UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

IMPACT OF ENVIRONMENTAL (PH) CHANGE IN A MODEL AQUATIC HERBIVORE: FROM GENES TO POPULATIONS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

BILLY WAYNE CULVER, JR.
Norman, Oklahoma
2016

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/34612

IMPACT OF ENVIRONMENTAL (PH) CHANGE IN A MODEL AQUATIC HERBIVORE: FROM GENES TO POPULATIONS

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF BIOLOGY

BY

_____
Dr. Lawrence J. Weider, Chair

_____
Dr. K. David Hambright

_____
Dr. Ingo Schlupp

_____
Dr. Richard Broughton

_____
Dr. Robert Cichewicz

© Copyright by BILLY WAYNE CULVER, JR. 2016
All Rights Reserved.

Acknowledgements

I want to thank my parents, Billy Culver, Sr. and Julia Culver, and my wife Kristen Culver for their love and support. I also, want to thank my Ph.D. advisor Larry Weider for his long suffering and mentorship. Input and insight from my committee members: Dave Hambright, Ingo Schlupp, Richard Broughton, and Robert Cichewicz were invaluable. I would also like to acknowledge several people for their contribution to this dissertation: Philip Morton and Francisco Acosta for their collaboration; Ann Harris for all the molecular work; and my lab mates – Rachel Hartnett, Weston Speer, and Aime'e Yousey for their support with rearing *Daphnia* and setting up experiments. In addition, Krystal Gaylor's and Zach Myers' knowledge of qRT-PCR was instrumental for this dissertation. I am grateful to the College of Arts and Science (travel grants), Department of Biology (Adams Scholarship and Loren Hill Excellence in research grant), and Biological Station (Summer research scholarship) for their financial support. Further, this study was supported by the U.S. National Science Foundation (NSF-IOS-OEI) grants to L. J. Weider (#0924289 and #1256881).

Table of Contents

Acknowledgements iv
Abstract vii

Chapter One: Population genetic structure and fitness of *Daphnia pulicaria* across a pH gradient in three North American lakes 1

Abstract 2
Introduction 3
Methods 7
Results 14
Discussion 17
References 25
Tables and Figures 34

Chapter Two: The evolutionary history of Daphniid α-Carbonic Anhydrase within Animalia 39

Abstract 40
Introduction 41
Materials and Methods 43
Results and Discussion 48
Conclusions 56
References 61
Tables and Figures 68

Chapter Three: An investigation of three Carbonic Anhydrase isoforms in *Daphnia pulicaria* with signatures of local adaptation to acid-base regulation across a pH gradient 72

Abstract 73
Introduction 74
Results 77
Discussion 81
Materials and Methods 88
References 95
Tables and Figures 104

Chapter Four: Evolutionary response of Carbonic Anhydrase (isoforms 1, 2, and 5) in *Daphnia pulicaria* populations across a pH gradient I three North American lakes 108

Abstract 109
Introduction 110
Materials and Methods 113

Results 119
Discussion 123
References 131
Tables and Figures 137

Synthesis 143

Appendix A: Chapter One Supplemental Material 152
Appendix B: Chapter Two Supplemental Material 156
Appendix C: Chapter Three Supplemental Material 165
Appendix D: Chapter Four Supplemental Material 167

Abstract

Organisms are subjected to a variety of environmental stressors in which they must respond in order to survive and reproduce. While some individuals are able to adjust to these stressors and live to produce offspring and propagate their genes, others do not and are extirpated. Although it is known that organisms can respond to environmental stress, the underlying physiological and genetic mechanisms are often not well understood. Elucidating the evolutionary responses of organisms to environmental gradients is important, especially in light of increasing anthropogenic changes to our environment. In this dissertation, I looked at the acidification and alkalization of three North American lakes (Frenchman, Hill, Madison). In particular, I was interested in the underlying genetic response (evolution) of populations of the keystone aquatic zooplankter, *Daphnia pulicaria*, to the pH gradient found in these three lakes. In Chapter one, I used ecological genetic tools to determine local adaptation of the model organism, *D. pulicaria*, across a pH gradient in three North American lakes. I predicted there would be genetic differentiation and local adaptation among the three *Daphnia* populations. I genotyped individuals, which were used to determine genetic structure of the three populations. To test for signatures of local adaptation, a survivorship experiment across a pH gradient under common garden conditions was performed. In Chapter two, I was interested in determining candidate genes that may be involved in acid-base regulation in *D. pulicaria*. Previous studies have shown that carbonic anhydrases (CAs), a family of zinc metallo-enzymes, are responsible for acid-base regulation in many organisms. Through the use of phylogenetic tools, Chapter two attempted to find homologous CA isoforms in *Daphnia* that are implicated in acid-base

regulation in closely related aquatic taxa. In Chapter three, I characterized the three isoforms of α-CAs found in Chapter two (CA1, CA2, and CA5). In addition, under common garden conditions, I investigated the differential expression of those CAs from *D. pulicaria* clones isolated from three North American lakes that exhibit a pH gradient. Finally, in Chapter four, I investigated the processes which affect genetic variation: neutral processes (i.e. genetic drift) versus natural selection (i.e. positive, purifying selection). I predicted that there will be evidence of selection at variants of these three CA loci and that specific CA genotypes will convey a fitness advantage via differential survivorship across a pH gradient. Populations were analyzed using population genetic tools. Further, five distinct CA genotypes were chosen for a common garden pH survival experiment to determine differential survivorship across a pH gradient. In summary, I identified three CAs that were homologous to CAs found to be implicated in acid-base regulation in other aquatic organisms. These isoforms were well-conserved across taxa and I found evidence that CA1 was differentially-expressed across a pH gradient and that CA5 was always up-regulated in the Frenchman population regardless of pH. In addition, I found evidence that *D. pulicaria* populations were locally adapted to native pH conditions and that sequence variation in the three CA isoforms are implicated in adaptive responses to pH environment in these populations. While, this dissertation provides support that CAs are involved in acid-base regulation in *Daphnia*, further study is warranted. In particular, RNA-seq experiments could implicate additional genes that are involved in acid-base regulation. In addition, protein structure analysis and activity assays of the CA isoforms and their variants could provide additional evidence to their role in acid-base regulation

Chapter One

Formatted for *Hydrobiologia*

Population genetic structure and fitness of *Daphnia pulicaria* across a pH gradient in three North American lakes

Billy W. Culver

And

Francisco Acosta

Abstract

Understanding the evolutionary responses of organisms to environmental gradients is important in light of increasing anthropogenic changes to our environment. In this study, we used ecological genetic tools to determine local adaptation of the model organism, *Daphnia pulicaria*, across a pH gradient in three North American lakes. We predicted there would be genetic differentiation and local adaptation among the three *Daphnia* populations. To assess genetic differentiation, we genotyped individual *D. pulicaria* using 15 microsatellite loci across the three populations and performed a STRUCTURE analysis corroborated with Principal Coordinates Analysis based upon Nei's genetic distance and multiple F_{st} comparisons. To test for signatures of local adaptation, a survivorship experiment across a pH gradient under common garden conditions was performed. We determined that each of the three populations was genetically differentiated from one another, with Hill and Madison Lake populations of *D. pulicaria* being more similar to each other than that of the Frenchman Lake population. The results of the survivorship experiment showed a signal of local adaptation, with Frenchman Lake showing higher survivorship at lower pH when compared to Hill and Madison populations, while both Hill and Madison had higher survivorship at higher pH when compared to the Frenchman population.

Keywords: Ecological Genetics, Population Structure, Survivorship, Environmental Stressors, Local Adaptation.

Introduction

The key objective of ecological genetics is to determine the interactions between genotypes and their environment (G×E interactions; Ford 1975) and consequently how phenotypic traits evolve across a given environmental gradient within a given genetic background. There have been numerous studies that have attempted to elucidate G×E interactions in different model systems and environments. Previous research ranges from studies of plant phenotypic plasticity along an altitudinal gradient (Clausen et al., 1940) to laboratory studies with the model organism *Drosophila melanogaster* that examined mutagenic effects and G×E interaction in five different environmental stressors (Fry et al., 1996). In aquatic systems, the model organism *Daphnia pulex* has been used to assess the interactions between genomes and environmental parameters (Stollewerk, 2010; Colbourne et al., 2011; Miner et al., 2012), such as studies on the coexistence of food source quality and genotypes (Weider et al., 2005).

An important sub-theme in ecological genetics is the characterization of mechanisms of local adaptation. Typically, organisms have an optimum range of environmental tolerance, which can be either broad or narrow (Cox et al., 1976; Raleigh et al., 1980). However, some individual populations within a given species can have a fitness advantage relative to other populations within narrow bands of the species tolerance, i.e. being locally adapted to narrow ranges within the overall tolerance range (Byars et al., 2007). Using ecological genetic tools, a number of researchers have studied the mechanisms of local adaptation. Some well-known examples include pea aphids that are locally adapted to specific host plants and are influenced by population differentiation and genetic variation (Ferrari et al., 2008); population structure of the

common frog, *Rana temporaria*, in the United Kingdom where temperature and spawning date are correlated (Phillimore et al., 2010); and finally, the erosion of local adaptation in the Atlantic salmon, *Salmo salar*, via introgression of escaped farmed salmon which are maladapted to their new habitat (Bourret et al., 2011). While evolutionary responses have been well-studied across a variety of environmental gradients and taxa, further study is still necessary to elucidate the mechanism(s) by which organisms respond to changing environments especially in light of increasing anthropogenic impact to ecosystems.

Here, we look at the effect of acidification and alkalization (pH) of water bodies as an important ecological stressor that can act as a strong selective force on the genetic composition and physiological response of plankton populations. There are a number of mechanisms that can alter acidification and alkalization in water bodies naturally, such as bedrock leaching (Eppinger et al., 2007), catchment runoff (Erlandsson et al., 2011), and chemical conversion (Satake et al., 1995; Ezoe et al., 2002). However, the increasing impacts from anthropogenic sources such as carbon dioxide emissions (Moya et al., 2012, Evans et al., 2013), cultural eutrophication (O'Brien & deNoyelles, 1972), and mining activities (Derry, 2007; Martins et al., 2010) are of great concern with regards to lake acidification and alkalization as these processes can negatively affect the osmoregulation, growth, and health of aquatic organisms that use calcium carbonate to build shells or exoskeletons (Hurd et al., 2011).

In this study, we looked at three North American lakes (Frenchman Lake, Hill Lake, and Madison Lake) that have been shown to represent distinct pH environments. Frenchman Lake is an oligotrophic lake found in the Sudbury region of Ontario,

Canada. The lake is located on bedrock consisting of Canadian Shield granite with low pH buffering capacity (Wetzel, 2001). The Sudbury region is host to thousands of lakes that have been affected by metal smelting that has lowered the pH (into the pH = 4 – 5 range) of these lakes due to sulfur dioxide (SO_2) emissions in the 1960's & 70's (Keller 2004). In the 1970's, standards were put in place to reduce SO_2 emissions and a fair number of the lakes now have recovered into pH 6 – 7 range. Frenchman Lake was found to have a relatively low pH in relation to the other lakes in this study and falls within the pH range of those lakes found to be in recovery (www.greatersudbury.ca). Studies have shown that the natural zooplankton communities in these recovered lakes have returned as well (Yan et al., 1996a, Derry & Anrnott, 2007). In contrast, Hill Lake is a mesotrophic lake found in the forested region of north-central Minnesota consisting of organic bedrock that allows this lake to buffer relatively well and has not had as much anthropogenic impact as the other lakes in this study (Wetzel, 2001). The pH of Hill Lake was found to be close to neutral and serves as the control lake in this study (www.pca.state.mn.us). The third lake is Madison Lake and is located in the corn-belt region of Minnesota. This lake is eutrophic, due to heavy runoff from agricultural fertilizers. Typically, eutrophic lakes have higher pH due to elevated phosphate concentrations (Wetzel, 2001). The pH of Madison Lake was found to be relatively high in relation to the other lakes in this study (www.pca.state.mn.us).

The keystone zooplankton herbivore, *Daphnia pulicaria*, is an emerging model organism (Stollewerk, 2010; Colbourne et al., 2011; & Miner et al., 2012) that has a number of "good" model organism characteristics for ecological genetic studies as set out by Ford (1975): (1) ecologically well-studied (Loaring & Hebert, 1981; Dodson,

1988; Lampert, 1993, Urabe & Sterner, 2001); (2) can be easily sampled using plankton nets; (3) can produce multiple generations in a year, given that one daphniid can produce greater than 10 generations in its life time (~40 days) (Lampert, 2011); (4) multiple studies have shown that *D. pulicaria* populations harbor significant genetic variation that can be easily interpreted across an environmental gradient (Baird & Barata, 1998; Dudycha & Tessier, 1999; Coors et al., 2009); (5) typically, *D. pulicaria* have very large populations (i.e., in the millions, billions) (Hebert, 1978; Lampert, 2011); and (6) they can be easily maintained in a laboratory environment. Another aspect of *Daphnia* biology that make them well-suited for ecological genetic studies is that they exhibit both asexual (parthenogenetic) and sexual reproductive modes (cyclical parthenogenesis), which makes it easy to set up both clonal and sexual (recombinant) lineages.

The tolerance to a pH gradient in the genus *Daphnia* is well known, with *Daphnia* being found in the pH range of 6.5 - 10 depending on species (O'Brien & deNoyelles, 1972; Havens et al., 1993). However, local adaptation and genetic response to local variations in pH is poorly understood at the intraspecific-level. We investigated the ecological/evolutionary responses of, *D. pulicaria*, across a pH gradient observed in Frenchman, Hill, and Madison Lakes, using a combination of population genetic tools, long-term monitoring programs, and common-garden experiments. Given the geographic distances among the three lakes and diverse ecological backgrounds, we predict that there should be genetically distinct *D. pulicaria* populations in each lake. Since the three lakes possess distinct buffering capabilities and historical pH ranges, we predict that each population has evolved in a distinct pH background and will exhibit

differential among-population fitness (survivorship) or local adaptation across a pH gradient. Specifically, we predict that the Frenchman Lake population should have higher survivorship at lower pH than Hill and Madison Lake, and at the higher pH, the Madison Lake population should have higher survivorship than Frenchman and Hill Lake.

Methods

Study sites

Madison Lake (44^{o} 11.549' N; 93^{o} 48.740' W) and Hill Lake (47° 0.741' N; 93° 35.845' W) are located in Minnesota and are two of 24 lakes under long-term monitoring as part of the Minnesota Sentinel Lakes Program (MSLP) administered by the Minnesota Pollution Control Administration in conjunction with the Minnesota Department of Natural Resources. Annual average pH of these lakes was determined from data collected from the MSLP over the past decade. Frenchman Lake (46°43.081' N; 80° 59.298' W) is located in the Greater Sudbury region of Ontario, Canada and is monitored as part of the Greater Sudbury Water Quality Program (GSWQP). Annual average pH of Frenchman Lake was determined from data collected from the GSWQP over the past decade. The mean (± 1 SD) annual pH values for the three lakes were as follows: Frenchman Lake (6.53 ± 0.50), Hill Lake (7.91 ± 0.52), and Madison Lake (8.63 ± 0.25) (Table1).

Daphnia collection

Two methods were used to collect *Daphnia* in these lakes: (1) we used a 158 uM mesh Wisconsin plankton net to take vertical tows through the full water column at the deepest part of the lakes in July 2010, 2011 and 2014 for Madison (n = 77) and Hill Lake (n =88) and in June 2011, 2013, and 2014 for Frenchman Lake (n = 67); (2) for Madison and Hill Lakes, we obtained and hatched animals (n = 15 and 5, respectively) from diapausing (ephippial) eggs harvested from core samples taken in July 2010 and 2011 (Table 1). The cores of Madison and Hill Lake were taken in duplicate using a 1.5m (6.93cm diameter) single drive Griffith sediment corer with Livingstone drive rods. One core was sliced into 4cm or 2cm sections for Madison and Hill, respectively. Core sections were sifted through a series of sieves of 710μm, 425μm, and 300μm mesh sizes to collect *D. pulicaria* ephippia. Individual eggs were removed from their ephippial casings and placed in COMBO medium (Kilham et al., 1998) for hatching. For unhatched eggs, we extracted DNA (see below). Eggs were hatched under the following conditions; eggs were stored in the dark at 4^{o}C for two weeks and then placed under direct lighting at 20^{o}C for 24 hours until hatched. Once eggs hatched, they were reared at 20°C with indirect lighting. The second core was sliced into 1cm sections and was used for ^{210}Pb dating at the St. Croix Watershed Research Station, MN using standard procedures (Engstrom & Schottler 2003). The eggs from Hill Lake were collected from the 0-2cm section of the sediment cores and were dated at AD 2006.7 ± 1.6 years at the bottom of the core section, while the eggs from Madison Lake were collected from the 0-4cm section and were dated at AD 2009.8 ± 1.3 years at the bottom of the core section. Further, another (n = 17) eggs were collected from sediment cores

in Hill Lake from which DNA was extracted. The final total sample sizes used in the population genetic analysis for the three lake populations were Frenchman Lake (n = 67), Hill Lake (n = 110), and Madison Lake (n = 92) (Table 1). See Frisch et al. (2014) for more details related to sediment core sample methods and analyses.

DNA extraction and amplification

DNA was extracted from either eggs that were removed from their ephippial casings, or adult *D. pulicaria*. DNA was extracted from eggs using the HotShot method (Montero-Pau et al., 2008) while the CTAB method (Hillis et al., 1990) was used on adult animals. Seventeen microsatellite loci were used to characterize genotypes; however two loci (Dp311 and Dp377) did not consistently amplify and were left out of the final analysis. The microsatellite loci used for this analysis are listed in Table S1, and were distributed throughout the entire genome (Frisch et al. 2014). Microsatellite loci were amplified in a single, 25μL multiplex reaction (type-it PCR kit, Qiagen Inc, Valencia, CA, USA) using a MJ research PTC-200 thermocycler at the manufacturer's recommended settings. Amplified microsatellites were genotyped using STRand 2.4.59 (Toonen & Hughes, 2001).

Population genetic structure

The genetic structure of *D. pulicaria* was inferred for the three lake populations using the 15 microsatellite loci within a clustering method utilizing a Bayesian framework,

executed in STRUCTURE 2.3.4 (Pritchard et al., 2000). Posterior likelihood values were computed for $K = 1$ to $K = 7$. These values represent the number of assumed ancestral genetic clusters from which the individuals analyzed were descended. The following parameters were used for the STRUCTURE analysis: (1) no admixture between populations, (2) allele frequencies are correlated, and (3) LOCPRIOR model (*a priori* population identity). A total of 10 simulations with a burn-in of 50,000 Markov Chain Monte Carlo (MCMC) iterations and 100,000 iterations after the initial burn-in for each *K* value were run. To elucidate the most likely *K* value, the K value with the greatest rate of change in the likelihood function (ΔK) was determined using the online software STRUCTURE HARVESTER (Evanno et al., 2005; Earl & von Holt, 2011). The most probable value of *K* as identified by STRUCTURE HARVESTER for data of 10 successive runs using this K value was merged using CLUMPP 1.1.2 (Jakobsson & Rosenberg, 2007) and the results were visualized using DISTRUCT 1.1 (Figure 1; Rosenberg, 2004).

To corroborate the results from STRUCTURE, a Principal Coordinate Analysis (PCA) was performed using Nei's genetic distance within GenAlEx 6.41 (Peakall & Smouse, 2006). To test for genetic differentiation of the three *D. pulicaria* populations, an AMOVA with 999 permutations was run on the raw genotypic data using GenAlEx 6.41.

To test for variation within populations using DNA from eggs (Hill Lake only), individuals hatched from eggs (referred to as hatchlings), individuals taken from the water column (referred to as lake samples) and individuals sampled in different years, an AMOVA with 999 permutations was run to compare pairwise F_{st} values for each

variable. The following comparisons were made within each population: Hill Lake – (1) eggs, hatchlings and lake samples and (2) year living (2006, 2011, 2013, 2014); Madison Lake – (1) hatchlings and lake samples and (2) year living (2010, 2011, and 2014); and Frenchman Lakes – year living (2011, 2013, and 2014).

Population genetic analysis

Allelic frequencies, number of alleles (N_a), private allele frequency (A_p), percentage of polymorphic loci, Shannon's Information Index (I), observed heterozygosity (H_o) and expected heterozygosity (H_e) for each *D. pulicaria* population were calculated within GenAlEx 6.41. Departure from Hardy-Weinberg Equilibrium (HWE) was tested for each locus for each of the three *D. pulicaria* populations using GenAlEx 6.41.

To evaluate the three *D. pulicaria* populations for recent genetic bottlenecks, a M-ratio test was performed. The M-ratio test compares the number of alleles to the range in allele size (*M*) (Garza & Williamson, 2001). It is suggested by Garza and Williamson (2001) that the number of alleles will rapidly decrease in relation to the allele size range and that the allele recovery is correlated with post-population size allowing for the identification of a bottleneck(s) long after it has occurred. A *M* –value > 0.68 suggests a population is in equilibrium and should approach 1.0. Using seven or more loci is recommended by the authors.

Survivorship experiment

To test for differential fitness of *D. pulicaria* from the three lakes, a survivorship experiment was performed. Four clones from each population (lake) were randomly selected from an established clone bank. For Hill Lake and Madison Lake, individual clones used for the experiment were initially established from hatched eggs that were recovered from sediment cores specifically from the Hill Lake 0-2cm core segment and Madison Lake 0-4cm core segment. We used the animals from the shallower portion of the cores, because they represent more recent populations and were from the population that experienced the pH regimes used in this study. For Frenchman Lake, clones were raised from animals initially harvested from the water column in 2011 (four years older than the oldest dated eggs from Hill and Madison Lake). During this five year period, the pH values of the individual lakes were relatively stable (see *Study Sites* section). All animals were raised under the same laboratory conditions using COMBO medium (~ pH = 7.5).

For each clone, a set of 40 gravid animals were used to set up as stem grandmothers. Once the stem grandmothers had released their clutches, 40 neonates from each clone were harvested to serve as stem mothers. All animals were fed 0.5 mL of approximately 1 – 2 mg C L^{-1} of chemostatically-cultured green algae, *Scenedesmus acutus*, daily to ensure that the *Daphnia* were not food limited. Neonates from the third clutch of the stem mothers were used for the experiment since previous work has shown that third-clutch offspring are more robust than first or second-clutch neonates (Glazier, 1992). In addition, using the grand-daughter generation minimizes maternal effects (Lampert 2011).

In a common-garden experimental design, five treatments of COMBO medium were adjusted, using 0.5M NaOH or 0.5M HCl (coarse adjustment), to the following pH: 6.0, 6.5, 7.5, 8.5, and 9.0. Five neonates from each clone were placed in 50mL of pH-adjusted COMBO medium per treatment. For each clone, five replicate 50mL jars were used per treatment. The pH of the treatments was measured twice a day, with a Milwaukee Instruments PH56 pH meter (the pH meter was calibrated before taking measurements), and if necessary adjusted with 0.05M NaOH or 0.05M HCl (fine adjustment). The experiment was terminated after 96-hour, prior to the *Daphnia* reaching maturity (i.e., becoming reproductive). During the experiment, animals were fed daily with 1mg C L^{-1} of chemostatically cultured green algae, *S. acutus*. When testing and adjusting pH, these values were measured after the feeding. The number of surviving animals in each replicate was recorded at the termination of the experiment.

Statistical analysis

Statistical analyses of the survivorship experiment were performed using the software R v2.15.0 (R Core Team, 2012), using the packages *lme4* v1.1-10 (Bates et al., 2014) and *lmerTest* v2.0-29 (Kuznetsova et al., 2015). Data for each organism in the experiment (1500 individuals in total) were fitted to a generalized linear mixed-effects model using the Laplace approximation, with a binomial distribution. The full model contained survivorship (dead or alive) as the response variable, pH treatment, lake of origin and their interactions as fixed factors, and the natural sources of variation (different clones within each lake and the jars in which the individuals were contained during the

experiment) were incorporated as random factors in the model. Over-dispersion was checked by comparing deviance of the model with the residual degrees of freedom.

Full and reduced models were compared based on their Akaike Information Criterion (AIC). The two best models were a model incorporating just the fixed factors and their interaction, and a model incorporating these and the jar as a random factor. We used the latter one in order to account for the potential small pH variations that might have taken place within each jar during the experiment.

Effects of lake on the survivorship of individuals under different pH conditions were calculated with the package *effects* v3.0-5 (Fox, 2003), using specific reduced models incorporating only one pH treatment and averaging over all other terms in the model.

Results

Population genetic analysis

The 15 microsatellite loci used in our final analysis were distributed on at least 6 of 12 chromosomes (Frisch *et al*., 2014). Two of the 15 loci (Dp 375 and 376) were found to be monomorphic across all three populations, while the remaining 13 polymorphic loci had allelic richness values ranging from 2 to 6 alleles per locus. On average, Frenchman Lake had a lower number of alleles per locus (1.33 ± 0.13), while both Hill and Madison Lake had higher numbers of alleles per locus (2.87 ± 0.26 and 2.93 ± 0.33), respectively) (Table 1, Figure S1). Further, a higher allelic diversity for Hill and Madison populations was indicated with each lake population having a significantly

higher *I* (Shannon Information Index, 0.442 ± 0.10 and 0.564 ± 0.11) respectively, than Frenchman Lake (*I* = 0.231 ± 0.09) (Table 1). Significant departures from Hardy Weinberg Equilibrium (HWE) were observed at 33.3% of the loci in Frenchman Lake, 66.7% of the loci in Hill Lake, and 40.0% of the loci in Madison Lake (Table S1). Private allele frequency (those alleles that are found only in a particular population), A_p, was observed to be low in Frenchman Lake (0.067 ± 0.70), while both Hill and Madison Lakes had significantly higher numbers of private alleles (0.600 ± 0.13 and 0.667 ± 0.19 respectively) (Table 1, S2). Further, observed heterozygosity (H_o) in Frenchman Lake (0.333 ± 0.13) was higher than expected heterozygosity (H_e; 0.167 ± 0.06; see *Discussion*), while in Hill and Madison Lakes, the H_o (0.239 ± 0.06 and 0.323 ± 0.07, respectively) were not significantly different than H_e (0.254 ± 0.06 and 0.313 ± 0.06, respectively). Across the three lakes, Frenchman Lake had a higher H_o than Hill and Madison Lakes (Table 1). A M-Ratio test was performed to test for bottlenecks within populations with all populations having an *M-value* greater than 0.68 (Table 1). However, the results for Frenchman Lake should be viewed with caution, since only five of the 15 loci were polymorphic, and thus, results could be biased due to this reduced genetic variability (as noted above, it is recommended that a minimum of seven loci be included in the M-Ratio test; Garza & Williamson, 2001).

The most likely number of genetic clusters using the Bayesian clustering analysis in STRUCTURE was found to be three (*K=3*), one cluster for each population, as determined by the *ΔK* method in STRUCTURE HARVESTER (Evanno et al., 2005; Earl & von Holt, 2011; Figure 1). The PCA based upon Nei's genetic distance provided additional support of three distinct populations that corresponded to the three lakes,

thus, corroborating the results of the STRUCTURE analysis (Figure 2). Additionally, a pairwise comparison of F_{st} values indicated that each population was significantly genetically different from each other: Frenchman vs Hill (AMOVA, F_{st} = 0.355, P < 0.001), Frenchman vs Madison (AMOVA, F_{st} = 0.428, P < 0.001), and Hill vs Madison (AMOVA, F_{st} = 0.245, P < 0.001). Pairwise comparisons within populations between years, and among eggs, hatchlings, and lake samples were not significantly different from each other (data not shown).

Fitness response to a pH gradient

As can be seen (Figure 3), at moderate pH values (7.5) survivorship was similar across all three populations. At extreme mean pH values, survivorship dropped dramatically (59% and 78% for pHs 6.0 and 9.0, respectively), although the effects at low pH were less severe for individuals from Lake Frenchman, with a survivorship of 70% at pH 6.

We evaluated these results using a generalized linear mixed-effects model to test the main effects (pH treatments and lake of origin), accounting for natural sources of variation (the different clones within each lake, and the individual jars which were used for the experiment).

Model comparison showed that the best model incorporated all the main effects and the variation induced by the individual jars; clones from within each lake did not differ significantly from one another (Figure S2), and models that did not incorporate them had the same likelihood and deviance that those who did (Table 2). The effect of pH was significant (Wald's test, z = 4.679, p < 0.001); Frenchman differed from the

other two populations (Wald's test, $z = 4.11$, $p < 0.001$), while Hill Lake did not differ from Madison Lake (Wald's test, $z = -0.467$, $p = 0.641$). In terms of interactions, only the interaction between pH treatment and Frenchman Lake was significant (Wald's test, $z = -4.072$, $p < 0.001$). A plot of the partial effects of lake origin in probability of survivorship under different pH treatments (Figure S3) corroborates our initial prediction of populations of Frenchman Lake being specially adapted, and apt to survive, at low pHs.

Discussion

The analyses performed in this study allowed us to explore how fitness and genetic composition in an aquatic organism, *D. pulicaria*, corresponds to local adaptation across a pH gradient in three lakes with contrasting ecological histories. Frenchman Lake, as predicted, had the lowest pH (Table 1). Like the other lakes in the Sudbury region, Frenchman seems to have been affected by SO_2 emissions due to historical smelting in the Sudbury region of Canada (Keller, 2004). In the 1960's, these lakes were acidified and many lakes typically had pH values in the 4-5 range. After regulation of the smelting industry in the 1970's, these lakes had begun to recover, so that now typical pH values range from 5-7. Frenchman Lake, while not a part of the hundreds of lakes that have been studied historically (Yan et al., 1996a; Keller, 2004; Derry & Anrnott, 2007), seems to have a pH history that has mirrored the other lakes in the region. On the other end of the spectrum, Madison Lake is located in the corn-belt region of Minnesota and is influenced heavily by fertilizer runoff (MSLP, www.pca.state.mn.us). Agricultural fertilizers, which have high concentrations of both nitrogen and

phosphorous also serve as major allochtonous nutrient sources in these lakes that can lead to massive algal blooms, resulting in deterioration of water quality (Verhoeven, 2006). Further, nitrogen (NO_3) and phosphorous (PO_4) inputs to aquatic environments have a tendency to cause lakes to become alkaline, which leads to higher pHs (O'Brien & DeNovelles, 1972; Schindler et al., 1985). Another effect of highly productive lakes is that algae respire at night, thus releasing CO_2 into the water column, which results in the production of carbonic acid, thus lowering the pH at night. However, during the day when the algae are photosynthesizing, the CO_2 is scrubbed from the water column, resulting in an increase in pH (Wetzel, 2001). So, in these highly eutrophic lakes not only do organisms need to deal with higher overall pHs, they need to deal with large diel fluctuations in pH values. Madison Lake, as predicted, also has a relatively high pH, but there are no available data on pH for Madison Lake to estimate how large the daily fluctuations in pH may be. Hill Lake, which is located in the northern forest region in Minnesota, seems to have had little anthropogenic impact (MSLP, www.pca.state.mn.us). This lake is mesotrophic, mainly due to its watershed and bedrock (Wetzel, 2001). Hill Lake, in terms of pH, is relatively neutral and serves as the control lake in this study.

In addition to the ecological gradients among these three lakes, physical distance and geological history likely play a role in influencing population genetic structure. Both Madison and Hill Lakes are ~500 km from each other, while Madison and Hill are greater than 1000 km from Frenchman Lake. Additionally, Madison and Hill were formed ~1000 years earlier than Frenchman due to retreating glaciers at the end of the Pleistocene (Dyke & Prest, 1987). Our genetic analysis did show that each lake had

distinct genetic structure (Figure 1 & 2), although, Hill and Madison Lakes were more genetically similar to each other (Figure 2), when compared with Frenchman Lake.

When looking at indicators of genetic diversity within each lake, again Hill and Madison Lake are similar; this can be explained by their relatively close geographic distance (Nei, 1972, Maruyama & Slatkin, 1975). Frenchman Lake, on the other hand, was formed about ~8000 years ago and has very low genetic diversity when compared to either Hill Lake or Madison Lake. In fact, in the analysis of the 67 individuals from Frenchman Lake, all shared the same multi-locus (microsatellite) genotype. One possible explanation for this greatly reduced genetic diversity is the smelting in this area, which reduced the lake pH to 4.0-5.0 (Pollard et al., 2003, Yan et al., 1996a; Derry & Anrnott, 2007). This could have caused a population bottleneck from which the species has not recovered. However, our M-ratio test revealed no bottleneck in Frenchman Lake. It should be noted that this analysis has lost power because only 5 of the 15 loci examined were polymorphic (which is lower than the minimum recommended seven loci for the M-ratio test; Garza & Williamson, 2001). Further, data (K. Milette, personal communication) suggests that the *Daphnia* from Frenchman Lake, are obligatory parthenogenetic hybrids between *D. pulicaria* and its closely-related sister taxon, *D. pulex*. This suggests that there has potentially been clonal selection for this specific genotype in Frenchman Lake. One can then speculate that the original *Daphnia* populations, which cannot tolerate pH values lower than 6.0, were extirpated from the lake. When the pH recovered, the lake was either recolonized from the egg bank in the sediment or colonized through invasion by this hybrid complex and clonal selection resulted in this 'superclone'. A similar phenomenon took place in Hannah

Lake, a lake recovering from acidification and metal contamination due to smelting in the Sudbury Region, which had its historically native species of *D. mendotae* and *D. ambigua* extirpated. These species were later replaced by *D. pulicaria*, after the pH decreased to near 6.0 (Pollard, 2003). In addition, researchers have shown invasions of obligately parthenogenetic hybrid *D. pulex* × *D. pulicaria* in African lakes with low genetic diversity in their native *D. pulex* sexual populations, which have been shown to be susceptible to total replacement by 'superclones'(Mergeay, Verschuren, & DeMeester, 2006). Of course, more extensive sampling of the extant population and the ephippial egg-bank in Frenchman Lake is needed to provide further support for this supposition.

Multiple studies have looked at the effects of environmental factors in the genetic differentiation of populations. For instance, the genetic structure among populations of maritime pines, *Pinus pinaster* in Europe was found to be influenced across a precipitation and thermal gradient with a signature of local adaptation for drought resistance (Eveno et al., 2008). Another study evaluated how the effects of thermal shock determined differences in evolutionary response in the heat shock protein Hsp70 through local adaptation over a thermal gradient in the Colorado potato beetle, *Leptinotarsa decemlineata*, (Lyytinen et al., 2012). These processes also hold true in *Daphnia*. Pantel et al. (2011) showed that nutrient gradients and predator densities in the asexual *D. pulex* × *D. pulicaria* hybrid complex can structure the populations. Specifically, pH gradients have been shown to genetically structure populations of *Daphnia*. Two particular studies have shown (1) that acid mine drainage can change the genetic diversity and structure of populations of *D. longispina* in Portugal (Martins

et al., 2009) and (2) a pH gradient across ponds in the Hungarian Great Plains contributed to ~80% of the genetic variability in populations of *D. atkinsoni-bolivari* (Nédli & Forró, 2013). While geological history and geographic distance certainly cannot be ignored when looking at the genetic structure of the three lakes in this study, one also cannot ignore the influence of environmental factors in shaping the genetic structure of populations. Eutrophication and pH have likely contributed to the genetic differentiation among Frenchman, Hill and Madison Lakes.

The results of our survivorship experiment indicated that there were significant differences among (pH) treatments; however there were no significant differences among clones within the populations. This suggests that all the clones studied within a population are equally adapted to their native pH conditions. Although the clones studied have different microsatellite genotypes, (except for the Frenchman Lake population), there is potentially evolutionary responses (i.e., effects on the mechanisms for pH homeostasis) unique to each population. The mechanism for pH homeostasis in *D. pulicaria* is the focus of ongoing research. As an organism deviates from the optimum point within its ecological range of tolerance for a given environmental stressor, one would expect that mortality would increase. In this case, all three lakes have an optimum pH of ~ 7.5, and presented significant differences in their response to pH extremes, particularly at the lower pH values. When looking at among population effects, there was an important signal of differential fitness, especially in Frenchman Lake. At pH = 6.0, Frenchman Lake *D. pulicaria* had a higher survivorship than Hill or Madison Lake, while at pH = 9.0, Frenchman had a lower survivorship than Hill or Madison. These results support the notion that Frenchman Lake *D. pulicaria* are locally

adapted to a lower pH environment relative to the Hill and Madison populations. Hill and Madison Lake have a more similar response to each other, but tended to have higher survivorships in the higher pH treatments

Local adaptation to pH is well supported in aquatic environments. In marine environments, which have been affected by acidification due to elevated atmospheric CO_2, it has been shown that some populations of the purple sea urchin (*Stromgylocentrous purpuratus*) in the northeast Pacific ocean have evolved a tolerance to lower pH environments by modification of their expression patterns during skeletogenesis (Evans et al., 2013). Another study (Weisse et al., 2007) measured fitness effects across a pH gradient using the aquatic asexual ciliate, *Meseres corlissi*, and found evidence of local adaptation to their native pH environment. While *Daphnia* seem to be well-studied in regard to local adaptation across nutrient (eutrophication) (DeClerck et al., 2001) and predator density gradients (DeMeester et al., 1999), there is a dearth of knowledge when it comes to local adaptation to pH (although see examples cited above).

