UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

AN INTEGRATIVE ECOLOGICAL AND EVOLUTIONARY GENOMIC STUDY OF LAKE *DAPHNIA* ACROSS TIME

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

MATTHEW JOSEPH WERSEBE

Norman, Oklahoma

2023

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/337170

AN INTEGRATIVE ECOLOGICAL AND EVOLUTIONARY GENOMIC STUDY OF LAKE *DAPHNIA* ACROSS TIME

A DISSERTATION APPROVED FOR THE

DEPARTMENT OF BIOLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Lawrence J. Weider, Chair

Dr. Hayley Lanier

Dr. Robert Nairn

Dr. Laura Stein

Dr. Caryn Vaughn

© Copyright by MATTHEW JOSEPH WERSEBE 2023
All Rights Reserved.

ACKNOWLEDGEMENTS

During the last five years I have been incredibly lucky to meet, befriend and work with so many amazing people. First and foremost, I would like to extend my gratitude to my advisor Dr. Larry Weider without whose invaluable advice and mentorship this work would not have been possible. I am fortunate to not only count him as a colleague and mentor but as a friend. Next, I would like to acknowledge the many friends I have made, who have made this entire experience worthwhile. These include Anitzane Santaquiteria, Alex Franzen, Carmen Padraza, Addison Allen, Jared Jager and Nick Buss and others. You all always had great advice and insight. I would like to extend my sincerest gratitude to my coauthors and coworkers over the last few years to whom I am grateful for their time, expertise, and patience; Mark Edlund, Puni Jeyasingh, Ryan Sherman, and Emily Kiehnau. Finally, I want to thank my family, including my Parents Tim and Theresa who somehow found no issue with their son moving halfway across the country to pursue some vague plans of a PhD. Thank you for always believing in me and supporting me. Also, I would like to thank my grandmother, Shirley, who always reminded me of how proud she was of me studying the "fish and bugs and other things I'll never understand."

I am indebted to several folks who also helped in my success. These include members of my doctoral committee, Dr. Hayley Lanier, Dr. Laura Stein, Dr. Robert Nairn, Dr. Caryn Vaughn, and Dr. Daniel Allen; Department of Biology staff, K. Baker, G. Martin, and E. Cooley who do amazing things with so little. I would like to thank Dr. C. Curry, Dr. M. Laufersweiler, and the folks of OU's OSCER group who helped on many analyses.

TABLE OF CONTENTS

Acknowledgments ……... iv

Table of Contents ……..... v

List of Tables ……. viii

List of Figures …….. ix

Abstract …….. xi

Chapter 1 The roles of recombination and selection in shaping genomic divergence in an incipient ecological species complex ……..…….....…… 1
.

Abstract ……... 2

Introduction ……... 3

Methods ……... 6

Results ……. 12

Discussion ……. 19

Acknowledgements ……. 28

Author Contributions ……. 29

Data Availability ……. 30

Works Cited ……. 31

Figures & Tables ……. 42

Chapter 2 Does salinization impact long-term *Daphnia* assemblage dynamics? Evidence from the sediment egg bank in a small hard-water lake …… 50

Abstract ……. 51

Introduction ……. 52

Methods ……. 54

Results ……. 56

Discussion 58

Acknowledgements 63

Works Cited 64

Figures & Tables 71

Chapter 3 Resurrection genomics provides molecular and phenotypic evidence of rapid adaptation to salinization in a keystone aquatic species 76

Abstract 77

Introduction 79

Results 81

Discussion 86

Methods 94

Acknowledgements 98

Data Availability 99

Works Cited 100

Figures & Tables 106

Synthesis 110

Appendix 1 Supplemental Materials to Chapter 1 113

Methods 114

Figures 116

Tables 147

Works Cited 156

Appendix 2 Supplemental Materials for Chapter 2 …… 157

Methods …… 158

Figures & Tables …… 160

Works Cited …… 166

Appendix 3 Supplemental Materials for Chapter 3 …… 167

Methods …… 168

Works Cited …… 172

Figures …… 174

Tables …… 181

List of Tables

Chapter 1

Table 1 …….. 47

Table 2 …….. 48

Table 3 …….. 49

Chapter 2

Table 1 …….. 75

Appendix 1

Table S1 …… 147

Table S2 …… 148

Table S3 …… 152

Table S4 …… 153

Table S5 …… 154

Table S6 …… 155

Appendix 2

Table S1 …… 165

Appendix 3

Table S1 …… 181

Table S2 …… 182

Table S3 …… 186

Table S4 …… 187

List of Figures

Chapter 1

Figure 1 …… 42
Figure 2 …… 43
Figure 3 …… 44
Figure 4 …… 45
Figure 5 …… 46

Chapter 2

Figure 1 …… 71
Figure 2 …… 72
Figure 3 …… 73
Figure 4 …… 74

Chapter 3

Figure 1 …… 106
Figure 2 …… 107
Figure 3 …… 108
Figure 4 …… 109

Appendix 1

Figure S1 …… 116
Figure S2 …… 117
Figure S3 …… 118
Figure S4 …… 119
Figure S5 …… 120
Figure S6 …… 121
Figure S7 …… 122
Figure S8 …… 123
Figure S9 …… 124
Figure S10 …… 125
Figure S11 …… 126
Figure S12 …… 127
Figure S13 …… 128
Figure S14 …… 129
Figure S15 …… 130
Figure S16 …… 131
Figure S17 …… 132
Figure S18 …… 134
Figure S19 …… 135

Figure S20 136
Figure S21 137
Figure S22 138
Figure S23 139
Figure S24 140
Figure S25 141
Figure S26 142
Figure S27 143
Figure S28 144
Figure S29 145
Figure S30 146

Appendix 2

Figure S1 160
Figure S2 160
Figure S3 161
Figure S4 162
Figure S5 163
Figure S6 164

Appendix 3

Figure S1 174
Figure S2 175
Figure S3 176
Figure S4 177
Figure S5 178
Figure S6 179
Figure S7 180

ABSTRACT

An undisputable fact of the modern age is that human activities are now a major force shaping the biosphere. Some have called for a new geological epoch called the "Anthropocene" or the age of humans— the search is underway for a reliable and unambiguous mark in the geological record for the designation of this new epoch. As a biologist, however, I am reminded daily of the "marks" this age has left on the world around us. The American philosopher and ecologist Aldo Leopold perhaps said it best when he described an "ecological education" as "living alone in a world of wounds." What follows in this dissertation is maybe best described as searching for wounds, for marks, in biological archives that allow us to better understand the ecosystems of the Anthropocene.

My dissertation is focused primarily on studying the widespread crustacean zooplankter, *Daphnia pulicaria.* A common refrain in the following chapters will be to point out the function of this and related species in lake ecosystems and their value to the humans that enjoy and benefit from lakes. *Daphnia* in lakes are important for two reasons; first, they are keystone species of the pelagic food webs, connecting primary production from algae to higher trophic levels namely fish. *Daphnia* thus support the recreational and commercial fisheries of freshwater lakes. A knock-on benefit of *Daphnia* trophic position is they control the standing crop of algae in freshwater ecosystems and maintain water clarity, so lakes are not choked with noxious algal blooms. My dissertation is separated into three chapters. In the first chapter, I sequence, assemble, and annotate a genome for *D. pulicaria* using the latest long-read DNA sequencing technology. We use the genomic resources developed for this species to better understand its evolutionary history, especially its split from the closely related "sister" species *D. pulex.* This

reference genome is important for enabling other work, a thread we pick up in the third and final chapter.

In my second chapter, I chronicled the 175-years history of *D. pulicaria* and other *Daphnia* species in a small lake that will be the primary focus of the last two chapters. Tanners Lake is an ecosystem replete with wounds from the numerous human activities that dominate the landscape surrounding it. Due to their landscape position, lakes integrate vast amounts of information about their watersheds in their sediments. Tanners Lake records in its sediments the history of a landscape dominated by humans with the development of a major city surrounding it. Located just outside of Saint Paul, MN, Tanners Lake is impacted by two major human impacts inflicted on Northern Temperate Lakes. It is not only eutrophic, from the export of unprecedented amounts of nutrients but also it is severely salinized. Freshwater salinization is a consequence of the widespread use of de-icing salts on impervious surfaces such as roads and parking lots- a ubiquitous feature of human-dominated landscapes. I collected sediment cores from Tanners Lake to reconstruct the ecological dynamics of the *Daphnia* community across time in this lake by examining the abundance and diversity of resting eggs (encased in durable sclerotized structures called ephippia) across time in the core. I found that only modest changes in diversity and abundance occurred in the lake during salinization. This result suggests that *Daphnia* may be resilient to the threat of salinization- perhaps maintaining the ecosystems they support despite salinization.

The results of my second chapter set up an important question to tackle in my third chapter. Since *D. pulicaria* remains in Tanners Lake despite salinization, is this population evolving higher tolerance to these conditions? We tackle this question using an approach that is somewhat unique to *Daphnia*, by hatching the eggs contained in the ephippia from across time.

This method known as Resurrection Ecology, allows us to sample individuals from across time and study their phenotypes and genotypes. I hatched *Daphnia* from across approximately 25 years (~1994-2019) and resequenced their genomes. In addition, we also evaluated these *D. pulicaria* clones for their tolerance to salinity using phenotypic (i.e., survivorship) assays. I compiled this data set together to understand the evolution of this population through time. The genomic data supports the idea that salinity is a driving force for evolution in this population. In particular, genes related to osmoregulation and salinity tolerance are enriched within the statistical outliers. The phenotypic data supported this finding, as we observed that the salinity tolerance of modern *Daphnia* was higher than that of the ancestors hatched from the sediment.

Interestingly, while I initially described this work as a search for ecological and or evolutionary wounds, I was surprised to find that while they exist in very real ways in the biological archives studied, they are not mortal. My research, taken together, suggests that *Daphnia* populations should have the potential to respond to human threats and evolve to maintain the ecosystems they support. However, this resiliency will be highly dependent on the speed and intensity of the threats that these systems will face, as well as the strength of the evolutionary forces (i.e., selection, drift, migration, mutation) that will shape the underlying genetic structure of these populations.

The roles of recombination and selection in shaping genomic divergence in an incipient ecological species complex

Published in: *Molecular Ecology*, 04 February 2022

https://doi.org/10.1111/mec.16383

Matthew J. Wersebe[1], Ryan E. Sherman[2], Punidan D. Jeyasingh[2], Lawrence J. Weider[1]

[1] Program in Ecology and Evolutionary Biology, Department of Biology, University of Oklahoma, Norman, OK

[2] Department of Integrative Biology, Oklahoma State University, Stillwater, OK

Abstract:

Speciation genomic studies have revealed that genomes of diverging lineages are shaped jointly by the actions of gene flow and selection. These evolutionary forces acting in concert with processes such as recombination and genome features such as gene density shape a mosaic landscape of divergence. We investigated the roles of recombination and gene density in shaping the patterns of differentiation and divergence between the cyclically parthenogenetic ecological sister-taxa, *Daphnia pulicaria* and *Daphnia pulex*. First, we assembled a phased chromosome-scale genome assembly using trio-binning for *D. pulicaria* and constructed a genetic map using an F2-intercross panel to understand sex-specific recombination rate heterogeneity. Finally, we used a ddRADseq data set with broad geographic sampling of *D. pulicaria*, *D. pulex*, and their hybrids to understand the patterns of genome-scale divergence and demographic parameters. Our study provides the first sex-specific estimates of recombination rates for a cyclical parthenogen, and unlike other eukaryotic species, we observed male-biased heterochiasmy in *D. pulicaria*, which may be related to this somewhat unique breeding mode. Additionally, regions of high gene density and recombination are generally more divergent than regions of suppressed recombination. Outlier analysis indicated that divergent genomic regions are probably driven by selection on *D. pulicaria*, the derived lineage colonizing a novel lake habitat. Together, our study supports a scenario of selection acting on genes related to local adaptation shaping genome-wide patterns of differentiation despite high local recombination rates in this species complex. Finally, we discuss the limitations of our data in light of demographic uncertainty.

Introduction:

One of the major goals of ecological and evolutionary genomics is to determine the origin of species and understand the mechanisms promoting diversification (Orsini et al., 2013; Ungerer et al., 2008). This pursuit is especially interesting when closely related species' ranges broadly overlap, but the species nevertheless appear to occupy distinct ecological niches (Campbell et al., 2018). Given the proximity of species with overlapping ranges, the potential for gene flow is high and a powerful force preventing diversification because it breaks down linkage between adaptive variants (Turelli et al., 2001). Despite the homogenizing effects of gene flow, in cases of so called "ecological speciation" divergent or disruptive selection on nascent species still drives the evolution of distinct ecotypes with traits sorted with respect to their environment (Rundle & Nosil, 2005; Schluter, 2009). As a result, the evolution of ecotypes under ecological speciation with gene flow will be associated with the genome becoming a mosaic of regions showing the footprints of selection and differing magnitudes of gene flow (Via, 2009). Presumably, the regions resistant to gene flow and coincidental with the signatures of selection host genes related to adaptive traits. Determining the factors that shape this mosaic of selection and gene flow in ecological species can reveal the action of diversification mechanisms.

Speciation genomic studies using large-scale resequencing of species pairs that have recently diverged or are undergoing divergence have identified several different genome-wide mosaic patterns. Specifically, many studies have identified so called "islands of speciation" or the now more widely accepted term "islands of divergence" (Cruickshank & Hahn, 2014; Feder et al., 2012; Nosil et al., 2009; Turner et al., 2005). Islands of divergence present themselves in genome scans as regions along chromosomes with population genetic summary statistics elevated above background levels (e.g., F_{ST}). In many cases such islands are thought to be

regions experiencing selection or hosting so-called “speciation genes” that reinforce nascent species boundaries (Feder et al., 2012; Hejase et al., 2020). Theory and empirical work suggest that the recombination landscape is critical for shaping the divergence landscape in species pairs, where low recombination (e.g., pericentromeric regions) is thought to promote the formation of islands by increasing the efficacy of selection to remove shared polymorphisms, while regions of high recombination provide opportunities for the same alleles to segregate in both species (Feder & Nosil, 2010; Felsenstein, 1981; Nachman & Payseur, 2012; Ravinet et al., 2017; Wolf & Ellegren, 2017). Additionally, there is potential for the nonuniform density of functional elements (e.g., genes) within the genome to also contribute to heterogenous divergence patterns. For instance, in the case of ecological speciation, divergent and/or disruptive selection is expected to concentrate genomic divergence in the genome around regions hosting genes associated with adaptation to different habitats. In general, gene-rich regions in the genome host higher recombination rates (Fu et al., 2001). Under this scenario, the relationship of recombination and linkage shaping divergence is expected to be more complex, especially so with increasingly polygenic adaptive architectures (Kautt et al., 2020; Rundle & Nosil, 2005). The relative importance of gene density and recombination rate in ecological speciation is not as well resolved, given that many previous studies have focused primarily on recombination (Carneiro et al., 2009; Nachman & Payseur, 2012).

The *Daphnia pulex* species complex is a widely distributed group of freshwater crustacean zooplankton (Class: Branchiopoda) that inhabit lakes and ponds across the Holarctic (Colbourne et al., 1998; Weider et al., 1999). The *D. pulex* species complex consists of a collection of morphologically similar, but ecologically distinct species (Colbourne et al., 1998; Crease et al., 2012), which includes species that have arisen from hybridization and

polyploidization events (Chin & Cristescu, 2021; Dufresne & Hebert, 1997). Although many of the evolutionary relationships within the *D. pulex* group remain poorly resolved at the genomic level (Ye et al.,2021), *Daphnia pulex* and *Daphnia pulicaria* are thought to represent a case of rapid ecological speciation (Chin & Cristescu, 2021; Pfrender et al., 2000). Specifically, *D. pulex* and *D. pulicaria* are largely restricted to distinct ecological niches, where *D. pulex* is the ancestral pond ecotype and *D. pulicaria* is the recently derived lake ecotype (Crease et al., 2011; Cristescu et al., 2012). D. pulicaria serves as a keystone species in many lake food webs, linking primary production from algae to higher trophic levels (Carpenter et al., 1987; Lampert & Sommer, 2007; e.g., fish) and ultimately provides key ecosystem services related to the maintenance of water clarity and supporting recreational and commercial fisheries (Walsh et al., 2016). Adaptation of *D. pulicaria* to fish predators and a permanent habitat has driven divergence in multiple physiological and life-history traits that distinguish it from *D. pulex* (Dudycha, 2004; Dudycha & Tessier, 1999; Heier & Dudycha, 2009). *D. pulex* usually inhabits fishless and temporary woodland or vernal ponds that are inhabited by gape-limited invertebrate predators (Spitze, 1991). Additionally, the species boundary between *D. pulex* and *D. pulicaria* is fluid, where F1 hybrids are readily found in nature and are able to successfully backcross onto both parental forms in the laboratory (Chin et al., 2019; Heier & Dudycha, 2009) with measurable gene flow across the species boundary (Omilian & Lynch, 2009). However, F_1 individuals and reticulated individuals are generally obligate asexuals, producing clonal offspring where an introgressed *D. pulicaria* haplotype suppresses meiosis on the *D. pulex* background (Tucker et al., 2013; Xu, Spitze, et al., 2015; Ye et al., 2019). Despite apparent genetic incompatibilities, these obligate asexuals are generally fit and exhibit heterosis. There are several examples of such lineages successfully invading new habitats and displacing native

zooplankton assemblages (Mergeay et al., 2006; Ye et al., 2021). Taken together, these findings suggest that the pulex-pulicaria relationship is an excellent model system to observe how gene and functional element density versus recombination shape patterns of genomic divergence during speciation.

Here, we present a phased chromosome-level genome assembly, gene and repeat annotation, and high-density linkage map for *Daphnia pulicaria*. First, we examined the pattern of recombination rate heterogeneity in the *D. pulicaria* genome. Next, we used our assembly, annotation, and linkage map to analyze a publicly available population genomics data set consisting of *D. pulicaria*, *D. pulex* and their hybrids to determine the patterns of genome-scale divergence between these incipient ecological species. Specifically, utilizing the unique confluence of genomic resources available for this species complex, we analyzed whether genomic divergence is driven primarily by increased efficacy of selection to remove shared polymorphism in regions of low-recombination or by selection in gene-rich regions during adaptation to a novel habitat. Specifically, we tested the prediction that islands of divergence would be primarily restricted to low recombination regions (e.g., centromeres; sensu Carneiro et al., 2009).

Materials and Methods:

Recombinant panel:

The grand-parental lines of the F2-intercross panel were hatched from diapausing eggs (a.k.a. ephippia) recovered from the sediments of South Center Lake (Chisago County, MN) in 2010. South Center Lake is a deep dimictic lake that has experienced significant human-induced eutrophication associated with the conversion of its watershed to agriculture by Euro-American

settlers beginning c. 1850 (Frisch et al., 2014, 2017). The grand-paternal line was recovered from the 20–24 cm depth interval within the core and was deposited in the 1960s (hereafter referred to as 2X). The grand-maternal line was established from the 4–8 cm depth interval and was deposited in the early 2000s (hereafter referred to as 13B). 2X was known to readily produce clonal male offspring, which made it an ideal candidate to serve as a sire clone in laboratory crossing experiments. The details of core collection, sediment processing, dating, and hatching can be found in Frisch et al. (2014). Each clone was maintained asexually for several years in the lab before they were crossed to form an initial F_1 (parental) clone, hereafter F1–1A. Further details about the recombinant panel are published in Sherman et al. (2021) and in the supplemental methods. Genomic DNA was extracted from isoclonal pools of approximately 25 adult individuals from the F0, F1 and 100 haphazardly selected F2 lines using a modified CTAB protocol (Doyle & Doyle, 1987). The resulting DNA was prepared as 350 bp average insert size libraries and sequenced on nine lanes of an Illumina HiSeq using 150 bp paired end reads by Novogene (Sacramento, CA) to a mean depth of 20×.

Primary genome assembly:

DNA was also subsequently extracted from F1-1A using the method of Athanasio et al. (2016) with slight modifications for the isolation of high molecular weight (HMW) DNA for de novo assembly sequencing (see Supporting Information). The HMW DNA was sent to the University of Maryland Institute of Genome Science for library preparation and Circular Consensus Sequencing (CCS) on a single 8 M SMRT cell of a Pacific Biosciences Sequel II instrument. The resulting reads were passed to Hifiasm (Cheng et al., 2021) and assembled using default parameters in trio-binning mode. Since short-read Illumina data were available for each of the F0 genotypes K-mer hash tables for HiFiasm's trio-binning mode were constructed using YAK

(https://github.com/lh3/yak). The haplotype-resolved assembly graphs were converted to FASTA format using Gfatools (https://github.com/lh3/gfatools).

Assembly scaffolding:

The 100 F2 Illumina short-read libraries that generated sufficient data were quality filtered using Adaptorremoval2 (Schubert et al., 2016) and aligned to the 13B haplotype FASTA file using Bowtie2 (Langmead & Salzberg, 2012). The resulting SAM files were then passed to Samtools (Li et al., 2009), where they were sorted, mates were fixed and PCR and optical duplicates marked. BAM files were finally passed to Freebayes (Garrison & Marth, 2012) for variant calling. The resulting VCF file was filtered using VCFtools (Danecek et al., 2011) to a set of high-confident biallelic variants. The filtered VCF from the 13B alignment was passed to Lepmap3 (Rastas, 2017) for the construction of a genetic linkage map using the F2-panel and the two F0-grandparents using the standard modules (see supplemental). The resulting genetic map was used to scaffold the 13B haplotype using Lepanchor (Rastas, 2020). We estimated recombination rates along Linkage Groups (LGs) using custom R scripts. After scaffolding was complete, any unplaced scaffolds were extracted from the assembly and we used Blastn (Camacho et al., 2009) to search these sequences against a database including previously published Daphnia genomes and the D. pulex mitochondrial reference (Table S1; Crease, 1999). Any sequences that did not show homology to previously published Daphnia sequences were further searched against the NCBI prokaryote database. Since none had significant homology to any sequences, they were kept in the final assembly. Sequences that had homology to the mitochondrial genome were removed from the final assembly. The primary assembly of the 2X haplotype was scaffolded by alignment using Ragtag and Nucmer (Alonge et al., 2019; Marçais et al., 2018). Contamination was removed from this assembly following the procedure above.

Completeness of the two assemblies was assayed using BUSCO (Simão et al., 2015) scanning against the Arthropod v10 gene set from OrthoDB (Kriventseva et al., 2019).

Assembly annotation:

We developed a repeat DNA library for Daphnia pulicaria de novo by using the program Repeatmodeler (Flynn et al., 2020) which calls on Recon (Bao & Eddy, 2002) and Repeatscout (Price et al.,2005) followed by Rmblast based on the 2X haplotype primary assembly. A soft masked FASTA file was created using the de novo library using the program Repeatmasker (http://www.repeatmasker.org) for each of the scaffolded haplotypes. We extracted Kimura two-parameter genetic distances (Kimura, 1980) for the annotated repeats using the auxiliary Perl scripts provided in Repeatmasker. To annotate each of the soft masked genomes, we trained Augustus (Stanke et al., 2006) following the protocols outlined in Hoff and Stanke (2018) using protein alignments derived from the D. pulex protein sequences from Ye et al. (2017). These proteins were aligned to the D. pulicaria references using Genomethreader (Gremme et al., 2005) to generate hints for predictions. The resulting gene predictions were provided with functional annotation using Interproscan (Jones et al., 2014).

Speciation genomics:

We downloaded a publicly available ddRADseq (Peterson et al., 2012) data set that consisted of 41 *D. pulicaria* libraries, 82 *D. pulex* libraries and 36 hybrids and reticulated individuals between these two sister species (Clifford, 2016; NCBI Study Accession: PRJNA288877).

Bioinformatics:

We aligned the paired-end data to the scaffolded 13B haplotype reference (see above) using the BWA mem algorithm, and as above, the output was piped directly to Samtools and was sorted,

duplicates were marked and RG headers were added to the BAM files (Li & Durbin, 2009; Li et al., 2009). We chose to use the 13B haplotype because it was more complete than 2X and was associated with our linkage map and recombination rate estimates. Next, we called variants using the ref_map.pl pipeline of Stacks (Catchen et al., 2013), designating *D. pulex*, *D. pulicaria* and their hybrids as separate populations in the pipeline. Finally, we constructed one VCF file with one SNP per RADtag (hereafter "single SNP"), and another containing all called variants in the population (hereafter "all SNPs").

Population structure:

We analysed the population structure of the samples in the single SNP data set using two approaches. First, we used the fastSTRUCTURE (Raj et al., 2014) to calculate posterior probabilities of population assignment for each sample to two nominal species groupings and next used principal components analysis (PCA) implemented in Adegenet (Jombart, 2008).

Relation between differentiation and recombination rate:

We computed the per site FST between *D. pulicaria* and *D. pulex* samples for each SNP included in the single SNP VCF file using VCFtools (Danecek et al., 2011; Weir & Cockerham, 1984). In R, we computed the 95th percentile outliers and extracted the position of these SNPs on each chromosome. Using custom R scripts (available on GitHub), we computed the recombination rate based on our genetic map and the gene density (coding DNA sequence - CDS content) present based on our annotation in 100 KB windows centred on the outlier SNP and 1000 permutations of randomly sampled background SNPs. Finally, we tested for differences in the mean recombination rates and the CDS content within the 100 kb windows using permutation tests.

Genomic differentiation:

To assay the patterns of genomic diversity, differentiation, and divergence we calculated the population genetic summary statistics π (nucleotide diversity), F_{ST} (differentiation), and D_{xy} (divergence) in 1 Mb sliding windows (stepping 10 kb) across the genome using the "all SNPs" VCF file. The summary statistics were computed using the Python scripts available in the genomics_general repository (https://github.com/simonhmartin/genomics_general). Additionally, we computed Tajima's D (Tajima, 1989), in the same sliding windows, across the genome using VCF-kit (https://github.com/Andersenlab/vcf-kit) to summarize the allele frequency spectrum.

Demographic analysis:

We fitted an isolation-with-migration (IM) model (Nielsen & Wakeley, 2001) of population divergence implemented as the maximum likelihood approach in Fastsimcoal2 (Excoffier et al., 2013). We filtered the single SNP VCF file to include only individuals identified as having ancestry from one parental background and created an observed 2-D folded site frequency spectra (SFS) between *D. pulex* and *D. pulicaria* using Easysfs (https://github.com/isaacovercast/easySFS). Subsequently, we estimated the ancestral effective population size (Ne; Nanc), the contemporary Ne for both species (NPa, NPx), the time since divergence (in generations; Tdiv) and migration rates (Mig12, Mig21) using 100 separate runs of fastsimcoal2 (see Supporting Information). In order to estimate these parameters, we fixed the mutation rate to $5.69 \times 10{-}9$, the per generation mutation rate estimated for *D. pulex* (Keith et al., 2016).

Adaptation genomics:

We determined the potential genes that were under selection in the transition from ponds to lakes using the R package PCadapt (Luu et al., 2017; Privé et al., 2020). First, we analysed the single SNP VCF file for outliers using the first two PCs as this sufficiently delineated the *D. pulex* and *D. pulicaria* in our data set. After computing the test statistic, we used Bonferroni adjustment to correct for multiple comparisons. Second, we extracted the location of the outlier SNPs and using BEDtools (Quinlan & Hall, 2010), extracted the genes within 1 kb windows centred on their position. Next, we then performed a gene ontology term (GO-term) overrepresentation test in Panther (Mi et al., 2021), searching against the Daphnia pulex reference list. Finally, we implemented a Fisher's exact test with Bonferroni-corrected p-values to determine enriched GO-terms, with the results visualized using Revigo (Supek et al., 2011).

Results:

Primary genome assemblies:

The Sequell II run produced 942,332 HiFi reads with a mean insert size of 12,513.4 BP (± 3489.3 BP) and generated approximately 60x coverage of the Daphnia pulicaria genome (estimated size 189 MB; Vergilino et al., 2009). Hifiasm produced two separately phased genome assemblies. Haplotype 13B, from the maternal F0 consisted of 185.8 Mb separated across 253 contigs. Haplotype 2X, from the paternal F0 consisted of 179.7 Mb separated across 180 contigs. The assembly statistics for each haplotype are reported in Table 1.

Genetic map, scaffolding and recombination landscapes:

We generated 103 successful Illumina libraries yielding sequences for both F0 genotypes, the F1 and 100 F2 genotypes. These libraries produced mean coverage of approximately 20× across all individuals (i.e., range of 15–50× per lineage). Alignment of the 100 F2, the F1 and the two F0

libraries to the 13B haplotype yielded 34,481,582 called variants consisting of SNPs and small Indels. After quality filtering, 1,088,069 variants remained. The Parentcall2 module of Lepmap3 (Rastas, 2017), further filtered these for segregation distortion to a set of 530,717 variants that were subsequently used to construct the map. After running Lepmap3, we constructed a linkage map consisting of 515,836 markers assigned to 12 linkage groups (LGs), ranging in size from 128,717 to 16,426 markers. See Table 2 for a full list of marker numbers included in each LG. After ordering, we generated and inspected the plot from the Lmplot module and after finding no apparent errors, we proceeded to use our map to construct scaffolds. Most individuals experienced 2–3 crossovers per chromosome ranging from 7 (LG 1) to 0 (all LGs but 1 and 6; Figure S1).

Scaffolding using Lepanchor (Rastas, 2020), generated 12 scaffolds from the 12 LGs designated in the map, nine of which are clearly chromosome-scale. Ultimately, 178.578 Mb (~96.4% of 185.227 Mb) of the primary assembly were placed into these 12 scaffolds which ranged in size from 43.5 Mb (LG 1) to 3.2 Mb (LG 12). See Tables 1 and 2 for additional assembly statistics and map details, respectively. Approximately 6 Mb across 148 contigs of the final assembly remained unplaced with respect to the scaffolds. After running Lepanchor, we reran the Ordermarkers2 module of Lepmap3 to finalize the genetic map. The outputs of which were parsed to develop files to plot Marey Maps (Chakravarti, 1991) to estimate recombination rates for both the male and female specific maps (Figure 1: panels 1–12). Our procedure generated LGs that ranged in size from 203.3 to 39.6 cM and 158.7 to 53.7 cM for male and female maps, respectively (Table 2). It is important to note that sex-specific maps can be compared directly in Daphnia because they do not have sex chromosomes.

The total male map length was longer than the female map length by 82.1 cM, with a total male map length of 1429.3 cM and a total female map length of 1347.2 cM. Estimates of genome-wide recombination rate (GWRR) placed the male-specific recombination at 8.04 cM/Mb and the female-specific recombination rate at 7.57 cM/Mb. GWRR was calculated by taking the ratio of the cumulative map lengths and the total physical length of the genome (Dukić et al., 2016). However, when averaged across all linkage groups, the mean recombination rates for both the male and female lineages were approximately equal at 11.14 cM/Mb and 11.13 cM/Mb, respectively. The overall increase in male map length appears to be driven by higher numbers of crossover events towards the very ends of chromosomes. In two LGs (2 and 12), however, the female map length is longer than the male map length. Visual inspection of the Marey Maps (Figure 1) revealed that all of the chromosome-sized LGs had at least one area of suppressed recombination located either in the centre or towards one end, which consisted of 60%–70% of the total physical length of the LG. However, these regions did not appear to be sex specific. The two largest LGs in terms of physical length (LGs 1 and 6, 43.5 and 24.2 Mb, respectively; Table 2) each had areas of suppressed recombination nearer the ends of the map indicating that suppressed recombination is not restricted to a single centromeric region on these LGs. Each of the smaller subchromosome sized LGs (i.e., LGs 9, 11, and 12; 4.6–3.2 Mb, Table 2) each had consistently high recombination suggesting that the scaffolds may approximate single arms of larger incompletely resolved chromosomes (Figure S2).

To better understand the factors shaping intrachromosomal recombination landscapes, we estimated recombination rates in cM/Mb for 1 Mb windows stepping 10 Kb along Marey maps. In the same windows, we estimated both coding DNA sequence (CDS) content and content of repeats sequences (e.g., transposons, LTR elements, etc.) along chromosomes. Across both the

male and female maps, higher proportions of repeat content were negatively correlated with recombination rate (rm = −0.207, p < 2.2e-16; rf = −0.251, p < 2.2e-16). Meanwhile, across both maps there was a positive correlation between CDS portion and recombination rates (rm = 0.269, p < 2.2e-16; rf = 0.350, p < 2.2e-16). Figure S3A–D shows the patterns in aggregate across both maps.

Genome annotation and completeness:

The D. pulicaria genome hosts a diverse set of elements ranging from DNA transposons to retrotransposons. Tables S2 and S3 show the output of Repeatmasker for each haplotype (13B and 2X). Interestingly, the majority of the repeat content remained unclassified to family (Tables S2 and S3; Figure S4). Surveying the Kimura two-parameter divergence between repeats and the consensus sequence indicated that transposition was active and ongoing, as upwards of 4% of the genome had repeats that were essentially un-diverged from their parental sequences (Figure S4).

The predicted proteome for the 13B haplotype consisted of 23,373 amino acid sequences of which 20,628 could be given functional annotations based on Interproscan (Jones et al., 2014). We computed 22,959 gene models for the 2X haplotype, 20,295 of which were given functional annotations. Total gene sizes, intron length, exon length, and intergene spaces were consistent across both haplotypes (Table S4). Benchmarking universal single-copy orthologues (BUSCO) analysis (Simão et al., 2015) of both haplotypes indicated that each was highly complete. The 13B haplotype resolved 1001 out of 1013 complete arthropod BUSCOs (C: 98.8%), 982 of which were single copy (S: 96.9%) with 19 duplicated, seven fragmented and five missing BUSCOs (Figure S4). The 2X haplotype resolved 999 complete BUSCOs (C: 98.6%), with 979 of these being present as single copies (S: 96.5%). Similar to the 13B haplotype, haplotype 2X had only nine fragmented and five missing BUSCOs (Figure S5).

Speciation genomics:

The "single SNP" data set consisted of 32,949 SNPs from across all 159 samples using the one SNP per RADtag option in Stacks. The "all SNPs" data set consisted of 244,020 SNPs from across all 159 samples which omitted the one-SNP option. We found that the D. pulex assembly and our 13B haplotype are largely colinear (see Supporting Information; Figure S30).

Population structure:

Both PCA and fastSTRUCTURE gave consistent results. For PCA, clusters representing "pure" pulicaria and "pure" pulex individuals were clearly delineated along the first two principal axes (Figure 2a). The structure of the pulex-grouping clearly indicated residual structure that could be expressed on other PCs as well. The hybrids appeared as a single clearly delineated central cluster located between the clusters for each of the parents. The results for fastSTRUCTURE (K = 2), are shown in Figure 2b; these results show that population assignment of each parental grouping was highly supported with the F1 hybrids also clearly delineated with approximately 50/50 posterior assignments. Several individuals that were identified in the data set as "admixed" by Clifford (2016), were also delineated clearly in the bar plot as having a greater proportion of ancestry derived from one parental lineage. The direction of backcrossing appears to favor D. pulex, where these individuals appear to have increased proportions of ancestry shared with this parental lineage (Paland et al., 2005).

Recombination landscape surrounding F_{ST} outliers:

We found that FST outlier loci did not preferentially fall in regions of low recombination. VCFtools computed F_{ST} estimates for 30,475 sites between D. pulex and D. pulicaria; we identified 762 outliers that had estimates of F_{ST} above the 95th percentile (FST > 0.935). Using

permutation tests, we assessed whether outlier SNPs had a greater average recombination rate in 100-kb windows centred on the SNP than 1000 groups of 762 randomly drawn background SNPs. For the male map, we determined that outlier SNPs with a mean recombination rate of 7.52 cM/Mb is not significant at the $p = .05$ level, nevertheless the observed recombination rate of the outlier SNPs exists in the upper tail of the distribution of the background SNPs ($p = .08$; Figure 3a). These results were echoed by the female map, where the mean recombination rate for outliers was 7.53 cM/Mb, here only 91 permutations were larger than the outliers ($p = .091$; Figure 3b). This relationship could be driven by the fact that outlier SNPs appear to be preferentially located in regions of high CDS density (Figure 3c). Calculating the proportion of CDS in the same 100 kb windows centred on the SNPs revealed that outliers had a mean CDS proportion of 0.252, while the 1000 background SNP permutations were surrounded on average by a proportion of CDS of 0.215 ($p \ll .001$). Mean FST was moderately elevated (mean = 0.1165 across the 30,475 sites) and the distribution showed a clear pattern that was consistent with outliers in the data (Figure 3d).

Patterns of genomic differentiation, divergence and diversity:

We calculated F_{ST} (relative divergence, aka differentiation), D_{xy} (absolute divergence), π (diversity) and Tajima's D (summary of allele frequency spectrum) in windows along each chromosome using the “all SNPs” data set. The landscape of all four measures along the genome were heterogenous with respect to comparisons between D. pulex and D. pulicaria. Chromosomes typically had two large (range 1–4) regions of high F_{ST}, overall F_{ST} was low with chromosome-wide averages ranging from ~0.1 to 0.35 (LGs 12 and 10, respectively). Chromosome F_{ST} -islands reached a high of 0.5 with most approaching 0.4. Each F_{ST} island corresponded well to regions of high Dxy and indicated that regions of high F_{ST} were also

regions of reduced gene flow (Cruickshank & Hahn, 2014). Figure 4 shows patterns of recombination, F_{ST}, nucleotide diversity, D_{xy}, and element density on LG 1. Chromosome-wide measures of nucleotide diversity (Table S5) were always larger for *D. pulex* than for *D. pulicaria*, which supports the notion that both lineages differ in effective population size (Ne; Omilian & Lynch, 2009). Regions of high F_{ST} and D_{xy} also generally correspond to troughs in π for *D. pulicaria*, indicating lineage-specific selection on these regions (Figures S6–S28). These signatures of selection in the sliding window data generally correspond to regions of high gene density (measured as proportion of CDS within the window), indicating that selection is restricted to functional regions of the genome. Genome-wide across both *D. pulex* and *D. pulicaria*, Tajima's D was always negative on average (Table S5) and indicated that both lineages were experiencing population expansion but were more negative on average for *D. pulex* (–0.878 vs. –0.632) when compared to *D. pulicaria* (Figure S17–S28). Interestingly, troughs in Tajima's D were not aligned with the other signatures of selection in *D. pulicaria* and, it is unlikely that these values are statistically significant, as D does not generally surpass ±2.5 (Tajima, 1989). These findings suggest that a better accounting of how demography and selection are jointly shaping the allele frequency spectrum is needed.

Demographic history:

We found that the ancestral population of modern-pulex/pulicaria had a small effective population size of only 11,313 haploid genome copies. Both *D. pulicaria* and *D. pulex* have experienced population expansion since their split, but *D. pulicaria* has an effective population size (221,377 haploid genome copies) approximately a third that of *D. pulex* (764,471 haploid genome copies). Divergence appears to have begun approximately 598,022 generations in the past, or approximately 149,505–119,604 years before present (assuming four to five asexual

generations per year (Omilian & Lynch, 2009)). Effective migration rates are low between species, and asymmetrical. *D. pulicaria* receives approximately 9.48 × 10−7 migrants per generation from *D. pulex*; meanwhile in the other direction *D. pulex* receives approximately 3.77 × 10−7 migrants per generation. See Table 3 for parameter estimates and confidence intervals.

PCAdapt outliers and GO-term enrichment:

Using single SNP data, which controls for elevated linkage disequilibrium (LD), we found 555 outlier loci using PCAdpat while retaining the first two PCs and after correction for multiple testing using the Bonferroni method (Figure 5a and S29; (Luu et al., 2017; Privé et al., 2020)). The outliers corresponded well with those regions previously identified with the site-wise FST outliers and the F_{ST} sliding window estimates. Using BEDtools (Quinlan & Hall, 2010), we extracted 401 genes that surrounded our outliers within 1 kb window centred on the outlier SNPs. Using Fisher's exact tests implemented in Panther (Mi et al., 2021), 43 GO-terms appeared enriched for biological processes in our data set after correcting for multiple testing using the Bonferroni method. We found divergence in genes related to light perception, cell cycle, and several other functions (Figure 5b; Table S6) most of which correspond with well-known environmental differences between ponds and lakes.

Discussion:

Facilitated by our chromosome-level genome assembly, genetic map, and annotation of *Daphnia pulicaria,* we investigated the patterns and processes of genomic differentiation present between the ecological sister species *D. pulex and D. pulicaria*. We were able to produce a highly complete phased genome assembly and ultra-dense genetic map showing a dynamic landscape of recombination in the genome of *D. pulicaria*. We applied these resources together with

previously available data and discovered that regions of the genome that exhibited high divergence and differentiation were largely restricted to portions of chromosomes that harbored high density of genes. These regions exhibit reduced diversity in *D. pulicaria*, supporting the interpretation that selection is acting more strongly on this lineage, which is exploiting a novel lake habitat. In addition, tests for enrichment of biological processes for genes in these regions showed that habitat transition from ponds to lakes was highly polygenic and related to genes with functions important for adaptive phenotypes present in lake populations. Overall, we observed that selection is widespread in the *Daphnia pulicaria* genome, probably due to the highly polygenic nature of adaptation to its novel lake habitat, driving increased divergence in regions of high gene density and recombination, falsifying our initial hypothesis.

Recombination is shaped by sequence content and life history:

Daphnia are known to have universally high recombination rates. For example, genetic maps constructed from an F_2 panel for *Daphnia magna* showed universally high recombination near the chromosome ends and suppressed recombination in the chromosome centers (Dukić et al., 2016). These results are mirrored in *D. pulex* for which both population genomic and sperm-based estimates have been produced (Lynch et al., 2020; Xu, Ackerman, et al., 2015). Our results are in general agreement with these previous studies; however, we did find several instances where large parts of linkage group arms exhibited suppressed recombination spanning several megabases (e.g., Figure 1). These differences may be accounted for by our choice of sequencing technology, where both the *D. pulex and D. magna* genomes were assembled primarily using short-read technology, which is known to have limitations when assembling highly repetitive regions of the genome (Jiao & Schneeberger, 2017; Michael & VanBuren, 2020; Ye et al., 2017). Analysis of the sequence content within these regions indicated that they were gene poor and

enriched with repetitive elements (Figure S3). The negative correlation between repeat content and recombination rate is common amongst eukaryotes (Kent et al., 2017). However, the causal relationship remains largely unresolved, as it is unknown whether repeat elements insert preferentially into regions of low recombination or whether their expansion causes reduced recombination. Unfortunately, our analysis does not provide an answer for *Daphnia*. Our linkage mapping approach does not rule out the possibility of mechanisms other than crossing over to be occurring in these regions. Linkage mapping detects crossing over events, while in pericentromeric regions other mechanisms operate (e.g., gene conversion: Shi et al., 2010). Gene conversion is but one of several mechanisms causing concerted evolution of repeat rich regions of the genome (Elder & Turner, 1995). An analysis of the divergence amongst repeat families indicates that transposition is active and ongoing in the *D. pulicaria* genome (Figure S3). Low effective population size (discussed below) is likely to promote the proliferation of deleterious genomic elements (e.g., repetitive elements; Lynch, 2007) and as such, the regions of high repeat density and low recombination may additionally be a consequence of the evolutionary history of D. pulicaria. Future work should include the use of ChIP-seq on centromere specific histone H3 (CENH3), which can resolve the location of centromeric regions (Malik & Henikoff, 2009; Naish et al., 2021). Additionally, using proximity mapping techniques such as HiC will help to identify genomic inversions, which could also be a source of suppressed recombination that has been observed widely across Daphnia species (Jackson et al., 2021).

Another novel aspect of our study is that we provide the first sex-specific estimates of recombination rates for any cyclical parthenogen. Interestingly, we found that recombination map lengths were higher in the male specific map than the female map (1429.3 cM vs. 1347.2 cM), a pattern that is in contrast to what has been generally observed for eukaryotes (Stapley et

al., 2017). The male bias in heterochiasmy is not consistent across all LGs. LG 2 and 12 show longer map lengths in females compared to males (Table 2). Additionally, for LG 9 the male "lags" behind the female map until it jumps towards the end. We believe this is an artefact of the stubby nature of the LG representing only an arm of a larger chromosome. For those LGs where male map length is longer, it is longer by several cM compared to the female map. There is only one other case of an organism with environmental sex determination (crocodile; Isberg et al., 2006; Miles et al., 2009) that exhibits heterochiasmy. In the crocodile, the direction of bias is in the female direction at a ratio of 8:1. Additionally, in the only case of a mixed direction of heterochiasmy in an invertebrate (coral: Acropora millepora), the reversal is restricted to a single chromosome where the male length is only slightly longer (Wang et al., 2009). Thus, our result of a near consistent case of male bias in heterochiasmy is striking. One possible explanation for our result is the unique life cycle of *Daphnia*.

In general, for most of the growing season *Daphnia* populations consist of clonal females producing ameiotic female offspring. Males are ephemeral and are produced when cues trigger their production (Kleiven et al., 1992; Spaak, 1995). The production of males and the subsequent engagement of sexual reproduction in *Daphnia* is linked to diapause (i.e., the production of resting eggs), and as such, occurs as the population experiences deteriorating environmental conditions, amplifying the benefit of sexual recombination (Gerber et al., 2018). Enhanced recombination rates and male-biased direction of heterochiasmy can be driven by such differences in life-history strategies. Interestingly, Neupane and Xu (2020) noted that there are differences in recombination rates between *D. pulex and D. pulicaria* presumably related to their divergent life histories. Because populations of *D. pulicaria* may undergo many generations of clonal reproduction, hidden genetic variance should build up, and can only be released via bouts

of sexual recombination (Lynch & Spitze, 1994). Presumably, the transient nature of males favors them recombining at higher rates because their function in the population is to facilitate the release of the hidden genetic variance in the population via sex. These hypotheses require further testing and would benefit from the development of genetic maps from both sexes in other cyclical parthenogens (e.g., aphids).

Heterogenous genomic divergence is not restricted to regions of low recombination:

Here, we present the first genome-wide survey of genomic divergence in pulex-pulicaria in the light of the expectations of "speciation genomics" (Nosil et al., 2009; Ravinet et al., 2017; Wolf & Ellegren, 2017). Our data show that in contrast to theoretical expectations, the pattern of genomic divergence is largely restricted to genic regions undergoing high recombination (Figure 4a–c). Additionally, the sliding window analysis along whole chromosomes showed that regions of high F_{st} were also regions that exhibited high D_{xy}, an indication that these regions are experiencing lower gene flow than the genome as a whole (Cruickshank & Hahn, 2014). Genetic diversity estimates (π), indicated that these regions were generally less diverse in *D. pulicaria*, thus implicating a role for lineage-specific positive selection. These results are in contrast to recent empirical work in other model systems that investigated ecotype divergence. Both Samuk et al. (2017) and Todesco et al. (2020) found that ecotype divergence was largely restricted to areas of low recombination. In sunflowers, Todesco et al. (2020) found that large structural variants (SVs) restricted geneflow and locked coadapted alleles together. Meanwhile, Samuk et al. (2017) using many population pairs of sticklebacks found strong statistical support for low recombination and selection increasing absolute and relative divergence under geneflow. In our study, we did not survey for SVs, nor did we try to specifically look at sympatric lake/pond populations in depth. However, given our results, future studies should leverage the potential for

long-read sequencing to survey for SVs and explicitly sample pond and lake systems arrayed sympatrically.

Our demographic model supports previous work showing that *D. pulicaria and D. pulex* have recently diverged and share low but non-negligible levels of gene flow. Omilian and Lynch (2009) determined that the two species diverged approximately 150–80 kya, have limited gene flow and that D. pulicaria has significantly reduced effective population size. Using genome-wide data, we have shown that the estimate for divergence should be much closer to 150 kya. Our estimates indicated that the mid-Pleistocene Illinoian glaciation, which is estimated to have covered North America from 191 to 130 kya (Richmond & Fullerton, 1986), may have been a contributing factor in the divergence seen in this species complex. What remains unclear is the exact nature of the demographic event that contributed to divergence between these two species. Some authors have argued that divergence between pulex and pulicaria represents rapid and ongoing ecological speciation stemming from *D. pulicaria* invading lake habitats and subsequent dispersal back to ponds (Pfrender et al., 2000). Nevertheless, prior work has not explicitly modeled demographic scenarios. Additionally, there is the potential for extensive reticulation and incomplete lineage sorting to confound signals throughout the genome (Vergilino et al., 2011).

The majority of studies reconstructing the diversification of the larger *D. pulex*-complex have focused primarily on mitochondrial markers (Colbourne et al., 1998; Weider et al., 1999). However, there is evidence that mitochondrial markers may not fully capture diversification in this species. Specifically, Marková et al. (2013) found evidence of mitonuclear discordance in the species-complex, where European isolates of *D. pulicaria* have distinct mitochondrial genomes but are more closely-related at nuclear markers to North American *D. pulicaria*. These

findings, taken with the presence of genetic variants originating from North *American D. pulicaria* that suppress meiosis when introgressed into the North American *D. pulex* background (Tucker et al., 2013; Xu, Spitze, et al., 2015; Ye et al., 2019), suggests a more complex demographic scenario for *D. pulex and D. pulicaria*.

Although we did not test different scenarios explicitly, models of divergence that include the possibility of secondary contact between the two lineages need to be tested. Clearly, such a shallow time divergence and sustained gene flow is likely to be important in shaping the patterns of divergence observed between these two species. Future demographic modelling should include isolates from across the Holarctic to illuminate potential demographic scenarios and resolve the taxonomic status of a number of lineages within the *Daphnia pulex* complex (Ye et al., 2021). Consideration of the recombination landscape present in diverse isolates, including admixed individuals, will also help in determining the source of the curious pattern of genomic divergence and recombination rate noted here.

Divergent regions host genes related to adaptation to a novel habitatI

Implementing a more conservative genome scan method than estimating F_{ST} either site-wise or in sliding windows using PCAdapt, we identified 555 outliers after Bonferroni correction, 207 fewer SNPs than our site-wise FST scan. Regardless, both methods detected outliers in the same regions and corresponded well with the sliding window peaks. Unique genes within 1 kb of outliers (401 in total) appeared to be enriched for 43 GO terms, including several of which that stand out as being plausibly related to key aspects of the habitat transition from ponds to lakes. One of the groupings showed enrichment of genes involved in light detection and phototransduction. One of the key differences between lake and pond *Daphnia* is that lake populations undergo extensive diel vertical migration (DVM – Lampert, 2011). DVM is a critical

Daphnia antipredator defense and shapes many aspects of their physiology and life history (Stich & Lampert, 1981). Given that DVM behavior involves moving deeper in the water column during daylight, the response to light is probably important in signaling this behavioral response. Additionally, *D. pulex and D. pulicaria* exhibit different life-history responses to photoperiod (Wen Deng, 1997), which has been hypothesized to contribute to the maintenance of species boundaries as a pre-reproductive isolating barrier.

Lakes and ponds also often differ in their predation regimes where many lakes that harbor *D. pulicaria* populations also contain visually oriented fish predators (Brooks & Dodson, 1965; Tessier & Welser, 1991). Meanwhile, many ponds that harbor *D. pulex* are fish-less but contain high population densities of gape-limited invertebrate predators such as phantom midge larvae (Chaoborus; Spitze, 1991). These different predator guilds have created contrasting selection pressures on growth rate and neonate body size in the two species where pulex tend to grow and "live" faster than pulicaria (Dudycha, 2004; Dudycha & Tessier, 1999) in order to more rapidly reach a body-size "refuge". Interestingly, we observed GO term enrichment for several different aspects of the cell cycle and mitosis which we believe may underlie the differences in growth and senescence observed between these two species. Finally, another set of highly enriched GO terms were related to various aspects of metabolism. This could be related to widespread metabolic rewiring related to various aspects of life-history as noted above, and the vast physicochemical and biological differences between ponds and lakes (Søndergaard et al., 2005). Several other enriched GO terms (e.g., vesicle-mediated transport processes) have less obvious connections to the ecology of these two sister species. Clearly, more experiments integrating functional genomics and life-history assays are needed to further investigate these apparent signatures of selection.

Conclusions:

We argue that the pattern and process of divergence is dynamic and shaped primarily by the underlying distribution of functional genomic elements rather than low recombination. In particular, lineage-specific selection in *D. pulicaria* prompted by invasion into a novel habitat appears to be shaping the genome-wide patterns of divergence in these sister ecological species. Selection and gene flow ultimately are opposing forces in speciation-with-geneflow, and their complex interplay is shaped by the genetic architecture that underlies adaptive phenotypes (Kautt et al., 2020; Martin et al., 2013). Several lines of evidence support our claims. First, there is broad overlap of our regions of divergence with regions in our genetic map showing high recombination and gene density. While the overlaps are striking, a careful examination of all chromosomes indicates several exceptions (e.g., Figure 4; 25–35 Mb). Why these occur is unclear, but they warrant future investigation. Next, we also observed an association between F_{ST} outliers and recombination rates, however, our permutation test falls short of statistical significance (at $p = .05$). However, we must also consider the biological significance of a pattern, and this relationship is nonetheless striking. Especially, when considered in conjunction with the fact that the outlier regions are surrounded by more genes on average than background SNPs.

We also contend that the nature of the habitat transition experienced by *D. pulicaria* must have been highly polygenic and this contributed to the patterns observed. We support this with model-based outlier analysis (PCadapt) and GO-term enrichment analysis. We acknowledge that GO enrichment tests are fraught with difficulties in developing appropriate null distributions for testing. Regardless, such tests are important tools for interrogating biological significance of large gene lists (Dessimoz & Skunca, 2017). While we invoke an adaptive explanation for the large number of outliers, demographic history cannot be ignored. Our demographic modeling

shows that *D. pulicaria* has low effective population size and may have experienced a bottleneck. Thus, many of these outliers may be due to drift. While this is a plausible explanation, we believe that life-cycle patterns in lake *D. pulicaria*, which often include increased clonality and delayed sex (Cáceres & Tessier, 2004; Tessier & Cáceres, 2004), can influence population genetic structure in lake populations by promoting clonal selection (Allen & Lynch, 2012). Further, given the current understanding of metapopulation dynamics for *Daphnia* (De Meester et al., 2006) and the nature of monopolization within lakes (De Meester et al., 2002) it is not surprising that strong clonal selection is at play despite low effective size. Additionally, new and independent evidence (Jackson et al., 2021) which used a novel *D. pulicaria* genome assembly and selection tests confirms a model where widespread polygenic selection is operating in this lake species.

