# DEVELOPMENT OF NEW TECHNIQUES TO DISSECT THE GENETIC BASIS OF POPULATION-LEVEL VARIATION OF RECOMBINATION RATES IN THE *DAPHNIA* MODEL SYSTEM

by

SWATANTRA NEUPANE

DISSERTATION

Submitted in partial fulfillment of the requirements

for the degree of Doctor of Philosophy at

The University of Texas at Arlington

December 2021

Arlington, Texas

Supervising Committee:

Sen Xu, Supervising Professor

Todd A Castoe

Jeffery P Demuth

Matthew Walsh

Mark Pellegrino

# Acknowledgment

I want to thank my advisor, Sen Xu, for first accepting me to the graduate program and guiding me through this dissertation adventure. His constant thoughts, motivation, and direction have guided me through this journey and made it possible. I also would like to convey my gratitude to all the lab members, Marelize Snyman, Trungh Hyunh, Thinh Pham, Hongjun Wang, Irene Moy, Matt Smith, and Kenny Tran along with all the undergraduates that were part of the lab at some point. Marelize Snyman deserves a handful of thanks not just from me but from all the graduate students in the lab. She constantly reminded us of the deadlines and meetings that otherwise would have been missed. Thanks, Trung Hyunh, for all the stories and talk that help from post-lunch dizziness, and of course, his strong suit, coding, that we all rely on. Thinh Pham has been a crucial part of the lab even before he joined as a graduate student. His help in running gels and PCR to MALBAC is greatly appreciated. It helped me ease my workload at times. Hongjun, I appreciate his patience in explaining things to me when I had many questions; I also learned a lot from him. The first year I joined as a graduate student, Alex Hall had made me feel more included in the ways of living here in the states. He helped me seep into the new culture that I would have had difficulty in.

I want to thank Kimberly Bowles for running the immense number of plates for genotyping. I want to thank Kim for all the patience, suggestion, training that she provided during my time here. The training surely helps me as it is an excellent addition to my skillset.

To all the lab members who have made my time here as a home away from home, Thinh, Marelize, Trung, Layla, Irene, Tiffany, Kenny, Hillary, Dustin, I would never forget the times we had, especially the bachelor's party, some part I would never remember.

My other international friends, Santosh Dhamala, Sarbjeet Niraula, Suman Shrestha, Adnan Qureshi, Balan Ramesh, Farah Shamma, Fatema Ruma, Joydeep Chatterjee, Isha Joglekar and Mithun Nag, Liqi Yao, I have had a great time with you guys here at UTA. Santosh Dhamala, thank you for always being there.

Many thanks to my committee members Drs. Jeff Demuth, Matt Walsh, Mark Pellegrino, and Todd Castoe for their insights to improve my research. I am fortunate to have them on my committee. Also, the Biology office deserves a shoutout for fixing all the problems that came from enrolling to disbursement.

In the states, I have my lovely sister Pallawi, my Vinaju, Subodha Khanal, and a notorious nephew, Nirvan. They helped me from applying here to making sure that I had a place to live in Arlington. No words can express how grateful I am. Also, my mother, Sheela Sharma, and my father, Pawan Kumar Sharma Neupane, I am lucky to have such blessed parents. My sister Nishika and Jwai Anurag, and nephew Addu, you guys are always intensely missed.

Finally, my wife, Ashmita Pokharel, I would not have been sane enough to carry out the journey without you. Thank you for always being by my side. Also, the Pokharel family in Nepal and here in the states have never let us feel any of our festivals missed. Moreover, of course, Dilisha Upadhayaya had always made us feel at home. I thank everyone.

# Dedication

*None of this would have been possible without the constant support and motivation from my family. I want to express my gratitude to my parents for everything they did to raise me and make me who I am today. Little did they know how much their kind words and motivation have helped me succeed in every part of my life. I want to dedicate this work to my family, who made my heart whole.*

# Abstract

We know very little about the determining genetic factors that control the recombination events within a genome. Moreover, the cause of recombination rate variation in population-level remains poorly understood, especially in non-mammalian species. The scarcity of population-level recombination rate data makes pinpointing genetic loci responsible for recombination rate variation even more challenging. Due to a lack of population-level recombination data, this hypothesis remains untested mainly in natural systems, like *Daphnia*. In response to the limitation of genetic tools used to dissect the genetic determinant of a trait in *Daphnia*, we put an effort to address these shortcomings in the *Daphnia* research. We started the exploration by first showing strong support for the adaptive divergence of meiotic recombination rate between two incipient sister species of *Daphnia*, *D. pulicaria* and *D. pulex*. Then, we explored optimizing two genetic tools that are crucial in population genetics. First, we fine-tuned and meticulously established a technique to develop an efficient biallelic and heritable gene-editing technique in *D. pulex* using CRISPR-Cas9. Further, we implemented different modifications to explore developing an inexpensive multiplexing protocol by re-engineering the whole genome amplification method before multiplex library preparation.

# Table of Contents

# Chapter 1
# Introduction

Meiosis is widespread in most sexual eukaryotes, and recombination is the hallmark of meiosis. Meiotic recombination is also a key determinant of many aspects of genomic evolution, including the distribution of genetic diversity, rate of adaptation, accumulation of deleterious mutations, and nucleotide substitution. Different molecular pathways that lead to recombination, including maintaining a certain number of crossovers (minimum and a maximum number of crossover) to controlling the crossover location in a chromosome, are tightly regulated by molecular mechanisms. However, recombination rate is a heritable quantitative trait that exhibits a remarkable variation at multiple biological levels in all species studied to date. The recombination rate variation has prompted many questions, including exploring the genetic factor that controls the recombination rate variation. Meiotic recombination rate varies greatly at multiple biological levels, e.g., within the genome, between individuals and populations, and between species.

Nevertheless, we know little about the genetic factors controlling recombination events within a genome in most model organisms, including *Daphnia*. The cause of recombination rate variation in population-level remains poorly understood, especially in non-mammalian species. Because of the scarcity of population-level recombination rate data, pinpointing genetic loci responsible for recombination rate variation remains challenging. Despite emphasizing the importance of different evolutionary forces in recombination rate evolution, these theories converge on predicting that the transition to a novel environment will lead to an increased recombination rate due to novel selection pressure. In the following dissertation, we investigate three topics related to exploring new and re-engineered techniques to help better estimate the recombination rate and dissect the genetic basis of recombination rate variation. The following dissertation first shows strong support

for the adaptive divergence of meiotic recombination rate between two incipient sister species of *Daphnia*, *D. pulicaria* and *D. pulex*. Second, we developed an efficient biallelic and heritable gene-editing technique in *D. pulex* using CRISPR-Cas9. Third, explore developing a re-engineered whole genome amplification method for multiplex library preparation to reduce cost significantly. Collectively, this dissertation works to develop and upgrade molecular techniques to help estimate genome-wide recombination rate and dissect the genetic basis of recombination rate variation in non-model organisms, like *Daphnia*.

Chapter 2

# Adaptive divergence of meiotic recombination rate in ecological speciation[1]

Swatantra Neupane and Sen Xu*

Department of Biology, University of Texas at Arlington, 501 S. Nedderman Dr., Arlington, TX 76019

*Correspondence author: Sen Xu, email: sen.xu@uta.edu

**Key words**: single-sperm genotyping, genetic maps, habitat transition, *Daphnia pulex*, *Daphnia pulicaria*, $Q_{st,}$ $P_{st}$

**Running title**: Adaptive divergence of meiotic recombination rate

[1]This article is published in Genome Biol. Evol. 12(10):1869–1881

Abstract

Theories predict that directional selection during adaptation to a novel habitat results in an elevated meiotic recombination rate. Yet the lack of population-level recombination rate data leaves this hypothesis untested in natural populations. Here we examine the population-level recombination rate variation in two incipient ecological species, the microcrustacean *Daphnia pulex* (an ephemeral-pond species) and *D. pulicaria* (a permanent-lake species). The divergence of *D. pulicaria* from *D. pulex* involved habitat shifts from pond to lake habitats as well as strong local adaptation due to directional selection. Using a novel single-sperm genotyping approach, we estimated the male-specific recombination rate of two linkage groups in multiple populations of each species in common garden experiments and identified a significantly elevated recombination rate in *D. pulicaria*. Most importantly, population genetic analyses show that the divergence in recombination rate between these two species does not follow the evolution of a neutral trait and is therefore most likely due to divergent selection in distinct ecological habitats.

## Introduction

Meiotic recombination is a hallmark of meiosis as it occurs in the majority of sexually reproducing eukaryotes (Cavalier-Smith 2002; Otto and Lenormand 2002). Although it remains contested as to why recombination originated in the last common ancestor of eukaryotes (Kondrashov 1988; Cavalier-Smith 2002), recombination plays an essential role in repairing the actively induced double-strand DNA breaks in the prophase I of meiosis (Pâques and Haber 1999). The presence of at least one recombination event (i.e., crossover event) per chromosome between homologous chromosomes ensures the correct segregation of chromosomes into daughter cells, preventing the otherwise disastrous outcome of chromosome non-disjunction and aneuploidy (Hassold and Hunt 2001).

Besides its well-known role in creating new haplotypes and in facilitating adaptation (Rice 2002), meiotic recombination is an important evolutionary force shaping the eukaryotic genomic architectures. Recombination rate is a determinant of the distribution of genetic diversity in the genomes, exhibiting a positive correlation with nucleotide diversity in many species including human (Lercher and Hurst 2002), mice (Booker et al. 2017), and *Drosophila* (Begun and Aquadro 1992; Charlesworth et al. 1993). Moreover, as genetic linkage reduces the efficacy of natural selection, i.e., Hill-Robertson effect (Hill and Robertson 1966), recombination, which breaks down genetic linkage, can greatly reduce selection interference between linked sites (Barton 1995a; Cutter and Payseur 2013) As a result, recombination can slow down the accumulation of deleterious mutations, i.e., Muller's ratchet (Felsenstein and Yokoyama 1976; Lynch et al. 1993), and of transposable elements (Rizzon et al. 2002; Dolgin and Charlesworth 2008; Kent et al. 2017). Moreover, recombination and associated biased gene conversion can also influence codon usage bias (Comeron et al. 1999; Pouyet et al. 2017) and base composition (Duret and Arndt 2008; Mugal

et al. 2015).

Meiotic recombination rate varies greatly at multiple biological levels, e.g., within genome, between individuals and populations, and between species (Smukowski and Noor 2011; Ortiz-Barrientos et al. 2016; Dapper and Payseur 2017; Ritz et al. 2017). Understanding the genetic basis and evolutionary forces underlying such variation is a major challenge to biologists. Striking progress has been made in mapping the genetic factors responsible for within-genome variation and for between-individual variation. For example, the zinc finger domain protein PRDM9 is a major determinant of recombination hotspots in the genomes of human and mice (Baudat et al. 2010; Myers et al. 2010; Grey et al. 2011; Brick et al. 2012). Also, promoters and transcription start sites have been identified to be associated with elevated recombination rate in dogs (Auton et al. 2013), the yeast *Saccharomyces cerevisiae* (Pan et al. 2011), birds (Singhal et al. 2015), and *Arabidopsis* (Choi et al. 2013). On the individual level, several meiosis-related genes (e.g., *Rnf212*, *Cplx1*, *Rec8*, *Prdm9*) have been identified to be responsible for variation of recombination rates in mammalian species including humans, cattle, and Soay sheep (Kong et al. 2008; Chowdhury et al. 2009; Sandor et al. 2012; Johnston et al. 2016; Halldorsson et al. 2019). However, it should be noted that these loci explain only a small portion (~3-11%) of the phenotypic variance between individuals (Kong et al. 2014; Johnston et al. 2016).

In contrast, the genetic factors governing the inter-specific variation of recombination rate remains understudied (Dapper and Payseur 2017), although many studies have compared recombination rate differences between closely related species at different genomic scales (Smukowski and Noor, 2011). We note that this research area has drawn increasing amount of attention, with a dicistronic gene mei-217/mei-218 recently identified to be responsible for recombination rate difference between *Drosophila melanogaster* and *D. mauritiana* (Brand et al.

2018). On the other hand, another equally understudied question is whether natural selection plays a role in shaping the between-species divergence. Despite numerous evolutionary theories have examined how natural selection can modulate the evolution and divergence of recombination rates between species, the lack of in-depth population-level data (see below) leaves these theories untested in natural systems, severely limiting our understanding of the evolutionary forces driving recombination rate divergence.

In populations undergoing divergence and incipient speciation, recombination rates could be driven to increase if the breakdown of overrepresented association of alleles, i.e., linkage disequilibrium, is beneficial. Generally speaking, three different situations can lead to the buildup of linkage disequilibrium and determine how recombination rate would response to natural selection. In the presence of linkage disequilibrium caused by weak negative epistasis selection favors increased recombination in a large population in stable environment (Otto and Lenormand 2002). Genetic drift can also lead to the accumulation of linkage between beneficial alleles and deleterious alleles in finite populations, and the increase of recombination rate would be favored by selection to bring together beneficial alleles (Otto 2009). Furthermore, temporal fluctuations in the environment favor different combinations of alleles, which could lead to increased recombination rate in environments with rapid and consistent temporal variation (Charlesworth 1976; Barton 1995b; Otto and Michalakis 1998), whereas in the absence of fluctuations recombination is selected against.

Despite the diverse views on the relative importance of these evolutionary forces in shaping the evolution of recombination rate, it is consistently predicted that transition to a novel environment would lead to an increase of recombination rate due to directional selection (Butlin 2005). Empirical work on indirect selection of physiology-related traits in *Drosophila* supports

this view (Korol and Iliadi 1994; Aggarwal et al. 2015). However, for domesticated animals that underwent strong directional selection, there seems to be no increase of recombination rate (Munoz-Fuentes et al. 2015), contradicting previous views of elevated recombination in domesticated plants (Ross-Ibarra 2004) and animals (Burt and Bell 1987; Poissant et al. 2010).

Notably, few studies have directly addressed whether habitat shift in natural populations results in elevation of recombination rate. A key challenge is that, for model organisms where recombination is heavily investigated, e.g., human, mice, Drosophila, and yeast, little is known about the ecological changes involved in speciation. Thus, the inter-specific difference between Drosophila species, e.g., ~2-fold difference between *D. melanogaster* and *D. mauritiana* (Brand et al. 2018), and the difference between yeast species, e.g., 40% lower recombination rate in *Saccharomyces paradoxus than in S. cerevisiae (Liu et al. 2019),* are unfortunately de-coupled from the consideration of ecology.

Another challenge in understanding the relationship between ecological shifts, directional selection, and recombination rate evolution is that multi-population data on recombination rate is lacking (but see Saleem *et al.* 2001). Recombination rate is laborious to measure, which usually involves producing and genotyping hundreds of recombinant progenies with a large number of genetic markers to generate only a single genetic map. Such practice is difficult to scale up to population-level studies. Thus, current estimates of recombination rates for most species are derived from the average recombination rates in the two lineages used for crossing-based map construction. Often the number of genetic maps for a single species remains below a handful except for some heavily studied model organisms and economically important crops and animals, yielding low statistical power for rigorously investigating the driving forces of interspecific differentiation of recombination rate in a population-genetic framework.