Our current study in this area is focused on the mechanisms of local adaptation among these populations of *D. pulicaria* using candidate genes. One particular enzyme, carbonic anhydrase (CA), is of particular interest since it is responsible for acid-base regulation in many aquatic species. We have found three isozymes of CA that are homologous to acid-base regulating CAs in other crustaceans (Culver & Morton, 2015). In addition, work looking at differential regulation and signals of positive selection in these genes across a pH gradient is in progress. The use of ecological genetic tools in

this study to explain local adaptation is particularly informative in clarifying the evolutionary response to a pH gradient.

Considering the anthropogenic alteration of pH in aquatic environments via increased land use for agriculture, CO_2 emissions, and industrial emissions, it is important to understand the adaptive capabilities of organisms in response to these environment perturbations. Of particular importance in our modern-day world of globally shifting climatic conditions, is to examine and try to understand the rate(s) of adaptation or extinction. Here, we observed that there was a marked difference in fitness (local adaptation) across a pH gradient, specifically in Frenchman Lake. The fitness advantage is associated with the respective genetic differentiation among Frenchman, Hill, and Madison Lakes. Considering the current trend of increasing anthropogenic eutrophication due to fertilizer runoff from increased land use for agriculture (Dodds, 2009), one could expect that more lakes will become more eutrophic, followed by an increase in pH from the NO_3 and PO_4 inputs. Another important factor is the rising anthropogenic input of CO_2 or SO_2 to the atmosphere (Solomon et al., 2007), particularly in East Asia (Lu et al., 2010), which lowers the pH of aquatic systems as these compounds are sequestered in water bodies. Elucidating the mechanisms for the evolutionary response of organisms to changing pH conditions is therefore of the utmost importance, in order to predict the impact of these widespread changes in aquatic ecosystems.

Acknowledgments

This study was supported by the U.S. National Science Foundation (NSF-IOS-OEI) grants to L. J. Weider (#0924289 and #1256881), the U. of Oklahoma Department of Biology Adams Scholarship to B.W Culver and the U. of Oklahoma Biological Station Summer Research Grant to B.W. Culver. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We thank the U. of Minnesota LacCORE staff (in particular R. O'Grady) at the University of Minnesota, for assistance with coring and providing *D. pulicaria* samples from Madison Lake, MN. A. Harris, J. Medders and F. Speer facilitated genotyping of *D. pulicaria* and R. Hartnett for logistical support with the survivorship experiment. We thank K. Milette (McGill University) and M. Edlund (St. Croix River Watershed Station, MN) for providing *Daphnia* samples. We also thank L.J. Weider for his advice, mentorship, and comments on this manuscript. In addition, we thank R. Broughton, K.D. Hambright, I. Schlupp, R. Cichewicz, J. Beyer, T. Hallidayschult, B. Tweedy, P. Morton, A. Yousey for their input. This manuscript constitutes a portion of BWC's doctoral dissertation at the U. of Oklahoma. There are no known conflicts of interests associated with this article and its authors.

References

Baird, D.J. & Barata, C., 1998 Genetic variation in the response of *Daphnia* to toxic substances: Implications for risk assessment. In Genetics and Ecotoxicology. Taylor & Francis, Inc, Ann Arbor, MI. 207-218.

Bourret, V., O'Reilly, P.T., Carr, J.W., Berg, P.R., & Bernatchez, L., 2011. Temporal change in genetic integrity suggests loss of local adaptation in a wild Atlantic salmon (*Salmo salar*) population following introgression by farmed escapees. Heredity 106: 500-510.

Byars, S.G., Papst, W., & Hoffmann, A.A., 2007. Local adaptation and cogradient selection in the alpine plant, *Poa hiemata*, along a narrow altitudinal gradient. Evolution 61: 2925-2941.

Colbourne, J.K., Pfrender, M.E., Gilbert, D., Thomas, W.K., Tucker, A., Oakley, T.H., et al., 2011. The Ecoresponsive Genome of *Daphnia pulex*. Science 331: 555-561.

Coors, A., Vanoverbeke, J., DeBie, T., & deMeester, L., 2009. Land use, genetic diversity and toxicant tolerance in natural populations of *Daphnia magna*. Aquatic Toxicology 95: 71-79.

Cox, B.C., Healy, I.N., & Moore, P.D., 1976. Biogeogrpahy: an Ecological and Evolutionary Approach 2nd ed. Blackwell Science Publication, Oxford.

Clausen, J., Keck, D.D., & Hiesey, W.M., 1940. Experimental studies on the nature of species. I. The effects of varied environments on western American plants. Carnegie Inst. Wash. Publ 520.

Culver, B.W., & Morton, P.K., 2015. The evolutionary history of Daphniid α-carbonic

anhydrase within Animalia. International Journal of Evolutionary Biology 2015: 1-11.

Dodds, W.K., Bouska, W.W., Eitzmann, J.L., Pilger, T.J., Pitts, K.L., Riley, A.J., et al., 2009. Eutrophication of U.S. Freshwaters: Analysis of Potential Economic Damages. Environmental Science & Technology 43: 12-19.

DeClerck, S., Cousyn, C., & DeMeester, L., 2001. Evidence for local adaptation in neighbouring *Daphnia* populations: a laboratory transplant experiment. Journal of Freshwater Biology 46: 187-198.

DeMeester, L., Boersma, M., & Spaak, P., 1999. Environmental stress and local adaptation in *Daphnia magna*. Limonology & Oceanography 44: 393-402.

Derry, A.M., & Arnott, S.E., 2007. Adaptive reversals in acid tolerance in copepods from lakes recovering from historical stress. Ecological Applications 17: 1116-1126.

Dodson, S., 1988. The ecological role of chemical stimuli for zooplankton: Predator-avoidance behavior in *Daphnia*. Limnology & Oceanography 33: 1431-1439.

Dudycha, J.L. & Tessier, A.J., 1999. Natural generic variation of life span, reproduction, and juvenile growth in *Daphnia*. Evolution 53: 1744-1756.

Dyke, A.S. & Prest, V.K., 1987. Late Wisconsinan and Holocene of the Laurentide Ice Sheet. Géographie physique et Quatenaire 41: 237-263.

Ear,l D.A. & von Holt, B.M., 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4: 359-361.

Engstrom, D.R. & Schottler, S., 2003. Dating methods for lake sediments and peat. St

Croix Watershed Research Station, Science Museum of Minnesota, special publication.

Eppinger, R.G., Briggs, P.H., Dusel-Bacon, C., Giles, S.A., Gough L.P., Hammarstrom, J.M., & Hubbard, B.E., 2007. Environmental geochemistry at Red Mountain, an unmined volcanogenic massive sulphide deposit in the Bonnifield district, Alaska Range, east-central Alaska. Geochemistry: Exploration, Environment, Analysis 7: 207-223.

Erlandsson, M., Cory, N., Fölster, J., Köhler, S., Laudon, H., Weyhenmeyer, G.A., & Bishop, K., 2011. Increasing dissolved organic carbon redefines the extent of surface water acidification and helps resolve a classic controversy. BioScience 61: 614-618.

Evanno, G., Regnaut, S. & Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation strudy. Molecular Ecology 14: 2611-2620.

Evans ,T.G., Chan, F., Menge, B.A., & Hofman, G.E., 2013, Transcriptomic responses to ocean acidification in larval sea urchins from a naturally variable pH environment. Molecular Ecology: 22, 1609-1625.

Eveno, E., Collada, C., Guervara, M.A., Léger, P., Ganzález-Martínez, S.C., Cervera, M.T., 2008.Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic differentiation analysis. Molecular Biology & Evolution 25: 417-437.

Ezoe, Y., Lin, C.H., Noto, M., Watanabe, Y., & Yoshimura, K., 2002. Evolution of

water chemistry in natural acidic environments in Yangmingshan, Taiwan. Journal of Environmental Monitoring 4: 533-540.

Ferrari, J., Via, S., & Godfrey, H.C.J., 2008: Population differentiation and genetic variation in performance on eight hosts in the pea aphid complex. Evolution 62: 2508-2524.

Ford, E.B., 1975. Ecological Genetics. Springer, Netherlands.

Frisch, D., Morton, P.K., Priyanka, R.C., Culver, B.W., Colbourne, J.K., Weider, L.J., et al., 2014. A millennial-scale chronicle of evolutionary responses to cultural eutrophication in *Daphnia*. Ecology Letters 17**:** 360-368.

Fry, J.D., Heinsohn, S.L., & MacKay, T.F., 1996. The contribution to genotype-environment interaction for fitness in *Drosophila melanogaster*. Evolution 50: 2316-2327.

Glazier, D.S., 1992. Effects of food, genotype, and maternal size and age of offspring investment in *Daphnia magna*. Ecology 73: 910-926.

Garza, J.C. & Williamson, E.G., 2001. Detection of reduction in population size using data from microsatellite loci. Molecular Ecology 10: 305-318.

Havens, K.E., Yan, N.D., & Keller, W., 1993. Lake acidification: effects on crustacean zooplankton populations. Environmental Science Technology 27: 1621-1624.

Hebert, P.D.N., 1978. The population biology of *Daphnia* (Crustacea, Daphnidae). Biological Reviews 53: 387-426.

Hillis, D.M., Moritz, C., & Mable, B.K., 1990. Molecular Systematics. Sinauer & Associates, Sunderland, MA, USA.

Hurd, C.L., Cornwall, C.E., Currie, K., Hepburn, C.D., McGraw, C.M., Hunter, K.A., &

Boyd, P.W., 2011. Metabolically induced pH fluctuations by some coastal calcifiers exceed projected 22nd century ocean acidification: a mechanism for differential susceptibility? Global Change Biology 17: 3254-3262.

Jakobsson, M. & Rosenberg, N.A., 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801-1806

Keller, W., Heneberry, J., Gunn, J.M., Snucins, E., Morgan, G., & Leduc, J., 2004. Recovery of Acid and Metal – Damaged Lakes Near Sudbury Ontario: Trends and Status. Laurentian University, Sudbury, Ontario, www3.laurentian.ca.

Kilham, S.S., Kreeger, D.A., Lynn, S.G., Goulden, C.E. & Herrara, L,. 1998. COMBO: a defined freshwater culture medium for algae and zooplankton. Hydrobiologia 377: 147-149.

Lampert, W., 1993. Phenotypic plasticity of the size at first reproduction in *Daphnia*: The importance of maternal size. Ecology 74: 1455-1466).

Lampert, W., 2011. Daphnia: development of a model organism in ecology and evolution. Oldendorf/Luhe: International Ecological Institute.

Loaring, J.M. & Hebert, P.D.N., 1981. Ecological differences among clones of *D. pulex* Leydig. Oecologia 51: 162-168.

Lu, Z., Streets, D.G., Zhang, Q., Wang, S., Carmichael, G.R., Cheng, F., et al., 2010. Sulfur dioxide emissions in China and sulfur trends in East Asia since 2000. Atmospheric Chemistry & Physics 10: 6311-6331.

Lyytinen, A., Mappes, J., & Linström, L., 2012. Variation in hsp70 levels after cold

shock: signs of evolutionary repsonses to thermal selection among *Leptinotarsa decemlineata* Populations. PLoS one 7: 1-4.

Martins, C.M.G., Almeida, D.V., Marins, L.F.F., Bianchini, A., 2010. mRNA expression and activity of ion-transporting proteins in gills of the blue crab *Callinectes sapidus*: effects of waterborne copper. Environmental Toxicology and Chemistry 30: 206-211.

Martins, N., Bollinger, C., Harper, R.M., & Ribeiro, R., 2009. Effects of acid mine drainage on the genetic diversity and structure of a natural population of *Daphnia longispina*. Aquatic Toxicology 92: 104-112.

Maruyama, T. & Slatkin, M., 1975. The influence of gene flow on genetic distance. The American Naturalist 109: 597-601.

Mergeay, J., Verschuren, D., & DeMeester, L., 2006. Invasion of an asexual American water flea clone throughout Africa and rapid displacement of a native sibling species. Proceeding of the Royal Society B 273: 2839-2844.

Miner, B.E., DeMeester, L., Pfrender, M.E., Lampert, W., & Hairston, Jr., N.G., 2012. Linking genes to communities and ecosystems: *Daphnia* as an ecogenomic model. Proceedings of the Royal Society B, 1-10.

Montero-Pau, J., Gomez, A., & Muñoz, J., 2008. Application of an inexpensive and high-throughput genomic DNA extraction method for molecular ecology of zooplankton diapausing eggs. Limnology & Oceanography – Methods 6: 218-222.

Morton, P.K., Frisch, D., Jeyasingh, P.D., & Weider, L.J., 2015. Out with the old, in

with the new? Younger *Daphnia* clones are competitively superior over centuries-old ancestors. Hydrobiologia 749: 43-52.

Moya, A., Huisman, L., Ball, E.E., Hayward, D.C., Grasso, L.C., Chua, C.M., Woo, H.N., Gattuso, J.P., Forêt S., & Miller, D.J., 2012. Whole transcriptome analysis of the coral *Acropora millepora* reveals complex responses to CO_2-driven acidification during the initiation of calcification. Molecular Ecology 21: 2440-2454.

Nédli, J., & Forró, L., 2013. Allozyme-based genetic variability of the *Daphnia atkinsoni-bolivari* species complex (Cladocera: Daphniidae) in the Hungarian Great Plains. Acta Zoologica Academiae Scietiarum Hungaricae 59: 67-79.

Nei, M., 1971. Genetic distance between populations. The American Naturalist 106: 283-292.

O'Brien, W.J., & deNoyelles, F., JR. 1972 . Photosynthetically elevated pH as a factor in zooplankton mortality in nutrient enriched ponds. Ecology 53: 605-614.

Peakall, R. & Smouse, P.E., 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6: 288-295.

Phillimore, A.B., Hadfield, J.D., Jones, O.R., & Smithers, R.J., 2010. Differences in spawning date between populations of common frog reveal local adaptation. Proceeding from the National Academy of Sciences 107: 8292-8297.

Pollard, H.G., Colbourne, J.K., & Keller, W., 2003. Reconstruction of centuries-old *Daphnia* communities in a Lake recovering from acidification and metal contamination. Ambio: A Journal of the Human Environment 23: 214-218.

Pritchard, J.K, Stephen, M., & Donnelly, P., 2000. Inference of population structure using mutlilocus genotype data. Genetics 155: 945-959.

Raleigh, R.F., Hickman, T.J., Nelson, K.L. & Maughan, O.E., 1980. Riverine habitat evaluation procedures for rainbow trout. In: Proceedings of the trout stream improvement workshop, USDA Forest Service and Trout Unlimited, pp. 50-59. Asheville, NC, November 1980.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org.

Rosenberg, N.A., 2004. DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes 4: 137-138.

Satake, K., Oyagi, A., & Iwao, Y., 1995. Natural acidification of lakes and rivers in Japan: The ecosystem of Lake Usoriko (ph 3.4-3.8). Water Air & Soil Pollution 85: 511-516.

Schindler, D.W., Turner, M.A. & Hesslein, R.H., 1985. Acidification and alkalization of lakes by experimental addition of nitrogen compounds. Biogeochemistry 1: 117-133.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marqui,s M., et al., 2007. Climate Change 2007: The Physical Science Basis: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge Univ. Press, New York.

Stollewerk, A., 2010. The water flea *Daphnia* – a 'new' model system for ecology and evolution? Journal of Biology 9: 1-4.

Toonen, R.J. & Hughes, S., 2001. Increased throughput for fragment analysis on ABI Prism 377 automated sequencer using a membrane comb and STRand software, Biotechniques 31: 1320-1324.

Urabe, J. & Sterner, W., 2001. Contrasting effects of different types of resource depletion on life-history traits in *Daphnia*. Functional Ecology 15: 165-174.

Verhoeven, J.T.A, Arheimer, B., Yin, C., & Hefting, M.M., 2006. Regional and global concerns over wetlands and water quality. Trends in Ecology & Evolution 21: 96-103.

Weider, L.J., Makino, W., Acharya, K., Glenn, K.I., Kyle, M., Urabe, J. & Elser, J.J., 2005. Gentotype X environment interactions, stoichiometric food quality effects and clonal coexistence in *Daphnia pulex*. Oecologia 143: 537-547.

Weisse, T., Scheffel, U., Stadler, P., & Foissner, W., 2007. Local adaptation among geographically distant clones of the cosmopolitan freshwater ciliate *Meseres corlissi.* II. Response to pH. Aquatic Microbial Ecology 47: 289-297.

Wetzel, R.G., 2001. Limnology: Lake and River Systems 3rd edition. Elsevier Academic Press, London.

Yan, N.D., Keller, W., Somers, K.M., Pawson, T.W., & Girard, R.G., 1996. Recovery of crustacean zooplankton communities from acid and metal contamination: comparing manipulation and reference lakes. Canadian Journal of Fisheries & Aquatic Science 53: 1301-1327.

Tables & Figures

Table 1. Average annual (± SD) pH, sample size and population genetic parameters measured in three North American lake population of *D. pulicaria*. Included in the sample size are data collected from the lake samples, animals hatched from eggs, and DNA extracted from eggs. N_a = average number of alleles per locus (± SE), A_p = average number of private alleles per locus (± SE), P = percentage of loci that are polymorphic, I = Shannon Information Index (± SE), H_0 = observed heterozygosity (± SE), H_e = expected heterozygosity (± SE; GenAlEx), and M = results of the M-ratio test for population bottlenecks.

Lake	*pH*	Lake Animals	Hatched	Eggs	N_a	A_p	*P*	*I*	H_o	H_e	M
	± 1SD	(n)	(n)	(n)	± SE	±SE	%	± SE	± SE	± SE	
Frenchman	6.53 ± 0.50	67	0	0	1.33 ± 0.13	0.067 ± 0.70	33.3	0.231 ± 0.09	0.333 ± 0.13	0.167 ±0.06	1.10
Hill	7.91 ± 0.52	88	5	17	2.87 ± 0.26	0.600 ± 0.13	100.0	0.442 ± 0.10	0.239 ± 0.06	0.254 ± 0.06	1.37
Madison	8.63 ±0.25	77	15	0	2.93 ± 0.33	0.667 ± 0.19	86.7	0.564 ± 0.11	0.323 ± 0.07	0.313 ± 0.06	1.16

Table 2. Results of the pH survivorship experiment fitted using a generalized linear mixed-effects model, using a binomial distribution. The fixed factors were the pH treatment and the lake of origin, and the random factors were the jars in which the experiment took place. First level for the lake of origin is Hill Lake (represented as the intercept). Significant results are shown in bold font.

	Estimate	**Std. Error**	**Z value**	**P (> | z |)**
(Intercept)	-2.2336	0.7554	-2.957	**0.00311**
Madison Lake	-0.5197	1.1134	-0.467	0.64068
Frenchman Lake	4.42	1.0755	4.11	**3.96E-05**
pH treatment	0.4898	0.1047	4.679	**2.89E-06**
Madison Lake × pH treatment	0.1003	0.1556	6.45E-01	0.51908
Frenchman Lake × pH treatment	-0.5892	0.1447	-4.07E+00	**4.66E-05**

Figure 1. Population genetic structure of *D. pulicaria* in three North American lakes (Frenchman, Hill, Madison) calculated using the program STRUCTURE (REF). Each vertical bar in the STRUCTURE plot represents an individual *Daphnia*. $K = 3$ was the estimated most likely number of clusters determined by the greatest rate of change in the likelihood function (ΔK) (Evanno *et al.*, 2005).

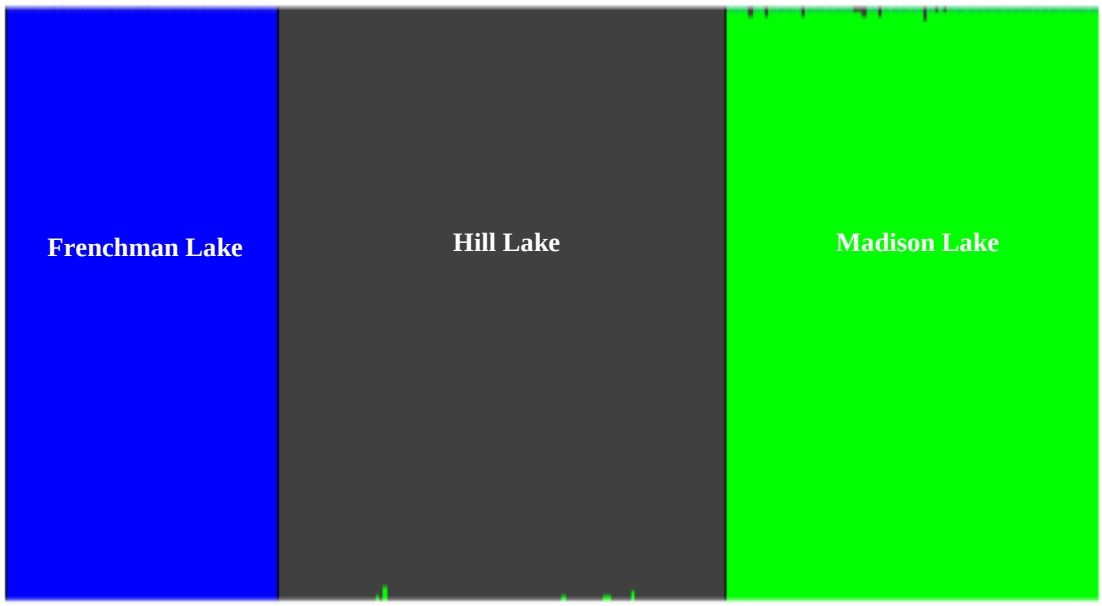


Figure 2. Principal Coordinates Analysis (PCA) based on Nei's genetic distance of 15 microsatellite loci from DNA extracted from *D. pulicaria* harvested from Frenchman (N=67), Hill (N=110), and Madison Lakes (N=92) implemented in GENEALEX 6.0.

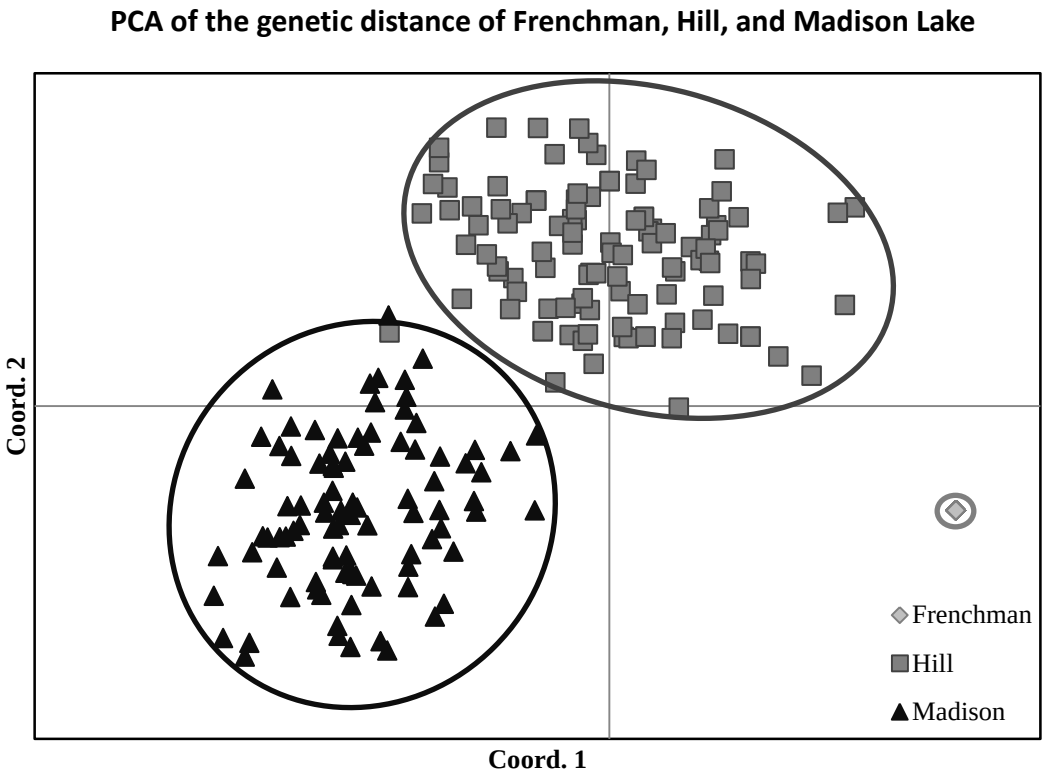
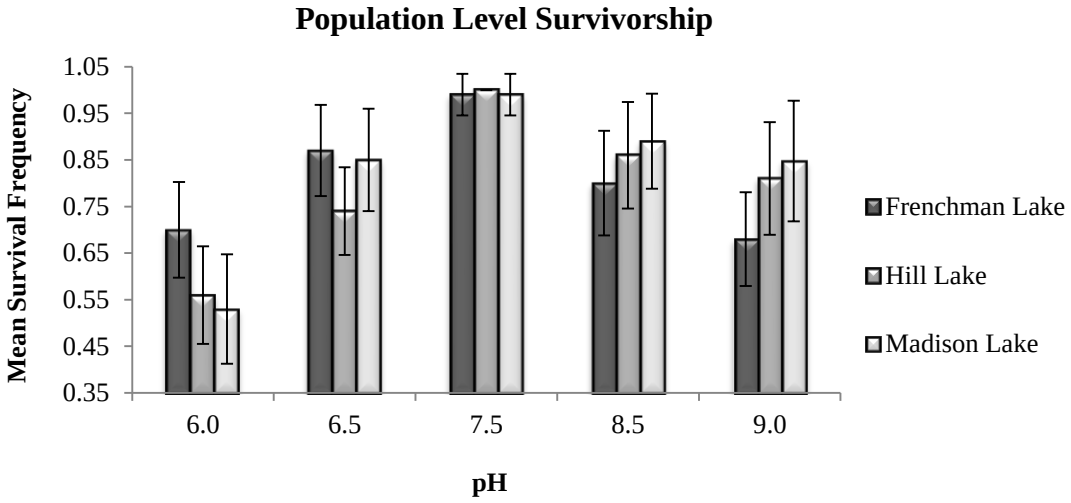


Figure 3. Results of pH survivorship experiments among populations. Mean survivorship (± 1 SD) overall population-level survivorship for Frenchman, Hill, and Madison Lakes.



Chapter Two

As published in *International Journal of Evolutionary Biology* 2015: 1-11, March 2015

The evolutionary history of Daphniid α-Carbonic Anhydrase within Animalia

Billy W. Culver

and

Philip K. Morton

Abstract

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 metallo-enzymes, 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 α-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 α-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 α-CAs have evolved, we find that contrary to expectations, amino acid conservation with bacterial α-CAs supports the supposition that extracellular α-CAs are the ancestral state of animal α-CAs. Furthermore, we show that two cytosolic and one GPI-anchored α-CA in *Daphnia* 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*, as compared with other organisms.

Keywords

Carbonic anhydrase, phylogeny, *Daphnia*, gene evolution, gene duplication, acid-base regulation

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 [1,2,3], 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 [7,8,9,10,11].

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 regards 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 [15,16,17], decapods [18,19], and aquatic insects [20]) has been linked to the enzyme carbonic anhydrase (CA). CAs are zinc metallo-enzymes that catalyze the reversible hydration/dehydration reaction: $CO_2 + H_20 \Leftrightarrow H_2CO_3 \Leftrightarrow HCO_3^- + H^+$ and are fundamental to many biological processes in addition to acid-base regulation; e.g.

photosynthesis [21], respiration [18,22], osmoregulation [18,22], bone resorption [23], and biominerization [24]. CAs are classified into five evolutionarily-distinct and unrelated super-families: α, β, γ, δ, and ζ, each super-family 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 α-CA super-family typically have 16 or 17 different isoforms within vertebrates, which are the primary contributors to acid-base regulation. The α-CA super-family 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 α-CAs in gills have been shown to regulate internal pH (Figure 1). The β-CAs are typically found only in bacteria, plants, algae, and fungi; however β-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 β-CAs in animals, but in plants they are catalytically similar to α-CAs in animals. The γ-CAs have only been found in archaea and bacteria, while δ-CAs and ζ-CAs have only been found in marine diatoms [24]. The ζ-CAs are unique among CAs since they replace the zinc ion with cadmium [30].

In this study, we investigated the evolutionary history of α-CAs in the microcrustacean *Daphnia* using phylogenetic methods. Since little is understood about β-CAs in animals, this study focuses on α-CAs. *Daphnia* 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* and nematodes, and 30% higher than humans [27]. Since *Daphnia* have 30 isoforms [22] of α-CAs, compared to the 15 in other organisms, this lends itself to the notion that there have been multiple duplication events within the *Daphnia* α-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) [31,32,33].

In addition, we used the phylogenetic analysis of the super-family of α-CAs to clarify which *Daphnia* α-CAs may be investigated further for their role in acid-base regulation. The criteria for this analysis involved examining *Daphnia* α-CA genes with functioning α-CA homologs in other crustaceans [18,19], aquatic insects [20], and fish [15,16,17]. Further, we investigated the evolutionary history of α-CAs in *Daphnia* to elucidate the functionality of duplicate α-CA genes, if they indeed exist.

Materials and methods

Sequence retrieval

All sequences, except *Daphnia* sequences, were obtained from the National Center for Biotechnology Information (NCBI). A key word search for "*Homo sapien* Carbonic Anhydrase" was performed for each of the 16 human isoforms of α-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) (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 α-CA amino acid sequences were retrieved from bacteria to use as an out-group. This search resulted in 213 amino sequences from taxa ranging from cnidarians to mammals (Table S1). 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 the some of the invertebrate clades.

Daphnia sequences were obtained from the *Daphnia* Genomics Consortium (DGC) (wFleaBase.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* α-CA was mapped to their respective chromosome to infer duplication history and duplication events. We mapped each *D. pulex* α-CA isoform using known scaffold positions on their respective chromosomes [31].

Sequence Alignment and Phylogenetic Analysis

The 271 amino acid sequences were uploaded into CLUSTALX 2.1 [36] and a multi-sequence 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 that 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, 250K 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 be found in Table S1. The aligned sequences were also used to determine residues that were conserved within each group of α-CAs (Table S2). A cutoff of 80% was used to determine if residues were conserved within an α-CA group across the entire phylogeny; however, if less than three species were in a group, then 100% conservation was used.

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 α-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 (most recent common ancestor) and the number of amino acid changes from the most recent common ancestor (including homoplasies).

N-terminus, GPI-anchor, and transmembrane prediction of *Daphnia* CA6 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 (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 gylcophosphatidylinositol (GPI)-anchored α-CAs. The *Daphnia* 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) non-plant 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 α-CAs with known N-terminus sequences. *Daphnia* 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* 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 α-CAs from the secretory and transmembrane α-CAs. Amino acids from *Daphnia* CA6 and CA7s were uploaded into GPI-SOM to predict *Daphnia* GPI-anchored α-CAs.

All of the prediction software was successfully able to predict their knowns from the vertebrate α-CA families.

Results and Discussion

General phylogenetic distribution of CA types in animals

Typically α-CAs are characterized by: (i) four active site residues - histidine (His)-316 (His-64 using nomenclature of vertebrate α-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 [47]. 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 [45]. 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 α-CAs resulted in two sister clades representing extracellular and intracellular α-CAs,

and appeared after the split of animals, plants, and fungi from bacteria. Before this early divergence, the most likely ancestral state of the α-CAs were extracellular (which include the (GPI)-anchored, transmembrane/membrane bound, and secretory α-CAs), as is evident by the bacterial α-CAs having similarly conserved residues as the extracellular α-CAs in animals (Table S3). Another line of evidence suggests that bacterial α-CAs are formed near or on the cytoplasmic membrane [46]. In addition, Le Roy et al. [24] found that porifera α-CAs were more similar to extracellular α-CAs, and were more basal phylogenetically than intracellular α-CAs. In particular, both bacterial α-CAs and extracellular animal α-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 α-CAs (Table 1, Table S2). Further, extracellular α-CAs and bacterial α-CAs share the following conserved residues that are not found in intracellular α-CAs (however they are found in some of the CA-RPs): Asparagine (Asn)-314, Asn-434, Tyrosine (Tyr)-491, Arginine (Arg)-492, Arg-578 (Table S3). These results are contrary to the commonly held notion that the intracellular α-CAs are the ancestral state [25].

The GPI-anchored α-CAs, found within the extracellular α-CAs, form three monophyletic clades: these clades consists of an invertebrate clade (including the chordate amphioxus), vertebrate clade, and an insect clade. Of note is 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 α-CAs are 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 [46, 47]. One of the weaknesses of constructing robust phylogenies of metazoans α-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 α-CA isoforms to increase coverage of both isoforms and taxa [24]. Even taking this approach, however, there is a severe lack of data on sequenced α-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 α-CAs further diverged into the secretory type α-CAs due to the loss of the C-terminus cleavage and anchoring site [22,45], which occurred after the appearance of amphioxus. This can be deduced, since the secretory α-CA appears in all vertebrates. The secretory α-CA is characterized by CA6, which is localized in the saliva of vertebrates. Membrane-bound α-CAs diverged from a common ancestor with the secretory α-CAs based on phylogenetic support that shows the divergence occurring after amphioxus, but before the amphibian/fish divergence. The transmembrane α-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 α-CAs diverged from intracellular α-CAs. The intracellular α-CAs are characterized by an amino acid change from the ancestral state at the following residues: 233 – Isoleucine (Ile) to Proline (Pro), 314 – Asn to Thr, 318 – Ile to Serine (Ser), 319 – Gln to Phenylalanine (Phe), 448 – Ille to Thr, 491 – Tyr to Trytophan (Trp), 492 – Arg to Thr, and 505 – Glysine (Gly) to Ser (Table S3). During the evolution of intracellular α-CAs, a duplication event likely occurred, which split intracellular α-CAs into two clades: CA-related proteins (CA-RP) and cytosolic α-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 [47]. 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 non-catalytic. These results suggest that CA8 may have an important biological function within deuterstomes [27]. Further, the CA8 subclade is sister to both CA11s and CA10s. The CA11s (including *Daphnia* CA3 and CA4 – nomenclature for the α-CAs of *Daphnia* and many invertebrates are not consistent with the nomenclature of α-CAs used for mammals) form two distinct groupings: protostomes and deuterstomes.

The cytosolic α-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 α-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 [24,45]. Also of interest within the cytosolic α-CAs 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, 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 [45]. In addition, vertebrate CA1 and CA13 has 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 α-CAs, while the others have varying degrees of lower activity [45]. Within the cytosolic α-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 [45], 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 [49, 50]; however, it is not universal to all teleosts. This is represented within the collapsed clade of teleost fish CA1/2 of the cytosolic α-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 α-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.

Daphnia CA isoforms

Of the *Daphnia* α-CAs, two fall within the cytosolic family (CA1 & CA2), clustering with other arthropods, echinoderms, and cnidarians (Figure1). *Daphnia* CA5, clusters with GPI-anchoring α-CAs of other arthropods (Figure 1). Two other α-CAs (CA3 & CA4) cluster within the CA-RP clade and are sister to hexapod CA-RPs (Figure 1). Specifically, CA3 is closely associated with hexapod CA11a, while CA4 is sister to hexapod CA11b.

The remaining 25 α-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 CA6 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 & 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 α-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 α-CAs in that they all had N-terminus signaling peptides and C-terminus cleavage and anchoring sites (Table 1, 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 α-CAs are secretory (Table2, Figure 2). The two remaining *Daphnia* α-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 α-CAs within their respective extracellular clade. They suggest that this may be an internalization of a formally secreted α-CA or they may be secreted proteins that are shuttled out the cell in a novel manner. Further research is warranted to verify the function and localization of these α-CAs.