Our findings contribute to a growing body of work that highlights *Daphnia* as an important ecogenomic model that will help revolutionize many aspects of population genomic studies centered on incipient species formation. However, further studies specifically addressing the roles of recombination versus gene density in determining the pattern off heterogenous genomic divergence in systems exhibiting ecological-speciation-with-gene-flow are needed. The sequencing of the first *Daphnia* genome (Colbourne et al., 2011) catalysed exciting discoveries in disparate areas of biology (e.g., Miner et al., 2012; Seda & Petrusek, 2011; Tkaczyk et al., 2021). This new high-quality near-chromosome scale phased genome assembly and development of an ultra-dense genetic map should further catalyse research illuminating the dynamic interplay between ecology and evolution across biological scales to understand processes ranging from genes to ecosystems.

Acknowledgements:

The work presented here would not have been possible without the tireless efforts of several generations of Weider laboratory students and post docs, including D. Frisch, P. Morton, B. Culver, R. Hartnett, and E. Kiehnau. Funding for this project was provided by The University of Oklahoma Board of Regents to LJW, and the U.S. National Science Foundation (IOS#0924289 and IOS#1256881 to LJW and IOS#0924401 and IOS#1256867 to PDJ) 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 U.S. National Science Foundation or the University of Oklahoma. Computing resources used in this research were provided by the University of Oklahoma's Oklahoma Supercomputing Center for Education and Research (OU OSCER). This manuscript represents a portion of MJW's doctoral dissertation at The University of Oklahoma. Finally, we dedicate this manuscript to the memory of Professor Dr Winfried Lampert (former Director of the Max-Planck-Institute for Limnology, Plön, Germany), who dedicated his career to developing *Daphnia* as an important international model species in the fields of ecophysiology, ecology, and evolution.

Author Contributions:

PDJ and LJW performed fieldwork and designed and implemented the crossing experiments and F2 panel breeding protocol. RES conducted the Illumina sequencing experiments, which MJW analysed and combined with the long-read PacBio genome sequencing, assembly, and annotation. MJW analysed all data sets and wrote the first draft of the manuscript. All authors contributed to interpretation, editing, and revising subsequent versions of the manuscript.

Data Availability Statement:

The ddRADseq data set used in this study is archived in the NCBI SRA under study Accession PRJNA288877. The whole genome sequence data from the F0s and F1individuals will be deposited in the NCBI SRA under accessions SAMN07108733, SAMN07108734, SAMN07108731 & SAMN21387128. The sequence data for the F2s will be available from SRA under SAMN25518473-SAMN25518575 after a 6-month embargo period. The genome assembly is available on NCBI Genbank under accession SAMN21387128. Metadata for this study is available on the Dryad depository (Wersebe et al., 2021). Scripts used to produce the analyses presented here are available on Github at: www.github.com/mwersebe/Daphnia_pulicaria_F2_genomics. This research is not subject to the Convention on Biological Diversity nor the Nagoya Protocol.

Works Cited:

Allen, D. E., & Lynch, M. (2012). The effect of variable frequency of sexual reproduction on the genetic structure of natural populations of a cyclical parthenogen. Evolution, 66(3), 919–926. https://doi.org/10.1111/j.1558-5646.2011.01488.x

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., Lippman, Z. B., & Schatz, M. C. (2019). RaGOO: Fast and accurate reference-guided scaffolding of draft genomes. Genome Biology, 20, 224. https://doi.org/10.1186/s13059-019-1829-6

Athanasio, C. G., Chipman, J. K., Viant, M. R., & Mirbahai, L. (2016). Optimisation of DNA extraction from the crustacean *Daphnia*. PeerJ, 4, e2004. https://doi.org/10.7717/peerj.2004

Bao, Z., & Eddy, S. R. (2002). Automated de novo identification of repeat sequence families in sequenced genomes. Genome Research, 12,1269–1276. https://doi.org/10.1101/gr.88502

Brooks, J. L., & Dodson, S. I. (1965). Predation, body size, and composition of plankton. Science, 150, 28–35. https://doi.org/10.1126/science.150.3692.28

Cáceres, C. E., & Tessier, A. J. (2004). Incidence of diapause varies among populations of *Daphnia pulicaria*. Oecologia, 141(3), 425–431. https://doi.org/10.1007/s00442-004-16575

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. BMC Bioinformatics, 10, 421. https://doi.org/10.1186/1471-2105-10-421

Campbell, C. R., Poelstra, J. W., & Yoder, A. D. (2018). What is speciation genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. Biological Journal of the Linnean Society, 124, 561–583. https://doi.org/10.1093/biolinnean/bly063

Carneiro, M., Ferrand, N., & Nachman, M. W. (2009). Recombination and speciation: Loci near centromeres are more differentiated than loci near telomeres between subspecies of the European rabbit (*Oryctolagus cuniculus*). Genetics, 181, 593–606. https://doi.org/10.1534/genetics.108.096826

Carpenter, S. R., Kitchell, J. F., Hodgson, J. R., Cochran, P. A., Elser, J. J., Elser, M. M., Lodge, D. M., Kretchmer, D., He, X., & von Ende, C. N. (1987). Regulation of lake primary productivity by food web structure. Ecology, 68, 1863–1876. https://doi.org/10.2307/1939878

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. Molecular Ecology, 22, 3124–3140. https://doi.org/10.1111/mec.12354

Chakravarti, A. (1991). A graphical representation of genetic and physical maps: The Marey map. Genomics, 11, 219–222. https://doi.org/10.1016/0888-7543(91)90123-V

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nature Methods, 18, 170–175. https://doi.org/10.1038/s41592-020-01056-5

Chin, T. A., Cáceres, C. E., & Cristescu, M. E. (2019). The evolution of reproductive isolation in *Daphnia*. BMC Evolutionary Biology, 19, 1–15. https://doi.org/10.1186/s12862-019-1542-9

Chin, T. A., & Cristescu, M. E. (2021). Speciation in *Daphnia*. Molecular Ecology, 30, 1398–1418. https://doi.org/10.1111/mec.15824

Clifford, B. J. (2016). Genomic and regulatory divergence between sister species of *Daphnia*. Doctoral Dissertation- University of Notre Dame. Retrieved from https://curate.nd.edu/show/t722h70577s

Colbourne, J. K., Crease, T. J., Weider, L. J., Hebert, P. D. N., Duferesne, F., & Hobaek, A. (1998). Phylogenetics and evolution of a circumarctic species complex (Cladocera: *Daphnia pulex*). Biological Journal of the Linnean Society, 65, 347–365. https://doi.org/10.1111/j.1095-8312.1998.tb01146.x

Colbourne, J. K., Pfrender, M. E., Gilbert, D., Thomas, W. K., Tucker, A., Oakley, T. H., Tokishita, S., Aerts, A., Arnold, G. J., Basu, M. K., Bauer, D. J., Cáceres, C. E., Carmel, L., Casola, C., Choi, J.-H., Detter, J. C., Dong, Q., Dusheyko, S., Eads, B. D., … Boore, J. L. (2011). The ecoresponsive genome of *Daphnia pulex*. Science (80), 331, 555–561. https://doi.org/10.1126/science.1197761

Crease, T. J. (1999). The complete sequence of the mitochondrial genome of *Daphnia pulex* (Cladocera: Crustacea). Gene, 233, 89–99. https://doi.org/10.1016/S0378-1119(99)00151-1

Crease, T. J., Floyd, R., Cristescu, M. E., & Innes, D. (2011). Evolutionary factors affecting Lactate dehydrogenase A and B variation in the *Daphnia pulex* species complex. BMC Evolutionary Biology, 11, 212. https://doi.org/10.1186/1471-2148-11-212

Crease, T. J., Omilian, A. R., Costanzo, K. S., & Taylor, D. J. (2012). Transcontinental phylogeography of the *Daphnia pulex* species complex K.A. Crandall [ed.]. PLoS One, 7, e46620. https://doi.org/10.1371/journal.pone.0046620

Cristescu, M. E., Constantin, A., Bock, D. G., Cáceres, C. E., & Crease, T. J. (2012). Speciation with gene flow and the genetics of habitat transitions. Molecular Ecology, 21, 1411–1422. https://doi.org/10.1111/j.1365-294X.2011.05465.x

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology, 23, 3133–3157. https://doi.org/10.1111/mec.12796

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. Bioinformatics, 27, 2156–2158. https://doi.org/10.1093/bioinforma tics/btr330

De Meester, L., Gómez, A., Okamura, B., & Schwenk, K. (2002). The Monopolization Hypothesis and the dispersal-gene flow paradox in aquatic organisms. Acta Oecologica, 23(3), 121–135. https://doi.org/10.1016/S1146-609X(02)01145-1

De Meester, L., Vanoverbeke, J., De Gelas, K., Ortells, R., & Spaak, P. (2006). Genetic structure of cyclic parthenogenetic zooplankton populations-a conceptual framework. Archiv Fur Hydrobiologie, 167, 1–4. https://doi.org/10.1127/0003-9136/2006/0167-0217

Dessimoz, C., & Škunca, N. (2017). The gene ontology handbook. Springer Nature.

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin, 19(1), 11–15.

Dudycha, J. L. (2004). Mortality dynamics of *Daphnia* in contrasting habitats and their role in ecological divergence. Freshwater Biology, 49, 505–514. https://doi.org/10.1111/j.1365-2427.2004.01201.x

Dudycha, J. L., & Tessier, A. J. (1999). Natural genetic variation of life span, reproduction, and juvenile growth in *Daphnia*. Evolution (N.Y.), 53, 1744–1756. https://doi.org/10.1111/j.1558-5646.1999.tb04559.x

Dufresne, F., & Hebert, P. D. N. (1997). Pleistocene glaciations and polyphyletic origins of polyploidy in an arctic cladoceran. Proceedings of the Royal Society B-Biological Sciences, 264, 201–206. https://doi.org/10.1098/rspb.1997.0028

Dukić, M., Berner, D., Roesti, M., Haag, C. R., & Ebert, D. (2016). A highdensity genetic map reveals variation in recombination rate across the genome of *Daphnia magna*. BMC Genetics, 17, 137. https://doi.org/10.1186/s12863-016-0445-7

Elder, J. F., & Turner, B. J. (1995). Concerted Evolution of repetitive DNA sequences in eukaryotes. The Quarterly Review of Biology, 70(3), 297–320. https://doi.org/10.1086/419073

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. PLoS Genetics, 9, 1003905. https://doi.org/10.1371/journal.pgen.1003905

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation with-gene-flow. Trends in Genetics, 28, 342–350. https://doi.org/10.1016/j.tig.2012.03.009

Feder, J. L., & Nosil, P. (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. Evolution (N. Y.), 64, 1729–1747. https://doi.org/10.1111/j.1558-5646.2009.00943.x

Felsenstein, J. (1981). Skepticism towards Santa Rosalia, or why are there so few kinds of animals? Evolution (N. Y), 35, 124–138. https://doi.org/10.2307/2407946

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences of the United States of America, 117, 9451–9457. https://doi.org/10.1073/pnas.1921046117

Frisch, D., Morton, P. K., Chowdhury, P. R., Culver, B. W., Colbourne, J. K., Weider, L. J., & Jeyasingh, P. D. (2014). A millennial-scale chronicle of evolutionary responses to cultural eutrophication in *Daphnia*. D. Post [ed.]. Ecology Letters, 17, 360–368. https://doi.org/10.1111/ele.12237

Frisch, D., Morton, P. K., Culver, B. W., Edlund, M. B., Jeyasingh, P. D., & Weider, L. J. (2017). Paleogenetic records of *Daphnia pulicaria* in two North American lakes reveal the impact of cultural eutrophication. Global Change Biology, 23, 708–718. https://doi.org/10.1111/gcb.13445

Fu, H., Park, W., Yan, X., Zheng, Z., Shen, B., & Dooner, H. K. (2001). The highly recombinogenic bz locus lies in an unusually gene-rich region of the maize genome. Proceedings of the National Academy of Sciences of the United States of America, 98, 8903–8908. https://doi.org/10.1073/pnas.141221898

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907. Retrieved from http://arxiv.org/abs/1207.3907

Gerber, N., Kokko, H., Ebert, D., & Booksmythe, I. (2018). *Daphnia* invest in sexual reproduction when its relative costs are reduced. Proceedings of the Royal Society B-Biological Sciences, 285, 20172176. https://doi.org/10.1098/rspb.2017.2176

Gremme, G., Brendel, V., Sparks, M. E., & Kurtz, S. (2005). Engineering a software tool for gene structure prediction in higher organisms. Information and Software Technology, 47, 965–978. https://doi.org/10.1016/j.infsof.2005.09.005

Heier, C. R., & Dudycha, J. L. (2009). Ecological speciation in a cyclic parthenogen: Sexual capability of experimental hybrids between *Daphnia pulex* and *Daphnia pulicaria*. Limnology and Oceanography, 54, 492–502. https://doi.org/10.4319/lo.2009.54.2.0492

Hejase, H. A., Salman-Minkov, A., Campagna, L., Hubisz, M. J., Lovette, I. J., Gronau, I., & Siepel, A. (2020). Genomic islands of differentiation in a rapid avian radiation have been driven by recent selective sweeps. Proceedings of the National Academy of Sciences of the United States of America, 117, 30554–30565. https://doi.org/10.1073/pnas.2015987117

Hoff, K. J., & Stanke, M. (2018). Predicting genes in single genomes with AUGUSTUS. Current Protocols in Bioinformatics, 65, e57. https://doi.org/10.1002/cpbi.57

Isberg, S. R., Johnston, S. M., Chen, Y., & Moran, C. (2006). First evidence of higher female recombination in a species with temperaturedependent sex determination: The saltwater crocodile. Journal of Heredity, 97(6), 599–602. https://doi.org/10.1093/JHERED/ESL035

Jackson, C. E., Xu, S., Ye, Z., Pfrender, M. E., Lynch, M., Colbourne, J. K., & Shaw, J. R. (2021). Chromosomal rearrangements preserve adaptive divergence in ecological speciation. BioRxiv, 2021.08.20.457158. https://doi.org/10.1101/2021.08.20.457158

Jiao, W. B., & Schneeberger, K. (2017). The impact of third generation genomic technologies on plant genome assembly. Current Opinion in Plant Biology, 36, 64–70. https://doi.org/10.1016/j.pbi.2017.02.002

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics, 24, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein

function classification. Bioinformatics, 30, 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kautt, A. F., Kratochwil, C. F., Nater, A., Machado-Schiaffino, G., Olave, M., Henning, F., Torres-Dowdall, J., Härer, A., Hulsey, C. D., Franchini, P., Pippel, M., Myers, E. W., & Meyer, A. (2020). Contrasting signatures of genomic divergence during sympatric speciation. Nature, 588, 106–111. https://doi.org/10.1038/s41586-020-2845-0

Keith, N., Tucker, A. E., Jackson, C. E., Sung, W., Lucas Lledó, J. I., Schrider, D. R., Schaack, S., Dudycha, J. L., Ackerman, M., Younge, A. J., Shaw,J. R., & Lynch, M. (2016). High mutational rates of large-scale duplication and deletion in *Daphnia pulex*. Genome Research, 26, 60–69. https://doi.org/10.1101/gr.191338.115

Kent, T. V., Uzunović, J., & Wright, S. I. (2017). Coevolution between transposable elements and recombination. Philosophical Transactions of the Royal Society B: Biological Sciences, 372, 20160458. https://doi.org/10.1098/rstb.2016.0458

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution, 16(2), 111–120. https://doi.org/10.1007/BF01731581

Kleiven, O. T., Larsson, P., Hobæk, A., & Hobaek, A. (1992). Sexual reproduction in *Daphnia* magna requires three stimuli. Oikos, 65, 197–206. https://doi.org/10.2307/3545010

Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., & Zdobnov, E. M. (2019). OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Research, 47, D807–D811. https://doi.org/10.1093/nar/gky1053

Lampert, W. (2011). *Daphnia*: Development of a model organism in ecology and evolution. International Ecology Institute (ECI).

Lampert, W., & Sommer, U. (2007). Limnoecology: The ecology of lakes and streams, 2nd ed. Oxford University Press.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 9, 357–359. https://doi.org/10.1038/nmeth.1923

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence Alignment/Map format and SAMtools. Bioinformatics, 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: An R package to perform genome scans for selection based on principal component analysis. Molecular Ecology Resources, 17, 67–77.

Lynch, M. (2007). The origins of genome architecture. Sinauer Associates.

Lynch, M., & Spitze, K. (1994). Evolutionary genetics of Daphnia. In L.Real (Eds.), Ecological Genetics (pp. 109–128). Princeton University Press.

Lynch, M., Ye, Z., & Maruki, T. (2020). The Recombinational Landscape in *Daphnia pulex*. bioRxiv 2020.03.03.974485. https://doi.org/10.1101/2020.03.03.974485

Malik, H. S., & Henikoff, S. (2009). Major evolutionary transitions in centromere complexity. Cell, 138(6), 1067–1082. https://doi.org/10.1016/J.CELL.2009.08.036

Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., & Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system A.E. Darling [ed.]. PLOS Computational Biology, 14,e1005944. https://doi.org/10.1371/journal.pcbi.1005944

Marková, S., Dufresne, F., Manca, M., & Kotlík, P. (2013). Mitochondrial capture misleads about ecological speciation in the Daphnia pulex complex D. Fontaneto [ed.]. PLoS One, 8, e69497. https://doi.org/10.1371/journal.pone.0069497

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., & Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Research, 23, 1817–1828. https://doi.org/10.1101/gr.159426.113

Mergeay, J., Verschuren, D., & De Meester, L. (2006). Invasion of an asexual American water flea clone throughout Africa and rapid displacement of a native sibling species. Proceedings of the Royal Society B-Biological Sciences, 273, 2839–2844. https://doi.org/10.1098/ rspb.2006.3661

Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L. P., Mushayamaha, T., & Thomas, P. D. (2021). PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Research, 49, D394–D403. https://doi.org/10.1093/nar/gkaa1106

Michael, T. P., & VanBuren, R. (2020). Building near-complete plant genomes. Current Opinion in Plant Biology, 54, 26–33. https://doi.org/10.1016/j.pbi.2019.12.009

Miles, L. G., Isberg, S. R., Glenn, T. C., Lance, S. L., Dalzell, P., Thomson, P. C., & Moran, C. (2009). A genetic linkage map for the saltwater crocodile (Crocodylus porosus). BMC Genomics, 10(1), 1–11. https://doi.org/10.1186/1471-2164-10-339

Miner, B. E., de Meester, L., Pfrender, M. E., Lampert, W., & Hairston, N. G. (2012). Linking genes to communities and ecosystems: *Daphnia* as an ecogenomic model. Proceedings of the Royal Society B-Biological Sciences, 279, 1873–1882. https://doi.org/10.1098/rspb.2011.2404

Nachman, M. W., & Payseur, B. A. (2012). Recombination rate variation and speciation: Theoretical predictions and empirical results from rabbits and mice. Philosophical Transactions of the Royal SocietyB: Biological Sciences, 367, 409–421. https://doi.org/10.1098/rstb.2011.0249

Naish, M., Alonge, M., Wlodzimierz, P., Tock, A. J., Abramson, B. W., Schmücker, A., & Henderson, I. R. (2021). The genetic and epigenetic landscape of the *Arabidopsis* centromeres. Science, 374(6569), eabi7489.

Neupane, S., & Xu, S. (2020). Adaptive divergence of meiotic recombination rate in ecological speciation C. Baer [ed.]. Genome Biology and Evolution, 12, 1869–1881. https://doi.org/10.1093/gbe/evaa182

Nielsen, R., & Wakeley, J. (2001). Distinguishing migration from isolation: A Markov chain Monte Carlo approach. Genetics, 158(2), 885–896.

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. Molecular Ecology, 18, 375–402. https://doi.org/10.1111/j.1365-294X.2008.03946.x

Omilian, A. R., & Lynch, M. (2009). Patterns of intraspecific DNA variation in the *Daphnia* nuclear genome. Genetics, 182, 325–336. https://doi.org/10.1534/genetics.108.099549

Orsini, L., Andrew, R., & Eizaguirre, C. (2013). Evolutionary ecological genomics. Molecular Ecology, 22, 527–531. https://doi.org/10.1111/mec.12177

Paland, S., Colbourne, J. K., & Lynch, M. (2005). Evolutionary history of contagious asexuality in *Daphnia pulex*. Evolution (N. Y.), 59, 800–813. https://doi.org/10.1111/j.0014-3820.2005.tb01754.x

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for De Novo SNP discovery and genotyping in model and non-model species. PLoS One, 7(5), e37135. https://doi.org/10.1371/journal.pone.003713

Pfrender, M. E., Spitze, K., & Lehman, N. (2000). Multi-locus genetic evidence for rapid ecologically based speciation in *Daphnia*. Molecular Ecology, 9, 1717–1735. https://doi.org/10.1046/j.1365-294X.2000.01062.x

Price, A. L., Jones, N. C., & Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. Bioinformatics, 21, i351–i358. https://doi.org/10.1093/bioinformatics/bti1018

Privé, F., Luu, K., Vilhjálmsson, B. J., Blum, M. G. B., & Rosenberg, M. (2020). Performing highly efficient genome scans for local adaptation with R package pcadapt version 4. Molecular Biology and Evolution, 37, 2153–2154. https://doi.org/10.1093/molbev/msaa053

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics, 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Raj, A., Stephens, M., & Pritchard, J. K. (2014). FastSTRUCTURE: Variational inference of population structure in large SNP data sets. Genetics, 197, 573–589. https://doi.org/10.1534/genetics.114.164350

Rastas, P. (2017). Lep-MAP3: Robust linkage mapping even for low coverage whole genome sequencing data. Bioinformatics, 33, 3726–3732. https://doi.org/10.1093/bioinformatics/btx494

Rastas, P. (2020). Lep-Anchor: Automated construction of linkage map anchored haploid genomes. Bioinformatics, 36, 2359–2364. https://doi.org/10.1093/bioinformatics/btz978

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. A. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. Journal of Evolutionary Biology, 30, 1450–1477. https://doi.org/10.1111/jeb.13047

Richmond, G. M., & Fullerton, D. S. (1986). Summation of quaternary glaciations in the United States of America. Quaternary Science Reviews, 5, 183–196. https://doi.org/10.1016/0277-3791(86)90184-8

Rundle, H. D., & Nosil, P. (2005). Ecological speciation. Ecology Letters, 8, 336–352. https://doi.org/10.1111/j.1461-0248.2004.00715.x

Samuk, K., Owens, G. L., Delmore, K. E., Miller, S. E., Rennison, D. J., & Schluter, D. (2017). Gene flow and selection interact to promote adaptive divergence in regions of low recombination. Molecular Ecology, 26(17), 4378–4390. https://doi.org/10.1111/MEC.14226

Schluter, D. (2009). Evidence for ecological speciation and its alternative. Science (80-), 323, 737–741. https://doi.org/10.1126/science.1160006

Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemovalv2: Rapid adapter trimming, identification, and read merging. BMC Research Notes, 9, 88. https://doi.org/10.1186/s13104-016-1900-2

Seda, J., & Petrusek, A. (2011). Daphnia as a model organism in limnology and aquatic biology: Introductory remarks. Journal of Limnology, 70, 337–344. https://doi.org/10.4081/JLIMNOL.2011.337

Sherman, R. E., Hartnett, R., Kiehnau, E. L., Weider, L. J., & Jeyasingh, P. D. (2021). Quantitative genetics of phosphorus content in the freshwater herbivore, Daphnia pulicaria S. Plaistow [ed.]. Journal of Animal Ecology, 90, 909–916. https://doi.org/10.1111/1365-2656.13419

Shi, J., Wolf, S. E., Burke, J. M., Presting, G. G., Ross-Ibarra, J., & Dawe, R. K. (2010). Widespread gene conversion in centromere cores. PLOS Biology, 8(3), e1000327. https://doi.org/10.1371/journal.pbio.1000327

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, 31, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Søndergaard, M., Jeppesen, E., Jensen, J., & Peder. (2005). Pond or lake: Does it make any difference? Archiv Für Hydrobiologie, 162(2), 143–165. https://doi.org/10.1127/0003-9136/2005/0162-0143

Spaak, P. (1995). Sexual reproduction in *Daphnia*: Interspecific differences in a hybrid species complex. Oecologia, 104, 501–507. https://doi.org/10.1007/BF00341348

Spitze, K. (1991). Chaoborus predation and life-history evolution in *Daphnia pulex*: Temporal pattern of population diversity, fitness, and mean life history. Evolution (N. Y.), 45, 82–92. https://doi.org/10.1111/j.1558-5646.1991.tb05268.x

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: Ab initio prediction of alternative transcripts. Nucleic Acids Research, 34, W435–W439. https://doi.org/10.1093/nar/gkl200

Stapley, J., Feulner, P. G. D., Johnston, S. E., Santure, A. W., & Smadja, C. M. (2017). Variation in recombination frequency and distribution across eukaryotes: Patterns and processes. Philosophical Transactions of the Royal Society B: Biological Sciences, 372, 20160455. https://doi.org/10.1098/rstb.2016.0455

Stich, H.-B., & Lampert, W. (1981). Predator evasion as an explanation of diurnal vertical migration by zooplankton. Nature, 293, 396–398. https://doi.org/10.1038/293396a0

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms C. Gibas [ed.]. PLoS One, 6, e21800. https://doi.org/10.1371/journal.pone.0021800

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics, 12(3), 585–595. https://doi.org/10.1093/genetics/123.3.585

Tessier, A. J., & Cáceres, C. E. (2004). Differentiation in sex investment by clones and populations of *Daphnia*. Ecology Letters, 7(8), 695–703. https://doi.org/10.1111/j.1461-0248.2004.00627.x

Tessier, A. J., & Welser, J. (1991). Cladoceran assemblages, seasonal succession and the importance of a hypolimnetic refuge. Freshwater Biology, 25, 85–93. https://doi.org/10.1111/j.1365-2427.1991.tb00475.x

Tkaczyk, A., Bownik, A., Dudka, J., Kowal, K., & Ślaska, B. (2021). *Daphnia magna* model in the toxicity assessment of pharmaceuticals: A review. Science of the Total Environment, 763, 143038. https://doi.org/10.1016/J.SCITOTENV.2020.143038

Todesco, M., Owens, G. L., Bercovich, N., Légaré, J.-S., Soudi, S., Burge, D. O., Huang, K., Ostevik, K. L., Drummond, E. B. M., Imerovski, I., Lande, K., Pascual-Robles, M. A., Nanavati, M., Jahani, M., Cheung, W., Staton, S. E., Muños, S., Nielsen, R., Donovan, L. A., … Rieseberg, L. H. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. Nature, 584(7822), 602–607. https://doi.org/10.1038/s41586-020-2467-6

Tucker, A. E., Ackerman, M. S., Eads, B. D., Xu, S., & Lynch, M. (2013). Population-genomic insights into the evolutionary origin and fate of obligately asexual *Daphnia pulex*. Proceedings of the National Academy of Sciences of the United States of America, 110, 15740–15745. https://doi.org/10.1073/pnas.1313388110

Turelli, M., Barton, N. H., & Coyne, J. A. (2001). Theory and speciation. Trends in Ecology & Evolution, 16, 330–343. https://doi.org/10.1016/S0169-5347(01)02177-2

Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae*. PLoS Biology, 3, 1572–1578. https://doi.org/10.1371/journal.pbio.0030285

Ungerer, M. C., Johnson, L. C., & Herman, M. A. (2008). Ecological genomics: understanding gene and genome function in the natural environment. Heredity, 100, 178–183. https://doi.org/10.1038/sj.hdy.6800992

Vergilino, R., Belzile, C., & Dufresne, F. (2009). Genome size evolution andpolyploidy in the *Daphnia pulex* complex (Cladocera: Daphniidae). Biological Journal of the Linnean Society, 97(1), 68–79. https://doi.org/10.1111/j.1095-8312.2008.01185.x

Vergilino, R., Markova, S., Ventura, M., Manca, M., & Dufresne, F. (2011). Reticulate evolution of the Daphnia pulex complex as revealed by nuclear markers. Molecular Ecology, 20, 1191–1207. https://doi.org/10.1111/j.1365-294X.2011.05004.x

Via, S. (2009). Natural selection in action during speciation. Proceedings of the National Academy of Sciences of the United States of America, 106(Suppl 1), 9939–9946. https://doi.org/10.1073/pnas.0901397106

Walsh, J. R., Carpenter, S. R., & Vander Zanden, M. J. (2016). Invasive species triggers a massive loss of ecosystem services through a trophic cascade. Proceedings of the National Academy of Sciences of the United States of America, 113, 4081–4085. https://doi.org/10.1073/pnas.1600366113

Wang, S., Zhang, L., Meyer, E., & Matz, M. V. (2009). Construction of a high-resolution genetic linkage map and comparative genome analysis for the reef-building coral *Acropora millepora*. Genome Biology, 10(11), 1–17. https://doi.org/10.1186/GB-2009-10-11-R126/FIGURES/5

Weider, L. J., Hobaek, A., Colbourne, J. K., Crease, T. J., Dufresne, F., & Hebert, P. D. N. (1999). Holarctic phylogeography of an asexual species complex I. Mitochondrial DNA variation in Arctic *Daphnia*. Evolution, 53, 777–792. https://doi.org/10.2307/2640718

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. Evolution (N. Y.), 38, 1358–1370. https://doi.org/10.2307/2408641

Wen Deng, H. (1997). Photoperiodic response of sexual reproduction in the *Daphnia pulex* group is reversed in two distinct habitats. Limnology and Oceanography, 42, 609–611. https://doi.org/10.4319/lo.1997.42.3.0609

Wersebe, M., Sherman, R., Jeyasingh, P. D., & Weider, L. J. (2021). Data For: The roles of recombination and selection in shaping genomic divergence in an incipient ecological species complex. Dryad Data Repository. https://doi.org/10.5061/dryad.xksn02vhs

Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. Nature Reviews Genetics, 18, 87–100. https://doi.org/10.1038/nrg.2016.133

Xu, S., Ackerman, M. S., Long, H., Bright, L., Spitze, K., Ramsdell, J. S., Thomas, W. K., & Lynch, M. (2015). A male-specific genetic map of the microcrustacean *Daphnia pulex* based on single-sperm whole-genome sequencing. Genetics, 201, 31–38. https://doi.org/10.1534/genetics.115.179028

Xu, S., Spitze, K., Ackerman, M. S., Bright, L., Keith, N., Jackson, C. E., Shaw, J. R., & Lynch, M. (2015). Hybridization and the origin of contagious asexuality in *Daphnia pulex*.

Molecular Biology and Evolution, 32, 3215–3225. https://doi.org/10.1093/molbev/msv190

Ye, Z., Molinier, C., Zhao, C., Haag, C. R., & Lynch, M. (2019). Genetic control of male production in *Daphnia pulex*. Proceedings of the National Academy of Sciences of the United States of America, 116, 15602–15609. https://doi.org/10.1073/pnas.1903553116

Ye, Z., Williams, E., Zhao, C., Burns, C. W., & Lynch, M. (2021). The rapid, mass invasion of New Zealand by North American *Daphnia "pulex"*. Limnology and Oceanography, 9999, 1–12. https://doi.org/10.1002/lno.11780

Ye, Z., Xu, S., Spitze, K., Asselman, J., Jiang, X., Ackerman, M. S., Lopez, J., Harker, B., Raborn, R. T., Thomas, W. K., Ramsdell, J., Pfrender, M. E., & Lynch, M. (2017). A new reference genome assembly for the microcrustacean *Daphnia pulex*. G3 (Bethesda), 7, 1405–1416. https://doi.org/10.1534/g3.116.038638

Figures & Tables:

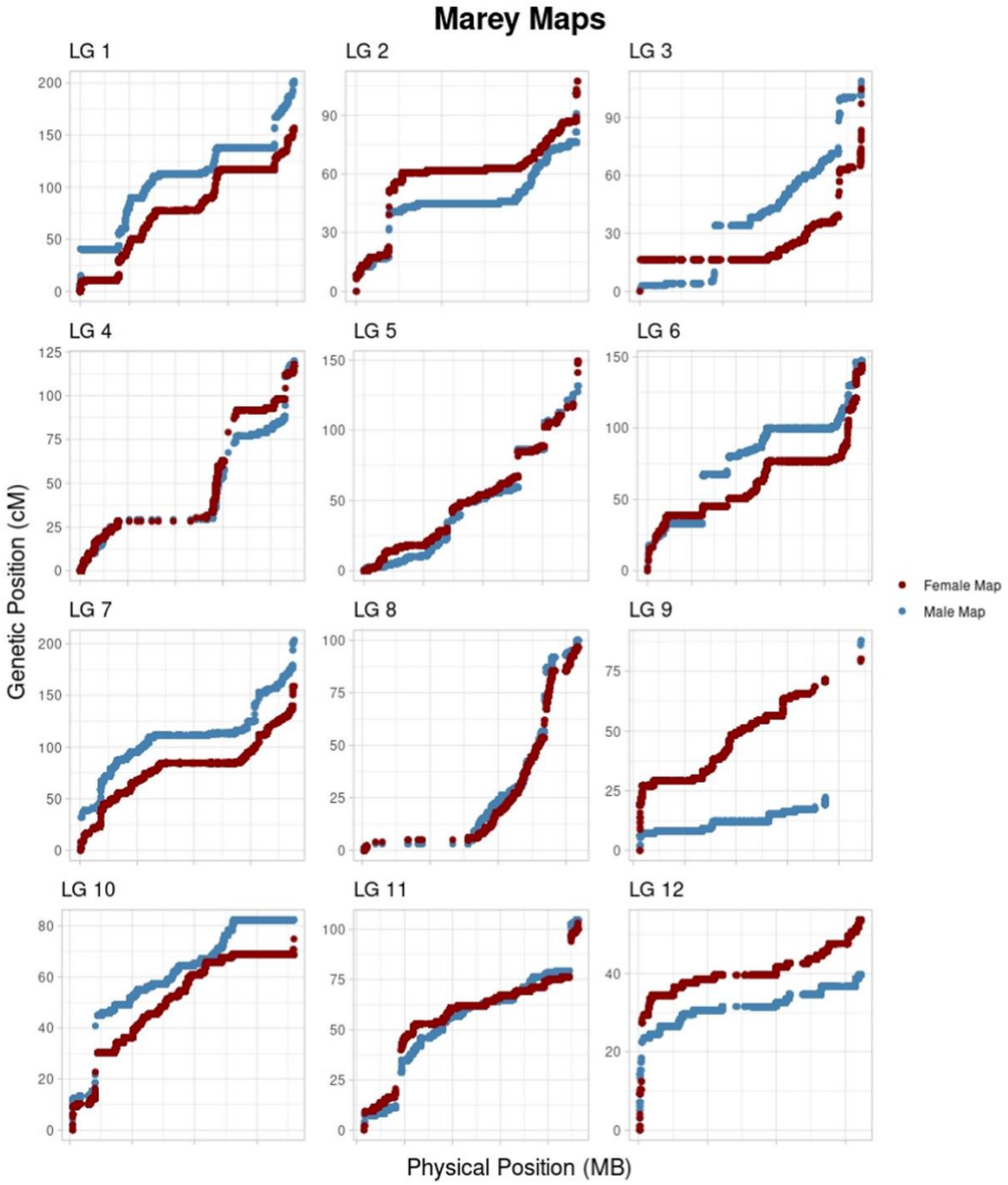


Figure 1: Linkage group (LG) Marey maps. The number of each LG map corresponds with the chromosome number assigned by LepMap3 and does not reflect reordering by total length. In each panel, the *x*-axis is the physical position (in base pairs) of markers along the chromosome and the *y*-axis represents the marker's corresponding genetic position (in cM). The color of the two lines corresponds to the sex-specific female (dark red) and male (light blue) maps. Marey maps are used to estimate the recombination rates along chromosomes by determining the slope of the line formed by plotting genetic and physical position of markers. Areas of positive slopes indicate regions of high recombination (e.g., LG2 ends), whereas regions of zero slopes indicate regions of suppressed recombination (e.g., LG2 center)

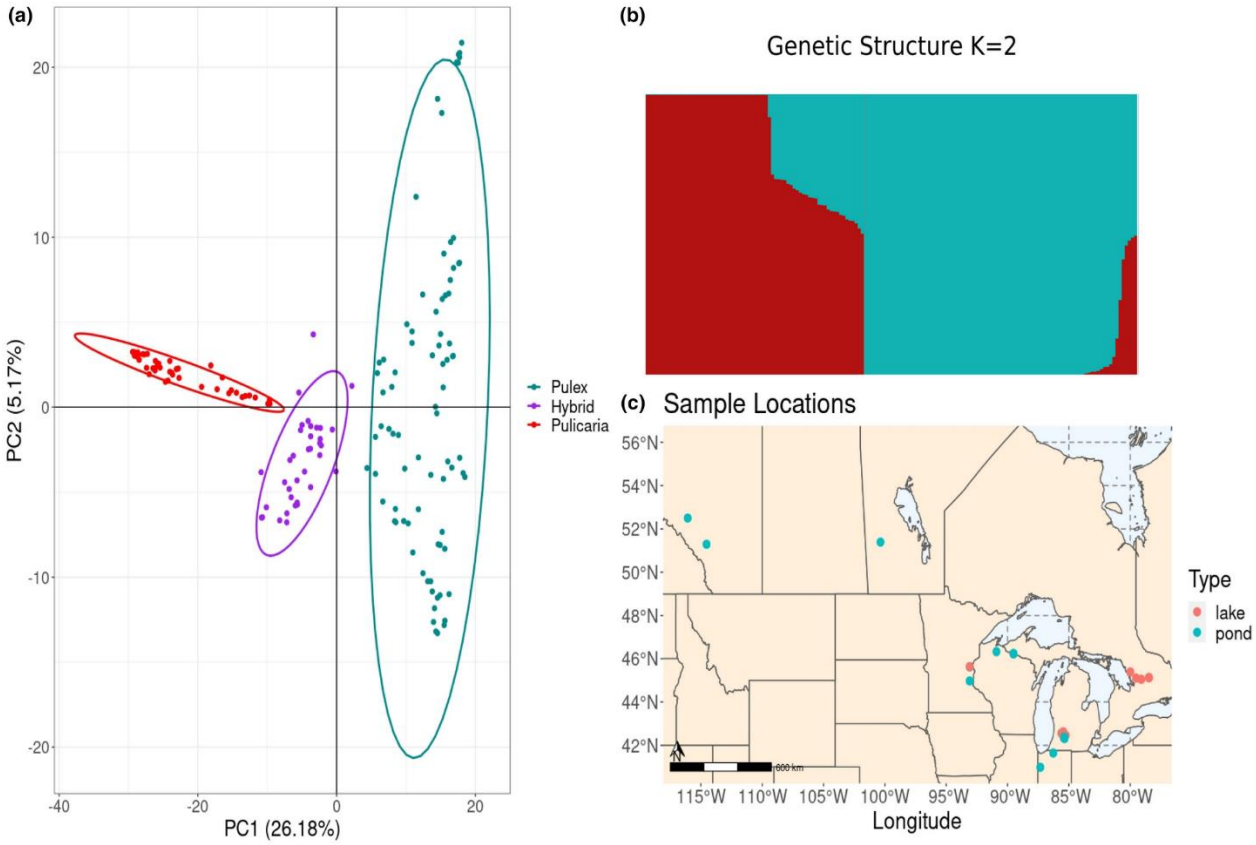


Figure 2: Population genetic structure of Daphnia pulex, Daphnia pulicaria, and their hybrids based on RADseq derived SNP genotypes. In all cases, SNP set analysed was pruned to include only one SNP per RAD locus to control for linkage disequilibrium. (a) Principal components analysis (PCA) biplot of sample genetic structure calculated using the R package Adegenet (Jombart, 2008). Ellipses indicated 95% confidence limits for groupings. (b) Bar plot of Bayesian posterior assignment probabilities of population using K = 2 for all samples calculated using Faststructure (Raj et al., 2014). (c) Sample collection locations (redrawn from Clifford (2016)). Lakes and ponds are denoted by the pink and blue dots respectively. Samples are primarily derived from the midwestern USA and the southern Canadian Provinces bordering the USA

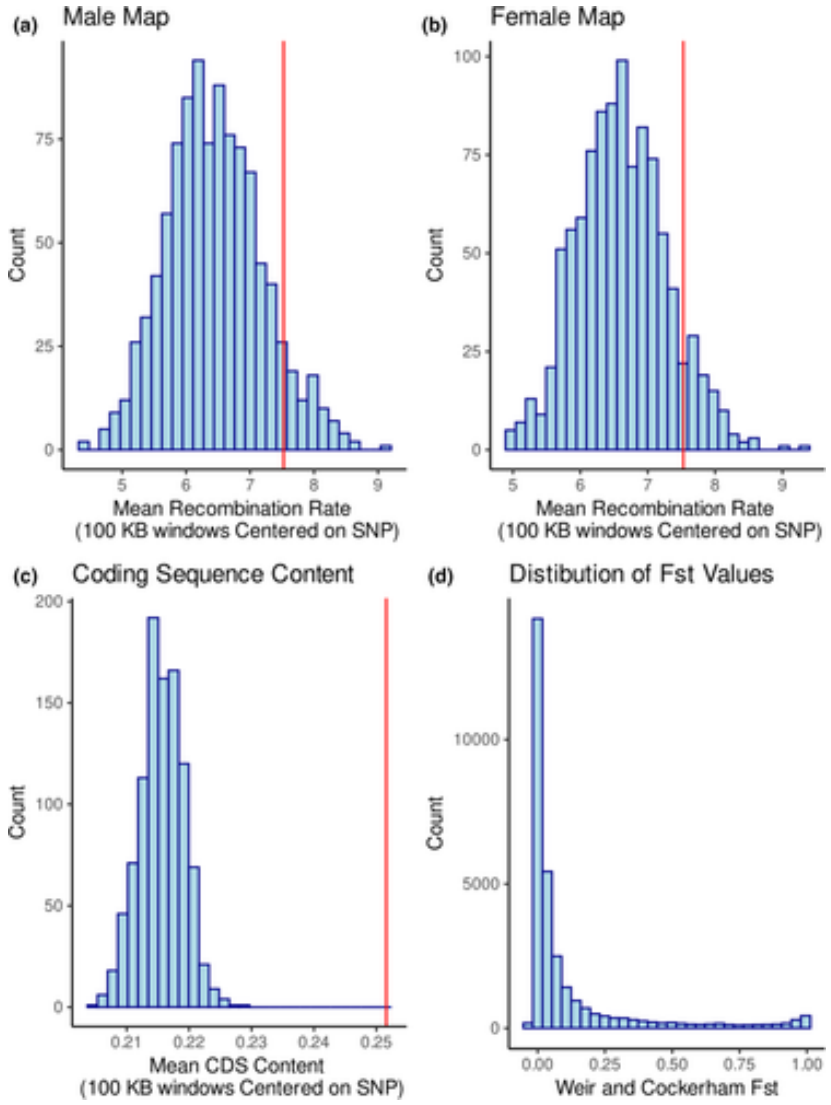


Figure 3: SNP outlier status and recombination or gene-density backgrounds. (a) Histogram of mean recombination rate surrounding 1000 permutations of 762 random background SNPs according to the male genetic map. Vertical red line indicates the mean recombination rate surrounding 762 outlier SNPs identified by VCFtools site-wise FST scan (Danecek et al., 2011). 80 of 1000 permutations are greater than the outlier SNPs. (b) Histogram of mean recombination rate surrounding 1000 permutations of 762 random background SNPs according to the female genetic map. Vertical red line indicates the mean recombination rate surrounding 762 outlier SNPs identified by VCFtools site-wise F_{ST} scan. 91 of 1000 permutations are greater than the outlier SNPs. (c) Histogram of mean coding DNA sequence (CDS) content around 1000 permutations of 762 randomly selected background SNPs. Vertical red line indicates the mean surrounding the 762 FST outlier SNPs. No permutation was larger than the observed outlier value. (d) Wright's FST distribution between *D. pulex* and *D. pulicaria* from all SNPs included in the single-SNP data sets.

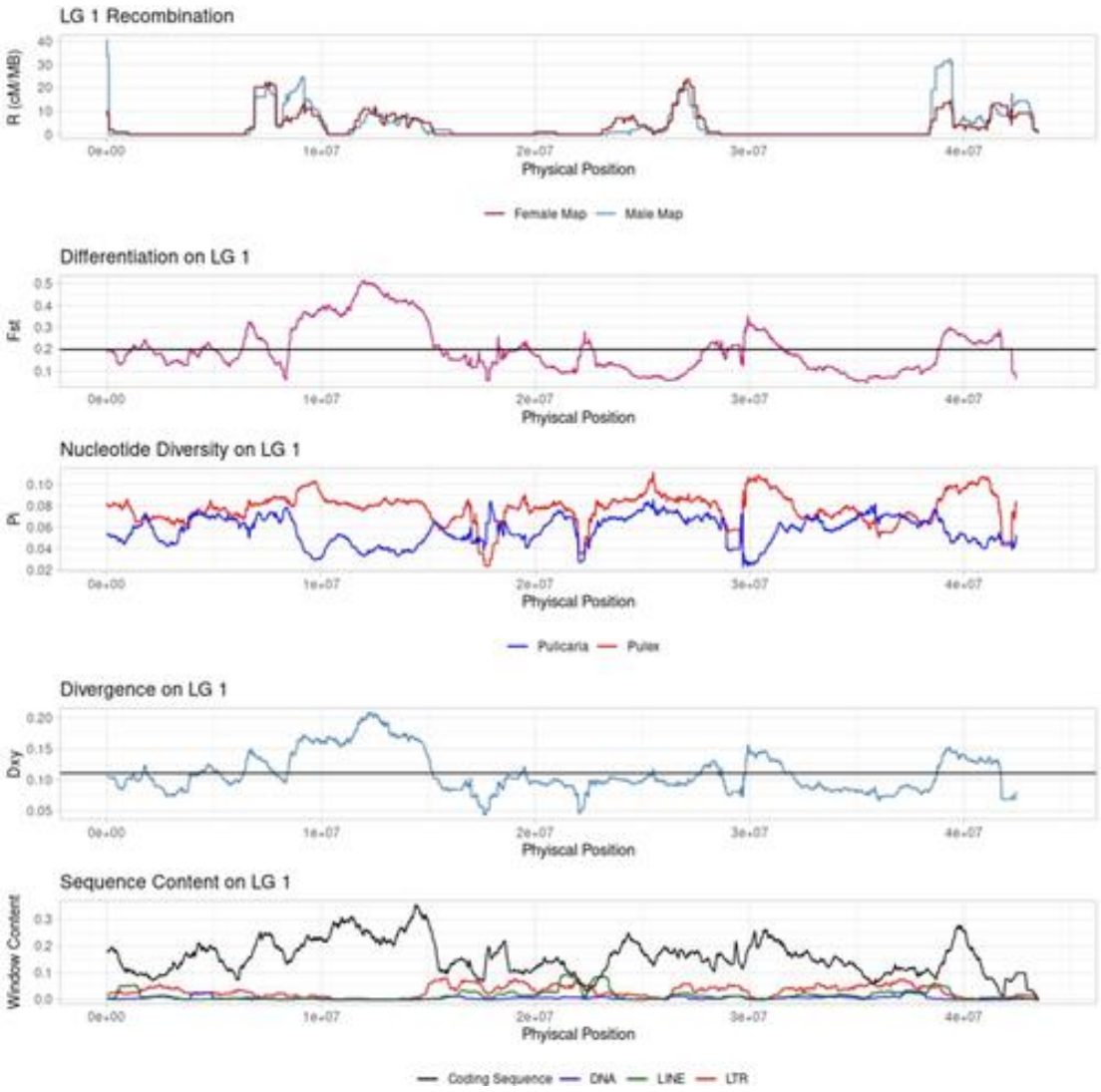


Figure 4: Sliding window analysis of linkage group (LG)-1 landscapes. Top: Recombination landscape in 1-Mb windows stepping 10 kb along the LG. Colours reflect the male (blue) and female (red) map positions. Second from top: Sliding window F_{ST} (relative differentiation) calculated between *D. pulicaria* and *D. pulex* in 1 Mb windows stepping 10 kb along the LG. Third from top: Sliding window π (nucleotide diversity) calculated for both *D. pulex* (red) and *D. pulicaria* (blue) in the 1 Mb windows stepping 10 kb along the LG. Fourth from top: Sliding window D_{xy} (absolute divergence) calculated between *D. pulicaria* and *D. pulex* in the 1 Mb windows stepping 10 kb along the LG. Bottom: Sliding window analysis of DNA sequence features along LG 1 in 1 Mb windows stepping 10 kb. The black line is coding DNA sequence (CDS) content, blue indicates the DNA transposons content, green long interspersed nuclear elements (LINEs) content and red indicates long terminal repeats (LTR) contents. Large highly differentiated and divergent regions (e.g., islands centred at ~11 and ~40 Mb) appear to align with regions of high recombination, high gene density, and low diversity in *D. pulicaria*. Smaller regions are more variable (e.g., island centred ~30 Mb) where gene density is high, but high repeat content probably is suppressing recombination.

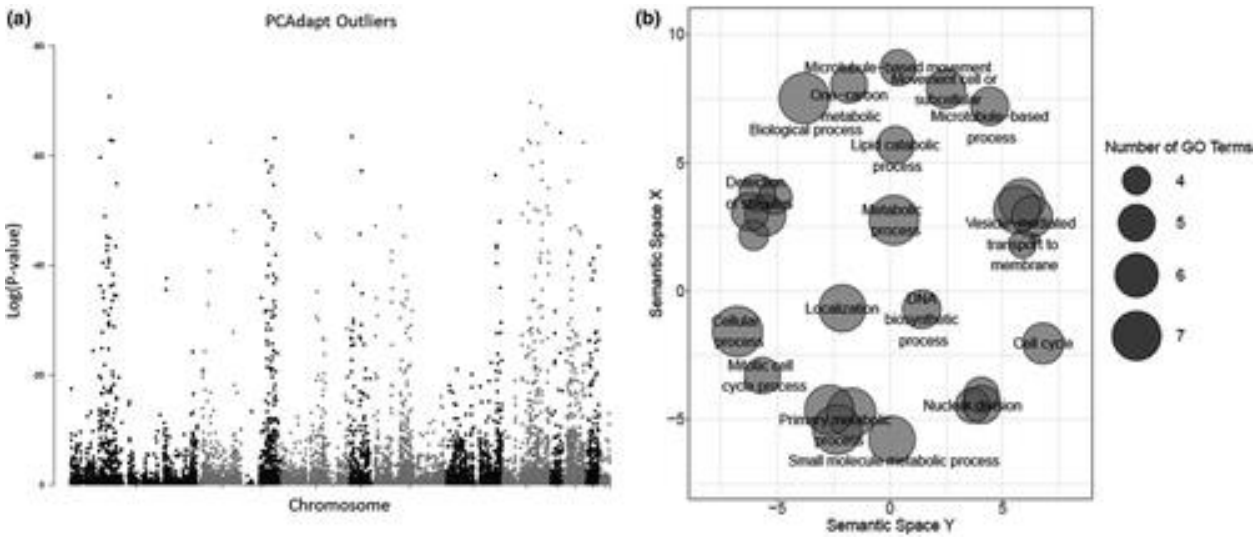


Figure 5: Statistical outlier analysis and gene ontology (GO) term enrichment. (a) PCAdapt Manhattan plot showing distribution of outliers throughout the genome. Locations of outliers correspond well with those detected via other methods (i.e., site-wise and sliding window F_{ST}). Note: chromosome 13 is made up of all 6 Mb of unplaced scaffolds. (b) REVIGO (Supek et al., 2011) semantic space clustering of GO terms. *x*- and *y*-axis represent clustering of GO terms in similar semantic space using multidimensional scaling (MDS). Semantic space is used to reduce the complexity of the input GO term list (here enriched terms among outliers) to group similar terms in reference to information content. Thus, the terms closely related to each other (e.g., hierarchically) share similar loadings along the *x*–*y* axis. The size of each circle is directly related to the number of enriched GO terms in the list inferred from PANTHER analysis (larger = more terms). The terms displayed as text are parent to the terms represented by unlabeled circles and are more general.

TABLE 1. Genome assembly statistics for each haplotype of the *Daphnia pulicaria* genome assembled by HiFiasm (columns 2 and 4)

Statistic	F_0 13B haplotype	Scaffolded assembly	F_0 2X haplotype	Scaffolded assembly
Total length	185,884,193 bp	185,227,602 bp	179,691,522 bp	179,392,649 bp
Longest contig/scaffold	11,563,873 bp	43,499,656 bp	17,778,422 bp	41,385,009 bp
Number contigs	253	160	180	205 [a]
N50	5.578 Mb	18.885 Mb	6.383 Mb	15.989 Mb

Note

Assembly statistics after scaffolding with the genetic map or by alignment (columns 3 and 5).

[a] Increased contig number represents an artefact of scaffolding by alignment where contigs are broken when they include poorly mapping regions.

TABLE 2. Sex-specific linkage map descriptive statistics for the *Daphnia pulicaria* genome separated by linkage group (LG)

Linkage group/chromosome	Male genetic length	Female genetic length	Physical length	Number of markers
1	201.935 cM	156.767 cM	43,499,656 bp	128,716
2	101.924	107.494	12,938,585	56,556
3	108.897	104.736	13,523,868	21,234
4	119.870	117.783	22,509,579	49,234
5	131.529	149.471	7,288,624	26,185
6	147.083	143.770	24,200,786	59,010
7	203.328	158.743	18,885,472	57,238
8	99.973	96.574	15,869,501	22,162
9	87.976	80.127	4,432,952	16,426
10	82.359	74.900	7,571,162	36,461
11	104.779	103.127	4,626,280	23,069
12	39.670	53.699	3,231,808	19,545

Note

Column 1: LG male genetic length in cM. Column 2: LG female genetic length in cM. Column 3: LG physical length in base pairs. Column 4: the number of markers in each LG in the final genetic map.

TABLE 3. Results of the isolation-with-migration (IM) demographic model for *Daphnia pulex* and *D. pulicaria* fit using Fastsimcoal2 (Excoffier et al., 2013)

	Nanc	NPa	NPx	Tdiv	Mig12 (Pa -> Px)	Mig21 (Px -> Pa)
Estimate	11,313	221,377	764,471	598,022	3.77×10^{-7}	9.48×10^{-7}
95% CI	11,023–11,378	216,978–225,729	749,244–780,081	581,444–606,220	3.76–3.92 ($\times 10^{-7}$)	9.30–9.65 ($\times 10^{-7}$)

Note

Top row: Maximum marginal likelihood estimates for model parameters in the IM model after selecting from 100 model runs using the observed folded site frequency spectrum (SFS). Bottom row: 95% confidence intervals for model estimates after performing simulations under 100 nonparametric bootstrapped SFS.

Does salinization impact long-term *Daphnia* assemblage dynamics? Evidence from the sediment egg bank in a small hard-water lake.