If a genetic map is constructed using computational methods based on linkage disequilibrium through population sequencing, we can obtain estimates of population recombination rates (4·Ne·c), which is an average of two sexes over large span of evolutionary time (McVean et al. 2004). The fact that population recombination rate is scaled by effective population size (Ne) makes it difficult to directly estimate recombination rate and can confound comparisons between diverging populations that may have distinct demographic histories. Moreover, studies using this method to compare multiple populations remain rare beyond the classic models, mostly with a focus on intragenomic variation and between-sex differences (e.g., Auton *et al.,* 2012; Bherer *et al.,* 2017). Therefore, we argue that the lack of understanding on the population- and individual-level variation of recombination rates ought to be addressed if we are to dissect the genetic basis of recombination rate variation.

The emergence of novel genomic sequencing techniques such as whole-genome sequencing of single-sperm cells (Xu et al. 2015) provides an efficient solution to estimating population-level recombination variation (albeit it only measures male-specific recombination rate). Taking advantage of this approach to investigate how ecological shifts and directional selection impact recombination rate, this study examines the male-specific recombination rate in two ecologically distinct, incipient microcrustacean species *Daphnia pulex* and *D. pulicaria*.

A well-known characteristic of the *Daphnia* system resides in its cyclically parthenogenetic reproduction. Under favorable environmental conditions female *Daphnia* produces directly developing embryos (i.e., live birth of neonates released from brood pouch) via apomictic parthenogenesis, generating genetically identical, diploid daughters. However, stressful conditions, e.g., food shortage (Deng 1996) and decrease in temperature, triggers *Daphnia* females to switch to sexual reproduction and also to parthenogenetically produce males via environmental

sex determination (Olmstead and Leblanc 2002). The parthenogenetic production of males allows us to amass a large number of males and sperm cells of the same genotype to examine recombination rate.

As members of the *D. pulex* species complex, *D. pulex* and *D. pulicaria* are estimated to have started diverging from 800,000 – 2,000,000 years ago (Colbourne and Hebert 1996; Omilian and Lynch 2009; Cristescu et al. 2012). These two species are morphologically nearly indistinguishable (Brandlova et al. 1972) but occupy distinct, overlapping freshwater habitats in North America, with *D. pulex* mostly living in ephemeral fishless ponds and *D. pulicaria* inhabiting stratified permanent lakes. Importantly, population genetic data suggest that the divergence of *D. pulicaria* from *D. pulex* most likely involved a habitat transition event from pond to lake systems (Cristescu et al. 2012).

As stratified permanent lakes and ephemeral ponds pose distinct selection regimes (e.g., distinct predators, environmental factors), these two species have most likely undergone strong local adaptation and divergent selection in their distinct habitats, resulting in clear physiological and behavioral differences. For example, compared to *D. pulicaria*, *D. pulex* grows faster to a larger size, reproduces at an earlier age. Also, *D. pulicaria* exhibits diurnal vertical migration in lakes, whereas *D. pulex* displays no such behavior. Interestingly, the frequency of sexual reproduction is also different between the two. *Daphnia pulex* goes through sexual reproduction producing resting eggs before ponds dry up in early summer every year, whereas *D. pulicaria* can persist in lakes largely without sex for a few years (Dudycha and Tessier 1999; Cáceres and Tessier 2004; Dudycha 2004). Notably, prezygotic isolation has developed between these two species (Deng 1997), with *D. pulex* switching to sexual reproduction at long-day hours (16 hours/day) and *D. pulicaria* switching to sexual at short-day hours (10 hours/day). Nonetheless, *D. pulex* and *D.*

*pulicaria* can still generate fertile cyclically parthenogenetic $F_1$ offspring in laboratory crossing experiments, indicating the absence of complete reproductive isolation (Heier and Dudycha 2009).

In this study we examine whether neutral evolution (i.e., genetic drift) is sufficient to explain the divergence of meiotic recombination rate between these two species, with the alternative hypothesis being that directional selection involved in ecological shifts better explains the between-species divergence. We estimated recombination rate for a 1.5-Mbp on linkage group 8 and a 0.5-Mbp region on linkage group 9 in three geographically isolated populations of each species. Most interestingly, our results yield strong support for significantly higher recombination rate in *D. pulicaria* than in *D. pulex*, and the between-species divergence in recombination rate cannot be accounted for by genetic drift and is most likely due to directional selection.

## Results

### *Recombination rate estimates*

We performed microsatellite genotyping on whole-genome amplified single-sperm cells to estimate recombination rates for a 1.5-Mbp on linkage group 8 and a 0.5-Mbp region on linkage group 9 in three populations of *D. pulex* and *D. pulicaria* each. Our recombination rate estimates show that *D. pulicaria* tends to recombine at a higher rate than *D. pulex* (Figs. 1 and 2). For the region on linkage group 8 alone, although the mean recombination rate is higher in *D. pulicaria*, no statistically significant difference (t-test P=0.10) is found between the mean of *D. pulex* (16.9 cM, SD=8.4) and that of *D. pulicaria* (28.4 cM, SD=4.5). However, for the region on linkage group 9, the average recombination rate of *D. pulicaria* (mean=24.5, SD=1.4) is higher (t-test P=0.046) than that of *D. pulex* (mean=13.0, SD=6.8). When we compared the recombination rates of both

linkage groups in these two species, *D. pulicaria* has an overall significantly higher recombination rate than *D. pulex* (t-test P=0.006).

Remarkably, our recombination rate estimates show that the within-species recombination rate variation is markedly higher in *D. pulex* than in *D. pulicaria* for regions on both linkage groups. In *D. pulex*, recombination rates of the three sampled populations range from 8.6 Kosambi cM to 25.4 Kosambi cM for the 1.5-Mbp region on linkage group 8 with a nearly 3-fold difference (Fig. 1). On the other hand, within-species variation in *D. pulicaria* for the interval in linkage group 8 is much lower, ranging from 24.5 to 33.3 Kosambi cM among the three examined populations (Fig. 1).

A resembling pattern of distinct within-species variation is also observed for linkage group 9. For *D. pulex*, the map distance of the 0.5-Mbp region on linkage group 9 range between 7.5 cM and 20.6 Kosambi cM, showing a nearly 3-fold difference among populations (Fig. 2). However, for *D. pulicaria*, the recombination rates of the same interval in linkage group 9 from the examined populations show little variation between 22.9 and 25.6 Kosambi cM (Fig. 2).

## Pst-Fst comparison

An important approach for determining whether the divergence of phenotypic traits is neutral is to compare $Q_{st}$ of phenotypic traits and $F_{st}$ of neutral molecular markers. As $F_{st}$ for molecular markers, $Q_{st}$ is a metric measuring the population differentiation for phenotypic traits (Prout and Barker 1993; Spitze 1993). In theory, the $Q_{st}$ of neutral traits on average should be equal to the mean $F_{st}$ of neutral molecular markers (Rogers and Harpending 1992; Whitlock and Mccauley 1999; Whitlock 2008). We calculated $P_{st}$ (Leinonen et al. 2006), an analog of $Q_{st}$, based on the recombination rates of both linkage groups 8 and 9 (see Discussion for the implications of using

$P_{st}$). The mean $P_{st}$ for recombination rate is 0.52 based on our ANOVA analyses of 1000 bootstrap replicates. We also estimated that the average genome-wide $F_{st}$ of four-fold degenerate sites between *D. pulex* and *D. pulicaria* is 0.15. Based on the distribution of $F_{st}$ of four-fold degenerate sites, we simulated the $Q_{st}$ of a neutrally evolving trait to estimate the distribution of the test statistic $Q_{st} - F_{st}$. Interestingly, our $P_{st} - F_{st}$ value is significantly higher than the $Q_{st} - F_{st}$ values of the simulated neutrally evolving trait (Fig. 3, P=0.03), leading us to reject the neutral hypothesis and to conclude that recombination rate divergence between these two species is adaptive.

## Discussion

### *A new model system for studying recombination rate divergence*

Meiotic recombination is one of the most laborious genetic parameters to estimate, with most species having no more than a handful of genetic maps each. Due to the lack of data on population-level recombination rate variation, many theories on the evolution of recombination remain untested in natural populations. Using novel single-sperm genotyping approach, this study examines the hypothesis that directional selection coupled with habitat shift leads to elevated recombination rate in the model system of microcrustacean *Daphnia*. Interestingly, based on $P_{st}$ - $F_{st}$ comparison analysis we find strong evidence that the divergence of recombination rates between these two species is adaptive and unlikely to be explained by genetic drift. With overall significantly higher recombination rates observed in *D. pulicaria* than in *D. pulex*, we argue that the directional selection that led to the local adaptation of *D. pulicaria* to permanent lake habitats most likely shaped the recombination rate divergence, providing unequivocal support to the theory that directional selection leads to elevated recombination rate.

Although prior work identified some empirical support for this theory, these work has largely been restricted to comparing domesticated animals and plants (Ross-Ibarra 2004; Munoz-Fuentes et al. 2015) with their wild progenitors and to examining laboratory populations (Korol and Iliadi 1994; Aggarwal et al. 2015). Although studies investigating interspecific recombination rate divergence are not uncommon (reviewed in Smukowski and Noor (2011)), this kind of studies are usually deficient in an ecological understanding of the speciation process or lack the population-level sampling required for inferring the driving forces of recombination rate differentiation. In contrast, our study is valuable in providing solid evidence in support of this hypothesis from the perspective of incipient species pairs undergoing ecological speciation. More importantly, the sampling strategy of examining multiple populations within each species is powerful in partitioning the recombination rate variation into between- and within-species level, allowing us to use established quantitative genetics methodologies (i.e., $Q_{st}$-$F_{st}$ comparison) to understand the evolutionary forces driving such divergences in recombination rate. We therefore argue that the well-understood ecology (distinct ephemeral pond vs. permanent habitats) and evolutionary history (speciation associated with transition from pond to lake habitats) of *D. pulex* and *D. pulicaria* set up an excellent framework for future in-depth investigation of the evolutionary and genetic basis of divergence in recombination rates.

*Single-sperm sequencing for studying recombination rate variation in emerging systems*

A major hurdle in studying recombination rate variation is the laborious process of generating genetic maps. This study greatly benefited from the novel whole-genome sequencing technique developed for single-sperm cells (Xu et al. 2015; Xu and Young 2017). Single sperm sequencing emerged in 1980's as a methodology for estimating localized recombination rates (Li et al. 1988; Cui et al. 1989). Nonetheless, empowered by whole-genome sequencing technologies, this

technique has recently been applied to human and mouse to examine whole-genome recombination patterns (Lu et al. 2012; Wang et al. 2012; Hinch et al. 2019).

We note that as collecting a large number of sperm/pollen cells is feasible in many species, our experimental procedure for single-sperm sequencing/genotyping can be applied to other emerging model systems. Although our protocol uses flow cytometry to isolate single cells and relies on whole-genome amplification of single cells, these are currently common laboratory procedures. We hope that an increasing number of researchers will take advantage of this approach to examine the divergence of recombination rates in a diverse set of emerging model systems with interesting ecological attributes.

*$P_{st}$-$F_{st}$ comparison*

We used $P_{st}$-$F_{st}$ comparison to determine whether the divergence of recombination rate between *D. pulex* and *D. pulicaria* is adaptive. This test is based on the observation that for neutral phenotypic traits that are controlled by purely additive genes the mean $Q_{st}$ (we used $P_{st}$ as a surrogate for $Q_{st}$) is equal to the mean $F_{st}$ of neutral genetic loci (Lande 1992; Whitlock 1999). Although this observation is based on several assumptions and this study likely violated some of them, we argue that the strong evidence pointing to the adaptive nature of the observed divergence in recombination rate is unlikely compromised (see below).

An important assumption of $Q_{st}$=$F_{st}$ for neutral phenotypic traits is that the loci from which $F_{st}$ is derived should be neutral. Although there have been concerns about whether $Q_{st}$=$F_{st}$ when the $F_{st}$ is based on markers such as microsatellites that have high mutation rate (Hendry 2002), the use of SNPs in our study alleviates this concern. Furthermore, despite that in other species four-fold degenerate sites have been shown to experience purifying selection, population genomic

analyses of *D. pulex* show that these sites evolve in a nearly neutral fashion (Lynch et al. 2017). Therefore, our use of genome-wide four-fold degenerate sites (n= 94711) should provide a meaningful estimate of the mean $F_{st}$ of neutral sites.

Our analysis differs from the standard $Q_{st}$-$F_{st}$ analysis in the use of $P_{st}$ as a substitute of $Q_{st}$. However, this study differs significantly from studies directly collecting phenotypic data from the field because the recombination rates were estimated in a common garden experiment. As recombination rate is known to be of great phenotypic plasticity due to biotic and abiotic factors such as age and temperature (Hunter, Robinson, et al. 2016; Lloyd et al. 2018), we estimated recombination rates from 2-week old males that were maintained under controlled temperature and photoperiod. Therefore, the obtained $P_{st}$ value is unlikely to be inflated by environmental variance.

Because $Q_{st}$ is defined based on additive variance of traits, one may wonder whether the $P_{st}$ - $F_{st}$ test in this study is biased towards rejecting the neutral hypothesis. Based on previous work that examines how dominance and epistatic effects may affect this test, we argue that our results are unlikely to be biased. Although the genetic basis of recombination rate variation (e.g., relative contribution of additive variance, dominance effects, and epistasis) is poorly understood, we consider the potential impact of epistasis and dominance effects in turns. It is true that our $P_{st}$ estimates could be affected by dominance and epistasis. However, it has been shown that epistasis tends to produce $Q_{st}$ values less than neutral $F_{st}$ (Whitlock 1999). Similarly, dominance makes $Q_{st}$ equal or less than neutral $F_{st}$ under the assumption of an island model (Goudet and Büchi 2006; Goudet and Martin 2007). Even though dominance under limited demographic circumstances can make $Q_{st}$ of neutral traits exceed neutral $F_{st}$, this is unlikely for traits affected by multiple loci (Goudet and Martin 2007) such as recombination rates. Taking all these into consideration, we argue that our use of $P_{st}$ in this study most likely makes our test conservative.

Due to limited resources this study only examined the recombination variation of two genomic regions in males. Despite the promising evidence that the positive selection is responsible for elevated recombination rate in *D. pulicaria*, it remains unclear whether this is true for the genome-wide variation in males and whether this is true for female-specific recombination rate. Sex-specific difference in recombination rates often shows that females recombine more frequently than males. For example, human females on average recombine 1.6 times as much as males (Kong et al. 2010), and sticklebacks show a similar pattern (Sardell et al. 2018). Considering that a minimum of one recombination event per chromosome must occur in meiosis and too many recombination are detrimental to genomic integrity (Ritz et al. 2017), it is likely that recombination rate in males has more potential room to be fine-tuned by natural selection compared to female recombination rate.