Duplication events in CA isoforms in *Daphnia*

Phylogenetic results also support the hypothesis of multiple duplication events in *Daphnia*. The first duplication event seems to be the result of gene-level duplication in an ancestral species that resulted in the divergence of cytosolic α-CAs from extracellular α-CAs and CA-RPs. This is supported by the fact that extracellular α-CAs and CA-RPs (the predicted ancestral state of α-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 α-CAs. Evidence in support of this is that the CA-RPs (and

extracellular α-CAs) are found on chromosome 7, while the cytosolic α-CAs are found on chromosome 4. Since, *Daphnia* 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 likelihood 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 α-CA in *Daphnia* (CA5), there is a radiation of 25 α-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 α-CAs (Table 1, Figure 2). Four isoforms, CA6F, CA7H, CA7K, and CA7O, later reverted to GPI-anchored α-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* and other invertebrate genes and genomes have unveiled duplicated genes that have led to neofunctionalization, such as the spooky genes in arthropods [51]. If these duplications prove to be neofunctional, then *Daphnia* would have a larger than expected number of neofunctional isoforms. Kondrashov [52], 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* has 30 isoforms of α-CA, while most vertebrates only have 15 or 16, lends support to previous work on the *Daphnia* genome, which found that *Daphnia* has a relatively high rate of gene duplication, at least within *D. pulex*. These high duplication rates are not novel to *Daphnia* and have been shown in another cyclically-parthenogenetic organism, the pea aphid, *Acyrthosiphon pisum* [53,54].

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 appear to be within the whole arthropod phylum; however, as more arthropod genomes are sequenced, this may fill in gaps in the phylogeny.

Conclusions

The results of this phylogenetic study support the previously held organization of the α-CA super-family of genes, namely that α-CAs are clustered into the following families: cytosolic, CA-RP, GPI-anchored, secretory, and membrane-bound [25]. Previous thought, however, was that intracellular α-CAs were the most likely ancestral state. In contrast, our results provide support that extracellular α-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 α-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 α-CAs are more closely related to the more basal vertebrate α-CAs, CA7 and CA5 (Table S3). Since the vertebrate CA5 is associated with mitochondria, invertebrate cytosolic α-CAs are therefore more similar to vertebrate CA7. This is also true of the extracellular α-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 α-CAs. Further, invertebrate groups have reduced diversity of cytosolic and extracellular α-CAs, when compared to vertebrates. However, this may be an artifact of the fact that invertebrate α-CAs have not been well investigated. A study of the purple sea urchin (*Strongylocentrotus purpuratus*) has uncovered 19 isoforms of α-CA, most of which are involved in acid-base regulation [55]. 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 α-CAs. Further investigation of invertebrate α-CAs may uncover a greater diversity of α-CAs within these families.

We had several overarching goals in performing this study with regards to the *Daphnia* genus. First, we were interested in the evolution of *Daphnia* α-CAs. Homologs of acid-base regulating α-CAs in organisms, such as crustaceans, aquatic insects, and fish were used to provide evidence in support of potential acid-base regulating α-CAs in *Daphnia*. Second, we were interested in gene duplication events in the α-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*, 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 α-CAs are in *Daphnia*. In other organisms, however, some experiments have shown differential expression of α-CAs across a pH gradient. For instance, Evans et al. [55] found that α-CA12 was differentially expressed in larval *S. purpuratus* at low pH conditions. Also, Lin et al. [16] found differential expression of α-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 α-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* lends support to the hypothesis that the eco-responsive nature of this organism may be due to possible neofunctionaliztion resulting from the high levels of gene duplication. Thus, the genome duplications in *Daphnia* may allow this organism to withstand an extensive range of environmental (e.g., pH) conditions that are encountered in aquatic habitat [27].

Acknowledgements

This study was supported by the U.S. National Science Foundation (NSF-IOS-OEI) grant to L. J. Weider (#0924289), the U. of Oklahoma Department of Biology Adams Scholarship to B.W Culver and the U. of Oklahoma Biological Station Summer Research Grant to B.W. Culver. We thank the U. of Minnesota LacCORE staff (in particular R. O'Grady) at the University of Minnesota, for assistance with coring and providing *D. pulicaria* samples from Madison Lake, MN. A. Harris and J. Medders facilitated sequencing of *D. pulicaria*. We also thank L.J. Weider for his advice, mentorship, and comments on this manuscript. In addition, we thank R. Broughton, D. Hambright, I. Schlupp, and R. Cichewicz for their input. We thank J. Cureton and R. Hartnett who reviewed this manuscript and provided invaluable comments on earlier drafts of this manuscript. We would also like to thank the two anonymous reviewers that greatly improved the quality of this manuscript. This manuscript constitutes a portion of BWC's doctoral dissertation at the U. of Oklahoma.

Supplementary Material

Supplementary Table S1, S2, S3 and Figures S1, S2 are available online at *International Journal of Evolutionary Biology* (http://www.hindawi.com).

References

[1] Tortell P. 2000. Evolutionary and ecological perspectives on carbon acquisition in phytoplankton. *Limnoly Oceanogr*, 45(3):744-750.

[2] Hairston, Jr NG, Holtmeier CL, Lampert W, Weider LJ, Post DM, Fischer JM, Cáceres CE, Fox JA, Gaedke U. 2001. Natural selection for grazer resistance to toxic cyanobacteria: evolution of phenotypic plasticity? *Evolution*, 55(11):2203-2214.

[3] Jeyasingh PD, Ragavendran A, Paland S, Lopez JA, Sterner RW, Colbourne JK. 2011. How do consumers deal with stoichiometric constraints? Lessons from functional genomics using *Daphnia pulex*. *Mol Ecol*, 20:2341-2352.

[4] Urabe J, Togari J, Elser JJ. 2003. Stoichiometric impacts of increased carbon dioxide on a planktonic herbivore. *Glob Change Biol*, 9:818-825.

[5] Pollard HG, Colbourne JK, Keller W. 2003. Reconstruction of Centuries-old *Daphnia* Communities in a Lake Recovering from Acidification and Metal Contamination. *AMBIO: J Human Environ*, 32(3):214-218.

[6] Martins CMG, Almeida DV, Marins LFF, Bianchini A. 2010. mRNA expression and activity of ion-transporting proteins in gills of the blue crab *Callinectes sapidus*: effects of waterborne copper. *Environmental Toxicology and Chemistry*, 30(1):206-211.

[7] O'Brien WJ, deNoyelles F JR. 1972. Photosynthetically elevated pH as a factor in zooplankton mortality in nutrient enriched ponds. *Ecology*, 53(4):605-614.

[8] Havens KE. 1992. Acidification effects on the plankton size spectrum: an *in situ* mesocosm experiment. *Journal of Plankton Research*, 14(12):1687-1696.

[9] Derry AM, Arnott SE. 2007. Adaptive reversals in acid tolerance in copepods from lakes recovering from historical stress. *Ecological Applications*, 17(4):1116-1126.

[10] Erlandsson M, Cory N, Fölster J, Köhler S, Laudon H, Weyhenmeyer GA, Bishop K. 2011. Increasing Dissolved Organic Carbon Redefines the Extent of Surface Water Acidification and Helps Resolve a Classic Controversy. *BioScience*, 61(8):614-618.

[11] Moya A, Huisman L, Ball EE, Hayward DC, Grasso LC, Chua CM, Woo HN, Gattuso JP, Forêt S, Miller DJ. 2012. Whole Transcriptome Analysis of the Coral *Acropora millepora* Reveals Complex Responses to CO_2-driven Acidification during the Initiation of Calcification. *Mol Ecol*, 21:2440-2454.

[12] Eppinger RG, Briggs PH, Dusel-Bacon C, Giles SA, Gough LP, Hammarstrom JM, Hubbard BE. 2007. Environmental geochemistry at Red Mountain, an unmined volcanogenic massive sulphide deposit in the Bonnifield district, Alaska Range, east-central Alaska. *Geochemistry: Exploration, Environment, Analysis*, 7:207-223.

[13] Satake K, Oyagi A, Iwao Y. 1995. Natural acidification of lakes and rivers in Japan: The ecosystem of Lake Usoriko (ph 3.4-3.8). *Water Air Soil Pollut*, 85(2):511-516.

[14] Ezoe Y, Lin CH, Noto M, Watanabe Y, Yoshimura K. 2002. Evolution of water chemistry in natural acidic environments in Yangmingshan, Taiwan. *J Environ Monit*, 4:533-540.

[15] Georgalis T, Perry SF, Gilmour KM. 2006. The role of branchial carbonic anhydrase in acid-base regulation in rainbow trout (*Oncorhynchus mykiss*). *J exp*

Biol, 209:518-530.

[16] Lin TY, Liao BK, Horng JL, Yan JJ, Hsiao CD, Hwang PP. 2008. Carbonic anhydrase 2-like a and 15a are involved in acid-base regulation and Na^+ uptake in zebrafish H^+-ATPase-rich cells. *Am J Physiol Cell Physiol*, 294:C1250-C1260.

[17] Gilmour KM, Perry SF. 2009. Carbonic anhydrase and acid-base regulation in fish. *J exp Biol*, 212:1647-1661.

[18] Henry RP, Cameron JN. 1983. The role of Carbonic Anhydrase in respiration, ion regulation and acid-base balance in the aquatic crab *Callinectes sapidus* and the terrestrial crab *Gecarnicus lateralis*. *J exp Biol*, 103:205-223.

[19] Henry RP. 1984. The Role of Carbonic Anhydrase in Blood Ion and Acid-Base Regulation. *Amer Zool*, 24:241-251.

[20] Cooper PD. 1994. Mechanisms of Hemolymph Acid-Base Regulation in Aquatic insects. *Physiol Zool*, 67(1):29-53.

[21] Badger MR, Price GD. 1994. The role of carbonic anhydrase in photosynthesis. *Annu Rev Plant Physiol Plant Mol Biol*, 45:369-392.

[22] Weber AK, Pirow R. 2009. Physiological responses of *Daphnia pulex* to acid stress. *BMC Physiol*, 9(9): 1-25.

[23] Teitelbaum SL. 2000. Bone Resorption by Osteoclasts. *Science*, 289(5484):1504-1508.

[24] Le Roy N, Jackson DJ, Marie, B, Ramon-Silva P, Marin F. 2014. The evolution of metazoan α-carbonic anhydrases and their roles in calcium carbonate biomineralization.

Frontiers in Zoology, 11: 1- 16.

[25] Hewett-Emmett D, Tashian RE. 1996. Functional Diversity, Conservation, and Convergence in the evolution of the α-, β-, and γ-Carbonic Anhydrase Gene Family. *Mol Phylo Evol*, 5(1):50-77.

[26] So AK, Espie GS. 2005. Cyanobacterial carbonic anhydrases. *Can J Bot* 83:721-734.

[27] Aspatwar A, Martti MEE, Parkkila S. 2010. Phylogeny and expression of carbonic anhydrase-related proteins. *BMC Mol Biol*, 11(25):1-19.

[28] Fasseas MK, Tsikou D, Flemetakis E, Katinakis P. 2010. Molecular and biochemical analysis of the class β class carbonic anhydrases in *Caenorhabitis elegans*. *Mol Biol Rep* 37:2941-2950.

[29] Linser PJ, Smith KE, Seron TJ, Oviedo MN. 2009. Carbonic anhydrases and anion transport in mosquito midgut pH regulation. *J Exp Biol* 212: 1662-1671.

[30] Xu Y, Feng L, Jeffrey PD, Shi Y, Morel FMM. 2008. Structure and metal exchange in the cadmium carbonic anhydrase of marine diatoms. *Nature* 452(6): 56-62.

[31] Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, Tokishita S, Aerts A, Arnold GJ, Basu MK, et al. 2011. The Ecoresponsive Genome of *Daphnia pulex*. *Science*, 331:555-561.

[32] Lynch M, Conery JS. 2000. The Evolutionary Fate and Consequences of Duplicate Genes. *Science*, 290(5494):1151-1155.

[33] Näsvall J, Sun L, Roth JR, Andersson DI. 2012. Real-Time Evolution of New Genes by Innovation, Amplification, and Divergence. *Science*, 388:384-387.

[34] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5:

Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol*, 28(10):2731-2739.

[35] Gertz EM, Yu YK, Agarwala R, Schäffer AA, Altschul SF. 2006. Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. *BMC Biol*, 4(41):1-14.

[36] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21):2947-2948.

[37] Whelan S, Goldman N. 2001. A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Mol Biol Evol*, 18(5): 691-699.

[38] Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. In *Proceedings of the Gateway Computing Environments Workshop (CGE)*; *New Orleans, LA*. 2010:1-8.

[39] Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12):1572-1574.

[40] Stamatakis A. 2014. RaxML version 8: A tool for phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*, 1-2.

[41] Rambaut A. 2009. FigTree v1.3.1: Tree figure drawing site. *FigTree website*.

[42] Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305(3):567-580.

[43] Emmanuelsson O, Brunak S, von Heijne G, Nielson H. 2007. Locating proteins in the cell using TargetP, Signal IP, and related tools. *Nature Protocols* 2: 953-971.

[44] Fankhauser N, Maeser P. 2005. Identification of GPI anchor attachment signals by a Kohonen self-organizing map. *Bioinformatics* 21(9): 1846-1852.

[45] Hilvo M, Tolvanen M, Clark A, Shen B, Shah GN, Waheed A, Halmi P, Hänninen M, Hämäläinen JM, Vihinen M, et al. 2005. Characterization of CA XV, a new GPI-anchored form of carbonic anhydrase. *Biochem J*, 392:83-92.

[46] Smith KS, Ferry JG. 2000. Prokaryotic carbonic anhydrases. *FEMS Microbiol Rev* 24:335-366.

[47] Esbaugh AJ, Tufts BL. 2006. The structure and function of carbonic anhydrase isozymes in the respiratory system of vertebrates. *Respiritory Physiology & Neurobiology* 154: 185-198.

[48] Barnea G, Silvennoinen O, Shaanan B, Honegger AM, Canoll PD, D'Eustachio P, Morse B, Levy JB, Laforgia S, Huebner K. 1993. Identification of a carbonic anhydrase-like domain in the extracellular region of RPTP gamma defines a new subfamily of receptor tyrosine phosphatases. *Mol Cell Biol*, 13(3):1497-1506.

[49] Hoegg S, Brinkmann H, Taylor JS, Meyer A. 2004. Phylogenetic Timing of the Fish-Specific Genome Duplication Correlates with the Diversification of Teleost Fish. *J Mol Evol* 59: 190-203.

[50] Meyer A, Van De Peer Y. 2005. From 2R to 3R: eveidence for a fish-specific genome duplication (FSGD). *BioEssays* 27.9: 937-945.

[51] Rewitz KF; Gilbert LI. 2008. *Daphnia* Halloween genes that encode sytochrome P450s mediating the synthesis of the arthropod molting hormone: Evolutionary

implications. *BMC Evolutionary Biology* 8: 1-8.

[52] Kondrashov F. 2012. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc R Soc B*, 1-10.

[53] International Aphid Genomics Consortium. 2010. Genome Sequence of the Pea Aphid *Acyrthosiphon pisum*. *PLoS Biology* 8(2):1-24.

[54] Simon J-C, Pfrender ME; Tollrian R, Tagu D, Colbourne JK. 2011. Genomics of Environmentally Induced Phenotypes in 2 Extremely Plastic Arthropods. *J Heredity* 102(5):512-525.

[55] Evans TG, Chan F, Menge BA, Hofmann GE. 2013. Transcriptomic responses to ocean acidification in larval sea urchins from a naturally variable pH environment. *Mol Ecol*, 22:1609-1625.

Tables and Figures

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 a transmembrane protein. N-terminus signaling peptides were elucidated using TargetP on the CBS Prediction Server, 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 a 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.

Daphnia 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

Table 2. Results of chromosome mapping which reflect the *D. pulex* α-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.

D. pulex 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

Figure 1. Phylogeny of α-CAs inferred from a maximum-likelihood analysis performed with Mr. Bayes, 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, Membrane bound, Secretory, CA-RP, and cytosolic. The bacterial α-CAs represents the outgroup.

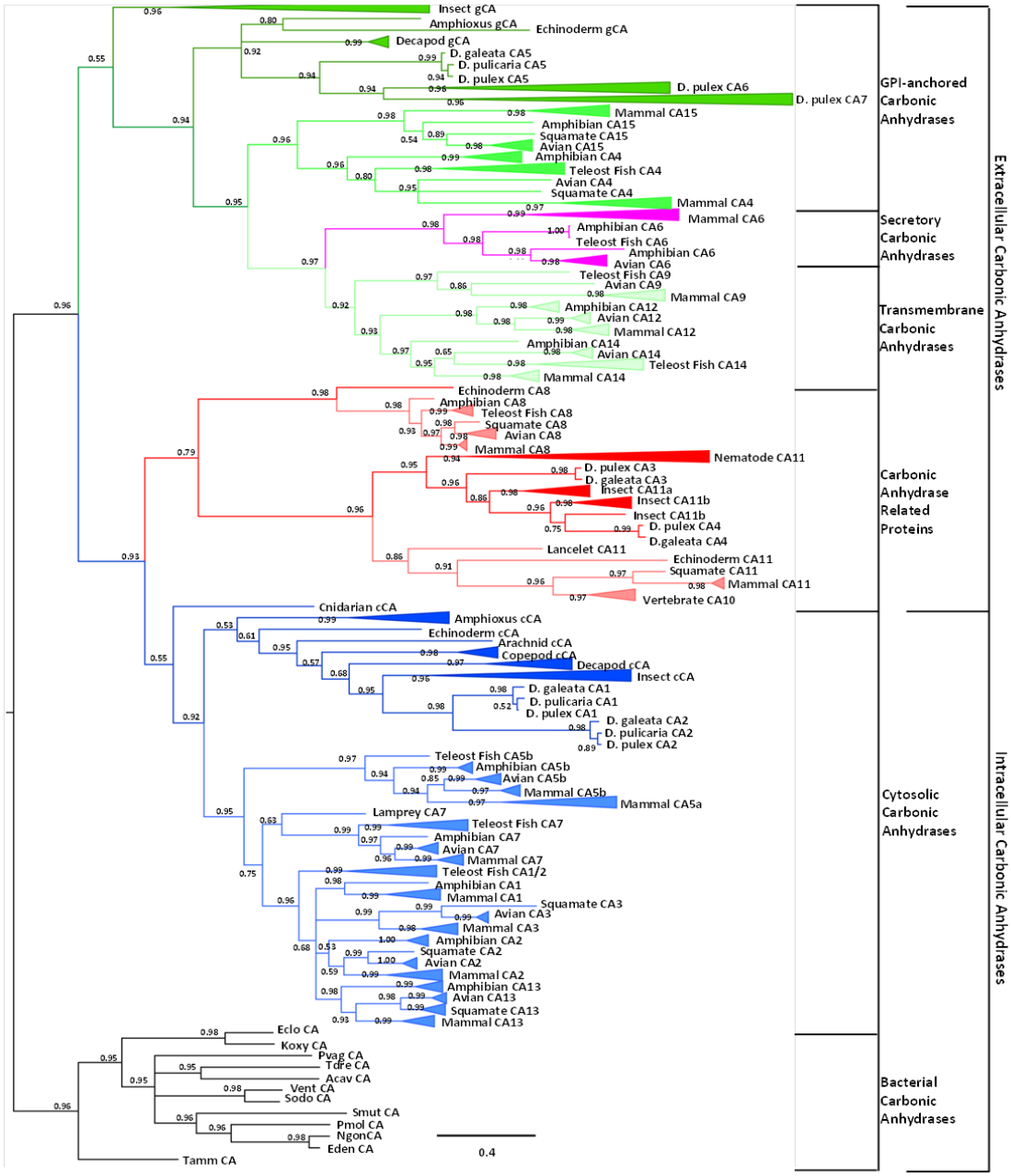
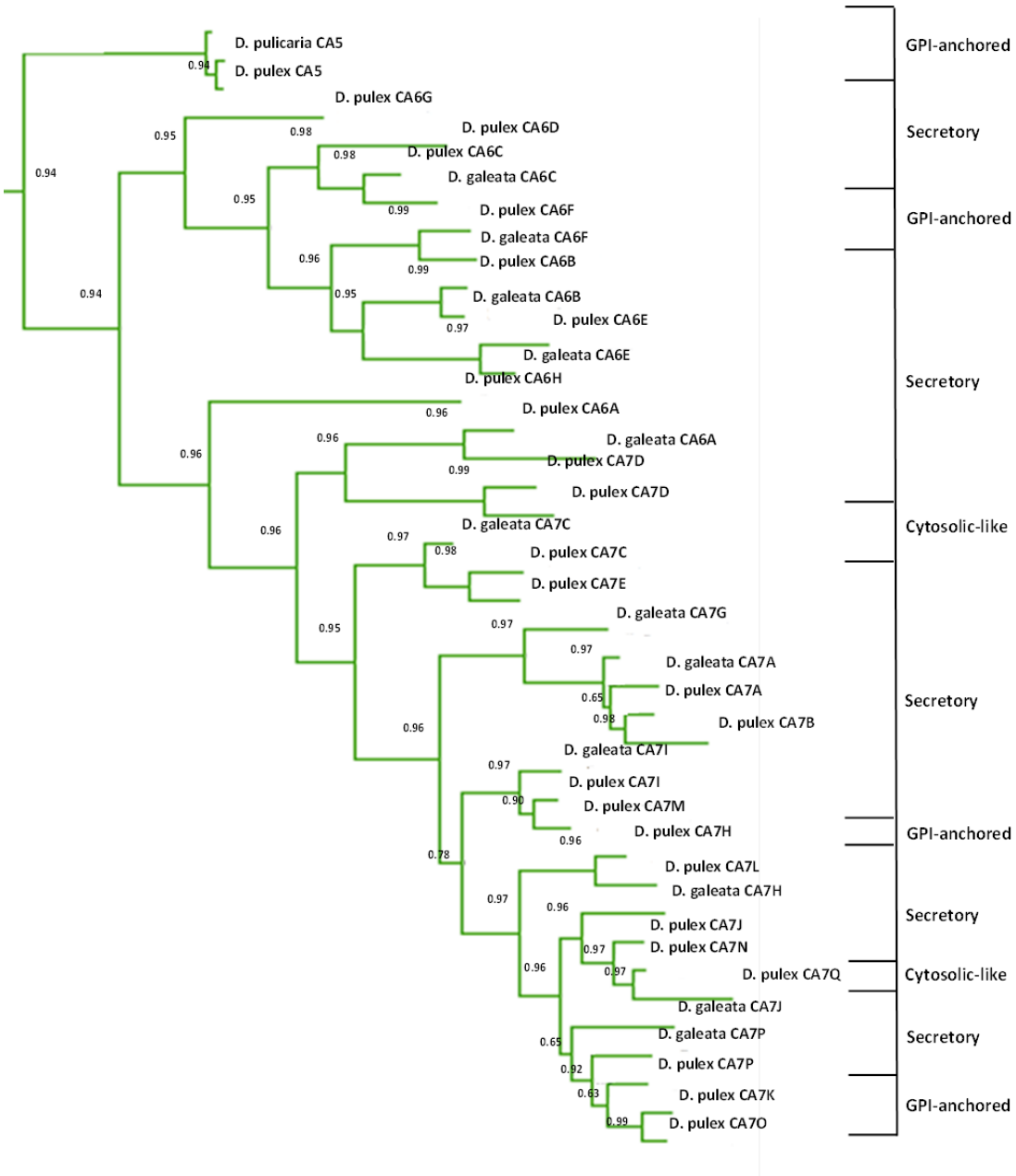


Figure 2. Isolated view of *Daphnia* CA5, CA6, and CA7s based on the phylogeny represented in Figure 1. Posterior probabilities of the branches are indicated at the nodes. On the right side of the phylogeny are the predicted states for *Daphnia* CA5, CA6, and CA7s.



Chapter Three

Formatted for *Molecular Biology and Evolution*

An investigation of three Carbonic Anhydrase isoforms in *Daphnia pulicaria* with signatures of local adaptation to acid-base regulation across a pH gradient

Billy W. Culver

and

Philip K. Morton

Abstract

Understanding the mechanisms that drive acid-base regulation in organisms that experience rapidly fluctuating pH conditions, such as in aquatic habitats, is important. Previous studies have shown that carbonic anhydrases (CAs) are responsible for acid-base regulation in many organisms. Here, we characterize three CA isoforms (CA1, CA2, and CA5) in the keystone aquatic herbivore, *Daphnia pulicaria*. In addition we investigate under common garden conditions the differential expression of those CAs from *D. pulicaria* clones isolated from three North American lakes which exhibit a pH gradient. The three *D. pulicaria* CAs were determined to have seven exons, six introns, and highly conserved motifs common to all Animalia CAs. Two of the three *D. pulicaria* CAs are cytosolic type CAs, while the third is a GPI-anchored CA. The cytosolic CAs have a key substitution at one of three active sites. In addition, we found that CA1 is differentially expressed across a pH gradient and among lake isolates, showing a signal of local adaptation in expression in two of the lakes (Frenchman and Madison). We also found significant up regulation of CA5 in the Frenchman Lake isolate relative to isolates from Madison Lake and a third lake (Hill) across all pH treatments. There are a number of anthropogenic and natural phenomena that have altered pH in aquatic environments; therefore it is critical to understand how organisms, such as *Daphnia*, respond with regards to fitness in these changing environments in order for conservation managers to understand how to properly manage these aquatic ecosystems.

Introduction

Carbonic anhydrase (CA) and its many isozymes are zinc metallo-enzymes that catalyze the reversible hydration/dehydration reaction: $CO_2 + H_2O \Leftrightarrow H_2CO_3 \Leftrightarrow HCO_3^- + H^+$ (Lindskog and Silverman 2000). CAs are fundamental to many physiological processes, such as photosynthesis (Badger and Price 1994), respiration (Henry and Cameron 1983), osmoregulation (Henry and Cameron 1983), bone resorption (Teitelbaum 2000), and biomineralization (Le Roy et al. 2014). CAs are classified into five evolutionarily-distinct and unrelated super-families: α, β, γ, δ, and ζ, each super-family has different active site amino acids, primary sequences, and protein structures (Hewett-Emmett and Tashian 1993, So and Espie 2005). These families are thought to be the result of convergent evolution. The α-CA super-family typically has 16 or 17 different isoforms within vertebrates, and is broken into four families: cytosolic, secretory, transmembrane/membrane-bound, and CA-related proteins (CA-RP). The latter family has have purportedly lost function due to the loss of at least one of the three active site histidine residues (Aspatwar et al. 2010). The β-CAs are typically found in bacteria, plants, algae, and fungi; however β-CAs have recently been found in a number of metazoans including *Caenorhabditis elegans* (Fasseas et al. 2010), *Anopheles gambiae* (Linser et al. 2009), *and Daphnia pulex* (Weber and Pirow 2009). There is a lack of knowledge on the catalytic activity and expression of β-CAs in animals, but in plants they are catalytically similar to α-CAs in animals. The γ-CAs have only been found in archaea and bacteria, while δ-CAs and ζ-CAs have only been found in marine diatoms (Le Roy et al. 2014). The ζ-CAs are unique among CAs since they replace the zinc ion with cadmium (Xu 2008).

Maintaining pH homeostasis in habitats with altered pH is critical for organisms to survive and reproduce. Acid-base regulation in a number of aquatic organisms (e.g. fish (Georgalis et al. 2006, Lin et al. 2008, Gilmour and Perry 2009), decapods (Henry and Cameron 1983, Henry 1984), and aquatic insects (Cooper 1994)) has been linked to CAs. In particular, cytosolic and membrane-bound α-CAs have been found in the gills of these organisms, and have been shown to regulate internal pH.

Acidification and alkalization are important ecological stressors that can act as strong selective forces on the genetic composition and physiological responses of aquatic populations. In addition to natural mechanisms of acidification and alkalization of water bodies, the increasing impacts from anthropogenic sources such as carbon dioxide emissions (Moya et al., 2012, Evans et al. 2013), cultural eutrophication (O'Brien & deNoyelles, 1972), and mining activities (Derry, 2007; Martins et al. 2010) are of great concern with regards to lake acidification and alkalization as these processes can negatively affect the osmoregulation, growth, and health of aquatic organisms that use calcium carbonate to build shells or exoskeletons (Hurd et al., 2011). Here we investigate the role of α-CAs in regulating pH in the keystone zooplankton herbivore, *Daphnia pulicaria*, which is known to be impacted by acidification and alkalization (fluctuating pH) of water bodies (Keller et al. 1990, Havens et al. 1993).

Daphnia pulicaria is an emerging model organism (Stollewerk 2010, Colbourne et al. 2011, Miner et al. 2012) that has a number of established characteristics that are considered ideal for genetic studies. For example, *Daphnia* exhibit both asexual and sexual reproduction (i.e., termed cyclical parthenogenesis) which allows us to control for the genetic background of clonal lineages and look at heritability in sexual

(recombinant) lineages. In addition, *Daphnia* are ecologically well-studied (Loaring and Hebert 1988; Lampert 1993, Urabe and Sterner 2001, with a plethora of studies indicating that *D. pulicaria* populations contain significant genetic variation that can be easily interpreted in an ecological or environmental context (Baird and Barata 1998, Dudycha and Tessier 1999, Coors et al. 2009). Further, they have very large population sizes (i.e., millions to billions; Hebert 1978, Lambert 2011), which reduces the impact of non-selective evolutionary forces (i.e., genetic drift) on the genetic composition of these populations. In addition, the genome of *D. pulex*, a sister species to *D.pulicaria*, has also been sequenced and annotated (Colbourne et al. 2011), which provides an important tool in genetic studies of *Daphnia*.

While α-CAs have been well characterized in many vertebrate species, i.e. *Mus musculus* (house mouse; Eicher et al. 1976), *Danio rerio* (zebrafish; Peterson et al. 1997), *Xenopus tropicalis* (pipid frog; Reece-Hoyes et al. 2002)), *Pan troglodytes* (chimpanzee; Cáceres et al. 2003) and *Homo sapiens* (Hilvo et al. 2003), with few exception α-CAs have not been well characterized among invertebrates. Here, we characterized three α-CA isoforms in the keystone aquatic herbivore, *Daphnia pulicaria*. The three α-CA isoforms (CA1, CA 2, and CA5) in question had previously been identified as potential candidate genes that may regulate pH in *Daphnia* (Culver and Morton 2015). We sequenced DNA for the three α-CA genes and translated these genes to the appropriate amino acid (AA) sequences with the goal of comparing the conserved sequence motifs with those of known α-CA isoforms that are involved in pH regulation in other taxa, i.e. *Calinectus Sapidus* (Martins et al. 2010), *Anopheles gambiae* (Smith et al. 2007), *Danio rerio* (Gilmour and Perry 2009), and *Homo sapiens*

(Hilvo et al. 2005). We predicted that if these isoforms of α-CA were indeed involved in pH regulation, then the DNA and AA sequences in *Daphnia pulicaria* should have similar architecture and conserved motifs as the other taxa.

In addition we investigated, under common garden conditions, the differential expression of those CAs from *D. pulicaria* clones isolated from three North American lakes that span a pH gradient: Frenchman Lake, Ontario, CN (average annual pH 6.5); Hill Lake, MN (average annual pH 7.9); and Madison Lake (average annual pH 8.6). Tolerance to a pH gradient in the genus *Daphnia* is well known, with *Daphnia* being found in the pH range of 6.5 - 10 depending on species (O'Brien & deNoyelles, 1972; Havens et al., 1993; Culver and Acosta *in review*). However, the genetic response to local variations in pH is poorly understood at the intraspecific-level. We investigated the differential expression (DE) of CA1, CA2, and CA5 in *D. pulicaria*, across a pH gradient observed in Frenchman, Hill, and Madison Lakes, using quantitative Real Time (qRT)-PCR. Having previously found evidence for local adaptation in these three *Daphnia* populations (Culver and Acosta *in review*), we predicted that there should be a difference in expression patterns among the *D. pulicaria* populations from each lake.

Results

Carbonic Anhydrase sequencing in Daphnia pulicaria

All *D. pulicaria* CA isoforms were found to have seven exons and six introns; however, each isoform had a distinct architecture (Figure 1). CA1 and CA2 are cytosolic sister isoforms that are likely the result of a tandem duplication (Culver and Morton 2015); however, they are strikingly different. In *D. pulicaria* CA1 is 2387bp, CA2 is 1596bp,

and CA5 is 6772bp, while in *D. pulex* they are 2220bp, 1606bp, and 6780bp respectively. The overall gene similarity in CA1 between *D. pulicaria* and *D. pulex* is 90.7%, while it is 97.5% among the exons and 87.6% among the introns; in CA2 there is a 94.6% overall similarity, with 95.6% similarity among exons, and 92.5% among introns; and in CA5 there is a 94.0% overall similarity, with 98.4% among exons, and 87.2% among introns. The major difference between the two isoforms of CA1 of both *Daphnia* species is that the first intron is longer than CA2's by ~780bp. Also, the first intron in CA1 is ~100bp shorter in *D. pulex* when compared with *D. pulicaria.* In addition, CA5, a gylcophosphatidylinositol (GPI)-anchored CA, in both *Daphnia* species has a large first intron, ~5300bp. (Note: it was not sequenced in *D. pulicaria*; therefore all analyses with introns and the gene, as a whole, do not account for the first intron). There were a number of small insertion/deletion (in/del) events (1 – 10bp) that were found in several of the introns; however there was no evidence of in/del events in any of the exons. However, we detected several point mutations in the exons: 16 in CA1, 32 in CA2, and 13 in CA5 in the comparison between *D. pulex* and *D. pulicaria*. Within the introns, there were large numbers of point mutations: 77 in CA1, 30 in CA2 and 50 in CA5 (excluding the first intron, as indicated above).

Comparison of Daphnia pulicaria Carbonic Anhydrases with other taxa

The results from the bioinformatic software Mauve analysis are illustrated by bar plots (Figure 2). The bar plots represent the similarity of the DNA sequences to *D. pulicaria* CA1, CA2 (Figure 2a), and CA5 (Figure 2b), respectively. The higher the histogram peaks, then the more similar the sequences are; regions of similarity (conserved regions

across taxa) are represented by the same colors. The Mauve plots showed that all the cytosolic CAs and all GPI-anchored CAs were homologous among taxa: *D. pulicaria*, *D. pulex*, *C. sapidus*, *A. gambiae*, *D. rerio*, and *H. sapiens*. Looking at the conserved regions among taxa in Mauve, there were indeed large regions of conservation among taxa (the purple areas in Figure 2). In the cytosolic CAs, there was a small region of conservation that was common among the crustaceans (light blue) towards the (3' terminal) end of the gene (700-750bp; Figure 2). In addition, a region that was only conserved in *Daphnia* CA2 was depicted as orange at the (5' end) beginning of the gene (0-100bp). Also of note was that *C. sapidus* was missing a conserved region (blue) at about the 150bp point in the gene. Among the CA5 isoforms, two regions of conservation were unique to *Daphnia* (light blue): the first was at the beginning of the gene (~0-100bp) and the second was at the 275-325bp point. There was also an area of conservation among the invertebrates (green and yellow) found at the 100-150bp point. In addition, whereas *Daphnia* had a common region of conservation (light blue) at 275-325bp, the remaining taxa had a similar region of conservation at the same point; however it was distinct (red) from that found in *Daphnia*. At the tail end (3' end) of the gene (850-925 bp), there was another region of conservation (orange) that was only found in the crustaceans and *D. rerio*.

The results from the aligned AA sequences were compared to the known conserved motifs, including active, zinc-binding, substrate-binding, and di-sulfide bonding sites in *H. sapiens* (Figure 3; Hilvo et al. 2005, Pilka et al. 2012). Among the cytosolic isoforms of CA (Figure 3a), *Daphnia* shared common motifs with all taxa analyzed, i.e. the QSP motif at residue 42 – 62, the GGPL motif at residue 101-106, the

GSEH motif at residue 127-130, the GLAVLG motif at residue 162-170, and the GSLTTPP motif at residue 224-230. All zinc-binding (residues 117, 119, and 140) and substrate-binding sites (residues 227-228) in the cytosolic CAs were conserved among all taxa, while all but one residue (86) was conserved among the active sites (residues 115, 130 and 227). At residue 86, the invertebrates had a Tyrosine (Y) or Alanine (A) residue, instead of the Histidine (H) residue found in vertebrates. The GPI-anchored isoforms of CA (Figure 3b) had several of the same motifs found in the cytosolic CAs, i.e. the QSP (residues 82-87), GSEH (residues 162-165), GLAVLG (residues 199-204) and GSLTTPP (residues 260-265). While the GPI-anchored CAs were missing the GGPL motif, they had another unique motif, i.e. NNGH (residue 117-122). All active sites (residues 120, 151, 163, and 263), zinc-binding sites (residues 153, 155, and 176), substrate-binding sites (residues 263-264) and di-sulfide bonding sites (residues 77 and 267) were conserved.