Published in: *Limnology and Oceanography Letters*, 13 October 2021

Available at: https://doi.org/10.1002/lol2.10217

Matthew J. Wersebe,[1] Mark B. Edlund,[2] Lawrence J. Weider[1]

[1] Program in Ecology and Evolutionary Biology, Department of Biology, University of Oklahoma, Norman, OK

[2] St. Croix Watershed Research Station, Science Museum of Minnesota, Marine-on-St. Croix, MN

Abstract:

Salinization of freshwater ecosystems threatens global aquatic biodiversity. There is a need for studies that follow populations in situ during salinization to understand the effects on species and ecosystems. We follow 170 years of *Daphnia* dynamics in the sediment ephippia archive of a small urban lake near St. Paul, Minnesota, to characterize effects of severe recent salinization on lake *Daphnia*. We found modest changes in the flux of ephippia in this lake; all three key *Daphnia* functional groups remained in the assemblage throughout the period of salinization. Reconstruction of the size distribution of *Daphnia pulicaria* demonstrated a consistent increase in body size in opposition to the expected trajectory during salinization. Our findings highlight that in hard-water lakes, the effects of salinization are nuanced and require further investigation to better understand overall impacts of salinization on hard-water lake *Daphnia* assemblages.

Scientific Significance Statement:

Many studies of recent anthropogenic salinization focus on establishing toxicological endpoints for species or use short-term mesocosm experiments that do not reflect ecological reality. Very little is known about the long-term dynamics of species and populations responding to salinization. We examine a 170-year record of *Daphnia* assemblage dynamics archived in lake sediments. Contrary to recent work published on lakes with divergent water chemistries, we find only modest shifts in the relative abundance of *Daphnia* ephippia. Our results indicate that a more detailed study of lake *Daphnia* responses to salinization is needed to disentangle the controls and drivers of species composition change during salinization.

Introduction:

Salinization is an emerging global threat to freshwater ecosystems (Reid et al. 2019). While the drivers of freshwater salinization are well understood (Kaushal et al. 2005, 2018; Dugan et al. 2017), the impacts to biological communities remain poorly resolved (Hintz and Relyea 2019; Iglesias 2020). Recent work on biological impacts of salinization has focused on single species toxicity assays or short-term mesocosm experiments (Hintz and Relyea 2019). These studies, which are critical for understanding the impacts of salinization, are often divorced from ecological reality and do not consider interactions with habitat conditions or long-term community dynamics.

Lake sediments are repositories that chronicle changes in organism abundance and diversity (Kerfoot et al. 1999; Rogalski et al. 2017; Burge et al. 2018) and allow us to explore the salinization impacts across broad temporal scales. For instance, durable cladoceran remains can be used to reconstruct historical shifts in zooplankton community dynamics related to various anthropogenic changes (e.g., fish stocking, pollution; Kerfoot et al. 1999; Jeppesen et al. 2002; Rogalski et al. 2017). This allows for reconstruction of community-level parameters and species-specific traits during periods of change. In the case of salinization, US Environmental Protection Agency guidelines have indicated that chloride concentrations beyond 230 mg Cl^{-} L^{-1} induce chronic effects, and chloride concentrations beyond 860 mg Cl^{-} L^{-1} induce acute negative effects in freshwater aquatic organisms such as crustacean zooplankton (e.g., *Daphnia*; US EPA 1988). Generally, lentic freshwater ecosystems have chloride concentrations below both acute and chronic thresholds (Dugan et al. 2017). However, questions remain whether lakes nonetheless change as a result of increasing salinity.

Studies using paleolimnological techniques have reconstructed shifts from *Bosmina spp*. dominated assemblages to those dominated by *Chydorus* and *Daphnia* in Canadian Shield lakes coincidental with road salting in Southern Canada (Arnott et al. 2020; Valleau et al. 2020). These shifts occurred despite only modest increases in chloride (~5–40 mg L−1) driven primarily by road salt application. However, Canadian Shield lakes are notable for their soft water (e.g., low [Ca2+] and [Mg2+]; Jeziorski et al. 2008; Valleau et al. 2020). Water hardness is an important parameter influencing chloride toxicity across many types of freshwater invertebrates (Elphick et al. 2011; Soucek et al. 2011). Thus, dramatic shifts in species composition during salinization may be unique to soft water lakes. In addition, some lakes in the Northern United States have seen more dramatic increases in salinities than those observed in Southern Canada (Dugan et al. 2017, 2020). In some cases, salinity increases have disrupted water column mixing dynamics of lakes, a condition known as "cultural meromixis" (Sibert et al. 2015; Wyman and Koretsky 2018). The transition to cultural meromixis impacts the quality and quantity of lake habitat available for key groups such as zooplankton. Loss of mixing and stratification in lakes can have major impacts on key zooplankton groups such as cladocerans, which undergo diel vertical migration to avoid predators and partition habitat vertically within the water column (Stich and Lampert 1981; Lampert et al. 2003). A lack of mixing reduces the size and persistence of an oxygenated hypolimnic refuge for large-bodied, predation-intolerant cladocerans, such as Daphnia. Such a refuge is critical for maintaining *Daphnia* populations throughout the growing season (Wright and Shapiro 1990). Thus, despite not reaching "toxic" salinity levels, lakes may experience loss of seasonally available habitats which could potentially lead to extirpation of key functional groups within plankton assemblages. To better understand the impacts of salinization

on key zooplankton groups we should jointly consider background water chemistry and seasonal mixing.

Here, we used the sediment subfossil archive from a salinized suburban Minnesota lake to document the impact that 70 years of salinization has had on *Daphnia* assemblage dynamics. *Daphnia* are keystone species in lake food webs linking primary production to higher trophic levels; large-bodied species provide key ecosystem services (Lampert and Sommer 2007; Ogorelec et al. 2021). Specifically, we identified and enumerated *Daphnia* ephippia preserved in sediments to characterize species abundance and diversity within the assemblage over the course of 170 years. In addition, we tested for a correlated shift that reflected a decrease in *Daphnia pulicaria* body size to illuminate potential drivers of community change (e.g., increasing predation) related to cultural meromixis. Our results highlight critical additional hypotheses needed to address the nuanced ecological impacts of recent salinization and concomitant altered physicochemical structure of lake habitats.

Methods:

Study site:

Tanners Lake (TL) is a small (27.66 ha) hard-water lake (>120 mg L− CaCO3) located in Oakdale, Minnesota (44.9509°N, −92.9813°W). The watershed (~230.5 ha) is dominated by suburban development with ~32% impervious surface, which includes residential development, parking lots, and interstate highways (U.S. Geological Survey 2016). TL receives an estimated 524 kg Cl− ha−1 yr−1 primarily from road deicing salts (primarily NaCl) from watershed runoff (Novotny and Stefan 2010). Previous studies have estimated the background pre-Euro-American colonization chloride concentration at 1–2 mg L−1 using diatom-inferred transfer functions

(Ramstack et al. 2003, 2004). Contemporary chloride concentrations range from ~150 mg L−1 at the surface to >350 mg L−1 in the bottom waters. Concomitant with increased salinity, TL has transitioned to a meromictic state, where bottom waters remain unmixed throughout the growing season (Fig. 1A–C; Novotny and Stefan 2012).

Sediment collection:

On 2 July 2019, we collected duplicate 1.5-m sediment cores using a piston corer and rigid drive rods operated from a pontoon anchored in ~14 m of water (Wright 1991). We extruded the top 20 cm of both cores in the field in 2-cm intervals (representing ~2–5 years) into 125-mL polycarbonate sampling cups and placed them in an ice-chest for transport and further processing in the lab. The remainder of the core was capped for transport to the lab, where it was sectioned into 2-cm intervals to a depth of 92-cm to capture a record from present-day through the Euro-American settlement (ca. 1850 AD). One core was retained for the collection of sub-fossil ephippia; the second was used for sediment geochemical analysis (see Supplemental Methods in Supporting Information Data S1).

Isolation of ephippial remains:

We took three subsamples from each core section (~5–15 g/subsample wet weight). Each subsample was washed through a series of nested metal sieves (710, 425, and 300 μm mesh sizes) using COMBO medium (Kilham et al. 1998). Each size fraction was surveyed for the isolation of Daphnia ephippia under a stereomicroscope (Olympus SZ61). All ephippia were collected and enumerated from the sediment and grouped by species-specific morphological characteristics consistent with previous studies (Mergeay et al. 2004; Rogalski et al. 2017). For each layer, the flux (#ephippia g sediment−1 yr−1) of each species' ephippia to the sediment was

calculated using a dating and sedimentation rate model (see Supporting Information Data S1). Furthermore, a subset of ephippia (up to 30) from each core section identified as *D. pulicaria* were measured for dorsal length (to the nearest 0.1 mm) using an ocular micrometer mounted to a Leica MZ-8 stereomicroscope to reconstruct body size trajectories of *D. pulicaria* (Jeppesen et al. 2002).

Statistical analysis:

We examined the relationship among ephippial fluxes for each species, estimated chloride concentration, and sediment geochemical attributes (% organic and % CaCO3) using generalized additive mixed models (GAMM) fitted using the package mgcv (version 1.8-33; Wood 2017) in R (version 4.0.2). Each species' GAMM model was fit using a continuous time first-order autoregressive process and restricted maximum likelihood (method = "REML") to account for the temporal (time-series) nature of the data and the uncertainty in the number of years contained within each sediment slice using Poisson errors. To assess size-structure changes of *D. pulicaria* over the full length of the core, we fit a linear model (LM), which analyzed ephippia dorsal length by depth of sediment using R (version 4.0.2; R Core Team 2020). Data for recreating this analysis are available in the Dryad Repository (Wersebe et al. 2021).

Results:

From the sediments, we recovered ephippia from four *Daphnia* species-complexes common to north temperate lakes: *D. pulicaria, Daphnia mendotae, Daphnia parvula, and Daphnia retrocurva* (Fig. 2A–C). Due to difficulty in distinguishing between ephippia from *D. parvula and D. retrocurva* (i.e., poor preservation; low hatching success), we collapsed this closely related species-complex into a single category which we call "small ephippia."

Daphnia ephippial flux:

Daphnia ephippia were present in every core layer sampled and ranged in total abundance from 83 to 3282 ephippia. The flux (#ephippia g sediment−1 yr−1) of *D. pulicaria* ephippia was positively correlated with chloride concentration (GAMM; edf = 2.799, r2 = 0.719, p = 0.0122; Fig. 3A) and significantly correlated with age of the sediment sampled. A peak in ephippia flux occurred around the late 1870s (~1878) and decreased until the late 1940s (~1948), at which point the flux stabilized until present day (GAMM; edf = 8.232, r2 = 0.719, p < 0.0001, Supporting Information Fig. S4A). The flux of "small ephippia" was only significantly correlated with chloride, with an initial increase in flux until approximately 125 mg L−1, after which there was a precipitous drop in small ephippial flux (GAMM; edf = 6.832, r2 = 0.385, p < 0.0001; Fig. 3B). The flux of *D. mendotae* ephippia showed a negative linear correlation with water-column chloride levels (GAMM; edf = 1.00, r2 = 0.363, p = 0.00347; Fig. 3C), along with a significant positive correlation with % CaCO3 in the sediments (GAMM; edf = 1.00, r2 = 0.363, p = 0.02576; Supporting Information Fig. S5). With the exception of *D. pulicaria*, models had low adjusted r^2 values (Table 1), indicating that other factors that remain unmeasured may also be important.

Size structure of D. pulicaria ephippia:

D. pulicaria ephippia showed a consistent average size increase (i.e., ~0.002 mm in each 2-cm layer) moving up-core from 90 to 2 cm sediment depth (LM; adjusted r2 = 0.1656, F1,778 = 155.6, p < 0.0001; Fig. 4).

Discussion:

Our goal was to reconstruct *Daphnia* assemblage dynamics in TL that were recorded in the sediment ephippial archive before and during salinization. Our results suggest that *Daphnia* assemblage structure has remained remarkably resilient in TL despite rapid and extreme changes in salinity during the last 75 years. The sediment ephippial bank revealed that all three common functional size classes of *Daphnia* (i.e., large-bodied *D. pulicaria*; medium-sized *D. mendotae*; "small ephippia" of *D. parvula/D. retrocurva*) remained detectable in the sediments throughout the period of salinization. All three species groups have remained active in the water column during salinization, in direct contrast to large detectible shifts observed in other paleolimnological studies of salinized lakes (Arnott et al. 2020; Valleau et al. 2020).

The concentration of chloride was significantly correlated with species' flux rates (i.e., delivery of ephippia to the sediments). *D. pulicaria* ephippia exhibited an increase in flux while small ephippia exhibited a decrease in flux as chloride levels exceeded an estimated 125 mg L−1. The flux of *D. mendotae* ephippia showed a negative linear relationship with changes in water column chloride concentration, along with a significant positive relationship with the % CaCO3 present in the core. Furthermore, we hypothesized that if large-bodied *D. pulicaria* were impacted by the loss of suitable hypolimnetic refuge, then their body size would decrease over the period of salinization, as the result of increasing predation selecting against larger-bodied animals. However, we observed a consistent increase in the dorsal length of *D. pulicaria* ephippia recovered from the sediments over the entire length of the sampled core. There are several potential explanations for the observed patterns.

Cultural meromixis and the loss of deep-water habitat:

One consequence of salinization in TL has been the transition to cultural meromixis with the increase in water column salinity (Novotny and Stefan 2012). This condition may lead to the loss or reduction of large-bodied *Daphnia* species because they generally require an oxygenated, albeit ephemeral, hypolimnetic predator refuge, which is critical for their persistence during the growing season (Wright and Shapiro 1990; Tessier and Welser 1991). If significant loss of a deep-water refuge were apparent in our data, one would expect to observe decreasing flux of *D. pulicaria* ephippia correlated with increased chloride concentrations and significant reductions in the dorsal length of the ephippia recovered. Our results showed neither of these trends and indeed exhibited the exact opposite patterns. The effects of a hypoxic refuge may be alleviated if some *Daphnia* clones increase hemoglobin production to enhance oxygen uptake and become hypoxia "specialists" (Weider and Lampert 1985). Weider and Lampert (1985) reported that *Daphnia* have a physiological limit 0.5–1.0 mg L−1 DO, below which animals could not survive regardless of hemoglobin production. In TL, summer depths >8 m (Fig. 1B) are below this threshold DO level, placing them outside the physiological limits of *Daphnia.* However, an anoxic hypolimnion may not be an absolute constraint on *Daphnia* populations. Mesocosm experiments have shown that *D. pulicaria* can exploit anoxic areas outside the physiological limits of fish predators as refugia, albeit temporarily, but with life-history costs (Larsson and Lampert 2011).

If habitat deeper than 8 m is generally unavailable for large predation-intolerant *Daphnia* species, then they must exploit areas higher in the water column where oxygen conditions are more suitable. However, inhabiting shallower depths comes with risks of higher predation pressure. Generally, clones that successfully exploit these riskier habitats exhibit smaller body sizes (Leibold and Tessier 1991; De Meester et al. 1995). The increase in body size of the largest

daphniid in this assemblage suggests that predation from visually oriented predators may not be strong in this system. Limited availability of fish community data prevents us from further speculating on the ecology of altered food web dynamics in TL. However, to fully evaluate whether these altered dynamics are operating, surveys of *Daphnia* diel vertical (and horizontal) migration behavior, the distribution of hemoglobin and body size phenotypes, and an accounting of predator density should be made in culturally meromictic lakes to understand how impactful non-mixing layers of the water column are in altering species interactions.

Contrasting responses of zooplankton in a hard-water system:

Perhaps the most interesting result from our study was the remarkable stability in assemblage composition during the period of salinization. While flux rates were correlated with increasing chloride in complex ways, each of the functional classes of *Daphnia* common to Northern Temperate Lakes remained detectible in the sediments. This is in direct contrast with evidence from Canadian Shield lakes, which have shown dramatic shifts in zooplankton assemblages reconstructed from various subfossil remains, despite modest shifts in the chloride content of the water column (Arnott et al. 2020; Valleau et al. 2020). The contradiction between our results and those from Canadian lakes should be a major focus of studies seeking to predict the long-term responses of lake ecosystems to salinization. Of particular importance may be buffering capacity (Jeziorski et al. 2008) or the contrasting trophic state of the lakes (Brown and Yan 2015), both of which impact zooplankton population and community responses under salinization.

Canadian Shield lakes are known for their low buffering capacity, that is, soft water, where [Ca2+] and [Mg2+] are low compared to other regions and related to underlying bedrock (Moser et al. 1998; Jeziorski et al. 2008). Studies have indicated that water hardness plays an important role in invertebrate chloride tolerance. For instance, Elphick et al. (2011) and Soucek

et al. (2011) both reported increased chloride sensitivity of cladoceran zooplankton (i.e., *Ceriodaphnia dubia* and *Daphna magna*) when water hardness was low. In both cases, of the animals tested, including species ranging from other crustaceans to mollusks, insects, and fish, cladocerans were the most or among the most sensitive to chloride in soft water. However, tests of chloride sensitivity under contrasting water hardness conditions are restricted to a few common toxicological models (e.g., *Ceriodaphnia*) and are unavailable for *Daphnia* species commonly found in Northern Temperate Lakes (Elphick et al. 2011; Soucek et al. 2011). Many Canadian Shield lakes are also far more oligotrophic and have low productivity (Brown and Yan 2015). TL is surrounded by urban development and therefore is far more eutrophic than Canadian Shield lakes (e.g., Supporting Information Fig. S1A). This may provide higher quantity and/or higher quality food resources for *Daphnia* populations to exploit and persist under increasing salinization, as previous work with *Daphnia* has shown (Brown and Yan 2015). Additionally, the increase in *D. pulicaria* size and over-all flux in recent sediments may reflect their superior competitive ability and release from competition with other smaller *Daphnia* species in the salinized and high productivity setting (Gliwicz 1990). This provides an additional potential explanation for the unexpected increase in *D. pulicaria* flux and size in recent sediments.

Caveats and conclusions:

Evidence from sediment egg banks must be considered critically. First and foremost, there is no empirical conversion from egg bank abundances to active animal abundance in the water column (Rogalski et al. 2017). However, it is generally thought that an individual should produce relatively fewer ephippia than other commonly assayed subfossil remains (e.g., shed carapaces) over a lifetime; thus our abundance estimate would be more conservative (Arnott et al. 2020).

However, given the unique cyclically parthenogenetic life-histories of *Daphnia*, lake populations may delay sex and thus ephippia production for several growing seasons if lake oxygen dynamics allow (Tessier and Cáceres 2004). Also, having only taken one core from a single pelagic station limits our ability to detect some species and within-lake variation. Lake sediment sampling from different locations within a lake would presumably increase the ability to detect species and their historical partitioning behavior via the sampling of the egg bank (Vandekerkhove et al. 2005). In addition, as with other subfossil indicators, differential preservation of some ephippia over others is important to consider. For instance, in our study, we had difficulty assigning small ephippia definitively to *D. parvula* or *D. retrocurva* because of poor preservation in deeper sediment layers and low hatching success, so we considered these species as a group. Cryptic changes in their relative abundances may provide important additional information that is unavailable from our study.

Previous studies using egg bank-derived clones have revealed idiosyncratic patterns when considered in conjunction with other paleo-indicators. For example, Elmarsafy et al. (2020) found that ephippia of the cladoceran, *Ceriodaphnia dubia*, remained in the egg bank of a naturally saline lake, despite predictions of salinity (from diatom-based transfer functions) exceeding estimated EC_{50} for salt during these time periods. Thus, considering additional evidence from other sources (e.g., genetic surveys) from sediment egg banks will be an important next step for increasing our understanding and interpretation of the sediment record in lakes (Frisch et al. 2017).

In conclusion, our data reveal remarkable stability in *Daphnia* assemblage dynamics in TL, despite a greater than 100- to 300-fold increase in water column chloride over the last 75 years. Our study suggests that despite approaching chronic limits for water column chloride,

only subtle shifts in flux rates occur in the daphnid egg bank in well-buffered and productive TL. These results imply that *Daphnia* communities in hard-water lakes in the continental United States may be relatively resilient and not experience outright loss of species richness, despite actual or projected increases in salinity. Clearly, future research is warranted to expand this work to encompass a greater number and diversity of lakes and integrate experimental tests (e.g., salinity tolerance assays on modern and resurrected clones) to better understand the adaptive response of *Daphnia* to salinity changes to determine the robustness of our findings.

Acknowledgments:

We thank the St. Croix Watershed Research Station and the University of Oklahoma Biological Station for logistical support. This work was supported in part by grants from the University of Oklahoma Graduate Student Senate and Adams Summer Fellowship to M.W. Field work was supported by the University of Oklahoma Faculty Investment Program to L.J.W. Laboratory assistance was provided in part by K. Gibbons and J. Agadagba. We thank J. Ulrich (SCWRS) for help with GIS rendering. Finally, we thank three anonymous reviewers for their constructive comments on an earlier version of the manuscript.

Works Cited:

Arnott, S. E., M. P. Celis-Salgado, R. E. Valleau, A. M. Desellas,A. M. Paterson, N. D. Yan, J. P. Smol, and J. A. Rusak. 2020. Road salt impacts freshwater zooplankton at concentrations below current water quality guidelines. Environ. Sci.Technol.54: 9398–9407. doi:10.1021/acs.est.0c02396

Brown, A. H., and N. D. Yan. 2015. Food quantity affects the sensitivity of *Daphnia* to road salt. Environ. Sci. Technol.49: 4673–4680. doi:10.1021/es5061534

Burge, D. R. L., M. B. Edlund, and D. Frisch. 2018. Paleolimnology and resurrection ecology: The future of reconstructing the past. Evol. Appl.11:42–59. doi:10.1111/eva.12556

De Meester, L., L. J. Weider, and R. Tollrian. 1995. Alternative antipredator defenses and genetic polymorphism in a pelagic predator–prey system. Nature 378: 483–485. doi:10.1038/378483a0

Dugan, H. A., and others. 2017. Salting our freshwater lakes. Proc. Natl. Acad. Sci. USA 114: 4453–4458. doi:10.1073/pnas.1620211114

Dugan, H. A., and others. 2020. Lakes at risk of chloride contamination. Environ. Sci. Technol. 54: 6639–6650. doi:10.1021/acs.est.9b07718

Elmarsafy, M., K. L. Tasky, and D. K. Gray. 2020. Can zooplankton on the North American Great Plains “keep up” with climate-driven salinity change? Limnol. Oceanogr.66: 865–877. doi:10.1002/lno.11647

Elphick, J. R. F., K. D. Bergh, and H. C. Bailey. 2011. Chronic toxicity of chloride to freshwater species: Effects of hard-ness and implications for water quality guidelines. Environ. Toxicol. Chem.30: 239–246. doi:10.1002/etc.365

Findlay, S. E. G., and V. R. Kelly. 2011. Emerging indirect and long-term road salt effects on ecosystems. Ann. N.Y. Acad.Sci.1223:58–68. doi:10.1111/j.1749-6632.2010.05942.x

Frisch, D., P. K. Morton, B. W. Culver, M. B. Edlund, P. D. Jeyasingh, and L. J. Weider. 2017. Paleogenetic records of *Daphnia pulicaria* in two North American lakes reveal the impact of cultural eutrophication. Glob. Chang. Biol.23: 708–718. doi:10.1111/gcb.13445

Gliwicz, Z. M. 1990. Food thresholds and body size in cladocerans. Nature 343: 638–640. doi:10.1038/343638a0

Hintz, W. D., and R. A. Relyea. 2019. A review of the species, community, and ecosystem impacts of road salt salinization in fresh waters. Freshw. Biol.64: 1081–1097. doi:10.1111/fwb.13286

Iglesias, M. C. A. 2020. A review of recent advances and future challenges in freshwater salinization. Limnetica. 39: 185–211.doi:10.23818/limn.39.13

Jeppesen, E., J. P. Jensen, S. Amsinck, F. Landkildehus, T. Lauridsen, and S. F. Mitchell. 2002. Reconstructing the historical changes in *Daphnia* mean size and planktivorous fish abundance in lakes from the size of *Daphnia* ephippia in the sediment. J. Paleolimnol.27: 133–143. doi:10.1023/A:1013561208488

Jeziorski, A., and others. 2008. The widespread threat of calcium decline in fresh waters. Science 322: 1374–1377. doi:10.1126/science.1164949

Kaushal, S. S., P. M. Groffman, G. E. Likens, K. T. Belt, W. P. Stack, V. R. Kelly, L. E. Band, and G. T. Fisher. 2005. Increased salinization of fresh water in the northeastern United States. Proc. Natl. Acad. Sci. USA 102: 13517–13520. doi:10.1073/pnas.0506414102

Kaushal, S. S., G. E. Likens, M. L. Pace, R. M. Utz, S. Haq, J.Gorman, and M. Grese. 2018. Freshwater salinization syndrome on a continental scale. Proc. Natl. Acad. Sci. USA 115:E574–E583. doi:10.1073/pnas.1711234115

Kerfoot, W. C., J. A. Robbins, and L. J. Weider. 1999. A new approach to historical reconstruction: Combining descriptive and experimental paleolimnology. Limnol. Oceanogr.44: 1232–1247. doi:10.4319/lo.1999.44.5.1232

Kilham, S. S., D. A. Kreeger, S. G. Lynn, C. E. Goulden, and L. Herrera. 1998. COMBO: A defined freshwater culture medium for algae and zooplankton. Hydrobiologia 377: 147–159. doi:10.1023/A:1003231628456

Lampert, W., E. McCauley, and B. F. J. Manly. 2003. Trade-offs in the vertical distribution of zooplankton: Ideal free distribution with costs? Proc. R. Soc. B Biol. Sci.270: 765–773.doi:10.1098/rspb.2002.2291

Lampert, W., and U. Sommer. 2007, Limnoecology: The ecology of lakes and streams, 2nd Edition. Oxford Univ. Press.

Larsson, P., and W. Lampert. 2011. Experimental evidence of a low-oxygen refuge for large zooplankton. Limnol. Ocean-ogr.56: 1682–1688. doi:10.4319/lo.2011.56.5.1682

Leibold, M., and A. J. Tessier. 1991. Contrasting patterns of body size for *Daphnia* species that segregate by habitat. Oecologia 86: 342–348. doi:10.1007/BF00317599

Mergeay, J., D. Verschuren, L. Van Kerckhoven, and L. DeMeester. 2004. Two hundred years of a diverse *Daphnia* community in Lake Naivasha (Kenya): Effects of natural and human-induced environmental changes. Freshw. Biol.49: 998–1013. doi:10.1111/j.1365-2427.2004.01244.x

Moser, K. A., J. P. Smol, D. R. S. Lean, and G. M. MacDonald.1998. Physical and chemical limnology of northern boreal lakes, Wood Buffalo National Park, northern Alberta and the Northwest Territories, Canada. Hydrobiologia 377:2543. doi:10.1023/A:1003225527053

Novotny, E. V., and H. G. Stefan. 2010. Projections of chloride concentrations in urban lakes receiving road de-icing salt. Water Air Soil Pollut. 211: 261–271. doi:10.1007/s11270-009-02970

Novotny, E. V., and H. G. Stefan. 2012. Road salt impact on lake stratification and water quality. J. Hydraul. Eng.138: 1069–1080. doi:10.1061/(ASCE)HY.1943-7900.0000590

Ogorelec,Ž., C. Wunsch, A. J. Kunzmann, P. Octorina, and J. I.Navarro. 2021. Large daphniids are keystone species that link fish predation and phytoplankton in trophic cascades. Fundam. Appl. Limnol.194:297–309. doi:10.1127/fal/2020/1344

Ramstack, J. M., S. C. Fritz, D. R. Engstrom, and S. A. Heiskary. 2003. The application of a diatom-based transfer function to evaluate regional water-quality trends in Minnesota since 1970. J. Paleolimnol.29:79–94. doi:10.1023/A:1022869205291

Ramstack, J. M., S. C. Fritz, and D. R. Engstrom. 2004. Twentieth century water quality trends in Minnesota lakes compared with pre-settlement variability. Can. J. Fish. Aquat.Sci.61: 561–576. doi:10.1139/f04-015

R Core Team. 2020.R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Version 4.0.2. https://www.R-project.org/

Reid, A. J., and others. 2019. Emerging threats and persistent conservation challenges for freshwater biodiversity. Biol. Rev.94: 849–873. doi:10.1111/brv.12480

Rogalski, M. A., P. R. Leavitt, and D. K. Skelly. 2017. Daphniid zooplankton assemblage shifts in response to eutrophication and metal contamination during the Anthropocene. Proc. R. Soc. B Biol. Sci.284: 20170865. doi:10.1098/rspb.2017.0865

Sibert, R. J., C. M. Koretsky, and D. A. Wyman. 2015. Cultural meromixis: Effects of road salt on the chemical stratification of an urban kettle lake. Chem. Geol.395: 126-137. doi:10.1016/J.CHEMGEO.2014.12.010

Soucek, D. J., T. K. Linton, C. D. Tarr, A. Dickinson, N. Wickramanayake, C. G. Delos, and L. A. Cruz. 2011. Influence of water hardness and sulfate on the acute toxicity of chloride to sensitive freshwater invertebrates. Environ. Toxicol. Chem.30: 930–938. doi:10.1002/etc.454

Stich, H.-B., and W. Lampert. 1981. Predator evasion as an explanation of diurnal vertical migration by zooplankton. Nature.293: 396–398. doi:10.1038/293396a0

Tessier, A. J., and J. Welser. 1991. Cladoceran assemblages, seasonal succession, and the importance of a hypolimnetic refuge. Freshw. Biol.25:85–93. doi:10.1111/j.1365-2427.1991.tb00475.x

Tessier, A. J., and C. E. Caceres. 2004. Differentiation in sex investment by clones and populations of *Daphnia*. Ecol.Lett.7: 695–703. doi:10.1111/j.1461-0248.2004.00627.x

U.S. Environmental Protection Agency. 1988. Aquatic life ambient water quality criteria for chloride—1988. US Government Publishing Office.

U.S. Geological Survey. 2016. The StreamStats program for Minnesota, http://water.usgs.gov/osw/streamstats/minnesota.html.

Valleau, R. E., A. M. Paterson, and J. P. Smol. 2020. Effects of road-salt application on Cladocera assemblages in shallow Precambrian Shield lakes in south-central Ontario. Canada. Freshw. Sci.39: 824–836. doi:10.1086/711666

Vandekerkhove, J., S. Declerck, E. Jeppesen, J. M. Conde-Porcuna, L. Brendonck, and L. De Meester. 2005. Dormant propagule banks integrate spatio-temporal heterogeneity in cladoceran communities. Oecologia142: 109–116. doi:10.1007/s00442-004-1711-3

Weider, L. J., and W. Lampert. 1985. Differential response of *Daphnia* genotypes to oxygen stress: Respiration rates, hemoglobin content and low-oxygen tolerance. Oecologia 65: 487–491. doi:10.1007/BF00379661

Wersebe, M., M. B. Edlund, and L. J. Weider. 2021. Does salinization impact long-term *Daphnia* assemblage dynamics? Evidence from the sediment egg bank in a small hard-waterlake. Dryad Dataset. doi:10.5061/dryad.tqjq2bw0j

Wood, S. N. 2017. Generalized additive models: An introduction with R, 2nd Edition. CRC Press.

Wright, D., and J. Shapiro. 1990. Refuge availability: A key to understanding the summer disappearance of *Daphnia*. Freshw. Biol.24:43–62. doi:10.1111/j.1365-2427.1990.tb00306.x

Wright, H. E. 1991. Coring tips. J. Paleolimnol.6:37–49. doi:10.1007/BF00201298

Wyman, D. A., and C. M. Koretsky. 2018. Effects of road salt deicers on an urban groundwater-fed kettle lake. Appl.Geochem.89: 265–272. doi:10.1016/J.APGEOCHEM.2017.12.023

Figures & Tables:

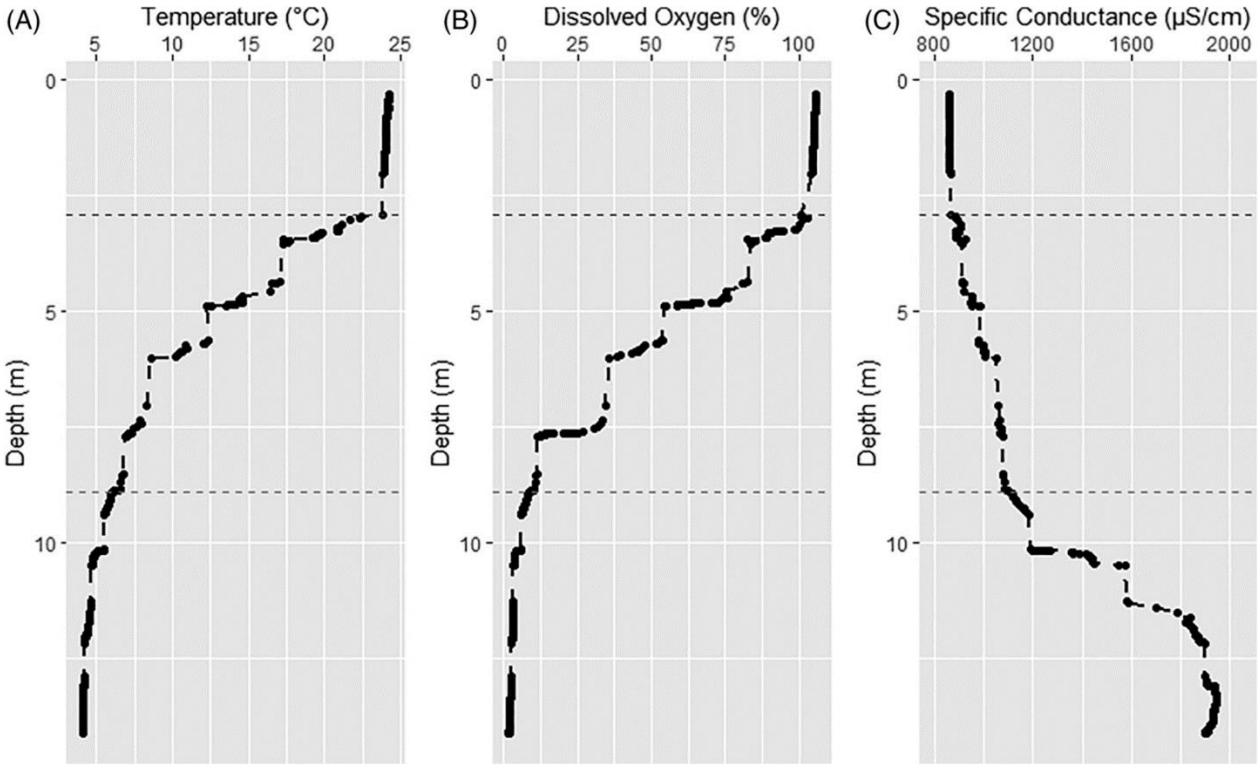


Figure 1: Lake profiles were collected by a YSI Sonde multi-parameter probe on 02 July 2019 at approximately 10:00 h above the deepest part of the lake basin. (A) Temperature (°C) profile; (B) dissolved oxygen (DO) profile (% saturation); (C) specific conductance profile (µS/cm). The thermocline occurs at approximate 2.9 m of depth. The blue dashed line indicated the Secchi disk depth recorded on 02 July 2019, DO approaches 0% at approximately 8.9 m, indicated by the red dashed line. The zone enclosed by the dashed lines indicated the "habitable" zone for large zooplankton such as *Daphnia pulicaria*. The salinity-induced chemocline occurs at approximately 10 m of depth.

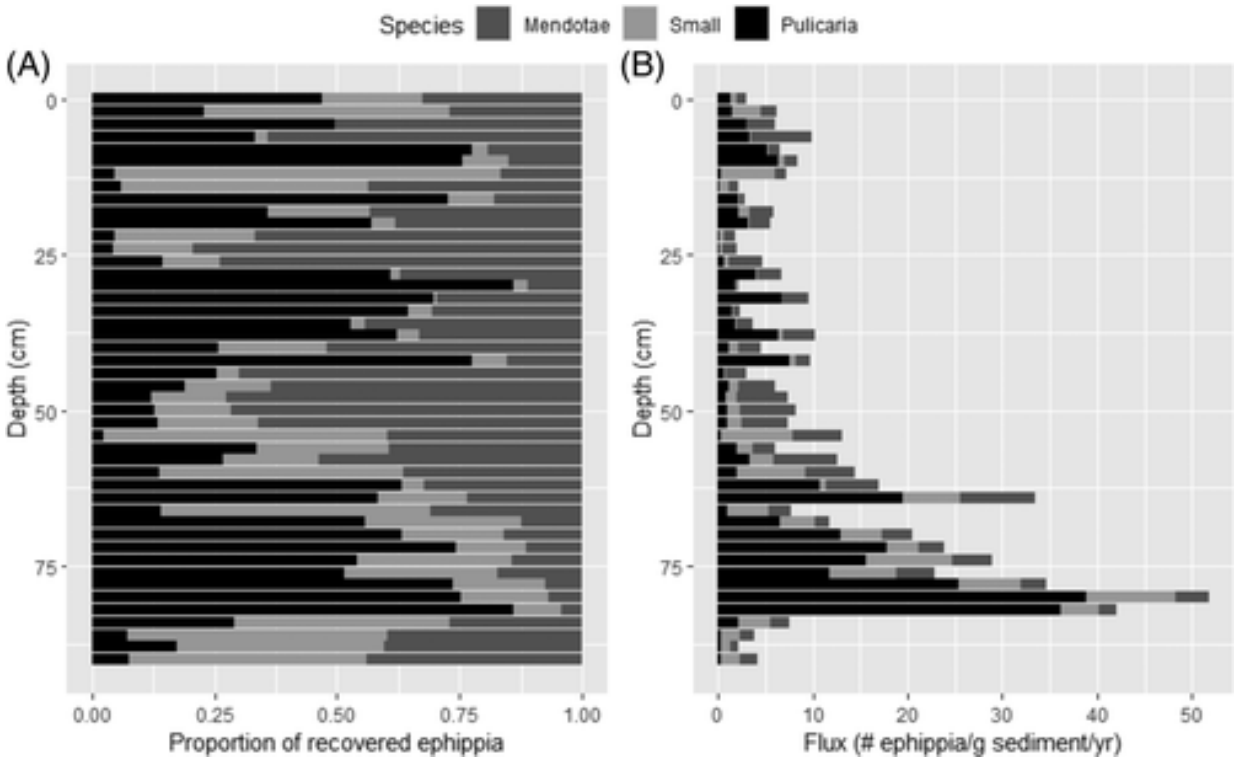


Figure 2: Relative and absolute abundances of ephippia recovered in the sediment core. (A) The relative abundance of the ephippia, as a proportion of total ephippia recovered, for each of the three species *Daphnia pulicaria* (black), *Daphnia mendotae* (dark gray), and *D*. "*small*" (light gray). (B) Absolute abundances of ephippia recovered, as total of their fluxes, for each of the three species *D. pulicaria* (black), *D. mendotae* (dark gray), and *D*. "*small*" (light gray). Peak abundance occurred at approximately 80 cm (approximately 1878).

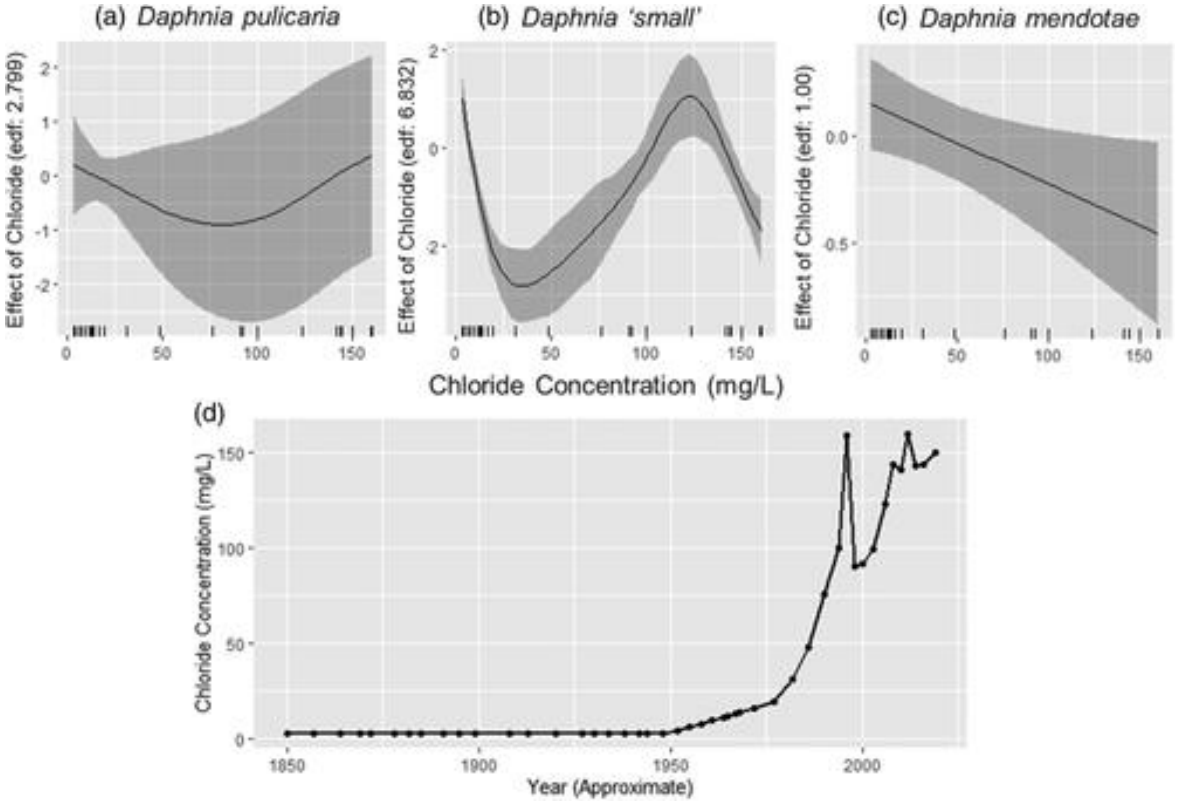


Figure 3: Estimated effects of chloride concentration on ephippial flux of the three species: (A) *Daphnia pulicaria*, (B) *Daphnia small*, (C) *Daphnia mendotae*. Rug along X-axis of A–C indicates the locations of estimated chloride measurements. (D) Estimated chloride concentration at the surface of Tanners Lake over the study period used in the GAMM model. Salinity was assumed to increase after 1950 in accordance with the beginning of widespread salt use on roadways in the United States (Findlay and Kelly 2011).

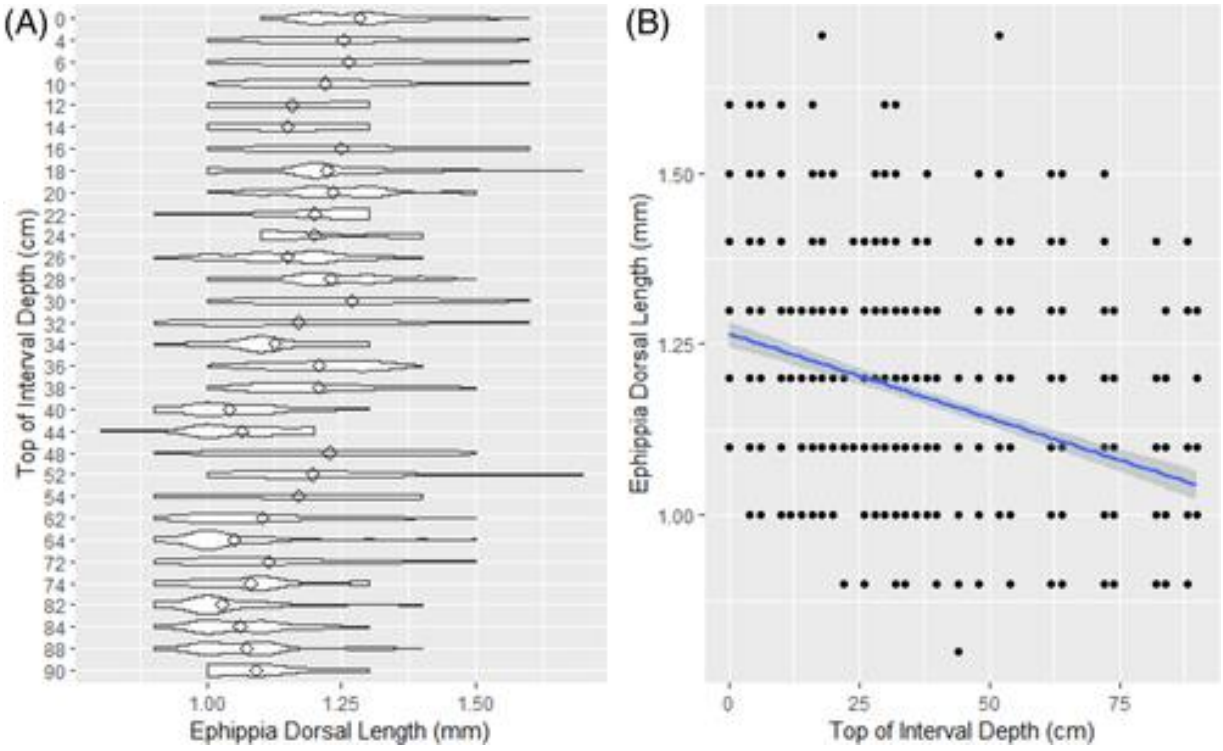


Figure 4: Dorsal length of *Daphnia pulicaria* ephippia recovered from the Tanners Lake sediment core. (A) Violin plot of ephippia length. Diamonds represent mean of distribution, the extremes of each violin represent the range of ephippia sizes, the width of each violin reflects the density of observations. (B) Raw data of the ephippia dorsal length. Trend line and 95% confidence interval for the regression slope. Length of each ephippium was measured to the nearest 0.1 mm on a stereomicroscope using an ocular micrometer (*see* "Methods" section for details).

Table 1. GAMM model results for each species (e.d.f., effective degrees of freedom).

Model	Coefficients (SE)		*t*-Value/*F*	*p*-value	Adj. R^2	e.d.f.
Daphnia pulicaria	Intercept	1.089 (0.09)	12.11	<0.0001	0.719	8.232
	s(year)	—	23.15	<0.0001		2.799
	s(chloride)	—	7.07	0.0122		2.518
	s(%-$CaC0_3$)	—	3.89	0.0306		
Daphnia mendotae	Intercept	0.979 (0.089)	11.05	<0.0001	0.363	1.00
	s(chloride)	—	9.539	0.00347		1.00
	s(%-$CaC0_3$)	—	5.328	0.02576		
Daphnia "small"	Intercept	0.394 (0.125)	3.145	0.00322	0.385	6.832
	s(chloride)	—	18.7	<0.0001		

Resurrection genomics provides molecular and phenotypic evidence of rapid adaptation to salinization in a keystone aquatic species.

Published in:

Proceedings of the National Academy of Sciences of the United States of America

02 February 2023

https://doi.org/10.1073/pnas.2217276120

Matthew J. Wersebe[1] and Lawrence J. Weider[1]

[1] Program in Ecology and Evolutionary Biology, Department of Biology, University of Oklahoma, Norman, OK

Abstract

Ecologists and evolutionary biologists are increasingly cognizant of rapid adaptation in wild populations. Rapid adaptation to anthropogenic environmental change is critical for maintaining biodiversity and ecosystems services into the future. Anthropogenic salinization of freshwater ecosystems is quickly emerging as a primary threat, which is well documented in the northern temperate ecoregion. Specifically, many northern temperate lakes have undergone extensive salinization because of urbanization and the associated increase in impervious surfaces causing runoff, and the extensive use of road deicing salts (e.g., NaCl). It remains unclear if increasing salinization will lead to extirpation of species from these systems. Using a "resurrection genomics" approach, we investigated whether the keystone aquatic herbivore, *Daphnia pulicaria*, has evolved increased salinity tolerance in a severely salinized lake located in Minnesota, USA. Whole genome resequencing of 54 *Daphnia* clones from the lake and hatched from resting eggs that represent a 25-year temporal contrast demonstrates that many regions of the genome containing genes related to osmoregulation are under selection in the study population. Tolerance assays of clones revealed that the most recent clones are more tolerant to salinity than older clones; this pattern is concomitant with the temporal pattern of stabilizing salinity in this lake. Together, our results demonstrate that keystone species such as *Daphnia* can rapidly adapt to increasing freshwater salinization. Further, our results indicate that rapid adaptation to salinity may allow lake *Daphnia* populations to persist in the face of anthropogenic

salinization maintaining the food webs and ecosystem services they support despite global environmental change.

Significance Statement

Rapid adaptation to human-induced environmental change is critical for preserving biodiversity and ecosystem services into the future. A key question is whether populations of keystone species can rapidly adapt to maintain the ecosystems they support. We investigated rapid adaptation to anthropogenic salinization in *Daphnia pulicaria*, a keystone aquatic herbivore in lake ecosystems. By resuscitating decades-old resting eggs, we investigate genomic changes across an approximately 25-year temporal contrast from a severely salinized lake. We report that the genes showing signatures of natural selection throughout the genome are related to osmoregulation and ion regulation. Phenotyping clones for salinity tolerance revealed evidence that genetic changes may underlie rapid evolution. We provide molecular genomic and phenotypic evidence for rapid adaptation to salinity in *D. pulicaria*.

Introduction

Ecologists and evolutionary biologists now recognize many examples of wild populations rapidly evolving in the face of environmental change (1, 2). A population's ability to rapidly evolve is critical for survival in the face of ever-increasing anthropogenic environmental change. This capacity is especially important for organisms that provision key ecosystem services or are keystone species because their extirpation would fundamentally alter ecosystem dynamics. Despite this, studies demonstrating a mechanistic basis for rapid adaptation that integrates information from the genome to the phenome of a population are rare (3). A key reason for the paucity of such studies is that many loci of small effect are thought to contribute most to rapid evolutionary change and do not align with classical "hard-sweep" models making their identification difficult (4, 5). Additional analytical challenges are exacerbated by different population-specific parameters (e.g., effective size or N_e) that influence the supply of new potentially beneficial mutations ($\theta = 4N_e\mu$) and the resulting distribution of fitness effects ($N_e s$) (6, 7).

One way that rapid adaptation can be studied is using temporal genomic contrasts. Most commonly, temporal contrasts take the form of so-called "evolve and re-sequence studies" which follow a population across time during experimental evolution trials that typically employ contrasting selection regimes (8–10). By finding the alleles that increase in frequency rapidly in different treatments, the molecular basis of phenotypic shifts can be explored (8). Such studies have been largely restricted to organisms such as bacteria (11) or *Drosophila* (10), which have rapid generation times and are easily manipulated in the lab. Other types of temporal contrasts encompass more natural experiments, such as the isolation and sequencing of ancient DNA,

which can give insight into past selection (12). However, with few exceptions, the genomes sampled are divorced from the phenomes they produced; thus, inference is based solely on the change in allele frequencies. A third way that temporal contrasts can be studied is through "resurrection" studies that seek to hatch or germinate seeds, cysts or other resting stages of organisms and compare genotypes and phenotypes from different points in time (13–15). For instance, resurrection ecology (16–18), commonly used in animals from the freshwater crustacean genus *Daphnia* has provided insight into the genetic basis of various traits (15, 19, 20). Thus far, however, this method has not allowed the identification of loci that can be plausibly related to the phenotype under study because genetic markers are either too sparse (21) or the traits under study are too highly integrated across the genome (19).

Daphnia are keystone species in freshwater food webs, connecting the flow of energy from algal production to higher trophic levels such as fish (22, 23). Specific to many North American lakes, *Daphnia pulicaria*, maintains water clarity, a key ecosystem service and supports recreational fisheries with values in the millions of USD per lake per annum (24). Freshwater ecosystems are among the most threatened ecosystems worldwide, impacted by various anthropogenic stressors such as pollution, climate change and invasive species (25). One issue threatening many freshwater ecosystems is salinization due to human activities (26), within northern temperate lakes specifically salinization is particularly acute (27, 28). The causes and scope of salinization have been well known (29, 30), while more recent studies have focused on deciphering the ecological impacts of salinization (31, 32). In addition, we lack a more general understanding of this widespread environmental issue from an evolutionary perspective, and of the specific genetic architecture of adaptative responses. Such a perspective is critical because

recent research has shown that current water quality guidelines do not sufficiently protect aquatic life from salinization (33).

To address this short coming, we sought to use resurrection ecology to study the evolutionary response of *D. pulicaria* from a severely salinized lake located in Minnesota, USA. Previous work on this lake has demonstrated the ecological dynamics of this population over the last 150 years (31). Towards this goal, we resurrected genotypes from across approximately 25 years from the sediment egg bank isolated from a dated sediment core. Using whole genome sequencing (WGS) of resurrected and extant individuals, we conducted numerous population genomic analyses to depict population structure over time, reconstruct the demographic history, and identify outlier genomic regions in the data. Additionally, we assayed a subset of genotypes for tolerance to salinity. Specifically, we were interested in testing two main hypotheses; first, we believed that the F_{st} outliers, or those genomic regions with extreme changes in frequency, would contain genes related to osmoregulation as selection would favor higher salinity tolerance. Second, this would be reflected in a higher mean tolerance of the more recent subpopulations. In addition, we identify a list of candidate genes, which likely influence phenotypic variation throughout the genome, and that can be targeted for further study.

Results

We studied the *D. pulicaria* population of Tanners Lake (TL; 44°57'02.2"N 92°58'54.2"W), a small suburban hardwater lake located in Oakdale, Minnesota. The watershed includes approximately 32% impervious surfaces including parking lots, interstate highways and residential development (31). TL has received significant inputs of chloride from the watershed primarily in

the form of the road deicer NaCl (524 kg Cl^- ha^{-1} yr^{-1}) (34). The upper waters (i.e., surface/epilimnetic) chloride concentration of TL has increased significantly in the last 75 years, from approximately 1-2 mg Cl^- L^{-1} on average to over 150 mg Cl^- L^{-1}(35, 36). TL has also transitioned to a state of cultural meromixis (37) with a persistent high salinity chemocline interrupting normal lake mixing dynamics. We isolated and sequenced 54 *D. pulicaria* clones including 10 from the water column and 44 resurrected from lake sediments representing an approximately 25-year temporal contrast (~1994-2019). In total we called 3,802,961 high confidence biallelic SNPs in the population.