*Within-species recombination rate divergence*

With much of the focus of this study probing whether between-species divergence in recombination rate is adaptive, it is necessary to provide some explanation about the contrasting pattern of within-species divergence in these two species. As mentioned in the Results, the intra-specific recombination rate of *D. pulex* varies by nearly 3-fold for both linkage groups 8 and 9, whereas the intraspecific variation within *D. pulicaria* is much lower with a ~1.3-fold difference on linkage group 8 and little variation on linkage group 9 (Figs. 1 and 2). The within-species divergence in *D. pulex* is larger than all the currently available within-species divergence (reviewed in Ritz *et al*. 2017), such as ~1.6-fold variation in both sexes of human (Coop et al. 2008), 1.9 fold in mice (Dumont et al. 2009), 1.1-2 fold in *Drosophila* (Brooks and Marks 1986; Hunter, Huang, et al. 2016), 1.3 fold in *Arabidopsis* (Sanchez-Moran et al. 2002), whereas the within-species divergence in *D. pulicaria* is in line with these available estimates.

One plausible explanation for this drastic difference between these two species is that selection pressure for maintaining recombination rates among different *D. pulicaria* populations is much more uniform than among *D. pulex* populations. To better understand this, we can use results of previous work on how spatially heterogeneous selection pressure influences the evolution recombination rate (Lenormand and Otto 2000). Regardless of the forms of epistasis and linkage disequilibrium and the amount of linkage between recombination rate modifier and the selected loci, when environmental selection pressures vary between populations with frequent migration it is predicted that more variation in recombination rate is expected in populations inhabiting highly spatially variable environments (Lenormand and Otto 2000). Although it is often said that the typical habitat of *D. pulex* is ephemeral pond habitats, we have to acknowledge that ecological conditions of each pond population probably differ substantially in terms of pond sizes, depths, hydrological conditions, habitat heterogeneity, predators, and other biotic and abiotic factors. On the other hand, the ecology of the different stratified permanent lake habitats of *D. pulicaria* may differ to a lesser extent. We therefore hypothesize that the greater variability of recombination within *D. pulex* is likely due to the greater amount of heterogeneity among the pond habitats. This hypothesis is certainly worth future investigation by examining a large number of populations of each species, which can provide insight into how spatially heterogeneous selection shapes the evolution of recombination rates.

## Materials and Methods

### *Daphnia culture and sperm extraction*

Males were collected for three isolates of *D. pulex* and *D. pulicaria* each (Table 1). Each individual represents a distinct population. To induce production of males, mature females with early sign of

carrying broods were collected and cultured at 20 °C in artificial lake water (Kilham et al. 1998) containing 400 nM Methyl Farnesoate (MF), a juvenile hormone that determines the sex of *Daphnia* offspring (Olmstead and Leblanc 2002). They were fed *ad libitum* with a suspension of *Scenedesmus obliquus*, and the offspring were screened for males. A total of 15-18 males were collected from each clone and were maintained in the lab for two weeks before sperm collection.

For analyzing recombination rate of each *Daphnia* isolate, we collected sperm from all the identified males because they have identical genotype. To extract sperm, each male immersed in a drop of double-distilled water ($ddH_2O$) was gently pressed with a cover-slip on a microscope slide. The $ddH_2O$ surrounding each individual was collected using Sigmacote-washed capillary needles and mouth pipettes into a 1.5 mL microcentrifuge tubes containing 50 µl of 1x PBS solution (Xu et al. 2015). To facilitate the sorting of single sperm cells by flow cytometry, we stained sperm cells using 8 µl of Hoechst 33342 (100 µg/mL) (Sigma-Aldrich) and incubated the sample in the dark at room temperature for 2 hours.

*Single-sperm cell sorting*

A BD FACS Aria-II cell sorter was used to isolate single sperm cells into individual wells of 96-well PCR plates containing cell lysis buffer. The specific settings of the FACS Aria II instrument were 488 nm 100 mW laser for light scatter detection and 355 nm 20 mW for Hoechst detection. A nozzle of 70 mm was used at 45 psi, and FSC-PMT was used for optimal small particle discrimination.

Each well of the PCR plate contained 5 µl of lysis buffer consisting of Tris (30 mM), EDTA (2 µM), potassium chloride (20 µM), Triton (0.2%), DTT (40 mM), and protease/ Proteinase K

(2.5 µg/µl). Cell lysis was performed in a thermal cycler at 50 °C for 3 hours, 75 °C for 20 minutes, and 80 °C for 5 minutes.

*Whole-Genome Amplification*

To obtain enough DNA from each sperm for genotyping, the lysed single sperm cell was used for MALBAC (Multiple annealing and looping-based amplification) whole-genome amplification (Zong et al. 2012). MALBAC consists of a pre-amplification stage and a standard PCR amplification. The pre-amplification is initiated with random primers, each having a common 27-nucleotide sequence (5′-GTGAGTGATGGTTGAGGTAGTGTGGAG-3′) and 8 variable nucleotides that can evenly hybridize to the templates.

*Pre-amplification stage*

A solution of 3.0 µl ThermoPol buffer (New England Biolabs), 1 µl dNTPs (10 mM), 0.75 ul each of two primers NT and NG (10 µM), and 19.5 µl $H_2O$ was added to each sperm sample. The samples were incubated at 95 °C for 5 min and quenched immediately on ice. 0.5 µl of *Bst* large fragment polymerase (New England Biolabs) was added to each sample and the following thermal amplification regime is performed: 10 °C for 45 sec, 15 °C for 45 sec, 20 °C for 45 sec, 30 °C for 45 sec, 40 °C for 45 sec, 50 °C for 45 sec, 65 °C for 2 min, 95 °C for 20 sec, followed by quenching on ice. Subsequently, five cycles of pre-amplification cycles were performed, consisting of 10 °C for 45 sec, 15 °C for 45 sec, 20 °C for 45 sec, 30 °C for 45 sec, 40 °C for 45 sec, 50 °C for 45 sec, 65 °C for 2 min, 95 °C for 20 sec, and 58 °C for 40 sec, followed by quenching on ice. 0.5 µl *Bst* large fragment polymerase was added to each sample before carrying out the next cycle.

*Standard PCR amplification stage*

A standard PCR amplification was performed on the amplicons from the pre-amplification stage using the 27mer as primer (5′-GTGAGTGATGGTTGAGGTAGTGTGGAG-3′) to generate the 1-2 μg DNA required for downstream genotyping. Each reaction consisted of the product from the pre-amplification, 3 μl ThermoPol Buffer (New England Biolabs), 1 μl dNTPs (10 mM), 23.5 ml $H_2O$, 1.5 μl 27mer (10 μM), and 1 μl DeepVentR exo- polymerase (New England Biolabs). The PCR thermal regime consisted of 22 rounds of 94 °C for 20 sec, 59 °C for 20 sec, 65 °C for 1 min, 72 °C for 2 min, which was followed by 72 °C for 5 min.

*Recombination rate estimation*

To examine recombination rate variation in *D. pulex* and *D. pulicaria*, we focused on two regions that have ~20 cM genetic distance on linkage groups 8 and 9 from the microsatellite-based genetic map by Cristescu *et al* (2006). For linkage group 8, located between the microsatellite markers d077 and d068, the interval is 1.5 Mbp, whereas on linkage group 9, the region spans ~0.5 Mbp lying between the microsatellite markers d171 and d118.

For detecting recombination events, two heterozygous markers are required. However, the four mapped microsatellite markers (i.e., d077, d068, d171, d118) are not heterozygous in all the *Daphnia* isolates. In cases where any of these markers are homozygous, new heterozygous microsatellites were identified within a 50-kb window centered at the mapped marker and were used for estimating recombination. The web-based platform WebSat (Martins et al. 2009) was used for identifying microsatellite markers and primer designs.

Our microsatellite genotyping followed the strategy outlined by (Schuelke 2000). Briefly, a M13 tail is added to the 5’ prime end of the forward primer, and a M13 sequenced labelled with one of the NED, PET, FAM and VIC fluorescent dye was used in the PCR. The thermal cycling

program for microsatellite amplification consisted of 3 min at 95 °C, 10 cycles of 35 sec at 95 °C, 35 sec at 56 °C (the temperature increased by 1 °C for each cycle) and 45 sec at 72 °C, 30 cycles of 35 sec at 95 °C, 35 sec at 48 °C, 45 sec at 72 °C, and a final 10 min at 72 °C. Fragment analysis was performed on an ABI 3130 Genetic Analyzer (Life Technologies) using 20x diluted PCR product. The four different M13 dyes allowed the pooled genotyping of different markers labelled with different dyes. The genotypes were called using the software GeneMapper 4.0 (Life Technologies).

To estimate the recombination rate for the two intervals of interests, 2-locus genotypes were examined for the pool of genotyped sperm for each *Daphnia* isolate. The two most abundant genotypes were identified as the parental genotypes, whereas the two rare genotypes were derived from recombination events. For example, the two locus genotypes for d077 (alleles: 227 and 232 bp) and d068 (alleles: 337 and 343 bp) are 227/337 (10 sperm cells), 232/343 (10 sperm cells), 232/337 (40 sperm cells), and 227/343 (40 sperm cells). Then, the genotypes 227/337 and 232/343 are recognized as recombinant genotypes with a recombinant frequency of 0.2. The frequency of recombinants is converted to Kosambi centiMorgan map distance. The standard error (SE) of recombination was calculated as $\sqrt{p(1-p)/n}$, where the p represents proportion of recombinant sperm cells and n represents the number of sampled sperm cells.

*$P_{st}$-$F_{st}$ comparison*

To investigate whether the divergence of recombination rate between these two species is adaptive, we performed $P_{st}$-$F_{st}$ comparison analysis. As the divergence of quantitative traits can be shaped by mutation, selection, and genetic drift, various methods have been developed for deciphering

whether the divergence of phenotypic traits is neutral (i.e., can be adequately explained by drift alone) or adaptive. An important approach among these is the comparison of $Q_{st}$ and $F_{st}$ values.

Analogous to the famous $F_{st}$ for measuring population differentiation using molecular markers (reviewed in Holsinger and Weir, (2009)), $Q_{st}$ (Prout and Barker 1993; Spitze 1993) is established as a measure of the genetic differentiation among populations for phenotypic traits. For a neutral quantitative trait with additive genetic basis, its $Q_{st}$ value on average should be equivalent to the mean $F_{st}$ of neutral loci (Rogers and Harpending 1992; Whitlock and Mccauley 1999; Whitlock 2008), providing an important means for distinguishing between neutral and adaptive divergence. Therefore, if the $Q_{st}$ of a trait is significantly higher than the mean $F_{st}$ of neutral loci, it would indicate divergent selection on this trait. On the contrary, if $Q_{st}$ of a trait is significantly smaller than the mean $F_{st}$ of neutral loci, it would indicate stabilizing selection on the trait in the presence of drift. Moreover, identical values of $Q_{st}$ and $F_{st}$ would indicate no evidence for selection acting in a spatially heterogeneous manner.

As specific breeding experimental designs in a common garden environment are required for estimating additive variance that is required for calculating $Q_{st}$, many studies on wild populations used another metric $P_{st}$ that is analogous to $Q_{st}$ (Leinonen et al. 2006). $P_{st}$ is a metric measuring total phenotypic variance (rather than additive variance) among populations, which could be confounded by environmental effects for phenotype data directly collected from the field. Although our recombination rates were measured in a controlled environment and in same-aged males, our experiments did not allow us to estimate the additive variance. Thus, we decided to use $P_{st}$ as a surrogate for $Q_{st}$ in this analysis.

To estimate $P_{st}$ of recombination rate, we used recombination rate data for both chromosomes 8 and 9 to quantify within- and between-species variances using ANOVA in R. This

strategy gave us a larger sample size and more statistical power than examining single linkage groups alone. The between-species variance was calculated using the equation Var(s) =(MSs - MSe)/n, where MSs and MSe represent the mean squares of between- and within-species, respectively, and n represents the number of data points for each species (n=6). The within-species variance Var(e) is equal to MSe, which is the mean squares of within-species. The $P_{st}$ value is calculated using the equation Var(s)/[Var(s) + 2Var(e)]. A total of 1000 bootstrap replicates were generated and analyzed using ANOVA to estimate the distribution and mean value of $P_{st}$.

To determine whether the divergence of recombination rates between *D. pulex* and *D. pulicaria* is adaptive, we followed the approach of (Whitlock and Guillaume 2009) to examine the difference between Qst and Fst with Qst – Fst as the test statistic. This approach rests upon the notion that the mean $Q_{st}$ value of neutral traits is expected to be the same as the mean $F_{st}$ of neutral makers under certain assumptions (Whitlock and Guillaume 2009). The Fst between *D. pulex* and *D. pulicaria* was estimated using genome-wide four-fold degenerate sites (n=94711) extracted from the whole-genome sequences of these isolates from (Tucker et al. 2013). To simulate the distribution of the $Q_{st}$ of a neutral trait, we calculated the expected between-species variance Var(s) using the formula $Var(s) = 2F_{st.bootstrap} \cdot Var(e)/(1-F_{st.bootstrap})$, where $F_{st.bootstrap}$ is the mean value of a bootstrap sample of four-fold degenerate sites and Var(e) is the observed within-species variance. Then we simulated the between- and within-species variance, Var(s).hat and Var(e).hat, respectively. Var(e).hat was calculated as $\frac{observed\ Var(e)}{DFwithin}$ multiplied by a random number drawn from a chi-square distribution with the degree of freedom at within-species level (i.e, DFwithin), whereas Var(s).hat was simulated as $\frac{expected\ Var(s)}{DFbetween}$ multiplied by a random number drawn from a chi-square distribution with the degree of freedom at between-species level (i.e., DFbetween). Furthermore, the simulated $Q_{st}$ was calculated as Var(s).hat/[Vars(s).hat + 2Var(e).hat]. The

simulation was repeated for 10,000 times to obtain a distribution of the test metric $Q_{st} - F_{st}$. Lastly, we determined whether the observed $P_{st} - F_{st}$ differs significantly from the neutral expectations by identifying the quantile of simulated distribution that had higher values than the observation, which gave us the P value of the test. This procedure was perform using a R script slightly modified from Lind *et al.* (2011).

Acknowledgements

We thank Marelize Snyman, Trung Huynh, Hongjun Wang, and Thinh Pham for their help and constructive comments on this manuscript. This work is partly supported by start-up funds from University of Texas at Arlington and is partly supported by NIH grant R35GM133730 to SX.

References

Aggarwal DD, Rashkovetsky E, Michalak P, Cohen I, Ronin Y, Zhou D, Haddad GG, Korol AB. 2015. Experimental evolution of recombination and crossover interference in *Drosophila* caused by directional selection for stress-related traits. *BMC Biol.* 13:1–14.