Differential expression (DE) experiment

For the duration of the common garden experiment, the pH treatments were maintained at the following means (+/- 1 SD): pH 6.0 (+/- 0.14), 7.52 (+/- 0.08), and 9.08 (+/- 0.16). For the results of qRT-PCR, a threshold of a fold change of 2.0 relative to the control indicated biologically significant results (Witten and Tibshirani 2007). An examination among treatments for each clone for each isoform (Figure 4) revealed that CA1 was the only isoform that showed significant DE relative to the control (pH 7.5). The Frenchman Lake clone showed a significant fold change in expression at pH 9.0, the Hill Lake clone showed a significant fold change in expression at both pH 6.0 and

9.0, and the Madison Lake clone showed a significant fold change in expression at pH 6.0. Looking among clones for each treatment for each isoform, CA5 was always significantly expressed in the Frenchman Lake clone relative to the control (Hill Lake) clone (Figure 4).

Discussion

It has been reported that 31 different isoforms of the α-CA genes exist in the *Daphnia pulex* genome (Weber and Pirow 2009; Colbourne et al. 2011; Culver and Morton 2015). Although these isoforms have been annotated, their functions have not been fully determined. Weber and Pirow (2009) posited that at least three of the isoforms (CA1, CA2, and CA5) should function with regards to acid-base regulation, since they have all the residues (active, zinc-binding, and substrate-binding sites) needed for a functional enzyme. In addition, Culver and Morton (2015) showed that these isoforms were homologs to those found in crustaceans and aquatic insects. Furthermore, these genes have not been explored in the sister species to *D. pulex*, *D. pulicaria*. *Daphnia pulex* and *D. pulicaria* are found in different aquatic environments, ponds and lakes respectively. The pH in ponds and lakes respond to the environment differently, i.e. buffering capacity, geology, and chemistry (Wetzel 2001). Therefore, these species may experience different environmental selective pressures with regards to variation in pH (Søndergaard et al. 2005). We predicted that *D. pulicaria* CA isoforms, while functionally similar to *D. pulex*, would exhibit significantly different DNA sequences. In addition, we hypothesized that if *D. pulicaria* CA isoforms were functional, then they would maintain conserved regions that are critical to acid-base regulation across taxa.

Lastly, we predicted that if the CA isoforms were involved in acid-base regulation, then they would be differentially expressed across a pH gradient. Further, if they evolved under different pH environments then we would expect to see signatures of local adaptation in the differential expression of the CA isoforms among clones from those different environments.

We did indeed find that the architecture of all three isoforms of CA between the two *Daphnia* species was conserved with all isoforms containing seven exons and six introns (Figure 1). However, the sequence differences were quite distinct among all three isoforms between the two *Daphnia* species. The largest difference was found in the first intron of CA1, where it was ~100bp shorter in *D. pulex*, when compared with *D. pulicaria*. Previous work on the functions of introns, has proposed that the first intron may control how quickly DNA is transcribed into mRNA (Chorev and Carmel 2012). This may be indicative of the differing selection pressures between the two different environments (ponds vs lakes) for *D. pulex* and *D. pulicaria*, respectively.

For instance, Søndergaard et al. (2005) found that the overall pH of water-bodies tends to decrease (i.e., becomes more acidic), as the size and volume of the water-body decreases. Additionally they found that smaller water bodies tended to fluctuate more in regards to pH magnitude and frequency. While not necessarily testing for pH response differences in ponds and lakes, Dudycha and Tessier (1999) tested for differential life-history traits between *D. pulex* living in ponds and *D. pulicaria* living in lakes. They found tradeoffs between early and late life histories, where *D. pulex* had increased fitness with regards to early life history and decreased fitness in regards to late life history. They found the opposite pattern for *D. pulicaria*. Dudycha and Tessier

(1999) were able to attribute these differences in life-history between these two sister species to the ecological differences associated with ponds and lakes.

When comparing conserved regions in the two cytosolic enzymes (CA1 and CA2) and the GPI-anchored enzyme (CA5) in *Daphnia* to other taxa (*C. sapidus*, *A. gambiae*, *D. rerio*, and *H. sapiens*), we found large regions of conservation (Figure 2). These regions of conservation in the genes are found near functionally constrained residues (active sites, zinc-binding sites, substrate-binding sites, etc.); thus they are not expected to be different in *Daphnia*, if these genes are functional. Additionally, those regions that are not conserved, are likely regions that are not constrained and free to evolve neutrally (Nei 1987) in that species. However, we still found regions that were conserved among groups of taxa in the cytosolic CAs, such as the region from 700-750bp that was found in the aquatic invertebrates, but not shared with vertebrates or *A. gambiae*. This may be due to some unique characteristics among aquatic invertebrates that functionally constrain this region of the gene, i.e. enhances the activity of the enzyme that gives the organism a fitness advantage, or is lethal without this conserved region in an aquatic environment. For instance, in a review of studies of thermal tolerance in snails of the genus *Tegula* and porcelain crabs of the genus *Petrolisthes* that have species that live in the subtidal zone and species that live in the intertidal zone, it was found that differences in the AA sequence of the proteins of Lactate dehydrogenase (LDH)–A and cytosolic malate dehydrogenase (cMDH) between subtidal species versus the intertidal species was related to differential fitness (Somero 2009). Of particular interest, it may be predicted that this conserved region should be seen in the cytosolic CA in *D. rerio* (CA2) since it also is an aquatic organism. However, *D. rerio* has seven cytosolic

isoforms (Gilmour and Perry 2009), whereas invertebrates typically have only one or maybe two cytosolic CAs (Culver and Morton 2015). Analysis of the other *D. rerio* cytosolic CA isoforms may turn up this shared conserved region in aquatic organism; therefore, further analysis is necessary. There are a number of potential similar examples of different conserved regions among the different taxa in this present study. For example, in CA5, the region between 250 – 300bp is conserved among the GPI-anchored CAs in all taxa, except in *Daphnia*. Does this difference have any functional significance that can impact fitness in *Daphnia* under different pH environments? This remains an open question, and clearly further research is warranted.

Looking further at evolutionary constraints among taxa, we assessed the similarity among aligned AA sequences for each isoform (Figure 3). While within the GPI-anchored CAs, we found that all the active, zinc-binding, substrate-binding, and di-sulfide binding sites were conserved, this was not the case in the cytosolic CAs. In the invertebrate cytosolic CAs, we found that the first active site (residue 86) had either a tyrosine (Y) or alanine (A) residue, rather than the histidine residue found among vertebrates. This histidine residue in vertebrates acts as a proton shuttle from the zinc-ion and is considered a rate-limiting step in the catalytic process (Esbaugh and Tufts 2006). Thus in vertebrates, those isoforms of cytosolic CAs that are missing these histidine residues have lower activity. Thus, this substitution of the histidine residue in invertebrates also likely has the consequence of lowering the activity level of these cytosolic enzymes. However, this supposition needs to be further studied with protein activity assays.

All other active, zinc-binding, and substrate-binding sites are conserved in the cytosolic CAs among taxa. There were several AA sequence motifs (QSP, GSEH, GSLTTPP, and GLAVLG) that were also conserved between the cytosolic and GPI-anchored CAs among all taxa. These motifs appear to be constrained because they are important in maintaining the conformation of the catalytic site of the enzyme (Le Roy et al. 2014). They are responsible for the regeneration of the active site and proton shuffling during catalytic activity (Silverman and Lindskog 1988). The cytosolic and GPI-anchored CAs each has a unique motif, GGPL and NNGH, respectively. In the GPI-anchored CAs, there is evidence that one of the asparagine (N) residues in the motif is the attachment site for the cleaved C-terminus that anchors the CA to the cell membrane (Hilvo et al. 2005; Weber and Pirow 2009).

The efficient regulation of acid-base homeostasis mediated by CAs is important in organisms. However, elucidating whether homeostasis is maintained by up or down regulation of the CA genes, or if AA sequence substitutions that subsequently tweak the protein structure, yet are expressed uniformly, regardless of environmental condition, is open for debate (Gilmour 2010). For instance, Pastoreková et al. (1997) looked at the differential expression of CA9 in the stomach of humans and rats, and found no differential expression across treatments, while Becker (unpublished data) used microarray technology and found no differential expression in any CA isoforms under a pH gradient in *D. pulicaria*. Still, other studies provide evidence for differential expression in CAs. For example Evans et al. (2013) showed that CA12 was differentially expressed in larval *Strongylocentrotus purpuratus* (purple sea urchin)

when exposed to low pH, while Lin et al. (2008) found differential expression in CA2 and CA15 in the gills of *Danio rerio* under differing pH gradients.

Here, we found evidence for both hypotheses (ubiquitous expression vs differential expression). We found that CA1 differentially expresses under a pH gradient, while CA2 and CA5 maintain relatively uniform expression levels regardless of pH. Of interest, we found that there were differences in expression among clones from our three study lakes that exhibit a pH gradient. The clone from Frenchman Lake (low pH) was found to significantly express CA1 differentially at high pH conditions, while for the Madison Lake clone, CA1 was significantly differentially expressed at low pH conditions. The control clone, from Hill Lake (medium pH), showed significant expression in CA1 at both low and high pH, which is what we would predict to occur (all things being equal) because CAs catalyze both carbonic acid and bicarbonate to regulate pH. In addition, under our experimental conditions, we found that the Frenchman Lake clone always had CA5 significantly up regulated in relation to the other clones, regardless of pH treatment. This provides support that the expression patterns in some of the CAs have evolved to local pH conditions (i.e., local adaptation).

Previous work (Culver and Acosta, *in review*) on the *D. pulicaria* populations from Frenchman, Hill, and Madison lakes showed that these populations were genetically differentiated based upon population genetics analysis using microsatellites. In addition, these authors showed that across a pH gradient, the Frenchman population had significantly greater survivorship at low pH than the Hill and Madison populations, while at higher pHs, the Hill and Madison Lake populations had significantly higher survivorships. This work lends support to our current study, specifically with regards to

the Frenchman clone, which has CA5 upregulated constitutively. This may be the mechanism by which local adaptation has occurred in the Frenchman population. Additionally while CA1 was significantly up regulated at high pH in the Frenchman clone, this did not translate into higher survivorship. The same relationship holds true for the Hill and Madison Lake clones, except at the low pH treatment, where CA1 is significantly upregulated with no corresponding increase in survivorship. Another avenue to investigate to detect possible signatures of local adaptation in these populations is to look at DNA sequences of the three CA isoforms among the three lakes for point substitutions that may be relevant to differential fitness in these populations (Culver and Morton, *in prep*.).

Elucidating the character of the CA genes allows us to make predictions on how *D. pulicaria* may evolve under a given set of environmental stressors (i.e,, pH). As the current trend of rising anthropogenic eutrophication owing to fertilizer runoff from increased agricultural production has demonstrated (Dodds, 2009), some lakes will be apt to become more eutrophic, followed by an increase in pH from the NO_3 and PO_4 inputs. Furthermore, the rising anthropogenic input of CO_2 or SO_2 to the atmosphere, which lowers the pH of aquatic systems as these compounds are sequestered in water bodies, is another contributing factor (Solomon et al., 2007) leading to variable and fluctuating pH environments in aquatic systems on a landscape-scale. Considering these anthropogenic alterations to pH in aquatic environments, it is important to understand the adaptive capabilities of organisms in response to these environmental perturbations. Therefore, elucidating the mechanisms for the evolutionary response of

organisms to changing pH conditions is therefore of the utmost importance, in order to predict the impact of these widespread changes in aquatic ecosystems.

Materials and Methods

Study sites

Madison Lake (44^{o} 11.549' N; 93^{o} 48.740' W) and Hill Lake (47° 0.741' N; 93° 35.845' W) are located in Minnesota and are two of 24 lakes under long-term monitoring as part of the Minnesota Sentinel Lakes Program (MSLP) administered by the Minnesota Pollution Control Administration in conjunction with the Minnesota Department of Natural Resources. Annual average pH of these lakes was determined from data collected from the MSLP over the past decade. Frenchman Lake (46°43.081' N; 80° 59.298' W) is located in the Greater Sudbury region of Ontario, Canada and is monitored as part of the Greater Sudbury Water Quality Program (GSWQP). Annual average pH of Frenchman Lake was determined from data collected from the GSWQP over the past decade. The mean (± 1 SD) annual pH values for the three lakes were as follows: Frenchman Lake (6.53 ± 0.50), Hill Lake (7.91 ± 0.52), and Madison Lake (8.63 ± 0.25).

Model organism

Two methods were used to collect *Daphnia* in our three study lakes: (1) we used a 158 uM mesh Wisconsin plankton net to take vertical tows through the full water column at the deepest part of Frenchman Lake in June 2011; and (2) for Madison and Hill Lakes, we obtained and hatched animals from diapausing (ephippial) eggs harvested from core

samples taken in 2011. The cores from Madison and Hill Lakes were taken using a 1.5m (6.93cm diameter) single drive Griffith sediment corer with Livingstone drive rods. The cores were sliced into 4cm or 2cm sections for Madison and Hill, respectively. Core sections were sifted through a series of sieves of 710μm, 425μm, and 300μm mesh sizes to collect *D. pulicaria* ephippia. Individual eggs were removed from their ephippial casings and placed in COMBO medium (Kilham et al., 1998) for hatching. Eggs were hatched under the following conditions; eggs were stored in the dark at 4^{o}C for two weeks and then placed under direct lighting at 20^{o}C for 24 hours until hatched. Once eggs hatched, they were reared at 20°C with indirect lighting. The eggs from Hill Lake were collected from the 0-2cm section of the sediment cores and were dated at AD 2006.7 ± 1.6 years at the bottom of the core section, while the eggs from Madison Lake were collected from the 0-4cm section and were dated at AD 2009.8 ± 1.3 years at the bottom of the core section. See Frisch et al. (2014) for more details related to sediment core sample methods, dating, and analyses.

Carbonic anhydrase sequencing

To obtain DNA sequences of CA1, CA2, and CA5 for *D.pulicaria*, gene-specific forward and reverse primers were designed using the *D. pulex* sequences as a template (Table S1; Figure 1). The *D. pulex* templates were obtained from the *Daphnia* Genomics Consortium (http://wFleabase.org; Gene IDs, CA1 – DappuDraft_222096, CA2 – DappuDraft_222141, and CA5 – DappuDraft_317362). DNA for *D. pulicaria* was isolated from a single clone that was hatched from a resting egg recovered from the 0-4cm sediment layer of Madison Lake following the CTAB method (Hillis et al. 1990).

PCR was performed using IQ^{TM} Supermix 2X (Bio-Rad Laboratories, Hercules, CA) in a 25µL reaction (12.5µL Supermix 2X, 0.2µL DMSO, 2µL forward primer (10µM), 2µL reverse primer (10µM), 7.3 µL ultra-pure water, and 2µL of template DNA) with thermocycler settings as per Table S1. Since all primers, except for primer CA2 F1, fell within the gene of question, single primer PCR needed to be performed to sequence the ends of the genes. Gene-specific single primers were designed as with the forward and reverse gene specific primers and amplified in 25µL reaction (12.0µL Supermix 2X, 0.3µL DMSO, 4µL forward primer (10µM), 5.3 µL ultra-pure water, and 3µL of template DNA) with thermocycler settings as per Table S1 (Table S1; Figure 1). Since the single primer PCR contained more than one band, bands were excised from the agarose (electrophoresis) gel and Wizard SV Gel and PCR Clean-Up System (Promega Corp., Madison, WI) was used according to manufacturer's instructions to extract PCR products from the gel. PCR products were purified and sequenced using an ABI-3130XL DNA analyzer (Applied Biosystems) in eight forward and reverse replicates to minimize sequencing error. Consensus sequences for *D. pulicaria* CA1, CA2, and CA5 were constructed using tools on MEGA 5.0 (Tamura et al. 2011). The *D. pulicaria* CA gene architecture was characterized by comparison with annotated *D. pulex* CA sequences and amino acid sequences were inferred using MEGA 5.0.

Comparison of Daphnia pulicaria Carbonic Anhydrases with other taxa

Cytosolic and GPI-anchored α-CAs sequences for *Calinectes sapidus*, *Anopheles gambiae*, *Danio rerio*, and *Homo sapiens* were obtained from the National Center for Biotechnology Information (NCBI; www.ncbi.nih.gov). A key word search for each

isoform (*Hs*CA2 - cytosolic and *Hs*CA9 – GPI-anchored) for *Homo sapiens* was performed; both cDNA and amino acid sequences for each isoform were obtained. For each *H. sapiens* isoform, a BLAST search was performed using the BLASTN algorithm (Gertz et al. 2006) with default settings from NCBI. For each taxon, the sequence with the lowest e-value was selected for analyses. Both the cDNA and amino acid sequence were retained for each taxon and isoform. Cytosolic α-CAs were uploaded into CLUSTALX 2.1 (Larkin et al. 2007) and a multi-sequence alignment was run with iterations after each alignment step for both DNA and amino acid sequences. The process was repeated with the GPI-anchored sequences.

Aligned cDNA sequences for CA1, CA2, and CA5 for all taxa were analyzed using the bioinformatics program Mauve (Darling et al. 2004) to determine whether the genes were homologous and to look for common areas of conservation among cDNA sequences for each taxon and isoform. Aligned AA sequences were used to determine common conserved motifs among taxa and isoforms; including active, zinc-binding, substrate-binding, and di-sulfide bond sites.

Differential expression experiment

To test for differential expression among the α-CA isoforms and clones in *D. pulicaria* a quantitative Real Time (qRT)-PCR experiment was performed. One clone was selected from each population (lake). For Hill Lake and Madison Lake, the individual clones used for the experiment were initially established form hatched eggs that were recovered form sediment cores from Hill Lake (0-2cm core layer) and Madison Lake (0-4cm core layer) (see *Model organism* section). We used animals from the shallower

portion of the cores, because they represent more recent populations and were from the population that experienced the pH regimes used in this study. For Frenchman Lake, the clone was raised from animals initially harvested from the water column in 2011 (four years older than the oldest dated eggs from Hill Lake and Madison Lake). During this five year period, the pH values of the individual lakes were relatively stable (see *Study sites* section). All animals were raised under the same laboratory conditions using (~ pH 7.5) COMBO medium (Kilham et al. 1998).

For each clone, a set of 90 gravid animals were used to set up as stem grandmothers. Once the stem grandmothers had released their clutches, 90 neonates from each clone were harvested to serve as stem mothers. All animals were fed 0.5mL of approximately 1-2mg C L^{-1} of chemostatically-cultured green algae, *Scenedesmus acutus*, daily to ensure that the *Daphnia* were not food limited. Neonates from the third clutch of the stem mothers were used for the experiment, since previous work has shown that third-clutch offspring are more robust than first or second-clutch neonates (Glazier 1992). In addition, using the grand-daughter generation reduces maternal effects (Lampert 2011).

In a common garden experimental design, three treatments of COMBO medium were adjusted using 1M NaOH or 1M HCl (coarse adjustment), to the following pHs: 6.0, 7.5, and 9.0. Thirty-five neonates from each clone were placed in 250mL of pH-adjusted COMBO medium per treatment. For each clone, four replicate 250mL jars were used per treatment. The pH of the treatments was measured at the beginning of the experiment, again at the 12-hr time point, and then finally at the conclusion (24-hrs) of the experiment using a Milwaukee Instruments PH56 pH meter (the pH meter was

calibrated before taking measurements). pH values were adjusted with 0.1M NaOH or 0.1M HCl (fine adjustment), if needed. At the end of the experiment (24hrs) animals were collected and pooled for each replicate. The pooled animals were crushed in 1mL of TRIzol (Life Technologies, Carlsbad, CA) and stored at -80° C until ready for RNA extraction.

Total RNA was isolated using a modified RNeasy Mini Kit (Qiagen, Hilden, Germany). RNA was precipitated from TRIzol by adding 200µL of chloroform and spinning in a centrifuge at 11,600rcf for 15 minutes. The upper aqueous phase was collected (~500µL) and 250µL of 100% ethanol was added. Total RNA was extracted from the resultant solution following the manufacturer's instruction for the RNeasy Mini Kit. Total RNA (quantity > 100 ng/µL and quality (260/280 value) > 2.00) was assessed using a Nanodrop spectrophotometer. QuantiTect Reverse Transcription Kit (Qiagen) was used according to manufacturer's instructions to synthesize cDNA using 1ng of total RNA. The glyceradehyde-3-phosphate dehydrogenase (GAPDH; Accession No.FJ668125) and actin (Accession No. AJ245732) genes were used as endogenous controls. However, it was discovered that actin was differentially expressing under the treatments; therefore the analysis was performed with GAPDH as the only control. Gene-specific primers were designed using the software Primer3 v0.4.0 (Table S2; Untergasser et al. 2012). A BioRad CFX Connect Real-Time System (Bio-Rad Corp) was used to run qRT-PCR and data were analyzed using the method described in Pfaffl (2001). Each 15 µL reaction contained 7.5µL SYBR green fastmix, 1µL of the forward and reverse primers, 4.5µL of molecular grade water, and 2µL of cDNA template (Life Technologies) and each reaction was run in triplicate.

Acknowledgements

This study was supported by the U.S. National Science Foundation (NSF-IOS-OEI) grants to L. J. Weider (#0924289 and #1256881), the U. of Oklahoma Department of Biology Adams Scholarship and the U. of Oklahoma Biological Station Summer Research Grant to BWC. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We thank the U. of Minnesota LacCORE staff (in particular R. O'Grady) at the University of Minnesota, for assistance with coring and providing *D. pulicaria* samples from Madison Lake, MN. A. Harris facilitated sequencing of *D. pulicaria* enzymes. We thank K. Milette (McGill University) for providing *Daphnia* samples. We also thank L.J. Weider for his advice, mentorship, and comments on this manuscript. In addition, we thank R. Broughton, K.D. Hambright, I. Schlupp, R. Cichewicz, and for their input. We thank the staff of the Holt Laboratory: K. Gaylor and Z. Myers for the assistance with the qRT-PCR. In addition, I would like to acknowledge F. Acosta for his assistance with Mauve. This manuscript constitutes a portion of BWC's doctoral dissertation at the U. of Oklahoma. There are no known conflicts of interests associated with this article and its authors.

References

Aspatwar A, Martti MEE, Parkkila S. 2010. Phylogeny and expression of carbonic anhydrase-related proteins. BMC Mol Biol. 11:1-19.

Badger MR, and Price GD. 1994. The role of carbonic anhydrase in photosynthesis. Annu Rev Plant Physiol Plant Mol Biol. 45:369-392.

Baird DJ and Barata C. 1998 Genetic variation in the response of *Daphnia* to toxic substances: Implications for risk assessment. In Genetics and Ecotoxicology. Taylor & Francis, Inc, Ann Arbor, MI. 207-218.

Cáceres M, Lachuer J, Zapala MA, Redmond JC, Kudo L, Geschwind DH, Lockhart DJ, Preuss TM, and Barlow C. 2003. Elevated gene expression levels distinguish human from non-human primate brains. PNAS. 100: 13030-13035.

Chorev M, and Carmel L. 2012. The function of introns. *Frontier in Genetics*, 3:1-15.

Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, Tokishita S, Aerts A,

Arnold GJ, Basu MK, et al. 2011. The Ecoresponsive Genome of *Daphnia pulex*. Science. 331: 555-561.

Cooper PD. 1994. Mechanisms of Hemolymph Acid-Base Regulation in Aquatic insects. Physiol Zool. 67(1):29-53.

Coors A, Vanoverbeke J, DeBie T, and deMeester L. 2009. Land use, genetic diversity and toxicant tolerance in natural populations of *Daphnia magna*. Aquatic Toxicology. 95: 71-79.

Culver BW and Morton PK, 2015. The evolutionary history of Daphniid α-carbonic

anhydrase within Animalia. International Journal of Evolutionary Biology. 2015: 1-11.

Culver BW and Acosta F. *In Review*. Population genetic structure of *Daphnia pulicaria* across a pH gradient in three North American lakes. Hydrobiologia.

Darling ACE, Mau B, Blattner FR, and Perna NT. 2004. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. Genome Res. 14:1394-1403.

Derry AM, and Arnott SE. 2007. Adaptive reversals in acid tolerance in copepods from lakes recovering from historical stress. Ecological Applications. 17: 1116-1126.

Dodds WK, Bouska WW, Eitzmann JL, Pilger TJ, Pitts KL, Riley AJ, Schloesser JT, and Thornbrugh DJ. 2009. Eutrophication of U.S. Freshwaters: Analysis of Potential Economic Damages. Environmental Science & Technology 43: 12-19.

Dodson S. 1988. The ecological role of chemical stimuli for zooplankton: Predator-avoidance behavior in *Daphnia*. Limnology & Oceanography. 33: 1431-1439.

Dudycha JL and Tessier AJ. 1999. Natural generic variation of life span, reproduction, and juvenile growth in *Daphnia*. Evolution. 53: 1744-1756.

Eicher EM, Stern RH, Womack JE, Davisson MT, Roderick TH, and Reynolds SC. 1976. Evolution of Mammalian Carbonic Anhydrase Loci by Tandem Duplication: Close Linkage of *Car-1* and *Car-2* of the centromere region of Chromosome 3 of the Mouse. Biocem Gen. 14:651-660.

Eppinger RG, Briggs PH, Dusel-Bacon C, Giles SA, Gough LP, Hammarstrom JM, and Hubbard BE. 2007. Environmental geochemistry at Red Mountain, an unmined

volcanogenic massive sulphide deposit in the Bonnifield district, Alaska Range, east-central Alaska. Geochemistry: Exploration,
Environment, Analysis. 7: 207-223.

Erlandsson M, Cory N, Fölster J, Köhler S, Laudon H, Weyhenmeyer GA, and Bishop, K. 2011. Increasing dissolved organic carbon redefines the extent of surface water acidification and helps resolve a classic controversy. BioScience 61: 614-618.

Esbaugh A and Tufts BL. 2006. The structure and function of carbonic anhydrase isozymes in the respiratory system of vertebrates. Respiritory Physiology & Neurobiology. 154: 185-198.

Evans TG, Chan F, Menge BA, and Hofman GE. 2013, Transcriptomic responses to ocean acidification in larval sea urchins from a naturally variable pH environment. Molecular Ecology: 22, 1609-1625.

Ezoe Y, Lin CH, Noto M., Watanabe Y, and Yoshimura, K. 2002. Evolution of water chemistry in natural acidic environments in Yangmingshan, Taiwan. Journal of Environmental Monitoring 4: 533-540.

Fasseas MK, Tsikou D, Flemetakis E, Katinakis P. 2010. Molecular and biochemical analysis of the class β class carbonic anhydrases in *Caenorhabitis elegans*. Mol Biol Rep. 37:2941-2950.

Frisch D, Morton PK, Chowdhury PR, Culver BW, Colbourne JK, Weider LW, and Jeyasingh PD. 2014. A millennial-scale chronicle of evolutionary responses to cultural eutrophication in *Daphnia*. Ecology Letters. 17: 360-368.

Georgalis T, Perry SF, Gilmour KM. 2006. The role of branchial carbonic anhydrase in acid-base regulation in rainbow trout (*Oncorhynchus mykiss*). J exp Biol. 209:518-530.

Gertz EM, Yu YK, Agarwala R, Schäffer AA, Altschul SF. 2006. Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. BMC Biol. 4:1-14.

Gilmour KM, Perry SF. 2009. Carbonic anhydrase and acid-base regulation in fish. J exp Biol. 212:1647-1661.

Gilmour KM. 2010. Perspectives on carbonic anhydrase. Comparative Biochemistry and PhysiologyA: Molecular & Integrative Physiology. 157: 193-197.

Havens KE, Yan ND, and Keller W. 1993. Lake acidification: effects on crustacean zooplankton populations. Environ Sci Technol. 27: 1621-1624.

Hebert PDN. 1978. The population biology of *Daphnia* (Crustacea, Daphnidae). Biological Reviews. 53: 387-426.

Henry RP, Cameron JN. 1983. The role of Carbonic Anhydrase in respiration, ion regulation and acid-base balance in the aquatic crab *Callinectes Sapidus* and the terrestrial crab *Gecarnicus lateralis*. J exp Biol. 103:205-223.

Henry RP. 1984. The Role of Carbonic Anhydrase in Blood Ion and Acid-Base Regulation. Amer Zool. 24:241-251.

Hewett-Emmett D, Tashian RE. 1996. Functional Diversity, Conservation, and Convergence in the evolution of the α-, β-, and γ-Carbonic Anhydrase Gene Family. Mol Phylo Evol. 5:50-77.

Hillis DM, Moritz C, and Mable BK. 1990. Molecular Systematics. Sinauer &

Associates, Sunderland, MA, USA.

Hilvo M, Tolvanen M, Clark A, Shen B, Shah GN, Waheed A, Halmi P, Hänninen M, Hämäläinen JM, Vihinen M, et al. 2005. Charaterization of CA XV, a new GPI-anchored form of carbonic anhydrase. Biochem J. 392: 83-92.

Hurd CL, Cornwall CE, Currie K, Hepburn CD, McGraw CM, Hunter KA, and Boyd PW. 2011. Metabolically induced pH fluctuations by some coastal calcifiers exceed projected 22nd century ocean acidification: a mechanism for differential susceptibility? Global Change Biology 17: 3254-3262.

Keller W, Yan ND, Holtze KE, Pitblado JR. 1990. Inferred effects of lake acidification on *Daphnia galeata mendotae*. Environ Sci Technol. 24: 1259-1261.

Kilham SS, Kreeger DA, Lynn SG, Goulden CE, and Herrara L. 1998. COMBO: a defined freshwater culture medium for algae and zooplankton. Hydrobiologia 377: 147-149.

Lampert W. 1993. Phenotypic plasticity of the size at first reproduction in *Daphnia*: The importance of maternal size. Ecology. 74: 1455-1466.

Lampert W. 2011. Daphnia: development of a model organism in ecology and evolution. Oldendorf/Luhe: International Ecological Institute.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947-2948.

Larsen PF, Nielsen EE, Williams TD, and Loeschcke V. 2008. Intraspecific variation in expression of candidate genes for osmoregulation, heme biosynthesis and stress resistance suggests local adaptation in European flounder (*Platichtys flesus*). Heredity. 101: 247-259.

Le Roy N, Jackson DJ, Marie B, Ramon-Silva P, Marin F. 2014. The evolution of metazoan α-carbonic anhydrases and their roles in calcium carbonate biomineralization. Frontiers in Zoology. 11:1-16.

Lin TY, Liao BK, Horng JL, Yan JJ, Hsiao CD, Hwang PP. 2008. Carbonic anhydrase 2-like a and 15a are involved in acid-base regulation and Na^+ uptake in zebrafish H^+-ATPase-rich cells. Am J Physiol Cell Physiol. 294:C1250-C1260.

Lindskog S and Silverman DN. 2000. The catalytic mechanism of mammalian carbonic anhydrases. In: Chegwidden WR, Carter ND, Edwards YH, editors. The Carbonic Anhydrases: New Horizons. Basel: Birkhäuser Verlag. P. 175-195.

Linser PJ, Smith KE, Seron TJ, Oviedo MN. 2009. Carbonic anhydrases and anion transport in mosquito midgut pH regulation. J Exp Biol. 212: 1662-1671.

Loaring JM and Hebert PDN. 1981. Ecological differences among clones of *D. pulex* Leydig. Oecologia 51: 162-168.

Lu Z, Streets DG, Zhang Q, Wang S, Carmichael GR, Cheng F, Wei C, Chin M, Diehl T, and Tan Q. 2010. Sulfur dioxide emissions in China and sulfur trends in East Asia since 2000. Atmospheric Chemistry & Physics 10: 6311-6331.

Martins CMG, Almeida DV, Marins LFF, and Bianchini, A. 2010. mRNA expression and activity of ion-transporting proteins in gills of the blue crab *Callinectes sapidus*: effects of waterborne copper. Environmental Toxicology and Chemistry 30: 206-211.

Miner BE, DeMeester L, Pfrender ME, Lampert W, and Hairston Jr N.G. 2012. Linking genes to communities and ecosystems: *Daphnia* as an ecogenomic model. Proceedings of the Royal Society B, 1-10.

Moya A, Huisman L, Ball EE, Hayward DC, Grasso LC, Chua CM, Woo HN, Gattuso JP, Forêt S, and Miller DJ. 2012. Whole transcriptome analysis of the coral *Acropora millepora* reveals complex responses to CO_2-driven acidification during the initiation of calcification. Molecular Ecology 21:2440-2454.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

O'Brien WJ, and deNoyelles F, JR. 1972 . Photosynthetically elevated pH as a factor in zooplankton mortality in nutrient enriched ponds. Ecology 53: 605-614.

Pastoreková S, Parkkila S, Parkkila A-K, Opavský R, Zelník V, Saarnio J, and Pastorek J. 1997. Carbonic Anhydrase IX, MN/CA IX: Analysis of Stomach Complementary DNA sequence and Expression in Human and Rat Alimentary Tracts. Gastroenterology. 112: 398-408.

Peterson RE, Tu C, and Linser PJ. 1997. Isolation and Characterization of a Carbonic Anhydrase Homologue from Zebrafish (*Danio rerio*). J Mol Evol. 44: 432-439.

Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucl. Acids Res. 29:2002-2007

Pilka ES, Kochan G, Oppermann U, and Yue WW. 2012. Crtystal structure of the secretory isozyme carbonic anhydrase CA VI: Implications for biological assembly and inhibitor development. Biochem Biophys Res Comm. 419: 485-489.

Reece-Hoyes JS, Keenan ID, and Isaacs HV. 2002. Cloning and Expression of the Cdx Family From the Frog *Xenopus tropicalis*. Developmental Dynamics. 223: 134-140.

Satake K, Oyagi A, and Iwao Y. 1995. Natural acidification of lakes and rivers in

Japan: The ecosystem of Lake Usoriko (ph 3.4-3.8). Water Air & Soil Pollution 85: 511-516.

Silverman D and Lindskog S. 1988. The catalytic mechanism of carbonic anhydrase implications of a rate-limiting protolysis of water. Accounts Chem Res. 21: 30-36.

Smith KE, VanEkeris LA, and Linser PJ. 2007. Cloning and characterization of AgCA9, a novel α-carbonic anhydrase from *Anopheles gambiae* Giles *sensu stricto* (Diptera:Culicidae) larvae. J Exp Biol. 210: 3919-3930.

So AK, Espie GS. 2005. Cyanobacterial carbonic anhydrases. Can J Bot. 83:721-734.

Somero GN. 2009. The physiology of climate change: how potentials for acclimatization and genetic adaptation will determine 'winners' and 'loser'. J Exp Biol. 213: 912-920.

Søndergaard M, Jeppesen E, and Jensen JP. 2005. Pond or Lake: does it make a difference? Archiv Für Hydrobiologie. 162: 143-165.

Stollewerk A. 2010. The water flea *Daphnia* – a 'new' model system for ecology and evolution? Journal of Biology 9: 1-4.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol. 28(10):2731-2739.

Teitelbaum SL. 2000. Bone Resorption by Osteoclasts. *Science*. 289:1504-1508.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, RemmM, and Rozen SG.

2012. Primer 3 – new capabilities and interfaces. Nucleic Acid Research. 40: e115.

Urabe J and Sterner W. 2001. Contrasting effects of different types of resource depletion on life-history traits in *Daphnia*. Functional Ecology. 15: 165-174.

Weber AK, Pirow R. 2009. Physiological responses of *Daphnia pulex* to acid stress. BMC Physiol. 9: 1-25.