Population Structure and Genetic Divergence:

Key to understanding the dynamics of the population across time is accurately describing population structure to rule out possible extinction and recolonization. The 54 clones selected for sequencing were separated into two clusters based on the first two principal component axes explaining 7.6% and 4.6% of the variance in the LD-pruned SNP data, respectively (Figure 1A). These two clusters largely separated the clones by depth with the older clones (layers 16-18 cm, 18-20 cm, and 22-24 cm) and more recent clones (Lake Clones, 2-4 cm, 6-8 cm, and 10-12 cm) forming the two groupings. Subsequently, we decided to assign the clones into three groups for the remainder of the analyses. We designated these as DEEP, encompassing all clones from 16-24 cm in depth (n = 18) and date from the mid to late 1990s. The second designation, called MID, encompassed all clones from 6-12 cm in depth (n = 18) and date from the mid to late 2000s. The third and final designation, referred to as TOP hereafter, included all clones collected from the water column and 2-4 cm in depth in the core (n = 18) spanning from 2016 to 2019. We based our assignment into the groupings on two observations: firstly, the TOP and DEEP subpopulations are

delineated by the PCA clusters, and thus warranted separate assignments. Secondly, the MID subpopulation, while closely related to TOP, was intermediate in the time-scale –approximately 10 years prior to TOP and approximately 10 years after DEEP– and thus formed an appropriate intermediate grouping. While the TOP and MID are indistinguishable using PCA (Fig 1A), their temporal separation is a key feature of our study. Hence, we included each as a separate temporal deme in our simulations and analysis. Conveniently, this scheme also allowed each grouping to achieve equal sample size (i.e., 18 samples). Using Discriminate Analysis of Principal Components (DAPC), which is a flexible group assignment method, we found good analytical support for assignment of clones to the three *a priori* designated groups (Figure 1B). However, as with the PCA results, generally MID clones had non-negligible assignment probabilities when compared to the TOP subpopulation. Site-wise F_{st} was low across time, with an overall estimate of approximately 0.016 (Figure 1C). However, site-wise F_{st} estimates ranged from around 0 to as high as 0.398, with most sites having an F_{st} of essentially 0. The pairwise genetic distance between the three temporal subpopulations was related to the temporal distance with the TOP vs DEEP comparison being highest (0.019) and the MID population being intermediate to both (Figure 1D). Patterns of nucleotide diversity (π) were similar across all temporal subpopulations. Mean nucleotide diversity ranged from 0.0066 for DEEP, to 0.0070 for MID and 0.0069 for TOP (Figures S3-5).

Estimation of Effective Population Size and Simulations:

To accurately parameterize tests for selection, we sought to estimate a demographic model for the TL population. We were able to estimate the effective population size of the TL *D. pulicaria* population using both linkage disequilibrium (LD) and coalescent simulations. The LD-based

results using just samples from the lake clone (LC) isolates showed that over the last few hundred generations, there was a period of population expansion and contraction (Figure S1). The population reached a peak N_e of 4000-6500 approximately 150 generations ago and the population has contracted recently to an N_e of around 2000. For this analysis, we interpret "generations" here to be sexual generations (i.e., LD is related to recombination and asexual generations are ameiotic). Since sex may occur once or at most a few times per year in stable lake habitats (38, 39), we interpret a single generation to be equivalent to one year. The coalescent-based FSC2.7 (40) run with the highest likelihood estimated N_e to be 2931 individuals and had a signal of population contraction with a population growth rate of 2.629×10^{-5}. The estimate of effective population size was within the 95% confidence for 100 parametric bootstraps; however, the estimate for population growth rate was not and the confidence intervals included zero (table S1). We ran simulations in FSC2.7 using the maximum likelihood estimates to establish expectations for F_{st} based on the modeled demographic parameters. The results from 100 independent runs of FSC2.7 were pooled to develop a distribution of expected F_{st} values from simulated ~110000 SNPs. Testing each LD-pruned SNP against this distribution and correcting for multiple testing using false discovery rate (FDR) resulted in 178 outlier SNPs with corrected one-tailed p-values above a $p = 0.05$ significance threshold (Figure 2). There were outliers on every chromosome, ranging from a low of 2 SNPs (CHR 08) to a high of 42 (CHR 04).

Genes surrounding F_{st} outliers and GO term enrichment and Variant Annotation:

One of our primary hypotheses was that the genes surrounding Fst outliers would be related to osmoregulation and salinity tolerance. We searched for genes within 10Kb of the Fst outlier SNPs (± 5 KB centered on the SNP) and in total we extracted 286 genes near these SNPs with known

function in the *D. pulicaria* genome. GO term enrichment analysis with PantherDB webtool (41) yielded 59 enriched terms for this list of genes after correction for FDR (Figure 3). The enriched terms and p-values are available in table S2. Notable among the enriched terms for molecular function are chloride channel activity (GO:0005254; p = 0.00925); however, many different ion and channel terms were enriched. After running variant effect prediction (42) on all the SNPs called in the population, we identified 17181 variants with "high" predicted effects. Intersecting this list with outlier genes, we found 78 of the 286 outlier genes had high effect variants, only one of these, Chloride Channel 2 isoform X2 (*clcn2-x2*), located on chromosome 5 had any obvious relation to osmoregulation or salinity tolerance. *Clcn2-x2* has five SNPs of high effect, including four premature stop codon changes and a splice donor change that would likely severely interrupt protein function. It is tightly linked to the F_{st} outlier site at CHR05:12238, which is an intronic SNP within *clcn2-x2*. In addition to the high effect mutations, this gene has a total of 28 missense mutations and 37 synonymous mutations classified as moderate and low impact respectively. Of the clones surveyed for chloride tolerance (see next section), we found that the two most tolerant clones were homozygous for the wildtype (i.e., functional) allele at *clcn2-x2.* Overall, 252 of the outlier genes had effects that included moderate impacts to function such as a missense SNP.

However, *clcn2-x2*'s behavior as an outlier is not consistent with a single large effect locus. The SNP at CHR05:12238 is the 88th most differentiated SNPs among the 178 outliers with an overall F_{st} of just 0.1985 (FDR corrected p = 0.0269). The top 11 outlier SNPs (F_{st} 0.3983 – 0.3175) all had F_{st} unobserved under neutral demography. One of the SNPs is an intergenic variant, while the 10 remaining SNPs are intronic SNPs associated with 13 different genes, 10 of which have annotated functions (table S3). The SNP located at CHR04:13394600 is the most differentiated

SNP observed and is associated with the genes *pank4* and *tda6.* The remaining genes span across functional categories including the structural protein collagen alpha-1 chain and the calcium-binding protein *fstl5*, to several genes related to development comprising *still life* (43), *daam1* (44), and *pnt* (45), which contains two outlier SNPs, *rap1gtp (46),* and *rotund (47).*

Chloride Tolerance LC_{50}:

We expected that increasing chloride pollution in TL would result in higher tolerance of more recent clones. Therefore, we conducted clone-specific assays to estimate 96-hour Lethal Concentration-50% (LC_{50}). Clonal tolerance ranged from a low of 584.91 mg/L Cl^- (clones 10-12-12A & 18-20-04A) to a high of 1047.16 mg/L Cl^- for (Clone LC-06). We observed a main effect of subpopulation in the Kruskal-Wallis test ($p = 0.006$; Figure 4). Post-hoc testing using a pairwise Wilcoxon test found that the Lake subpopulation (i.e., TOP) was more tolerant on average than either the DEEP (22-24 cm & 18-20 cm) or MID (10-12 cm) subpopulations ($p = 0.0076$). The DEEP and MID subpopulations did not differ from one another ($p = 0.9073$).

Discussion

Our resurrection ecology (RE) study across ~25-year temporal contrast has provided phenotypic and molecular evidence of rapid evolution of salinity tolerance in the TL *Daphnia pulicaria* population. Using whole genome resequencing to analyze population genetic structure and demographic history, we have identified genomic regions putatively under selection, with support from LC_{50} chloride tolerance assays. We found that Chloride Channel-2 isoform X2 (*clcn2-x2*) has a unique mutational history and may be affecting salinity tolerance in this population, as one of many genes under selection in this population. We find support for both of

our main hypotheses — genes related to osmoregulation are enriched in outlier regions and that the population shows increasing tolerance to salinity over time.

Demographic History and Subpopulation Structure:

A goal of our study was to establish a demographic model for the TL population. This would enable the modeling of population genetic summary statistics (i.e., Fst) for establishment of a distribution of expected or null estimates of Fst-based on the demographic history (discussed below). We used two complementary methods – LD-based methods and coalescent simulation to understand the recent demographic history of the TL population. Each method provides incomplete information, and each has unique biases that need to be considered. Furthermore, despite Ne being a critical parameter in population genetics, it is notoriously difficult to estimate (48). For instance, it appears that SFS (coalescent) based methods are underpowered for recent demographic history; however larger sample sizes may ameliorate some of these effects (49). This may also explain why it was difficult to accurately estimate the population growth rate parameter. The LD-based method we employed here appears to be strongly affected by the number of homozygotes in the sample (50), and thus, is sensitive to analyzing genomes from different generations. This prevented us from modeling each subpopulation separately. Additionally, it also appears to be inaccurate for the first few generations, giving unrealistically small numbers (see Figure S1). As such, we used these methods in tandem to increase the confidence in our estimation of the demographic parameters. Both methods were concordant that recent Ne was approximately 2000-3000 individuals, and both showed signatures of recent population contraction. This effective population size is not unexpected for D. pulicaria because populations are thought to delay sex and recombination for long periods of time (a year or more)

(38, 39). During this time, clonal selection is thought to winnow down the population to a collection of ecologically equivalent clones (51). These facts, taken together with the current understanding of zooplankton metapopulation dynamics (52, 53) suggests that Daphnia effective populations should be small and insular.

Despite the small and insular nature of the TL population, it appears that drift does not always predominate. Our PCA results demonstrate at least on a decadal scale (i.e., between MID and TOP) that the population is temporally consistent. This likely reflects the dominance of the few ecologically equivalent clones predicted by Lynch and Spitze (51). It is likely that only after a rapid or pronounced change in environment can one expect to see a population structure across time, as reflected in other studies of Daphnia utilizing RE (15, 21). Salinization in TL is ongoing and likely started in the mid-20th century with the onset of widespread use of road deicing salts in the 1950s (54). The sediment intervals dating to the early 1990s from which we recovered the oldest samples in this study reflect a period of change in the sediment egg bank (e.g., low ephippial fluxes) (31) and suggests a period of rapid environmental change in this lake. Wersebe et al. (31), reported that by the early-1990s, TL had already reached a surface (waters) chloride concentration of at least 100 mg/L. Thus, all the source periods examined in the present study are typified by elevated chloride levels. Since we were unable to hatch eggs from before salinization commenced or was comparatively low in TL, it is impossible to determine if the structuring we observed was ongoing or more sudden. It is most likely that our temporal samples represent a process of ongoing adaptation to very rapidly increasing salinization. Regardless, since the clones are closely related across time at both the genomic and mitochondrial levels, it remains unlikely that the population became extinct and was recolonized by migrant genotypes (Fig. 1C; S2).

Outliers and the Identification of Candidate Genes:

The specification of a demographic model for the TL lake population allowed us to establish the presence of statistical significance for each site in the genome scan. This analysis revealed 178 regions with Fst beyond what we could reasonably expect based on demography alone. Genes surrounding these outliers had many functions, but as expected, genes involved in osmoregulation were among the most enriched in the dataset according to GO terms for molecular function. The presence of hundreds of outlier genes suggests that salinity tolerance has a complex genetic architecture and that tolerance to increased salinity requires changes at many loci of small effect. Transcriptomic studies of clonal isolates from within the D. pulex complex (to which D. pulicaria belongs) support this assertion (55). Indeed, cross-referencing these outlier genes with different mutational types (e.g., missense SNPs) showed that many of these genes have sites that may be under selection. However, the only gene with a known functional role in osmoregulation with high effect mutations was clcn2-x2. Chloride channels play a key role in osmoregulatory physiology of all animals. Daphnia are known as hypo-osmoregulators, meaning they attempt to maintain their hemolymph solute concentration above the ambient media concentration (56). Osmoregulation occurs within the gill-epithelium of Daphnia, and chloride channels like clcn2-x2 play a major role in shuttling Cl- ions into the hemolymph across the basolateral membrane of the gill epithelium to maintain a hypo-osmotic concentration in the hemolymph (57). The constellation of premature stop codon mutations in this gene means that the protein is very unlikely to function properly because the channel is too short to pass through the cell membrane. Individuals are at least heterozygous for each of the four premature stop codon SNPs and the wild-type gene sequence annotated in the reference. Thus, they can produce a functioning protein. The two most tolerant clones in this study were both homozygous for the

functional allele. This suggests that clcn2-x2 is a critical gene requiring at least one functional copy and it explains some portion of the variance in acute salinity tolerance. However, a larger sample of genotyped and phenotyped individuals would be required to further validate this statistically.

The most differentiated genes in the genome were involved in a few key functions. There was a concentration of outlier SNPs in genes involved in regulating Rho-like GTPases which are involved in the regulation of actin filaments and neuronal development. The reason why these genes are the most differentiated is not initially clear. One potential explanation involves phenotypic plasticity. Increased salinity tolerance may require the accommodation of new developmental trajectories through phenotypic plasticity. Evolution via plasticity (58) requires that central developmental pathways serve as fuel for phenotypic differentiation (59). Furthermore, plasticity is thought to play an important role in rapid adaptation in Daphnia, a pattern we observe in the TL population (60–62). Regardless, testing of this assertion would be difficult because unlike other model arthropods, Daphnia development is not as well explored.

Chloride Tolerance and Rapid Evolution:

The clones observed in this study vary nearly two-fold in salinity tolerance (585 – 1047 mg/L), with the most tolerant clones detected in the most recent (TOP) temporal subpopulation. It is important to note that all clones in this study come from natal conditions that included elevated salinity with all subpopulations likely experiencing surface chloride conditions of 100-150 mg/L (31). Surface chloride conditions in TL are unlikely to exceed 150 mg/L in a year because of the dynamic equilibrium between annual loading and flushing of chloride in the system (63). However, TL more recently has transitioned to a state known as “cultural meromixis” were the accumulation of Cl at depth has interrupted normal lake mixing. For instance, in July 2019, we

observed inferred Cl- concentrations of approximately 275 mg/L directly above the chemocline (47% of the lowest LC50). Below the chemocline, inferred concentrations approached 483 mg/L (82% of the lowest LC50). Large-bodied D. pulicaria, undergo diel vertical migration to avoid visually oriented predators like fish. This vertical movement means that a clone may have to accommodate fluctuations in the ambient chloride concentration of 100s-of-miligrams over the course of 24-hrs. We attempted to study the degree to which this might occur in the summer of 2021 (see the supplemental materials for details). We conducted a Diel Vertical Migration (DVM) study in TL to track the spatial and temporal patterns of D. pulicaria distribution in the water column. We observed that D. pulicaria do inhabit the deepest, saltiest parts of the water column up to 12-m (Figure S6). However, in 2021, the chemical stratification of TL was much weaker than what we observed in 2019 and DO was not depleted at depth (see 31, Figure 1A-C & Figure S7 A-C). We believe the reason for the weaker chemical stratification is that the area surrounding TL (Washington and Ramesy Counties, MN) were classified as being abnormally dry on 06/29/2021 and had been classified as such intermittently since 09/29/2020. Thus, our abilities to determine the extent to which this might be reflected in a higher salinity year are diminished. Regardless, this shows that a portion of the Daphnia population is moving to more anoxic and saltier layers as part of the normal DVM behavior.

Such elevated Cl- concentrations routinely observed in TL at depth are likely to have several sublethal effects that reduce clonal fitness. Presumably, a clone that has high acute tolerance has the physiological capacity to ameliorate the sublethal effects of increased chloride and we attempted to approximate this physiological response with LC50 assays. LC50 may be a quick measure of acute tolerance and it has an uncertain correlation with fitness components. As such, our phenotyping results should be interpreted with caution when considering the potential

impact(s) on relative clonal fitness. Insofar as the pattern of rapid evolution of acute tolerance observed may be not actively reflect the phenotype actually under selection across time.

Caveats, Analytical Roadmap, and Conclusions:

All studies integrating historical data require caution to avoid over-interpretation of the emergent patterns. Our study is no different and relies heavily on a single core from a single lake. Our previous studies with this lake have explored many of these caveats (31). Specific to resurrection ecology (RE) studies, an acute limitation has always been hatching from the egg bank in the deepest layers. There may be non-random patterns of egg mortality in the sediments or non-random propensities in hatching success that skew phenotypic estimates and prevent accurate estimation of allele frequencies. Some have termed this the "invisible fraction" (sensu (64)). In our study this is most notable in our inability to hatch genotypes predating the most substantial increases in salinity (e.g., from 1950 or before). In other resurrection studies that have been able to hatch truly 'ancient' eggs (15), the utility of these samples in constructing a framework of phenotypic evolution is limited by sample size because only one or a few isolates survive to be cultured. In many circumstances, however, directly sequencing the eggs is an alternative but will prevent any possible phenotypic characterization of sequenced individuals because eggs are destructively sampled (65). These latter methods are still nonetheless difficult and involve costly genome amplification steps with variable success rates and high levels of exogenous contamination in the final libraries (65). This approach may help reduce the impacts of the 'invisible fraction' but will not eliminate it completely. Another caveat that we must highlight is that we have only analyzed a single population, and as such, we cannot place our results within a metapopulation context. We assume that gene flow should be negligible in producing allele frequency changes. To accurately account for gene flow, one would need to sample many

additional spatially-distributed populations in addition to hatching temporal samples. While such data sets are within technical reach, the practicality of such a sampling regime would be difficult to amass for species such as D. pulicaria.

Regardless of the issues with RE data sets, we believe we can provide some insights into an analytical framework for future resurrection genomic studies. One major goal of the field of ecological and evolutionary genomics is to ultimately produce a comprehensive phenotype-to-genotype map- especially for those traits that one deems "ecologically relevant." Relevant critiques of a QTL or QTN-centric research program aside (66), temporally-sampled genomic datasets may provide some very convincing examples of phenotype-genotype maps (see (67). Resurrection-type studies are poised to do this in natural populations as well. This will require robust demographic analysis of resurrected populations, integrated with simulations in a flexible manner. One potential way forward is the use of flexible-forward genetic simulations (e.g., those available in SLiM; (68)), which can allow more analytical power in genomic analysis of temporally sampled populations.

In summary, resurrection ecology (RE), when paired with whole genome sequencing, provides a unique and powerful way to study rapid evolution of populations in situ. Integration of candidate loci identified with RE into study designs including breeding experiments (69), forward mutation screens (70) or CRISPR technology (71) will provide unrivaled insight for the genetic architecture of complex, ecologically relevant traits in the wild. Such an approach would be beneficial here, for example in validating the effect of the null clcn2-x2 allele on salinity tolerance in the TL population. Overall, however, here RE provides both molecular and phenotypic evidence of rapid adaptation in the TL population. Thus, the persistence of D. pulicaria in this severely salinized lake is likely the result of this rapid adaptation. Our findings

indicate that keystone aquatic species such as lake Daphnia may continue to thrive in lakes that exceed current water quality limits for salinization and thus maintain the stability of the food webs and ecosystem services they support.

Materials and Methods

Clone bank and Sequencing:

On 2 July 2019, we collected duplicate sediment cores in TL from a 14 m deep station following Wright (72). During the same period, we also collected D. pulicaria from the active plankton community using several vertical tows of a Wisconsin net at the core sampling station. Animals were isolated as single individuals in 125 mL plastic (screw-capped) cups in COMBO media (73). A total of 10 of these clones were established in laboratory culture. Resting eggs (ephippia) collected from throughout the cores were collected according to the methods outlined in Wersebe et al. (31). D. pulicaria are cyclically parthenogenetic, meaning they produce clonal offspring during the growing season and may occasionally engage in sex to produce resting eggs encased in durable structures called ephippia (74). Ephippia identified as D. pulicaria were subjected to hatching protocol described (15) (see supplemental). These hatchling individuals were expanded in culture to establish upwards of 10 clones per sediment layer to establish a clone bank. Fifty-four isoclonal lineages from the clone bank were selected for DNA extraction and whole genome sequencing (average 10X) on an Illumina NovaSeq by the Oklahoma Medical Research Foundation.

Bioinformatics:

Raw sequencing reads were quality trimmed and adaptor contamination removed using Trimmomatic (75). Quality trimmed reads were aligned to the chromosome-level D. pulicaria

genome assembly (76) using the BWA mem algorithm (63). The resulting files were piped through samtools (78) to mark duplicates, fix mates and sort the bam files. We called variants using the bcftools mpileup and call pipeline using all individuals together (79). Using bcftools, the resulting BCF files were concatenated together into a single genome-wide file and quality filtered to a set of high confidence biallelic single nucleotide polymorphisms (SNPs).

Population Structure and Genetic Divergence:

The variants in the final quality-filtered VCF file were pruned for linkage disequilibrium (LD) using Plink (80) independent pairwise function (settings: 50, 10, 0.1), providing an independent and essentially random set of SNPs. The resulting variants were further filtered to SNPs with 0% missingness to a set 27854 genome-wide SNPs. We conducted Principal Components Analysis (PCA) in R (Version 4.2.0) using the packages adegenet and vcfr (81 - 83). We also conducted Discriminant Analysis of Principal Components (DAPC); a flexible population assignment method also implemented in adegenet (84). Using the cross-validation procedure outlined in the vignette, we found that retention of 10 PCs performed best in population assignment. We performed population assignment tests for each clone retaining 10 PCs and 2 discriminant functions and plotted the results as a bar plot to visual the probabilities of population assignment. DAPC is flexible enough to handle mixed clonal and temporal sampling - two factors that violate other assignment techniques (e.g., STRUCTURE (85)). From the PCA and the DAPC, results, we determined that the samples could be assigned to three “sub-populations” according to the depth of their recovery (see results). Using these three subpopulations as designations, we estimated overall site-wise Fst using the basic.stats function and mean pairwise genetic distance genet.dist function in the R package hierfstat (86). In addition, we estimated nucleotide diversity

(π) in 10 kb windows throughout the genome for each of the temporal subpopulations using the program PIXY (87).

Estimation of Effective Population Size and Simulations:

To parameterize our tests for selection, we sought to identify the recent demographic history of the TL D. pulicaria population. To accomplish this, we used two different methods to estimate effective population size (Ne) and growth trajectory of the population. The first method, implemented in GONE (88) uses LD to estimate the recent population history. This method is robust to non-equilibrium histories such as selection (89); however, it is suitable only for sample pools collected from the same generation. Thus, for this method we used only the samples collected from the water column in 2019. We ran six independent runs of this method using random subsets of 600,000 SNPs from all the SNPs called in the population and a constant recombination rate of 7.2 cm/mb estimated from the D. pulicaria genetic map (76). This analysis does not assume a given model a priori, instead it produces a population size trajectory that when inspected graphically can hint at different events (e.g., bottlenecks). In addition to LD-based methods, we estimated demographic parameters of the TL population - estimating Ne and the population growth rate - by fitting a demographic model to the folded site frequency spectrum (SFS) implemented in FastSimcoal 2.7 (FSC2.7) (40). With the LD pruned SNPs used above in the PCA analysis, we created folded 2-D SFS from the three temporal subpopulations. Using 100 independent runs of FSC2.7, we chose the run that maximized the likelihood of the observed data. We fixed the sampling points in time for the temporal subpopulation (MID and DEEP) to be 50 and 100 generations in the past. This assumes approximately 4 asexual generations a year and a single sexual generation for a total of five generations a year. Each run conducted 1-million coalescent simulations and used 40 brent maximization cycles. Further, we estimated

empirical p-values for the site-wise Fst estimates using simulation (90). We chose the best fitting demographic model estimated in FSC2.7 and conducted 100 separate runs of this model to simulate approximately 1100 SNPs for each run. For each run, we estimated Fst using basic.stats function in hierfstat (86). We pooled each simulation into an empirical distribution of probable Fst values under the best fitting demographic model. We tested each observed site-wise Fst estimate against this empirical distribution to estimate a p-value. We corrected these p-values for multiple testing using a false discovery rate in R using the p.adjust function.

Genes surrounding Fst outliers, GO term enrichment, and Variant Effects:

We extracted the genes surrounding Fst outliers ($p \leq 0.05$ after correction) in 10 Kb windows using the D. pulicaria RefSeq annotation (release 100, SC_F0-13Bv2) using bedtools (91). Next, we created Panther Generic Mappings for the genes with known annotations using each gene's protein sequences following the method outlined in (41). Using the generic mappings, we tested for molecular function Gene Ontology (GO) term enrichment by testing against the Daphnia pulex gene list using the PantherDB webtool using a false discovery rate correction (41). In addition to testing for GO term enrichment, we also surveyed all genes for potential SNPs and small indels mutations potentially driving selection using Ensembl's Variant Effect Predictor (VEP) (42). We built a custom database using the RefSeq annotations in GFF format following the developer's protocol. We extracted all "High" (e.g., premature stops) and "Moderate" (e.g., missense SNPs) impact mutations predicted and cross-referenced these with the genes near Fst outliers.

Chloride Tolerance:

We selected a subset of 30 clones from the Lake (n = 10), 10-12 cm sediment layer (n =10), 18-20 cm (n = 3), and 22-24 cm (n = 7) subpopulations for estimating clone-specific tolerance to chloride. See the Supplemental Materials for details on the experimental set-up. We estimated LC50 for each clone separately by fitting a reduced-bias generalized linear model (GLM) to the survival curve using the R package brglm (92). Data were not normally distributed, nor did they meet the assumption of equal variances, so we performed a non-parametric ANOVA (Kruskal-Wallis) test on the LC50 estimates to test for differences in the mean LC50 value for each subpopulation

Acknowledgments

We would like to thank M. Edlund who facilitated the collection of the Tanners Lake sediment cores. T. Curb, Z. Arnold, A. Twumasi-Mensah, and A. Hemani, helped with the LC_{50} experiments. G. Wiley aided in the preparation of sequencing libraries. The University of Oklahoma Biological Station and the St. Croix Watershed Research Station facilitated field work. Funding for this study was provided by the University of Oklahoma Department of Biology Adams Summer Scholarship, Robberson Graduate College Grant, Hill Fund for Research in Biology, Graduate Student Senate Research Grant, AMNH Theodore Roosevelt Grant and the Biogeography of Behavior student seed grant (NSF DBI-2021880; PI L. Stein) awarded to MJW in support of graduate research. 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, the American Museum of Natural History or the University of Oklahoma. This manuscript represents a portion of MJW's doctoral dissertation at The University of Oklahoma. We thank G. Robinson and three anonymous reviewers for constructive comments on earlier versions of the manuscript.

Data Availability:

Sequencing reads are available on the NCBI SRA under BioProject PRJNA925094. Data and appropriate metadata from are archived on Harvard Dataverse (https://doi.org/10.7910/DVN/IGSXBN) (93). Code used in the analysis is available on GitHub at https://github.com/mwersebe/Tanner_lake_genomics.

References:

1. S. M. Rudman, et al., What genomic data can reveal about eco-evolutionary dynamics. Nature Ecology & Evolution 2, 9–15 (2018).
2. N. G. Hairston, et al., Lake ecosystems: Rapid evolution revealed by dormant eggs. Nature 401, 446–446 (1999).
3. F. Mallard, V. Nolte, R. Tobler, M. Kapun, C. Schlötterer, A simple genetic basis of adaptation to a novel thermal environment results in complex metabolic rewiring in Drosophila. Genome Biology 19, 119 (2018).
4. P. W. Messer, D. A. Petrov, Population genomics of rapid adaptation by soft selective sweeps. Trends in Ecology and Evolution 28, 659–669 (2013).
5. P. W. Messer, S. P. Ellner, N. G. Hairston, Can Population Genetics Adapt to Rapid Evolution? Trends in genetics : TIG 32, 408–418 (2016).
6. A. Eyre-Walker, P. D. Keightley, The distribution of fitness effects of new mutations. Nature Reviews Genetics 8:8 8, 610–618 (2007).
7. P. D. Keightley, A. Eyre-Walker, What can we learn about the distribution of fitness effects of new mutations from DNA sequence data? Philosophical Transactions of the Royal Society B: Biological Sciences 365, 1187–1193 (2010).
8. A. Long, G. Liti, A. Luptak, O. Tenaillon, Elucidating the molecular architecture of adaptation via evolve and resequence experiments. Nature Reviews Genetics 16, 567–582 (2015).
9. C. Schlötterer, R. Kofler, E. Versace, R. Tobler, S. U. Franssen, Combining experimental evolution with next-generation sequencing: A powerful tool to study adaptation from standing genetic variation. Heredity 114, 431–440 (2015).
10. T. L. Turner, A. D. Stewart, A. T. Fields, W. R. Rice, A. M. Tarone, Population-Based Resequencing of Experimentally Evolved Populations Reveals the Genetic Basis of Body Size Variation in Drosophila melanogaster. PLoS Genetics 7, e1001336 (2011).
11. R. E. Lenski, Convergence and divergence in a long-term experiment with bacteria. American Naturalist 190, S57–S68 (2017).
12. M. Dehasque, et al., Inference of natural selection from ancient DNA. Evolution Letters 4, 94–108 (2020).
13. L. J. Weider, P. D. Jeyasingh, D. Frisch, Evolutionary aspects of resurrection ecology: Progress, scope, and applications-An overview. Evolutionary Applications 11, 3–10 (2018).
14. M. Gros-Balthazard, et al., The genomes of ancient date palms germinated from 2,000 y old seeds. Proceedings of the National Academy of Sciences of the United States of America 118 (2021).
15. D. Frisch, et al., A millennial-scale chronicle of evolutionary responses to cultural eutrophication in Daphnia. Ecology Letters 17, 360–368 (2014).
16. W. C. Kerfoot, J. A. Robbins, L. J. Weider, A new approach to historical reconstruction: Combining descriptive and experimental paleolimnology. Limnology and Oceanography 44, 1232–1247 (1999).
17. W. C. Kerfoot, L. J. Weider, "Experimental paleoecology (resurrection ecology): Chasing Van Valen's Red Queen hypothesis" *Limnology and Oceanography*, 4, 2. (2004).

18. L. Orsini, et al., Temporal genetic stability in natural populations of the waterflea Daphnia magna in response to strong selection pressure. Molecular Ecology 25, 6024–6038 (2016).
19. A. Chaturvedi, et al., Extensive standing genetic variation from a small number of founders enables rapid adaptation in Daphnia. Nature Communications 12, 4306 (2021).
20. C. Cousyn, et al., Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. Proceedings of the National Academy of Sciences of the United States of America 98, 6256–60 (2001).
21. D. Frisch, et al., Paleogenetic records of Daphnia pulicaria in two North American lakes reveal the impact of cultural eutrophication. Global Change Biology 23, 708–718 (2017).
22. W. Lampert, U. Sommer, Limnoecology: The Ecology of Lakes and Streams (Oxford University Press, 2007).
23. Ž. Ogorelec, C. Wunsch, A. J. Kunzmann, P. Octorina, J. I. Navarro, Large daphniids are keystone species that link fish predation and phytoplankton in trophic cascades. Fundamental and Applied Limnology 194, 297–309 (2021).
24. J. R. Walsh, S. R. Carpenter, M. J. Vander Zanden, Invasive species triggers a massive loss of ecosystem services through a trophic cascade. Proceedings of the National Academy of Sciences of the United States of America 113, 4081–5 (2016).
25. A. J. Reid, et al., Emerging threats and persistent conservation challenges for freshwater biodiversity. Biological Reviews 94, 849–873 (2019).
26. S. S. Kaushal, et al., Freshwater salinization syndrome on a continental scale. Proceedings of the National Academy of Sciences (2018) https:/doi.org/10.1073/pnas.1711234115.
27. H. A. Dugan, et al., Lakes at Risk of Chloride Contamination. Environmental Science and Technology 54, 6639–6650 (2020).
28. H. A. Dugan, et al., Salting our freshwater lakes. Proceedings of the National Academy of Sciences of the United States of America 114, 4453–4458 (2017).
29. W. D. Hintz, R. A. Relyea, A review of the species, community, and ecosystem impacts of road salt salinisation in fresh waters. Freshwater Biology 64, 1081–1097 (2019).
30. D. Cunillera-Montcusí, et al., Freshwater salinisation: a research agenda for a saltier world. Trends in Ecology & Evolution 0 (2022).
31. M. J. Wersebe, M. B. Edlund, L. J. Weider, Does salinization impact long-term Daphnia assemblage dynamics? Evidence from the sediment egg bank in a small hard-water lake. Limnology and Oceanography Letters n/a (2021).
32. S. Arnott, et al., Road salt impacts freshwater zooplankton at concentrations below current water quality guidelines. Environmental Science & Technology (2020) https:/doi.org/10.1021/acs.est.0c02396 (August 11, 2020).
33. W. D. Hintz, et al., Current water quality guidelines across North America and Europe do not protect lakes from salinization. Proceedings of the National Academy of Sciences 119, e2115033119 (2022).
34. E. V. Novotny, D. Murphy, H. G. Stefan, Increase of urban lake salinity by road deicing salt. Science of The Total Environment 406, 131–144 (2008).

35. J. M. Ramstack, S. C. Fritz, D. R. Engstrom, Twentieth century water quality trends in Minnesota lakes compared with presettlement variability. Canadian Journal of Fisheries and Aquatic Sciences 61, 561–576 (2004).
36. J. M. Ramstack, S. C. Fritz, D. R. Engstrom, S. A. Heiskary, The Application of a Diatom-based Transfer Function to Evaluate Regional Water-Quality Trends in Minnessota Since 1970. Journal of Paleolimnology 29, 79–94 (2003).
37. R. J. Sibert, C. M. Koretsky, D. A. Wyman, Cultural meromixis: Effects of road salt on the chemical stratification of an urban kettle lake. Chemical Geology 395, 126–137 (2015).
38. D. E. Allen, M. Lynch, The effect of variable frequency of sexual reproduction on the genetic structure of natural populations of a cyclical parthenogen. Evolution 66, 919–926 (2012).
39. C. E. Cáceres, A. J. Tessier, Incidence of diapause varies among populations of Daphnia pulicaria. Oecologia 141, 425–431 (2004).
40. L. Excoffier, et al., fastsimcoal2: demographic inference under complex evolutionary scenarios. Bioinformatics 37, 4882–4885 (2021).
41. H. Mi, et al., PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Research 49, D394–D403 (2021).
42. W. McLaren, et al., The Ensembl Variant Effect Predictor. Genome Biology 17, 122 (2016).
43. S. Masaki, et al., Still life, a Protein in Synaptic Terminals of Drosophila Homologous to GDP-GTP Exchangers. Science 275, 543–547 (1997).
44. P. Aspenström, N. Richnau, A.-S. Johansson, The diaphanous-related formin DAAM1 collaborates with the Rho GTPases RhoA and Cdc42, CIP4 and Src in regulating cell morphogenesis and actin dynamics. Experimental Cell Research 312, 2180–2194 (2006).
45. C. Klambt, The Drosophila gene pointed encodes two ETS-like proteins which are involved in the development of the midline glial cells. Development 117, 163–176 (1993).
46. C. Fangli, B. Margaret, R. K. T., Q. Adrian, H. I. K., Biological characterization of Drosophila Rapgap1, a GTPase activating protein for Rap1. Proceedings of the National Academy of Sciences 94, 12485–12490 (1997).
47. S. E. St Pierre, M. I. Galindo, J. P. Couso, S. Thor, Control of Drosophila imaginal disc development by rotund and roughened eye: differentially expressed transcripts of the same gene encoding functionally distinct zinc finger proteins. Development 129, 1273–1281 (2002).
48. J. Wang, E. Santiago, A. Caballero, Prediction and estimation of effective population size. Heredity (Edinb). **117**, 193–206 (2016).
49. J. D. Robinson, A. J. Coffman, M. J. Hickerson, R. N. Gutenkunst, Sampling strategies for frequency spectrum-based population genomic inference. BMC Evol. Biol. **14**, 254 (2014).
50. E. Santiago, et al., Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium. Mol. Biol. Evol. **37**, 3642–3653 (2020).
51. M. Lynch, K. Spitze, "Evolutionary Genetic of Daphnia" in Ecological Genetics, L. Real, Ed. (Princeton University Press, 1994).

52. L. De Meester, A. Gómez, B. Okamura, K. Schwenk, The Monopolization Hypothesis and the dispersal-gene flow paradox in aquatic organisms. Acta Oecologica **23**, 121–135 (2002).
53. L. De Meester, J. Vanoverbeke, K. De Gelas, R. Ortells, P. Spaak, Genetic structure of cyclic parthenogenetic zooplankton populations a conceptual framework. Arch. für Hydrobiol., 217–244 (2006).
54. S. E. G. Findlay, V. R. Kelly, Emerging indirect and long-term road salt effects on ecosystems. Ann. N. Y. Acad. Sci. **1223**, 58–68 (2011).
55. L. C. Latta, L. J. Weider, J. K. Colbourne, M. E. Pfrender, The evolution of salinity tolerance in Daphnia: a functional genomics approach. Ecol. Lett. **15**, 794–802 (2012).
56. E. S. Chang, M. Thiel, J. Hervé Lignot, G. Charmantier, "Chapter 8: Osmoregulation and Excretion" *in Natural History of Crustacea Volume IV: Physiology*, E. S. Chang, M. Thiel, Eds. (Oxford University Press, 2015).
57. M. B. Griffith, Toxicological perspective on the osmoregulation and ionoregulation physiology of major ions by freshwater animals: Teleost fish, crustacea, aquatic insects, and Mollusca. Environ. Toxicol. Chem. **36**, 576–600 (2017).
58. N. A. Levis, D. W. Pfennig, Evaluating "Plasticity-First" Evolution in Nature: Key Criteria and Empirical Approaches. Trends Ecol. Evol. **31**, 563–574 (2016).
59. M. J. West-Eberhard, Developmental plasticity and the origin of species differences. Proc. Natl. Acad. Sci. **102**, 6543–6549 (2005).
60. K. I. Brans, L. De Meester, City life on fast lanes: Urbanization induces an evolutionary shift towards a faster lifestyle in the water flea Daphnia. Funct. Ecol. **32**, 2225–2240 (2018).
61. N. G. Hairston, et al., Natural selection for grazer resistance to toxic cyanobacteria: Evolution of phenotypic plasticity? Evolution (N. Y). **55**, 2203–2214 (2001).
62. A. Scoville, M. Pfrender, Phenotypic plasticity facilitates recurrent rapid adaptation to introduced predators. Proc. Natl. Acad. Sci. **107**, 4260–4263 (2010).
63. E. V. Novotny, H. G. Stefan, Projections of Chloride Concentrations in Urban Lakes Receiving Road De-icing Salt. *Water, Air, & Soil Pollution* **211**, 261–271 (2010).
64. A. E. Weis, Detecting the "invisible fraction" bias in resurrection experiments. Evolutionary Applications 11, 88–95 (2018).
65. J. B. Lack, L. J. Weider, P. D. Jeyasingh, Whole genome amplification and sequencing of a Daphnia resting egg. Molecular Ecology Resources 18, 118–127 (2018).
66. M. V. Rockman, The qtn program and the alleles that matter for evolution: all that's gold does not glitter. Evolution 66, 1–17 (2012).
67. N. O. Therkildsen, et al., Contrasting genomic shifts underlie parallel phenotypic evolution in response to fishing. Science (New York, N.Y.) 365, 487–490 (2019).
68. B. C. Haller, P. W. Messer, SLiM 3: Forward Genetic Simulations Beyond the Wright-Fisher Model. Molecular Biology and Evolution 36, 632–637 (2019).
69. R. E. Sherman, R. Hartnett, E. L. Kiehnau, L. J. Weider, P. D. Jeyasingh, Quantitative genetics of phosphorus content in the freshwater herbivore, Daphnia pulicaria. Journal of Animal Ecology 00, 1365-2656.13419 (2021).

70. M. Snyman, T. V. Huynh, M. T. Smith, S. Xu, The genome-wide rate and spectrum of EMS-induced heritable mutations in the microcrustacean Daphnia: on the prospect of forward genetics. Heredity 127, 535–545 (2021).
71. C. Hiruta, K. Kakui, K. E. Tollefsen, T. Iguchi, Targeted gene disruption by use of CRISPR/Cas9 ribonucleoprotein complexes in the water flea Daphnia pulex. Genes to Cells 23, 494–502 (2018).
72. H. E. Wright, Coring tips. Journal of Paleolimnology 6, 37–49 (1991).
73. S. S. Kilham, D. A. Kreeger, S. G. Lynn, C. E. Goulden, L. Herrera, COMBO: a defined freshwater culture medium for algae and zooplankton. Hydrobiologia 377, 147–159 (1998).
74. W. Lampert, Daphnia: Development of a Model Organism in Ecology and Evolution (International Ecology Institute (ECI), 2011).
75. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014).
76. M. J. Wersebe, R. E. Sherman, P. D. Jeyasingh, L. J. Weider, The roles of recombination and selection in shaping genomic divergence in an incipient ecological species complex. Molecular Ecology n/a (2022).
77. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).
78. H. Li, et al., The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
79. P. Danecek, et al., Twelve years of SAMtools and BCFtools. GigaScience 10, 1–4 (2021).
80. C. C. Chang, et al., Second-generation PLINK: Rising to the challenge of larger and richer datasets. GigaScience 4, 7 (2015).
81. R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2022). https://www.R-project.org/
82. T. Jombart, adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405 (2008).
83. B. J. Knaus, N. J. Grünwald, vcfr: a package to manipulate and visualize variant call format data in R. Molecular Ecology Resources 17, 44–53 (2017).
84. T. Jombart, S. Devillard, F. Balloux, Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC Genetics 11, 94 (2010).
85. T. A. Joseph, I. Pe'er, Inference of Population Structure from Time-Series Genotype Data. American Journal of Human Genetics 105, 317–333 (2019).
86. J. Goudet, hierfstat, a package for r to compute and test hierarchical F-statistics. Molecular Ecology Notes 5, 184–186 (2005).
87. K. L. Korunes, K. Samuk, pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. Molecular Ecology Resources 21, 1359–1368 (2021).
88. E. Santiago, et al., Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium. Molecular Biology and Evolution 37, 3642–3653 (2020).

89. I. Novo, E. Santiago, A. Caballero, The estimates of effective population size based on linkage disequilibrium are virtually unaffected by natural selection. PLOS Genetics 18, e1009764 (2022).
90. K. E. Lotterhos, M. C. Whitlock, The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Molecular Ecology 24, 1031–1046 (2015).
91. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).
92. I. Kosmidis, E. C. Kenne Pagui, N. Sartori, Mean and median bias reduction in generalized linear models. Statistics and Computing 30, 43–59 (2020).
93. M. Wersebe, L. Weider, Replication Data for: Resurrection genomics provides molecular and phenotypic evidence of rapid adaptation to salinization in a keystone aquatic species. Harvard Dataverse. V1. https://doi.org/10.7910/DVN/IGSXBN

Figures & Tables:

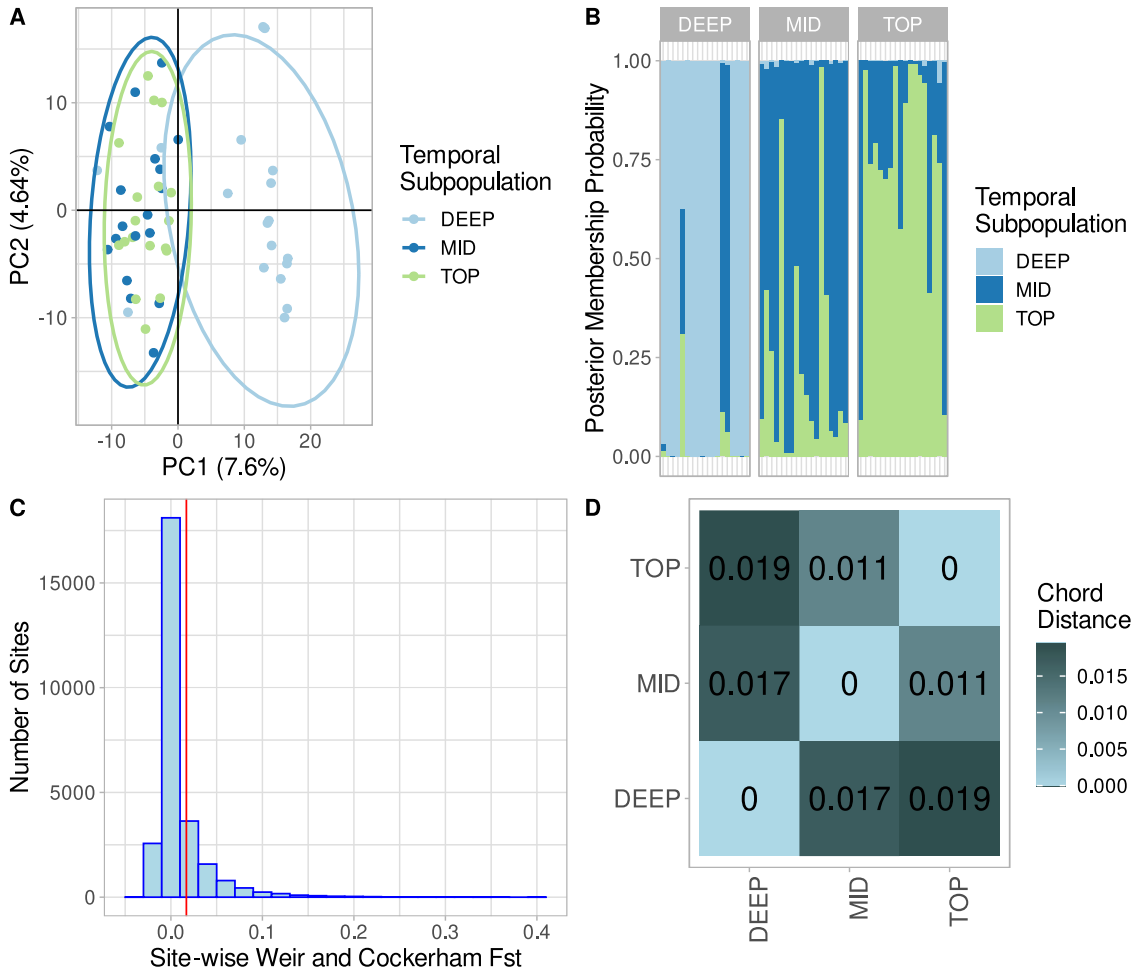


Figure 1: Genetic structure, divergence, and distance across time in Tanners Lake. A) Principal Component Analysis (PCA) biplot of all clones color-coded according to the depth of recovery in the core (cm). PC1 and PC2 explain 7.6 % and 4.6 % of variance observed in the SNP data, respectively. Older clones (22-24 cm, 18-20 cm, & 16-18 cm) form a cluster distinct from more recent clones (LC, 2-4 cm, 6-8 cm and 10-12 cm). B) Discriminant Analysis of Principal Components (DAPC) bar plot showing the posterior probability of assignment to the three *a priori* defined temporal subpopulations. C) Observed Weir and Cockerham site-wise F_{st} estimates. Overall F_{st} was low (red vertical line) at only 0.0169. D) Pair-wise genetic chord distances between the three temporal subpopulations (TOP, MID and DEEP)

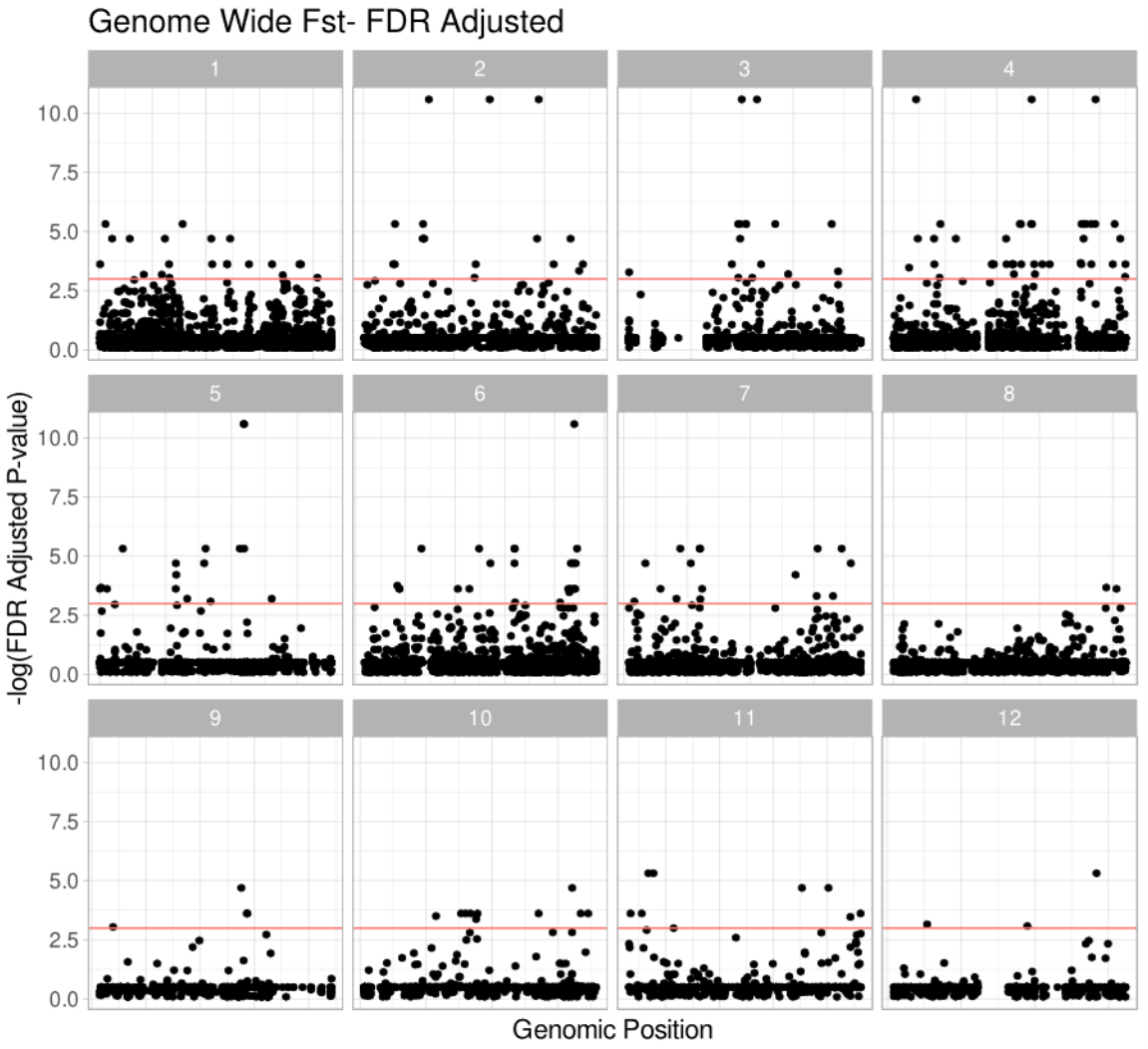


Figure 2: Genome-wide F_{st} Manhattan plots. Each panel represents a chromosome (1-12), each point designates a Single Nucleotide Polymorphism (SNP), the horizontal line indicates genome-wide significance at p = 0.05. Points above the red line are statistical outliers. Chromosome lengths are normalized across panels.

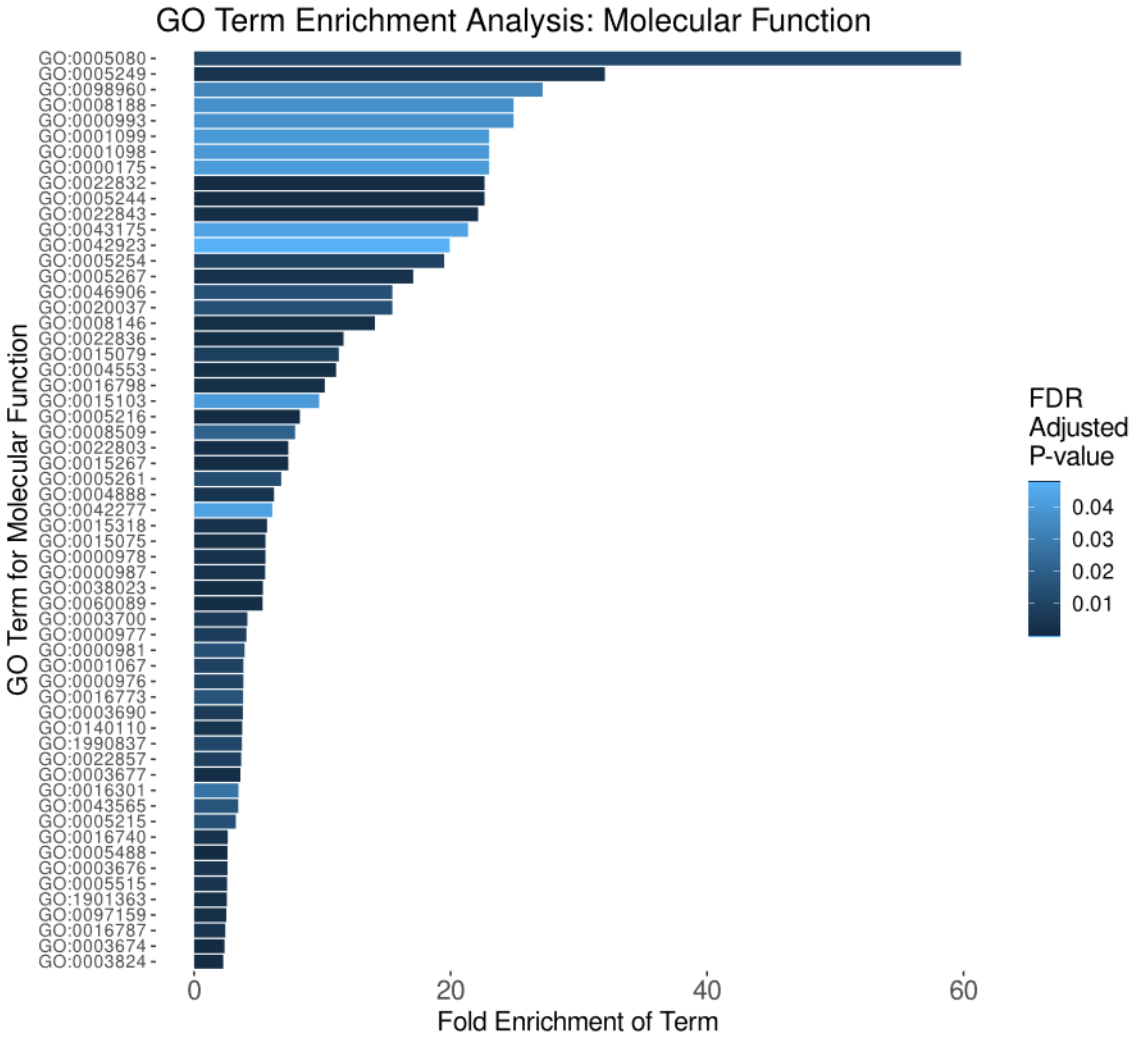


Figure 3: Enriched Gene Ontology (GO) terms for molecular function. Enriched terms are set along the Y-axis. The length of each bar indicates the term's fold enrichment in the analyzed gene list, and the bar color denotes its FDR corrected p-value. In total, 59 terms were enriched among the 286 genes found near outlier SNPs. GO term mappings are provided in table S2.