Auton A, Fledel-Alon A, Pfeifer S, Venn O, Ségurel L, Street T, Leffler EM, Bowden R, Aneas I, Broxholme J, et al. 2012. A fine-scale chimpanzee genetic map from population sequencing. *Science.* 336:193–198.

Auton A, Rui Li Y, Kidd J, Oliveira K, Nadel J, Holloway JK, Hayward JJ, Cohen PE, Greally JM, Wang J, et al. 2013. Genetic recombination is targeted towards gene promoter regions in dogs. *PLoS Genet.* 9.

Barton N. 1995a. Linkage and the limits to natural selection. *Genetics* 140:821–841.

Barton N. 1995b. A general model for the evolution of recombination. *Genet. Res.* 65:123–144.

Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, Massy B de. 2010. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Nord. Stud. Alcohol Drugs* 327:836–840.

Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* 356:242–244.

Bherer C, Campbell CL, Auton A. 2017. Refined genetic maps reveal sexual dimorphism in human meiotic recombination at multiple scales. *Nat. Commun.* 8.

Booker TR, Ness RW, Keightley PD. 2017. The recombination landscape in wild house mice inferred using population genomic data. *Genetics* 207:297–309.

Brand CL, Cattani MV, Kingan SB, Landeen EL, Presgraves DC. 2018. Molecular evolution at a meiosis gene mediates species differences in the rate and patterning of recombination. *Curr. Biol.* 28:1289-1295.e4.

Brandlova J, Brandl Z, Fernando CH. 1972. The Cladocera of Ontario with remarks on some species and distribution. *Can. J. Zool.* 50:1373–1403.

Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova G V. 2012. Genetic recombination is directed away from functional genomic elements in mice. *Nature* 485:642–645.

Brooks LD, Marks RW. 1986. The organization of genetic variation for recombination in *Drosophila melanogaster*. *Genetics* 114:525–547.

Burt A, Bell G. 1987. Mammalian chiasma frequencies as a test of two theories of recombination. *Nature* 326:803–805.

Butlin RK. 2005. Recombination and speciation. *Mol. Ecol.* 14:2621–2635.

Cáceres CE, Tessier AJ. 2004. Incidence of diapause varies among populations of *Daphnia pulicaria*. *Oecologia* 141:425–431.

Cavalier-Smith T. 2002. Origins of the machinery of recombination and sex Introduction: the fundamental difference in recombination biology of bacteria and eukaryotes. *Heredity (Edinb).* 88:125–141.

Charlesworth B. 1976. Recombination modification in a fluctuating environment. *Adv. Appl. Probab.* 8:2–4.

Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134:1289–1303.

Choi K, Zhao X, Kelly KA, Venn O, Higgins JD, Yelina NE, Hardcastle TJ, Ziolkowski PA, Copenhaver GP, Franklin FCH, et al. 2013. *Arabidopsis* meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. *Nat. Genet.* 45:1327–1338.

Chowdhury R, Bois PRJ, Feingold E, Sherman SL, Cheung VG. 2009. Genetic analysis of variation in human meiotic recombination. *PLoS Genet.* 5.

Colbourne JK, Hebert PDN. 1996. The systematics of North American *Daphnia* (Crustacea: Anomopoda) : A molecular phylogenetic approach. *Philos. Trans. R. Soc. B Biol. Sci.* 351:349–360.

Comeron JM, Kreitman M, Aguadé M. 1999. Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*. *Genetics* 151:239–249.

Coop G, Wen X, Ober C, Pritchard JK, Przeworski M. 2008. High resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among Humans. *Science.* 319:1395–1398.

Cristescu ME, Constantin A, Bock DG, CÚceres CE, Crease TJ. 2012. Speciation with gene flow and the genetics of habitat transitions. *Mol. Ecol.* 21:1411–1422.

Cui X, Li H, Goradia TM, Lange K, Kazazian HH, Galas D, Arnheim N. 1989. Single-sperm typing: Determination of genetic distance between the $^{G}\gamma$-globin and parathyroid hormone loci by using the polymerase chain reaction and allele-specific oligomers. *Proc. Natl. Acad. Sci. U. S. A.* 86:9389–9393.

Cutter AD, Payseur BA. 2013. Genomic signatures of selection at linked sites: Unifying the disparity among species. *Nat. Rev. Genet.* 14:262–274.

Dapper AL, Payseur BA. 2017. Connecting theory and data to understand recombination rate evolution. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20160469.

Deng H-W. 1996. Environmental and genetic control of sexual reproduction in *Daphnia* . *Heredity (Edinb).* 76:449–458.

Deng H-W. 1997. Photoperiodic response of sexual reproduction in the *Daphnia pulex* group is reversed in two distinct habitats. *Limnol. Oceanogr.* 42:609–611.

Dolgin ES, Charlesworth B. 2008. The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics* 178:2169–2177.

Dudycha JL. 2004. Mortality dynamics of *Daphnia* in contrasting habitats and their role in ecological divergence. *Freshw. Biol.* 49:505–514.

Dudycha JL, Tessier AJ. 1999. Natural Genetic Variation of Life Span, Reproduction, and Juvenile Growth in *Daphnia* . *Evolution (N. Y).* 53:1744.

Dumont BL, Broman KW, Payseur BA. 2009. Variation in genomic recombination rates among heterogeneous stock mice. *Genetics* 182:1345–1349.

Duret L, Arndt PF. 2008. The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet.* 4.

Felsenstein J, Yokoyama S. 1976. The evolutionary advantage of recombination. II. Individual selection for recombination. *Genetics* 83:845–859.

Goudet J, Büchi L. 2006. The effects of dominance, regular inbreeding and sampling design on $Q_{st}$, an estimator of population differentiation for quantitative traits. *Genetics* 172:1337–1347.

Goudet J, Martin G. 2007. Under neutrality, $Q_{st} \leq F_{st}$ when there is dominance in an island model. *Genetics* 176:1371–1374.

Grey C, Barthès P, Friec G, Langa F, Baudat F, de Massy B. 2011. Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. *PLoS Biol.* 9:1–9.

Halldorsson B V., Palsson G, Stefansson OA, Jonsson H, Hardarson MT, Eggertsson HP, Gunnarsson B, Oddsson A, Halldorsson GH, Zink F, et al. 2019. Human genetics: Characterizing mutagenic effects of recombination through a sequence-level genetic map. *Science.* 363.

Hassold T, Hunt P. 2001. To err (meiotically) is human: The genesis of human aneuploidy. *Nat. Rev. Genet.* 2:280–291.

Heier CR, Dudycha JL. 2009. Ecological speciation in a cyclic parthenogen: Sexual capability of experimental hybrids between *Daphnia pulex* and *Daphnia pulicaria*. *Limnol. Oceanogr.* 54:492–502.

Hendry AP. 2002. QST > = ≠ ST? *Trends Ecol. Evol.* 17:502.

Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genet. Res. (Camb).* 8:269–294.

Hinch AG, Zhang G, Becker PW, Moralli D, Hinch R, Davies B, Bowden R, Donnelly P. 2019. Factors influencing meiotic recombination revealed by whole-genome sequencing of single sperm. *Science.* 363.

Holsinger KE, Weir BS. 2009. Genetics in geographically structured populations: Defining, estimating and interpreting $F_{ST}$. *Nat. Rev. Genet.* 10:639–650.

Hunter CM, Huang W, Mackay TFC, Singh ND. 2016. The Genetic Architecture of Natural Variation in Recombination Rate in Drosophila melanogaster. *PLoS Genet.* 12:1–31.

Hunter CM, Robinson MC, Aylor DL, Singh ND. 2016. Genetic background, maternal age, and interaction effects mediate rates of crossing over in *Drosophila melanogaster* females. *G3 Genes, Genomes, Genet.* 6:1409–1416.

Johnston SE, Bérénos C, Slate J, Pemberton JM. 2016. Conserved genetic architecture underlying individual recombination rate variation in a wild population of soay sheep (*Ovis aries*).

Kent T V., Uzunović J, Wright SI. 2017. Coevolution between transposable elements and recombination. *Philos. Trans. R. Soc. B Biol. Sci.* 372.

Kilham SS, Kreeger DA, Lynn SG, Goulden CE, Herrera L. 1998. COMBO: A defined freshwater culture medium for algae and zooplankton. *Hydrobiologia* 377:147–159.

Kondrashov AS. 1988. Deleterious mutations and the evolution of sex. *Nature* 336:435–440.

Kong A, Thorleifsson G, Frigge ML, Masson G, Gudbjartsson DF, Villemoes R, Magnusdottir E, Olafsdottir SB, Thorsteinsdottir U, Stefansson K. 2014. Common and low-frequency variants associated with genome-wide recombination rate. *Nat. Genet.* 46:11–16.

Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir Aslaug, Walters GB, Jonasdottir Adalbjorg, Gylfason A, Kristinsson KT, et al. 2010. Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* 467:1099–1103.

Kong A, Thorleifsson G, Stefansson H, Masson G, Helgason A, Gudbjartsson DF, Jonsdottir GM, Gudjonsson SA, Sverrisson S, Thorlacius T, et al. 2008. Sequence variants in the *RNF212* gene associate with genome-wide recombination rate. *Science.* 319:1398–1401.

Korol AB, Iliadi KG. 1994. Increased recombination frequencies resulting from directional selection for geotaxis in *Drosophila*. *Heredity (Edinb).* 72:64–68.

Lande R. 1992. Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. *Evolution (N. Y).* 46:381–389.

Leinonen T, Cano JM, Mäkinen H, Merilä J. 2006. Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of threespine sticklebacks. *J. Evol. Biol.* 19:1803–1812.

Lenormand T, Otto SP. 2000. The evolution of recombination in a heterogeneous environment. *Genetics* 156:423–438.

Lercher MJ, Hurst LD. 2002. Human SNP variability and mutation rate are higher in regions of high recombination. *Trends Genet.* 18:337–340.

Li H, Gyllensten UB, Cui X, Saiki RK, Erlich HA, Arnheim N. 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* 335:414–417.

Lind MI, Ingvarsson PK, Johansson H, Hall D, Johansson F. 2011. Gene flow and selection on phenotypic plasticity in an island system of *rana temporaria*. *Evolution (N. Y).* 65:684–697.

Liu H, Maclean CJ, Zhang J. 2019. Evolution of the yeast recombination landscape. *Mol. Biol. Evol.* 36:412–422.

Lloyd A, Morgan C, Franklin FCH, Bomblies K. 2018. Plasticity of meiotic recombination rates in response to temperature in *Arabidopsis*. *Genetics* 208:1409–1420.

Lu S, Zong C, Fan W, Yang M, Li J, Chapman AR, Zhu P, Hu X, Xu L, Yan L, et al. 2012. Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. *Science.* 338:1627–1631.

Lynch M, Butcher D, Bürger R, Gabriel W. 1993. The mutational meltdown in asexual populations. *J. Hered.* 84:339–344.

Lynch M, Gutenkunst R, Ackerman M, Spitze K, Ye Z, Maruki T, Jia Z. 2017. Population genomics of *Daphnia pulex*. *Genetics* 206:315–332.

Martins W, Lucasa DCS, Neves KFD, Bertioli DJ. 2009. WebSat - A web software for microsatellite marker development. *Bioinformation* 2063:282.

McVean GAT, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. 2004. The fine-scale structure of recombination rate variation in the Human genome. *Science.* 304:581–584.

Mugal CF, Weber CC, Ellegren H. 2015. GC-biased gene conversion links the recombination

landscape and demography to genomic base composition: GC-biased gene conversion drives genomic base composition across a wide range of species. *BioEssays* 37:1317–1326.

Munoz-Fuentes V, Marcet-Ortega M, Alkorta-Aranburu G, Forsberg CL, Morrell JM, Manzano-Piedras E, Soderberg A, Daniel K, Villalba A, Toth A, et al. 2015. Strong artificial selection in domestic mammals did not result in an increased recombination rate. *Mol. Biol. Evol.* 32:510–523.

Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, MacFie TS, Mcvean G, Donnelly P. 2010. Drive against hotspot motifs in primates implicates the *PRDM9* gene in meiotic recombination. *Science.*:876–879.

Olmstead AW, Leblanc GA. 2002. Juvenoid hormone methyl farnesoate is a sex determinant in the crustacean *Daphnia magna*. *J. Exp. Zool.* 293:736–739.

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

Ortiz-Barrientos D, Engelstädter J, Rieseberg LH. 2016. Recombination Rate Evolution and the Origin of Species. *Trends Ecol. Evol.* 31:226–236.

Otto SP. 2009. The evolutionary enigma of sex. *Am. Nat.* 174.

Otto SP, Lenormand T. 2002. Resolving the paradox of sex and recombination. *Nat. Rev. Genet.* 3:252–261.

Otto SP, Michalakis Y. 1998. The evolution of recombination in changing environments. *Trends Ecol. Evol.*

Pan J, Sasaki M, Kniewel R, Murakami H, Blitzblau HG, Tischfield SE, Zhu X, Neale MJ, Jasin M, Socci ND, et al. 2011. A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* 144:719–731.

Pâques F, Haber JE. 1999. Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiol. Mol. Biol. Rev.* 63:349–404.

Poissant J, Hogg JT, Davis CS, Miller JM, Maddox JF, Coltman DW. 2010. Genetic linkage map of a wild genome: Genomic structure, recombination and sexual dimorphism in bighorn

sheep. *BMC Genomics* 11.

Pouyet F, Mouchiroud D, Duret L, Sémon M. 2017. Recombination, meiotic expression and human codon usage. *Elife* 6:1–19.

Prout T, Barker JSF. 1993. F statistics in *Drosophila buzzatii*: Selection, population size and inbreeding. *Genetics* 134:369–375.

Rice WR. 2002. Experimental tests of the adaptive significance of sexual recombination. *Nat. Rev. Genet.* 3:241–251.

Ritz KR, Noor MAF, Singh ND. 2017. Variation in Recombination Rate: Adaptive or Not? *Trends Genet.* 33:364–374.

Rizzon C, Marais G, Gouy M, Biemont C. 2002. Recombination rate and the distribution of transposable elements in the *Drosophila melanogaster* genome. *Genome Res.* 12:400–407.

Rogers AR, Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* 9:552–569.

Ross-Ibarra J. 2004. The evolution of recombination under domestication: A test of two hypotheses. *Am. Nat.* 163:105–112.

Saleem M, Lamb BC, Nevo E. 2001. Inherited differences in crossing over and gene conversion frequencies between wild strains of *Sordaria fimicola* from "Evolution Canyon." *Genetics* 159:1573–1593.

Sanchez-Moran E, Armstrong SJ, Santos JL, Franklin FCH, Jones GH. 2002. Variation in chiasma frequency among eight accessions of *Arabidopsis thaliana*. *Genetics* 162:1415–1422.