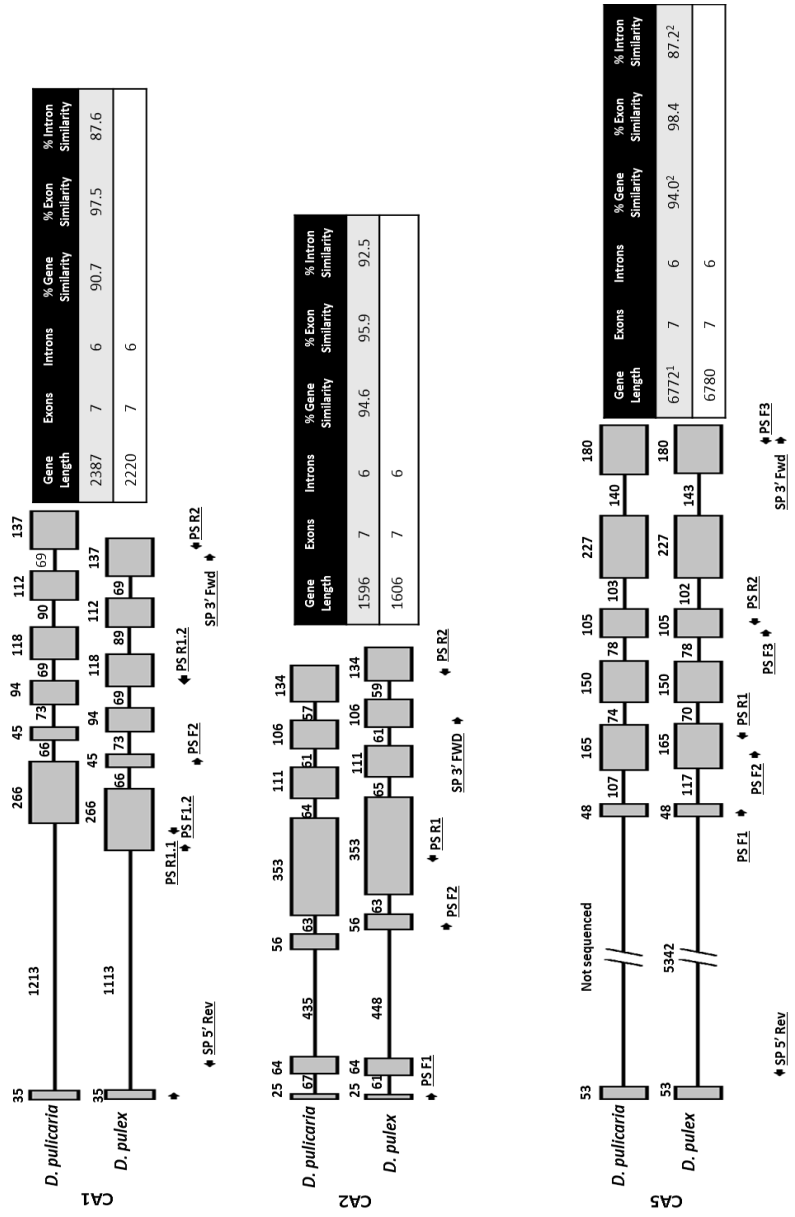
Wetzel RG. 2001. Limnology: Lake and River Systems 3rd edition. Elsevier Academic Press, London.

Witten DM and Tibshirani R. 2007.A comparison of fold-change and t-statistic for microarray data analysis. Stanford University. 1-13.

Xu Y, Feng L, Jeffrey PD, Shi Y, Morel FMM. 2008. Structure and metal exchange in the cadmium carbonic anhydrase of marine diatoms. Nature 452: 56-62.

Tables and Figures

Figure 1. Comparison of genetic architecture (5' end to 3' end) for CA1, CA2, & CA5 for *Daphnia pulicaria* and *D. pulex* using line and box diagrams with exon and intron lengths. Primer locations (Table S11) are designated beneath the line and box diagram. All *Daphnia* CA isoforms have seven exons and six introns which are consistent with CAs among other taxa.



	Gene Length	Exons	Introns	% Gene Similarity	% Exon Similarity	% Intron Similarity
CA1	2387	7	6	90.7	97.5	87.6
	2220	7	6			
CA2	1596	7	6	94.6	95.9	92.5
	1606	7	6			
CA5	6772[1]	7	6	94.0[2]	98.4	87.2[2]
	6780	7	6			

Figure 2. Comparison of CA1, CA2, & CA5 cDNA sequences among various taxa using the program Mauve (Darling et al. 2004). Bar plots represent how similar the DNA sequences are when compared to *Daphnia pulicaria* CA1 (A.), and *D. pulicaria* CA5 (B.). The higher the histogram (plot) peaks, the more similar the sequences are. The colors represent areas of conservation among taxa. The x-axis indicates number of base pairs (5' end to 3' end).

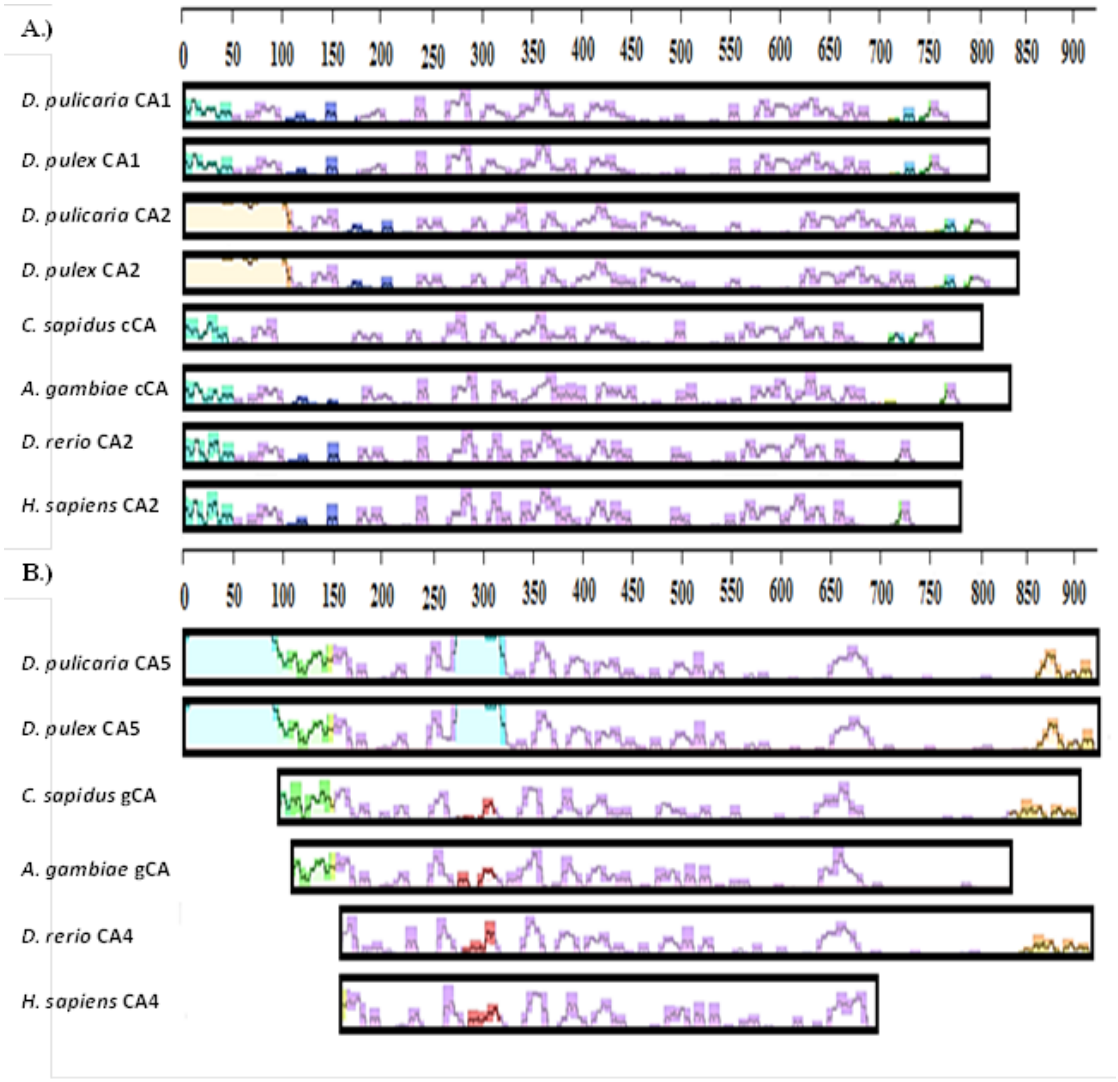


Figure 3. Alignment of the amino acid sequences among *Daphnia* and four other taxa: *C. sapidus*, *A. gambiae*, *D. rerio*, and *H. sapiens*. Amino Acids were aligned using ClustalX2 (Larkin et al. 2007). The green boxes indicate active sites, purple boxes indicate zinc-binding sites, red boxes indicate substrate-binding sites, blue boxes indicate disulfide bonds, and the gray boxes indicate conserved residues. Conserved motifs are highlighted by a black box around the motif. Cytosolic CAs are represented by panel (A) and the GPI-anchored CAs are shown in panel (B).

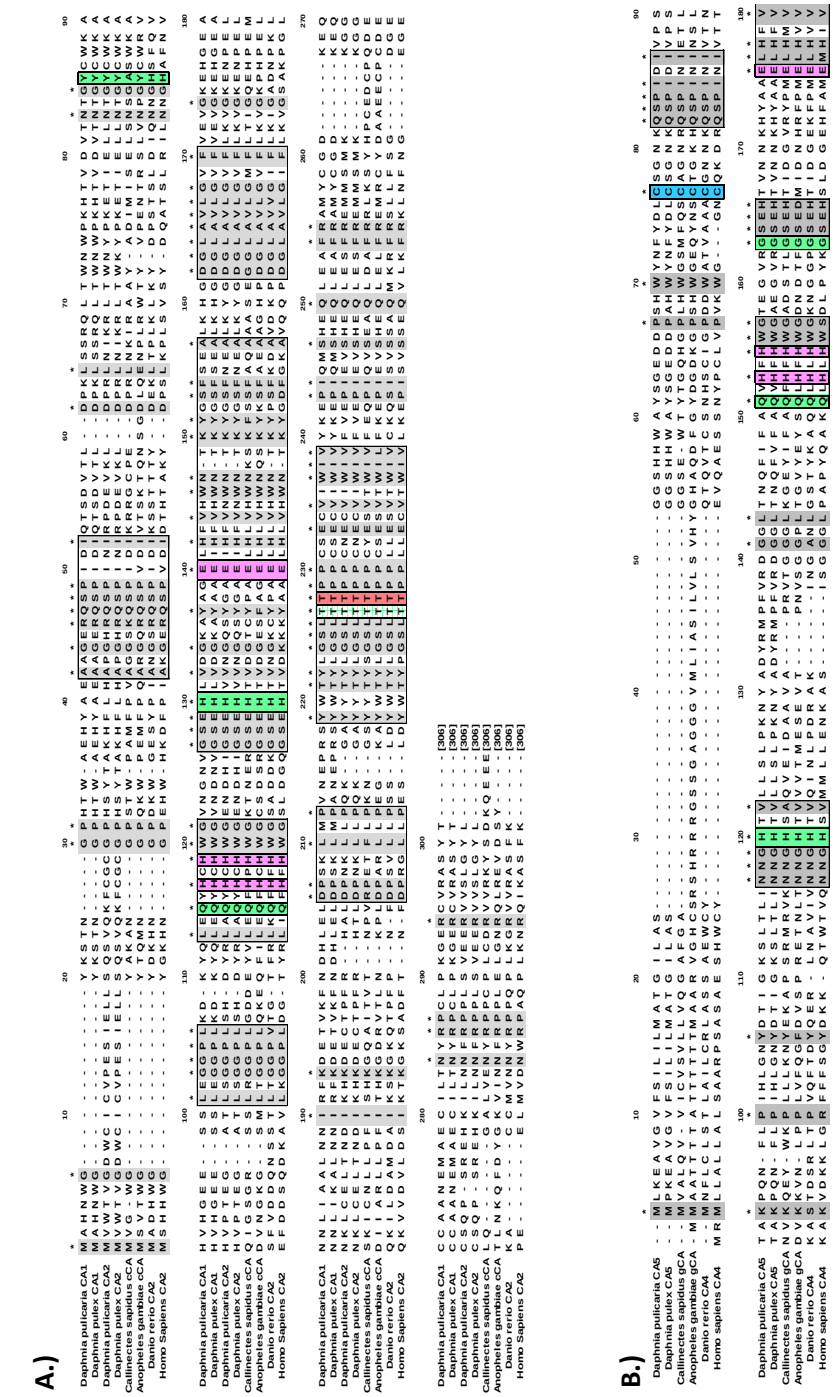
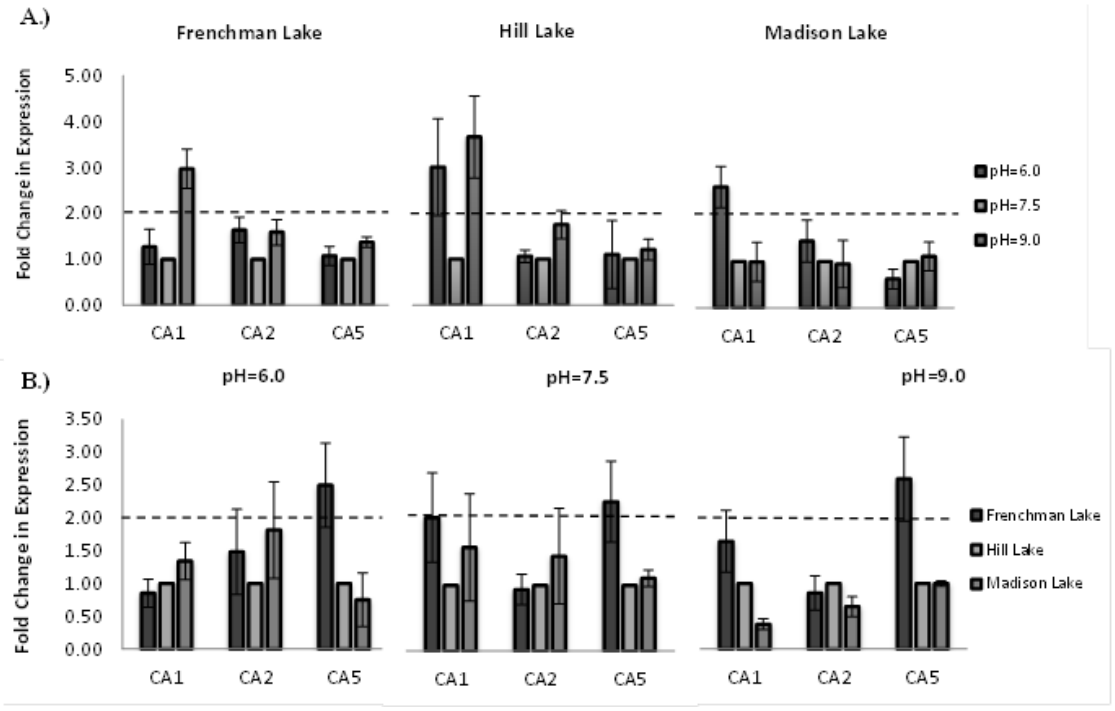


Figure 4. Results of qRT-PCR of three clones from three North American lakes that have a pH gradient. Frenchman Lake, CN is a low pH lake, Hill Lake, MN is a medium pH lake (control), and Madison Lake, MN is a high pH lake. qRT-PCR was performed using a BioRad CFX Connect Real-Time System and analyzed using the Pfaffl method (Pfaffl 2001). Panel (A) represents the fold change in relative expression of each CA isoform among pH treatment for each clone with pH= 7.5 being the control treatment, while panel (B) represents the fold change in relative expression for each CA isoform among clones for each treatment with the Hill Lake clone acting as the control. Fold changes (dotted line) above 2.0 are deemed to be functionally significant (Witten and Tibshirani 2007).



Chapter Four

Evolutionary response of Carbonic Anhydrase (isoforms 1, 2, and 5) in *Daphnia pulicaria* populations across a pH gradient in three North American lakes

Billy W. Culver

and

Francisco Acosta

Abstract

Understanding how the environment affects standing genetic variation is an important goal in biology. Numerous studies have tried to discern the processes which affect genetic variation: neutral processes (i.e. genetic drift) versus natural selection (i.e. positive, purifying selection). Here, we predict that in three North American *D. pulicaria* populations from local pH environments, we will see sequence divergence across three α-carbonic anhydrase (CA) loci. We predict that there will be evidence of selection at variants of these three loci and those specific genotypes will convey differential survivorship across a pH gradient. We sequenced the three CA loci for 15 individuals from each population for population genetic analysis. Further, five distinct CA genotypes were chosen for a common garden pH survival experiment to determine differential survival across a pH gradient. We found that the three populations were divergent across the three CA loci and that there were high levels of predicted heterozygosity. In addition we found evidence for balancing selection, which supports our finding of high heterozygosity. We also found that one particular genotype (represented by several clonal lineages from two of the three lakes) had significantly higher survivorship at low pH, with low survivorship at high pH. In contrast, other genotypes were found to have significantly lower survivorship at low pH, but had higher survivorship at high pH indicating clear genotype x environment interactions. Future research is warranted to further characterize the functional relationship of CA gene polymorphisms and effects on organism fitness in this system.

Introduction

Acidification and alkalization of aquatic ecosystems has increasingly become a concern that can negatively impact ecosystem "health". Such impacts on pH are commonly the direct consequence of human activity. While carbon dioxide emissions (Moya et al., 2012, Evans et al. 2013) and mining activities (Derry and Arnott 2007; Martins et al. 2010) often result in acidification, other human impacts can lead to alkalization, such as cultural eutrophication (O'Brien & deNoyelles, 1972). These changes in pH can act as strong selective forces that affect the genetic variation and physiological responses of natural aquatic populations. Analyses of genetic variation of populations using DNA sequences have been used in a variety of studies to elucidate patterns of selection pressures (neutral, positive, purifying, and balancing selection) on the evolutionary dynamics of natural populations. Many studies have shown that there are disproportionate levels of variation at non-coding or silent sites with regard to non-synonymous sites (or amino acid substitutions), implying that most amino acid substitutions are deleterious (Charlesworth 2010, Crease et al. 2011). However, patterns that imply neutral processes are often found too (Avise 2004). In contrast, isoform variants at specific gene loci have been found that affect fitness across different environmental gradients, such as alcohol dehydrogenase and glucose-6- phosphate dehydrogenase in *Drosophila melanogaster* and phosphoglucose isomerase in butterflies (Eanes 1999). Elucidating the adaptive significance of polymorphisms in DNA sequences at specific gene loci is a major goal of molecular evolutionary biologists.

The enzyme α-carbonic anhydrase (CA) catalyzes the reversible hydration/dehydration reaction: $CO_2 + H_2O \Leftrightarrow H_2CO_3 \Leftrightarrow HCO_3^- + H^+$ and belongs to a class of zinc-metallo-enzymes that has been shown to be involved in pH homeostasis in many organisms. While there are a number of studies that show local adaptation in different populations that inhabit ecosystems with pH gradients, such as purple sea urchins (*Strongylocentrotus purpuratus*) that showed differential transcriptomic and tolerance to differing pH environments (Evans et al 2013), there is a general lack of research looking at selection on α-CA isoform variants across taxa and pH gradients. A few studies have shown selection on α-CA isoform variants in vertebrates; for example, Hunt and Fierke (1997) examined polymorphisms at amino acid residues of human CAII (a specialized cytosolic α-CA) and found that under different pH conditions certain CAII variants conferred a selective advantage. However, there is an even greater scarcity of research on selection in α-CA genes, particularly in invertebrates.

Here, we investigated variants in three CA isoforms (CA 1, 2, and 5) from three lake populations (Frenchman Lake, Ontario, CN; Hill Lake, MN; and Madison Lake, MN) of the freshwater zooplankter, *Daphnia pulicaria*. These particular lakes have significantly different pH regimes and there is evidence in these populations of adaptation to local pH conditions (Culver and Acosta, *in review*). *Daphnia pulicaria* is a keystone freshwater microcrustacean that is a model organism (Stollewerk 2010, Colbourne et al 2011, Miner et al 2012) and that has a number of traits that are invaluable for genetics studies: (i) *Daphnia* exhibit both asexual and sexual reproduction (i.e., termed cyclical parthenogenesis) which allows us to control for the genetic background of clonal lineages and look at heritability in sexual (recombinant)

lineages; (ii) they tend to have extremely large populations (millions to billions) – this characteristic reduces that impact of random evolutionary forces (such as genetic drift) on the genetic composition of these populations (Hebert 1978, Lambert 2011), and (iii) populations of *Daphnia* contain substantial genetic variation that can be easily interpreted in an ecological context (Baird and Barata 1998, Dudycha and Tessier 1999, Coors et al. 2009). In addition, the genome of *D. pulex*, a sister species to *D.pulicaria*, has also been sequenced and annotated (Colbourne et al. 2011), which provides an important tool in genetic studies of this genus.

While there are thirty known α-CA isoforms in *D. pulicaria*, the three isoforms in this study (CA1, 2, and 5) have been previously studied (Weber and Pirow 2009, Culver and Morton 2015) and have been shown to be homologous to α-CAs that have been implicated in acid-base regulation in other aquatic organisms (e.g. decapods (Henry and Cameron 1983, Henry 1984) and aquatic insects (Cooper 1994)). In addition, these isoforms of α-CA have been characterized in a previous study (Culver and Morton, *in review*). CA1 and CA2 were determined to be cytosolic CAs and lacked one of the three histidine residues that act as a proton shuttle from the zinc ion (Esbaugh and Tufts 2006). CA5 is a gylcophosphatidylinositol (GPI)-anchored enzyme that is attached to the interstitial surface of the cell. Furthermore, experiments have shown differential expression of CA1 across a pH gradient and that CA5 in one population (Frenchman Lake) was always differentially expressed (up-regulated) compared to the other two populations across all pH treatments (Culver and Morton, *in review*).

In this study, we analyzed sequence variation in three CA isoforms across three *D. pulicaria* populations from lakes that differ in pH regimes in order to: (i) unveil the

evolutionary history of these three genes in the context of population differentiation; (ii) elucidate if there is selection acting on the variation of each CA locus among the three populations; and (iii) determine whether specific *Daphnia* CA genotypes exhibit a fitness (measured as survivorship) advantage across a pH gradient.

Materials and Methods

Study sites

Our study sites included Madison Lake (44^{o} 11.549' N; 93^{o} 48.740' W) and Hill Lake (47° 0.741' N; 93° 35.845' W), which are located in Minnesota and are two of 24 lakes under long-term monitoring as part of the Minnesota Sentinel Lakes Program (MSLP) administered by the Minnesota Pollution Control Administration in conjunction with the Minnesota Department of Natural Resources. Annual average pH of these lakes was determined from data collected from the MSLP over the past decade. Our third lake was Frenchman Lake (46°43.081' N; 80° 59.298' W) which is located in the Greater Sudbury region of Ontario, Canada and is monitored as part of the Greater Sudbury Water Quality Program (GSWQP). Annual average pH of Frenchman Lake was determined from data collected from the GSWQP over the past decade. The mean (± 1 SD) annual pH values for the three lakes was previously determined (Culver and Acosta, *in review*) and were as follows: Frenchman Lake (6.53 ± 0.50), Hill Lake (7.91 ± 0.52), and Madison Lake (8.63 ± 0.25).

Model organism

Two methods were used to collect *D. pulicaria* in our three study lakes: (1) we used a 158 uM mesh Wisconsin plankton net to take vertical tows through the full water

column at the deepest part of the lakes in July 2011 for Madison Lake; in July 2010, 2011, and 2014 for Hill Lake; and in June 2011 and 2013 for Frenchman Lake, and (2) for Madison and Hill Lakes, we obtained and hatched animals from diapausing (ephippial) eggs harvested from sediment core samples taken in July 2010 and 2011. The cores from Madison and Hill Lakes were taken in duplicate using a 1.5m (6.93cm diameter) single drive Griffith sediment corer with Livingstone drive rods. The cores were sliced into 4cm or 2cm sections for Madison and Hill, respectively. Core sections were sifted through a series of sieves of 710μm, 425μm, and 300μm mesh sizes to collect *D. pulicaria* ephippia. Individual eggs were removed from their ephippial casings and placed in COMBO medium (Kilham et al., 1998) for hatching. Eggs were hatched under the following conditions; eggs were stored in the dark at 4^{o}C for two weeks and then placed under direct lighting at 20^{o}C for 24 hours until hatched. Once eggs hatched, they were reared at 20°C with indirect lighting. The eggs from Hill Lake were collected from the 0-2cm section of the sediment cores and were dated at AD 2006.7 ± 1.6 years at the bottom of the core section, while the eggs from Madison Lake were collected from the 0-4cm, 4-8cm, and 8-12cm sections and the oldest section was dated at AD 2009.8 ± 1.3 years at the bottom of the core section. See Frisch et al. (2014) for more details related to sediment core sample methods, dating, and analyses.

Carbonic anhydrase sequencing and genotyping

DNA was extracted individually from 15 adult *D. pulicaria* from each population using the CTAB method (Hillis et al., 1990). To obtain DNA sequences of CA1, CA2, and CA5 for *D. pulicaria*, gene-specific forward and reverse primers were designed using

the *D. pulex* sequences as a template (Table S1; Figure S1; Culver and Morton, *in review*). The *D. pulex* templates were obtained from the *Daphnia* Genomics Consortium (http://server7.wFleabase.org; Gene IDs, CA1 – DappuDraft_222096, CA2 – DappuDraft_222141, and CA5 – DappuDraft_317362). PCR was performed using IQ^{TM} Supermix 2X (Bio-Rad Laboratories, Hercules, CA) in a 25µL reaction (12.5µL Supermix 2X, 0.2µL DMSO, 2µL forward primer (10µM), 2µL reverse primer (10µM), 7.3 µL ultra-pure water, and 2µL of template DNA) with thermocycler settings as per Table S1. PCR products were purified and sequenced using an ABI-3130XL DNA analyzer (Applied Biosystems) in two forward and reverse replicates to minimize sequencing error. Consensus sequences for *D. pulicaria* CA1, CA2, and CA5 were constructed using the software MEGA 5.0 (Tamura et al. 2011). The ends of the genes were trimmed to make all sequences for each isoform the same length. All sequences for each population and isoform were uploaded into CLUSTALX 2.1 (Larkin et al. 2007) and a multi-sequence alignment was run with iterations after each alignment step. Aligned sequences were then analyzed in Mega 5.0 for single nucleotide polymorphisms. Genotypes were determined by using a neighbor-joining tree with 500 bootstraps under a Jukes Cantor model with a gamma distribution (the best fit model) using Mega 5.0 of the concatenated DNA sequences of all three isoforms from each *D. pulicaria* isolate.

Population genetic analysis and selection

Genetic structure was determined between populations in a pair-wise approach by calculating G_{st} (Nei 1973) using concatenated DNA sequences for each isoform from

each isolate and a chi-square test was used determine significance of the pair-wise comparisons with 1000 permutations. Allele frequencies, predicted homozygosity ($F = \Sigma p^2$), and predicted heterozygosity ($h = 1\text{-}F$) were calculated for each isoform within each population. Additional measures of population genetic structure were also determined including number of segregating sites (*S*), nucleotide diversity (*Π*), and haplotype diversity (H_d). Furthermore, signatures of selection were elucidated using *Tajima's D* (Tajima 1989) and *Fu and Li's D* and F** (Fu and Li 1993) for each gene isoform within each population by looking at significant departure from neutrality. Each test for neutrality was run with a sliding window of 3 base pairs (bp) with a 3bp step for segregating sites to determine whether there was selection at particular DNA sites within the gene isoform within each population. We ran *Fu and Li's* test (Fu and Li 1993) for neutrality; however, the results were similar to the *Tajima's D* test for neutrality, and therefore, are not reported here. A *Tajima's D* score of zero indicates the gene or site along the gene is under neutral evolution. A significant positive deviation from zero indicates evidence of balancing selection, while a significant negative deviation from zero indicates either positive or purifying selection. All analyses were performed in the software package DnaSP v5.0 (Librado and Rozas 2009).

Survivorship experiment

To test whether specific CA genotypes in *D. pulicaria* from the three lakes convey a fitness (survivorship) benefit across a pH gradient, a survivorship experiment was performed. Eleven *D. pulicaria* isolates (i.e., clones) representing five CA genotypes were chosen from an established clone bank (Table 1). If a genotype was represented

from more than one population, then an isolate from each population was paired for analysis. Genotypes will be referred to hereafter by color codes (Figure S2). Five different genotypes were selected for the survivorship experiment: blue, red, yellow, green, and purple. For Hill Lake and Madison Lake, only hatched established clones from the upper sediment layers (0-4 cm) were used because they represented more recent populations that experienced the pH regimes used in this study. Those animals initially harvested from the water column, of which the latest was harvested in 2014, are five years older than the oldest dated eggs from Hill and Madison Lake. During this 5-year period, the pH values of the individual lakes were relatively stable (see *Study Sites* section). All animals were raised under the same laboratory conditions using COMBO medium (~ pH = 7.5; Kilham et al. 1998).

For each clone, a set of 80 gravid animals were used to set up as stem grandmothers. Once the stem grandmothers had released their clutches, 80 neonates from each clone were harvested to serve as stem mothers. All animals were fed daily 0.5 mL of approximately 1 – 2 mg C L^{-1} of chemostatically-cultured green algae, *Scenedesmus acutus*, to ensure that the *Daphnia* were not food limited. Neonates from the third clutch of the stem mothers were used for the experiment, since previous work has shown that third-clutch offspring are more robust than first or second-clutch neonates (Glazier, 1992). In addition, using the grand-daughter generation reduces maternal effects (Lampert 2011). Note that two clones (one red genotype and one purple genotype) died during the set-up and were left out of the analysis, thus leaving only 9 clones for the final analysis.

In a common-garden experimental design, three treatments of COMBO medium were adjusted, using 0.5M NaOH or 0.5M HCl (coarse adjustment), to the following pHs: 6.0, 7.5, and 9.0. Ten neonates from each clone were placed in 100mL of pH-adjusted COMBO medium per treatment. For each clone, four replicate 100mL jars were used per treatment. The pH of the treatments was measured twice a day, with a Milwaukee Instruments PH56 pH meter (the pH meter was calibrated before taking measurements), and if necessary, adjusted with 0.05M NaOH or 0.05M HCl (fine adjustment). The experiment was terminated after 6 days, prior to the *Daphnia* reaching maturity (i.e., becoming reproductive). During the experiment, animals were fed daily with 1mg C L^{-1} of *S. acutus*. When testing and adjusting pH, these values were measured after the feeding. The number of surviving animals in each replicate was recorded at the termination of the experiment.

Statistical analysis

The statistical analyses for the genotype fitness experiment were done using the software R v2.15.0 (R Core Team 2012) with the packages *lme4* v1.1-10 (Bates et al. 2014) and *lmerTest* v2.0-29 (Kuznetsova 2015). The results for the 1080 *Daphnia* individuals were fitted to a generalized linear mixed-effects model using the Laplace approximation, with a binomial distribution. The full model contained survival (dead or alive) as the response variable; pH treatment, genotype for the individual and the interaction between those terms as fixed factors; and the different clones and the jars in which the experiment took place as random factors. We checked for over-dispersion by comparing deviance of the model with the residual degrees of freedom, and none of our

models presented over-dispersion. Post-hoc tests of the GLM were performed and p-values were corrected for multiple comparisons.

Full and reduced models were compared based on their Akaike Information Criterion (AIC). The full model had a slightly better AIC score than models incorporating clone as a random factor, or neither random factor (1065.2, 1152.1 and 1150.2, respectively), and a very similar score than the model incorporating jar as a random factor (1063.1). Therefore, we kept the full model for the final analysis.

Results

Carbonic anhydrase sequencing and genotyping

We obtained a total of 45 α-CA sequences for each isoform, 15 from each population. The gene lengths for each α-CA isoform were as follows: CA1 was 927bp and spanned 6 exons and 5 introns; CA2 was 749bp and spanned 4 exons and 4 introns; and CA5 was 1110bp and spanned 6 exons and 5 introns (Figure S1). For the population genetics analysis, the introns were excised from the final alignment, thus the resultant cDNA lengths for each isoform were as follows: 561bp (CA1), 573bp (CA2), and 600bp (CA5). The cDNA for CA1 included two of the four active sites, one of the three zinc-binding sites, and both substrate-binding sites; CA2 included all four active sites, all three zinc-binding sites, and both substrate-binding sites; and CA5 also included all four active sites, all three zinc-binding sites, both substrate-binding sites, and one of the two cysteine residues that form a disulfide bond. Of interest, in CA1 we found a non-synonymous substitution at site 214 (of the resultant cDNA; **Note:** from here on out when referring to DNA sites, it is with regards to the cDNA) that causes an amino acid

substitution one residue downstream of the first active site from a Leucine to an Isoleucine. This amino acid substitution was found in 80% of the Hill population, 20% of the Madison population, and was not found in the Frenchman population. The other two non-synonymous sites (343 and 372) were not found within conserved motifs (those regions that are conserved across the gene family; see Culver and Morton *in review*) of the resulting protein. In CA2, we found one rare (13% of the Hill population) non-synonymous substitution at site 91 that altered the amino acid four residues upstream of the first active site from a Leucine to Methionine. In CA5, we found two non-synonymous substitutions at sites 52 and 64 that altered the conserved amino acid motif NNGHT to CNGHA (note: the H in this motif represents the first active-binding site). These amino acid substitutions in CA5 were detected in 73% of the Madison population and in 20% of the Hill population, but not in the Frenchman population. In addition, we found: (i) CA1 had six different alleles; (ii) CA2 had five alleles; (iii) CA5 had three alleles (Table 1 & 2). The Frenchman population was fixed for allele one for all three CA isoforms. See Table 2 for the frequencies of the alleles at each locus.

Population Genetic Analysis

A total of eight genotypes were found among the three populations (Table 1, Figure S2). Phylogenetic analysis showed that genotypes cluster within their respective populations, with some exceptions. For example, there are threeindividual isolates (i.e., clones) from the Madison population that have an identical genotype to clones found in the Frenchman population. In addition three other clones from the Madison population are

identical to clones found in the Hill population. A pair-wise analysis of the population structure, G_{st}, also confirmed unique overall population differentiation (for each comparison X^2_{14}, $p < 0.001$; Table 3), with the Hill and Madison populations being more similar to each other, and both were significantly different from the Frenchman population.

Levels of predicted heterozygosity were relatively high for the Hill and Madison populations at each locus: heterozygosity was 0.792 and 0.738 for CA1; 0.516 and 0.605 for CA2; and 0.498 and 0.627 for CA5 in the Hill and Madison populations, respectively. The Frenchman population was monomorphic (i.e., fixed homozygosity) for allele 1 for each isoform; therefore, no genetic variation was detected across the three CA isoforms in this population. Only the CA5 locus in the Madison population exhibited a significant *Tajima's D* score, 1.905 ($p < 0.01$). Using the *Tajima's D* analysis with a sliding window (3 bp) for each population and locus (Figure 1), several sites were found to be significant ($p < 0.01$) in the different populations. In the Hill population, CA1 had two sites (343 and 372) that had significantly positive *Tajima* D values. These two sites were the non-synonymous sites that were not in conserved regions. Further, these same two sites have significantly positive values in the Madison population. Additionally, CA5 in the Madison population also had a significantly positive value at site 64. This non-synonymous site changed the last amino acid residue in the conserved motif NNGHT from a Threonine to an Alanine. The CA2 locus had no significant results; however, the trend between the Hill and Madison populations at the CA2 locus were markedly different, with several sites with near significant negative *Tajima's D* values at synonymous sites, while all the results at these sites were near zero

or slightly positive in the Madison population. Since the Frenchman population was fixed for allele one at all three loci (i.e., no nucleotide variation), the *Tajima's D* results are undefined. Additionally, the haplotype diversity was relatively high in the Hill and Madison populations (ranging from 0.552 to 0.790).

Fitness (survivorship) of genotypes under different pH conditions

In order to determine the effect of different CA isoforms to pH adaptation, the results from our common garden experiment are shown in Figure 2. As can be seen, at moderate pH values (7.5), survivorship was similar for all genotypes, with a mean survivorship frequency of 96.8%. The other two more extreme pH treatments decreased survivorship for all genotypes; this effect was more severe for the pH 6.0 treatment, where only 33.3% of all individuals survived, while the survival frequency at pH 9.0 was 65.3%. The results were evaluated using a generalized linear mixed-effects model to test the main effects (pH treatment and genotype), accounting for natural sources of variation (the different clones within each genotype and the individual jars used for the experiment). The effect of the pH treatments was statistically significant, with a p value <0.001 (Table 4).

