, pp. 818–825, 2003.
- [5] H. G. Pollard, J. K. Colbourne, and W. Keller, "Reconstruction of centuries-old *Daphnia* communities in a lake recovering from acidification and metal contamination," *Ambio*, vol. 32, no. 3, pp. 214–218, 2003.
- [6] C. M. G. Martins, D. V. Almeida, L. F. F. Marins, and A. Bianchini, "mRNA expression and activity of ion-transporting proteins in gills of the blue crab *Callinectes sapidus*: effects of waterborne copper," *Environmental Toxicology and Chemistry*, vol. 30, no. 1, pp. 206–211, 2011.
- [7] W. J. O'Brien and F. deNoyelles Jr., "Photosynthetically elevated pH as a factor in zooplankton mortality in nutrient enriched ponds," *Ecology*, vol. 53, no. 4, pp. 605–614, 1972.
- [8] K. E. Havens, "Acidification effects on the plankton size spectrum: an in situ mesocosm experiment," *Journal of Plankton Research*, vol. 14, no. 12, pp. 1687–1696, 1992.
- [9] A. M. Derry and S. E. Arnott, "Adaptive reversals in acid tolerance in copepods from lakes recovering from historical stress," *Ecological Applications*, vol. 17, no. 4, pp. 1116–1126, 2007.
- [10] M. Erlandsson, N. Cory, J. Fölster et al., "Increasing dissolved organic carbon redefines the extent of surface water acidification and helps resolve a classic controversy," *BioScience*, vol. 61, no. 8, pp. 614–618, 2011.
- [11] A. Moya, L. Huisman, E. E. Ball et al., "Whole transcriptome analysis of the coral acropora millepora reveals complex responses to CO<sub>2</sub>-driven acidification during the initiation of calcification," *Molecular Ecology*, vol. 21, no. 10, pp. 2440–2454, 2012.
- [12] R. G. Eppinger, P. H. Briggs, C. Dusel-Bacon et al., "Environmental geochemistry at Red Mountain, an unmined volcanogenic massive sulphide deposit in the Bonfield district, Alaska Range, east-central Alaska," *Geochemistry: Exploration, Environment, Analysis*, vol. 7, no. 3, pp. 207–223, 2007.
- [13] K. Satake, A. Oyagi, and Y. Iwao, "Natural acidification of lakes and rivers in Japan: the ecosystem of Lake Usoriko (pH 3.4–3.8)," *Water, Air, and Soil Pollution*, vol. 85, no. 2, pp. 511–516, 1995.
- [14] Y. Ezoe, C. H. Lin, M. Noto, Y. Watanabe, and K. Yoshimura, "Evolution of water chemistry in natural acidic environments in Yangmingshan, Taiwan," *Journal of Environmental Monitoring*, vol. 4, no. 4, pp. 533–540, 2002.
- [15] T. Georgalis, S. F. Perry, and K. M. Gilmour, "The role of branchial carbonic anhydrase in acid-base regulation in rainbow trout (*Oncorhynchus mykiss*)," *Journal of Experimental Biology*, vol. 209, no. 3, pp. 518–530, 2006.
- [16] T.-Y. Lin, B.-K. Liao, J.-L. Horng, J.-J. Yan, C.-D. Hsiao, and P.-P. Hwang, "Carbonic anhydrase 2-like a and 15a are involved in acid-base regulation and Na<sup>+</sup> uptake in zebrafish H<sup>+</sup>-ATPase-rich cells," *The American Journal of Physiology—Cell Physiology*, vol. 294, no. 5, pp. C1250–C1260, 2008.
- [17] K. M. Gilmour and S. F. Perry, "Carbonic anhydrase and acid-base regulation in fish," *The Journal of Experimental Biology*, vol. 212, no. 11, pp. 1647–1661, 2009.
- [18] R. P. Henry and J. N. Cameron, "The role of Carbonic Anhydrase in respiration, ion regulation and acid-base balance in

- the aquatic crab *Callinectes sapidus* and the terrestrial crab *Gecarcinus lateralis*,” *The Journal of Experimental Biology*, vol. 103, pp. 205–223, 1983.