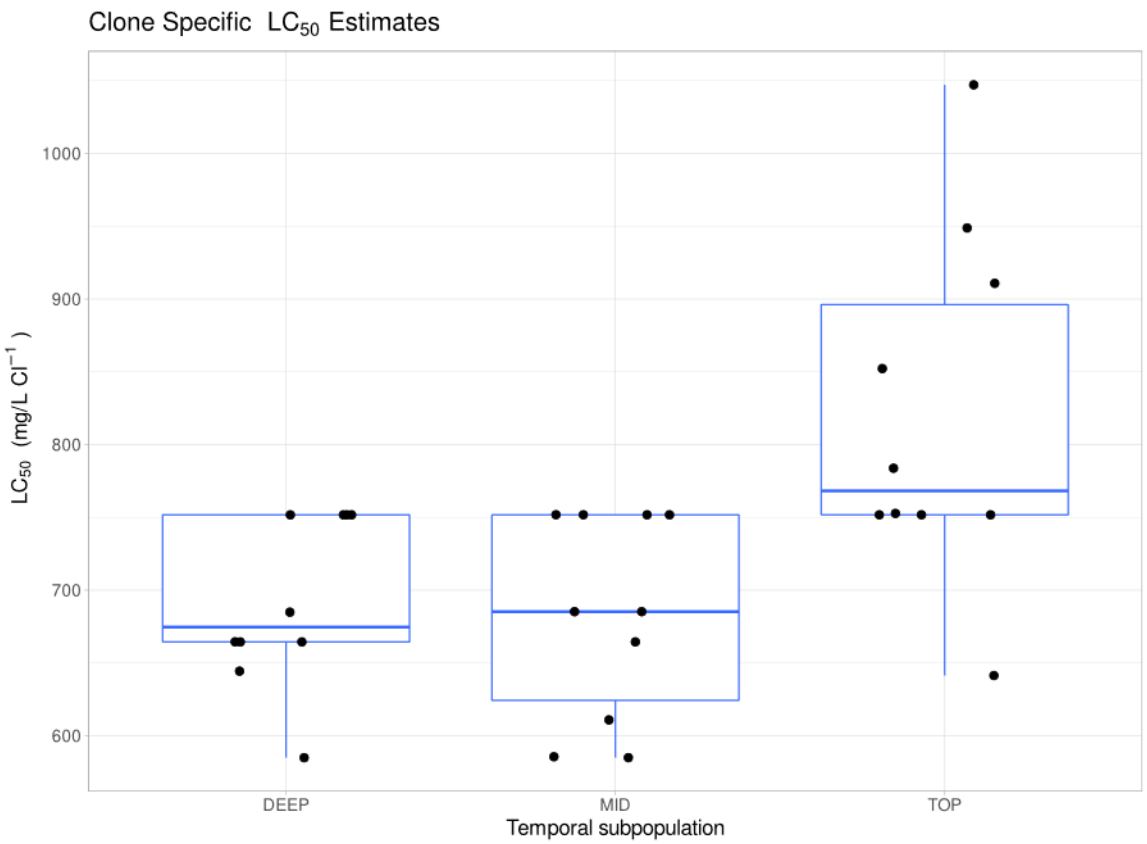


Figure 4: 96-hour lethal concentration-50% assay (LC_{50}) estimates for clones evaluated in the salinity tolerance experiments. Box plots indicate population medians and variances, each point is a clone. Overall, non-parametric ANOVA (Kruskal-Wallis) indicated a significant main effect of population. Post-hoc testing (pairwise Wilcox) confirmed that the TOP population was more tolerant than either the MID or DEEP populations, which were not different from one another.

Synthesis:

Biological archives are incredible sources of information that we can study to better understand ecosystems and species from the past to hopefully predict their futures. In this dissertation, I have presented three chapters that scoured two different types of biological archives to understand the lake daphniid, *Daphnia pulicaria.* In the first chapter, I constructed a reference genome sequence of this species and mapped its genome. Using DNA sequence variation from across the genome from *D. pulicaria* and *D. pulex*, we can understand occurrences of the distant past- and develop knowledge of *D. pulicaria's* evolutionary history. This work sets the stage for the continued development of *D. pulicaria* as an eco-genomic model. Next, I turned to an altogether different archive to understand the long-term temporal dynamics of the *Daphnia* community in Tanners Lake, a small suburban lake that has undergone extensive salinization recently because of human activity. I showed that ephippial egg banks recorded in lake sediments can be used to reconstruct the local history and ecological dynamics of *Daphnia* populations. Interestingly, we found that specific to this lake, salinization, despite how extreme it changes in this lake, does not exert a huge influence on the dynamics of the population. Finally, the egg bank can also be used in another way to look closely at the genomes of samples from across time to better understand the information in the sediment archive. In my third and final chapter, we hatched *D. pulicaria* from across ~25 years from Tanners Lake. We then sequenced their genomes and used the DNA sequence variation encoded in them to shed light on why we might observe very few impacts of salinization in this population, and we looked at the emergent phenotypic characteristics over time. The genomes and phenomes paint a rich picture of evolution of tolerance over time. Thus, it seems that evolutionary responses over relatively short times scales may help buffer population from the worst impacts of salinization.

However, such a conclusion requires an important caveat. Lake populations of *Daphnia* may exhibit small effective population sizes resulting from long-term clonal selection and reduced recombination, and reduced gene flow due to their inherent insular nature. Thus, these features may limit such populations in their capacity to evolve, as such some lake populations will not meet the challenges of their environment and will be extirpated. The spread of adaptive alleles will be highly dependent on the underlying genetic structure of the *D. pulicaria* metapopulation. This, of course, could not be explored by surveying a single lake in this present work.

The pursuit of research does not conclude with one's dissertation. As I complete my PhD, I am left with many more questions than when I started. I am lucky that as of this writing, I get to continue in pursuit of these with generous funding from the State of Minnesota with excellent collaborators. Soon, I will explore the evolutionary responses of populations not across time, but across space and assess the role of underlying genetic structure on adaptation to salinity. This work will extend much of what was done here and drive it in new directions. While these first steps are important, one aspect of the adaptation puzzle remains elusive in my dissertation. Specifically, the mappings of specific genetic variants or alleles to the salinity tolerance phenotype. Determining the specific mappings of allelic variation to variance in quantitative traits has been seen as the "ultimate" goal of geneticists since the "dawn" of the Modern Evolutionary Synthesis. New technologies offer a promising avenue for this pursuit, particularly the use of Genome Editing (e.g., CRISPR-Cas9) to introduce mutations onto novel genetic backgrounds to explore their impact on phenotypes. As this technique becomes increasingly trackable in *Daphnia*, the significance of the mutations that were highlighted in my third chapter can be further explored. In the future, unifying the powerful temporal approach of Resurrection Ecology with ecological and evolutionary genomics to identify a list of selected variants and

validating the phenotypic effects with CRISPR-edited clones, offer a robust way to expand our understanding of the "marks" that humans continue to leave in biological archives. This becomes increasingly more critical in the face of the environmental challenges that the future will bring.

Appendix 1

Supplemental Materials to Chapter 1:

The roles of recombination and selection in shaping genomic divergence in an incipient ecological species complex.

Matthew J. Wersebe[1,*], Ryan E. Sherman[2], Punidan D. Jeyasingh[2], Lawrence J. Weider[1]

[1] Department of Biology, Program in Ecology and Evolutionary Biology, University of Oklahoma, Norman, OK, USA

[2] Department of Integrative Biology, Oklahoma State University, Stillwater, OK, USA

[*]Corresponding Author: matthew.wersebe@ou.edu, 730 Van Vleet Oval, Rm 314 Norman, OK 73071

Methods:

Recombinant Panel Breeding:

Briefly, the above F0 (grandparental 2X and 13B) lineages were sexually crossed to produce the F1 (parental) clone, which was subsequently propagated parthenogenetically in the laboratory. The F1 clone (F1-1A) was induced to produce males under "stressful" conditions including high population densities in "crowding" water (L.J. Weider, unpublished data). F1 males were crossed with F1 females to produce potential F2 resting eggs (ephippia). F2 ephippial eggs (i.e., usually paired) were carefully removed from their ephippial casings and were placed in 24-well cell culture plates containing 1-2 mL of sterile COMBO medium per well (Kilham et al. 1998). The eggs were then subjected to hatching cues, which consisted of incubation for two weeks in cold and dark conditions at 4°C, followed by exposure to warmth and 24-hr light conditions at 20°C to establish the F2 clone bank. Approximately 7000 F2 eggs were submitted to this hatching protocol. This produced an F2 population of 255 clones; further details about the recombinant panel are detailed in Sherman et al. (2021).

Isolation of HMW DNA for Genome Assembly:

To isolate sufficient DNA for Genome Sequencing, individuals of clone F_1-1A were grown in mass cultures in 5 gallon buckets and feed *ad libitum* (~25 mL/ Day/Bucket) on chemostat cultured *Scenedesmus acutus* grown in High Ortho-phosphate (HOP) COMBO media (Kilham et al. 1998). To reduce the amount of exogeneous contamination in the DNA extraction from epibionts and the gut microflora, a total of 1,100 adult individuals from F1-1A clone were incubated in solution containing 50 mg L^{-1} each of ampicillin and tetracycline (Sigma Cat. # A9518 & T7660, respectively), for 72 hours and fed 1-mL of 5 g L^{-1} Sephadex beads (® Sigma Cat. # S5772) once a day. Approximately 950 individuals survived the antibiotic incubation and were isolated into 18 1.5 mL microfuge tubes containing 48-50 adult individuals per tube. First *Daphnia* were placed on a small (1 cm x 1 cm) Nitex® mesh patch and closed inside a 1.5 mL microfuge tube lid. These tubes were spun at 10,000 rpm for 2 minutes in centrifuge (Eppendorf 5415 C) to remove access liquid from the carapace and prevent ice crystal formation upon flash freezing. Next, tissues were then transferred to a clean 1.5 mL tube and flash frozen in a freshly charged liquid nitrogen dry shipper (Air Liquide Voyageur 12) for 30 minutes. After 30 minutes, tissues were lysed using the MasterPure DNA extraction kit with the modification that incubation with proteinase K at 60°C was extended to 1-hour with gentle tube rocking every 15 minutes. Next, proteins were precipitated according to the manufacturer's instructions. Instead of precipitating DNA into separate tubes, the remaining liquid supernatant from 9 1.5 mL tubes was pooled into a 15 mL Falcon tube and the DNA was precipitated with 4.5 mL of cold isopropanol at -20°C overnight and collected by centrifugation at 4°C for 10 minutes. DNA was then washed with cold 70% ethanol and resuspended in TE buffer.

Linkage Map Construction:

Due to the computational burden of calling all 105 individuals together, we chose to call variants in groups of 10 F2s together with the two F0s and the F1 using Freebayes (Garrison and Marth 2012). The unfiltered vcf file from each group was merged and redundant samples removed

using BCFtools (Danecek et al. 2021). Using VCFtools (Danecek et al. 2011), the fully merged vcf file was filtered to a set of high quality SNPs and small Indels (min Q = 30, mac =3, max-missing = 0.8, minDP = 3, maxDP=200).

The genetic linkage map was constructed using LepMap3 (Rastas 2017). The filtered set of variants was turned into LepMap3 data format using the ParentCall2 module with the argument removeNonInformative=1. Next, we used the SeparateChromosomes2 module of LepMap3 to assign markers to linkage groups using the distortionLod=1 argument, since our pedigree included only one family. According to LepMap3's creator (M. Wersebe, Personal Communication), the heuristic (n/10 * 3; e.g., 102/10 * 3 = 30.6), was used to define the start of our search for the best LOD to use in the SeparateChromosomes2 module of LepMap3. We ran SeparateChromosomes2 separately for values ranging from 32 to 20 for the lodLimit argument. Next, we calculated the number or markers assigned to the first 12 linkage groups (LGs) and chose runs of SeparateChromosomes2 that maximized the placement of markers into the first 12 LGs and had the greatest difference between LG 12 and LG 13. Then, we singled out all markers forming LGs above the first 12 and submitted this map to the module JoinSingles2All to place more markers into the 12 defined LGs, with the arguments lodLimit=6, lodDifference=1, and iterate=1. From these, we selected maps that placed most of the markers into the 12 LGs and prevented one large and several small LGs (e.g., all LGs between 150,000 and 10,000 markers) to form. LGs from Maps with these characteristics were passed to the OrderMarkers2 module separately and ordered with 12 phasing iterations. After 12 phasing iterations, the LepMap graphs were produced using the module LMPlot and inspected for any irregularities. If another run of OrderMarkers2 was warranted after inspection of the LMPlots, these were run with the parameters above. The completed maps were finally passed to LepAnchor to scaffold the genomes assembly using the default parameters (Rastas 2020). A single map was chosen based on the results from LepAnchor which best matched the distribution in hypothesized chromosome sizes of the *D. pulex* group.

Demographic Analysis:

Each run of Fastsimcoal2 (Excoffier et al. 2013) consisted of 40 optimization cycles implementing 1000000 coalescent simulations. We determined the best model parameters by extracting the run with highest likelihood. Next, we estimated 95% confidence intervals for model parameter estimates by generating non-parametric bootstrapped SFS. Then we simulated 100 SFS for the model estimated with the best likelihood. For each of the 100 simulated SFS, we re-ran fastsimcoal2 40 times with each run undergoing 40 optimization cycles implementing 250000 coalescent simulations. The parameter estimates for the best run based on the likelihood were extracted and 95% confidence intervals determined using a T-test.

Genome Co-linearity:

To assess the co-linearity of our assembly with that of *D. pulex*, we conducted a whole genome alignment and constructed a dot plot to survey for any large-scale rearrangements that could be suppressing recombination. We aligned the recent *D. pulex* assembly PA42 version 4.1 from Ye et al. (2017) to the 13B haplotype assembly (scaffolded by our genetic map) using Minimap2 (Li

2018) on the D-GENIES webserver to construct a dot plot (Cabanettes and Klopp 2018). The resulting dot plot is presented in supplemental figure S30.

Supplemental Figures:

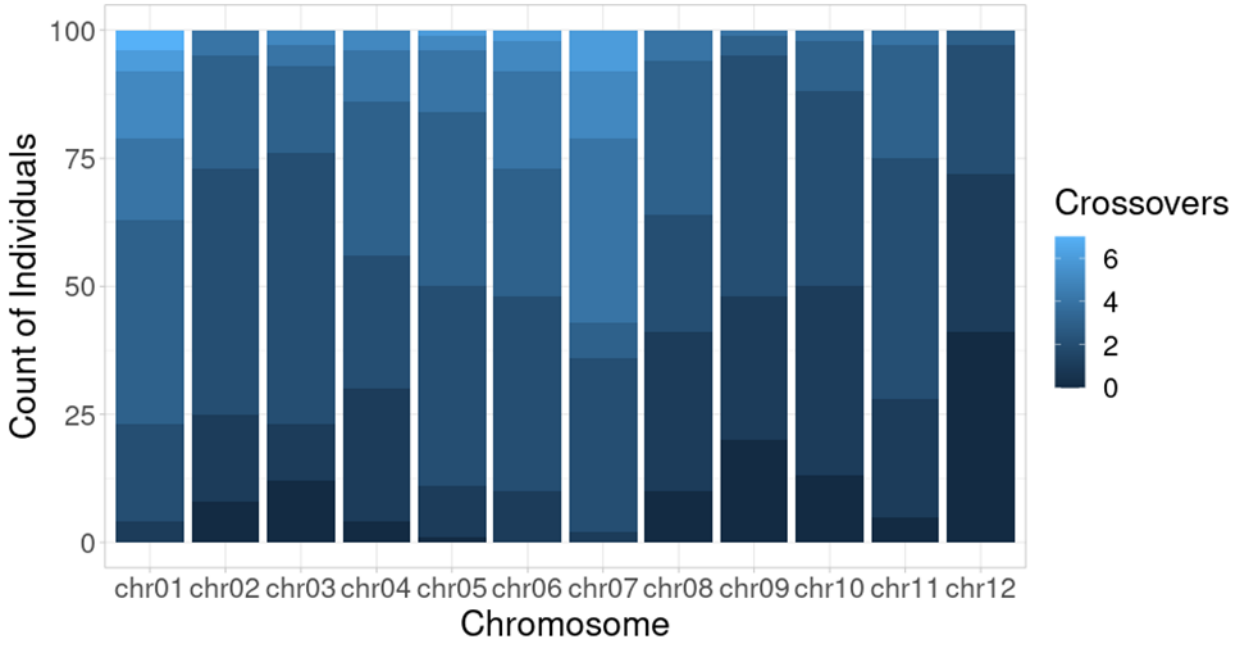


Figure S1: Crossover events per linkage group. The number of crossovers per linkage group per individual ranged from a high of 7 (on LG 1), to 0 (all but LGs 1 & 6). Most individuals experienced 2 to 3 crossovers per chromosome.

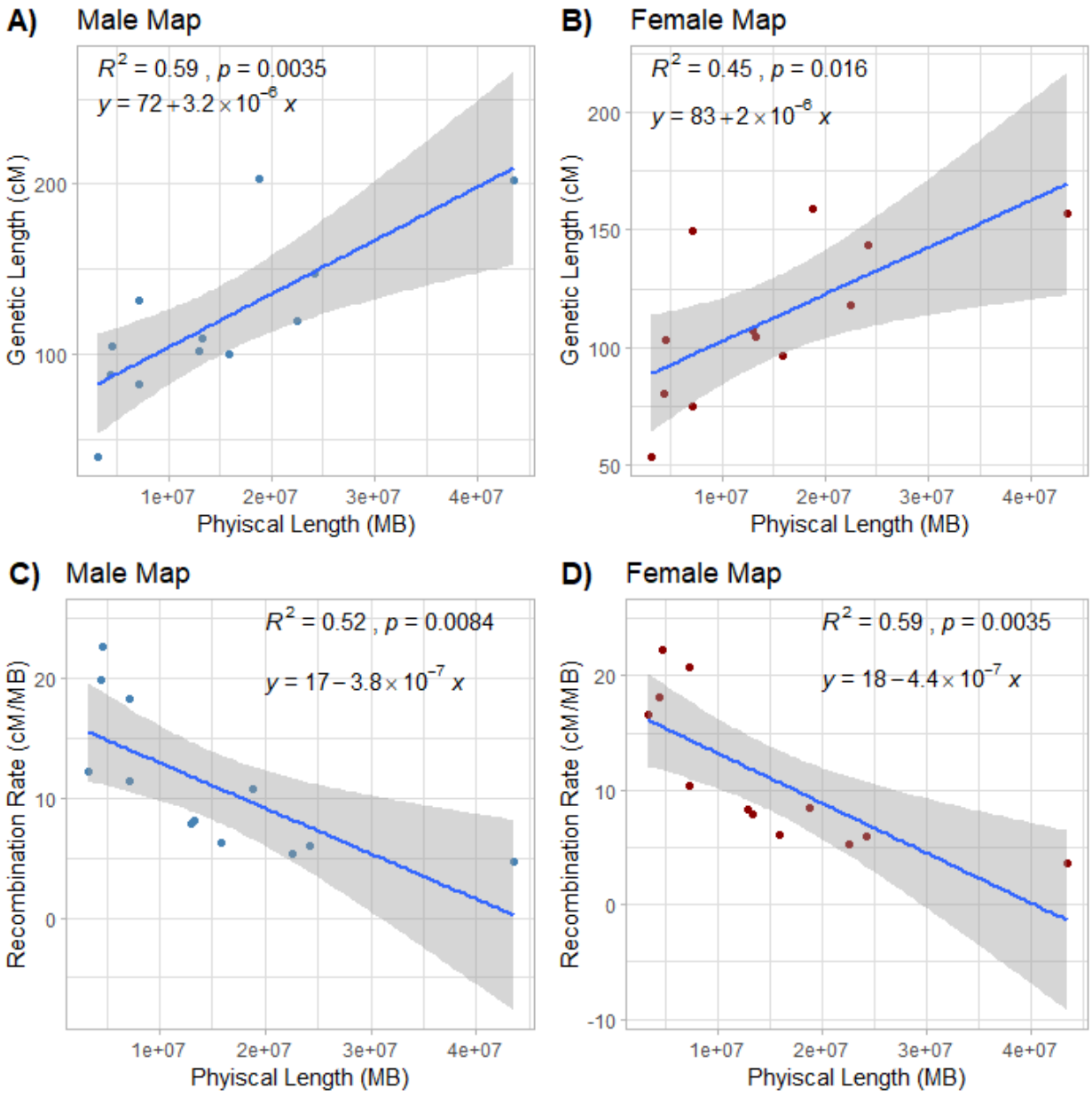


Figure S2: Correlations between Linkage Group (LG) physical lengths, genetic lengths and recombination rates. A) Correlation between LG physical lengths and their corresponding genetic lengths for the male map. B) Correlation between LG physical lengths and their corresponding genetic lengths for the female map. Positive slopes of the correlation across both maps indicate that the physically longest chromosomes are hosting the greatest number of recombination events. C) Correlation between LG physical lengths and their corresponding recombination rates in cM/MB for the male map. D) Correlation between LG physical lengths and their corresponding recombination rates in cM/MB for the female map. Negative correlations indicate that larger LGs host regions of suppressed recombination. This effect is likely driven in part by the fact that the smallest LG may only approximate arms of chromosomes.

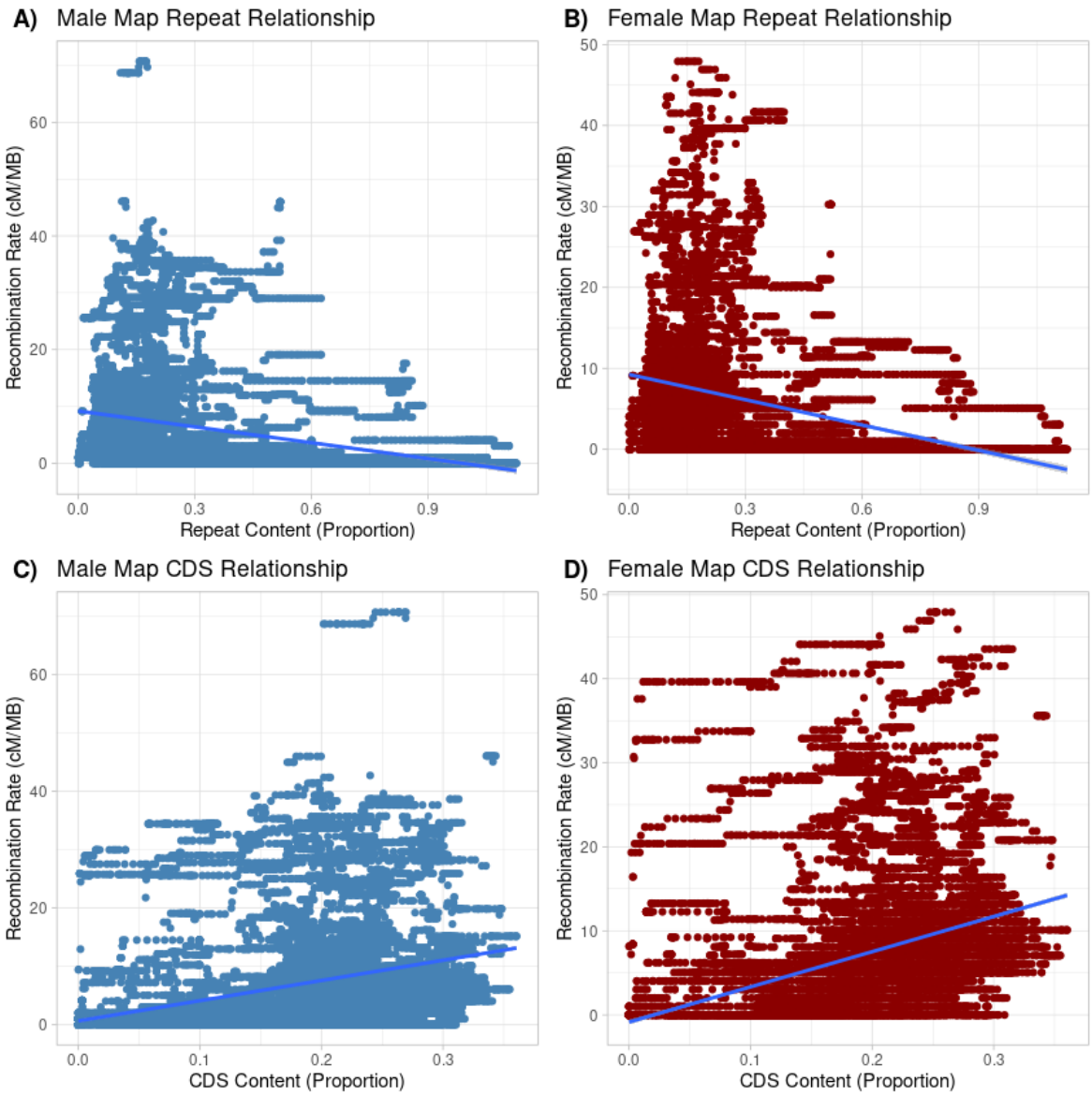


Figure S3: Aggregate correlations between sequence content and recombination rates in the *Daphnia pulicaria* genome. A) Negative correlation between recombination rate (cM/Mb; calculated using the male genetic map) and the repeat sequence content in 1-Mb windows through the genome. B) Negative correlation between recombination rate (cM/Mb; calculated using the female genetic map) and the repeat sequence content in 1-Mb windows through the genome. C) Positive correlation between recombination rate (cM/Mb; calculated using the male genetic map) and Coding DNA Sequence (CDS) content in 1-Mb windows throughout the genome. D) Positive correlation between recombination rate (cM/Mb; calculated using the female genetic map) and Coding DNA Sequence (CDS) content in 1-Mb windows throughout the genome. In all cases windows were 1-Mb in length stepping 10-Kb along chromosomes. Recombination rates were determined by subtracting the ending genetic location (in cM) of the window from the starting marker and dividing by physical distance (i.e., 1-Mb). CDS content and Repeat was extracted from the predicted gene models from Augustus (CDS) or from

RepeatMasker (Repeats) in the same windows in which recombination rates were estimated across the genome.

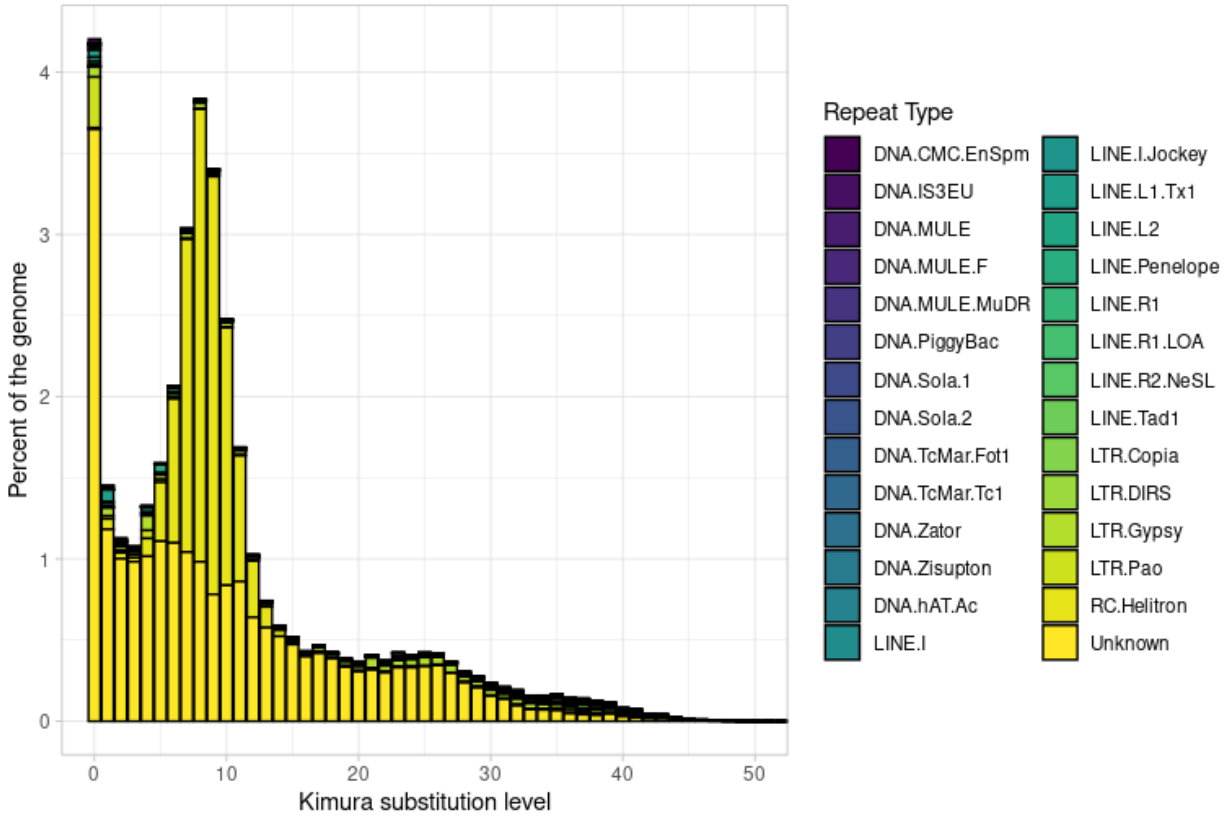


Figure S4: Kimura 2-parameter distances and percent genome coverage of repeat families. Color of stacked bar plot indicates the identity of the repeat family. There is a peak of > 4% of the genome covered in low divergence (genetic distance = 0) indicating that transposition is active and ongoing in the genome. A more ancient transposition burst also occurred, indicated by the bi-modal distribution primarily driven by RC Helitron superfamily of repeat elements.

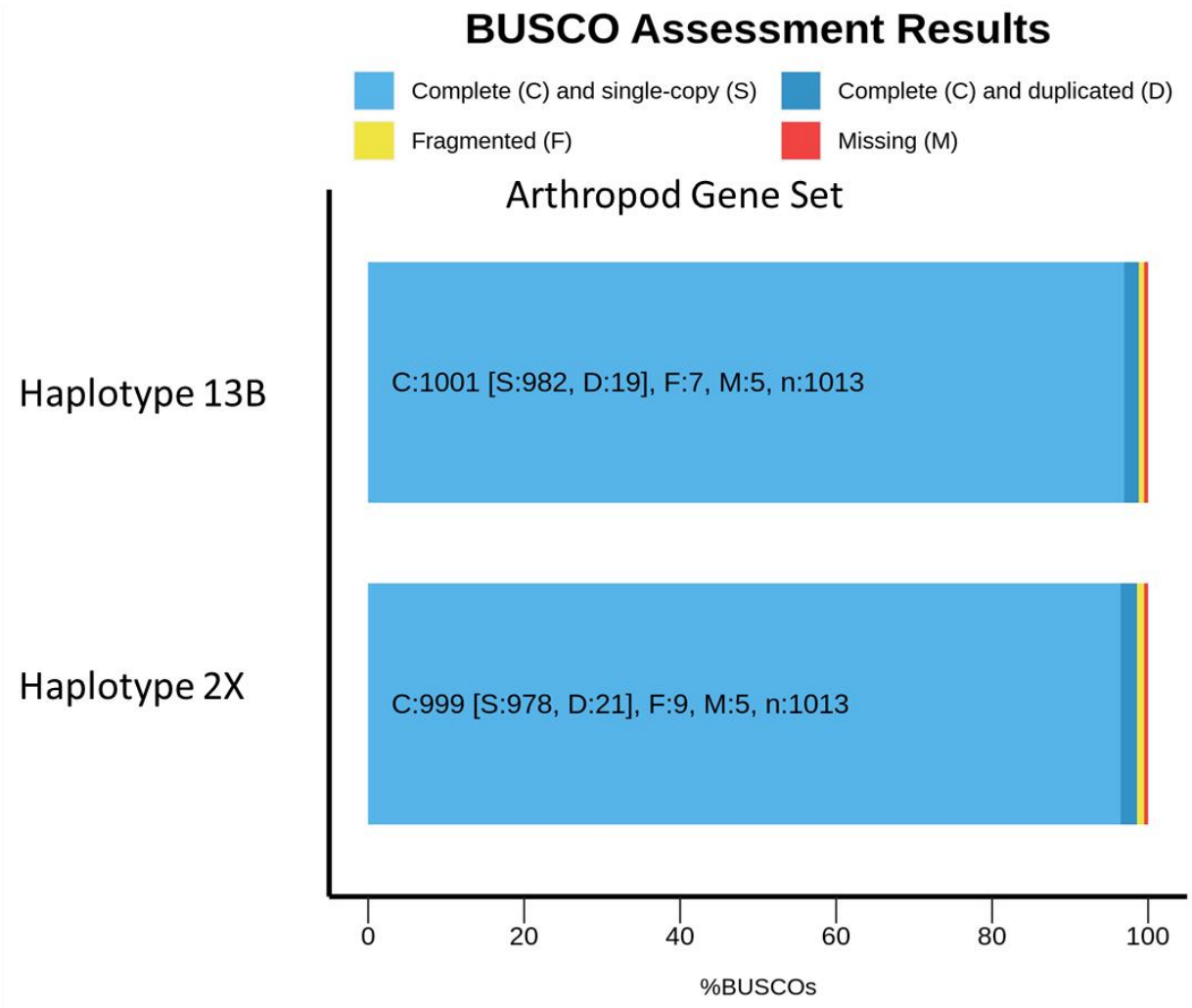


Figure S5: Benchmarking Universal Single-Copy Orthologs (BUSCO) results for each assembly haplotype. Stacked bar plot indicates the percent of BUSCO genes in each category; C: complete, S: single copy, D: duplicated, F: fragmented, M: missing. Haplotypes refer to the phased assembly inherited by the F1 genotype from each of the F0 parental genotypes. Genomes were assayed against the Arthropod Gene set downloaded from OrthoDB version 10 (Kriventseva et al. 2019). Figure was produced using the supplemental Python and R scripts packaged with BUSCO (Simão et al. 2015).

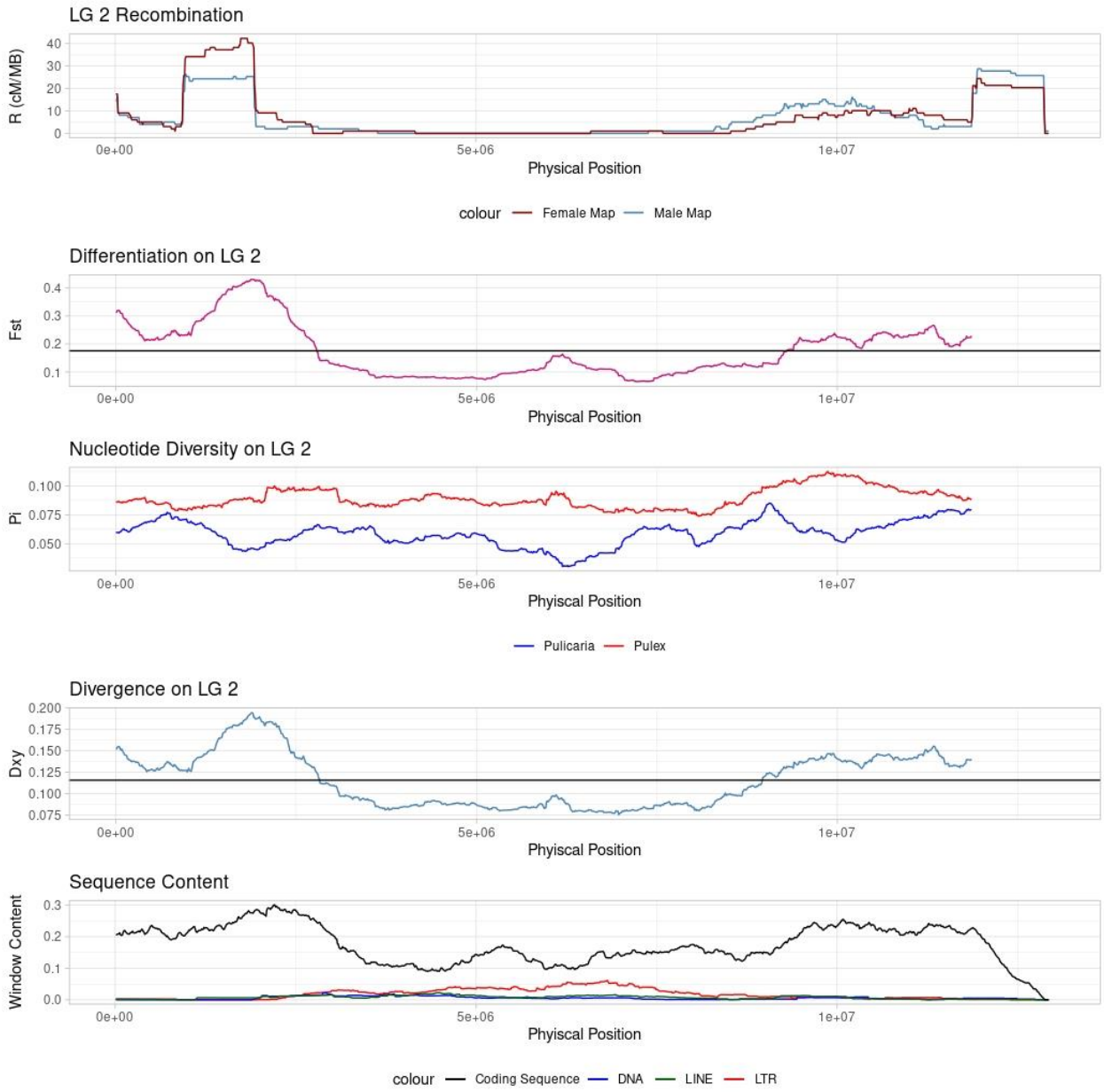


Figure S6: Sliding window analysis of Linkage Group (LG)-1 landscapes. (Top) Recombination landscape in 1-Mb windows stepping 10-Kb along the LG. Colors reflect the male (blue) and female (red) map positions. (Second from top) Sliding window F_{st} (relative differentiation) calculated between *D. pulicaria* and *D. pulex* in 1-Mb windows stepping 10-Kb along the LG. (Third from top) Sliding window π (nucleotide diversity) calculated for both *D. pulex* (red) and *D. pulicaria* (blue) in the 1-Mb windows stepping 10-Kb along the LG. (Fourth from top) Sliding window D_{xy} (absolute divergence) calculated between *D. pulicaria* and *D. pulex* in the 1-Mb windows stepping 10-Kb along the LG. (Bottom) Sliding window analysis of DNA sequence features along LG 1 in 1-Mb windows stepping 10-Kb. The black line is Coding DNA Sequence (CDS) content, blue indicates the DNA transposons content, green Long Interspersed Nuclear Elements (LINEs) content and red indicates Long Terminal Repeats (LTR) contents.

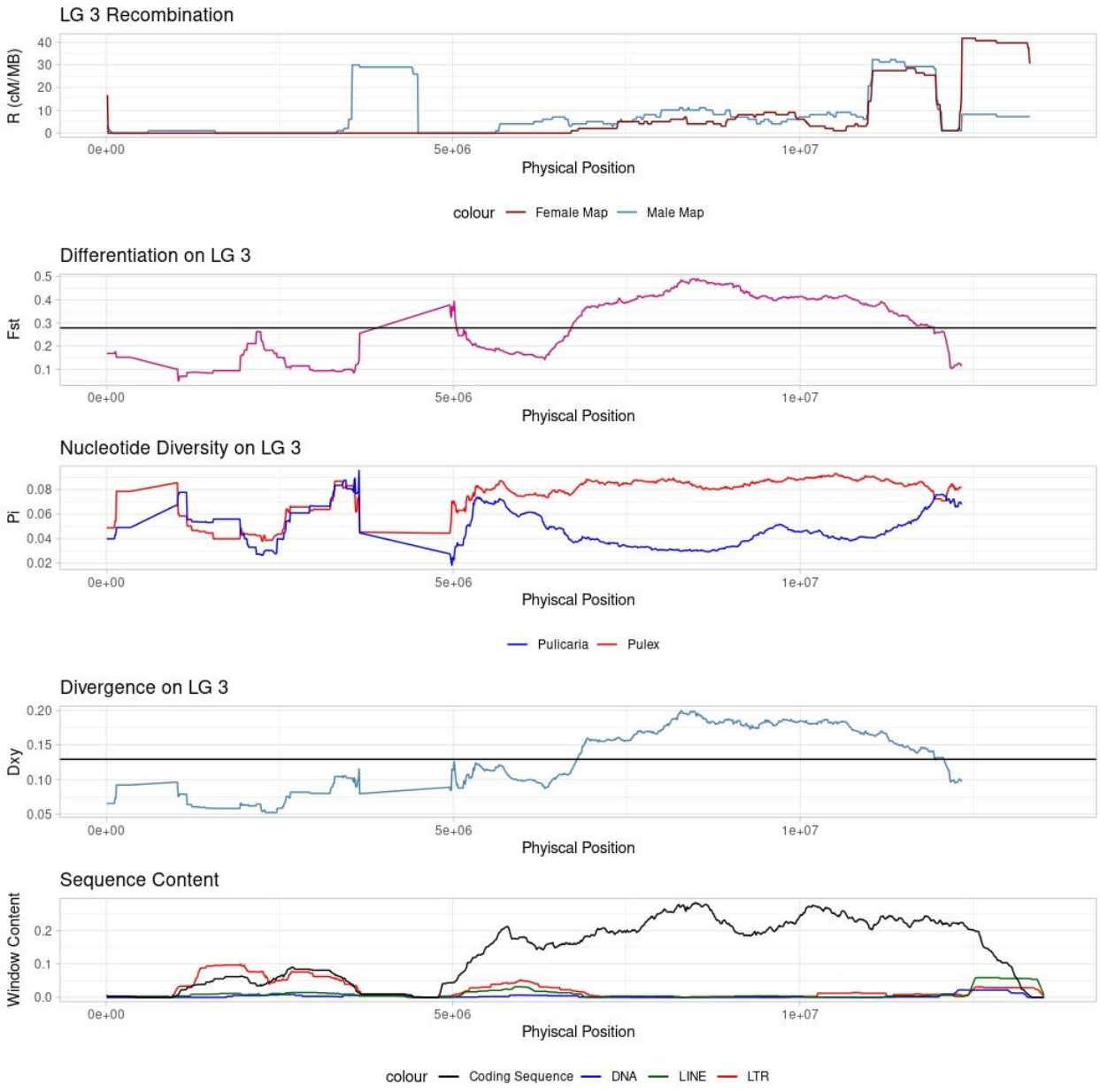


Figure S7: Sliding window analysis of Linkage Group (LG)-3 landscapes. Description same as above (Figure S6).

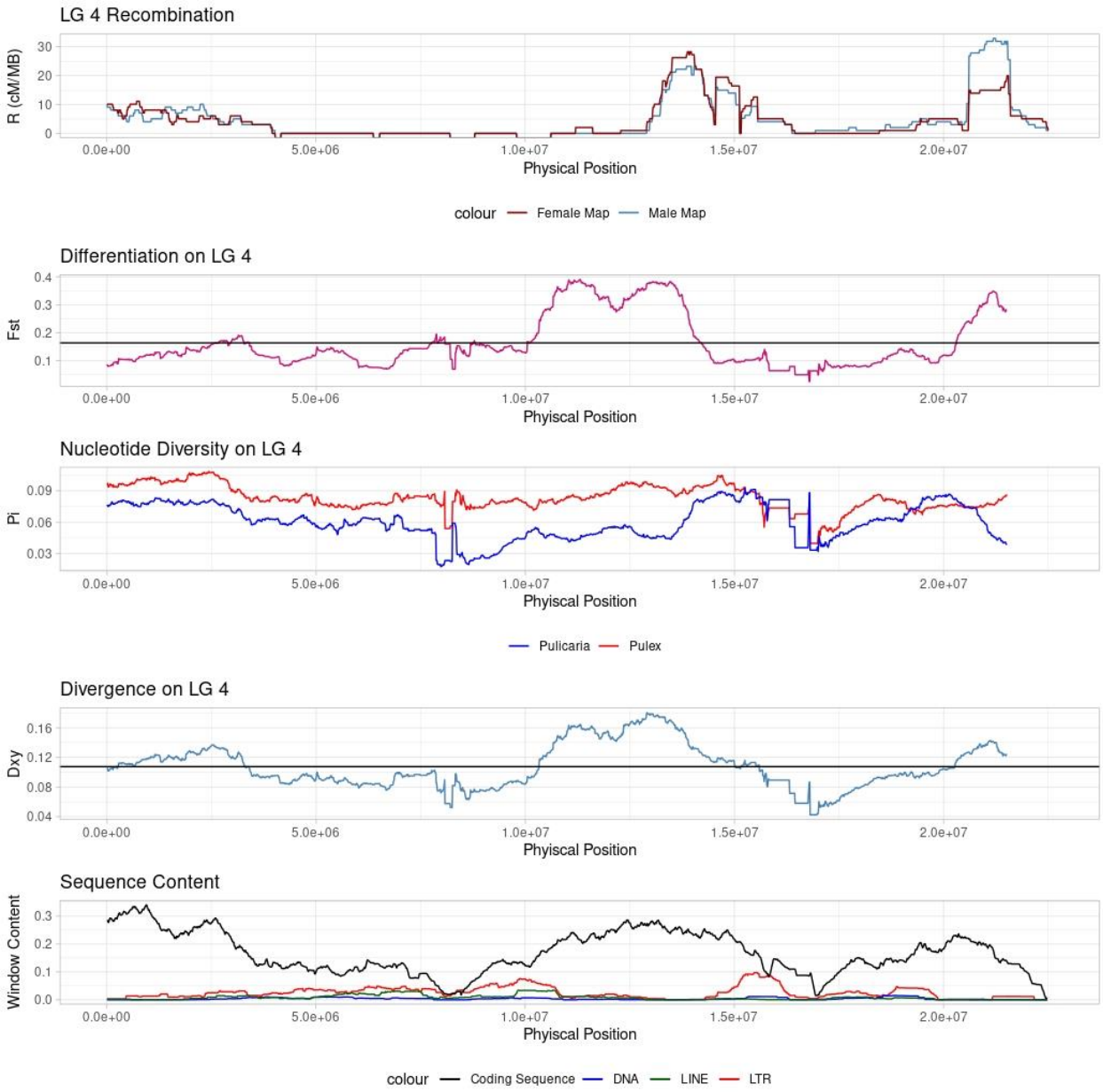


Figure S8: Sliding window analysis of Linkage Group (LG)-4 landscapes. Description same as above (Figure S6).

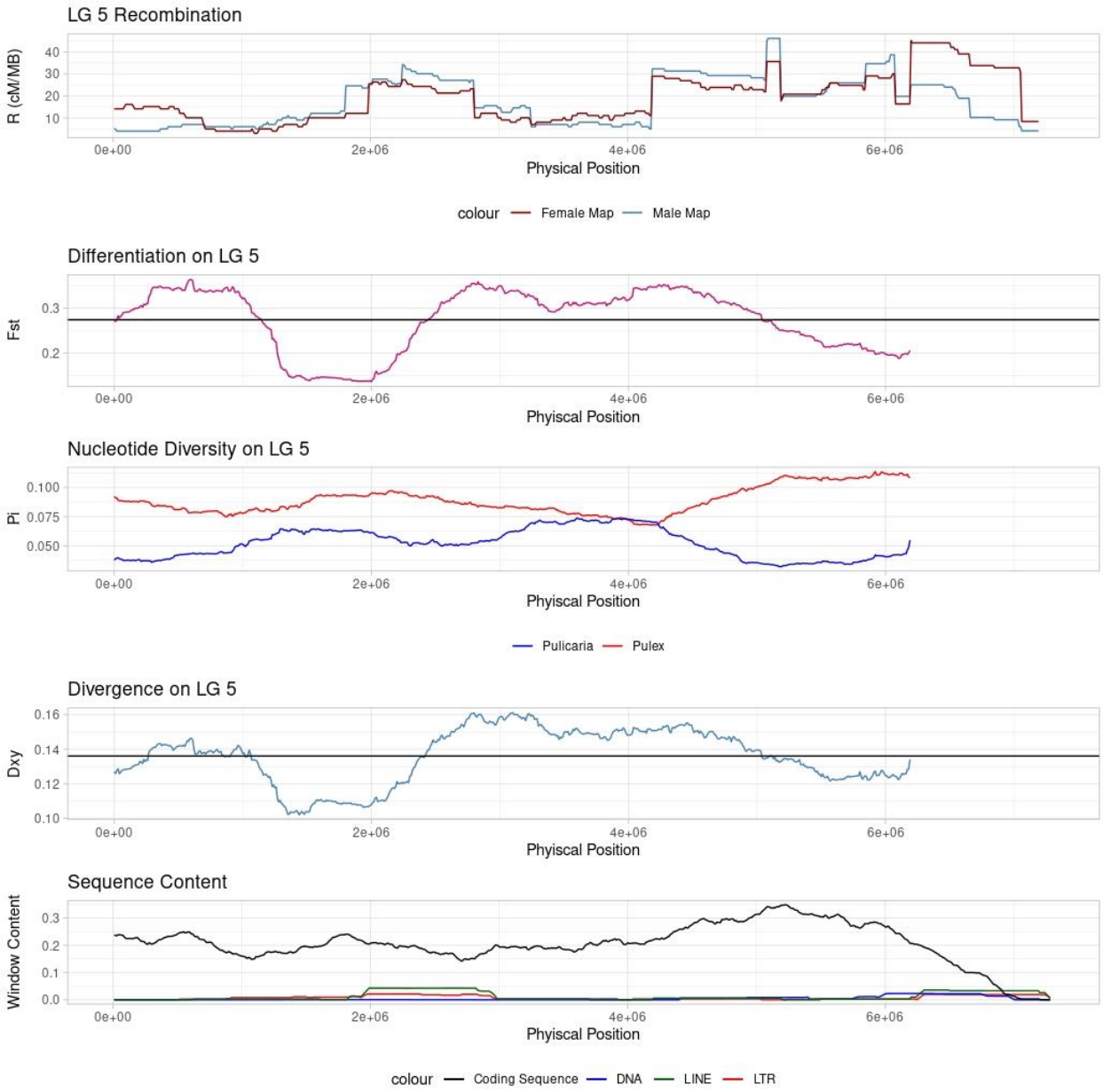


Figure S9: Sliding window analysis of Linkage Group (LG)-5 landscapes. Description same as above (Figure S6).

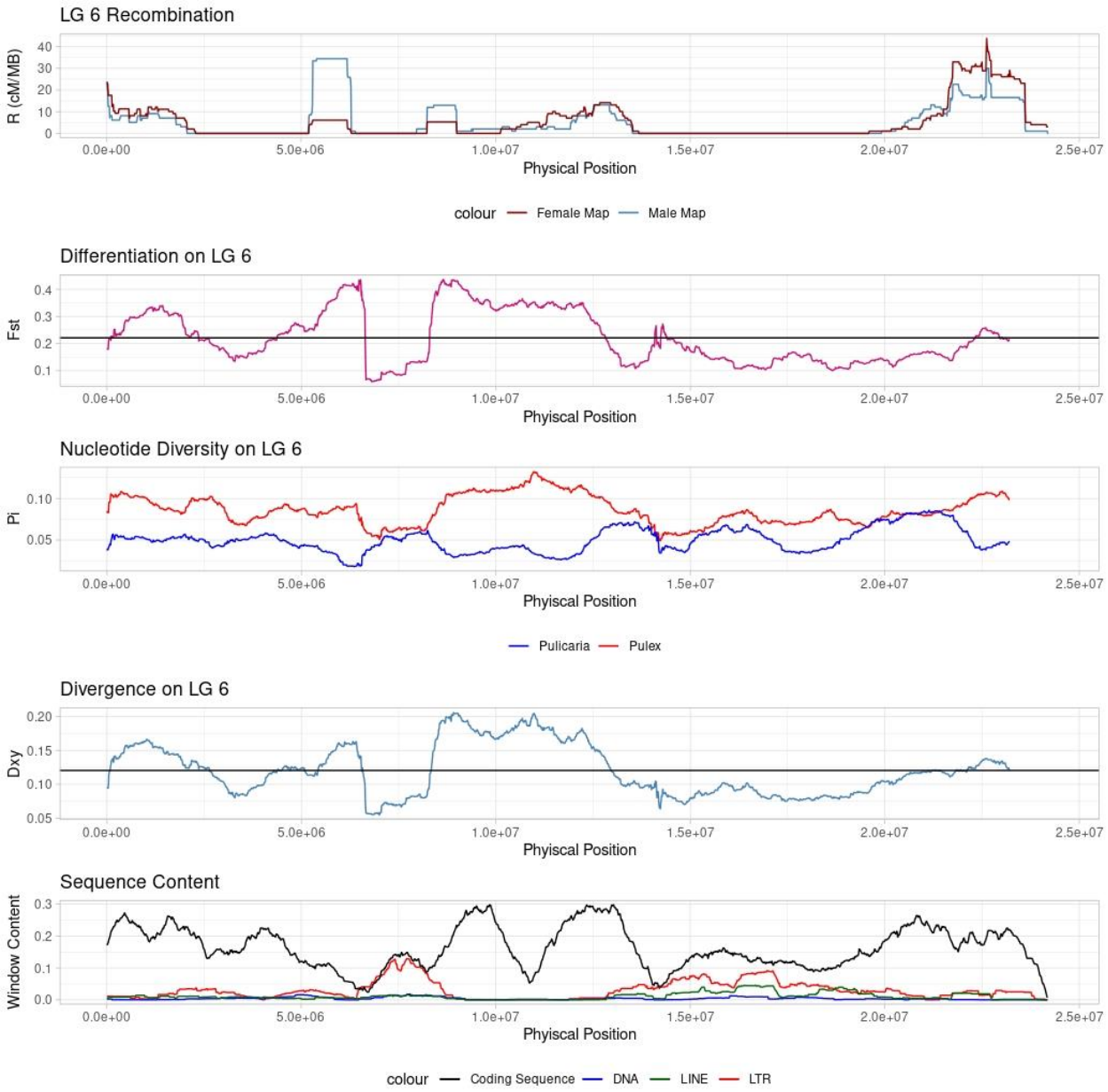


Figure S10: Sliding window analysis of Linkage Group (LG)-6 landscapes. Description same as above (Figure S6).