Sandor C, Li W, Coppieters W, Druet T, Charlier C, Georges M. 2012. Genetic variants in *REC8*, *RNF212*, and *PRDM9* influence male recombination in cattle. *PLoS Genet.* 8.

Sardell JM, Cheng C, Dagilis AJ, Ishikawa A, Kitano J, Peichel CL, Kirkpatrick M. 2018. Sex differences in recombination in sticklebacks. *G3 Genes, Genomes, Genet.* 8:1971–1983.

Schuelke M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* 18:233–234.

Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, Strand AI, Li Q, Raney B, Balakrishnan CN, et al. 2015. Stable recombination hotspots in birds. *Science.* 350:928–932.

Smukowski CS, Noor MAF. 2011. Recombination rate variation in closely related species. *Heredity (Edinb).* 107:496–508.

Spitze K. 1993. Population structure in *Daphnia obtusa*: Quantitative genetic and allozymic variation. *Genetics* 135:367–374.

Tucker AE, Ackerman MS, Eads BD, Xu S, Lynch M. 2013. Population-genomic insights into the evolutionary origin and fate of obligately asexual *Daphnia pulex*. *Proc. Natl. Acad. Sci. U. S. A.* 110:15740–15745.

Wang J, Fan HC, Behr B, Quake SR. 2012. Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell* 150:402–412.

Whitlock MC. 1999. Neutral additive genetic variance in a metapopulation. *Genet. Res.* 74:215–221.

Whitlock MC. 2008. Evolutionary inference from $Q_{ST}$. *Mol. Ecol.* 17:1885–1896.

Whitlock MC, Guillaume F. 2009. Testing for spatially divergent selection: Comparing $Q_{ST}$ to $F_{ST}$. *Genetics* 183:1055–1063.

Whitlock MC, Mccauley DE. 1999. Indirect measures of gene flow and migration: $F_{st} \neq 1/(4Nm + 1)$. *Heredity (Edinb).* 82:117–125.

Xu S, Ackerman MS, Long H, Bright L, Spitze K, Ramsdell JS, Thomas WK, Lynch M. 2015. A male-specific genetic map of the microcrustacean *Daphnia pulex* based on single-sperm whole-genome sequencing. *Genetics* 201:31–38.

Xu S, Young K. 2017. Whole-Genome Haplotyping of Single Sperm of *Daphnia pulex* (Crustacea, Anomopoda).

Zong C, Lu S, Chapman AR, Xie X. 2012. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science.* 338:1622–1626.

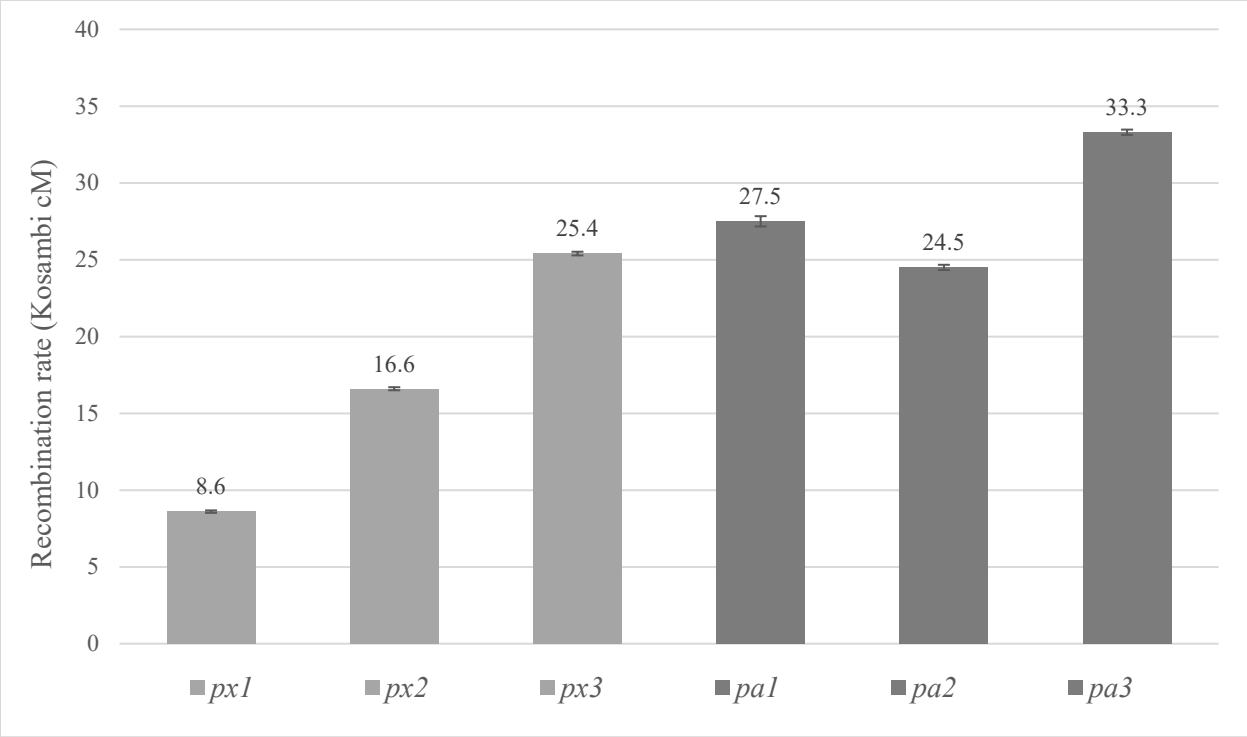


**Figure 1**. Estimated recombination rates for a 1.5-Mbp region on linkage group 8 for three isolates of *D. pulex* (px) and *D. pulicaria* (pa) each. Each grey bar represents the recombination estimate from a specific *Daphnia* isolate with error bar representing standard error.

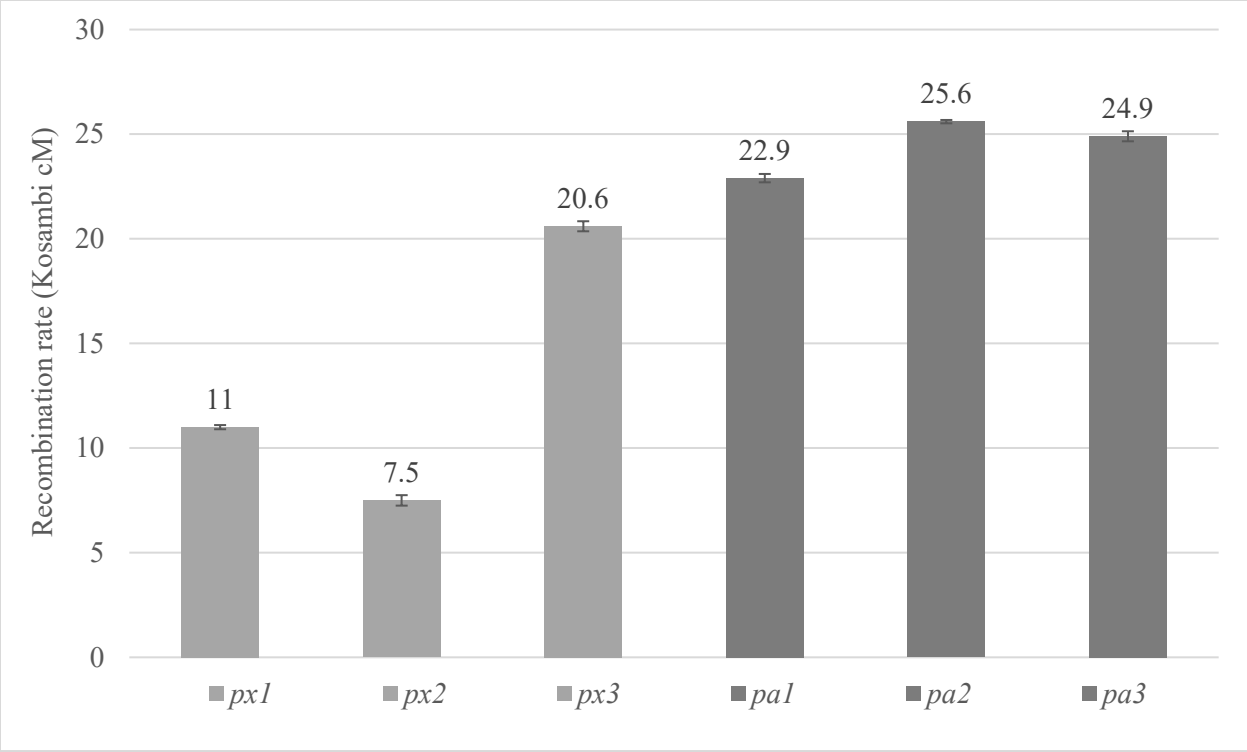


**Figure 2**. Estimated recombination rates for a 0.5-Mbp region on linkage group 9 for three isolates of *D. pulex* (px) and *D. pulicaria* (pa) each. Each grey bar represents the recombination estimate from a specific *Daphnia* isolate with error bar representing standard error.

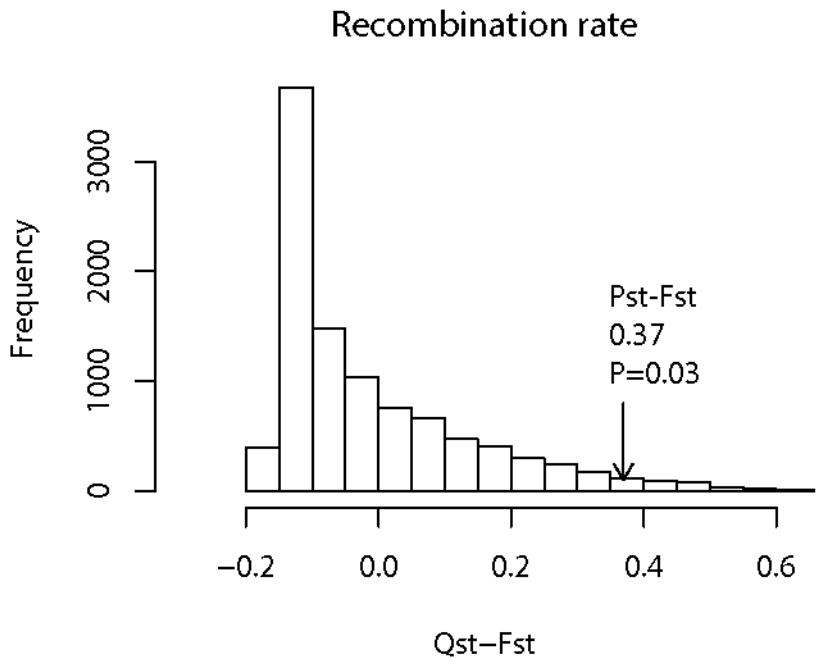


**Figure 3.** The simulated distribution of $Q_{st} - F_{st}$ for a neutral trait and the observed $P_{st} - F_{st}$ (indicated by an arrow) for recombination rate.

**Table 1:** Summary of the *Daphnia* isolates used for the recombination rate estimates.

| **Species** | **Isolates** | **Lab code** | **Location** |
|---|---|---|---|
| *D. pulex* | *px1* | *SW4* | *Illinois* |
| | *px2* | *LPB17* | *Long point, Ontario, Canada* |
| | *px3* | *Tex21* | *42°12, -83°12, Textile Road, Michigan* |
| *D. pulicaria* | *pa1* | *Little Curtis* | *45°43, -122°44, Curtis Lake, Oregon* |
| | *pa2* | *RLSD26* | *44°57, -96°49, Round Lake, South Dakota* |
| | *pa3* | *AroMoose* | *44°50, -69°16, Sebasticook Lake, Maine* |

Chapter 3

# An efficient CRISPR-Cas9 based heritable gene-modification in *D. pulex*

Swatantra Neupane[§], Hongjun Wang[§], Thinh Phu Pham, Sen Xu*

Department of Biology, University of Texas at Arlington, Arlington, Texas, 76019, USA

[§]Equally contributing authors.

*Corresponding author: Sen Xu. Email: sen.xu@uta.edu

**Key words**: *scarlet*, embryos, one-cell stage, phenotype, off-target mutation

Abstract

*Daphnia* offers a variety of benefits for the study of ecology, toxicology, speciation genetics, epigenetics, sex, and recombination. Also, *Daphnia*'s parthenogenetic life cycle allows the genetic study in the absence of confounding genetic differences. The availability of a reference genome has allowed further exploration in the field of genetics in *Daphnia.* Here, we present an efficient genetic engineering tool based on CRISPR-Cas9 editing in *Daphnia* pulex. After meticulous exploration of factors responsible for an efficient gene knockout, we streamlined a technique that would be top of the shelf in the *Daphnia* genomics toolkit. We successfully disrupted an eye-pigment transporter gene, *Scarlet*, that phenotypically stands out as a white-eyed *Daphnia.* We coinjected two RNPs targeting two exons of the *scarlet* gene, creating a heritable mutation. This study is the first reported heritable biallelic gene disruption in *D. pulex* and will help advance the dissection of other genetic factors in *Daphnia.*

## Introduction

CRISPR (clustered regularly interspaced short palindromic repeats) has gained recognition in the genetic modification field for its ease and versatility (Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013). Various revolutionary ambitions have been ignited using this technique; an exemplary is to create a malaria-free mosquito population (Gantz et al., 2015), to cure genetic diseases such as muscular dystrophy (Nelson et al., 2016), and to understand the genetics of a mammalian system (Grunwald et al., 2019). Cas9 nucleases, when guided by the gRNAs to the complementary sequence region, create double-strand breaks (DSBs) at a targeted location. The double-strand breaks are usually short-lived, repaired by either one of the two methods, depending on the availability of a repair template. In the absence of a DNA repair template, the DSBs can be repaired by an error-prone non-homologous end joining (NHEJ) pathway, which results in indels leading to often a disruption of gene function (Rodgers & Mcvey, 2016). In the presence of a homologous DNA repair template, DSBs can be repaired through homology-directed repair (HDR) (Liang et al., 1998; Sekelsky, 2017). CRISPR-cas9 gene editing has revolutionized genomic research and proves to be a valuable tool for emerging model organisms.

One emerging model organism where CRISPR/Cas9 genetic editing could be beneficially exploited is the freshwater crustacean *Daphnia pulex*, reviewed in (Xu et al., 2020). Well-known for its genomic studies scope, population genetic studies, and toxicology, this model organism has a lot to offer to the evolutionary and population genetic field. The development of genetically identical asexual embryos from *Daphnia* provides ideal embryos for genetic editing (Figure 1). *Daphnia* can provide enough clonal asexual embryos which can be easily collected for injection. After microinjection, the injected embryos can quickly hatch into neonates ($G_0$ generation) within 48 hours and have asexual offspring ($G_1$ generation) within ten days.