- [19] R. P. Henry, “The role of carbonic anhydrase in blood ion and acid-base regulation,” *American Zoologist*, vol. 24, no. 1, pp. 241–251, 1984.
- [20] P. D. Cooper, “Mechanisms of hemolymph acid-base regulation in aquatic insects,” *Physiological Zoology*, vol. 67, no. 1, pp. 29–53, 1994.
- [21] M. R. Badger and G. D. Price, “The role of carbonic anhydrase in photosynthesis,” *Annual Review of Plant Physiology and Plant Molecular Biology*, vol. 45, pp. 369–392, 1994.
- [22] A. K. Weber and R. Pirow, “Physiological responses of *Daphnia pulex* to acid stress,” *BMC Physiology*, vol. 9, no. 1, article 9, 25 pages, 2009.
- [23] S. L. Teitelbaum, “Bone resorption by osteoclasts,” *Science*, vol. 289, no. 5484, pp. 1504–1508, 2000.
- [24] N. Le Roy, D. J. Jackson, B. Marie, P. Ramos-Silva, and F. Marin, “The evolution of metazoan  $\alpha$ -carbonic anhydrases and their roles in calcium carbonate biomineralization,” *Frontiers in Zoology*, vol. 11, article 75, pp. 1–16, 2014.
- [25] D. Hewett-Emmett and R. E. Tashian, “Functional diversity, conservation, and convergence in the evolution of the  $\alpha$ -,  $\beta$ -, and  $\gamma$ -carbonic anhydrase gene families,” *Molecular Phylogenetics and Evolution*, vol. 5, no. 1, pp. 50–77, 1996.
- [26] A. K.-C. So and G. S. Espie, “Cyanobacterial carbonic anhydrases,” *Canadian Journal of Botany*, vol. 83, no. 7, pp. 721–734, 2005.
- [27] A. Aspatwar, M. E. E. Tolvanen, and S. Parkkila, “Phylogeny and expression of carbonic anhydrase-related proteins,” *BMC Molecular Biology*, vol. 11, article 25, 19 pages, 2010.
- [28] M. K. Fasseas, D. Tsikou, E. Fletmetakis, and P. Katinakis, “Molecular and biochemical analysis of the  $\beta$  class carbonic anhydrases in *Caenorhabditis elegans*,” *Molecular Biology Reports*, vol. 37, no. 6, pp. 2941–2950, 2010.
- [29] P. J. Linser, K. E. Smith, T. J. Seron, and M. N. Oviedo, “Carbonic anhydrases and anion transport in mosquito midgut pH regulation,” *Journal of Experimental Biology*, vol. 212, no. 11, pp. 1662–1671, 2009.
- [30] Y. Xu, L. Feng, P. D. Jeffrey, Y. Shi, and F. M. M. Morel, “Structure and metal exchange in the cadmium carbonic anhydrase of marine diatoms,” *Nature*, vol. 452, no. 6, pp. 56–62, 2008.
- [31] J. K. Colbourne, M. E. Pfrender, D. Gilbert et al., “The ecoresponsive genome of *Daphnia pulex*,” *Science*, vol. 331, no. 6017, pp. 555–561, 2011.
- [32] M. Lynch and J. S. Conery, “The evolutionary fate and consequences of duplicate genes,” *Science*, vol. 290, no. 5494, pp. 1151–1155, 2000.
- [33] J. Näsval, L. Sun, J. R. Roth, and D. I. Andersson, “Real-time evolution of new genes by innovation, amplification, and divergence,” *Science*, vol. 338, no. 6105, pp. 384–387, 2012.
- [34] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, “MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods,” *Molecular Biology and Evolution*, vol. 28, no. 10, pp. 2731–2739, 2011.
- [35] E. M. Gerts, Y.-K. Yu, R. Agarwala, A. A. Schäffer, and S. F. Altschul, “Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST,” *BMC Biology*, vol. 4, article 41, 2006.
- [36] M. A. Larkin, G. Blackshields, N. P. Brown et al., “Clustal W and Clustal X version 2.0,” *Bioinformatics*, vol. 23, no. 21, pp. 2947–2948, 2007.