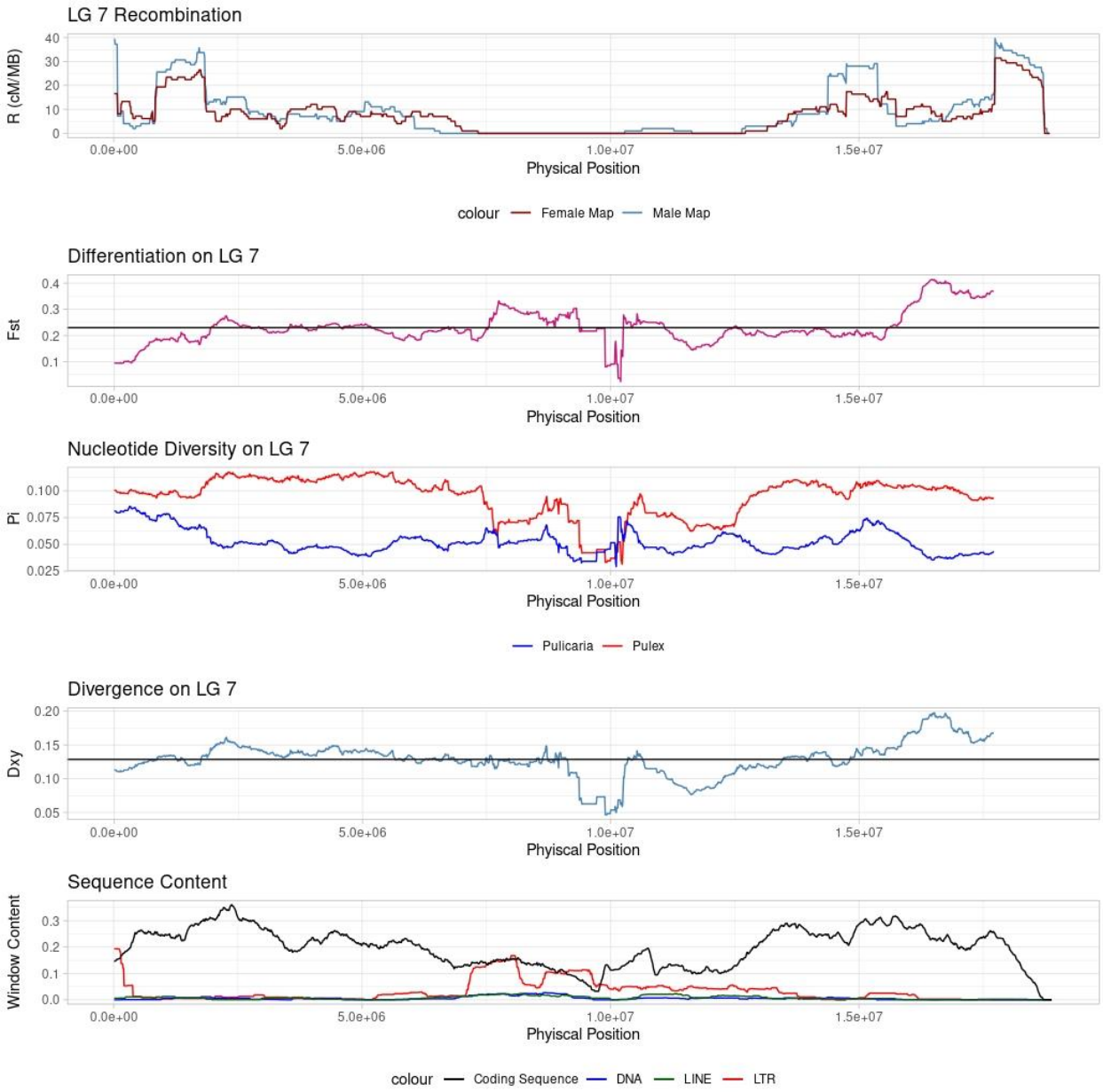


Figure S11: Sliding window analysis of Linkage Group (LG)-7 landscapes. Description same as above (Figure S6).

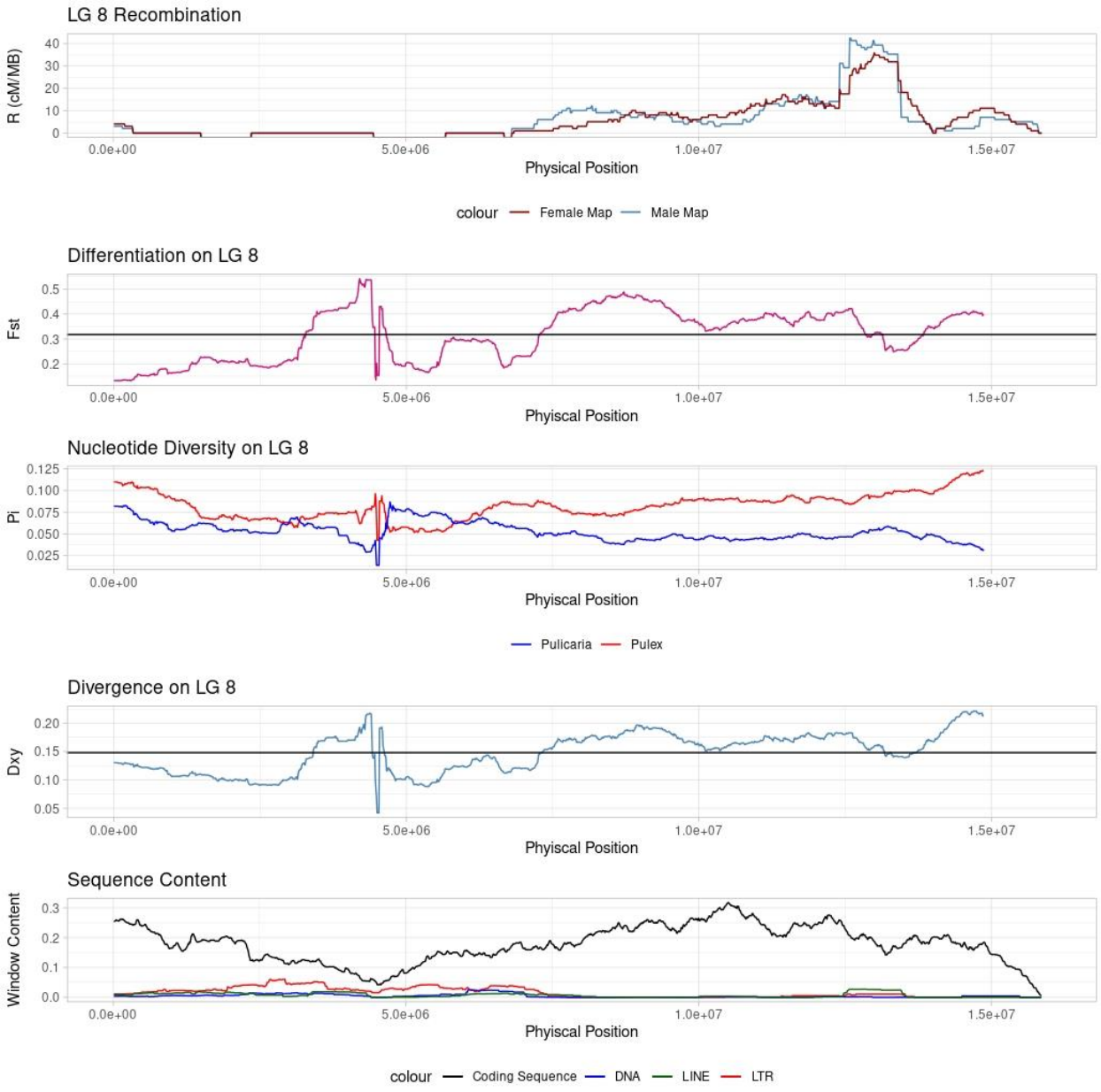


Figure S12: Sliding window analysis of Linkage Group (LG)-8 landscapes. Description same as above (Figure S6).

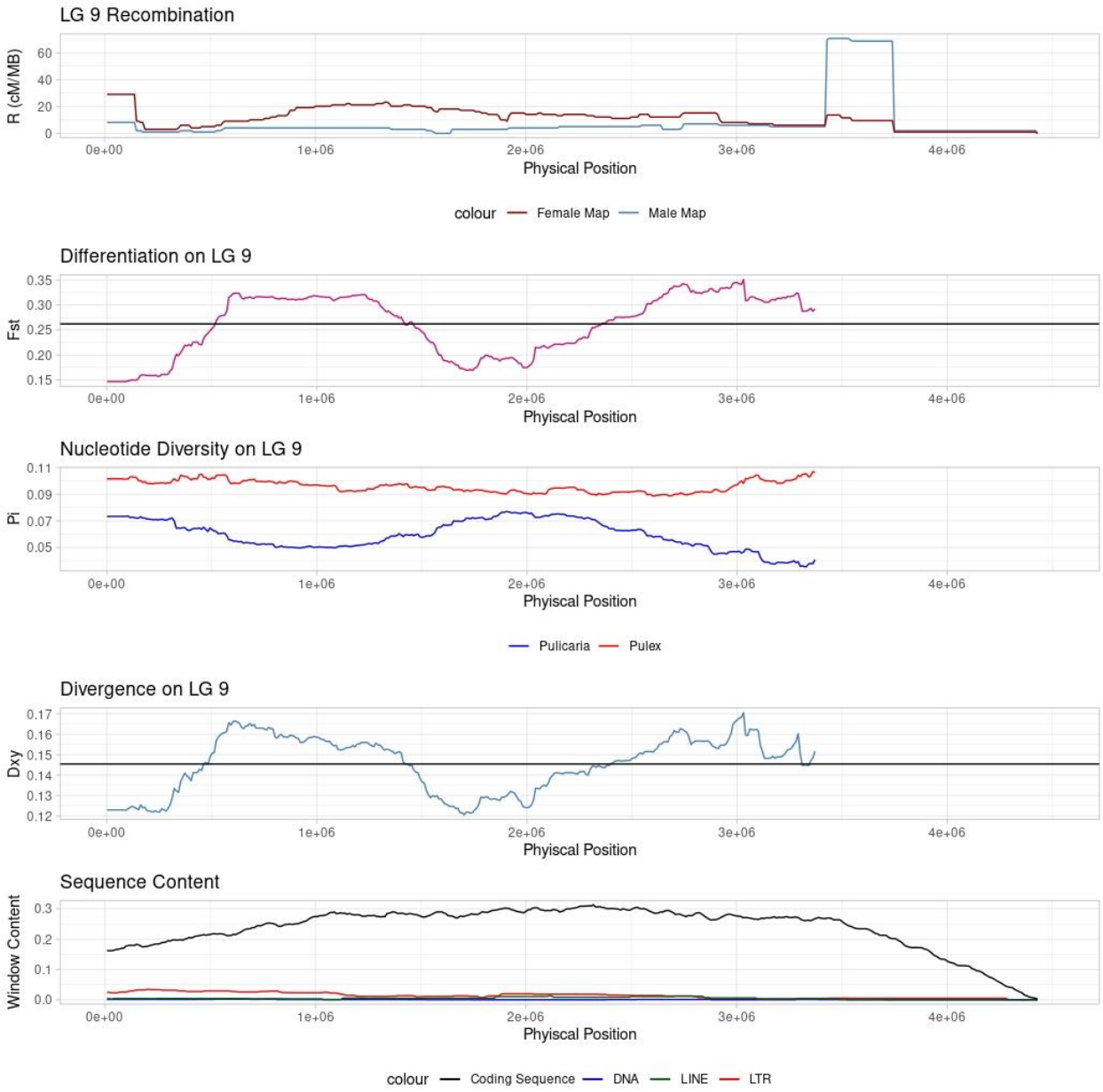


Figure S13: Sliding window analysis of Linkage Group (LG)-9 landscapes. Description same as above (Figure S6).

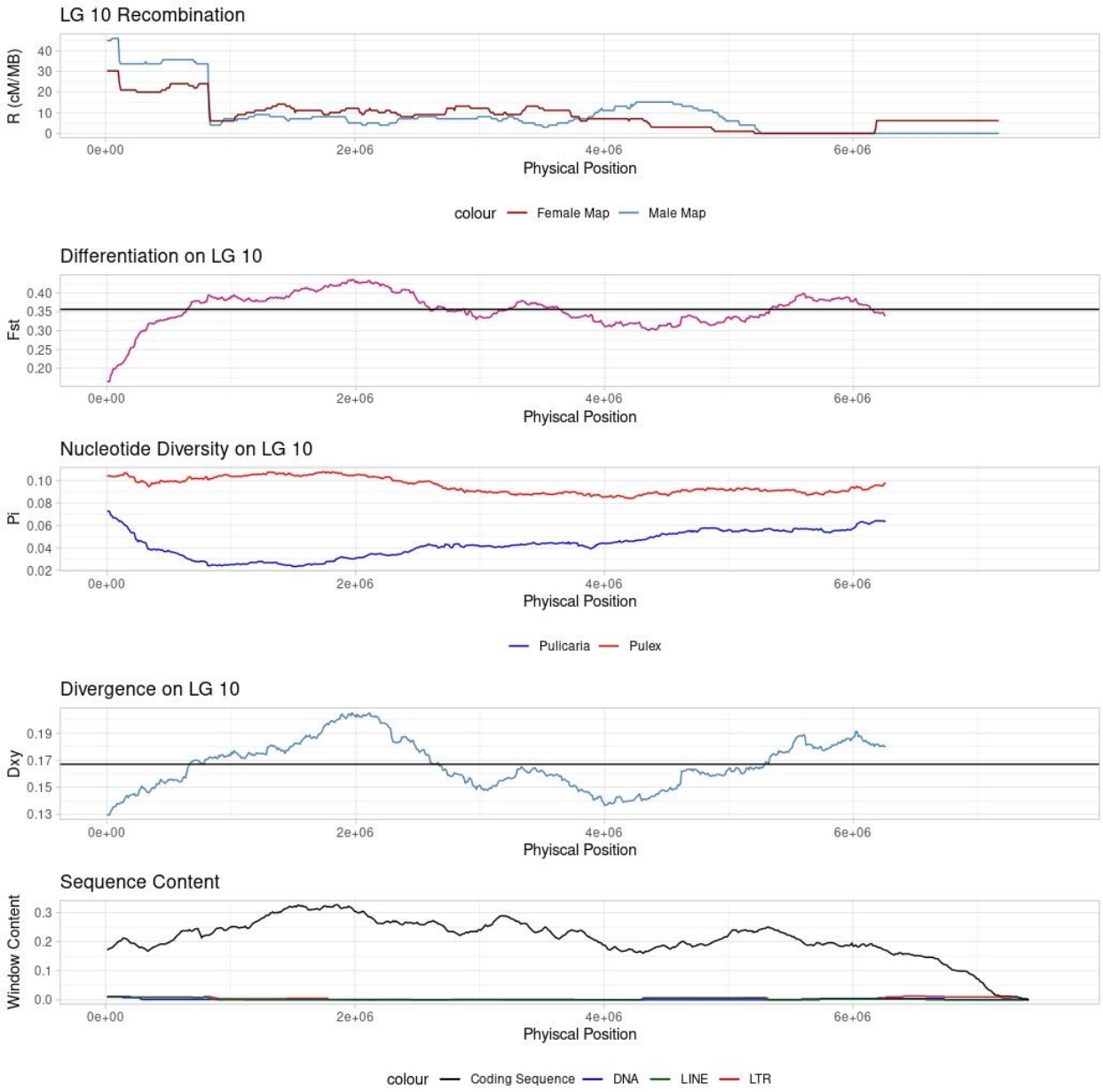


Figure S14: Sliding window analysis of Linkage Group (LG)-10 landscapes. Description same as above (Figure S6).

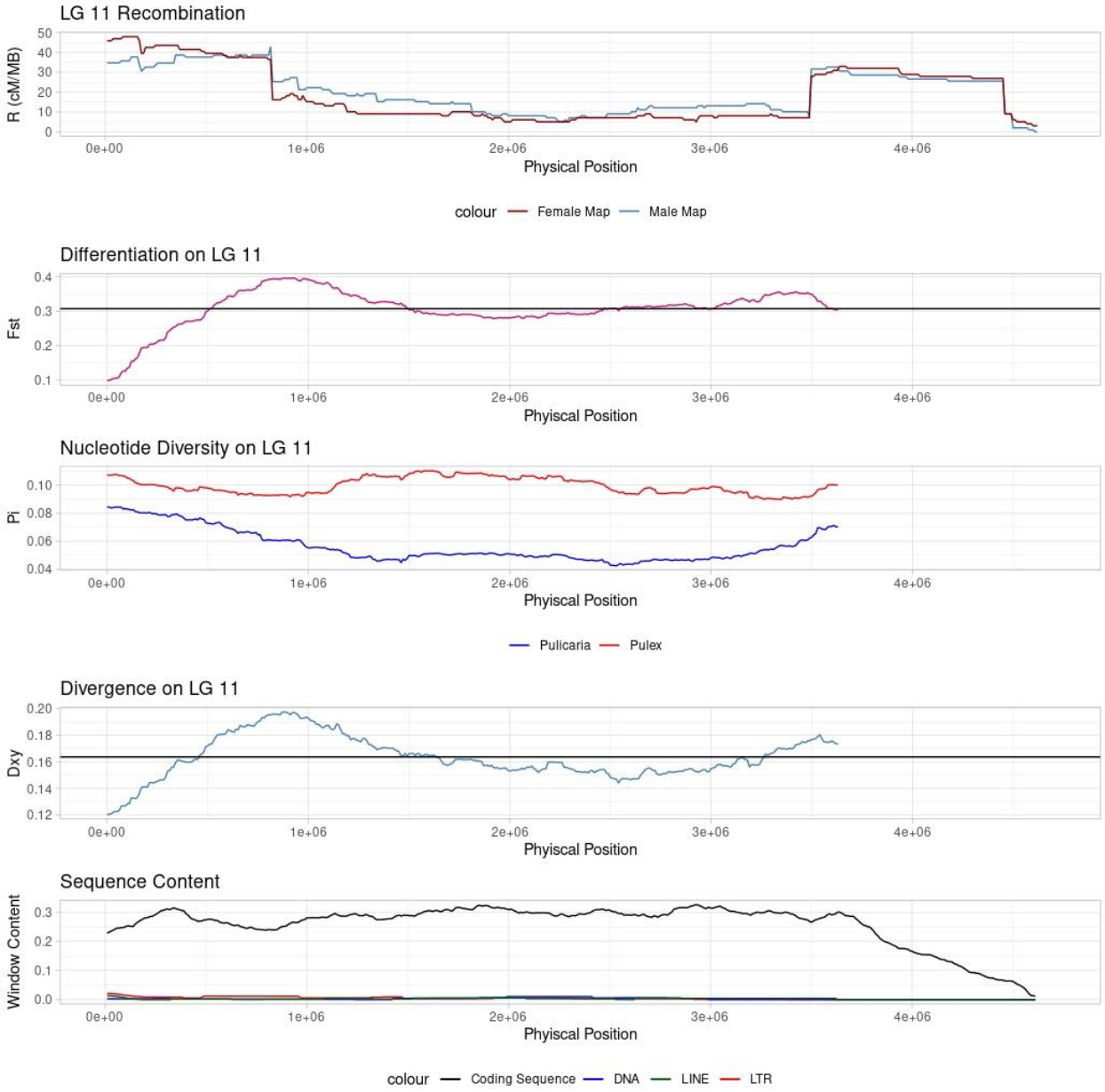


Figure S15: Sliding window analysis of Linkage Group (LG)-11 landscapes. Description same as above (Figure S6).

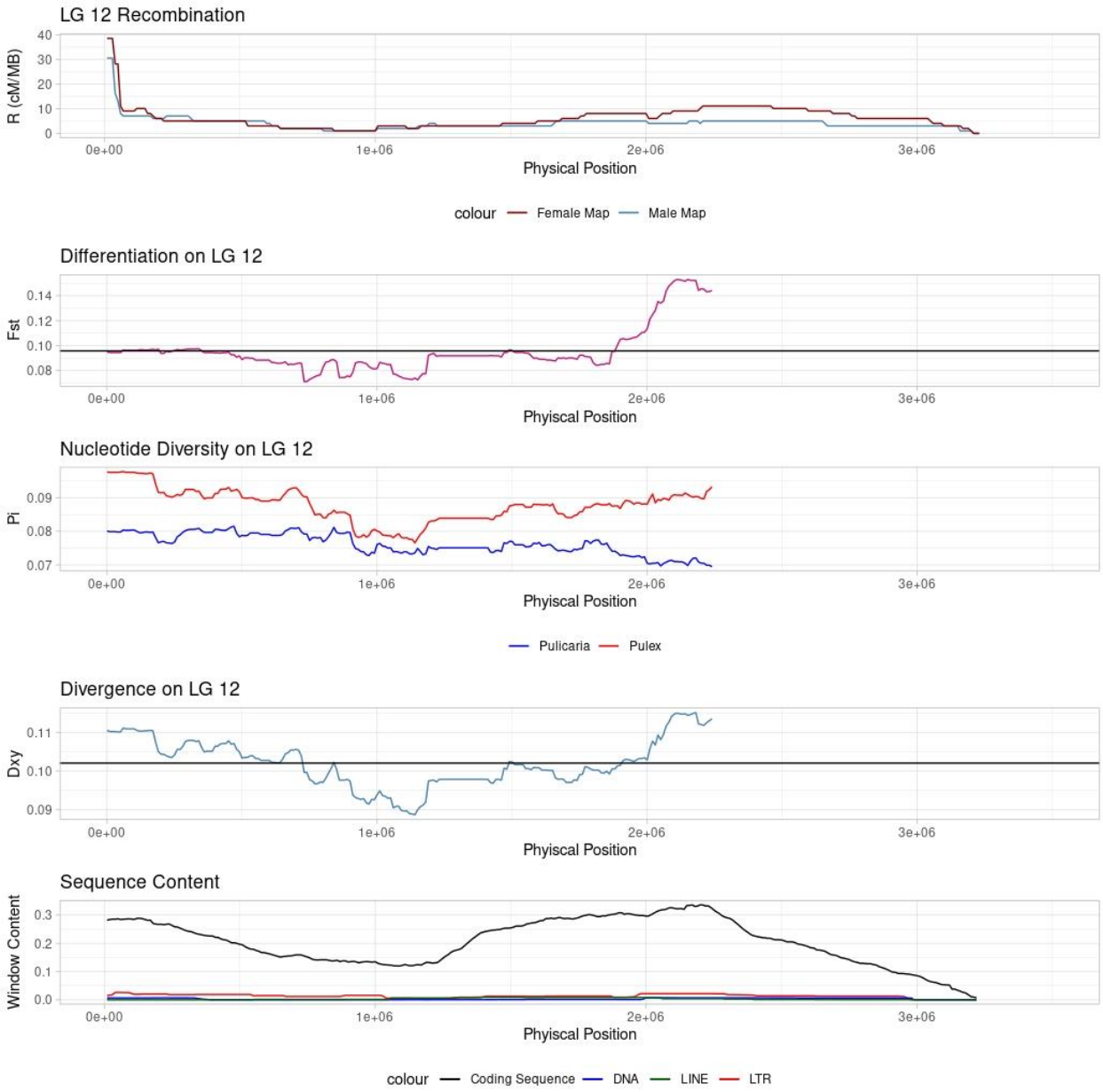
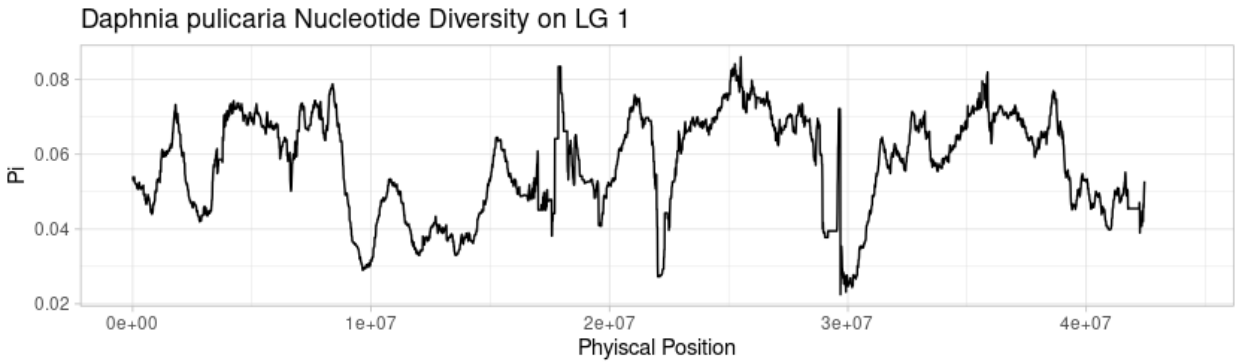
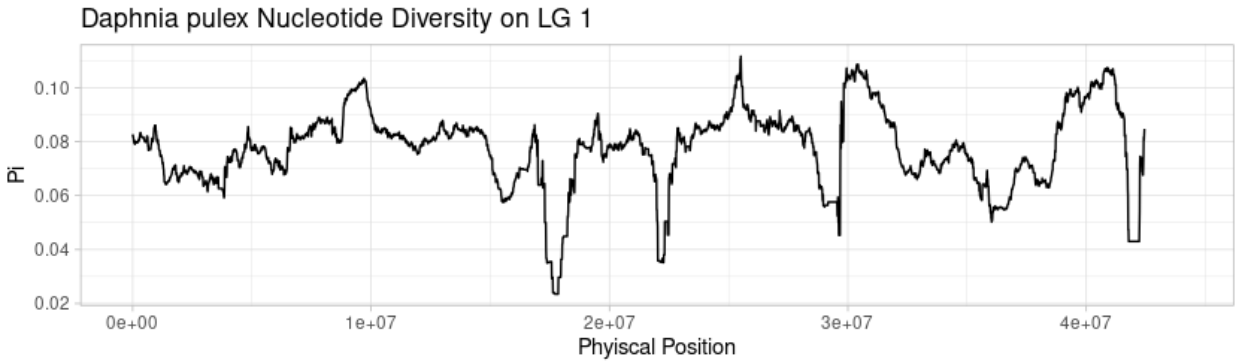


Figure S16: Sliding window analysis of Linkage Group (LG)-12 landscapes. Description same as above (Figure S6).


Daphnia pulicaria Nucleotide Diversity on LG 1
Pi
0.08
0.06
0.04
0.02
0e+00
1e+07
2e+07
3e+07
4e+07
Phyiscal Position


Daphnia pulicaria Tajima's D on LG 1
Tajima's D
0.0
-0.5
-1.0
-1.5
0e+00
1e+07
2e+07
3e+07
4e+07
Physical Position


Daphnia pulex Nucleotide Diversity on LG 1
Pi
0.10
0.08
0.06
0.04
0.02
0e+00
1e+07
2e+07
3e+07
4e+07
Phyiscal Position

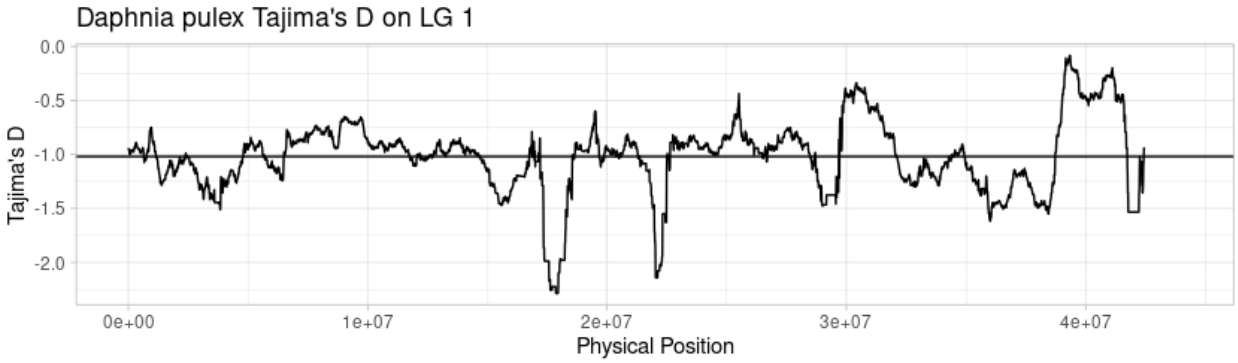

Daphnia pulex Tajima's D on LG 1
Tajima's D
0.0
-0.5
-1.0
-1.5
-2.0
0e+00
1e+07
2e+07
3e+07
4e+07
Physical Position

Figure S17: Sliding window nucleotide diversity and Tajima's D on Linkage Group (LG)-1 for *Daphnia pulicaria* and *Daphnia pulex.* (Top) *D. pulicaria* landscape of nucleotide diversity in landscape in 1-Mb windows stepping 10-Kb along the LG. (Second) *D. pulicaria* landscape of Tajima's in landscape in 1-Mb windows stepping 10-Kb along the LG. (Third) *D. pulex* landscape of nucleotide diversity in landscape in 1-Mb windows stepping 10-Kb along the LG. (Bottom) *D. pulex* landscape of Tajima's in landscape in 1-Mb windows stepping 10-Kb along the LG. Troughs in Pi along the chromosome for *D. pulicaria* generally correspond well with peaks in F_{st} and D_{xy} when comparing *D. pulex* & *D. pulicaria.* This suggests that selection is operating primarily in *D. pulicaria.* Estimates of Pi are also consistently higher in *D. pulex* (note the difference in the Y-axis in first and third panels), indicating that *D. pulex* has a consistently higher effective population size when compared to *D. pulicaria.* Tajima's D for both species is on average negative, indicating population expansion in both species.

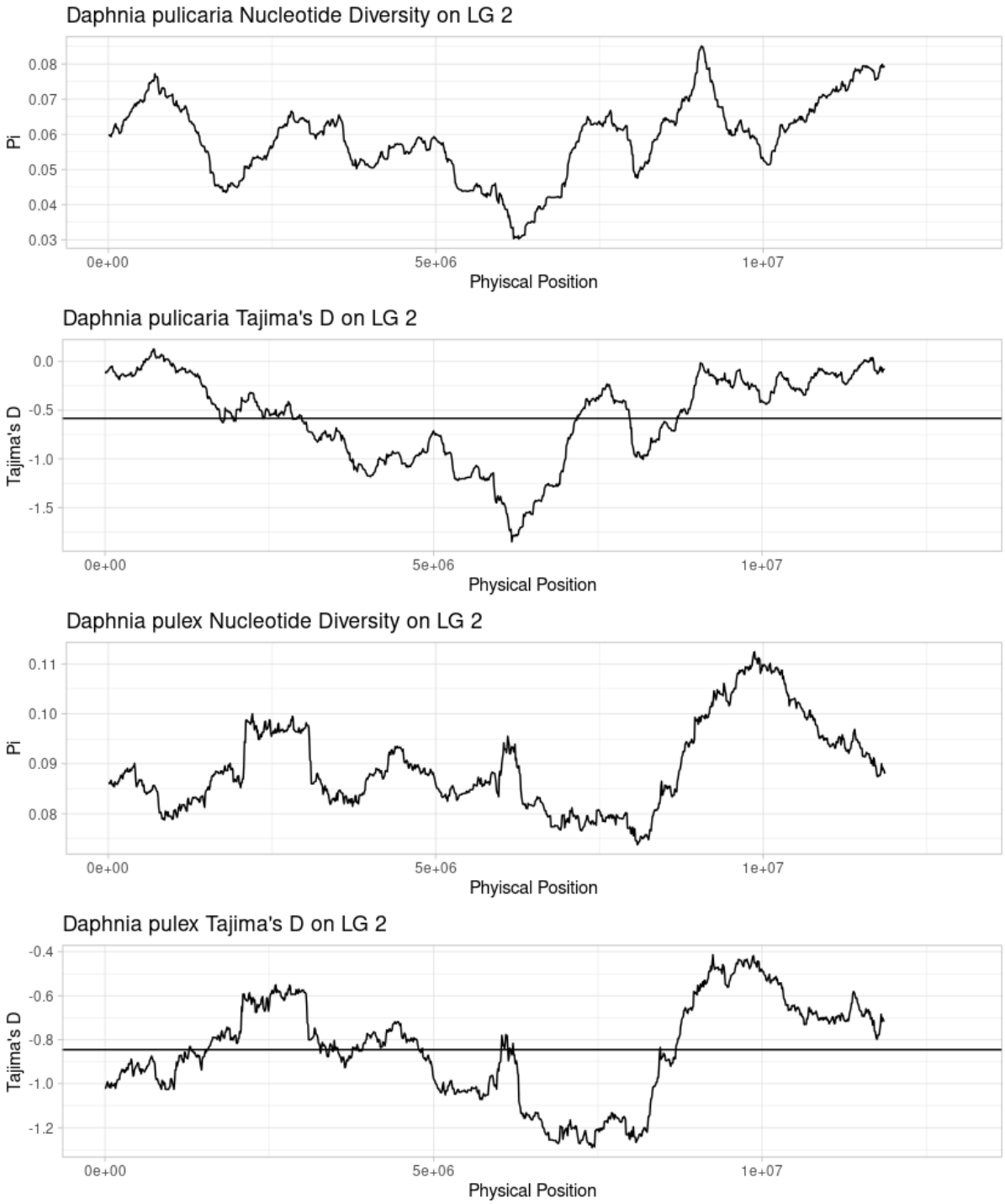


Figure S18: Sliding window nucleotide diversity and Tajima's D on Linkage Group (LG)-2 for *Daphnia pulicaria* and *Daphnia pulex.* Description same as above (Figure S17).

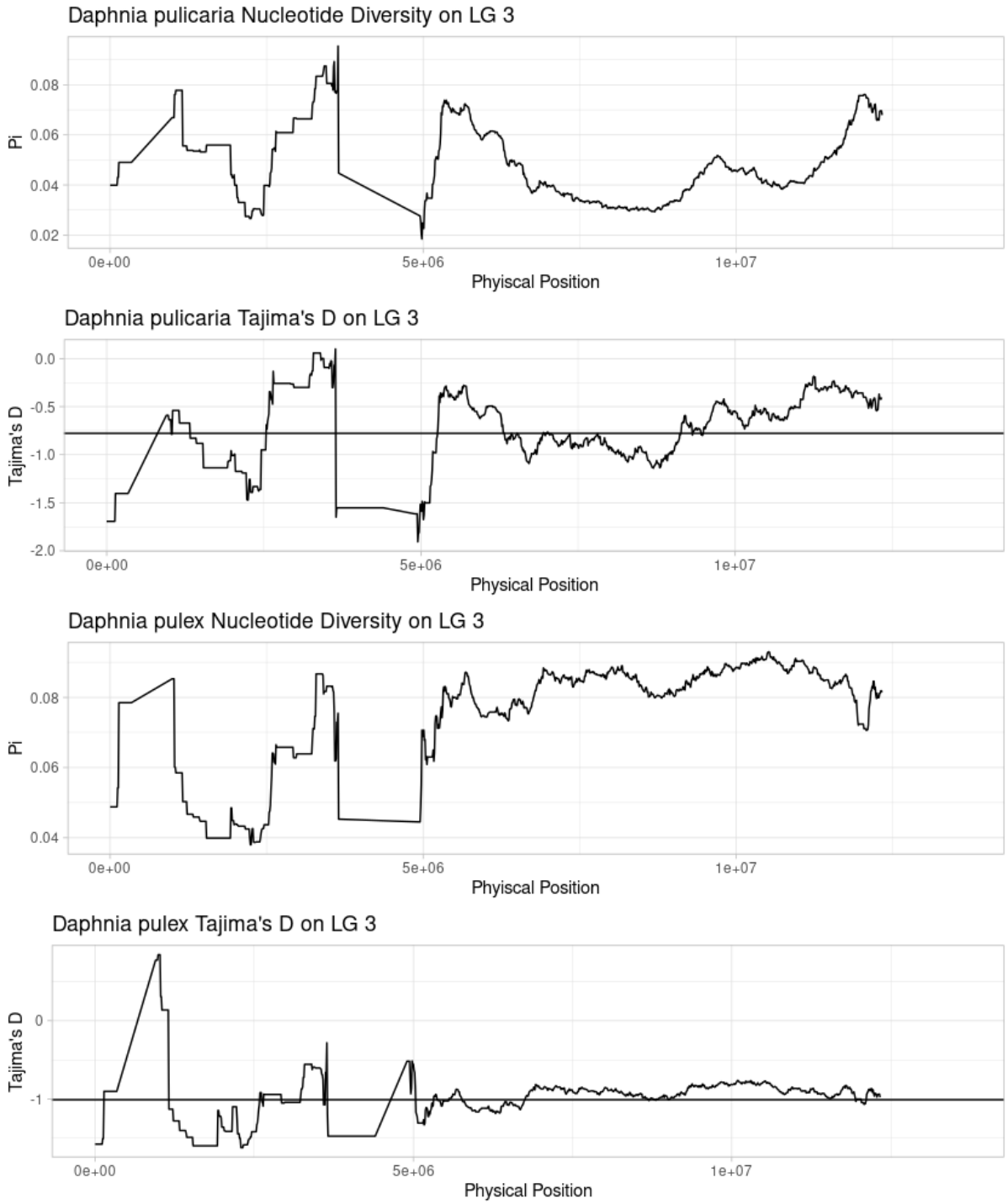


Figure S19: Sliding window nucleotide diversity and Tajima's D on Linkage Group (LG)-3 for *Daphnia pulicaria* and *Daphnia pulex.* Description same as above (Figure S17).

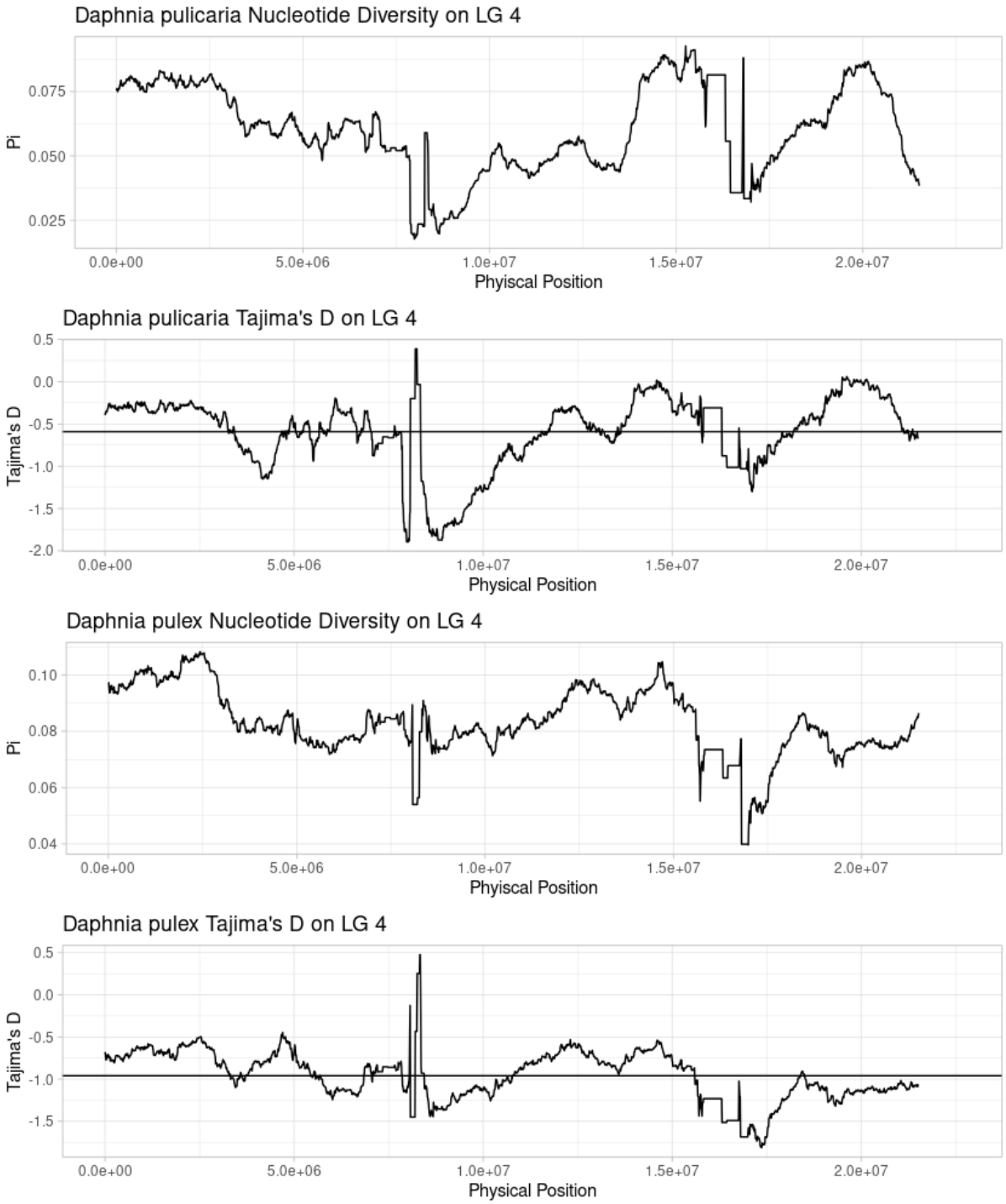


Figure S20: Sliding window nucleotide diversity and Tajima's D on Linkage Group (LG)-4 for *Daphnia pulicaria* and *Daphnia pulex.* Description same as above (Figure S17).

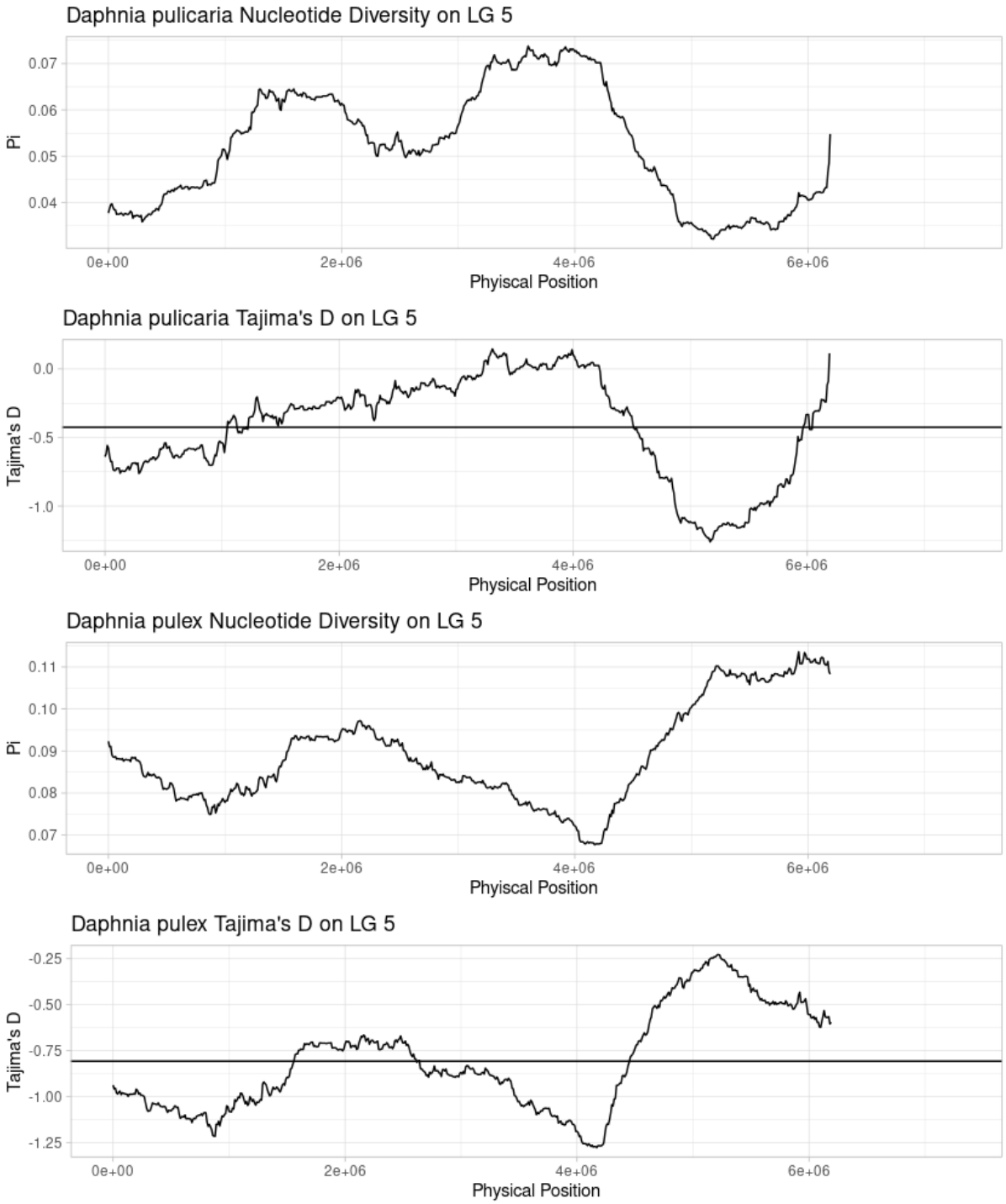


Figure S21: Sliding window nucleotide diversity and Tajima's D on Linkage Group (LG)-5 for *Daphnia pulicaria* and *Daphnia pulex.* Description same as above (Figure S17).

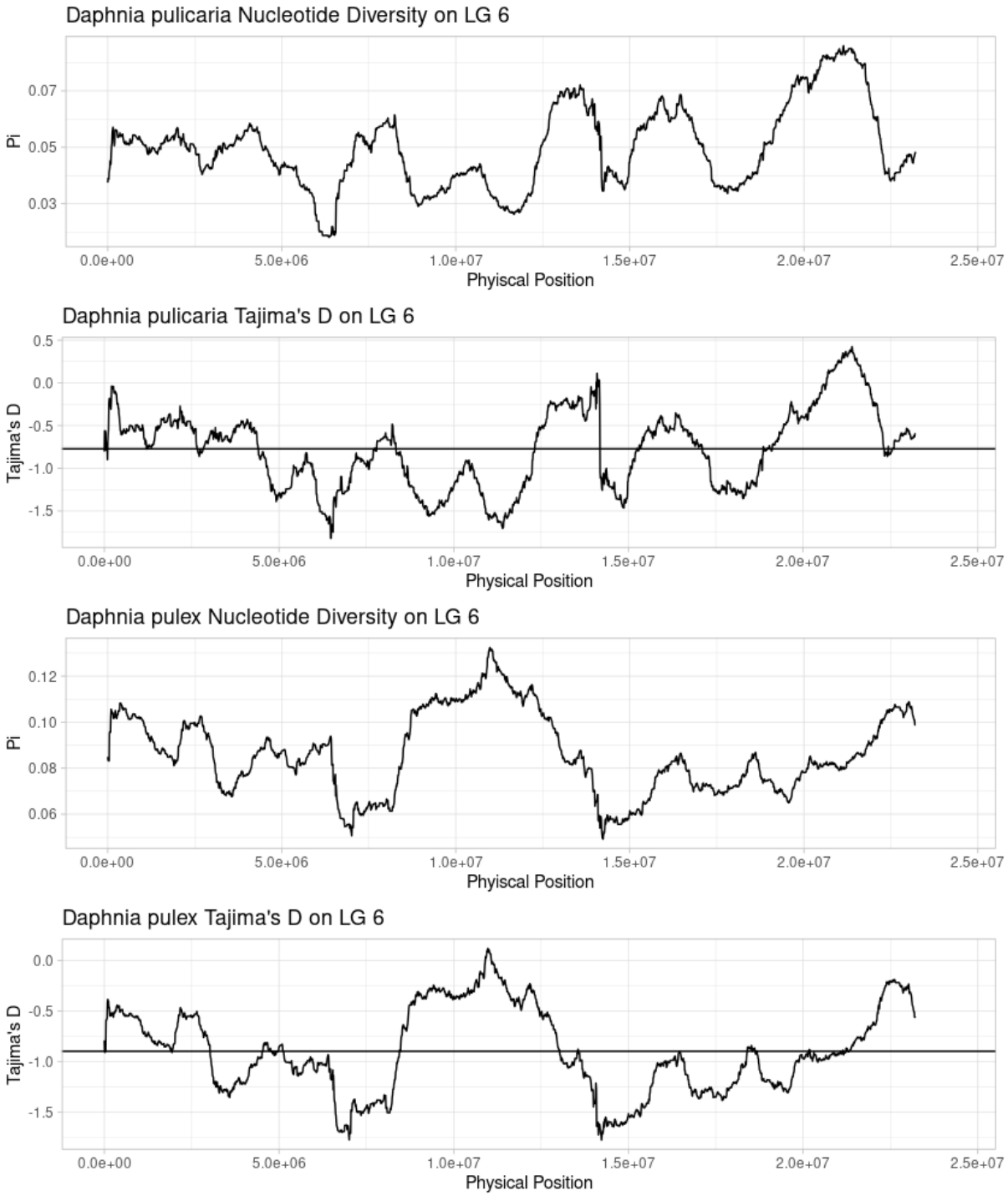


Figure S22: Sliding window nucleotide diversity and Tajima's D on Linkage Group (LG)-6 for *Daphnia pulicaria* and *Daphnia pulex.* Description same as above (Figure S17).

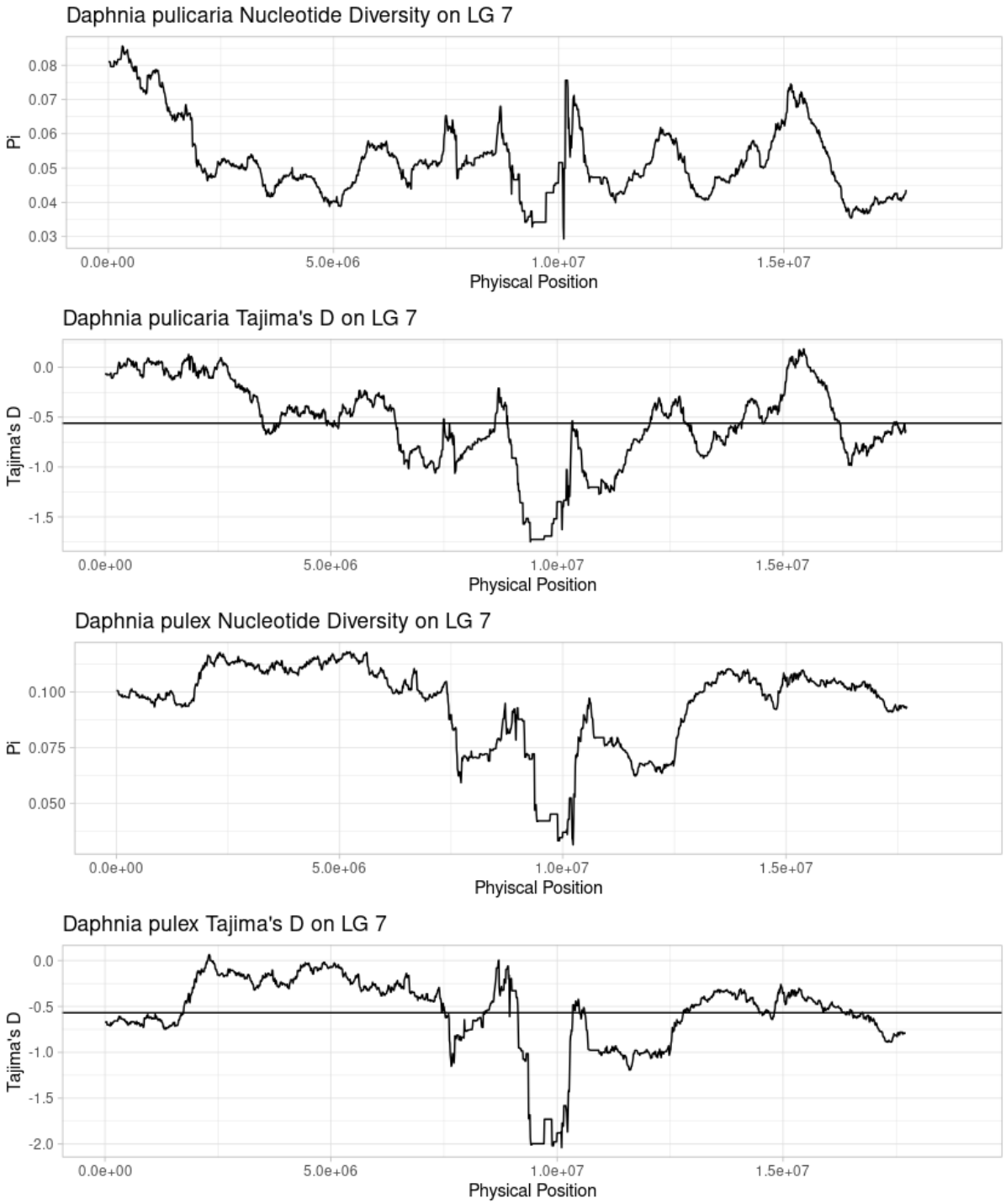


Figure S23: Sliding window nucleotide diversity and Tajima's D on Linkage Group (LG)-7 for *Daphnia pulicaria* and *Daphnia pulex.* Description same as above (Figure S17).