It would be fair to state that only the tip of the iceberg has been touched in the gene-editing field on the *Daphnia* model system. One of the widely studied species of *Daphnia*, *D. pulex*, has become a critical genomics model system for understanding gene-environment interaction (Altshuler, 2012), epigenetics (Harris et al., 2012), and evolutionary genomics (Lynch et al., 2017). Few studies so far have edited (knockout) the desired gene and obtained a phenotype using either CRISP/Cas9 (Binti Ismail et al., 2018) or RNAi (Hiruta et al., 2013; Kato et al., 2011) and TALEN mediated gene editing (Nakanishi et al., 2016). The escalated necessity to further improve the CRISPR/Cas9 editing in *Daphnia* after successful but limited agility of RNAi and TALEN systems needs prompt addressing. Significant improvement of CRISPR/Cas9 has already opened to discovering the molecular mechanism of environmental sex determination in *Daphnia* magna (Kato et al., 2018).

This study presents a highly efficient way to successfully generate a heritable knockout mutation in the *Daphnia* pulex gene using CRISPR/Cas9 via microinjection. The previous efforts on the species *Daphnia* magna have ~200-million-year divergence from *D. pulex* (Colbourne & Hebert, 1996). They also have a larger body size and embryo size compared to *D. pulex*, making *D. pulex* even challenging to handle. (Hiruta et al., 2018) attempted to knock out the Hox gene *distaless* in *D. pulex*, but biallelic knockout was impossible to generate because it would be lethal.

*Establishing a workflow to achieve an efficient gene modification technique*

We meticulously streamlined a process based on various physical stages and timelines of an asexual oocyte development to answer the timing and location to inject. We looked into the cytological timeline of ameiosis in *Daphnia* to determine when to introduce the biomolecules. In asexual reproduction, female *Daphnia* produces chromosomally unreduced diploid embryos when oocytes undergo modified meiosis (i.e., ameiosis). For a heritable modification, the RNPs

(ribonucleotide proteins – Cas9 fused with gRNAs) should be delivered to the nucleus of a one-cell stage embryo, else it would be a mosaic. We first considered the timing of critical events in the oogenesis of asexually produced *D. pulex* embryos.

During asexual embryo production, first, the germinal vesicles break down when the eggs are still in the ovary. At the same time, females molt. Molting provides an easy visual cue to know that ovulation will start in the next 10-15 mins. After ovulation, the cells enter anaphase I; at this time, the chromosome is supposed to be in the periphery of the cell membrane (Ojima, 1958; Zaffagnini & Sabelli, 1972). Within 20-60 minutes post-ovulation, the chromosomes move to the center of the embryo. The ameiotic division proceeds to anaphase II in 10 minutes post-ovulation. Based on these key timings, the first 10 minutes post ovulation provides a good window for delivering RNPs to establish biallelic heritable modifications. The chromosome is less condensed during this time interval, providing a good chance of RNPs binding to the chromosomes. Oocytes rapidly lose the elasticity of their membrane (chorion), making the early stage perfect for microinjecting since the membranes can withstand the piercing needles (Kato et al., 2011).

The size of the asexual embryo is small (diameter of ~200-300 μm). Considering that the chromosomes move from the periphery of the embryo membrane to the central part of the embryo after the ameiotic division (Ojima, 1958; Hiruta et al., 2010), we aimed to deliver RNPs to the center of the embryo, essentially trapping the reagents within the fat droplets.

We successfully disrupted (knockout) the *Scarlet* gene involved in eye pigmentation on *Daphnia* pulex. *Scarlet* is responsible for transporting eye pigment precursors, and disrupting *scarlet* gene results in white-eyed daphniids providing an easy phenotypic difference to screen for mutants (Ismail et al., 2018). Here, we present the improved microinjection-based CRISPR/Cas9 gene-editing technique in *D. pulex*, increasing efficiency (Figure 2).

Materials and Methods

*CRISPR-Cas9 reagents*

We chose two exons as our target and designed two sgRNAs (Figure 3, Table 2) targeting the exons based on the DNA sequence of a *D. pulex* isolate, EB1. The sgRNAs were custom ordered (Alt-RTM custom sgRNAs, Integrated DNA Technologies). RNP was prepared by mixing each sgRNA and the crRNA (Alt-RTM crRNA, Integrated DNA Technologies) and incubated at 95° C at room temperature to form the guide RNA. The guide RNA was fused with the Cas9 enzyme (catalog no. 1081058, Integrated DNA Technologies) at room temperature for 15 minutes. We co-injected two different RNPs into embryos, each at a concentration of 250ng/ul, with a concentration of 1000ng/ul for the Cas9 enzyme.

*Microinjection equipment*

The microinjection needles were prepared prior to dissection. The needles used were custom made and pulled using Sutter needle puller P1000 (Sutter Instrument, Novato, CA, USA) and aluminosilicate glass capillary, injection parameters (constant pressure, injection time, and injection pressure) were calibrated for each needle. We chose aluminosilicate glass over borosilicate glass because it neatly penetrates the chorion and membrane of *Daphnia* embryos. We pulled microinjection needles to have a final specification of ~1.5-µm tip size and ~7-mm taper length on a P-1000 needle puller (Sutter Instrument), using the following pulling parameters: heat 535 (ramp test value 525 + 10), pull 65, velocity 70, time 200, and pressure 250. Then, the pulled needles were beveled on a BV-10 micropipette beveler (Sutter instrument) over a fine 104D grinding plate (Sutter Instrument) to forge a 30-degree bevel at the tip, followed by the tip thoroughly cleaning with 70% Ethanol and 100% Ethanol to remove any debris.

*Daphnia maintenance and screening*

The microinjection utilized the 2-3 weeks old *Daphnia pulex* females with asexual ovaries as described previously by (Nakanishi et al., 2014). The *Daphnia* were fed algae *ad libitum* every day and cultured in artificial lake water (Kilham et al., 1998) in 18° C and 12:12 (light: dark hours) photoperiod. The babies were cleared out every other day, accompanied by the artificial lake water change once every week to prevent overcrowding that can trigger the *Daphnia* to switch to sexual reproduction.

*Embryo collection*

The healthy sexual females with dark and distinct ovaries were collected prior to the injection. The collected females were then screened until they started molting. Molted females were separated, placing them in a drop of a solution of COMBO lake water containing 60mM sucrose. These animals were regularly checked to see if they started to ovulate (i.e., oocytes starting to enter the brood chamber), which usually occur 10-15 minutes after molting (Figure 2). Once the ovulation was almost ~80% completed (Figure 2), we transferred that individual female to an ice-cold (~1.5°C) COMBO solution with 60mM sucrose.

*Daphnia* was then left in iced sucrose combo water for around 5 minutes, followed by dissection to obtain the embryos for microinjection. We also microinjected embryos with a 2-minute wait time and a 10-minute wait time for comparison, revealing a less hatching rate and no expected phenotype. We dissected the female on the bottom surface of the Petri dish (60 mm x 15 mm, cat no. FB0875713A, Fisher Scientific). Throughout this process, the dissection and dissected embryos were in 60 mM Sucrose containing artificial lake water shown previously by (Hiruta et al., 2018). These dissected embryos were ready for injection.

*Microinjection*

The microinjection parameters in microinjector (FemtoJet 4i; Eppendorf, Hamburg, Germany) and micromanipulator (Injectman 4; Eppendorf, Hamburg, Germany) were adjusted to flow 1-2 nL for each custom pulled needle. At the beginning of each microinjection session, 2-3 uL of the Cas9/sgRNA (RNP mix) was loaded onto the needle from the blunt rear end using the Eppendorf needle loader, Microloader™ (Eppendorf, cat no. 930001007). The injection pressure is generally between 250 and 450 psi, where the constant pressure is between 200 and 400 psi (injection pressure at least 50 psi more than constant pressure), with an injection duration around 0.8 to 1 sec.

We injected immediately after the embryos were dissected to avoid the resuming of ameiotic division at room temperature. We delivered the RNPs near the center of the round or pear-shaped embryo. After injection, we added 60 mM Sucrose containing lake combo to the embryos. We incubated at room temperature for about 30 minutes before transferring them to 12 well-plate containing fresh artificial lake water. At the end of the microinjection session (~1-2 hours), the injected embryos were incubated at 25℃ (12:12 light: dark) for 24 hours. After 48 hours, the neonates hatch.

*Screening for phenotype*

After 48 hours of incubation post-injection, the hatched neonates were screened as tentative knockout mutants for white-eyed individuals. We maintained and cultured the white-eyed individuals. If the asexually produced $G_1$ offspring of the $G_0$ mutant were white-eyed, we concluded the mutation to be heritable. For each heritable knockout mutant, we maintained a clonal line. We performed Sanger sequencing to examine the DNA sequences at the target *scarlet* regions.

## Results

### *Scarlet knockout mutants*

We optimized the CRISPR-Cas9 gene-editing process successfully as validated by the generation of biallelic heritable mutation in *D. pulex*. We targeted the first two exons of the *scarlet* gene by microinjecting a mixture of RNPs carrying two different sgRNAs (sgRNA1 and sgRNA2, sgRNA2 and sgRNA3) (Figure 3). Our microinjection successfully generated *scarlet* germline knockout mutant lines with an average success rate of 3% (Table 1). All of the white-eye we observed were heritable (Figure 8). We also did observe mosaic mutants having partial black pigments in the eyes (Figure 9).

### *Genotypes of scarlet knockout mutants*

We coinjected two RNPs and observed only one mutant with a double-stranded DNA break. This mutant had a ~230bp homozygous deletion (Figure 7). The remainder of the white-eyed mutants carried heterozygous insertion/deletions alleles at the targeted location (i.e., two alleles carry different insertions/deletions, Figure 5-7). The lengths of insertions range between 3 and 237 bp.

## Discussion

Establishing an effective heritable biallelic gene modification technique in *Daphnia* pulex is crucial for future research avenues, and here we meticulously established individual parameters to fine-tune the microinjection technique in *D. pulex*. Carefully establishing each factor guiding the successful microinjection and delivery of reagents to the embryo, we tried to tackle each factor at a time. Dissecting the critical events in oogenesis in asexual reproduction, we tested if the initial 10 minutes post ovulation is an effective time window for microinjecting RNPs (ribonucleoproteins). We further tested if another time window gives us a similar or better

phenotype; an earlier (2-minute) and later (15-minutes) post ovulation timepoints yielded no observable white-eye phenotype along with a low hatching rate post-injection. We find that all individuals showing white-eye phenotype due to insertion or deletion at a *scarlet* locus are germline mutants, as all their asexually produced progenies inherit the white-eye phenotype.

We then transferred the ovulating females to the ice-cold (~1.5 °C) artificial lake water with 60mM sucrose solution after about 80% ovulation. Sucrose helps balance the osmotic pressure between the inside of embryos and the surrounding solution, essentially creating an isotonic environment. The low-temperature solution is crucial as it slows down the embryo's development, providing an extended time window to inject. Furthermore, a transfer at a late or early time point makes the embryo unsuitable for microinjection. Earlier time would have fragile embryos and late time would have hardened embryos. Also, at the 80% ovulated stage, there is a mixture of embryos at different stages: early elongated, oval-shaped, and round embryos allowing us to obtain many embryos.

Then, we optimized the protocol further by testing the storing of the ovulating females in ice-cold water for ~5 minutes before injection. A shorter (1-3 minutes) storage time made them irreversibly vulnerable during dissection and, for a longer storage time (~10 minutes), we observed a low hatching rate.

Finally, we figured out that the needle pulling parameter that gave us a higher hatching rate using aluminosilicate glass capillary with a ~1.5 μm diameter tip size. The aluminosilicate glass capillaries offer more hardiness to penetrate the *Daphnia* embryos' chorion than the borosilicate capillary. Beveling the tip also helps reduce clogging of the injection needle and easy penetration during injection, significantly increasing efficiency.

We set up an efficient microinjection-based delivery of RNPs to introduce heritable biallelic mutation in the *D. pulex*. This toolkit will be of immense value in *Daphnia* research, allowing us to directly test the effects of genetic variants that potentially modulate recombination rate by introducing a subtle change in gene sequences through homology-directed repair. Recombination rate variation is an intensively studied field; however, it still has a lot unexplored in the *Daphnia* system, allowing us to dissect genetic factors responsible for this variation. We can also develop a mutant line that can be clonally maintained. Any casual genetic variants pointed out from prior association mapping that is suspected to regulate recombination rate variation can be identified by introducing identified variants to a common genetic background and testing their effects on recombination rates. Furthermore, the technique established here needs to be further tuned to effectively introduce our choice of nucleotide through homology-directed repair, leaving significant room for further improvement.

References

Altshuler, I. (2012). *Gene expression under multiple stressors in Daphnia pulex*. http://scholar.uwindsor.ca/etd/358

Binti Ismail, N. I., Kato, Y., Matsuura, T., & Watanabe, H. (2018). Generation of white-eyed *Daphnia* magna mutants lacking *scarlet* function. *BioRxiv*, 1–11. https://doi.org/10.1101/313395

Colbourne, J. K., & Hebert, P. D. N. (1996). The systematics of North American *Daphnia* (Crustacea: Anomopoda) : A molecular phylogenetic approach. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *351*(1337), 349–360. https://doi.org/10.1098/rstb.1996.0028

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Hsu, P. D., Wu, X., Jiang, W., & Marraffini, L. a. (2013). Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., … Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science (New York, N.Y.). *Science (New York, N.Y.)*, *339*(6121), 819–823. https://doi.org/10.1126/science.1231143.Multiplex

Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E., & James, A. A. (2015). Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(49), E6736–E6743. https://doi.org/10.1073/pnas.1521077112

Grunwald, H. A., Gantz, V. M., Poplawski, G., Xu, X. S., Bier, E., & Cooper, K. L. (2019). Super-Mendelian inheritance mediated by CRISPR/Cas9 in the female mouse germline

Hannah. *Nature*, *176*(3), 139–148. https://doi.org/10.1038/s41586-019-0875-2.Super-Mendelian

Harris, K. D. M., Bartlett, N. J., & Lloyd, V. K. (2012). *Daphnia* as an Emerging Epigenetic Model Organism. *Genetics Research International*, *2012*, 1–8. https://doi.org/10.1155/2012/147892

Hiruta, C., Kakui, K., Tollefsen, K. E., & Iguchi, T. (2018). Targeted gene disruption by use of CRISPR/Cas9 ribonucleoprotein complexes in the water flea *Daphnia* pulex. *Genes to Cells*, *23*(6), 494–502. https://doi.org/10.1111/gtc.12589

Hiruta, C., Nishida, C., & Tochinai, S. (2010). Abortive meiosis in the oogenesis of parthenogenetic *Daphnia* pulex. *Chromosome Research*, *18*(7), 833–840. https://doi.org/10.1007/s10577-010-9159-2

Hiruta, C., Toyota, K., Miyakawa, H., Ogino, Y., Miyagawa, S., Tatarazako, N., Shaw, J. R., & Iguchi, T. (2013). Development of a microinjection system for RNA interference in the water flea *Daphnia* pulex. *BMC Biotechnology*, *13*, 1–7. https://doi.org/10.1186/1472-6750-13-96

Ismail, N. I. B., Kato, Y., Matsuura, T., & Watanabe, H. (2018). Generation of white-eyed *Daphnia* magna mutants lacking *scarlet* function. *PLoS ONE*, *13*(11), 1–11. https://doi.org/10.1371/journal.pone.0205609

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). *A Programmable Dual-RNA – Guided*. *337*(August), 816–822.