Not all genotypes responded equally to the pH treatments. Genotype Blue was the main outlier. It appeared to be particularly well-adapted to low pH conditions; this was observed in the three clonal lines (i.e., two Frenchman, one Madison) present in our experiment. It had a much higher mean survivorship (67.5%) at pH 6.0, when compared to the average survivorship (16.2%) of the other four genotypes. Conversely, at pH 9.0, Genotype Blue had a much lower mean survivorship (39.2%), when compared to the

average survivorship (78.3%) of the other four genotypes. This overall difference with the other genotypes was significant ($p < 0.0001$, Table 4). It also showed a significant interaction with pH treatment, reflecting its differential response compared to the other genotypes ($p < 0.0001$, Table 4). In addition, the other four genotypes had consistently lower mean survivorships at pH 6.0, but were not drastically affected at high pHs, suggesting an increased tolerance to alkaline (higher pH) conditions. Post-hoc results showed that the yellow, green, purple and red genotypes were not significantly different from each other (Table S3). (note: Genotype Green, which was represented by two clones, one each from Hill and Madison, showed nearly identical survivorship responses across the pH gradient; see Figure 2).

Discussion

The overarching goal of this study was to elucidate whether sequence variation in three α-CA isoforms would allow us to determine whether selective or neutral processes contribute to the adaptive capabilities of these populations to local pH regimes. In particular, we looked at three lake populations of *D. pulicaria*: Frenchman Lake (low pH), Hill Lake (near neutral pH), and Madison Lake (high pH). Previous studies in these lakes have shown evidence of local adaptation with regards to juvenile survivorship (Culver and Acosta *in review*). Studies in other taxa have shown support for adaptation to local pH regimes, such as in the purple sea urchin (Evans et al. 2013) and to other environmental stressors, such as across thermal clines in Europe in the Colorado potato beetle (*Leptinotara decemlineata*; Lyytinen et al. 2012). Lyytinen et al. (2012) was also able to trace the adaptive mechanism to variation in heat shock

protein (Hsp70) expression levels. Other studies have looked at variation in protein-coding genes for patterns of selection across and between species. For example, Crease et al. (2011) were able to determine patterns of neutral and selective processes of evolution between *D. pulex* and *D. pulicaria* on variants of the lactate dehydrogenase (*Ldh*) gene across different habitat types, i.e. temporary ponds vs. large stratified lakes. Here, we hypothesized that if the three α-CA isoforms have diverged in populations under different local pH regimes, then there should be evidence of selection on the α-CA genes. Further, there should be genotypes that exhibit differential fitness (survivorship) across a pH gradient. Indeed, we found evidence of divergence in the three populations, along with some signature of non-neutral (selection) processes occurring in the three populations. This was manifested in genotype-specific differential (survivorship) fitness across a pH gradient.

Population differentiation

We found that our three populations of *D. pulicaria* did indeed represent distinct lineages using both a clustering approach (phylogeny; Figure S2) and population structure analysis using genetic distance (G_{st}; Table 3) based upon sequence variation in the three α-CA isoforms. While the populations were significantly differentiated using both approaches, there was evidence for some genetic admixture among the populations, especially between the Hill and Madison populations which are geographically closer to each other (~320 km) than they are to Frenchman Lake (~1120-1280 km). Previous studies using microsatellite markers, which are assumed to be evolving under neutral processes, also found a similar pattern of differentiation among

the three populations (Culver and Acosta *in review*). The pattern of divergence found in this study seems to be driven by the divergence of alleles that segregate into their own subclusters within populations. Others (Crease et al. 2011; Vergilino et al. 2011) have found similar patterns of divergence in other species of *Daphnia* for other gene families. However, these previous studies looked at variation in sequences among *Daphnia* species. For instance, Vergilino et al. (2011) found that *D. pulex, D. tenebrosa,* and *D. arenata* clustered into individual clades based upon variation in the *Rab4* gene. Also, as noted above, Crease et al. (2011) found similar patterns of divergence in variants of *Ldh* genes between *D. pulex* and *D.pulicaria*. This divergence is speculated to be based upon habitat variation among the species (i.e., *D. pulex* is found in ponds; *D. pulicaria* is a lake species).

Sequence Diversity

We found that the Frenchman population was fixed for a single allele at each CA locus. Previous work (Culver and Acosta *in review*) determined that this population consists of a single obligately parthenogenetic hybrid clone of *D. pulex* x *D. pulicaria*. This work was based on detecting a single genotype based on 15 microsatellite markers. Our findings here support this earlier work, in that each individual in the Frenchman population had the exact same genotype based upon three CA loci. Clonal selection is a likely mechanism to create this population genetic structure in Frenchman, as has been shown in other *Daphnia* systems (Mergeay et al. 2006). Given that this clone is particularly well-suited for the low pH in this lake, we can speculate that this provided a selective advantage.

One mechanism that has been shown to greatly reduce genetic diversity in populations is extreme alteration of native habitat. For example, in a study of acid mine drainage by Martins et al. (2009), it was found that the acidification of a drainage basin reduced the genetic diversity in a lake population of *Daphnia longispina* in southern Portugal, resulting in a genetically-depauperate population adapted to the new environment. A similar phenomenon may have happened in the Frenchman population when heavy-metal smelting during the 1960s in the Sudbury region of Canada reduced the lake pH to 4.0-5.0 (Pollard et al., 2003, Yan et al., 1996a; Derry & Arnott, 2007), likely wiping out the resident populations of *Daphnia*. When the pH recovered after the government/industry regulation of smelting operations in the 1970s, the lake was then re-colonized by the current hybrid complex, which was more tolerant to a lower pH. Of course, additional work would be needed (e.g. paleolimnology/resurrection ecology; regional genetic surveys) to further test this hypothesis.

The Hill and Madison Lake populations had significantly different histories. Madison Lake is located in the corn-belt region of Minnesota and has been subjected to large inputs of both nitrogen (NO_3) and phosphorous (PO_4) from fertilizer runoff. Large inputs of these nutrients can alkalize lakes, and likely, this has happened in Madison Lake (O'Brien and DeNovelles 1972, Schindler et al. 1985), while Hill Lake has seen relatively little disturbance. Since both populations have to deal with similar annual pH fluctuations (~ 1.6 pH change, while the change in Frenchman is half that ~ 0.8), having higher genetic diversity (heterozygosity) may allow for the populations to adapt to these relatively large annual fluctuations in pH (Gillespie and Turelli 1989).

There have been a number of studies that have looked at the relationship between levels of heterozygosity and niche-width variation that impacts the ability of populations to respond to fluctuating/changing environmental conditions (Hedrick et al. 1976; Hedrick and Cockerham 1986). For example, Yampolsky and Kalabushkin (1991) found that heterozygotes exhibited higher fitness than homozygotes, when exposed to fluctuating food concentrations. Very little research has been done to test the response of heterozygotes/homozygotes to pH fluctuations, let alone look at the fitness effects. The work that has been done usually looks at human physiology with regards to acidosis. For instance, Jarilim et al. (1998) found that heterozygotes for variants of AE1 anion exchanger in lumen of human renal tubes was found to function normally with regards to acid-base homeostasis, while homozygotes functioned abnormally (loss-of-function). If our study system holds to a similar pattern with regards to heterozygosity, we speculate that heterozygotes may have a broader pH niche.

It has also been hypothesized that balancing selection maintains high levels of heterozygosity (Black and Salzano 1981, Hedrick and Thomson 1983). While our *Tajima's D* scores for each locus within each population showed no significant evidence of balancing selection (with the exception of CA5 in the Madison population), the trend in Tajima's D scores shows evidence of balancing selection in all but the CA2 locus in the Hill population (the value for Hill was slightly negative, but not significantly different from neutral expectations). Indeed, if we look at the individual loci at finer scales (i.e., sliding windows), we find significant evidence for balancing selection at two non-synonymous sites (343 and 372) in CA1 in both Hill and Madison (Figure 1).

Additionally we found a non-synonymous site (64) in CA5 in the Madison population. This non-synonymous site changed the last amino acid residue in the conserved motif NNGHT from a Threonine to an Alanine and has the potential of changing the activity of this CA5 variant. Activity assays would likely help us to assess this prediction. Our results are supported by other studies, such as Jeyasingh et al. (2009), who found evidence for balancing selection for heterozygotes at the phosphoglucose isomerase (*Pgi*) locus in *Daphnia pulicaria*. These authors showed that heterozygotes were competitively superior to homozygotes under different food quality regimes.

Survivorship of genotypes

We had hypothesized that if *D. pulicaria* genotypes came from populations with localized pH conditions, then we would expect to see differential fitness (survivorship) across a pH gradient. We indeed did find support for our hypothesis. Clones (i.e., two from Frenchman, one from Madison) that had the blue genotype, had significantly higher survivorships in our low pH (6.0) treatment, but had higher mortalities in our high pH (9.0) treatment showing that this genotype is particularly well suited to a lower pH. All other genotypes, which were isolated from the Madison/Hill populations, had significantly higher mortalities at the low pH treatment; however they had higher survivorships at the high pH. The blue genotype consists of allele one from each locus, so it seems having all three of these alleles conveys fitness (survivorship) advantage at low pHs. The blue genotype is also found in the Madison population at low frequencies due to the homozygote being deleterious at high pH; however, it appears to be maintained in the population at low frequencies, possibly due to balancing selection.

Conclusion

We found support for our hypotheses that the populations were genetically differentiated; we found evidence of balancing selection maintaining high levels of heterozygosity in the Hill and Madison populations, and found evidence that certain genotypes convey local adaptation to local pH environments.

Further research that would help elucidate the mechanism(s) through which fitness is increased in local pH conditions, would be to look at protein structure and enzymatic activity of the different variants at each locus. In particular, some of the sequence variation has led to amino acid substitutions in conserved areas near active and zinc-binding sites. It would be interesting to determine if these amino acid substitutions change the conformation of the enzymes allowing for increased or decreased activity under varying pH conditions.

Acknowledgements

This study was supported by the U.S. National Science Foundation (NSF-IOS-OEI) grants to L. J. Weider (#0924289 and #1256881), the U. of Oklahoma Department of Biology Adams Scholarship and the U. of Oklahoma Biological Station Summer Research Grant to BWC. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We thank the U. of Minnesota LacCORE staff (in particular R. O'Grady) at the University of Minnesota, for assistance with coring and providing *D. pulicaria* samples from Madison Lake, MN. A. Harris facilitated sequencing of *D. pulicaria* enzymes. We thank K. Milette (McGill University) and M. Edlund (St. Croix River Watershed Station) for providing *Daphnia* samples and P. Morton for his assistance with primer design. We also thank L.J. Weider for his advice, mentorship, and comments on this manuscript. In addition, we thank R. Broughton, K.D. Hambright, I. Schlupp, and R. Cichewicz for their input. This manuscript constitutes a portion of BWC's doctoral dissertation at the U. of Oklahoma. There are no known conflicts of interests associated with this article and its authors.

References

Avise J. 2004. Molecular Markers, Natural History, and Evolution 2nd ed. Sunderland, MA: Sinauer Associates.

Baird DJ and Barata C. 1998 Genetic variation in the response of *Daphnia* to toxic substances: Implications for risk assessment. In Genetics and Ecotoxicology. Taylor & Francis, Inc, Ann Arbor, MI. 207-218.

Bates D, Mächler M, Bolker B, and Walker S. 2014. Fitting Linear Mixed-Effects Models using lme4. Submitted: Journal of Statistical Software 1-51.

Black FL and Salzano FM. 1981. Evidence of heterosis in the HLA system. Genetics. 33: 894-899.

Charlesworth B. 2010. Molecular population genomics: a short history. Genet Res Comb 92: 397-411.

Colbourne JK and Hebert PDN. 1996. The systematics of North America *Daphnia* (Crustacea: Anomopoda): A molecular phylogeny approach. Philos Trans R Soc London, Ser B. 351: 349-360.

Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, Tokishita S, Aerts A, Arnold GJ, Basu MK, *et al*. 2011. The Ecoresponsive Genome of *Daphnia pulex*. Science. 331: 555-561.

Cooper PD. 1994. Mechanisms of Hemolymph Acid-Base Regulation in Aquatic insects. Physiol Zool. 67(1):29-53.

Coors A, Vanoverbeke J, DeBie T, and deMeester L. 2009. Land use, genetic diversity and toxicant tolerance in natural populations of *Daphnia magna*. Aquatic Toxicology. 95: 71-79.

Crease TJ, Floyd R, Cristescu ME, and Innes D. 2011. Evolutionary factors affecting *Lactate dehydrogenase* A and B variation in the *Daphnia pulex* species complex. BMC Evol Biol. 11: 1-12.

Culver BW and Acosta F. *In Review*. Population genetic structure of *Daphnia pulicaria* across a pH gradient in three North American lakes. Hydrobiologia.

Culver BW and Morton PK, 2015. The evolutionary history of Daphniid α-carbonic anhydrase within Animalia. International Journal of Evolutionary Biology. 2015: 1-11.

Derry AM and Arnott SE. 2007. Adaptive reversals in acid tolerance in copepods from lakes recovering from historical stress. Ecological Applications. 17: 1116-1126.

Dudycha JL and Tessier AJ. 1999. Natural generic variation of life span, reproduction, and juvenile growth in *Daphnia*. Evolution. 53: 1744-1756.

Dufresne F and Hebert PDN. 1997. Pleistocene glaciations and polyphyletic origins of polyploidy in artic cladoceran. Proc R Soc London, Ser B. 264: 201-206.

Eanes WF. 1999. Analysis of selection on enzyme polymorphisms. Annu Rev Ecol Syst. 30: 301-326.

Esbuagh AJ and Tufts BL. 2006. The structure and function of carbonic anhydrase isozymes in the respiratory system of vertebrates. Respiratory Physiology and Neurobiology. 154: 185-198.

Frisch D, Morton PK, Chowdhury PR, Culver BW, Colbourne JK, Weider LW, and Jeyasingh PD. 2014. A millennial-scale chronicle of evolutionary responses to cultural eutrophication in *Daphnia*. Ecology Letters. 17: 360-368.

Fu YX and Li WH. 1993. Statistical tests of neutrality of mutations. Genetics. 133: 693-

709.

Gillespie JH and Turello M. 1989. Genotype-environment interaction and the maintenance of polygenic variation. Genetics. 121: 129-138.

Glazier DS.1992. Effects of food, genotype, and maternal size and age of offspring investment in *Daphnia magna*. Ecology. 73: 910-926.

Hebert PDN. 1978. The population biology of *Daphnia* (Crustacea, Daphnidae). Biological Reviews. 53: 387-426.

Hedrick PW, Ginevan ME, and Ewing EP. 1976. Genetic Polymorphism in Hetergeneous Environments. Annual Review of Ecology and Systematics 7: 1-32.

Hedrick PW and Thomson G. 1983. Evidence for Balancing Selection at HLA. Genetics. 104: 449-456.

Hedrick PW and Cockerham CC. 1986. Partial Inbreeding Equilibrium Heterozygosity and the Heterozygosity paradox. Evolution. 40: 856-861.

Henry RP, Cameron JN. 1983. The role of Carbonic Anhydrase in respiration, ion regulation and acid-base balance in the aquatic crab *Callinectes Sapidus* and the terrestrial crab *Gecarnicus lateralis*. J exp Biol. 103:205-223.

Henry RP. 1984. The Role of Carbonic Anhydrase in Blood Ion and Acid-Base Regulation. Amer Zool. 24:241-251

Hunt JA and Fierke CA. 1997. Selection of Carbonic Anhydrase Variants Displayed on Phage Aromatic Residies in Zinc Binding Site Enhance Metal Affinity and Equilibrium Dynamics. J Biol Chem. 271: 20364-20372.

Jeyasingh PD, Weider LW, and Sterner RW. 2009. Genetically-based trade-offs in

response to stoichiometric food quality influence competition in a keystone aquatic herbivore. Ecology Letters. 12: 1229-1237.

Kilham SS, Kreeger DA, Lynn SG, Goulden CE, and Herrara L. 1998. COMBO: a defined freshwater culture medium for algae and zooplankton. Hydrobiologia. 377: 147-149.

Kuznetsova A, Brockhoff P, and Christensen. 2015. R. Package 'lmerTest': Tests in linear mixed-effects models (Version 2.0–25).

Lampert W. 2011. *Daphnia*: development of a model organism in ecology and evolution. Oldendorf/Luhe: International Ecological Institute

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947-2948.

Librado P and Rozas J. 2009. DnaSP v5: A software package for comprehensive analysis of DNA polymorphism data. Bioinformatics. 25: 1451-1452.

Lyytinen A, Mappes J, and Linström L. 2012. Variation in Hsp70 Levels after cold shock: signs of Evolutionary Responses to Thermal Selection among *Leptinotarsa decemlineata* Populations. PLoS one. 7: 1-4.

Martins CMG, Almeida DV, Marins LFF, and Bianchini, A. 2010. mRNA expression and activity of ion-transporting proteins in gills of the blue crab *Callinectes sapidus*: effects of waterborne copper. Environmental Toxicology and Chemistry 30: 206-211.

Martins N, Bollinger C, Harper RM, and Ribeiro R. 2009. Effects of mine drainage on the genetic diversity and structure of a natural population of *Daphnia longispina*. Aquatic Toxicology. 92: 104-112.

Mergeay J, Vershuren D, and DeMeester L. 2006. Invasion of an asexual American water flea clone throughout Africa and rapid displacement of a native sibling species. Proc R Soc Londong, Ser B. 1-10.

Miner BE, DeMeester L, Pfrender ME, Lampert W, and Hairston Jr N.G. 2012. Linking genes to communities and ecosystems: *Daphnia* as an ecogenomic model. Proceedings of the Royal Society B, 1-10.

Moya A, Huisman L, Ball EE, Hayward DC, Grasso LC, Chua CM, Woo HN, Gattuso JP, Forêt S, and Miller DJ. 2012. Whole transcriptome analysis of the coral *Acropora millepora* reveals complex responses to CO_2-driven acidification during the initiation of calcification. Molecular Ecology 21:2440-2454

Nei M. 1973. Analysis of gene diversity in subdivided populations. PNAS. 70: 3321-3323.

O'Brien WJ and DeNoyelles, JR. F. 1972. Photosynthetically elevated pH as a factor in zooplankton mortality in nutrient enriched ponds. Ecology. 53: 605-614.

Pollard HG, Colbourne JK, and Keller W. 2003. Reconstruction of centuries-old *Daphnia* communities in a lake recovering from acidification and metal contamination. Ambio: A journal of the Human Environment 23: 214-218.

R Core Teamm, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org.

Schindler DW, Turner MA and Hesslein RH. 1985. Acidification and Alkalizationof lakes by experimental addition of nitrogen compounds. Biogeography. 1:117-133.

Stollewerk A. 2010. The water flea *Daphnia* – a 'new' model system for ecology and evolution? Journal of Biology 9: 1-4.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123: 585-595.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol. 28(10):2731-2739.

Vergilino R, Markova S, Ventura M, Manca M, and Dufresne F. 2011. Reticulate evolution of the *Daphnia pulex* complex as revealed by nuclear markers. Mol Ecol. 20: 1191-1207.

Weber AK, Pirow R. 2009. Physiological responses of *Daphnia pulex* to acid stress. BMC Physiol. 9: 1-25.

Wetzel RG. 2001. Limnology: Lake and River Systems 3rd Ed. London: Elsevier Academic Press.

Yampolsky LY and Kalabushkin BA. 1991. The components of life-history traits variation in *Daphnia magna* population. Hydrobiologia. 225: 255-261.

Yan ND, Keller W, Somers KM, Pawson TW, and Girard RG. 1996. Recovery of crustacean zooplankton communities from acid and metal contamination: comparing manipulation and reference lakes. Canadian Journal of Fisheries and Aquatic Science. 53: 1301-132.

Tables and Figures

Table 1. *Daphnia pulicaria* genotypes from the three lake populations (Frenchman, Hill and Madison). Alleles for each isoform (CA1, CA2, and CA5) are color-coded for ease-of-identification. The numbers reflect site polymorphisms in the DNA sequence for the particular locus, with * indicating non-synonymous sites. The left column indicates the clone ID and the clones used for the survivor experiment are highlighted by color of their genotype as indicated by figure S2. Note: only five of the eight genotypes were used.

	CA1									CA2										CA5						
Clone	39	99	214*	285	294	324	343*	372*	492	15	91*	117	123	162	204	216	382	406	430	4*	5	18*	42*	46	52*	53	64*
French_13	C	A	C	A	C	T	A	T	T	T	C	T	G	C	G	C	G	T	T	A	T	T	T	G	A	A	A
French_14	C	A	C	A	C	T	A	T	T	T	C	T	G	C	G	C	G	T	T	A	T	T	T	G	A	A	A
French_15	C	A	C	A	C	T	A	T	T	T	C	T	G	C	G	C	G	T	T	A	T	T	T	G	A	A	A
French_18	C	A	C	A	C	T	A	T	T	T	C	T	G	C	G	C	G	T	T	A	T	T	T	G	A	A	A
French_20	C	A	C	A	C	T	A	T	T	T	C	T	G	C	G	C	G	T	T	A	T	T	T	G	A	A	A
French_22	C	A	C	A	C	T	A	T	T	T	C	T	G	C	G	C	G	T	T	A	T	T	T	G	A	A	A
French_23	C	A	C	A	C	T	A	T	T	T	C	T	G	C	G	C	G	T	T	A	T	T	T	G	A	A	A
French_26	C	A	C	A	C	T	A	T	T	T	C	T	G	C	G	C	G	T	T	A	T	T	T	G	A	A	A
French_27	C	A	C	A	C	T	A	T	T	T	C	T	G	C	G	C	G	T	T	A	T	T	T	G	A	A	A
French_30	C	A	C	A	C	T	A	T	T	T	C	T	G	C	G	C	G	T	T	A	T	T	T	G	A	A	A
French_35	C	A	C	A	C	T	A	T	T	T	C	T	G	C	G	C	G	T	T	A	T	T	T	G	A	A	A
French_M3	C	A	C	A	C	T	A	T	T	T	C	T	G	C	G	C	G	T	T	A	T	T	T	G	A	A	A
French_M8	C	A	C	A	C	T	A	T	T	T	C	T	G	C	G	C	G	T	T	A	T	T	T	G	A	A	A
French_M16	C	A	C	A	C	T	A	T	T	T	C	T	G	C	G	C	G	T	T	A	T	T	T	G	A	A	A
French_M23	C	A	C	A	C	T	A	T	T	T	C	T	G	C	G	C	G	T	T	A	T	T	T	G	A	A	A
Hill-10_Lake_1	A	G	A	C	C	C	A	G	C	C	C	A	C	C	A	T	A	C	A	A	T	T	T	G	A	A	A
Hill-10_Lake_10	C	A	C	A	C	T	A	T	T	C	A	T	C	T	G	C	G	C	A	T	G	G	G	T	T	G	G
Hill-10_Lake_11	A	G	A	C	C	C	A	G	C	C	C	A	C	C	A	T	A	C	A	A	T	T	T	G	A	A	A
Hill-10_Lake_12	C	A	C	C	C	C	G	T	T	T	C	T	G	C	G	C	G	T	T	T	G	G	G	T	T	G	G
Hill-10_Lake_13	A	G	A	C	C	C	A	T	C	C	C	A	C	C	A	T	A	C	A	A	T	T	T	G	A	A	A
Hill-10_0-2_2	A	G	A	C	C	C	A	G	C	C	C	A	C	C	A	T	A	C	A	A	T	T	T	G	A	A	A
Hill-10_0-2_4	A	G	A	C	T	G	C	G	C	C	C	A	C	C	A	T	A	C	A	A	T	T	T	G	A	A	A
Hill-10_0-2_6	A	G	A	C	T	G	C	G	C	C	C	A	C	C	A	T	A	C	A	A	T	T	T	G	A	A	A
Hill-11_Lake_2	A	G	A	C	T	G	C	G	C	C	C	A	C	C	A	T	A	C	A	A	T	T	T	G	A	A	A
Hill-11_Lake_13	A	G	A	C	T	G	C	G	C	C	C	A	C	C	A	T	A	C	A	A	T	T	T	G	A	A	A
Hill-11_0-2_3H	A	G	A	C	C	C	A	G	C	C	C	A	C	C	A	T	A	C	A	A	T	T	T	G	A	A	A
Hill-11_0-2_12H	A	G	A	C	C	C	A	G	C	C	C	A	C	C	A	T	A	C	A	A	T	T	T	G	A	A	A
Hill-11_0-2_15B	C	A	C	A	C	T	A	T	T	C	A	T	C	T	G	C	G	C	A	T	G	G	G	T	T	G	G
Hill-14_Lake_58	C	G	A	C	T	G	C	G	C	C	C	A	C	C	A	C	G	C	A	T	G	T	G	T	A	A	A
Hill-14_Lake_71	C	G	A	C	T	G	C	G	C	C	C	A	C	C	A	C	G	C	A	T	G	T	G	T	A	A	A
Mad-10 0-4cm 1B	C	A	C	C	C	C	G	T	T	T	C	T	G	C	G	C	G	T	T	T	G	G	G	T	T	G	G
Mad-10 0-4cm 3B	C	A	C	C	C	C	G	T	T	T	C	T	G	C	G	C	G	T	T	T	G	G	G	T	T	G	G
Mad-10 0-4cm 4A	A	G	C	C	C	C	A	T	T	T	C	T	G	T	G	C	G	T	T	T	G	G	G	T	T	G	G
Mad-10_0-4cm_4B	A	G	C	C	C	C	A	T	T	T	C	T	G	T	G	C	G	T	T	T	G	G	G	T	T	G	G
Mad-10_0-4_5A	C	A	C	C	C	C	G	T	T	T	C	T	G	C	G	C	G	T	T	T	G	G	G	T	T	G	G
Mad-10_0-4_5B	C	A	C	C	C	C	G	T	T	T	C	T	G	C	G	C	G	T	T	T	G	G	G	T	T	G	G
Mad-10 0-4cm 16A	A	G	C	C	C	C	A	T	T	T	C	T	G	T	G	C	G	T	T	T	G	G	G	T	T	G	G
Mad-10 0-4cm 16B	A	G	C	C	C	C	A	T	T	T	C	T	G	T	G	C	G	T	T	T	G	G	G	T	T	G	G
Mad-10 0-4 20	C	G	A	C	T	G	C	G	C	C	C	A	C	C	A	C	G	C	A	A	T	T	G	T	A	A	A
Mad-10_4-8_10	C	A	C	A	C	T	A	T	T	T	C	T	G	C	G	C	G	T	T	A	T	T	T	G	A	A	A
Mad-10 8-12cm 1A	C	A	C	A	C	T	A	T	T	T	C	T	G	C	G	C	G	T	T	A	T	T	T	G	A	A	A
Mad-10 8-12cm 2A	C	A	C	A	C	T	A	T	T	T	C	T	G	C	G	C	G	T	T	A	T	T	T	G	A	A	A
Mad-10 8-12cm 13A	C	G	A	C	T	G	C	G	C	C	C	A	C	C	A	C	G	C	A	T	G	T	G	T	A	A	A
Mad-11_Lake_8	C	A	C	C	C	C	G	T	T	T	C	T	G	C	G	C	G	T	T	T	G	G	G	T	T	G	G
Mad-11_Lake_10	C	G	A	C	T	G	C	G	C	C	C	A	C	C	A	C	G	C	A	T	G	T	G	T	A	A	A

Table 2. Results of the population genetic analysis using DnaSP v5 (Librado and Rozas 2009). Results are partitioned by locus (CA1, CA2, CA5) and *D. pulicaria* population (Frenchman, Hill, and Madison). Results are for allele frequency, predicted *Homozygosity*, predicted *Heterozygosity*, number of segregating site (*S*), nucleotide diversity (*ΠT*), Tajima's D, Non-synonymous site/ synonymous site ratio, and haplotype diversity (H_d) and H_d standard deviation.

Gene Isoform ->	**CA1**			**CA2**			**CA5**		
Population ->	**Frenchman**	**Hill**	**Madison**	**Frenchman**	**Hill**	**Madison**	**Frenchman**	**Hill**	**Madison**
Allele 1	1.000	0.133	0.200	1.000	0.067	0.533	1.000	0.667	0.267
Allele 2		0.400			0.667			0.200	0.533
Allele 3		0.267			0.133			0.133	0.133
Allele 4		0.067	0.333		0.133	0.200			
Allele 5		0.133	0.200			0.267			
Allele 6			0.267						
Homozygosity ($F=\Sigma p^2$)	1.000	0.288	0.262	1.000	0.484	0.395	1.000	0.502	0.373
Heterozygosity (1-F)	0.000	0.792	0.738	0.000	0.516	0.605	0.000	0.498	0.627
S = segregating site	0	9	9	0	10	7	0	8	6
Π=nucleotide mismatches	0.0000	0.0069	0.0070	0.0000	0.0047	0.0043	0.0000	0.0055	0.0062
Tajima's D	0.000	1.460	1.591	0.000	-0.506	0.541	0.000	1.220	1.905*
NonSyn/Syn	0.000	0.827	1.015	0.000	0.844	n/a	0.000	0.965	2.62
Haplotype Diversity (H_d)	0.000	0.781	0.790	0.000	0.552	0.648	0.000	0.781	0.590
H_d Std Dev	0.000	0.005	0.003	0.000	0.019	0.008	0.000	0.006	0.011

Table 3. Results of pairwise comparisons of each of the three *D. pulicaria* population genetic structures (G_{st}) based upon genetic distance among the concatenated sequences of the three CA isoforms (Nei 1973). Each comparison is significantly different ($X^2_{14,}$ $p < 0.001$).

Population 1	Population 2	G_{st}
Frenchman	Hill	0.438
Frenchman	Madison	0.328
Hill	Madison	0.094

Table 4. Results of the genotype survivorship (fitness) experiment. The data were fitted to a generalized linear mixed-effects model, using a binomial distribution. Fixed factors were the pH treatment and the CA genotype, while the random factors were the jars in which the experiment took place and the different clonal lines under each genotype. First level for the genotype factor is the Green genotype (Green genotype is the reference for GLM). Significant results are shown in **bold** font. Results of post-hoc tests are given in Table S3.

	Estimate	Std Error	Z value	*P* (> |z|)
(Intercept)	-6.166	2.0384	-3.015	**0.0025**
pH	0.881	0.267	3.3	**<0.0010**
Blue Genotype	10.737	2.606	4.12	**<0.0001**
Yellow Genotype	-3.458	2.982	-1.16	0.2462
Red Genotype	-0.5431	3.557	-0.153	0.8787
Purple Genotype	-4.075	3.744	-1.088	0.2764
pH: Blue Genotype	-1.339	0.341	-3.929	**<0.0001**
pH: Yellow Genotype	0.557	0.394	1.414	0.1574
pH: Red Genotype	0.103	0.466	0.22	0.8259
pH: Purple Genotype	0.662	0.499	1.327	0.1844

Figure 1. *Tajima's D* results for each Population and CA isoform (CA1, CA2, and CA5). The sequences were analyzed using DnaSP 5.1 (Rozas et al. 2010) with a sliding window of 3 base pairs (bp). A *Tajima's D's* score that deviates from zero indicates departure from neutrality, with positive values indicating balancing selection and negative values indicating positive or purifying selection. The * indicate areas of significant departure from neutrality. The left facets are from the Hill Lake population (n=15) and the right facets from the Madison Lake population (n=15). Frenchman Lake could not be analyzed since all sites were fixed in each isoform. The top facets indicate CA1, the middle facets CA2, and the bottom facets CA5. **(Note the scales on the axes are not the same)**.

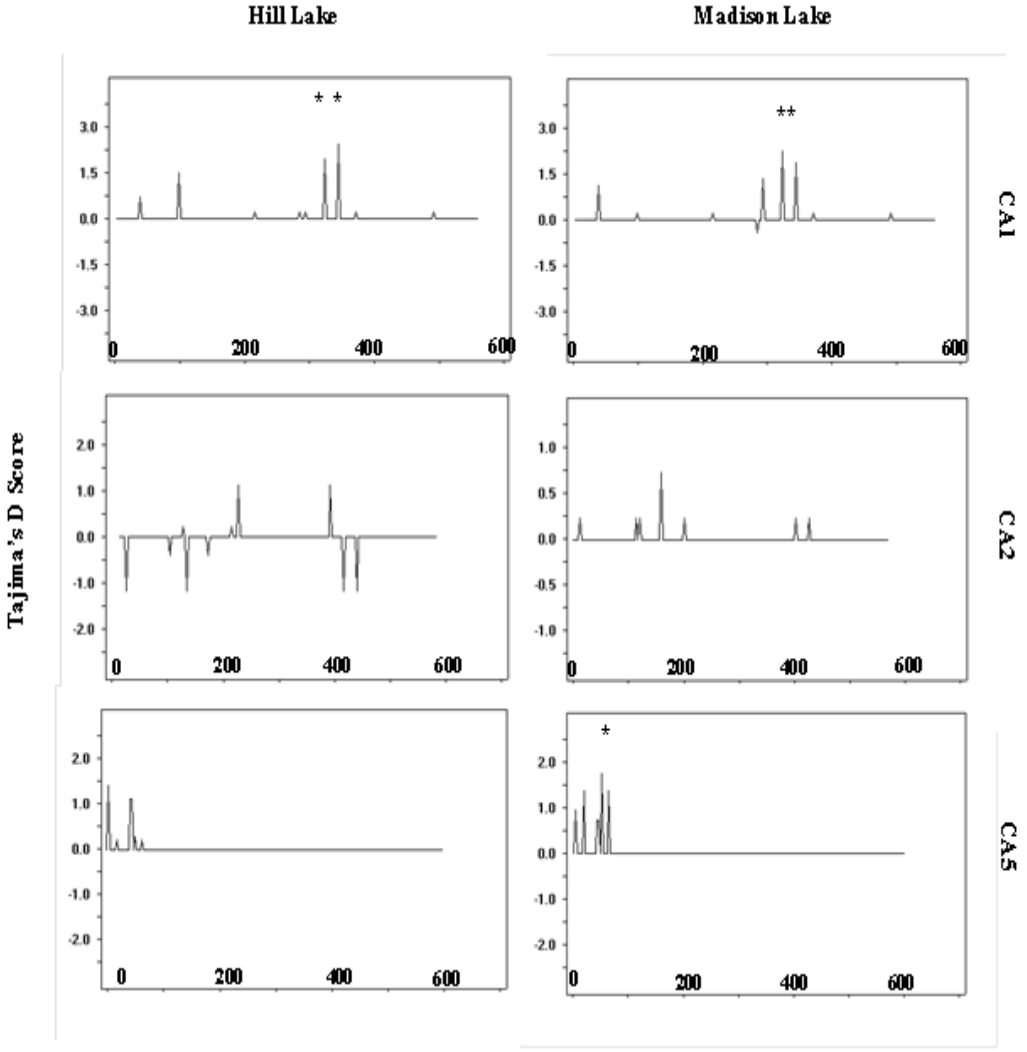
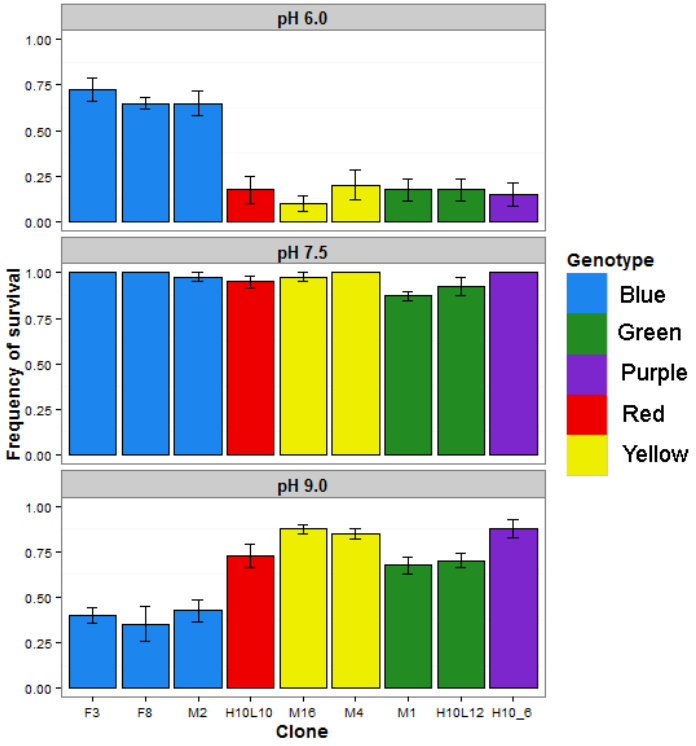


Figure 2. Mean (± 1 S.E.) survival frequency of 9 different clonal lines of *Daphnia pulicaria* that exhibited five different CA genotypes (indicated by colors according to legend) in three different pH conditions (pH 6.0, 7.5, 9.0) in a common garden experiment. Clones from the three study lakes are coded as follows with the prefix F (Frenchman), H (Hill) and M (Madison). Note that two sets of genotypes (blue and green) contain clones from two different lakes (blue – Frenchman & Madison; green – Hill & Madison).