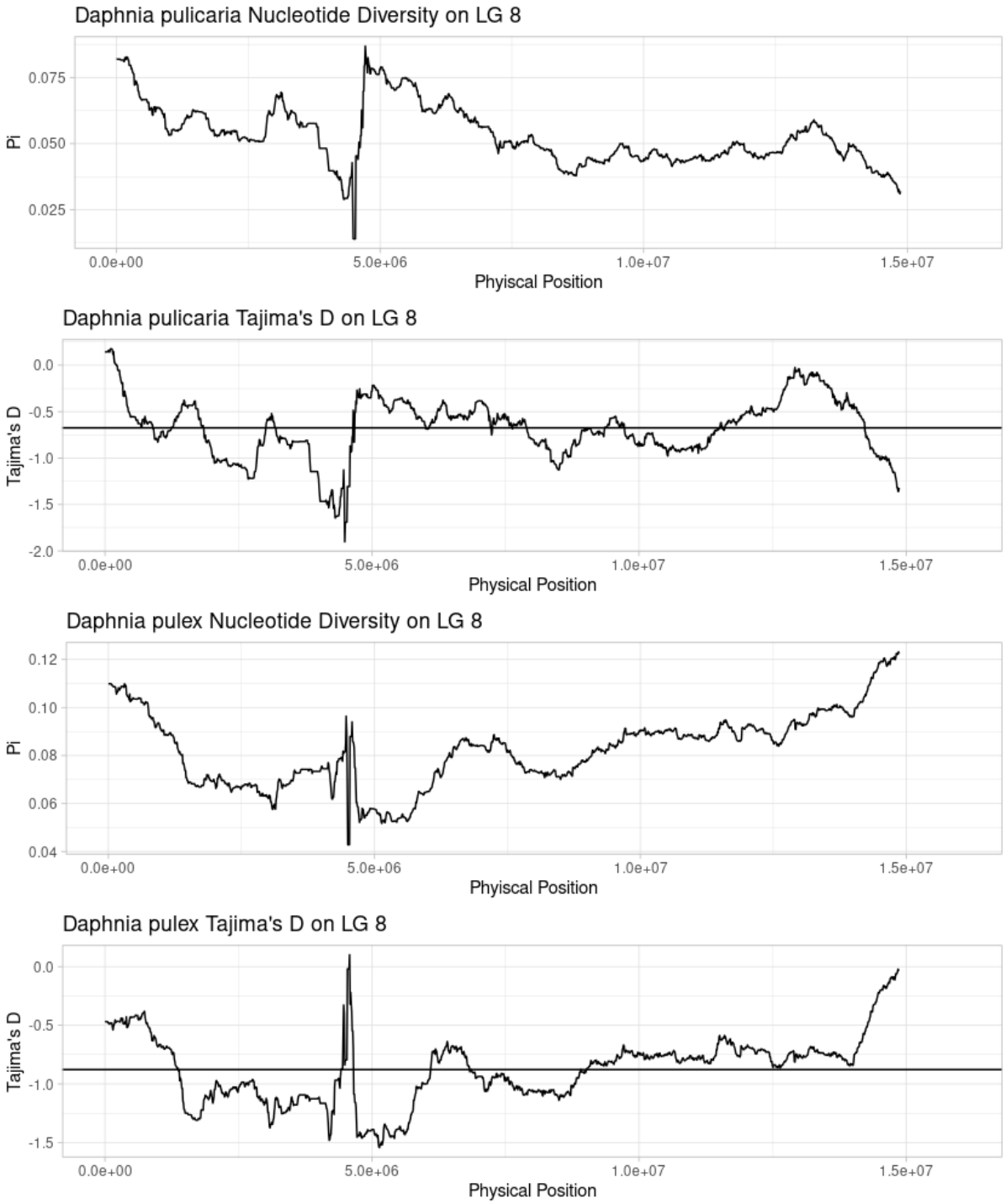


Figure S24: Sliding window nucleotide diversity and Tajima's D on Linkage Group (LG)-8 for *Daphnia pulicaria* and *Daphnia pulex.* Description same as above (Figure S17).

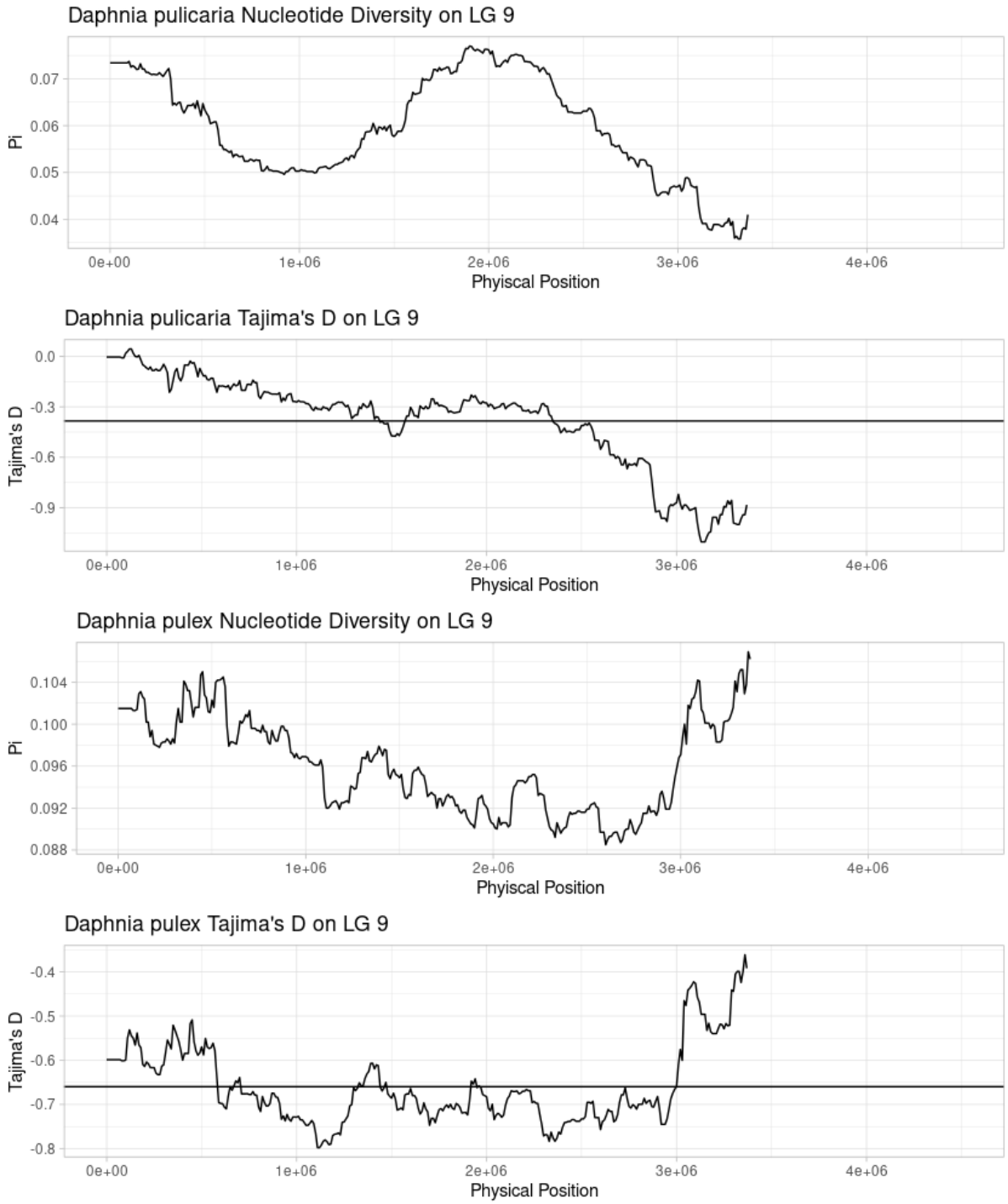


Figure S25: Sliding window nucleotide diversity and Tajima's D on Linkage Group (LG)-9 for *Daphnia pulicaria* and *Daphnia pulex.* Description same as above (Figure S17).

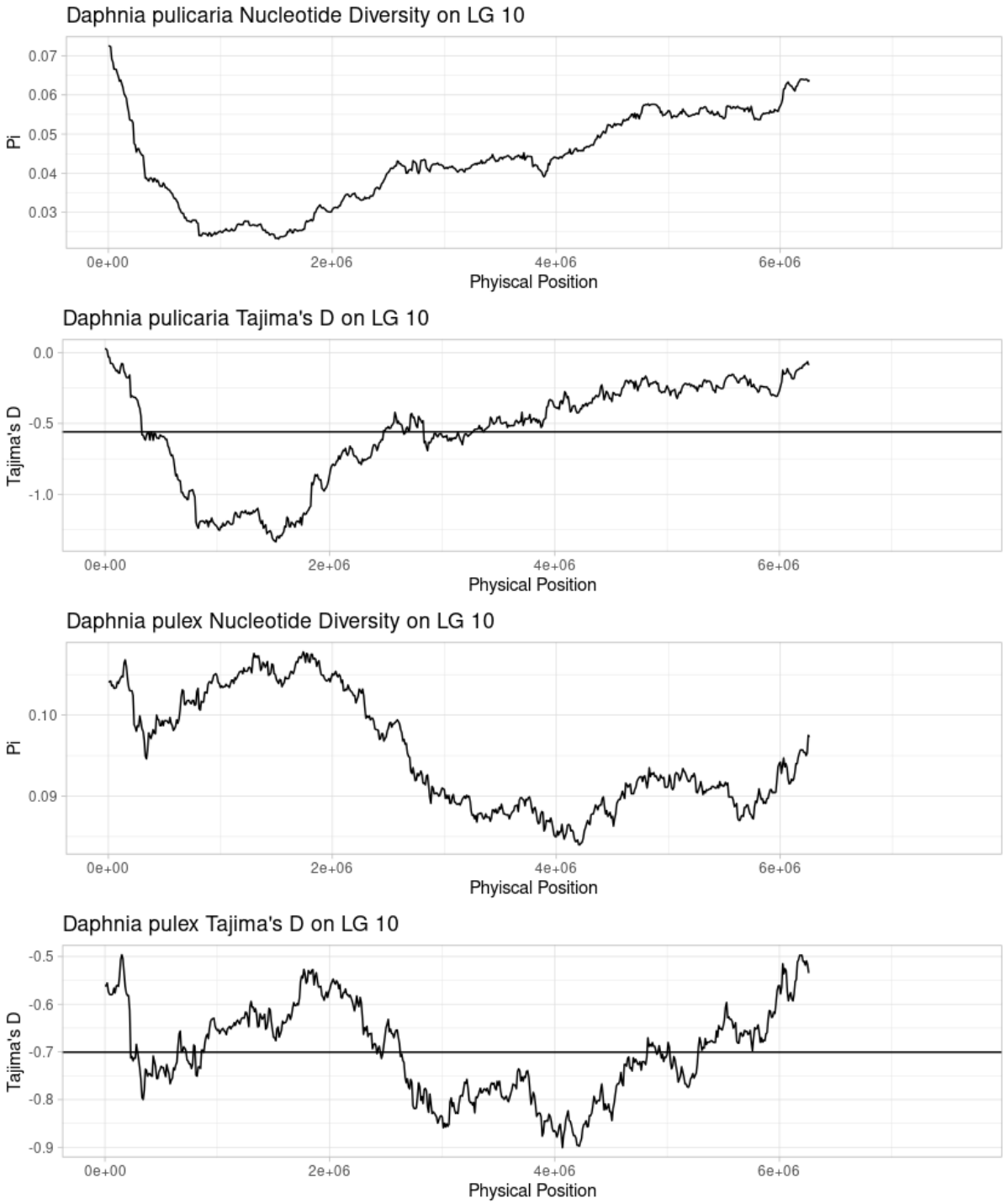


Figure S26: Sliding window nucleotide diversity and Tajima's D on Linkage Group (LG)-10 for *Daphnia pulicaria* and *Daphnia pulex.* Description same as above (Figure S17).

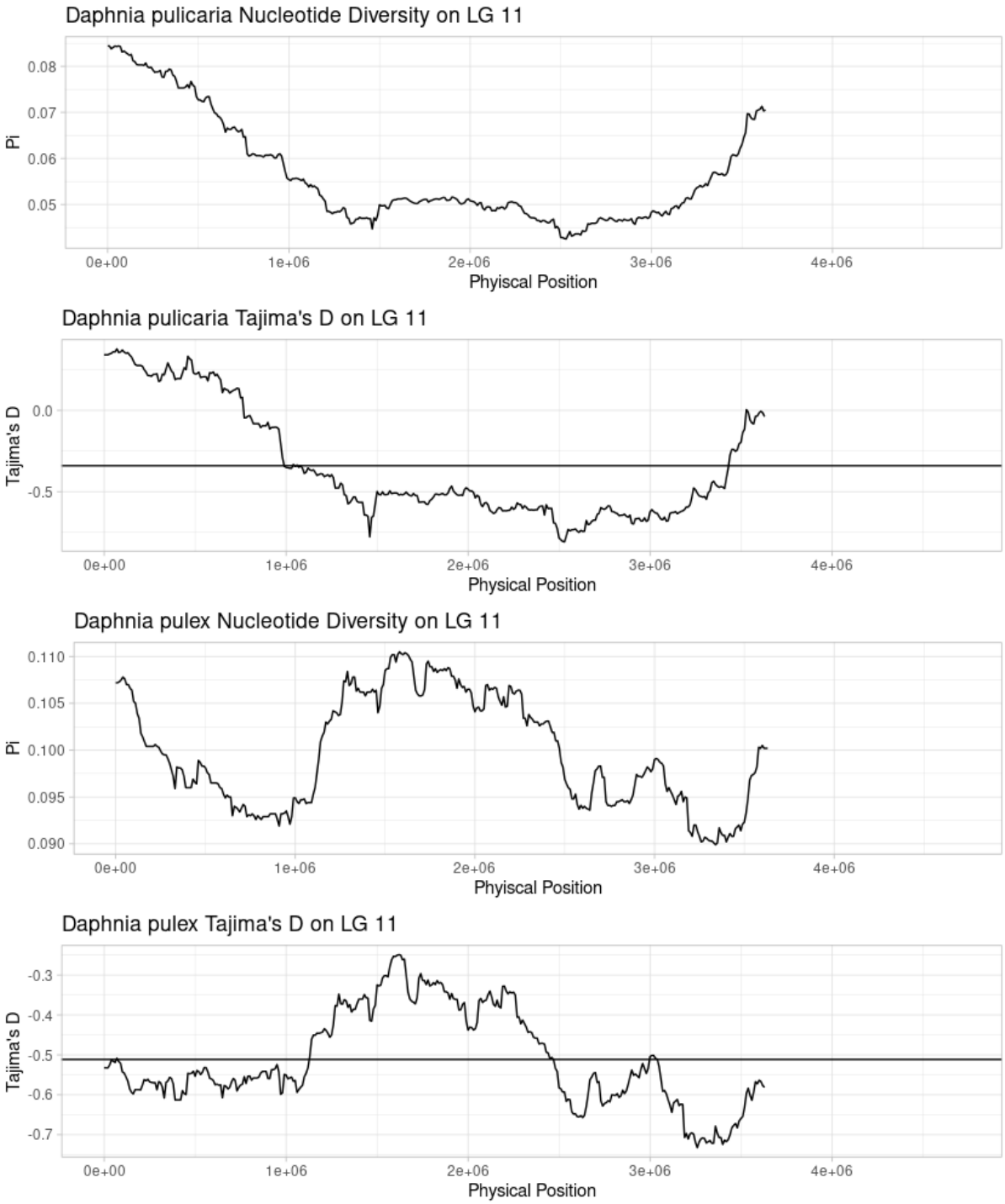


Figure S27: Sliding window nucleotide diversity and Tajima's D on Linkage Group (LG)-11 for *Daphnia pulicaria* and *Daphnia pulex.* Description same as above (Figure S17).

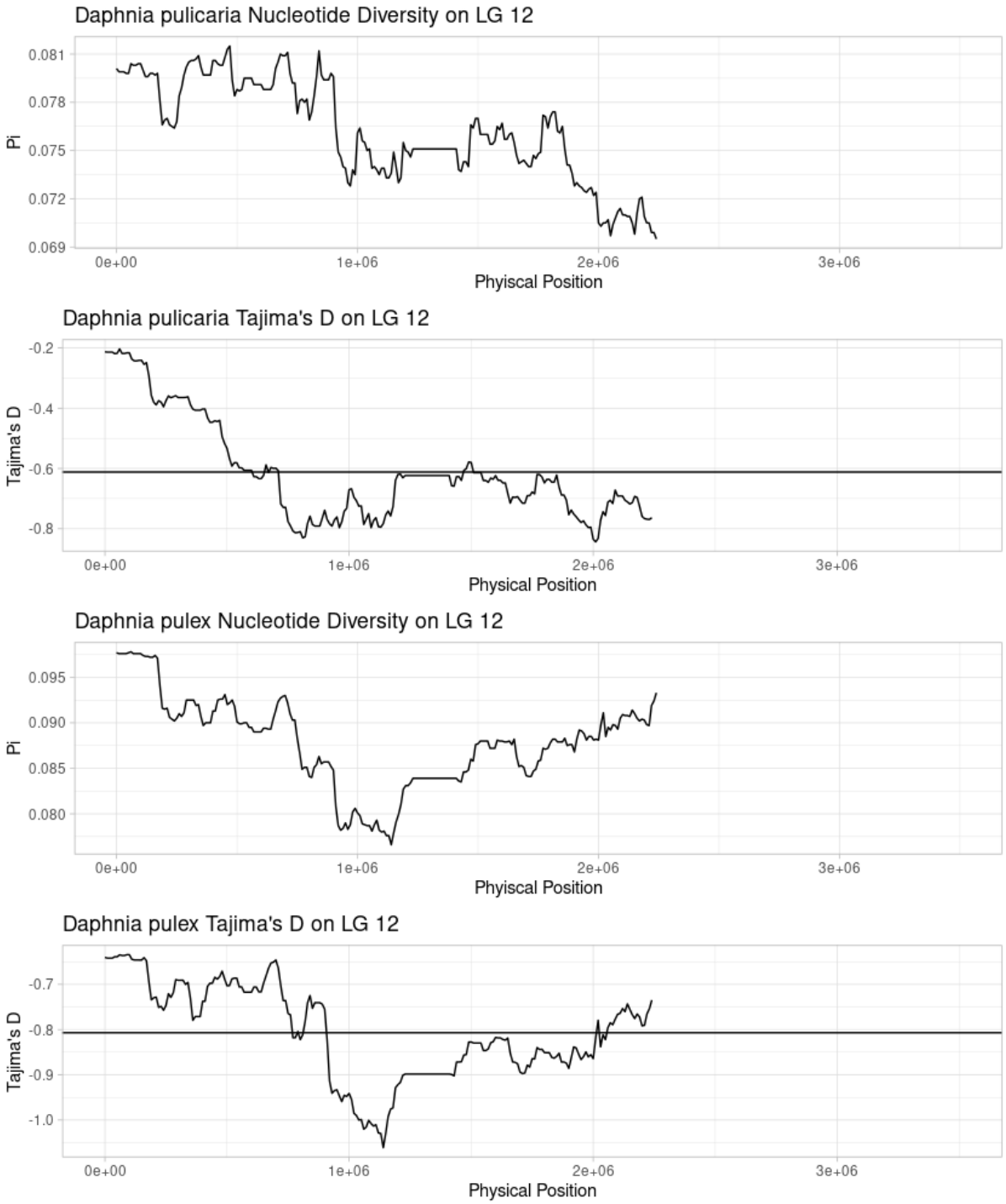


Figure S28: Sliding window nucleotide diversity and Tajima's D on Linkage Group (LG)-12 for *Daphnia pulicaria* and *Daphnia pulex.* Description same as above (Figure S17).

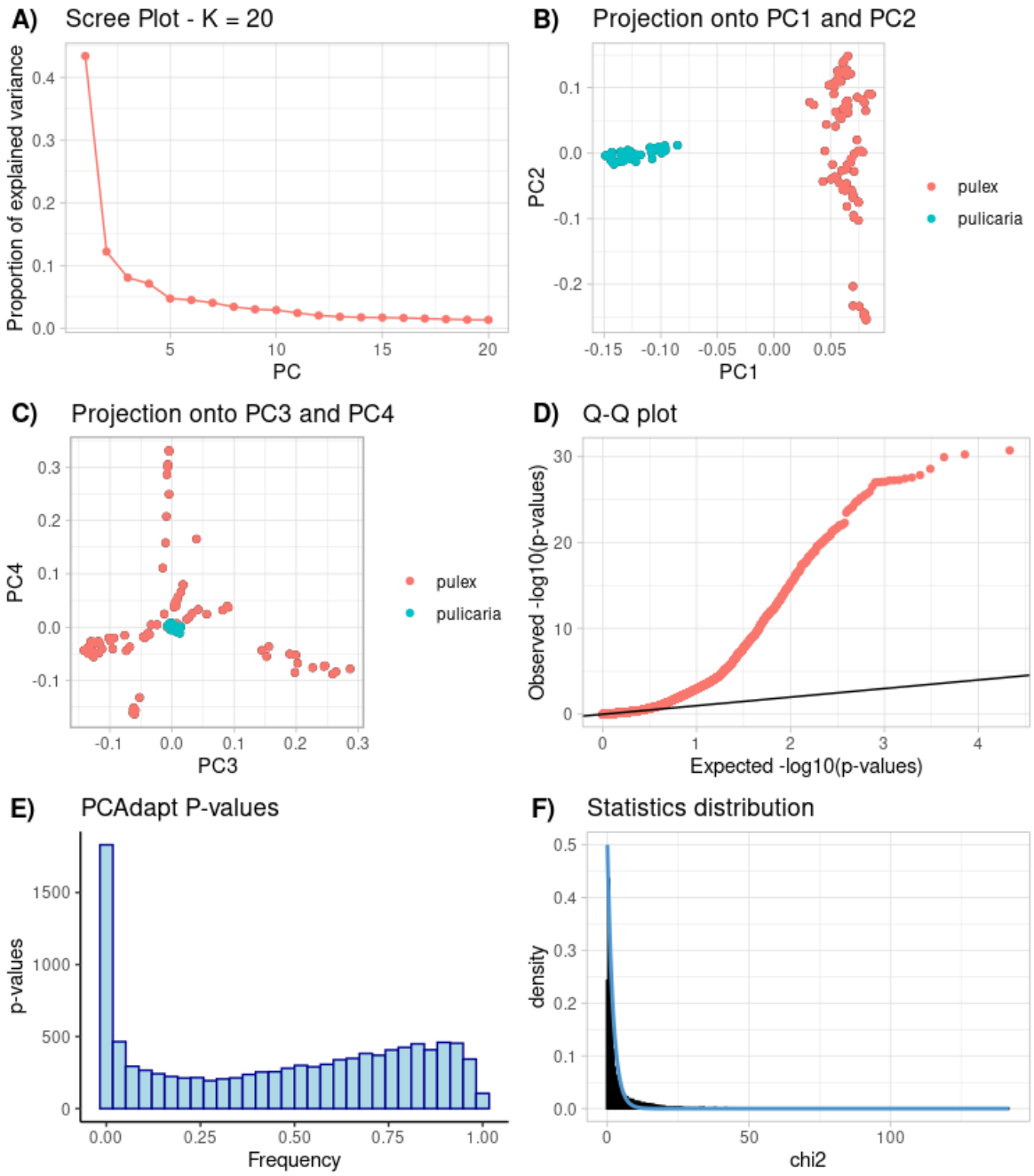


Figure S29: Supplementary Plots of PCAdapt (Privé et al. 2020) Outlier Tests. A) Scree plot for the Principal Components Analysis (PCA). The first 2 PCs explain the majority of the variance in the SNP data. B) Individuals plotted on PCs 1 and 2, clearly indicate the distinct groupings of *D. pulex* versus *D. pulicaria.* C) Projection of data onto PCs 3 and 4 indicate that there may be residual structure present in the *D. pulex* data. D) Q-Q plot of observed P-values for the test statistic. Deviation away from the slope of the expected P-values indicate the presence of outliers in the data. E) Test P-value histogram, inflation of P-values near zero indicate the presence of outliers. F) Fit of the observed χ^2 statistic distribution to the expected distribution.

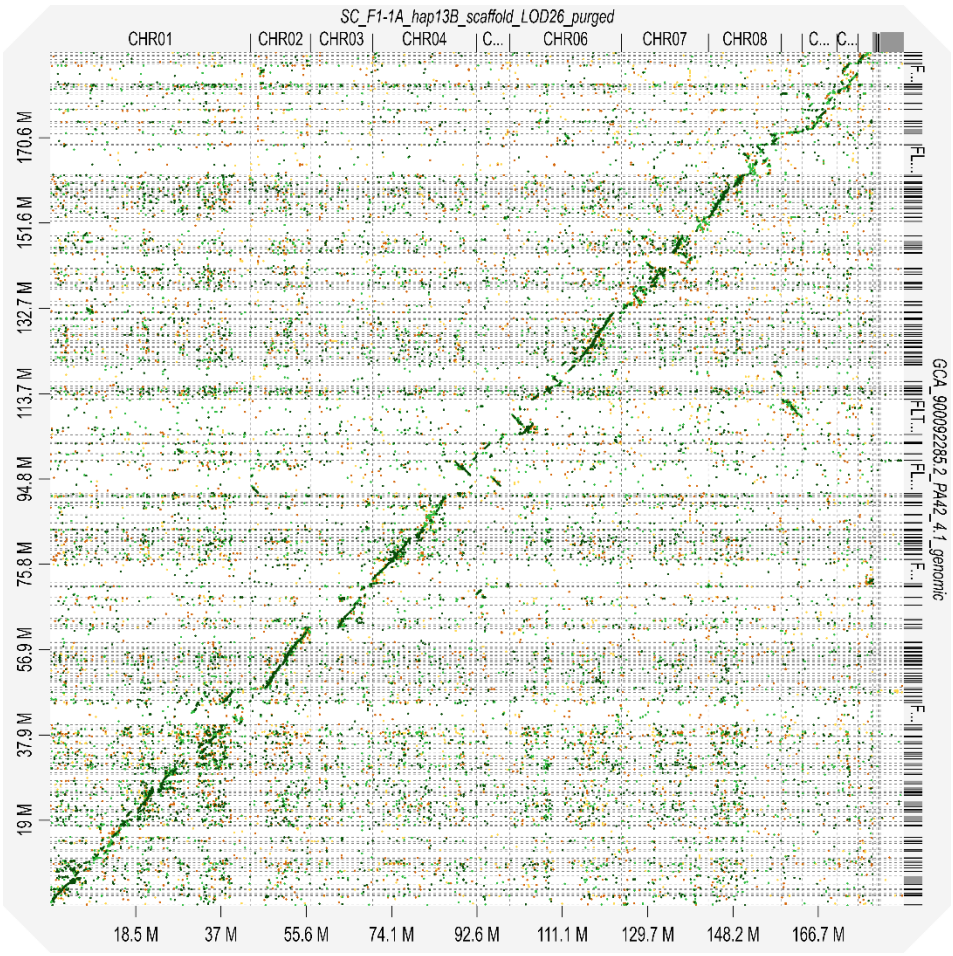


Figure S30: Dot plot of the whole genome alignment of SC F1-1A haplotype 13B to *Daphnia pulex* assembly PA42v4.1. Top axis: Scaffold IDs belonging to SC F1-1A haplotype 13B. Right: Alignment of scaffolds from PA42v4.1. Bottom: cumulative assembly length for PA42v4.1. Left: cumulative assembly length for SC F1-1A 13B. Color of alignment indicated pairwise alignment score. Scores with slopes of 1 demonstrate co-linearity of regions. Scores with slopes of -1 are probable rearrangements.

Species	NCBI Accession	Genome Type
Daphnia pulex	PRJNA11866	Mitochondrial genome
Daphnia pulex	GCA_900092285.2	Nuclear Genome
Daphnia magna	GCF_003990815.1	Nuclear Genome
Daphnia carinata	PRJNA300739*	Nuclear Genome
Daphnia dubia	GCA_013387435.1	Nuclear Genome

Table S1: *Daphnia* sequences used to filter contamination from Assembly. * The *Daphnia carinata* is no longer available on NCBI, however the raw sequencing reads are archived at the noted bioproject accession (see Jia et al. 2021)

file name: SC_F1-1A_hap13B_scaffold_LOD26_purged.fa

sequences: 160

Total length: 185226602 bp (185212202 bp excl N/X-runs)

GC level: 41.29 %

bases masked: 50605297 bp (27.32 %)

==========

	number of elements*	length occupied	percentage of sequence
Retroelements	7427	5830358 bp	3.15 %
SINEs:	0	0 bp	0.00 %
Penelope	391	378464 bp	0.20 %
LINEs:	2717	1738572 bp	0.94 %
CRE/SLACS	0	0 bp	0.00 %
L2/CR1/Rex	1573	526273 bp	0.28 %
R1/LOA/Jockey	262	282928 bp	0.15 %
R2/R4/NeSL	129	167694 bp	0.09 %
RTE/Bov-B	0	0 bp	0.00 %
L1/CIN4	0	0 bp	0.00 %
LTR elements:	4710	4091786 bp	2.21 %
BEL/Pao	776	1038509 bp	0.56 %
Ty1/Copia	1029	493453 bp	0.27 %
Gypsy/DIRS1	2905	2559824 bp	1.38 %
Retroviral	0	0 bp	0.00 %
DNA transposons	1803	781358 bp	0.42 %
hobo-Activator	168	67673 bp	0.04 %
Tc1-IS630-Pogo	197	112120 bp	0.06 %
En-Spm	0	0 bp	0.00 %
MuDR-IS905	0	0 bp	0.00 %

PiggyBac 64 87732 bp 0.05 %

Tourist/Harbinger 0 0 bp 0.00 %

Other (Mirage, 0 0 bp 0.00 %

P-element, Transib)

Rolling-circles 6322 8830819 bp 4.77 %

Unclassified: 98861 35162762 bp 18.98 %

Total interspersed repeats: 41774478 bp 22.55 %

Small RNA: 0 0 bp 0.00 %

Satellites: 0 0 bp 0.00 %

Simple repeats: 0 0 bp 0.00 %

Low complexity: 0 0 bp 0.00 %

==========

Table S2: Repeatmasker output for Haplotype 13B (primary haplotype). *Most repeats fragmented by insertions or deletions have been counted as one element.

==========

file name: SC_F1-1A_hap2X_scaffold_ragtag_purged.fa

sequences: 205

total length: 179420049 bp (179392649 bp excl N/X-runs)

GC level: 41.25 %

bases masked: 48806325 bp (27.20 %)

==========

	number of elements*	length occupied	percentage of sequence
Retroelements	7299	5972014 bp	3.33 %
SINEs:	0	0 bp	0.00 %
Penelope	635	558944 bp	0.31 %
LINEs:	2515	1796601 bp	1.00 %
CRE/SLACS	0	0 bp	0.00 %
L2/CR1/Rex	1144	486799 bp	0.27 %
R1/LOA/Jockey	256	262670 bp	0.15 %
R2/R4/NeSL	141	174768 bp	0.10 %
RTE/Bov-B	0	0 bp	0.00 %
L1/CIN4	0	0 bp	0.00 %
LTR elements:	4784	4175413 bp	2.33 %
BEL/Pao	792	1005151 bp	0.56 %
Ty1/Copia	994	492221 bp	0.27 %
Gypsy/DIRS1	2998	2678041 bp	1.49 %
Retroviral	0	0 bp	0.00 %
DNA transposons	1848	855237 bp	0.48 %
hobo-Activator	172	85740 bp	0.05 %
Tc1-IS630-Pogo	224	130960 bp	0.07 %
En-Spm	0	0 bp	0.00 %
MuDR-IS905	0	0 bp	0.00 %

PiggyBac 64 83495 bp 0.05 %

Tourist/Harbinger 0 0 bp 0.00 %

Other (Mirage, 0 0 bp 0.00 %

P-element, Transib)

Rolling-circles 6355 8463072 bp 4.72 %

Unclassified: 96413 33516002 bp 18.68 %

Total interspersed repeats: 40343253 bp 22.49 %

Small RNA: 0 0 bp 0.00 %

Satellites: 0 0 bp 0.00 %

Simple repeats: 0 0 bp 0.00 %

Low complexity: 0 0 bp 0.00 %

==========

Table S3: Repeatmasker output for Haplotype 2X (secondary haplotype). *Most repeats fragmented by insertions or deletions have been counted as one element.

Feature	Haplotype 13B	Haplotype 2X
Mean Gene Length (SE)	2749.6 (28.2)	2747.0 (27.4)
Mean Intron Length (SE)	292.9 (3.5)	294.9 (3.5)
Mean Exon Length (SE)	247.4 (0.97)	244.1 (0.97)
Mean Inter-gene Space (SE)*	4862.1 (131.2)	4694.9 (110.0)

Table S4: Gene Structure Information by Haplotype. All lengths are in base pairs (BP). *Only genes placed on chromosomes were used to calculate inter-gene length. Omitting 197 genes from the 13B assembly and 584 genes from the 2X assembly

Chr	**01**	**02**	**03**	**04**	**05**	**06**	**07**	**08**	**09**	**10**	**11**	**12**
$\mathbf{F_{st}}$	0.198	0.175	0.278	0.164	0.275	0.221	0.230	0.318	0.262	0.357	0.306	0.096
$\mathbf{D_{xy}}$	0.111	0.116	0.129	0.107	0.136	0.120	0.129	0.148	0.145	0.167	0.164	0.102
$\boldsymbol{\pi}_{\mathbf{pa}}$	0.057	0.059	0.049	0.060	0.052	0.050	0.054	0.053	0.060	0.043	0.057	0.057
$\boldsymbol{\pi}_{\mathbf{px}}$	0.076	0.089	0.075	0.083	0.089	0.086	0.094	0.082	0.096	0.095	0.099	0.095
$\mathbf{D_{pa}}$	-0.655	-0.584	-0.775	-0.591	-0.425	-0.772	-0.560	-0.674	-0.383	-0.559	-0.340	-0.612
$\mathbf{D_{px}}$	-1.020	-0.845	-1.010	-0.960	-0.808	-0.899	-0.569	-0.878	-0.650	-0.700	-0.512	-0.807

Table S5: Population Genetic Summary Statistics per chromosome. All summary statistics are presented as chromosome-wide means calculated using sliding windows (see methods).

Analysis Type: PANTHER Overrepresentation Test (Released 20210224)

Annotation Version and Release Date: GO Ontology database DOI: 10.5281/zenodo.4495804 Released 2021-02-01

Analyzed List: upload_1 (Daphnia pulex)

Reference List: Daphnia pulex (all genes in database)

Test Type: FISHER

Correction: BONFERRONI

Bonferroni count: 1814

GO biological process complete	Daphnia pulex - REFLIST (30012)	upload_1 (152) (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)		upload_1 (P-value)
Unclassified (UNCLASSIFIED)	18853	17	95.48	-	.18	0.00E00
biological process (GO:0008150)	11159	135	56.52	+	2.39	1.52E-36
one-carbon metabolic process (GO:0006730)	50	25	.25	+	98.72	3.53E-36
cellular process (GO:0009987)	9112	122	46.15	+	2.64	2.15E-33
metabolic process (GO:0008152)	6119	83	30.99	+	2.68	4.14E-17
small molecule metabolic process (GO:0044281)	797	32	4.04	+	7.93	3.72E-16
microtubule-based movement (GO:0007018)	110	16	.56	+	28.72	6.21E-15
cellular metabolic process (GO:0044237)	5239	69	26.53	+	2.60	3.49E-12
movement of cell or subcellular component (GO:0006928)	224	16	1.13	+	14.10	1.86E-10
microtubule-based process (GO:0007017)	335	16	1.70	+	9.43	5.97E-08
retrograde transport, endosome to plasma membrane (GO:1990126)	13	6	.07	+	91.13	6.91E-07
mitotic nuclear division (GO:0140014)	57	8	.29	+	27.71	2.48E-06
phototransduction (GO:0007602)	52	7	.26	+	26.58	3.57E-05
detection of light stimulus (GO:0009583)	52	7	.26	+	26.58	3.57E-05
detection of abiotic stimulus (GO:0009582)	52	7	.26	+	26.58	3.57E-05
detection of external stimulus (GO:0009581)	52	7	.26	+	26.58	3.57E-05
vesicle-mediated transport to the plasma membrane (GO:0098876)	32	6	.16	+	37.02	6.50E-05
detection of stimulus (GO:0051606)	61	7	.31	+	22.66	9.76E-05

lipid catabolic process (GO:0016042)	136	9	.69	+	13.07	1.01E-04	
endocytic recycling (GO:0032456)	16	5	.08	+	61.70	1.05E-04	
transport (GO:0006810)	1983	30	10.04	+	2.99	1.34E-04	
organic substance metabolic process (GO:0071704)	5806	58	29.41	+	1.97	1.48E-04	
establishment of localization (GO:0051234)	2005	30	10.15	+	2.95	1.69E-04	
response to light stimulus (GO:0009416)	70	7	.35	+	19.74	2.33E-04	
mitotic cell cycle (GO:0000278)	208	10	1.05	+	9.49	3.15E-04	
response to radiation (GO:0009314)	76	7	.38	+	18.19	3.93E-04	
localization (GO:0051179)	2218	31	11.23	+	2.76	4.41E-04	
response to external stimulus (GO:0009605)	219	10	1.11	+	9.02	4.96E-04	
mitotic cell cycle process (GO:1903047)	170	9	.86	+	10.45	6.13E-04	
cell cycle (GO:0007049)	388	12	1.97	+	6.11	1.75E-03	
nuclear division (GO:0000280)	154	8	.78	+	10.26	3.21E-03	
mitotic spindle assembly (GO:0090307)	14	4	.07	+	56.41	3.26E-03	
primary metabolic process (GO:0044238)	5575	53	28.24	+	1.88	3.39E-03	
organelle assembly (GO:0070925)	216	9	1.09	+	8.23	4.12E-03	
intracellular transport (GO:0046907)	591	14	2.99	+	4.68	4.58E-03	
organelle fission (GO:0048285)	162	8	.82	+	9.75	4.61E-03	
mitotic sister chromatid segregation (GO:0000070)	40	5	.20	+	24.68	5.73E-03	
DNA biosynthetic process (GO:0071897)	44	5	.22	+	22.44	8.80E-03	
ion transport (GO:0006811)	1466	22	7.42	+	2.96	1.06E-02	
endosomal transport (GO:0016197)	83	6	.42	+	14.27	1.11E-02	
response to abiotic stimulus (GO:0009628)	130	7	.66	+	10.63	1.16E-02	
establishment of localization in cell (GO:0051649)	688	14	3.48	+	4.02	2.47E-02	
spindle assembly (GO:0051225)	28	4	.14	+	28.21	3.62E-02	

Table S6: Output of Panther (Mi et al. 2021) Gene Ontology (GO) enrichment test.

Works Cited:

Cabanettes, F., and C. Klopp. 2018. D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. PeerJ **2018**: e4958. doi:10.7717/PEERJ.4958/TABLE-2

Danecek, P., A. Auton, G. Abecasis, and others. 2011. The variant call format and VCFtools. Bioinformatics **27**: 2156–2158. doi:10.1093/bioinformatics/btr330

Danecek, P., J. K. Bonfield, J. Liddle, and others. 2021. Twelve years of SAMtools and BCFtools. Gigascience **10**: 1–4. doi:10.1093/gigascience/giab008

Excoffier, L., I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa, and M. Foll. 2013. Robust Demographic Inference from Genomic and SNP Data. PLoS Genet. **9**: 1003905. doi:10.1371/journal.pgen.1003905

Garrison, E., and G. Marth. 2012. Haplotype-based variant detection from short-read sequencing.

Kilham, S. S., D. A. Kreeger, S. G. Lynn, C. E. Goulden, and L. Herrera. 1998. COMBO: a defined freshwater culture medium for algae and zooplankton. Hydrobiologia **377**: 147–159. doi:10.1023/A:1003231628456

Kriventseva, E. V., D. Kuznetsov, F. Tegenfeldt, M. Manni, R. Dias, F. A. Simão, and E. M. Zdobnov. 2019. OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Res. **47**: D807–D811. doi:10.1093/nar/gky1053

Li, H. 2018. Minimap2: pairwise alignment for nucleotide sequences I. Birol [ed.]. Bioinformatics **34**: 3094–3100. doi:10.1093/bioinformatics/bty191

Mi, H., D. Ebert, A. Muruganujan, C. Mills, L. P. Albou, T. Mushayamaha, and P. D. Thomas. 2021. PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res. **49**: D394–D403. doi:10.1093/nar/gkaa1106

Privé, F., K. Luu, B. J. Vilhjálmsson, M. G. B. Blum, and M. Rosenberg. 2020. Performing Highly Efficient Genome Scans for Local Adaptation with R Package pcadapt Version 4. Mol. Biol. Evol. **37**: 2153–2154. doi:10.1093/molbev/msaa053

Rastas, P. 2017. Lep-MAP3: Robust linkage mapping even for low-coverage whole genome sequencing data. Bioinformatics **33**: 3726–3732. doi:10.1093/bioinformatics/btx494

Rastas, P. 2020. Lep-Anchor: Automated construction of linkage map anchored haploid genomes. Bioinformatics **36**: 2359–2364. doi:10.1093/bioinformatics/btz978

Sherman, R. E., R. Hartnett, E. L. Kiehnau, L. J. Weider, and P. D. Jeyasingh. 2021. Quantitative genetics of phosphorus content in the freshwater herbivore, *Daphnia pulicaria* S. Plaistow [ed.]. J. Anim. Ecol. **00**: 1365-2656.13419. doi:10.1111/1365-2656.13419

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31**: 3210–3212. doi:10.1093/bioinformatics/btv351

Ye, Z., S. Xu, K. Spitze, and others. 2017. A new reference genome assembly for the microcrustacean Daphnia pulex. G3 Genes, Genomes, Genet. **7**: 1405–1416. doi:10.1534/g3.116.038638

Appendix 2

Supplemental Materials for Chapter 2:

Does salinization impact long-term *Daphnia* assemblage dynamics? Evidence from the sediment egg bank in a small hard-water lake

Matthew J. Wersebe[1, †], Mark B. Edlund[2], and Lawrence J. Weider[1]

[1]Program in Ecology and Evolutionary Biology, Department of Biology, University of Oklahoma, Norman, OK

[2]St. Croix Watershed Research Station, Science Museum of Minnesota, Marine on St Croix, Minnesota

[†]matthew.wersebe@ou.edu; 730 Van Vleet Oval, Rm 314 Norman, OK 73019

Supplemental Methods

Field Methods:

We used a YSI EXO1 multiparameter sonde (Yellowsprings, OH) to collect temperature, dissolved oxygen, and conductivity profiles over the deepest part of the lake basin (Fig 1 A-C in main manuscript).

Loss-on-Ignition and Core Dating:

As noted in the main text (Methods), the second core underwent loss-on-ignition (LOI) analysis following standard protocols (Heiri et al. 2001). Briefly, wet sediment was massed on an analytical balance, dried overnight at 100°C and sequentially burned in a muffle furnace at 550 °C and 1000 °C to determine water content, bulk density, % organic-carbon, % $CaCO_3$, and % inorganic content (Fig. S1 A-C). The LOI profile of the 2019 core was used to estimate the age of each core section by aligning unique sediment geochemistry markers to similar markers from a $^{210}Pb/^{137}Cs$ radiometric dating model from a core collected in January 2015 from the same location in Tanners Lake (Blumentritt et al. 2013). The January 2015 sediment core was previously analyzed for LOI following standard protocols (Heiri et al. 2001) and for ^{210}Pb activity to determine a sediment depth-age model and sediment accumulation rates for the past 150 to 200 years (Table S1). Lead-210 activity was measured from its daughter product, ^{210}Po, which is considered to be in secular equilibrium with the parent isotope. Aliquots of freeze-dried sediment were spiked with a known quantity of ^{209}Po as an internal yield tracer and the isotopes distilled at 550°C after treatment with concentrated HCl. Polonium isotopes were then directly plated onto silver planchets from a 0.5 N HCl solution. Activity was measured for 1-3 x 10^5 s using an Ortec alpha spectrometry system. Supported ^{210}Pb was estimated by mean activity in the

lowest core samples and subtracted from up-core activity to calculate unsupported ^{210}Pb. Core dates and sedimentation rates were calculated using the constant rate of supply model (Appleby and Oldfield 1978, Appleby 2001). Dating and sedimentation errors represented first-order propagation of counting uncertainty (Binford 1990).

Estimation of Water Quality Trends:

Approximately 30 years of chloride concentration and specific conductance data were downloaded from the National Water Quality Monitoring Council online portal (waterqualitydata.us) for Tanners Lake (Monitoring site ID: MNPCA-82-0115-00-100, 101 & 201). Specific conductance was converted to chloride concentration using the equation provided by Novotny et al. (2008; Equation (1), see also Fig. 3-D). Surface water chloride concentration was estimated by linear interpolation from 1988 through 1950. Lake chloride concentrations are thought to have begun to increase starting in 1950 with the wide-spread use of road deicing salts in this region (Kelly et al. 2010). Prior to 1950, chloride concentration was assumed to reflect the background levels (1-2 mg L^{-1}) estimated with a fossil diatom-inferred Cl^- model by Ramstack et al. (2004).

(1) $[Cl^-] = 0.26 * \text{Specific Conductance} - 37.25$ (Novotny et al. 2008)

Supplement Figures and Tables:

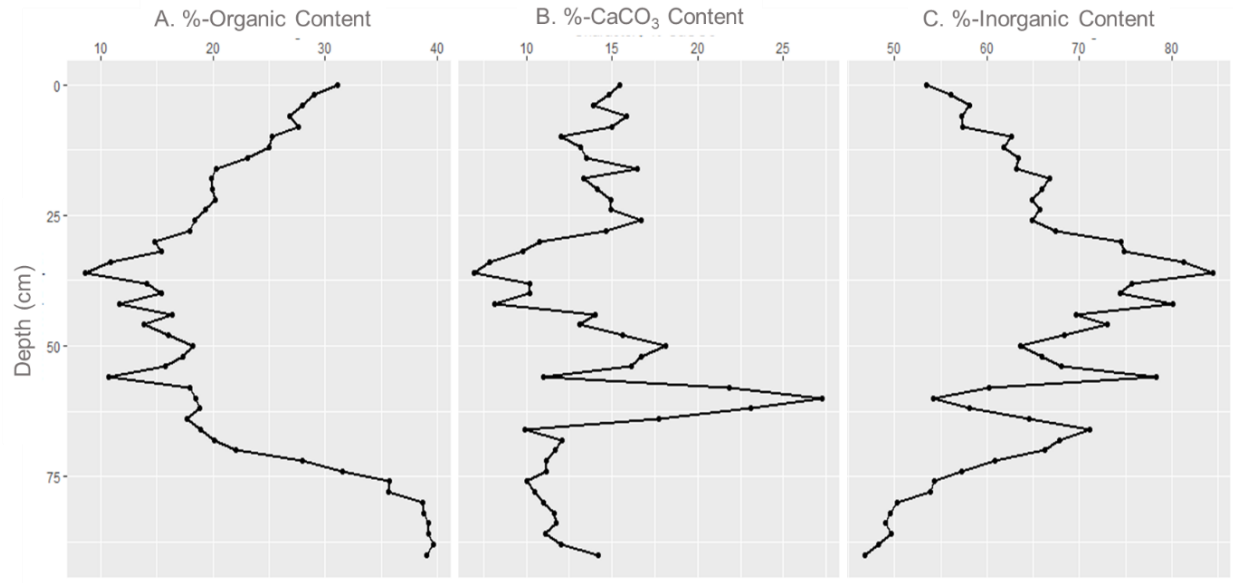


Figure S1: Geochemical attributes determined by Loss-on-Ignition of the July 2019 core: (A) %-Organic Carbon content (B) %-$CaCO_3$ content, and (C) %-Inorganic Content.

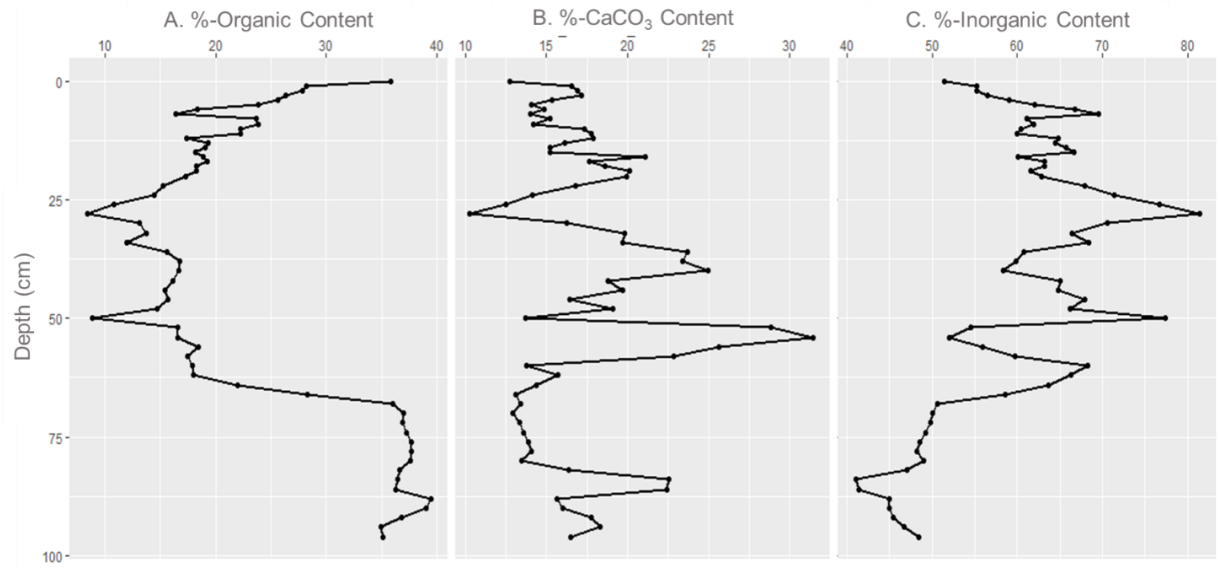


Figure S2: Geochemical attributes determined by Loss-on-Ignition (LOI) of the January 2015 Core: (A) %-Organic Carbon content (B) %-$CaCO_3$ content, and (C) %-Inorganic Content. Our 2019 LOI results (Fig S1 A-C above) were aligned to the major peaks observed in this data set.

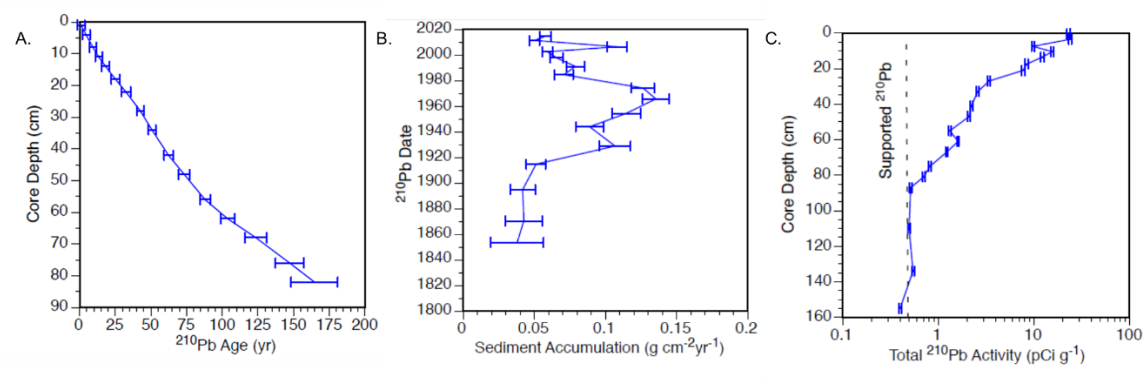


Figure S3: Results of the January 2015 Tanners Lake Pb-210 dating model: (A) Estimated age of layers based on the dating, (B) estimated sediment accumulation rates across the core, and (C) Pb-210 activity observed for each core layer tested.

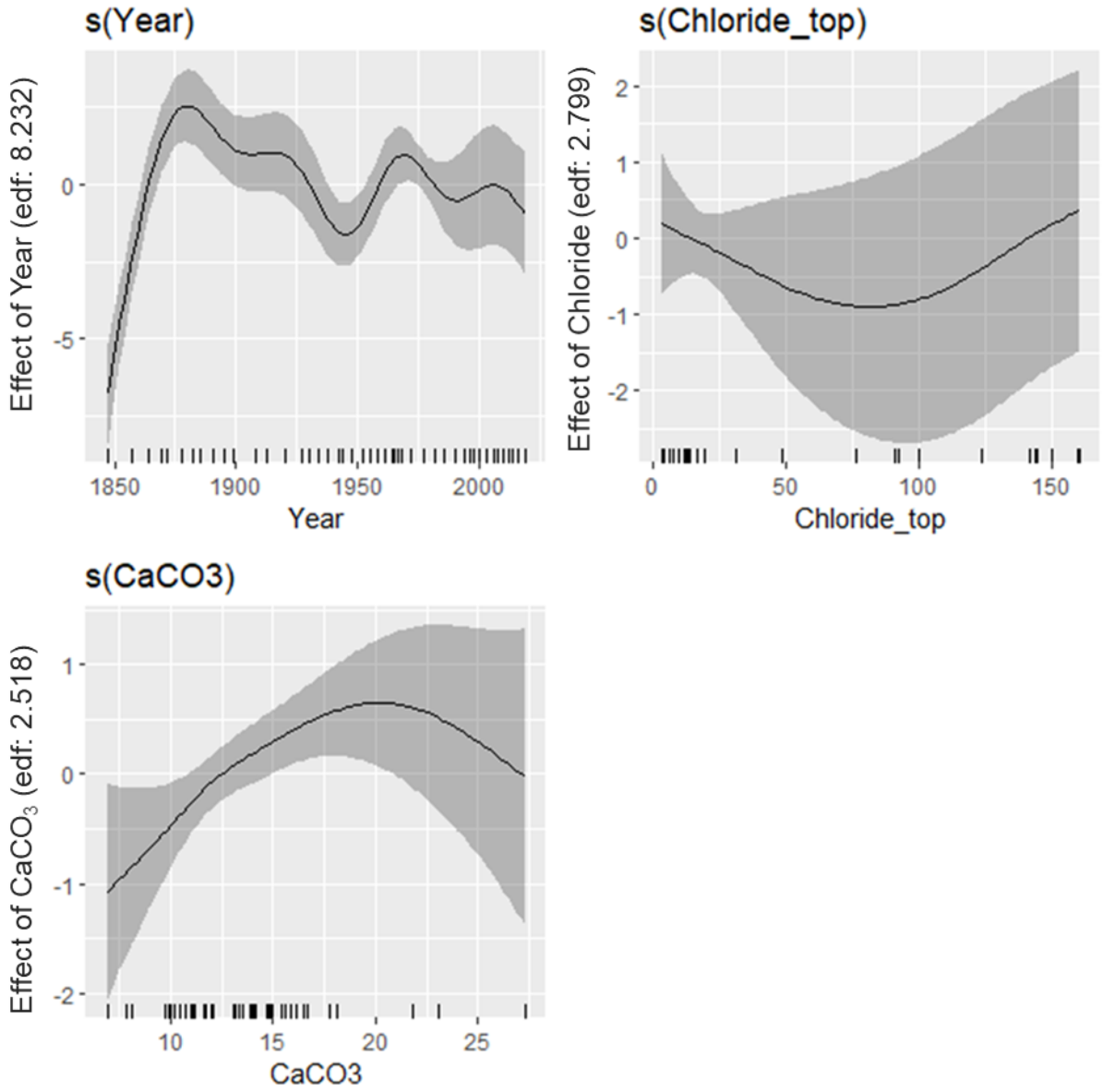


Figure S4: GAMM model plots for *Daphnia pulicaria*. (A) Effect of year within the model, (B) Effect of Chloride within the model (e.g., Figure 3A) and (C) Effect of %-Organic content. Rug along X-axis of A-C indicates the locations of estimated chloride measurements.