- [37] S. Whelan and N. Goldman, “A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach,” *Molecular Biology and Evolution*, vol. 18, no. 5, pp. 691–699, 2001.
- [38] M. A. Miller, W. Pfeiffer, and T. Schwartz, “Creating the CIPRES Science Gateway for inference of large phylogenetic trees,” in *Proceedings of the Gateway Computing Environments Workshop (GCE ’10)*, pp. 1–8, New Orleans, La, USA, November 2010.
- [39] F. Ronquist and J. P. Huelsenbeck, “MrBayes 3: Bayesian phylogenetic inference under mixed models,” *Bioinformatics*, vol. 19, no. 12, pp. 1572–1574, 2003.
- [40] A. Stamatakis, “RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies,” *Bioinformatics*, vol. 30, no. 9, pp. 1312–1313, 2014.
- [41] A. Rambaut, *FigTree v1.3.1: Tree Figure Drawing Site*, FigTree Website, 2009.
- [42] A. Krogh, B. Larsson, G. von Heijne, and E. L. L. Sonnhammer, “Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes,” *Journal of Molecular Biology*, vol. 305, no. 3, pp. 567–580, 2001.
- [43] O. Emanuelsson, S. Brunak, G. von Heijne, and H. Nielsen, “Locating proteins in the cell using TargetP, SignalP and related tools,” *Nature Protocols*, vol. 2, no. 4, pp. 953–971, 2007.
- [44] N. Fankhauser and P. Mäser, “Identification of GPI anchor attachment signals by a Kohonen self-organizing map,” *Bioinformatics*, vol. 21, no. 9, pp. 1846–1852, 2005.
- [45] A. J. Esbaugh and B. L. Tufts, “The structure and function of carbonic anhydrase isozymes in the respiratory system of vertebrates,” *Respiratory Physiology & Neurobiology*, vol. 154, no. 1–2, pp. 185–198, 2006.
- [46] M. Hilvo, M. Tolvanen, A. Clark et al., “Characterization of CA XV, a new GPI-anchored form of carbonic anhydrase,” *Biochemical Journal*, vol. 392, no. 1, pp. 83–92, 2005.
- [47] K. S. Smith and J. G. Ferry, “Prokaryotic carbonic anhydrases,” *FEMS Microbiology Reviews*, vol. 24, no. 4, pp. 335–366, 2000.
- [48] S. Hoegg, H. Brinkmann, J. S. Taylor, and A. Meyer, “Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish,” *Journal of Molecular Evolution*, vol. 59, no. 2, pp. 190–203, 2004.
- [49] A. Meyer and Y. van de Peer, “From 2R to 3R: evidence for a fish-specific genome duplication (FSGD),” *BioEssays*, vol. 27, no. 9, pp. 937–945, 2005.
- [50] K. F. Rewitz and L. I. Gilbert, “*Daphnia* Halloween genes that encode cytochrome P450s mediating the synthesis of the arthropod molting hormone: evolutionary implications,” *BMC Evolutionary Biology*, vol. 8, no. 1, article 60, 8 pages, 2008.
- [51] F. A. Kondrashov, “Gene duplication as a mechanism of genomic adaptation to a changing environment,” *Proceedings of the Royal Society B: Biological Sciences*, vol. 279, no. 1749, pp. 5048–5057, 2012.
- [52] International Aphid Genomics Consortium, “Genome sequence of the pea aphid *Acyrtosiphon pisum*,” *PLoS Biology*, vol. 8, no. 2, Article ID e1000313, 24 pages, 2010.
- [53] J.-C. Simon, M. E. Pfrender, R. Tollrian, D. Tagu, and J. K. Colbourne, “Genomics of environmentally induced phenotypes in 2 extremely plastic arthropods,” *Journal of Heredity*, vol. 102, no. 5, pp. 512–525, 2011.
- [54] T. G. Evans, F. Chan, B. A. Menge, and G. E. Hofmann, “Transcriptomic responses to ocean acidification in larval sea urchins from a naturally variable pH environment,” *Molecular Ecology*, vol. 22, no. 6, pp. 1609–1625, 2013.