Synthesis

The genome of *Daphnia pulex* was published in 2011 (Colbourne et al. 2011). This gave ecological geneticists that work with *Daphnia*, a valuable tool box to tackle questions with regards to how organisms respond to changes in the environment. This dissertation uses a multi-disciplinary approach to elucidate physiological and genetic mechanisms involved in adaptation of *D. pulicaria* populations to a particular environmental factor, in this case pH. (Note: *D. pulicaria* is a sister species of *D. pulex* and together, they form a recently divergent species complex that can hybridize; Omilian and Lynch 2009). Further, this approach has allowed me to look into the processes of evolution that may be implicated in maintaining local adaptation across a range of pHs.

Using three North American populations of *D. pulicaria* that are representative from across their pH tolerance range (pH 6 - 10; O'Brien & deNoyelles, 1972; Havens et al., 1993), I investigated several hypotheses. First, if candidate genes were found in closely-related taxa (decapods and aquatic insects; Henry and Cameron 1983, Henry 1984, Cooper 1994) that were found to be implicated in acid-base regulation, then I would expect to find homologs in *D. pulicaria*. I took a phylogenetic approach to elucidate candidate genes by undertaking an extensive literature review to identify homologs of the gene α-carbonic anhydrase (CA) in *D. pulicaria*. Second, if the candidate genes were indeed involved in acid-base regulation in *Daphnia*, then I would expect that these genes would have all the functioning amino-acid (AA) residues found in the homologous genes that were implicated in acid-base regulation in these other taxa. I used phylogenetic comparisons in both DNA and AA sequences of the candidate

genes to determine conservation of the functional elements of the resultant enzymes. Third, if CA gene regulation is a function of changes in pH, then it would be expected that genes would be differentially-expressed across a pH gradient. To test this hypothesis, I conducted a common garden experiment across a pH gradient and differential expression was measured using quantitative real-time PCR (qRT-PCR). Fourth, since the *Daphnia* populations used in this study are found across a breadth of ecological gradients and geographic distances, I expected to find genetic divergence among the populations. I used both microsatellite loci and polymorphisms in the DNA sequences of the CA (candidate) genes to perform population genetic analysis to determine population divergence. Lastly, if populations of *D. pulicaria* are adapted to their native habitat (lake), then I would expect to see evidence of differential fitness, localized differential expression, advantageous CA genotypes to be more common in their native population, and detect the signatures of non-neutral evolutionary processes (i.e. purifying, positive, or balancing selection) operating in these populations.

Candidate Genes

In the broad literature survey, I found that α-CA is implicated in acid-base regulation in most organisms. In particular, I was interested in α-CAs localized in the gills of decapods and aquatic insects. In these organisms, it had been determined that a few different isoforms of α-CA were enzymatically active in acid-base regulation, in particular cytosolic and GPI-anchored isoforms. Using a phylogenetic approach (Chapter two), I found that three different isoforms of CA in *Daphnia* were homologous to those CA isoforms in decapods and aquatic insects: two cytosolic (CA1 and CA2)

and one GPI-anchored (CA5). These three isoforms in *D. pulicaria* were characterized (Chapter three) and found to have similar gene architecture and were conserved in functional regions (active-sites, substrate-binding sites, and zinc-binding sites) of the protein across taxa ranging from *Daphnia* to *Homo sapiens.* In fact, an analysis using Mauve bioinformatics software (Darling et al. 2004) predicted all three isoforms to be homologous in function across the taxa.

CA isoform regulation

There has been some debate in the literature about whether differential expression of CAs occurs, or if CAs are ubiquitously expressed (i.e., represent housekeeping genes; Gilmour 2010). Using qRT-PCR (Chapter three) on *D. pulicaria* isolates from three North American lakes that were identified to have a pH gradient: Frenchman Lake (6.53 ± 0.50), Hill Lake (7.91 ± 0.52), and Madison Lake (8.63 ± 0.25) under common garden conditions, it was determined that isoform CA1 was differentially-expressed among populations across a pH gradient. However, CA2 and CA5 were not differentially-expressed across a pH gradient. In addition, it was found that in the Frenchman population, CA5 was always differentially up-regulated relative to isolates from the other two populations (Hill and Madison) under all pH conditions. This partially supports the third hypothesis that I would expect to see differential expression across the pH gradient for all isoforms. These findings also lend support to the notion that populations are locally adapted to their environment through the mechanism of differential expression.

Population Differentiation

Given that the three *D. pulicaria* populations examined in this dissertation are separated by considerable geographic distances (i.e., hundreds of kms) and have different geological histories, I expected to find significant population genetic differentiation among the three populations. Two approaches were used to determine population differentiation: (i) using microsatellite loci (Chapter one) and (ii) polymorphisms in the three CA isoforms (Chapter four).

(i) *Microsatellite approach*: Three methods were used to determine population structure: STRUCTURE analysis, PCoA of the genetic distance, and an AMOVA of pairwise population F_{st}. All analyses indicated that each population was genetically distinct; however, the Madison and Hill populations appeared to be most similar. In addition, population genetic analysis indicated that the Hill and Madison populations had higher allelic diversity than the Frenchman population. Furthermore, there was no evidence that any of the populations have experienced a bottle-neck in the recent past (however, see the caveat about Frenchman Lake – Chapter one and below).

(ii) *DNA polymorphism approach*: Two methods were used to determine population structure using the three CA isoforms: phylogenetic clustering and pairwise comparison of G_{st} based upon genetic distance. Again, all three populations were distinct; however, the relationship of the Madison population to either the Hill or Frenchman population was ambiguous. Population genetic analysis predicted high levels of heterozygosity at each CA locus for the Madison and Hill populations; however, in the Frenchman population, each CA locus was found to be monomorphic

for a single allele. Again, there is high haplotype diversity in the Madison and Hill populations, while there is no diversity in the Frenchman population.

Evidence for local adaptation and evolution

Organisms typically are found within an optimum range of environmental tolerance, which can be either broad or narrow (Cox et al., 1976; Raleigh et al., 1980). However, some individual populations within a given species can have a fitness advantage relative to other populations within narrow bands of the species tolerance, i.e. being locally adapted to narrow ranges within the overall tolerance range (Byars et al., 2007). In particular, changes in environment (pH) can act as a strong selective force that can affect the genetic variation and physiological responses of natural aquatic populations. To investigate the evolution of the three *D. pulicaria* populations within their native environments, I looked at (i) differential fitness in common garden survivorship experiments, both at the population-level (Chapter one) and CA genotype-level (Chapter four); (ii) examined localized differential expression across the CA isoforms (Chapter three); and (iii) detected signatures of non-neutral evolutionary processes (Chapter four).

I performed two different survivorship experiments to look for (1) local adaptation at the population level, and (2) differential fitness of CA genotypes across a pH gradient. At the population level, there was evidence that the Frenchman population was locally adapted to low pH conditions, while the Madison and Hill populations were locally adapted to higher pHs. At the genotypic level, one genotype (homozygous for the first allele at each CA locus, which was fixed in the Frenchman population and rare

in the Madison population) had higher fitness (survivorship) under low pH conditions, while the other genotypes exhibited higher fitness at higher pHs. In fact, evidence suggests that balancing selection is maintaining high levels of heterozygosity in the Madison and Hill populations due to the high levels of annual pH variation in these lakes.

Further avenues of study

While this dissertation shows strong support that the α-CA isoforms investigated here are implicated in acid-base regulation in *D. pulicaria*, further research is still needed to determine the actual function of these genes. For instance, isolating the CA proteins and investigating the activity of the different isoforms and their variants could lend more support for the hypotheses presented here. Also, deciphering the protein structure could give us an understanding of the molecular mechanism(s) involved in how each isoform variant functions at a given pH condition. Furthermore, CAs are undoubtedly not the only genes involved in acid-base regulation, as has been shown in other studies, such as sodium/potassium ATPase subunit alpha-3 in purple sea urchins (Evans et al. 2013). For instance, I determined that there was differential expression across a pH gradient in CA1. Thus, the co-factors involved in regulating CA1 expression may also be subjected to evolutionary forces (i.e., drift, selection). In order to more fully characterize potential other genes involved in acid-base regulation, more advanced molecular methods (beyond the scope of this dissertation), would need to be conducted (e.g., Next Generation Sequencing (NGS) methods such as RNA-seq; Ansorge 2009), following a similar experimental approach, as shown here. This approach could also be

augmented with other surrogate measures of differential fitness (i.e., fecundity, life-history shifts) to further characterize the impact of pH variation in natural populations of aquatic organisms like *Daphnia*.

References

Ansorge WJ. 2009. Next-generation DNA sequencing techniques. New Biotechnology. 25:195-203.

Byars, S.G., Papst, W., & Hoffmann, A.A., 2007. Local adaptation and cogradient selection in the alpine plant, *Poa hiemata*, along a narrow altitudinal gradient. Evolution 61: 2925-2941.

Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, Tokishita S, Aerts A, Arnold GJ, Basu MK, *et al*. 2011. The Ecoresponsive Genome of *Daphnia pulex*. Science. 331: 555-561.

Cooper PD. 1994. Mechanisms of Hemolymph Acid-Base Regulation in Aquatic insects. *Physiol Zool*, 67(1):29-53.

Cox, B.C., Healy, I.N., & Moore, P.D., 1976. Biogeogrpahy: an Ecological and Evolutionary Approach 2nd ed. Blackwell Science Publication, Oxford.

Darling ACE, Mau B, Blattner FR, and Perna NT. 2004. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. Genome Res. 14:1394-1403.

Evans TG, Chan F, Menge BA, and Hofmann GE. 2013. Transcriptomic responses to ocean acidification in larval sea urchins from a naturally variable pH environment. Molecular Ecology. 22:1609-1625.

Gilmour KM. 2010. Perspectives on carbonic anhydrase. Comparative Biochemistry and PhysiologyA: Molecular & Integrative Physiology. 157: 193-197.

Havens, K.E., Yan, N.D., & Keller, W., 1993. Lake acidification: effects on crustacean zooplankton populations. Environmental Science Technology 27: 1621-1624.

Henry RP, Cameron JN. 1983. The role of Carbonic Anhydrase in respiration, ion regulation and acid-base balance in the aquatic crab *Callinectes sapidus* and the terrestrial crab Gecarnicus lateralis. J exp Biol, 103:205-223.

Henry RP. 1984. The Role of Carbonic Anhydrase in Blood Ion and Acid-Base Regulation. *Amer Zool*, 24:241-251.

O'Brien, W.J., & deNoyelles, F., JR. 1972 . Photosynthetically elevated pH as a factor in zooplankton mortality in nutrient enriched ponds. Ecology 53: 605-614.

Omilian AR and Lynch M. 2009. Patterns of intraspecific DNA variation in the *Daphnia* nuclear genome. Genetics. 182: 325-336.

Raleigh, R.F., Hickman, T.J., Nelson, K.L. & Maughan, O.E., 1980. Riverine habitat evaluation procedures for rainbow trout. In: Proceedings of the trout stream improvement workshop, USDA Forest Service and Trout Unlimited, pp. 50-59. Asheville, NC, November 1980.

Appendix A: Chapter One Supplemental Material

Table S1. Results of HWE analysis for each microsatellite loci within each population (GenAlEx 6.41).

Pop	Locus	DF	ChiSq	Prob	Signif
Frenchman	Dp162	Monomorphic			
Frenchman	Dp173	1	67.000	0.000	***
Frenchman	Dp283	1	67.000	0.000	***
Frenchman	Dp291	Monomorphic			
Frenchman	Dp369	Monomorphic			
Frenchman	Dp375	Monomorphic			
Frenchman	Dp376	Monomorphic			
Frenchman	Dp38	Monomorphic			
Frenchman	Dp401	1	67.000	0.000	***
Frenchman	Dp43	1	67.000	0.000	***
Frenchman	Dp433	Monomorphic			
Frenchman	Dp437	Monomorphic			
Frenchman	Dp446	Monomorphic			
Frenchman	Dp461	1	67.000	0.000	***
Frenchman	Dp90	Monomorphic			
Hill	Dp162	1	0.021	0.885	ns
Hill	Dp173	3	9.360	0.025	*
Hill	Dp283	1	0.597	0.440	ns
Hill	Dp291	1	110.000	0.000	***
Hill	Dp369	3	38.588	0.000	***
Hill	Dp375	1	110.000	0.000	***
Hill	Dp376	1	110.000	0.000	***
Hill	Dp38	1	48.215	0.000	***
Hill	Dp401	10	115.107	0.000	***
Hill	Dp43	1	0.009	0.923	ns
Hill	Dp433	6	58.158	0.000	***
Hill	Dp437	6	2.691	0.846	ns
Hill	Dp446	3	69.748	0.000	***
Hill	Dp461	6	33.542	0.000	***
Hill	Dp90	3	0.423	0.935	ns
Madison	Dp162	1	92.000	0.000	***
Madison	Dp173	10	22.680	0.012	*
Madison	Dp283	1	1.220	0.269	ns
Madison	Dp291	3	14.140	0.003	**
Madison	Dp369	3	10.118	0.018	*
Madison	Dp375	Monomorphic			
Madison	Dp376	Monomorphic			
Madison	Dp38	1	0.291	0.589	ns
Madison	Dp401	6	24.896	0.000	***
Madison	Dp43	1	0.160	0.689	ns
Madison	Dp433	3	3.615	0.306	ns
Madison	Dp437	10	13.180	0.214	ns
Madison	Dp446	6	47.083	0.000	***
Madison	Dp461	6	6.572	0.362	ns
Madison	Dp90	3	0.624	0.891	ns

Key: ns=not significant, * P<0.05, ** P<0.01, *** P<0.001

Table S2. Individual alleles for each loci within the populations that unique that population and at what frequency (GenAlEx 6.41).

Pop	Locus	Allele	Freq
Frenchman	Dp43	323	0.500
Hill	Dp173	256	0.032
Hill	Dp291	149	0.009
Hill	Dp369	398	0.395
Hill	Dp375	144	0.009
Hill	Dp376	192	0.009
Hill	Dp43	329	0.009
Hill	Dp433	184	0.100
Hill	Dp461	179	0.005
Hill	Dp90	231	0.005
Madison	Dp173	270	0.283
Madison	Dp173	274	0.005
Madison	Dp291	161	0.027
Madison	Dp291	164	0.065
Madison	Dp369	400	0.418
Madison	Dp43	315	0.087
Madison	Dp437	270	0.011
Madison	Dp446	269	0.087
Madison	Dp461	187	0.022
Madison	Dp90	229	0.005

Figure S1. Allelic richness, diversity, and frequencies of polymorphic loci among populations (GenAlEx 6.41). The microsatellite loci Dp38, Dp 43, Dp90, Dp162, Dp283, Dp291, Dp375, and DP376 were not displayed because they were either monomorphic or only had two alleles.

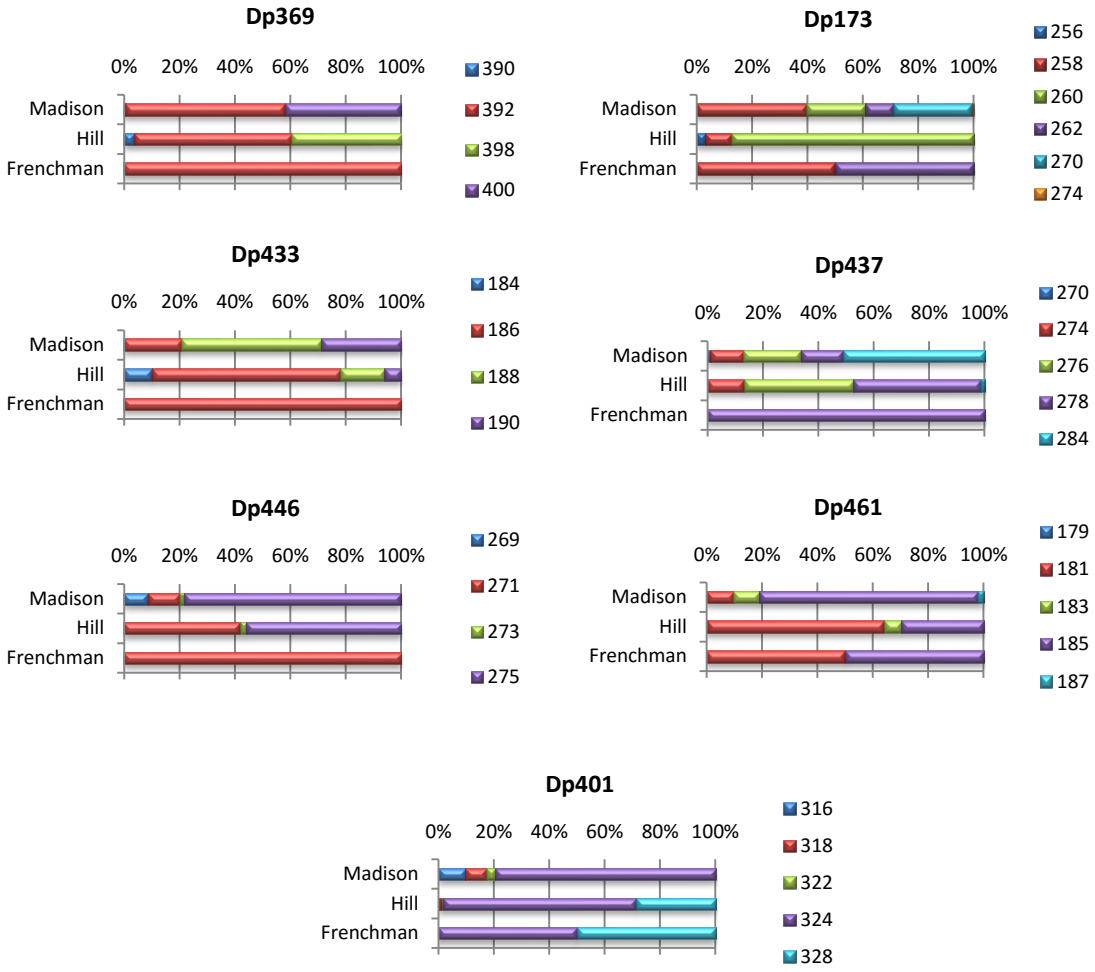
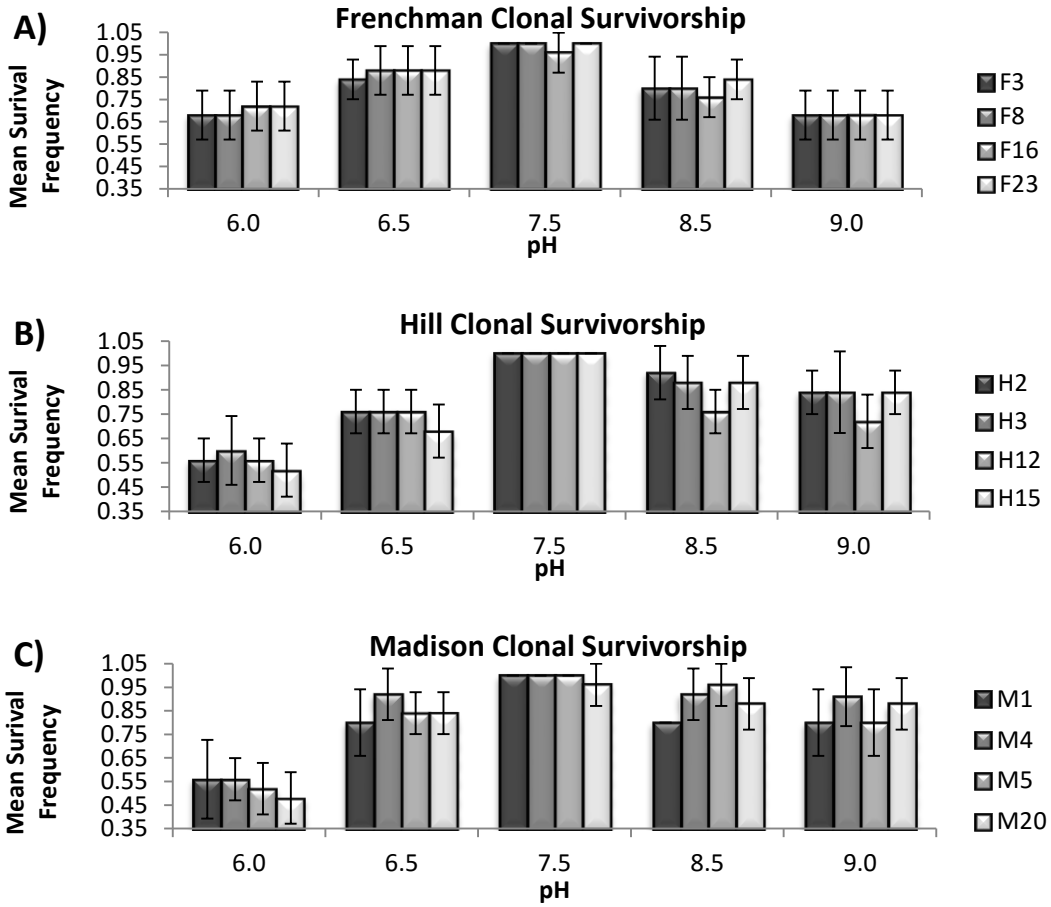


Figure S2. Results among clones for each population for the pH survivorship experiments . Mean survivorship (± 1 SD) are given for: (A.) clone-level survivorship for Frenchman Lake, (B.) clone-level survivorship for Hill Lake, and (C.) clone-level survivorship for Madison Lake.



Appendix B: Chapter Two Supplemental Material

Table S1 List of the α-CA sub-families with the species and accession number of the amino acid sequences used in this study.

α-CA Sub-Family	Species Name	Accession #
CA 10	Homo sapiens_CA10	AAH47456.1
	Mus musculus_CA10	AHH17606.1
	Macaca mulatta_CA10	AFE64410.1
	Equus caballus_CA10	XP_001503286.2
	Bos taurus_CA10	AAI26508.1
	Gallus gallus_CA10	XP_415644.1
	Taeniopygia guttata_CA10	XP_002198310.1
	Anolis carolinesis_CA10	XP_003216966.1
	Rattus norvegicus_CA10	XP_002727818.1
	Xenopus tropicalis_CA10	XP_002932546.1
	Danio rerio_CA10	NP_001032198.1
	Oreochromis niloticus_CA10	XP_003438532.1
CA11	Macaca mulatta_CA11	AFE64868.1
	Bos taurus_CA11	AAI23828.1
	Homo sapiens_CA11	AAH02662.1
	Mus musculus_CA11	AAH19392.1
	Rattus norvegicus_CA11	AAL78173.1
	Loxodonta africana_CA11	XP_003406877.1
	Anolis carolinesis_CA11	XP_003225067.1
	Strongiocentrotus purpuratus_CA11	XP_784796.2
	Branchiostoma floridae_CA11	XP_002611575
	Drosophila melanogaster_CA11a	NP_572581.1
	Tribolium castaneum_CA11a	XP_972474.2
	Nasonia vitripennis_CA11a	XP_003424234.1
	Drosophila melanogaster_CA11b	NP_572407.3
	Aedes aegyti_CA11	XP_0016588761.1
	Tribolium castaneum_CA11b	XP_974173.2
	Daphnia pulex_CA4	EFX88377.1
	Daphnia galeata_CA4	unpublished
	Daphnia pulex_CA3	EFX72738.1
	Daphnia galeata_CA3	unpublished
	Loa loa_CA11	XP_003137734.1
	Brugia malayi_CA11	XP_001900835.1
	Caernorhabditis elegans_CA11a	NP_495567.3
	Trichinella spiralis_CA11	XP_003376534.1
	Caernorhabditis elegans_CA11b	NP_498083.1
CA 8	Homo sapiens_CA8	EAW86826.1
	Macaca mulatta_CA8	EHH28528.1
	Mus musculus_CA8	AAH10773.1
	Rattus norvegicus_CA8	AAH87586.1
	Bos taurus_CA8	NP_001077159.1
	Equus caballus_CA8	XP_001496523.1
	Gallus gallus_CA8	XP_419221.3
	Meleagris gallopavo_CA8	XP_003205121.1
	Taeniopygia guttata_CA8	XP_002197204>1
	Anolis carolinesis_CA8	XP_003219630.1
	Xenopus tropicalis_CA8	NP_001011213.1
	Danio rerio_CA8	AAH92740.1
	Oncorhynchus mykiss_CA8	NP_001118116.1
	Strongiocentrotus purpuratus_CA8	XP_795365.3
CA 2	Homo sapiens_CA2	NP_000058.1
	Macaca mulatta_CA2	NP_001182346.1
	Equus caballus_CA2	XP_001488540.1
	Bos taurus_CA2	NP_848667.1
	Mus musculus_CA2	EDL05130.1
	Rattus norvegicus_CA2	NP_062164.1
	Gallus gallus_CA2	NP_990648.1
	Taeniopygia guttata_CA2	XP_002199833.1
	Anolis carolinesis_CA2	XP_003219585.1
	Xenopus tropicalis_CA2	AAH89661.1
	Xenopus laevis_CA2	NP_001080080.1

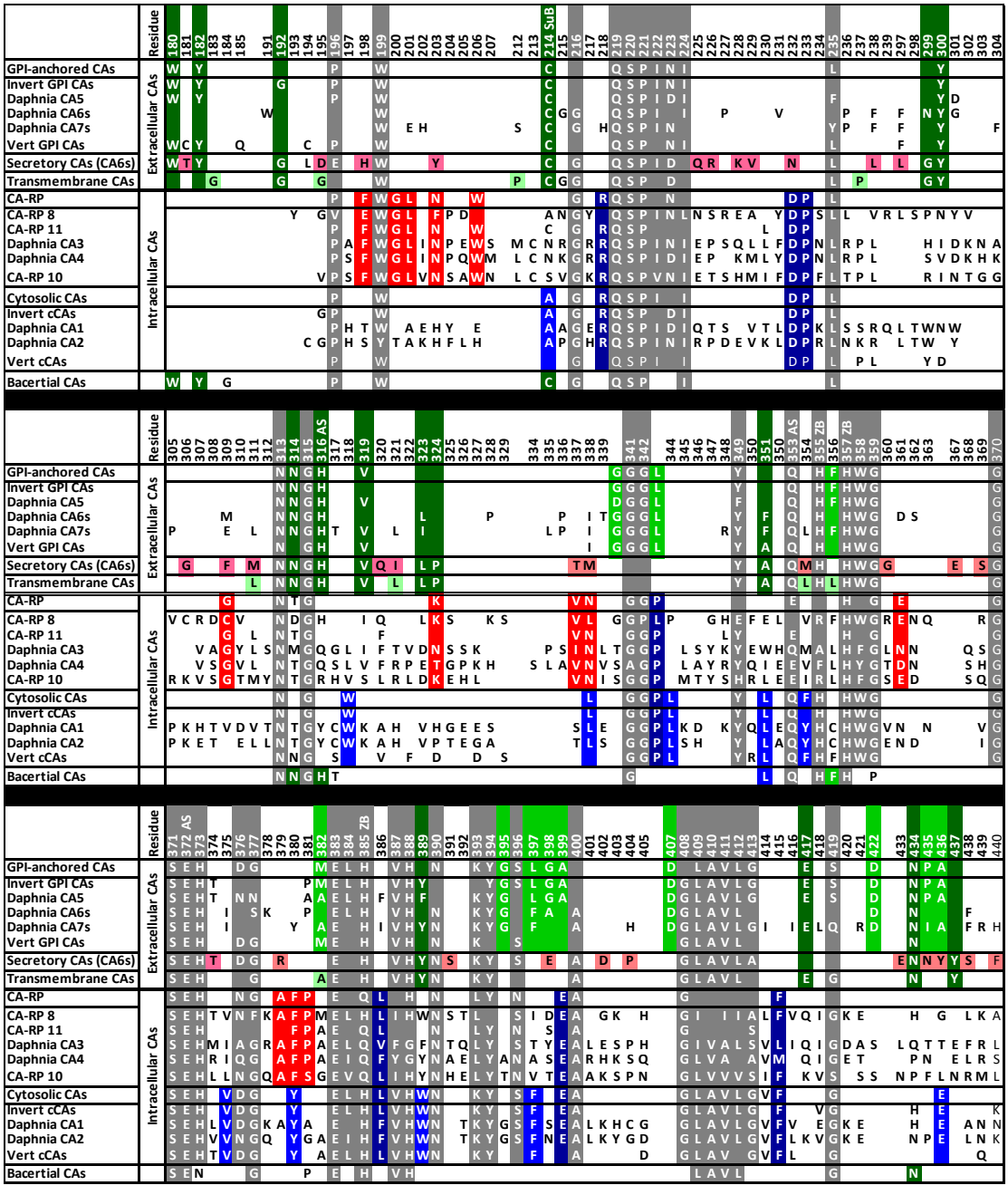
Piscean CA1/2	Oncorhynchus mykiss_CA1	NP_001117692.1
	Dicentrarchus labrax_CA1	CBN82139.1
	Oncorhynchus mykiss_CA2	NP_001117693.1
	Ictalurus punctatus_CA2	NP_001187560.1
	Danio rerio_CA2	NP_954685.1
CA 1	Homo sapiens_CA1	NP_001729.1
	Macaca mulatta_CA1	EHH28617.1
	Equus caballus_CA1	XP_001488605
	Mus musculus_CA1	AAA37354.1
	Rattus norvegicus_CA1	AAI58889.1
	Bos taurus_CA1	AAI16127.1
	Xenopus tropicalis_CA1	XP_002939198
CA 13	Homo sapiens_CA13	AAH52602.1
	Macaca mulatta_CA13	XP_001095487.1
	Equus caballus_CA13	XP_001489984.2
	Mus musculus_CA13	AAK16672.1
	Rattus norvegicus_CA13	NP_001128465.1
	Bos taurus_CA13	DAA22563.1
	Gallus gallus_CA13	XP_003640859.1
	Meleagris gallopavo_CA13	XP_003205200.1
	Python molurus_CA13	AEA49961.1
	Anolis carolinesis_CA13	XP_003219587.1
	Xenopus tropicalis_CA13	NP_001072448.1
	Xenopus laevis_CA13	NP_001086981.1
CA7	Petromyzon marinus_CA2	
	Homo sapiens_CA7	NP_005173.1
	Macaca mulatta_CA7	XP_001085299.2
	Mus musculus_CA7	NP_444300.1
	Rattus norvegicus_CA7	NM_001106165.1
	Loxodonta africana_CA7	XM_003417135.1
	Bos taurus_CA7	NP_001179451.1
	Gallus gallus_CA7	XP_41415.2
	Meleagris gallopavo_CA7	XP_003209867.1
	Taeniopygia guttata_CA7	XP_002190292.1
	Xenopus tropicalis_CA7	NP_001015903.1
	Danio rerio_CA7	NM_200813.1
	Ictalurus punctatus_CA7	NP_001187680.1
	Oreochromis niloticus_CA7	XM_003445759.1
CA 5	Homo sapiens_CA5a	NP_001730.1
	Macaca mulatta_CA5a	EHH31930.1
	Equus caballus_CA5a	XP_001500362.2
	Bos taurus_CA5a	DAA20167.1
	Mus musculus_CA5a	AAH30174.1
	Rattus norvegicus_CA5a	AAH88147.1
	Homo sapiens_CA5b	NP_00915.1
	Macaca mulatta_CA5b	EHH30558.1
	Equus caballus_CA5b	XP_001490399.1
	Bos taurus_CA5b	NP_001074377.2
	Mus musculus_CA5b	CAM26357.1
	Rattus norvegicus_CA5b	AAH81872.1
	Gallus gallus_CA5b	XP_414195.3
	Meleagris gallopavo_CA5b	XP_003209935.1
	Taeniopygia guttata_CA5b	XP_002192230.2
	Xenopus tropicalis_CA5b	CAJ82585.1
	Xenopus laevis_CA5b	NP_001084659.1
	Danio rerio_CA5b	NP_001104671.1
CA 3	Homo sapiens_CA3	3UYN
	Macaca mulatta_CA3	EHH28618.1
	Mus musculus_CA3	AAH11129.1
	Rattus norvegicus_CA3	AAH61980.1
	Equus caballus_CA3	NP_001157426.1
	Bos taurus_CA3	DAA22526.1

	Gallus gallus_CA3	NP_001264340.1
	Meleagris gallopavo_CA3	XP_003205196.1
	Taeniopygia guttata_CA3	XP_002199827.1
	Anolis carolinesis_CA3	XP_003219614.1
Invertebrate Cytosolic CA	Branchiostoma floridae_cCA2	XP_00259304.1
	Branchiostoma floridae_cCA3	XP_002605865.1
	Ixodes scapularis_cCA	XP_002416504.1
	Anthopleura elegantissima_cCA	AAD32675.1
	Strongylocentrotus purpuratus_cCA	XP_782997.2
	Drosophila melanogaster_cCA	AAF53332.1
	Anopheles gambiae_cCA	ABF66618.1
	Danaus plexippus_cCA	EHJ77985.1
	Tribolium castaneum_cCA	XP_974322.1
	Daphnia pulicaria_CA1	unpublished
	Daphnia pulex_CA1	EFX86103.1
	Daphnia galeata_CA1	unpublished
	Daphnia pulicaria_CA2	unpublished
	Daphnia pulex_CA2	EFX86394.1
	Daphnia galeata_CA2	unpublished
	Panaeus monodon_cCA	ABV65904.1
	Litopenaeus vannamei_cCA	ADM16544.2
	Callinectes sapidus_cCA	ABN51213.1
	Portunus trituberculatus_cCA	AFV46144.1
	Carcinus maenus_cCA	ABX71208.1
	Caligus rogercresseyi_cCA	ACO11717.1
	Caligus clemensi_cCA	ACO15216.1
	Lepeophtheirus salmonis_cCA	ACO12707.1
CA 4	Homo sapiens_CA4	NP_000708.1
	Macaca mulatta_CA4	XP_001107970.2
	Bos taurus_CA4	AAB09466.1
	Equus caballus_CA4	XP_001501182.1
	Mus musculus_CA4	AAH12704.1
	Rattus norvegicus_CA4	AAH97329.1
	Gallus gallus_CA4	XP_415893.1
	Python molurus_CA4	AEA49062.1
	Danio rerio_CA4	AAI09406
	Oreochromis niloticus_CA4	XP_003456174.1
	Xenopus tropicalis_CA4	XP_002983882.1
	Xenopus laevis_CA4	NP_001079788.1
CA15	mus musculus_CA15	NP_085035.1
	Rattus norvegicus_CA15	NP_001099371.1
	Equus caballus_CA15	XP_005613908.1
	Gallus gallus_CA15	XP_415218
	Meleagris gallopavo_CA15	XP_003211059.1
	Taeniopygia guttata_CA15	XP_002195938.2
	Anolis carolinesis_CA15	XP_003225153.1
	Xenopus tropicalis_CA15	XP_002931952.2
CA 6	Homo sapiens_CA6	NP_001206.2
	Macaca mulatta_CA6	XP_001099188.2
	Bos taurus_CA6	DAA21262.1
	Equus caballus_CA6	NP_001137421.2
	Mus musculus_CA6	NP_033932.2
	Rattus norvegicus_CA6	NP_001128313.1
	Gallus gallus_CA6	XP_425745.3
	Meleagris gallopavo_CA6	XP_003212306.1
	Taeniopygia guttata_CA6	XP_002187446.1
	Xenopus laevis_CA6	NP_001123372.1
	Xenopus tropicalis_CA6	NP_001085550.1
	Danio rerio_CA6	XP_002666525.1
	Homo sapiens_CA9	NP_001207.2
	Macaca mulatta_CA9	XP_001088481.1
	Equus caballus_CA9	XP_001504561.3

CA 9	Bos taurus_CA9	XP_002689744.1
	Mus musculus_CA9	AAI20545.11
	Rattus norvegicus_CA9	EDL98746.1
	Gallus gallus_CA9	XP_001233320.2
	Danio rerio_CA9	XP_694982.2
CA 12	Homo sapiens_CA12	NP_001209.1
	Macaca mulatta_CA12	EHH27385.1
	Mus musculus_CA12	AAH35941.1
	Rattus norvegicus_CA12	NP_001074225.1
	Bos taurus_CA12	XP_870878.3
	Gallus gallus_CA12	XP_413756.2
	Meleagris gallopavo_CA12	XP_003209429.1
	Xenopus tropicalis_CA12	CAJ81489.1
	Xenopus laevis_CA12	NP_001091227.1
CA14	Homo sapiens_CA14	BAA85002.1
	Macaca mulatta_CA14	NP_001244832.1
	Equus caballus_CA14	XP_001489157.1
	Bos taurus_CA14	NP_001179134.1
	Mus musculus_CA14	NP_035927.1
	Rattus norvegicus_CA14	NP_001103125.1
	Gallus gallus_CA14	XP_003642713.1
	Meleagris gallopavo_CA14	XP_003212828.1
	Xenopus tropicalis_CA14	NP_001103521.1
	Danio rerio_CA14	NP_001032782.1
	Oreochromis niloticus_CA14	XP_005810348.1
Daphnia CA 5/6/7	Daphnia pulex_CA6B	EFX88008.1
	Daphnia galeata_CA6B	unpublished
	Daphnia pulex_CA6E	EFX88011.1
	Daphnia galeata_CA6E	unpublished
	Daphnia pulex_CA6F	EFX88012.1
	Daphnia galeata_CA6F	unpublished
	Daphnia pulex_CA6C	EFX88009.1
	Daphnia galeata_CA6C	unpublished
	Daphnia pulex_CA6D	EFX88010.1
	Daphnia pulex_CA6G	EFX88013.1
	Daphnia pulicaria_CA5	unpublished
	Daphnia pulex_CA5	EFX81683.1
	Daphnia galeata_CA5	unpublished
	Daphnia pulex_CA6A	EFX88007.1
	Daphnia galeata_CA6A	unpublished
	Daphnia pulex_CA7D	EFX88103.1
	Daphnia galeata_CA7D	unpublished
	Daphnia pulex_CA7A	EFX88100.1
	Daphnia galeata_CA7A	unpublished
	Daphnia pulex_CA7B	EFX880101.1
	Daphnia galeata_CA7B	unpublished
	Daphnia pulex_CA7I	EFX88114.1
	Daphnia pulex_CA7M	EFX88119.1
	Daphnia galeata_CA7I	unpublished
	Daphnia galeata_CA7G	unpublished
	Daphnia pulex_CA7H	EFX88113.1
	Daphnia pulex_CA7L	EFX88118.1
	Daphnia pulex_CA7K	EFX88116.1
	Daphnia pulex_CA7O	EFX88121.1
	Daphnia pulex_CA7P	EFX88123.1
	Daphnia galeata_CA7P	unpublished
	Daphnia pulex_CA7J	EFX88115.1
	Daphnia pulex_CA7N	EFX88120.1
	Daphnia pulex_CA7Q	EFX77528.1
	Daphnia galeata_CA7H	unpublished
	Daphnia galeata_CA7J	unpublished
	Daphnia pulex_CA7C	EFX88102.1
	Daphnia pulex_CA7E	EFX88104.1

	Daphnia galeata_CA7C	unpublished
	Daphnia pulex_CA6H	EFX88180.1
Invertebrate GPI-anchored CA	Callinectes sapidus_gCA	ABN51214.1
	Portunus trituberculatus_gCA	AFV46145.1
	Carcinus maenus_gCA	ABX71209.1
	Branchiostoma floridae_gCA	XP_002601262.1
	Strongiocentrotus purpuratus_gCA	XP_796525.1
	Anopheles gambiae_gCA	AC528257.1
	Danaus plexippus_gCA	EHJ70026.1
	Patella vulgata_gCA	CCJ509593.1
Bacteria CA	Neisseria gonorrhoeae_CA	YP_207719.1
	Kingella denitrificans_CA	WP_003782874.1
	Pastuerella multocida_CA	WP_005752370.1
	Streptococcus mutans_CA	WP_002278508.1
	Thiorhodococcus drewsii_CA	WP_007041651.1
	Aeromonas caviae_CA	WP_010675444.1
	Yersinia enterocolotica_CA	CCQ39919.1
	Serratia sp_CA	YP_006027343.1
	Enterobacter cloacae_CA	EUL61977.1
	Klebsiella oxytoca_CA	EHS90795.1
	Pantoea vagans_CA	WP_13358087.1
	Thermovibrio ammonificans_CA	YP_004152175.1

Table S2 Listing of conserved amino acid residues for each group of α-CAs along with *Daphnia* α-CAs. The gray highlighted residues represent conserved sites throughout the phylogeny. Dark green represents sites conserved in all extracellular α-CAs, medium green the GPI-anchored CAs, light green the transmembrane CAs, pink the secretory CAs. The dark blue represents sites that are conserved in all intracellular α-CAs, medium blue the cytosolic CAs and red the CA-RPs. The bacterial group shows conserved sites across α-CAs. AS = active site, ZB = zinc-binding site, SB = substrate-binding site, GPI = purported GPI-anchoring site, SuB = disulfide bond site.