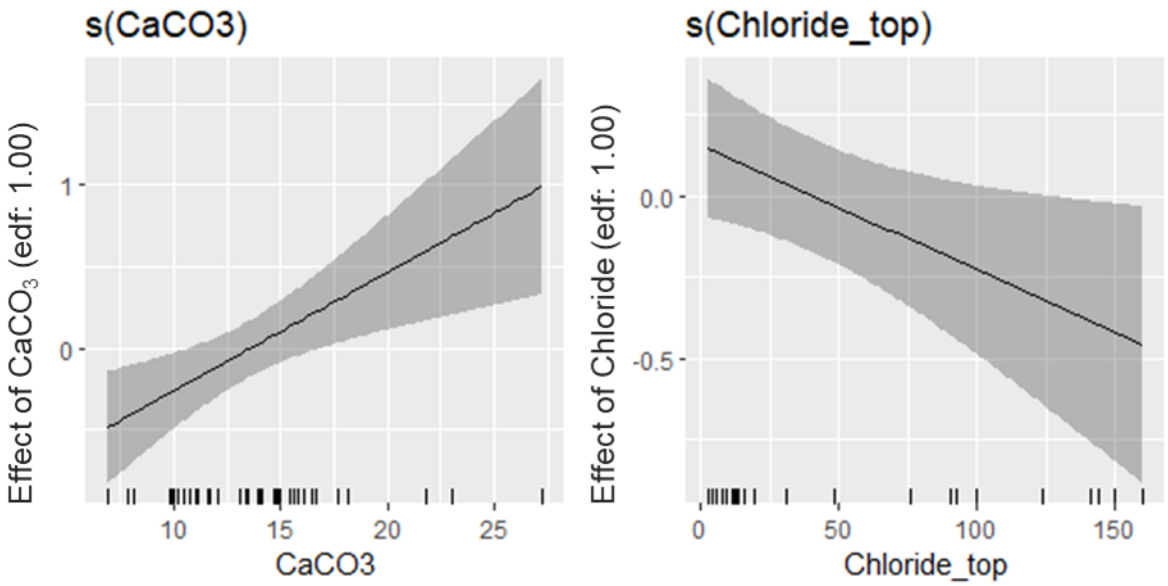


Figure S5: GAMM model plots for *Daphnia mendotae*. (A) Effect of %-$CaCO_3$ content within the model and (B) Effect of Chloride within the model (e.g., Figure 3C). Rug along X-axis of A & B indicates the locations of estimated chloride measurements.

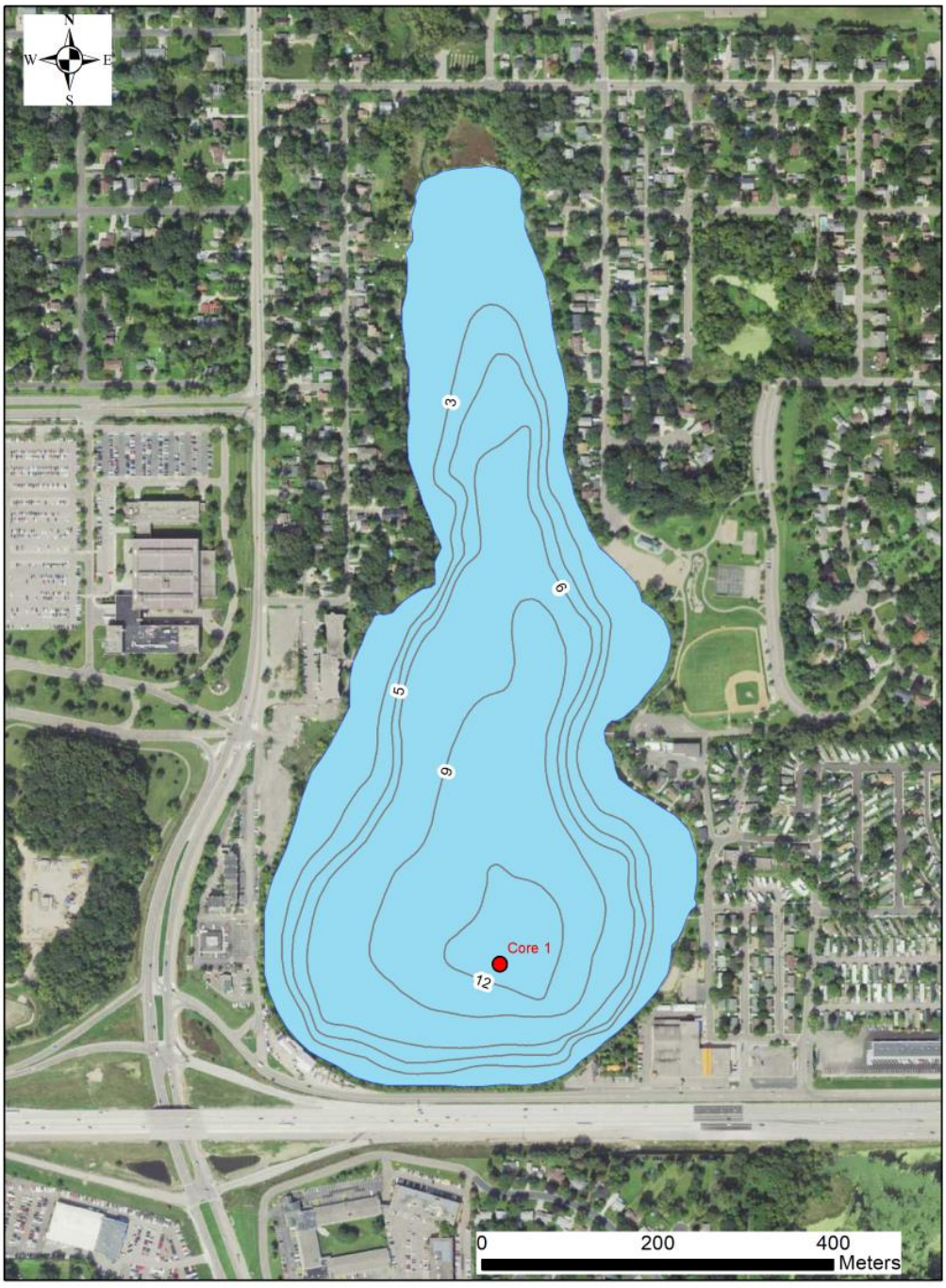


Figure S6: Bathymetry of Tanners Lake (in meters). Core 1 indicates the approximate location where we collected the cores and obtained the lake vertical water column profiles.

Top of Interval (cm)	Base of Interval (cm)	Mid of Interval (cm)	Total 210Pb (pCi/g)	Error of Total Pb (±s.d.)	Cum. Dry Mass (g/cm2)	Unsup. 210Pb (pCi/g)	Error of Unsup Pb (±s.d.)	Age: Base of Int. (yr)	Error of Age (±s.d.)	Date: Base A.D.	Date: Mid A.D.	Sediment DMAR (g/cm2 yr)	Error of DMAR (±s.d.)
0	1	0.5	23.205	0.925	0.056	22.730	0.926	0.96	2.57	2014.1	2014.6	0.0577	0.0040
3	4	3.5	24.074	0.805	0.240	23.598	0.806	4.46	2.72	2010.6	2011.3	0.0502	0.0036
7	8	7.5	9.861	0.352	0.631	9.385	0.353	9.29	2.46	2005.8	2006.4	0.1082	0.0069
10	11	10.5	15.625	0.524	0.926	15.149	0.526	13.33	2.53	2001.7	2002.5	0.0594	0.0038
13	14	13.5	12.316	0.406	1.234	11.840	0.407	18.18	2.71	1996.9	1997.8	0.0656	0.0046
17	18	17.5	8.436	0.348	1.737	7.961	0.349	25.00	2.86	1990.1	1991.0	0.0790	0.0063
20	22	21	7.772	0.319	2.275	7.297	0.321	32.36	3.32	1982.7	1984.6	0.0708	0.0065
26	28	27	3.425	0.113	3.318	2.949	0.119	42.46	2.47	1972.6	1974.2	0.1264	0.0080
32	34	33	2.589	0.082	4.401	2.113	0.089	50.67	2.60	1964.4	1965.7	0.1355	0.0094
40	42	41	2.231	0.080	5.821	1.755	0.087	62.26	3.07	1952.8	1954.4	0.1147	0.0101
46	48	47	2.107	0.069	6.884	1.632	0.077	73.20	3.90	1941.9	1943.9	0.0890	0.0099
54	56	55	1.329	0.033	8.334	0.853	0.048	87.91	3.49	1927.2	1928.9	0.1068	0.0110
60	62	61	1.617	0.040	9.370	1.141	0.053	103.69	4.81	1911.4	1914.7	0.0513	0.0068
66	68	67	1.231	0.031	10.260	0.755	0.047	123.54	7.35	1891.5	1894.9	0.0419	0.0087
74	76	75	0.821	0.022	11.240	0.345	0.041	147.30	9.92	1867.8	1870.4	0.0427	0.0129
80	82	81	0.705	0.023	11.923	0.229	0.042	164.64	16.31	1850.4	1853.5	0.0380	0.0186
86	88	87	0.516	0.017									
108	112	110	0.497	0.015									
132	136	134	0.546	0.017									
154	156	155	0.400	0.013									

Supported Pb-210:	0.4757 ± 0.0349 pCi/g
Number of Supported Samples:	4

Cum. Unsup. Pb-210:	42.7966 pCi/cm2
Unsup. Pb-210 Flux:	1.3644 pCi/cm2 yr

Table S1. Core depth (cm) – estimated date (AD) – dry mass sediment accumulation rate (DMAR) model for Tanners Lake sediment core collected January 2015 based on Pb-210 inventory.

Works Cited:

Appleby, P.G. and Oldfield, F. 1978. The calculation of lead-210 dates assuming a constant rate of supply of the unsupported lead-210 to the sediment. Catena 5:1-8.

Appleby, P.G. 2001. Chronostratigraphic techniques in recent sediments. In: Last, W.M. and Smol, J.P. (Eds) Tracking Environmental Change Using Lake Sediments. Volume 1: Basin analysis, Coring, and Chronological Techniques. Kluwer Academic Publishers, Dordrecht, pp 171-203.

Binford, M.W. 1990. Calculation and uncertainty analysis of 210-Pb dates for PIRLA project lake sediment cores. Journal of Paleolimnology 3: 253-267.

Blumentritt, D.J., Engstrom, D.R. and Balogh, S.J., 2013. A novel repeat-coring approach to reconstruct recent sediment, phosphorus, and mercury loading from the upper Mississippi River to Lake Pepin, USA. Journal of Paleolimnology, 50(3), pp.293-304.

Heiri, O., A. F. Lotter, and G. Lemcke. 2001. Loss on ignition as a method for estimating organic and carbonate content in sediments: reproducibility and comparability of results. J. Paleolimnol. 25: 101–110. doi:10.1023/A:1008119611481

Kelly, V., S. Findlay, W. Schlesinger, A. Chatrchyan, and K. Menking. 2010. Road Salt: Moving Toward the Solution. The Cary Institute of Ecosystem Studies.

Novotny, E. V., D. Murphy, and H. G. Stefan. 2008. Increase of urban lake salinity by road deicing salt. Sci. Total Environ. 406: 131–144. doi:10.1016/J.SCITOTENV.2008.07.037

Ramstack, J. M., S. C. Fritz, and D. R. Engstrom. 2004. Twentieth century water quality trends in Minnesota lakes compared with presettlement variability. Can. J. Fish. Aquat. Sci. 61: 561–576. doi:10.1139/f04-015

Appendix 3

Supplemental Materials for Chapter 3:

Resurrection genomics provides molecular and phenotypic evidence of rapid adaptation to salinization in a keystone aquatic species

Matthew J. Wersebe[1,2] and Lawrence J. Weider[1]

[1] Program in Ecology and Evolutionary Biology, Department of Biology, University of Oklahoma

[2]Corresponding Author: 730 Van Vleet Oval, Richards Hall 304, Norman, OK 73019

matthew.wersebe@ou.edu

Supplemental Methods:

Hatching Protocol:

In brief, eggs (often paired) were excised from the ephippial casing using fine forceps and transferred into a well of a sterile 24-well cell culture plate filled with 2 mL of sterile COMBO media using a Pasteur pipette (1). Plates were covered in foil and incubated in a 4^{o} C refrigerator for 14 days and subsequently transferred to constant light at 20° C. Plates were checked daily for hatchlings for up to two weeks. Hatchlings as single individuals were carefully transferred to a 60 mL glass jar and fed a 50:50 mix of High Orthophosphate (HOP) COMBO grown *Scenedesmus* algae and diluted commercial *Nannochloropsis* concentrate (Reed Mariculture; 1 µL into 1 mL COMBO).

DNA Extraction, Library Preparation and Sequencing:

Clones were expanded in 3.8 L jars so that approximately 100 adult females could be isolated. These 100 adults were subjected to a shortened antibiotic treatment described in (2). Genomic DNA was extracted from the resulting isoclonal pools (~50 adults per tube) using a MasterPure DNA extraction kit following the manufacturer's directions (3). DNA was checked for quality by running a 2 µL aliquot on a 1% agarose electrophoresis gel and quantified using a Qubit fluorometer (ThermoFisher). Illumina-compatible sequencing libraries were constructed using the Ultra FS II DNA library preparation kit (New England Biolabs) and enriched using 4 cycles of PCR. Pooled and normalized libraries were sequenced across three separate runs of an Illumina NovaSeq instrument by the Oklahoma Medical Research Foundation. We sequenced the population to approximately 10X coverage on average (range 3- 20X).

Chloride Tolerance:

We reduced maternal effects among clones by raising experimental animals under standardized conditions. Ten to twenty gravid females per clone were isolated as single individuals into 60 mL jars filled with COMBO media from mass cultures (900 mL jars), for the establishment of experimental populations. These gravid females were allowed to release their clutches, the newborn animals (~25-35) were collected and isolated as single animals in fresh 60 mL jars and grown to adulthood and allowed to release their first clutches. Mothers were then transferred to fresh jars and allowed to release their second clutches. Animals from the second or third clutch for each clone that were born within 24 hours were isolated and pooled together into 900 mL jars and used in the experiments. We conducted 96-hour lethal concentration-50% assays (LC_{50}) to estimate clone-specific tolerance to chloride. Experimental animals less than 24 hours old were haphazardly pipetted from the pooled (900-mL) jars and placed as single animals in 60 mL jars with COMBO media amended with either 0 mg (control), 250 mg, 500 mg, 1000 mg, or 1500 mg of Cl^- from NaCl. Each clone by concentration combination was replicated five times when animal numbers permitted (minimum was triplicate jars per treatment). Each jar across clones and treatments was assigned a random number and jars were placed in covered plastic shoe boxes in numerical order to reduce evaporative loss, as well as any bias due to positional effects. Experimental animals were maintained in a temperature-controlled room (20^o C) on a 12:12 light: dark cycle. We monitored animals daily (every 24 hours) for survival during the experiment. We recorded an animal as dead, when no movement of internal organs and/or appendages were observed after viewing under a dissecting microscope.

mtDNA Analysis:

To assess whether the TL population consisted of a single continuous population through time as the nuclear genome suggested, we analyzed the mtDNA of all clones by constructing a time-scaled phylogeny and inferred a Bayesian Skyline plot of the demographic history. Following the same bioinformatic approach outlined in the main text, we aligned reads to the *Daphnia pulex* mitochondrial reference genome (4). We called variants for each sample individually using BCFtools and applied SNPs variants to create a unique mitochondrial consensus sequence for each sample (5). We aligned each unique consensus sequence using MAFFT v7 (6). Next, we inferred a maximum likelihood phylogenetic tree in iqtree (7), and obtained branch support with 1000 ultra-fast bootstraps (8). We time-scaled out phylogeny using default parameters in TimeTree (9) and implemented a Bayesian Skyline analysis (10) to infer the demographic history from the mitochondrial sequences using BEAST2 (11). Tip dates were inferred from the sediment dating model proposed in Wersebe et al. (12), to the closest year (i.e., LC inferred to be from 2019, 22-24 cm from 1994, etc.).

2021 Tanners Lake Diel Vertical Migration (DVM):

On 29 June 2021 we returned to Tanners Lake to investigate the temporal and spatial heterogeneity of *Daphnia* occupancy of the water column. We quantitatively sampled the *Daphnia* population using a 30-L Schindler-Patalas trap at three stations covering the north-south axis of Tanners Lake. Beginning at 0600 hours (local time), we sampled the deep pelagic station (44.95076° N, -92.98135° W) trapping plankton at depths of 1-m, 6-m, 8-m and 12-m from on board an anchored canoe. We collected samples at a mid-depth pelagic station (44.95287° N, -

92.98108° W) sampling the plankton density at 1-m, 6-m 9-m. Finally, we also collected samples from a shallow littoral station (44.95681° N, -92.98141° W) at 1-m and 2.5-m. In addition to the plankton samples, while at each station we also collected vertical profiles of temperature (°C), specific conductance (SPC; μS/cm) and dissolved oxygen (%) using a multiparamter sonde (YSI, Yellow Springs, OH) and the approximate Secchi depth (m).

We conducted two additional sampling bouts throughout the day, at 1200 hours and 2000 hours on 29 June 2021. During these times we only collected animals from the water column. Once on shore, we preserved animals according to the method outlined in Black and Dodson (14). Each sample was standardized to 50 mL of 70% EtOH. We estimated the number of collected *Daphnia* by subsampling each 50 mL sample 3 times with a 2 mL Hensen-Stempel Pipette and enumerating the animals over a grid. We identified animals to species using the key provided by Haney et al. (15). Using the average of the three Hensen-Stempel subsamples we calculated an estimate of the number of animals collected. Finally, this number was divided by 30-L (size of trap) to determine the density of animals (number per liter) present in the water column at the time of collection.

2021 Tanners Lake DVM Study Results:

We limit our DVM results here to just the focal species of our genetic study, *Daphnia pulicaria.* However, several other species were collected during our surveys. We detected *D. pulicaria* in every sample except from the shallow littoral sites (0600 2.5-m& 1200 1-m). In samples where we did detect *D. pulicaria,* densities ranged from a low of 0.185 individuals per liter (0600 12m

deep) to a high of 27.225 animals per liter (0600 6 m deep). We consistently observed that the highest density estimates were at 6-m at the two pelagic sampling stations. Animals were observed at depth (12 m) throughout the day, ranging from 0.185 (0600) to 5.65 (1200) animals per liter. The density estimates of *D. pulicaria* are presented in Figure S6. Thermal and chemical stratification was much weaker in 2021 compared to our profiles taken in 2019 (see 12). The thermocline begins at approximately 3-m and ends at approximately 5.2-m of depth. SPC ranged from approximately 880 μS/cm at the surface (~190 mg/L Cl-) to a high at 13-m of 1108 μS/cm (~255 mg/L Cl-). Also, likely as a result of the weaker chemical stratification, DO was not completely depleted at depth and ranged from over 100% saturation at the surface to 5.1% saturation at 13-m (approximately 0.64 m/L O_2). Temperature, SPC and DO profiles are available in Figure S7 A-C. Secchi depths at all stations ranged from ~2.5 m to 3.0 m and are available in Table S4.

Works Cited:

1. S. S. Kilham, D. A. Kreeger, S. G. Lynn, C. E. Goulden, L. Herrera, COMBO: a defined freshwater culture medium for algae and zooplankton. *Hydrobiologia* **377**, 147–159 (1998).

2. M. J. Wersebe, R. E. Sherman, P. D. Jeyasingh, L. J. Weider, The roles of recombination and selection in shaping genomic divergence in an incipient ecological species complex. *Molecular Ecology* **n/a** (2022).

3. C. G. Athanasio, J. K. Chipman, M. R. Viant, L. Mirbahai, Optimisation of DNA extraction from the crustacean *Daphnia*. *PeerJ* **4**, e2004 (2016).

4. T. J. Crease, The complete sequence of the mitochondrial genome of Daphnia pulex (Cladocera: Crustacea). *Gene* **233**, 89–99 (1999).

5. P. Danecek, *et al.*, Twelve years of SAMtools and BCFtools. *GigaScience* **10**, 1–4 (2021).

6. K. Katoh, D. M. Standley, MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772–780 (2013).

7. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**, 268–274 (2015).

8. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* **35**, 518–522 (2018).

9. P. Sagulenko, V. Puller, R. A. Neher, TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evolution* **4**, vex042 (2018).

10. A. J. Drummond, A. Rambaut, B. Shapiro, O. G. Pybus, Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. *Molecular Biology and Evolution* **22**, 1185–1192 (2005).

11. R. Bouckaert, *et al.*, BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology* **15**, e1006650 (2019).

12. M. J. Wersebe, M. B. Edlund, L. J. Weider, Does salinization impact long-term Daphnia assemblage dynamics? Evidence from the sediment egg bank in a small hard-water lake. *Limnology and Oceanography Letters* **n/a** (2021).

13. K. L. Korunes, K. Samuk, pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Molecular Ecology Resources* **21**, 1359–1368 (2021).

14. A. R. Black, S. I. Dodson, Ethanol: a better preservation technique for *Daphnia*. *Limnology and Oceanography: Methods* **1**, 45–50 (2003).

15. Haney, J.F. *et al*. An-Image-based Key to the Zooplankton of North America. version 5.0 released 2013. University of New Hampshire Center for Freshwater Biology <cfb.unh.edu>

Supplemental Figures:

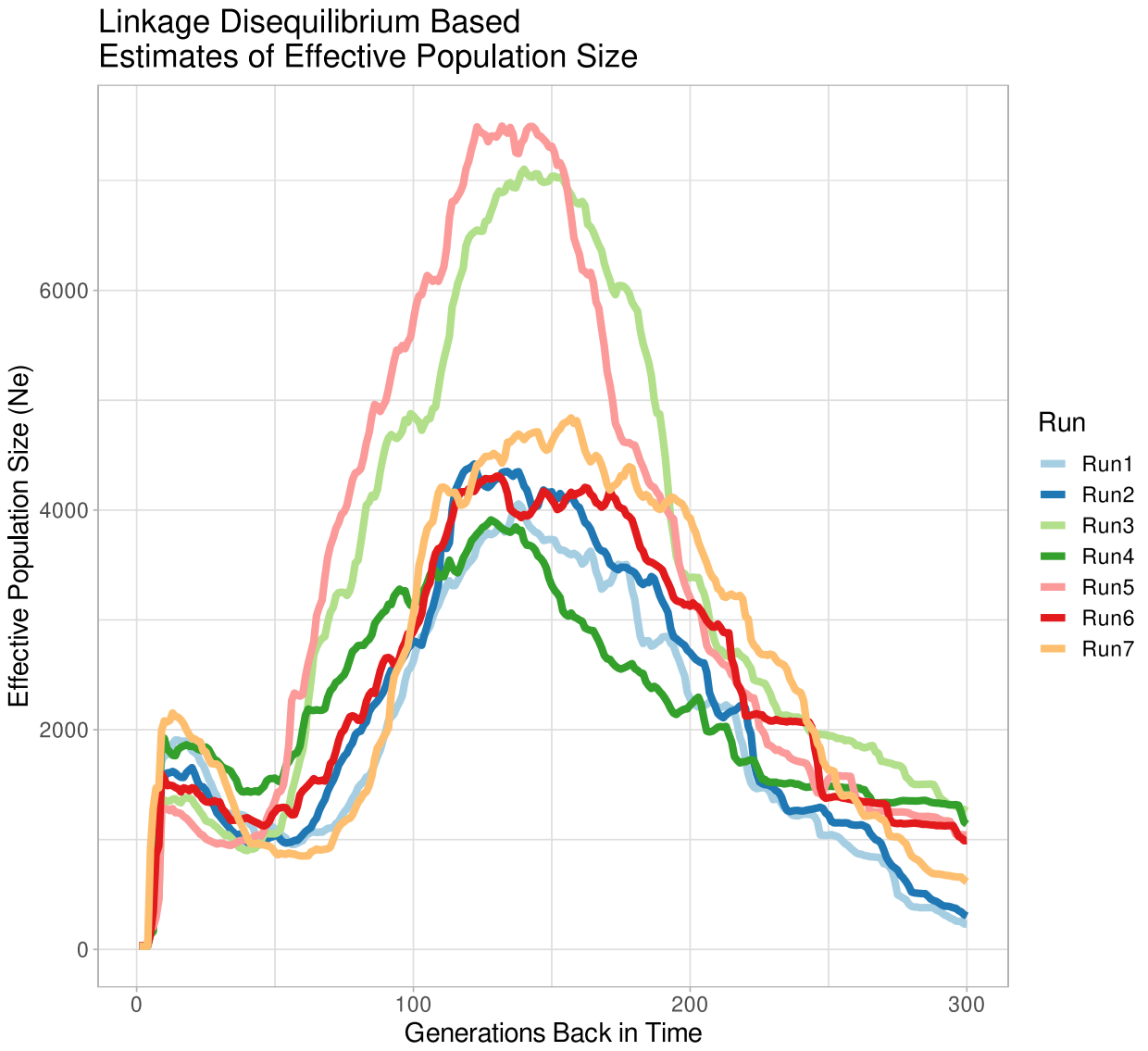


Figure S1: Effective population size (N_e) trajectory inferred using Linkage Disequilibrium. The program GONE was run 7 random samples of 600,000 SNPs sampled from all the SNPs called in the Lake Clones (LC) subpopulation. Results were pruned to show the first 300 generations (method accurate up to ~200-300 generations back in time).

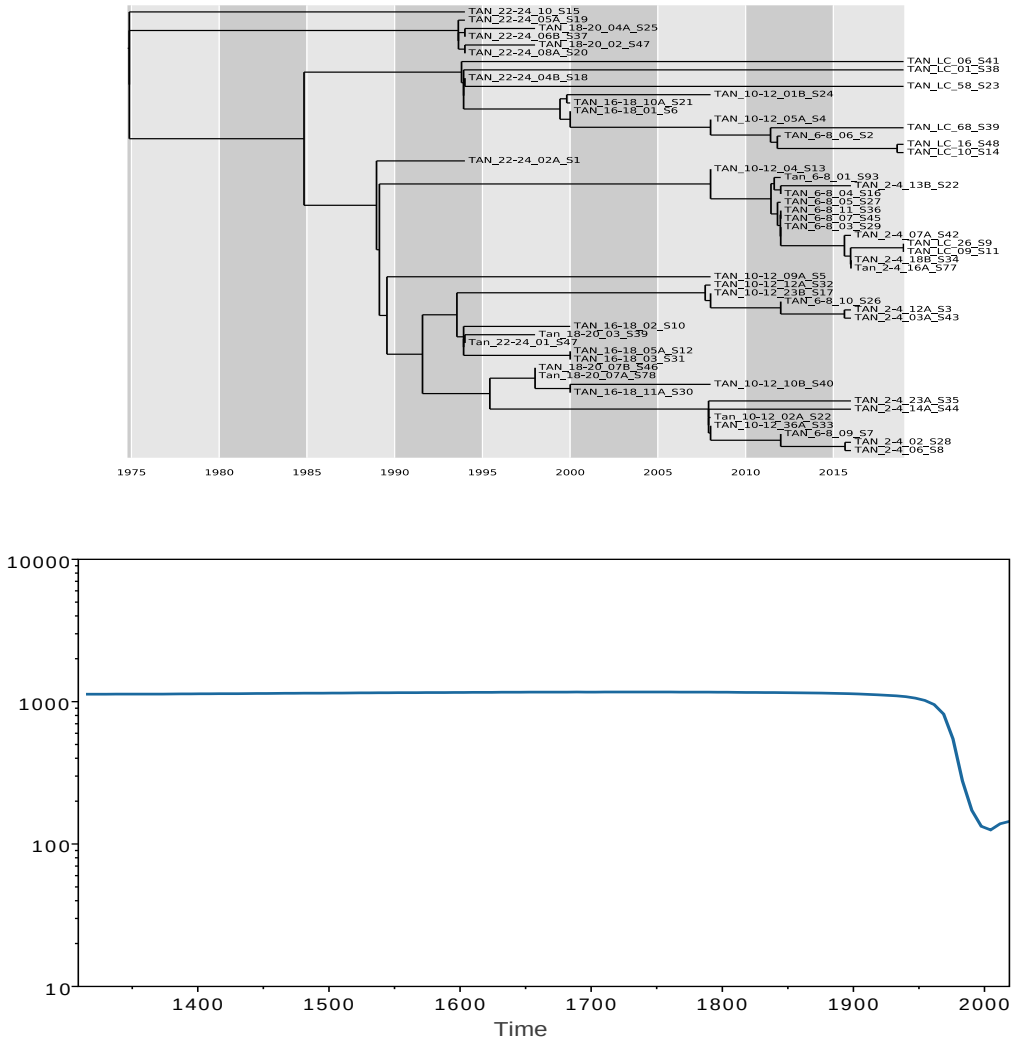


Figure S2: (TOP) Time-scaled maximum-likelihood phylogeny for full-length Tanners Lake (TL) mitochondrial genome sequences. Tip dates were assigned based on the age of the sediment layer of ephippial recovery inferred by the dating model for TL (see above) to nearest year. The tree was rooted with TimeTree to maximize the temporal signal using the default method. Nodes coalesced samples typically across time periods, indicating a constant population across time. **(BOTTOM)** Bayesian skyline plot inferred using BEAST2. The shape of the demographic history is similar to that inferred by GONE (see S1 above); however N_e sizes differed by an order of magnitude.

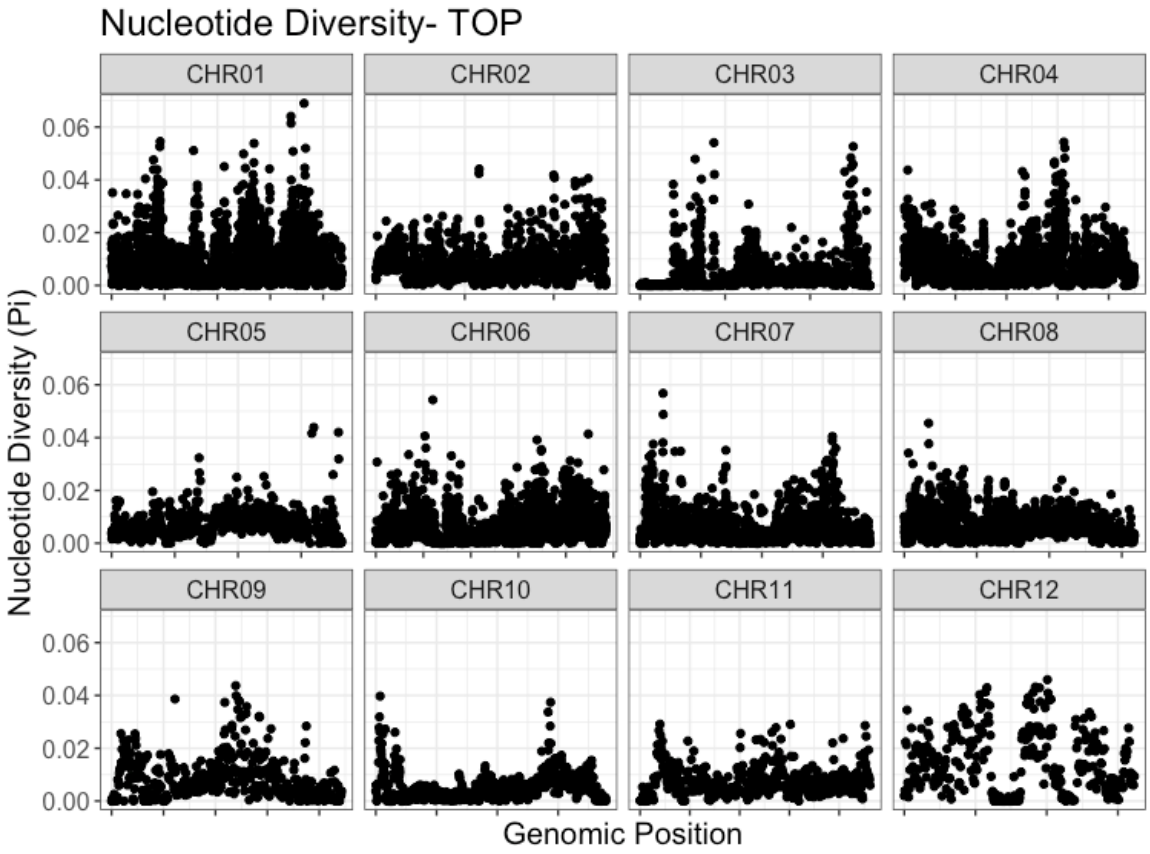


Figure S3: Windowed (10kb) nucleotide diversity genome-wide for the TOP population.

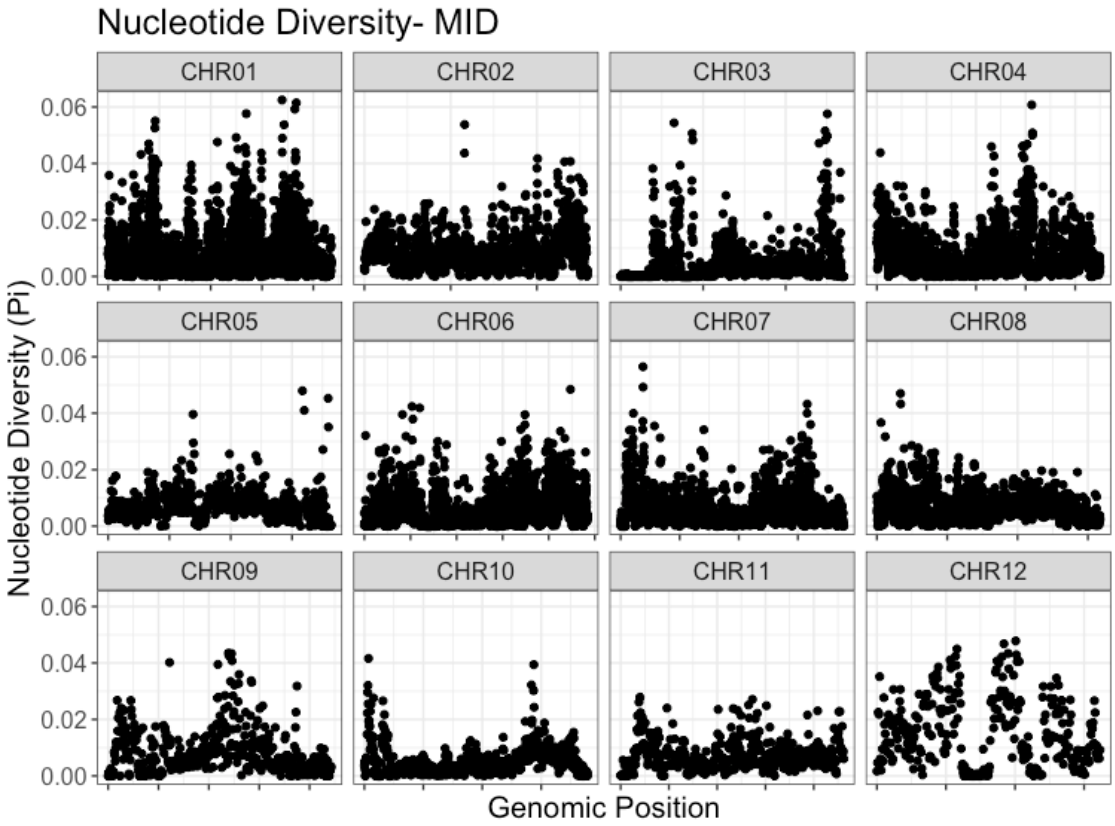


Figure S4: Windowed (10kb) nucleotide diversity genome-wide for the MID population.

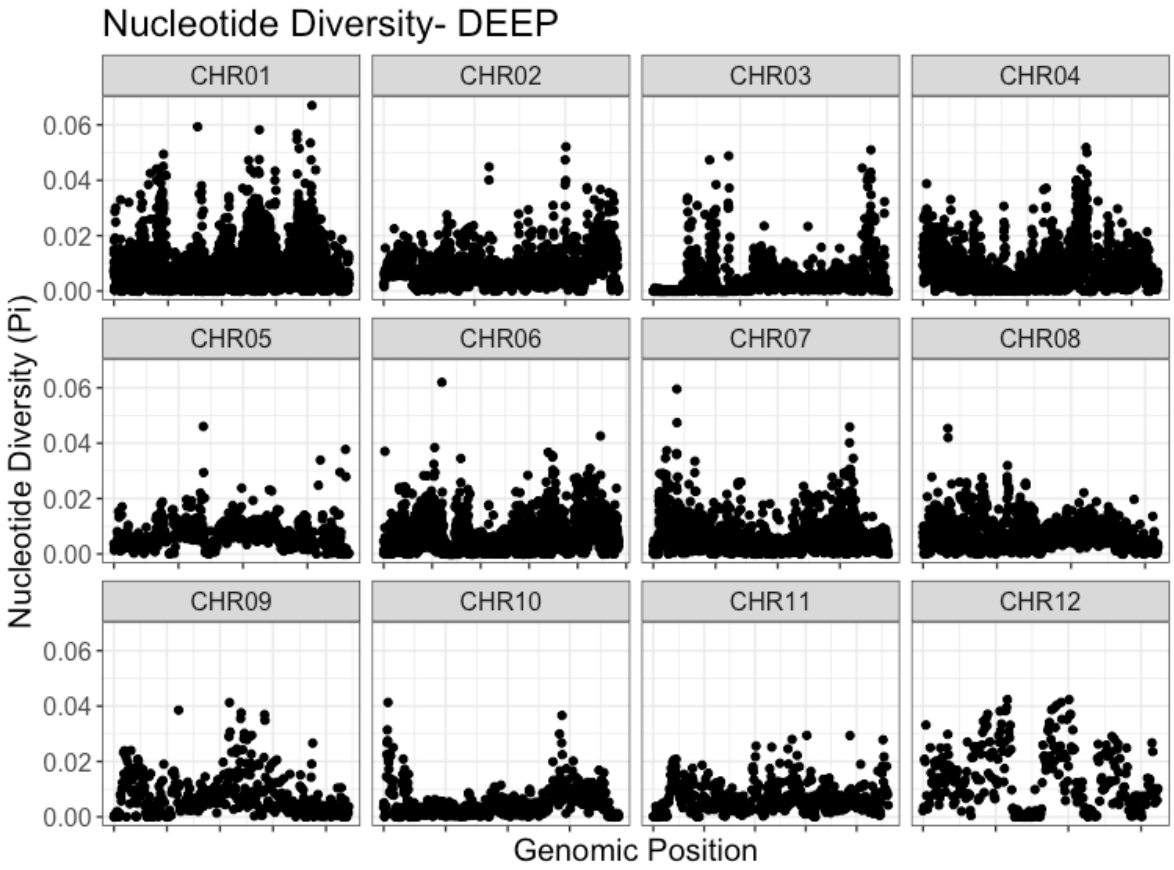


Figure S5: Windowed (10kb) nucleotide diversity genome-wide for the DEEP population.

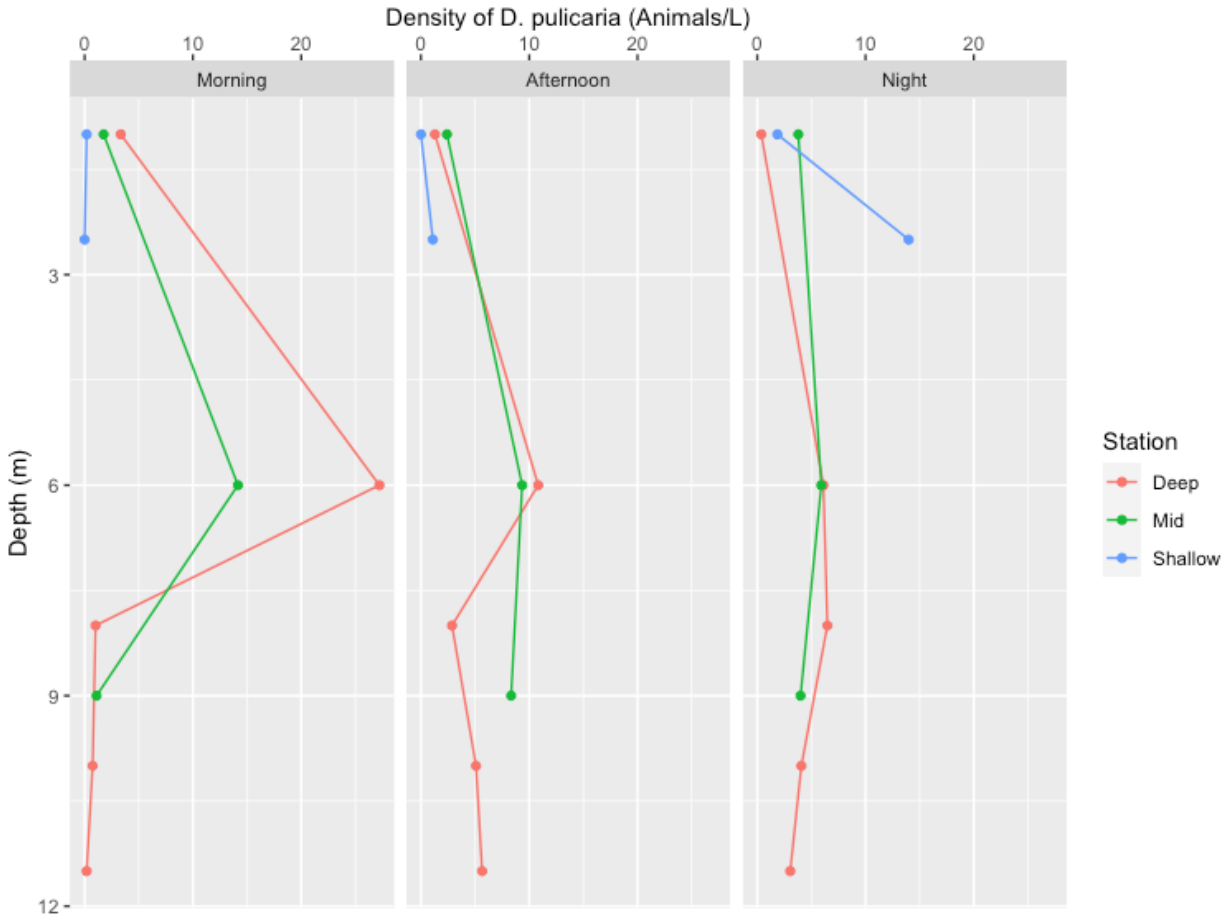


Figure S6: Spatial and temporal heterogeneity in *Daphnia pulicaria* densities. Colors represent the three sampling stations. Left Panel: *D. pulicaria* densities across depths at the three stations at 0600 hours. Middle Panel: *D. pulicaria* densities across depths at the three stations at 1200 hours. Right Panel: *D. pulicaria* densities across depths at the three stations at 2200 hours.

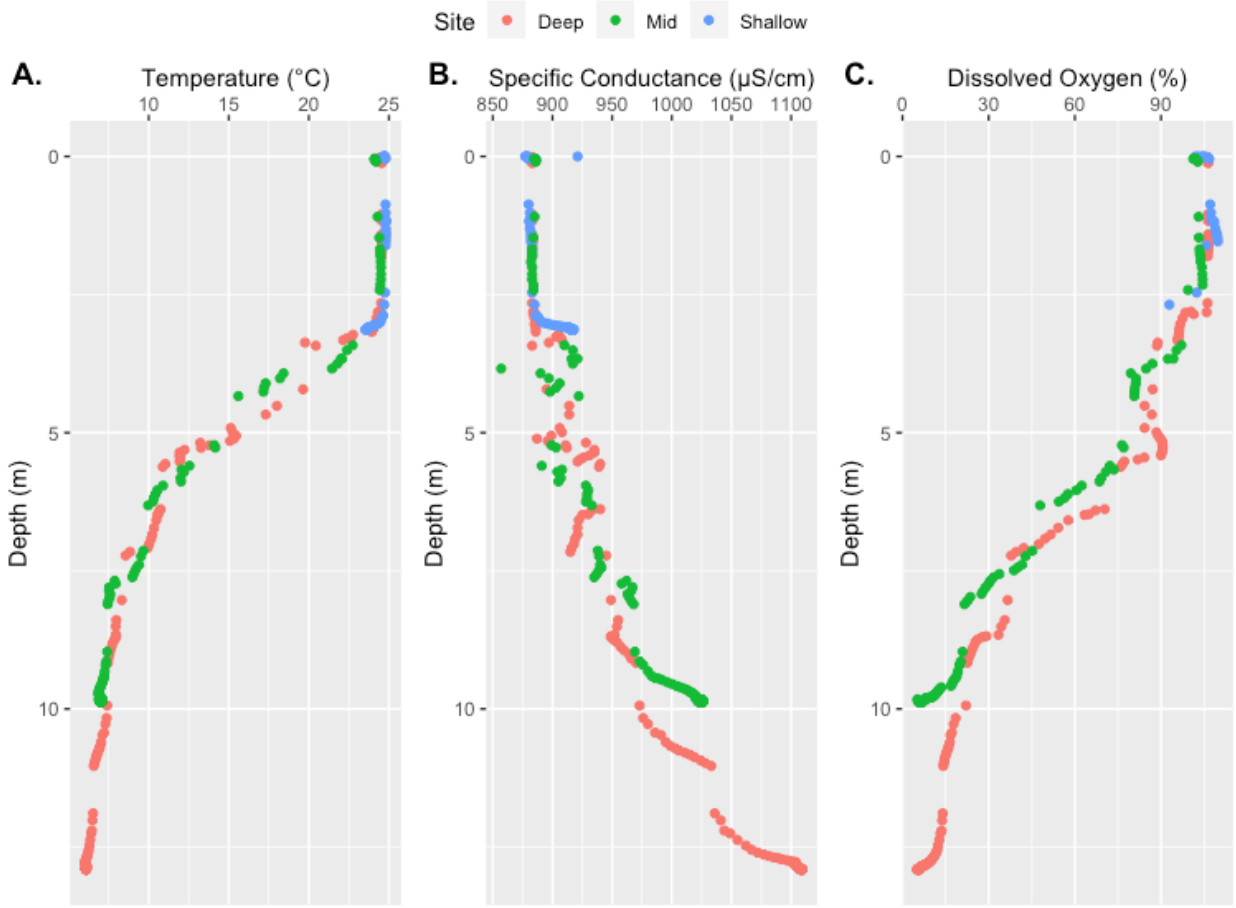


Figure S7: Abiotic profiles collected on 29 June 2021 collected with a YSI multiparameter sonde. Colors represent profiles taken at each sampling station. A) Temperature profile (°C). B) Specific conductance profile (µS/cm). C) Dissolved Oxygen profile (% saturation).

Table S1: 95% confidence intervals for the demographic model fit using FSC2.7.

	Effective Population Size	**Growth Rate**
Model	2931	2.6269×10^{-5}
+ 95% Confidence	2939.162	2.6479×10^{-6}
- 95% Confidence	2921.498	-5.529×10^{-6}
Mean	2930.33	-1.44×10^{-6}

Table S2: GO term enrichment analysis.

#Analysis Type: PANTHER Overrepresentation Test (Released 20220202)

#Annotation Version and Release Date: PANTHER version 17.0 Released 2022-02-22

#Analyzed List: outlier_prots.panther.txt

#Reference List: Daphnia pulex (all genes in database)

#Test Type: FISHER

#Correction: FDR

#PANTHER GO-Slim Molecular Function Daphnia pulex - REFLIST (30047) outlier_prots.panther.txt (201) outlier_prots.panther.txt (expected) outlier_prots.panther.txt (over/under) outlier_prots.panther.txt (fold Enrichment) outlier_prots.panther.txt (raw P-value) outlier_prots.panther.txt (FDR)

protein kinase C binding (GO:0005080) 5 2 .03 + 59.80 9.03E-04 1.12E-02

voltage-gated potassium channel activity (GO:0005249) 14 3 .09 + 32.03 1.84E-04 3.47E-03

postsynaptic neurotransmitter receptor activity (GO:0098960) 11 2 .07 + 27.18 3.27E-03 3.27E-02

RNA polymerase II complex binding (GO:0000993) 12 2 .08 + 24.91 3.79E-03 3.72E-02

neuropeptide receptor activity (GO:0008188) 12 2 .08 + 24.91 3.79E-03 3.65E-02

3'-5'-exoribonuclease activity (GO:0000175) 13 2 .09 + 23.00 4.36E-03 4.04E-02

basal RNA polymerase II transcription machinery binding (GO:0001099) 13 2 .09 + 23.00 4.36E-03 3.97E-02

basal transcription machinery binding (GO:0001098) 13 2 .09 + 23.00 4.36E-03 3.90E-02

voltage-gated channel activity (GO:0022832) 33 5 .22 + 22.65 5.18E-06 5.28E-04

voltage-gated ion channel activity (GO:0005244) 33 5 .22 + 22.65 5.18E-06 4.40E-04

voltage-gated cation channel activity (GO:0022843) 27 4 .18 + 22.15 5.17E-05 1.47E-03

RNA polymerase core enzyme binding (GO:0043175) 14 2 .09 + 21.36 4.96E-03 4.29E-02

neuropeptide binding (GO:0042923) 15 2 .10 + 19.93 5.60E-03 4.76E-02

chloride channel activity (GO:0005254) 23 3 .15 + 19.50 6.71E-04 9.25E-03

potassium channel activity (GO:0005267) 35 4 .23 + 17.08 1.30E-04 3.01E-03

tetrapyrrole binding (GO:0046906) 29 3 .19 + 15.46 1.24E-03 1.44E-02

heme binding (GO:0020037) 29 3 .19 + 15.46 1.24E-03 1.41E-02

sulfotransferase activity (GO:0008146) 53 5 .35 + 14.10 4.24E-05 1.44E-03

gated channel activity (GO:0022836) 77 6 .52 + 11.65 1.97E-05 1.26E-03

potassium ion transmembrane transporter activity (GO:0015079) 53 4 .35 + 11.28 5.68E-04 8.04E-03

hydrolase activity, hydrolyzing O-glycosyl compounds (GO:0004553) 81 6 .54 + 11.07 2.58E-05 1.20E-03

hydrolase activity, acting on glycosyl bonds (GO:0016798) 88 6 .59 + 10.19 4.00E-05 1.46E-03

inorganic anion transmembrane transporter activity (GO:0015103) 46 3 .31 + 9.75 4.25E-03 4.02E-02

ion channel activity (GO:0005216) 145 8 .97 + 8.25 8.98E-06 6.54E-04

anion transmembrane transporter activity (GO:0008509) 76 4 .51 + 7.87 2.02E-03 2.10E-02

passive transmembrane transporter activity (GO:0022803) 163 8 1.09 + 7.34 2.01E-05 1.14E-03

channel activity (GO:0015267) 163 8 1.09 + 7.34 2.01E-05 1.03E-03

cation channel activity (GO:0005261) 110 5 .74 + 6.79 1.05E-03 1.27E-02

transmembrane signaling receptor activity (GO:0004888) 168 7 1.12 + 6.23 1.78E-04 3.48E-03

peptide binding (GO:0042277) 98 4 .66 + 6.10 4.84E-03 4.26E-02

inorganic molecular entity transmembrane transporter activity (GO:0015318) 210 8 1.40 + 5.69 1.12E-04 2.72E-03

RNA polymerase II cis-regulatory region sequence-specific DNA binding (GO:0000978) 215 8 1.44 + 5.56 1.31E-04 2.90E-03

ion transmembrane transporter activity (GO:0015075) 242 9 1.62 + 5.56 4.99E-05 1.50E-03

cis-regulatory region sequence-specific DNA binding (GO:0000987) 216 8 1.44 + 5.54 1.35E-04 2.86E-03

signaling receptor activity (GO:0038023) 279 10 1.87 + 5.36 2.58E-05 1.10E-03

molecular transducer activity (GO:0060089) 280 10 1.87 + 5.34 2.66E-05 1.04E-03

DNA-binding transcription factor activity (GO:0003700) 324 9 2.17 + 4.15 4.14E-04 6.40E-03

RNA polymerase II transcription regulatory region sequence-specific DNA binding (GO:0000977) 330 9 2.21 + 4.08 4.71E-04 7.07E-03

DNA-binding transcription factor activity, RNA polymerase II-specific (GO:0000981) 304 8 2.03 + 3.93 1.21E-03 1.43E-02

transcription cis-regulatory region binding (GO:0000976) 350 9 2.34 + 3.84 7.09E-04 9.51E-03

transcription regulatory region nucleic acid binding (GO:0001067) 350 9 2.34 + 3.84 7.09E-04 9.27E-03

phosphotransferase activity, alcohol group as acceptor (GO:0016773) 313 8 2.09 + 3.82 1.45E-03 1.57E-02

double-stranded DNA binding (GO:0003690) 393 10 2.63 + 3.80 3.92E-04 6.24E-03

transcription regulator activity (GO:0140110) 438 11 2.93 + 3.75 2.25E-04 3.83E-03

sequence-specific double-stranded DNA binding (GO:1990837) 361 9 2.41 + 3.73 8.77E-04 1.12E-02

transmembrane transporter activity (GO:0022857)	407	10	2.72	+	3.67	5.11E-04	7.45E-03
DNA binding (GO:0003677)	579	14	3.87	+	3.61	4.72E-05	1.51E-03
kinase activity (GO:0016301)	346	8	2.31	+	3.46	2.65E-03	2.71E-02
sequence-specific DNA binding (GO:0043565)	391	9	2.62	+	3.44	1.50E-03	1.60E-02

Table S3: Extreme outlier SNPs and nearest genes (within 5Kb).

Chromosome	**Position**	**Fst**	**Type**	**Gene**
04	13394600	0.3983	Intronic	*pank4; tda6*
02	6974770	0.3645	Intergenetic	NA
02	3614856	0.3635	Intronic	*col1a1; tcaim*
03	8229625	0.3609	Intronic	*rap1gap*
03	7485508	0.3480	Intronic	*daam1*
05	4372154	0.3284	Intronic	*ptn*
06	21993985	0.3221	Intronic	*rotund*
04	2181789	0.3220	Intronic	*still life*
04	19599247	0.3211	Intronic	*fstl5*
05	4380601	0.3206	Intronic	*ptn*
02	9684091	0.3175	Intronic	Uncharacterized: LOC124326411 LOC124326443 LOC124326498

Table S4: Secchi Depths at sampling stations on Tanners Lake

Time	**0600**	**1200**	**2000**
Deep Pelagic	3.0 m	2.9 m	2.9 m
Mid Pelagic	3.0 m	2.85 m	2.8 m
Shallow Littoral	3.1 m	2.9 m	2.9 m