Kato, Y., Perez, C. A. G., Mohamad Ishak, N. S., Nong, Q. D., Sudo, Y., Matsuura, T., Wada,

T., & Watanabe, H. (2018). A 5′ UTR-Overlapping LncRNA Activates the Male-Determining Gene doublesex1 in the Crustacean *Daphnia* magna. *Current Biology*, *28*(11), 1811-1817.e4. https://doi.org/10.1016/j.cub.2018.04.029

Kato, Y., Shiga, Y., Kobayashi, K., Tokishita, S. I., Yamagata, H., Iguchi, T., & Watanabe, H. (2011). Development of an RNA interference method in the cladoceran crustacean *Daphnia* magna. *Development Genes and Evolution*, *220*(11–12), 337–345. https://doi.org/10.1007/s00427-011-0353-9

Kilham, S. S., Kreeger, D. A., Lynn, S. G., Goulden, C. E., & Herrera, L. (1998). COMBO: A defined freshwater culture medium for algae and zooplankton. *Hydrobiologia*, *377*(1–3), 147–159. https://doi.org/10.1023/A:1003231628456

Kumagai, H., Nakanishi, T., Matsuura, T., Kato, Y., & Watanabe, H. (2017). CRISPR/Cas-mediated knock-in via non-homologous end-joining in the crustacean *Daphnia* magna. *PLoS ONE*, *12*(10), 1–12. https://doi.org/10.1371/journal.pone.0186112

Liang, F., Han, M., Romanienko, P. J., & Jasin, M. (1998). Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(9), 5172–5177. https://doi.org/10.1073/pnas.95.9.5172

Lynch, M., Gutenkunst, R., Ackerman, M., Spitze, K., Ye, Z., Maruki, T., & Jia, Z. (2017). Population genomics of *Daphnia pulex*. *Genetics*, *206*(1), 315–332. https://doi.org/10.1534/genetics.116.190611

Mahato, S., Morita, S., Tucker, A. E., Liang, X., Jackowska, M., Friedrich, M., Shiga, Y., & Zelhof, A. C. (2014). Common Transcriptional Mechanisms for Visual Photoreceptor Cell

Differentiation among Pancrustaceans. *PLoS Genetics*, *10*(7). https://doi.org/10.1371/journal.pgen.1004484

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., & Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science*, *339*(6121), 823–826. https://doi.org/10.1126/science.1232033

Nakanishi, T., Kato, Y., Matsuura, T., & Watanabe, H. (2014). CRISPR/Cas-mediated targeted mutagenesis in *Daphnia* magna. *PLoS ONE*, *9*(5). https://doi.org/10.1371/journal.pone.0098363

Nakanishi, T., Kato, Y., Matsuura, T., & Watanabe, H. (2016). TALEN-mediated knock-in via non-homologous end joining in the crustacean *Daphnia* magna. *Scientific Reports*, *6*(November), 1–7. https://doi.org/10.1038/srep36252

Nelson, C. E., Hakim, C. H., Ousterout, D. G., Thakore, P. I., Moreb, E. A., Rivera, R. M. C., Madhavan, S., Pan, X., & Ann, F. (2016). In vivo editing improves muscle function in mouse of DMD. *Science*, *351*(6271), 403–407. https://doi.org/10.1126/science.aad5143.In

Nong, Q. D., Mohamad Ishak, N. S., Matsuura, T., Kato, Y., & Watanabe, H. (2017). Mapping the expression of the sex determining factor Doublesex1 in *Daphnia* magna using a knock-in reporter. *Scientific Reports*, *7*(1), 1–13. https://doi.org/10.1038/s41598-017-13730-4

Rivera, A. S., Pankey, M. S., Plachetzki, D. C., Villacorta, C., Syme, A. E., Serb, J. M., Omilian, A. R., & Oakley, T. H. (2010). Gene duplication and the origins of morphological complexity in pancrustacean eyes, a genomic approach. *BMC Evolutionary Biology*, *10*(1), 1–16. https://doi.org/10.1186/1471-2148-10-123

Rodgers, K., & Mcvey, M. (2016). Error-Prone Repair of DNA Double-Strand Breaks. *Journal of Cellular Physiology*, *231*(1), 15–24. https://doi.org/10.1002/jcp.25053

Sekelsky, J. (2017). DNA repair in Drosophila: Mutagens, models, and missing genes. *Genetics*, *205*(2), 471–490. https://doi.org/10.1534/genetics.116.186759

Xu, S., Pham, T. P., & Neupane, S. (2020). Delivery methods for CRISPR/Cas9 gene editing in crustaceans. *Marine Life Science & Technology*, *2*(1), 1–5. https://doi.org/10.1007/s42995-019-00011-4

Zaffagnini, F., & Sabelli, B. (1972). Karyologic observations on the maturation of the summer and winter Eggs of *Daphnia* pulex and *Daphnia* middendorffiana. *Chromosoma*, *36*(2), 193–203. https://doi.org/10.1007/BF00285213

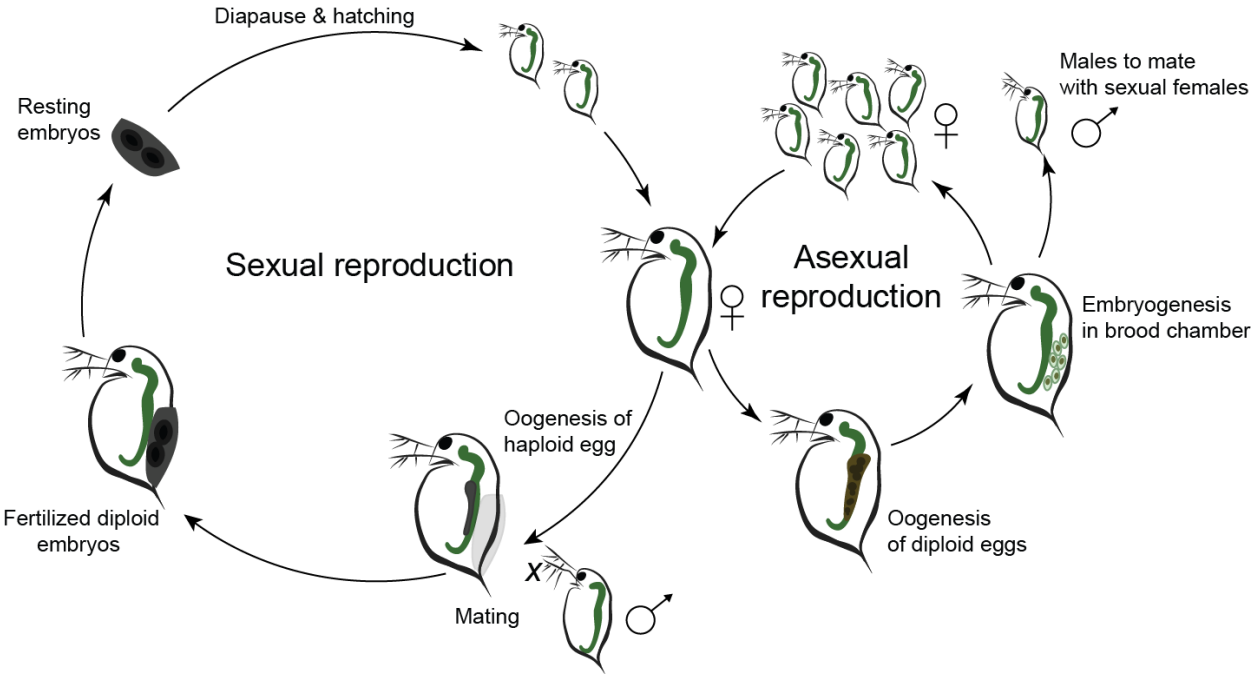


**Figure 1.** The parthenogenetic life cycle of *Daphnia* . Asexual reproduction produces clonal offspring.

Figure adapted from Xu *et al.* 2021, https://doi.org/10.1101/2021.08.26.457823 (preprint, bioRxiv)

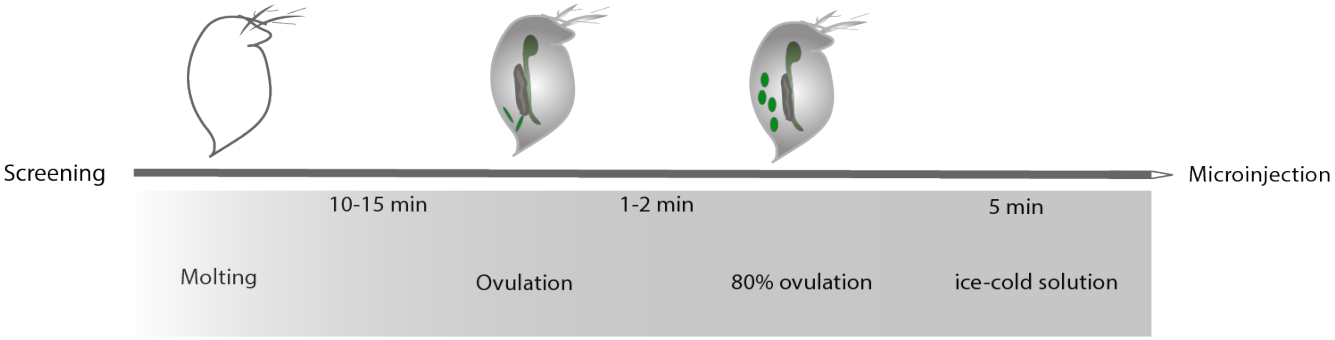


**Figure 2.** Tentative timeline of oogenesis in *Daphnia* along with injection workflow. Females with distinct and dark ovaries are screened and observed until they molt. Once molted the female will ovulate within 10-15 min, after ovulation start, there is a window of 1-2 min until they almost complete ovulating. Once 80% of ovulation is complete, the female is transferred to ice-cold solution for 5min. The animal is then ready for dissection and microinjection.

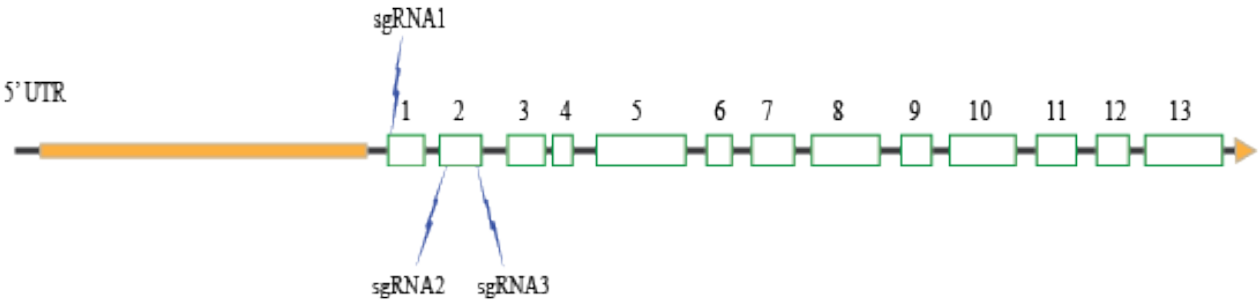


**Figure 3.** A schematic gene structure of *Scarlet.* The sgRNA1, sgRNA2, and sgRNA3 target exon 1 and 2 in the *scarlet* gene. Green boxes represent the exons. The black bar represents the intron.

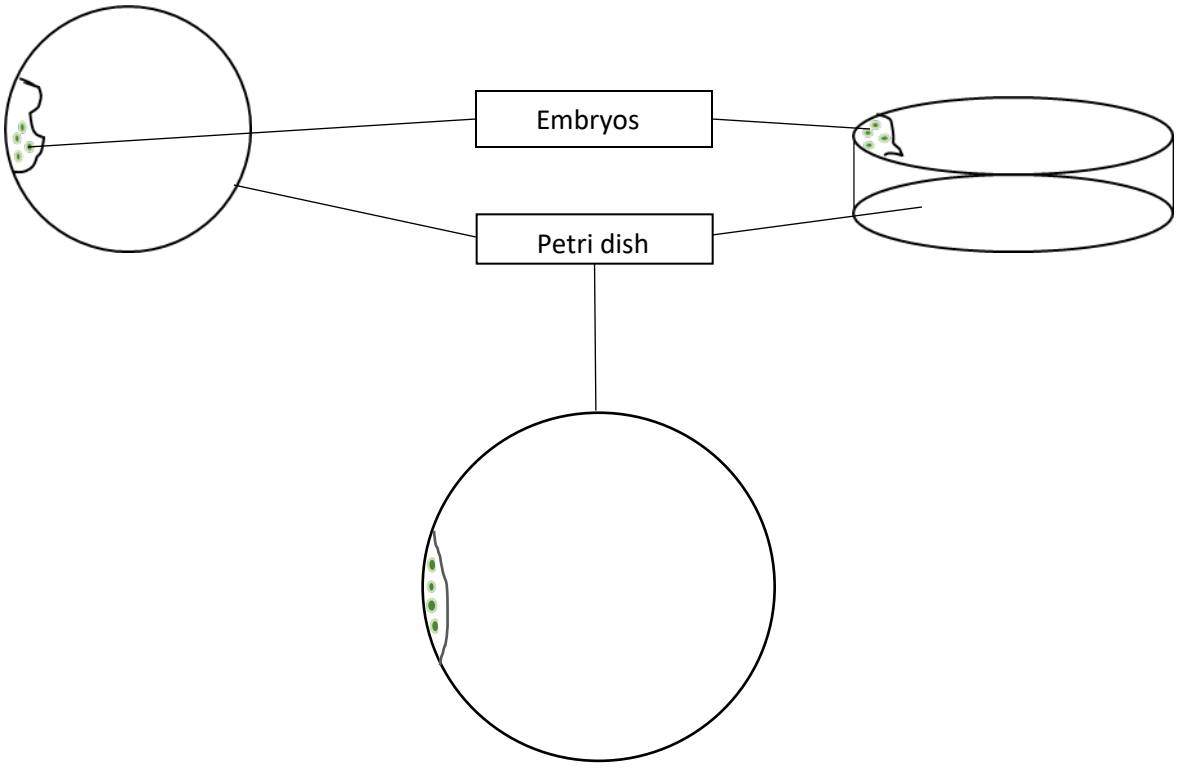


**Figure 4.** Top view (left) and side view (right) of organizing and positioning embryos on the petri dish after dissection **(**Top panel). Top view of organized and lined up embryos for microinjection (Bottom panel).