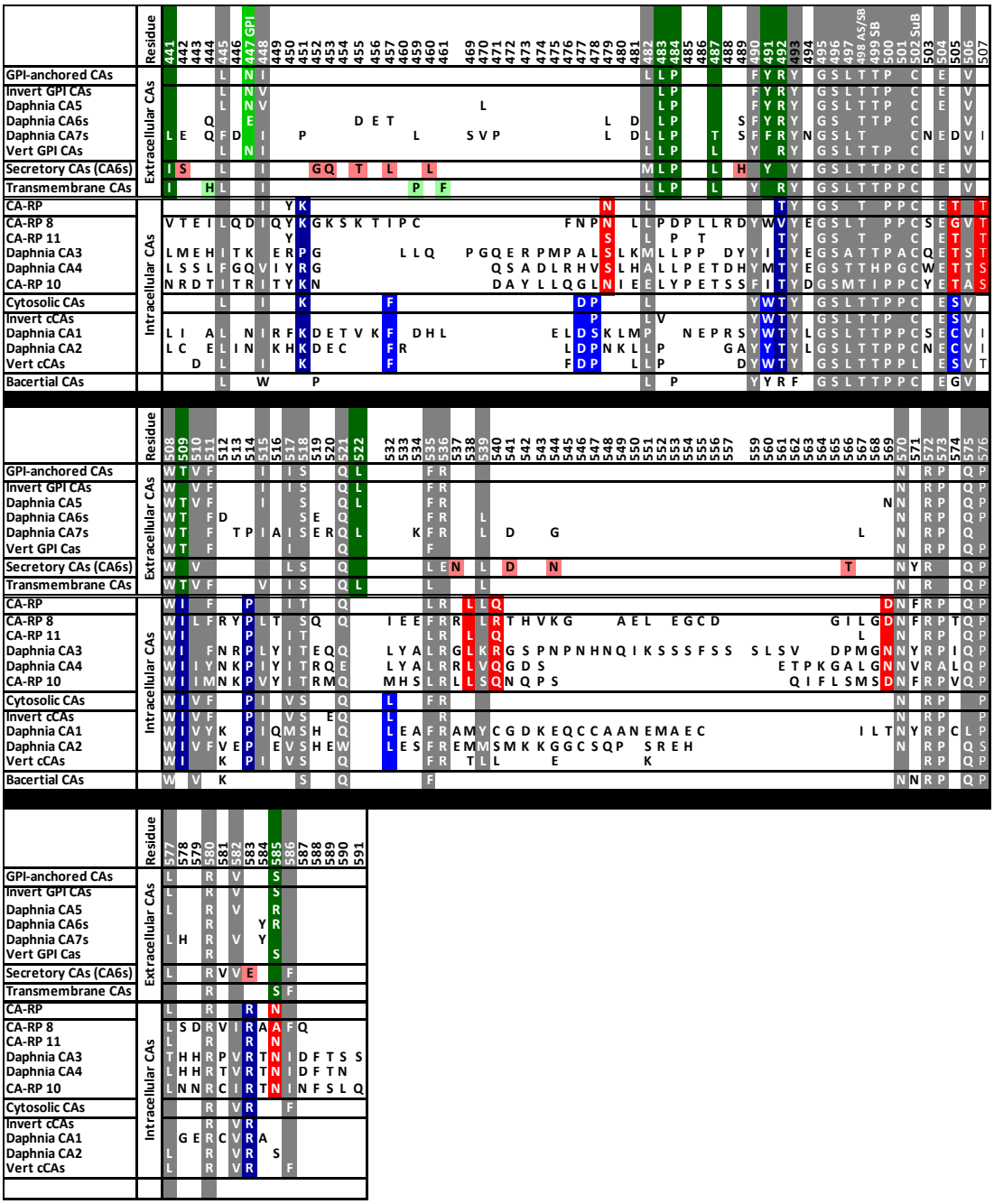
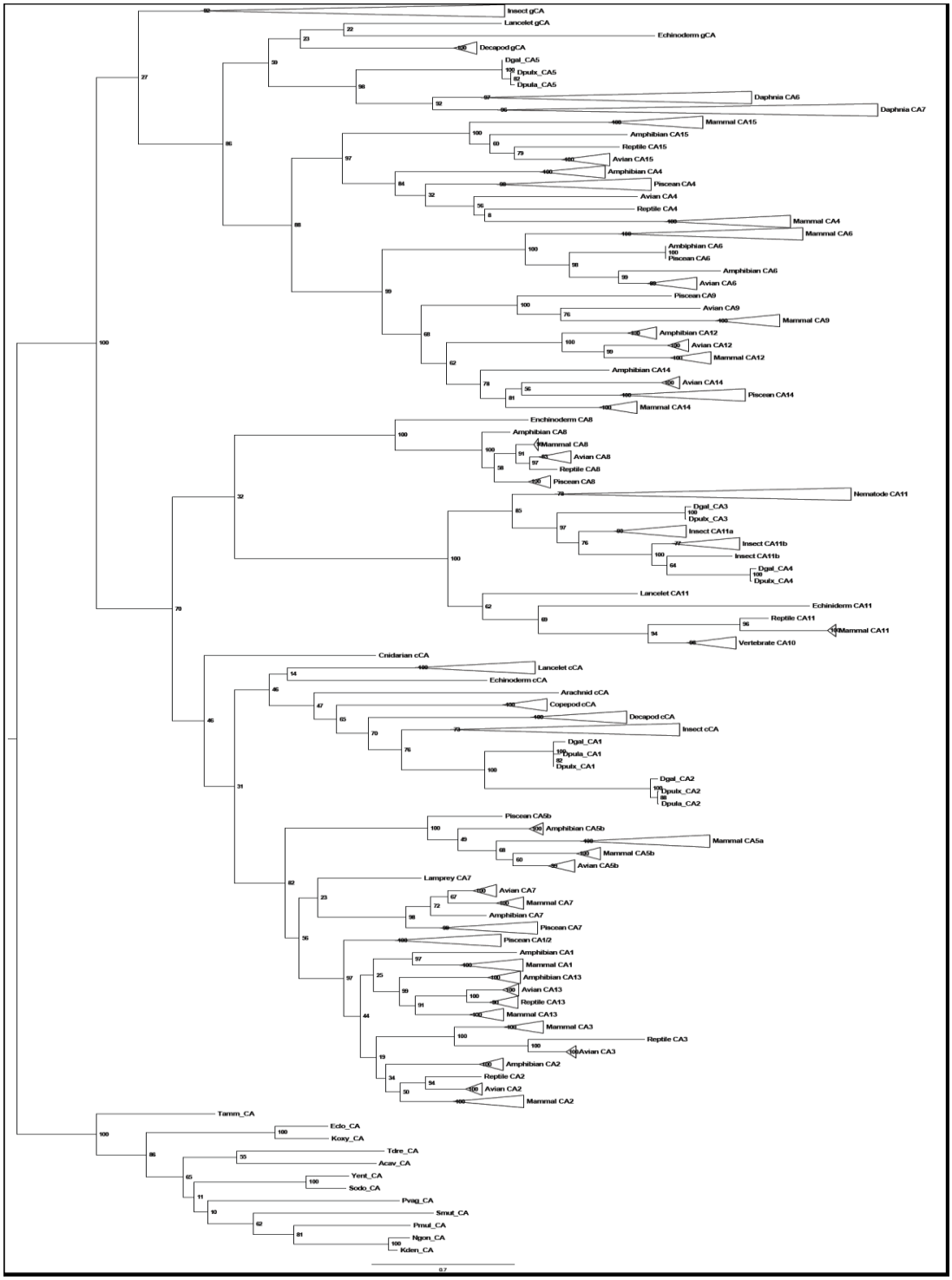

Residue
Extracellular CAs
GPI-anchored CAs
Invert GPI CAs
Daphnia CA5
Daphnia CA6s
Daphnia CA7s
Vert GPI CAs
Secretory CAs (CA6s)
Transmembrane CAs
Intracellular CAs
CA-RP
CA-RP 8
CA-RP 11
Daphnia CA3
Daphnia CA4
CA-RP 10
Cytosolic CAs
Invert cCAs
Daphnia CA1
Daphnia CA2
Vert cCAs
Bacertial CAs

Table S3 Ancestral states of amino acids at informative residues. Ancestral states were inferred using a Maximum Likelihood model in MEGA 5.0 and the most probable states at a given residue were selected. CA groups are based on the most recent common ancestor as determine by the phylogenic analysis, for example the hypothetical ancestral Animalia α-CA state is the predicted sequence of the most recent common ancestor of the extracellular α-CAs and intracellular α-CAs, while the extracellular α-CA hypothetical ancestor is the most recent common ancestor to the invertebrate and vertebrate clades of Extracellular α-CAs (Figure1). The residues are based on the alignment performed for phylogeny construction. Codes on the residues are as follows: SuB = disulfide bond site, AS = active site, ZB = zinc-binding site, and SB = substrate-binding site. Shaded amino acids reflect an amino acid change from the most common recent ancestor and the shaded amino acids in a box are amino acid change that resulted in convergent evolution. The no. of amino acid changes reflects the number of changes from the most common recent ancestor for that group.

Residue / CA Group	214 SuB	232	233	314	316 AS	318	319	351	353 AS	354	355 ZB	356	381	385 ZB	387	419	434	445	448	491	492	499 SB	501	505	512	573	578	No. of Amino Acid Changes
Bacterial	C	P	I	N	H	I	Q	L	Q	F	H	F	P	H	V	G	N	L	W	Y	R	T	P	G	K	P	N	
Ancestral Animalia CA	C	D	I	N	H	I	Q	L	Q	F	H	F	P	H	V	G	N	L	I	Y	R	T	P	G	K	P	N	2
Extracellular	C	D	I	N	H	V	D	L	Q	F	H	F	P	H	V	S	N	Y	I	Y	R	T	P	V	K	P	N	5
Invertebrate	C	D	I	N	H	V	T	A	Q	L	H	F	P	H	V	S	N	F	I	Y	R	T	T	V	E	P	N	6
Vertebrate	C	D	S	N	H	V	D	A	Q	F	H	F	P	H	V	S	N	Y	I	Y	R	T	P	V	E	P	N	3
GPI	C	D	S	N	H	V	D	A	Q	F	H	F	P	H	V	S	N	Y	I	Y	R	T	P	V	E	P	N	0
CA 4	C	D	S	N	H	V	D	A	Q	F	H	F	P	H	V	S	N	Y	I	Y	R	T	P	T	E	P	N	1
CA 15	C	D	S	D	H	L	S	A	Q	F	H	F	P	H	V	S	N	Y	V	Y	R	T	D	V	E	P	K	6
Secretory																												
CA6	C	D	P	N	H	I	S	A	Q	M	H	L	V	H	V	G	E	Y	I	Y	R	T	P	S	D	A	N	11
Transmembrane	C	D	P	N	H	L	S	A	Q	L	H	L	P	H	V	G	N	Y	L	Y	R	T	P	S	N	A	N	10
CA9	C	S	P	N	H	L	T	A	Q	L	H	L	P	H	V	G	N	Y	L	F	R	T	P	S	N	A	N	3
CA12	C	D	S	N	H	M	S	A	Q	L	H	L	A	H	V	G	N	Y	I	Y	R	T	P	S	R	Q	N	6
CA14	C	D	P	N	H	L	S	A	Q	L	H	L	P	H	V	G	N	Y	L	Y	R	T	P	S	N	A	N	0
Intracellular	C	D	P	T	H	S	F	L	Q	F	H	F	P	H	V	G	N	L	T	W	T	T	P	S	K	P	N	8
CA-RP	C	D	P	T	H	T	V	L	E	I	R	F	P	H	I	G	N	L	T	W	T	T	P	S	K	P	N	6
CA8	N	D	P	I	H	T	V	L	E	V	R	F	P	H	I	G	H	L	T	W	T	T	P	G	R	P	S	7
CA10	C	D	P	R	H	S	D	L	E	I	R	L	S	Q	I	G	N	N	N	I	T	I	P	T	N	P	N	12
CA11	C	D	P	R	S	S	D	L	E	I	R	L	P	Q	I	G	N	L	T	M	T	T	A	T	N	P	H	11
Cytolsolic	A	D	P	T	N	S	F	L	Q	F	H	F	P	H	V	G	H	L	T	W	T	T	P	S	K	P	K	4
Invertebrate	A	D	P	T	T	S	F	L	Q	F	H	A	P	H	V	G	H	L	T	W	T	T	P	S	K	P	K	2
Vertebrate	A	D	P	S	N	S	F	L	Q	F	H	F	P	H	V	G	H	L	T	W	T	T	P	S	K	P	K	1
CA5	G	D	P	W	N	S	F	L	Q	F	H	F	P	H	V	G	H	L	V	W	T	T	P	S	K	P	K	3
CA7	E	D	P	S	H	S	V	L	Q	F	H	F	P	H	V	G	H	L	T	W	T	T	P	S	K	P	K	3
Fish CA1/2	N	D	A	L	H	S	F	L	Q	F	H	F	P	H	V	G	N	L	L	W	T	T	P	S	K	P	K	5
CA1	N	D	P	V	H	S	F	L	Q	F	H	F	A	H	V	G	N	L	L	W	T	H	P	C	K	P	K	4
CA2	N	D	P	V	H	S	F	L	Q	F	H	F	A	H	V	G	K	L	L	W	T	T	P	C	K	P	K	4
CA3	K	D	P	L	N	T	C	L	Q	F	H	F	A	H	V	G	K	M	L	W	T	T	P	C	K	P	K	7
CA13	N	D	P	L	H	S	F	L	Q	F	H	F	A	H	V	G	N	L	T	W	T	V	P	S	K	P	K	3

Figure S1. Phylogeny of α-CAs inferred from a maximum-likelihood analysis performed with RaxML version 8.0 with 1000 iterations. Bootstrap values are indicated at the nodes. Species are collapsed within a larger taxonomical grouping.



Appendix C: Chapter Three Supplemental Material

Table S1. Primers for the amplification of CA1, CA2, and CA5 DNA in *D. pulicaria*. The table contains the primer name (Loci and location; Figure 1), primer sequence, and thermocycler settings.

Name	Primer Sequence (5' - 3')	Thermocycle Settings
CA1 F1.1	ATGGCTCACAACTGGGGTTAC	95° C - 3min;(95° C - 30sec; 70° (-1° C)- 40sec; 72°C- 90sec)x 6; (95° C- 30sec; 60° C- 40sec; 72°C- 90sec) x34; 72°C- 6 min
CA1 R1.1	CTACTCAGTTTAGGATCGAGCG	95° C - 3min;(95° C - 30sec; 70° (-1° C)- 40sec; 72°C- 90sec)x 6; (95° C- 30sec; 60° C- 40sec; 72°C- 90sec) x34; 72°C- 6 min
CA1 F1.2	GACCTCATACCTGGCTGAACAC	95° C - 3min;(95° C - 30sec; 63° (-1° C)- 40sec; 72°C- 90sec)x 6; (95° C- 30sec; 52° C- 40sec; 72°C- 90sec) x34; 72°C- 6 min
CA1 R1.2	GGATCTAGCTCTAGGTGATC	95° C - 3min;(95° C - 30sec; 63° (-1° C)- 40sec; 72°C- 90sec)x 6; (95° C- 30sec; 53° C- 40sec; 72°C- 90sec) x34; 72°C- 6 min
CA1 F2	GGATCTGAACATCTGGTCGACG	95° C - 3min;(95° C - 30sec; 69° (-1° C)- 40sec; 72°C- 90sec)x 6; (95° C- 30sec; 57° C- 40sec; 72°C- 90sec) x34; 72°C- 6 min
CA1 R2	CTGGCTCGAACACATCGTTCACC	95° C - 3min;(95° C - 30sec; 69° (-1° C)- 40sec; 72°C- 90sec)x 6; (95° C- 30sec; 57° C- 40sec; 72°C- 90sec) x34; 72°C- 6 min
CA2 F1	ATGGTCTGGACGGTTGGTG	95° C - 3min;(95° C - 30sec; 68° (-1° C)- 40sec; 72°C- 90sec)x 6; (95° C- 30sec; 57° C- 40sec; 72°C- 90sec) x34; 72°C- 6 min
CA2 R1	CGTGAGACAATGGACCTCC	95° C - 3min;(95° C - 30sec; 68° (-1° C)- 40sec; 72°C- 90sec)x 6; (95° C- 30sec; 57° C- 40sec; 72°C- 90sec) x34; 72°C- 6 min
CA2 F2	GTCCACACTCGTATACTGC	95° C - 3min;(95° C - 30sec; 70° (-1° C)- 40sec; 72°C- 90sec)x 6; (95° C- 30sec; 59° C- 40sec; 72°C- 90sec) x34; 72°C- 6 min
CA2 R2	GTGCTCACGACTTGGTTGGCTGC	95° C - 3min;(95° C - 30sec; 70° (-1° C)- 40sec; 72°C- 90sec)x 6; (95° C- 30sec; 59° C- 40sec; 72°C- 90sec) x34; 72°C- 6 min
CA5 F1	TTGGCCAGTGGCGGCTCGCATCATT	95° C - 3min;(95° C - 30sec; 70° (-1° C)- 40sec; 72°C- 90sec)x 6; (95° C- 30sec; 60° C- 40sec; 72°C- 90sec) x34; 72°C- 6 min
CA5 R1	GATCAGCGTCAAGGACTTGC	95° C - 3min;(95° C - 30sec; 70° (-1° C)- 40sec; 72°C- 90sec)x 6; (95° C- 30sec; 60° C- 40sec; 72°C- 90sec) x34; 72°C- 6 min
CA5 F2	GATCAGCGTCAAGGACTTGC	95° C - 3min;(95° C - 30sec; 70° (-1° C)- 40sec; 72°C- 90sec)x 6; (95° C- 30sec; 59° C- 40sec; 72°C- 90sec) x34; 72°C- 6 min
CA5 R2	CGTGCACCACGTGATCCAGGAC	95° C - 3min;(95° C - 30sec; 70° (-1° C)- 40sec; 72°C- 90sec)x 6; (95° C- 30sec; 59° C- 40sec; 72°C- 90sec) x34; 72°C- 6 min
CA5 F3	CGTCCAAGGAGGATAATCC	95° C - 3min;(95° C - 30sec; 70° (-1° C)- 40sec; 72°C- 90sec)x 6; (95° C- 30sec; 59° C- 40sec; 72°C- 90sec) x34; 72°C- 6 min
CA5 R3	GAGGATCGGACGTCAACGGTG	95° C - 3min;(95° C - 30sec; 70° (-1° C)- 40sec; 72°C- 90sec)x 6; (95° C- 30sec; 59° C- 40sec; 72°C- 90sec) x34; 72°C- 6 min
CA1 SP 5' Rev	CTGTGTCAGTCATCTGACGGATGTC	95° C - 3min;(95° C - 30sec; 55° C- 30sec; 72°C- 90sec) x 20; (95° C- 30sec; 44° C- 30sec; 72°C- 90sec) x30; (95° C- 30sec; 55° C- 30sec; 72°C- 90sec) x 30; 72°C- 6 min
CA1 SP 3' Fwd	GCAATGTATTGCGGAGACAAGG	95° C - 3min;(95° C - 30sec; 55° C- 30sec; 72°C- 90sec) x 20; (95° C- 30sec; 44° C- 30sec; 72°C- 90sec) x30; (95° C- 30sec; 55° C- 30sec; 72°C- 90sec) x 30; 72°C- 6 min
CA2 SP 3' Fwd	GCAATGTATTGCGGAGACAAGG	95° C - 3min;(95° C - 30sec; 55° C- 30sec; 72°C- 90sec) x 20; (95° C- 30sec; 42° C- 30sec; 72°C- 90sec) x30; (95° C- 30sec; 55° C- 30sec; 72°C- 90sec) x 30; 72°C- 6 min
CA5 SP 5' Rev	CGTTCGTTGTTAGTTGCGTCACG	95° C - 3min;(95° C - 30sec; 55° C- 30sec; 72°C- 90sec) x 20; (95° C- 30sec; 46° C- 30sec; 72°C- 90sec) x30; (95° C- 30sec; 55° C- 30sec; 72°C- 90sec) x 30; 72°C- 6 min
CA5 SP 3' Fwd	CACCGTTCACGTCCGATCCTC	95° C - 3min;(95° C - 30sec; 55° C- 30sec; 72°C- 90sec) x 20; (95° C- 30sec; 46° C- 30sec; 72°C- 90sec) x30; (95° C- 30sec; 55° C- 30sec; 72°C- 90sec) x 30; 72°C- 6 min

Table S2. Primers for qRT-PCR of CA1, CA2, CA5, and GAPDH in *D. pulicaria*. The table contains the primer name and primer sequence.

Name	Primer Sequence (5' - 3')
GAPDH Fwd	ACCACTGTCCATGCCATCACT
GAPDH Rev	CACGCCACAACTTTCCAGAA
CA1-qPCR_Fwd	ACTTATGCCAGCCAACGAAC
CA1-qPCR_Rev	ATTCGGCCATCTCATTTGCG
CA2-qPCR_Fwd	AACGAAGCACTCAAGTACGG
CA2-qPCR_Rev	ATGGGGTGCACTCATCTTTG
CA5-qPCR_Fwd	TACGGCAGTTTAGGCAATGC
CA5-qPCR_Rev	ACGACAGAGTCTCGTTCAACTC

Appendix D: Chapter Four Supplemental Material

Figure S1. Line and box diagrams with exons and introns of *D. pulicaria* Carbonic Anhydrase (CA) isoforms (CA1, CA2, and CA5). Primer locations (Table S11) are designated beneath the line and box diagram. Length (base pairs -bp) of the genes are designated from the 5' to 3' end. Diagram modified from Culver and Morton (*in review*).

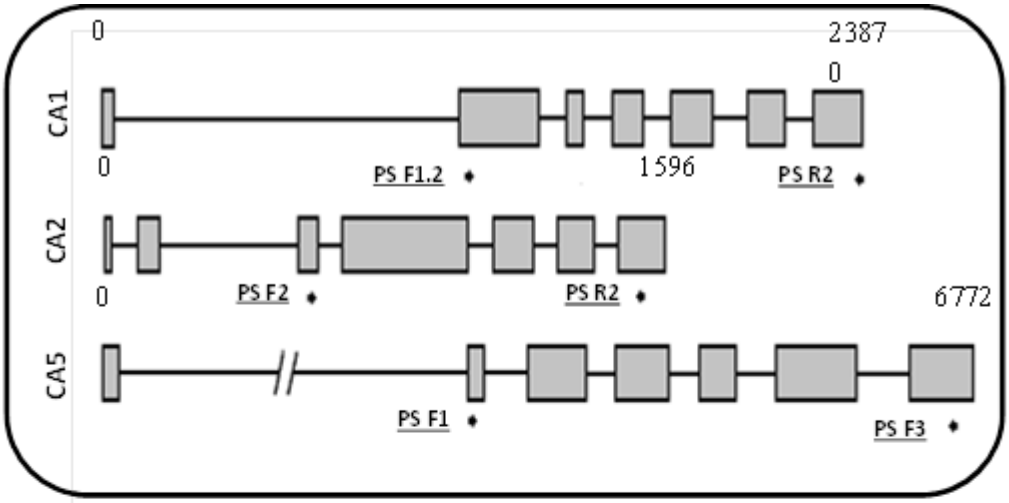


Figure S2. The evolutionary history of *Daphnia pulicaria* clonal lineages from the three study lakes (Frenchman – French; Madison – Mad; Hill) for the three Carbonic Anhydrase (CA) isoforms (CA1, CA2, CA5) was inferred using the Neighbor-Joining method. Bootstrap values with support >50% (based on 500 iterations) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Evolutionary analyses were conducted in MEGA5. Genotypes were inferred based upon clustering and were color-coded, with eight distinct genotypes detected. Those clones used for the survival experiment are boxed.

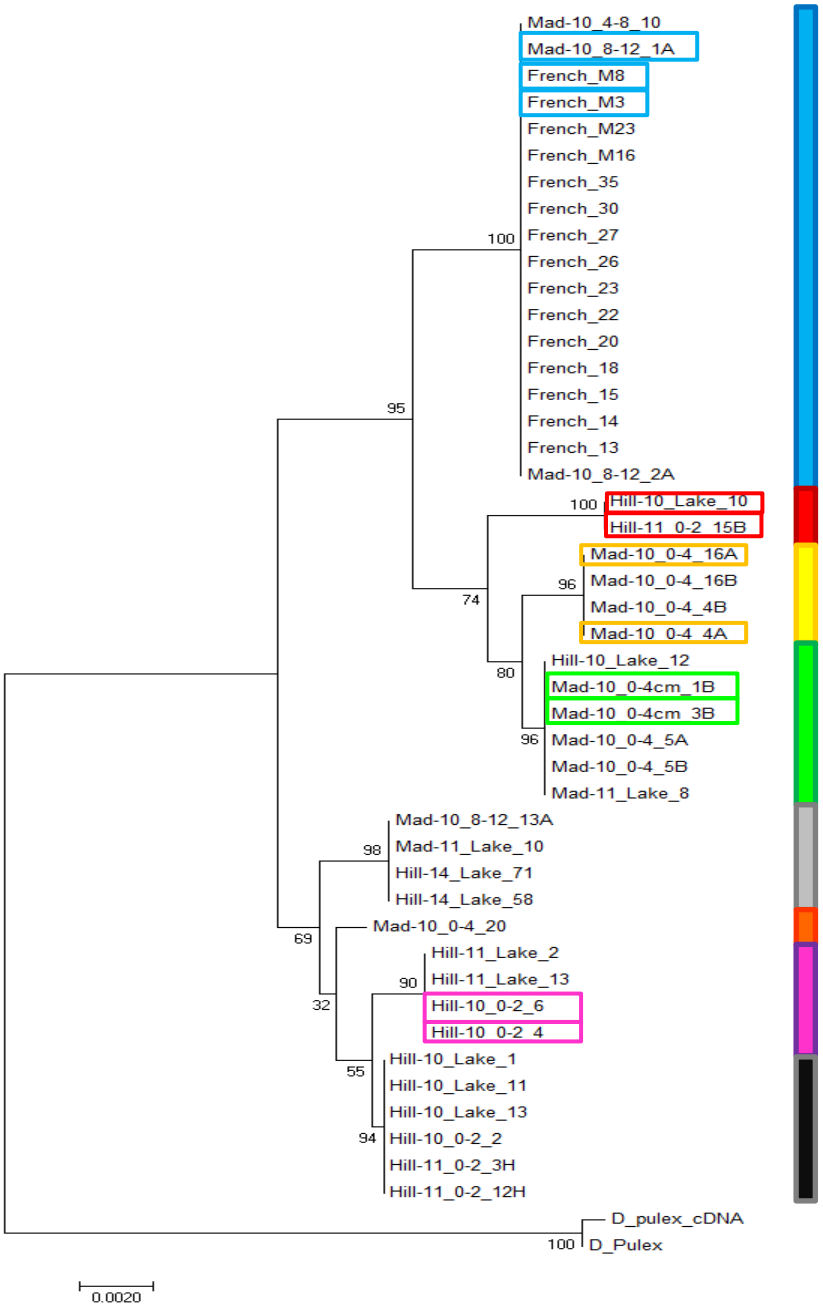


Table S3. Post-hoc comparisons of the genotype survivorship (fitness) experiment. As mentioned in the text, data were fitted to a generalized linear mixed-effects model, using a binomial distribution. Fixed factors were the pH treatment and the CA genotype, while the random factors were the jars in which the experiment took place and the different clonal lines under each genotype. We used four of the different genotypes as the first level in the CA genotype factor, as indicated in the subheadings, in order to make a pair-wise comparison between all genotypes. P-values were adjusted for multiple comparisons using the false discovery rate adjustment. Significant results are shown in bold font.

Green genotype as first level of genotype					
	Estimate	**Std Error**	**Z value**	***p* (> |z|)**	**Adjusted *p***
(Intercept)	-6.145	2.043	-3.008	0.003	**0.006**
pH	0.877	0.268	3.280	0.001	**0.003**
Blue genotype	10.683	2.608	4.096	< 0.001	**< 0.001**
Yellow genotype	-3.378	2.985	-1.132	0.258	0.397
Red genotype	-0.617	3.559	-0.173	0.862	0.931
Purple genotype	-4.141	3.744	-1.106	0.269	0.398
pH: Blue Genotype	-1.331	0.341	-3.904	< 0.001	**< 0.001**
pH: Yellow Genotype	0.547	0.394	1.387	0.165	0.288
pH: Red Genotype	0.112	0.467	0.240	0.810	0.926
pH: Purple Genotype	0.672	0.499	1.347	0.178	0.297
Blue genotype as first level of genotype					
	Estimate	**Std Error**	**Z value**	***p* (> |z|)**	**Adjusted *p***
(Intercept)	4.532	1.617	2.803	0.005	**0.011**
pH	-0.453	0.212	-2.142	0.032	**0.059**
Green genotype	-10.646	2.607	-4.083	< 0.001	**< 0.001**
Yellow genotype	-14.216	2.722	-5.222	< 0.001	**< 0.001**
Red genotype	-11.156	3.338	-3.342	0.001	**0.002**
Purple genotype	-14.775	3.536	-4.178	< 0.001	**< 0.001**
pH: Green Genotype	1.327	0.341	3.892	< 0.001	**< 0.001**
pH: Yellow Genotype	1.899	0.360	5.271	< 0.001	**< 0.001**
pH: Red Genotype	1.425	0.438	3.258	0.001	**0.003**
pH: Purple Genotype	1.996	0.472	4.229	< 0.001	**< 0.001**
Yellow genotype as first level of genotype					
	Estimate	**Std Error**	**Z value**	***p* (> |z|)**	**Adjusted *p***
(Intercept)	-9.403	2.174	-4.325	< 0.001	**< 0.001**
pH	1.408	0.290	4.861	< 0.001	**< 0.001**
Green genotype	2.913	2.984	0.976	0.329	0.439
Blue genotype	13.951	2.711	5.146	< 0.001	**< 0.001**
Red genotype	2.609	3.639	0.717	0.473	0.557
Purple genotype	-0.386	3.797	-0.102	0.919	0.931
pH: Green Genotype	-0.486	0.394	-1.232	0.218	0.348
pH: Blue Genotype	-1.863	0.359	-5.192	< 0.001	**< 0.001**

pH: Red Genotype	-0.416	0.480	-0.866	0.386	0.475
pH: Purple Genotype	0.074	0.508	0.145	0.885	0.931
Red genotype as first level of genotype					
	Estimate	**Std Error**	**Z value**	***p* (> |z|)**	**Adjusted *p***
(Intercept)	-6.499	2.906	-2.236	0.025	**0.048**
pH	0.956	0.381	2.506	0.012	**0.024**
Green genotype	0.306	3.544	0.086	0.931	0.931
Blue genotype	11.078	3.327	3.330	0.001	**0.002**
Yellow genotype	-3.107	3.630	-0.856	0.392	0.475
Purple genotype	-3.757	4.273	-0.879	0.379	0.475
pH: Green Genotype	-0.072	0.465	-0.154	0.877	0.931
pH: Blue Genotype	-1.415	0.436	-3.245	0.001	**0.003**
pH: Yellow Genotype	0.479	0.479	1.001	0.317	0.437
pH: Purple Genotype	0.589	0.568	1.037	0.300	0.428