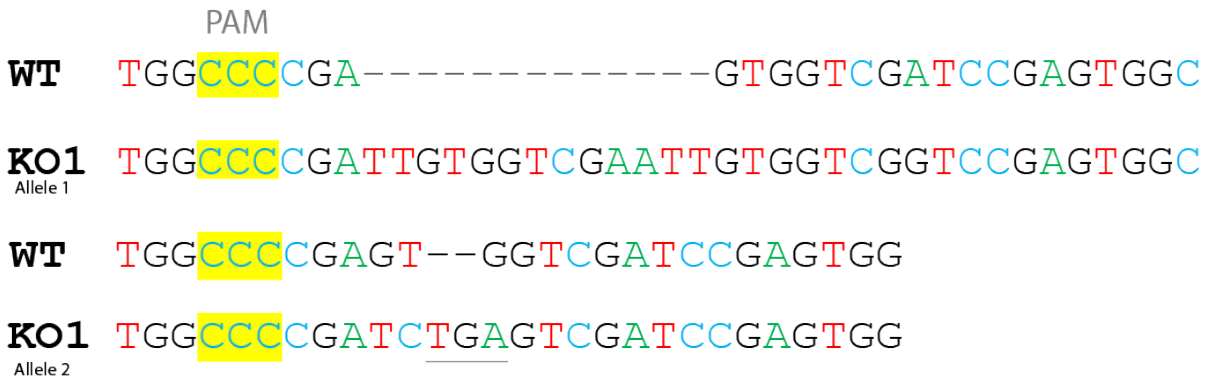


**Figure 5.** Alignment of mutant line KO1 with the wildtype EB1. Allele 1 has 13 bp insertion. Allele 2 has 3 bp insertion.

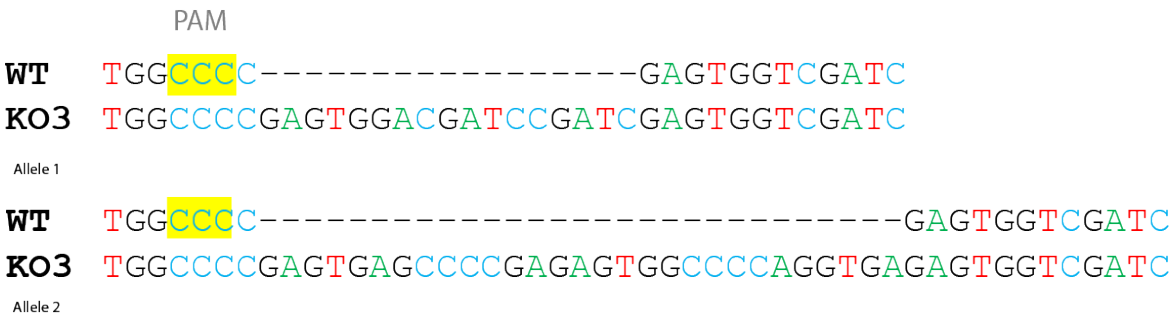


**Figure 6.** Alignment of mutant line KO3 with wildtype EB1. Allele 1 has 17 bp insertion. Allele 2 has 29 bp insertion.

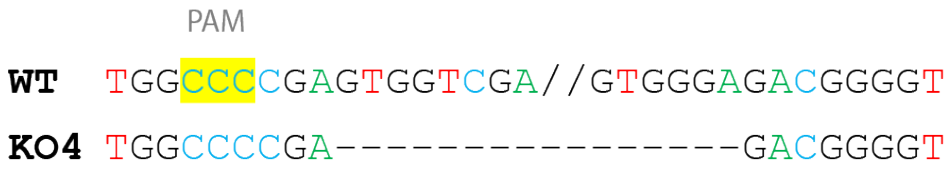


**Figure 7.** Alignment of mutant line KO4 with wildtype EB1. Biallelic deletion of 237 bp.

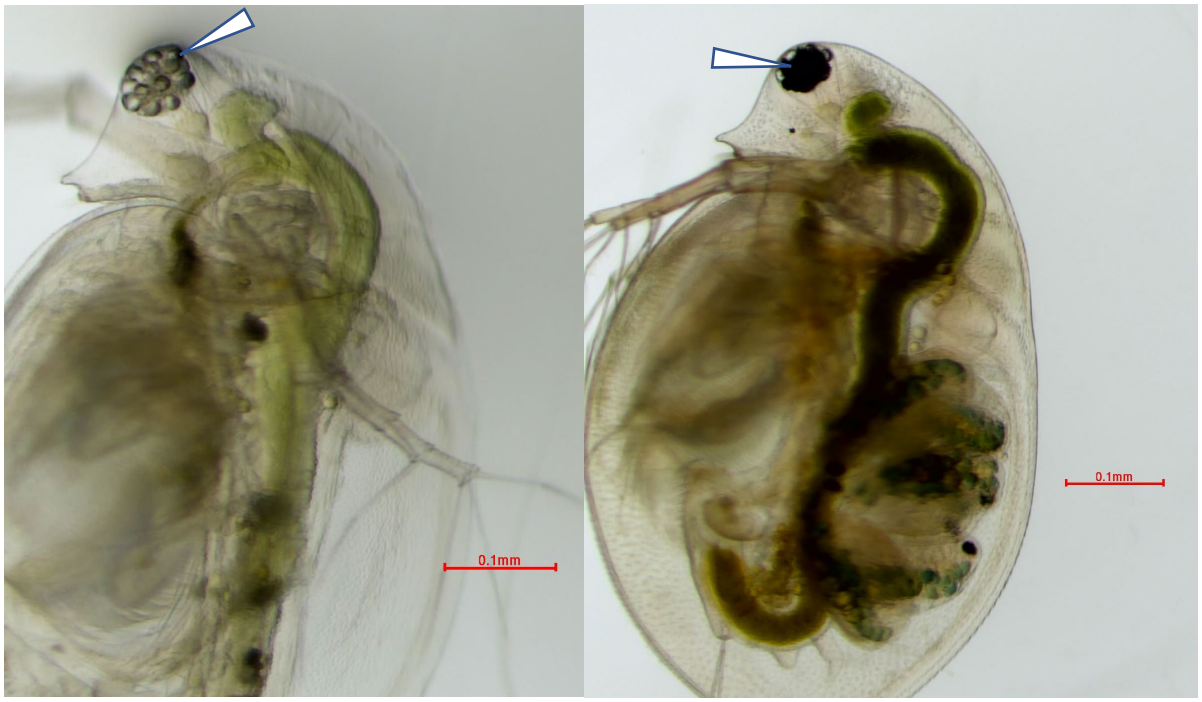


**Figure 8.** A white-eye mutant lacking a functioning *scarlet* gene (Left) labelled with a white triangle. A wildtype with normal black eye (Right).

**Figure 9.** A mosaic phenotype with a black and white eye (blue triangle), note that the brood the female carries has a normal black eye (white triangle), making the mutation somatic.

**Table 1.** Summary of CRISPR-Cas9 microinjection with *Scarlet* gene knockouts.

| Injected embryos | Hatching embryos | White-eye Phenotype | Mosaic phenotype |
|---|---|---|---|
| 80 | 28 | 1 (3.6%) | - |
| 176 | 45 | 7 (15.6%) | - |
| 147 | 67 | 1 (1.5%) | - |
| 63 | 24 | 1 (4.1%) | - |
| 107 | 35 | 2 (5.7%) | 1 (2.9%) |
| 245 | 100 | 1 (1.0%) | - |
| 140 | 80 | 1 (1.3%) | - |
| 200 | 117 | 4 (3.4%) | 3 (2.6%) |
| 235 | 130 | 2 (1.5%) | - |
| 204 | 92 | 1 (1.1%) | - |
| 1597 | 718 | 21 | 4 |

**Table 2.** The sgRNA sequence chosen on exon 1 and exon 2. sgRNA1 is present on exon 1, sgRNA2 and sgRNA3 are present on exon 2.

| sgRNA | Sequence |
|---|---|
| sgRNA1 | cgtctcccacgtccgacacg |
| sgRNA2 | ccattcggatcgaccgctcg |
| sgRNA3 | gtgacgacccccaaagccaa |

# Chapter 4.
# Indexing of re-engineered Multiple Annealing and Looping-based Amplification Cycle (MALBAC) for multiplexing.

Swatantra Neupane and Sen Xu*

Department of Biology, University of Texas at Arlington, Arlington, Texas, 76019, USA

*Corresponding author: Sen Xu. Email: sen.xu@uta.edu

Abstract

The advancement in next-generation sequencing techniques having high throughput and benchtop instruments has considerably reduced the time and cost of sequencing; however, in the field where many samples need to be sequenced, the cost and laboriousness of library preparation emerge as a bottleneck. Library preparation protocols require costly reagents and substantial hands-on time making it impractical to sequence even a hundred samples, thus limiting the study where genomic data on a population is required. Here, we present a strategical approach where 96 single sperm samples can be uniquely indexed during the whole genome amplification process, making it suitable for a pooled library preparation, reducing cost by at least 100 times. The barcoding strategy simultaneously tags and whole-genome amplify a single sperm cell, reducing the time required for separate tagmentation, as often found in various methods. This strategy exploits the multiple annealing and looping-based amplification cycle (MALBAC) process and provides a tool for low quantity DNA samples. We present a practical approach to introducing a unique barcode to a single cell genome as a proof of concept.

## Introduction

The emergence of new cost-efficient next-generation sequencing (NGS) platforms has opened new doors for advancing research, especially in single-cell genomics. However, the availability of abundant genomic DNA content is critical for many available sequencing platforms; the same is valid for the majority of the downstream molecular biology assays. Often due to the nature of sampling or the availability of samples, the number of tissue/cells is limited. Several whole-genome amplification (WGA) methodologies have been developed to mitigate this shortcoming in the past few decades (Gawad et al., 2016; Hughes et al., 2005).

Single-cell genomics is one such study where the quantity of DNA is low. Single-cell genomics advances our understanding of genetics by increasing the resolution of genome study. With the improvement in genome sequencing, single-cell genome sequencing is exploring investigation in cancer research (Wang et al., 2014), haplotype analysis (Amini et al., 2014), and genomic variability. The primary technical challenge in this single-cell research field is to amplify a whole-genome by recovering a high percentage of the genome with a minimal amplification bias (Blainey, 2013; Gawad et al., 2016; Lasken, 2013).

NGS includes sample generation, library preparation, sequencing, and bioinformatics. Even though next-generation sequencing has developed, reducing the costs per base (National Human Genome Research Institute, 2019), the library preparation cost still takes a considerable portion (Gawad et al., 2016; Rohland & Reich, 2012). Studies centered around single-cell genomics are usually sample hungry, requiring hundreds of samples to make valid inferences. Especially when working with gametes, the studies need hundreds of gametes (sperms/eggs) to be sequenced.

Generating high-quality single-cell sequencing data has four primary technical aspects: efficient physical isolation of individual cells; amplification of single cells to obtain sufficient material for downstream analyses; cost-effective analysis for hypothesis testing; and interpreting the data (Gawad et al., 2016). The first step, isolating individual cells, is a rigorous study sector. With the advanced development in Fluorescent Activated Cell Sorting (FACS), single-cell isolation has become accessible to various sampling types. Different single-cell isolation technologies, including laser capture (Bagnell, 2005), micropipetting (Zong et al., 2012), have recently been reviewed (Blainey, 2013; Gawad et al., 2016; Navin, 2014; Shapiro et al., 2013) in different contexts. The current study focuses on the second step, where labs can efficiently amplify the low-content DNA to obtain enough starting material for further downstream analyses. We propose incorporating the third step to cost-effectively query the genome by adding DNA barcodes to the single-cell DNA samples while performing WGA. Each unique barcode is associated with a single cell. It would allow us to multiplex multiple single barcoded DNA to create a sequencing library that can be NGS sequenced as a single sample, thus drastically lowering the cost involved in the library preparation.

Multiplexing involves pooling and sequencing a large number of libraries simultaneously during a single run (Amini et al., 2014; Di et al., 2020; Rohland & Reich, 2012), reviewed in (Adey, 2021). The ultimate purpose is to increase sample throughput, reducing time and effort. While multiplexing, unique "barcode" or “tag” sequences are added to each DNA fragment during next-generation sequencing (NGS) library preparation. These barcodes, or index adapters, as of now, can follow one of two indexing strategies; examples are oligo indexing, barcodes introduced with the help of Tn5 endonucleases to tag RNA, or even single cells (Amini et al., 2014; Di et al., 2020). Tn5 tagmentation method uses an active transposase to fragment target DNA before

incorporating universal adapters (Di et al., 2020), also see review (Adey, 2021). In a follow-up de-multiplexing step, the reads need to be separated using the attached barcode (sample marker) sequences, done by the sequencing facility, and the barcodes we used that can be identified with unique tags. The strategy we propose is to add in-house indices or barcodes (Table 1). One unique barcode is added to a single sperm—ninety-six of these unique barcodes tag a 96 well-plate consisting of a single sperm cell in each well. The tagging happens when a specially designed MALBAC primer containing a barcode is used to amplify a single sperm cell genome using the MALBAC procedure (described later). Then standard library preparation and NGS can be performed. The revolutionary aspect of this procedure is that at least 96 individual sperm samples can be library prepped as a single sample, significantly reducing the cost by at least 100-fold.

Results

MALBAC consists of five cycles of linear amplification, making a pre-amplification stage and a final standard PCR amplification (Figure 1). The pre-amplification is initiated with a pair of quasi-degenerate primers (NT and NG, see Figure 1), both having a common 27mer sequence (5'-GTG AGT GAT GGT TGA GGT AGT GTG GAG- 3') followed by eight variable nucleotides that can evenly hybridize to DNA templates and create an overlapped amplicons throughout the whole genome. Five cycles of pre-amplification are performed, where the *Bst* large fragment polymerase (New England Biolabs) having strand displacing activity is added to each reaction. A full-amplicon is formed at the end of each cycle, which loops at 58°C. The looping is due to the complementary sequence of 27mer at the one end (green bar in Figure**.** 1). This looping prevents further pre-amplification reducing the overrepresentation of some genomic regions in the final pool of amplicons. After the pre-amplification stage, a standard PCR amplification is performed with only a common 27mer to generate 1–2 μg DNA used for downstream applications. Our first approach

is to verify that the deviation from the original common 27mer while maintaining the primer criteria would amplify as effectively as the original primer. So, we changed the 8-bp region in the original 27mer and proceeded to MALBAC using ~1 picogram of genomic DNA as a template (Table 1). The modified 27mers generated comparable DNA content (μg) compared to the original one when observed in electrophoresis gel intensity and multiple microsatellite PCR. The indices (barcodes) are introduced in the first pre-amplification stage of MALBAC; the final stage (amplification with common 27mer), however, only amplifies the generated full amplicons exponentially. We first developed a list of 8-bp (base pair) barcode indices incorporated in the 27mer to perform indexed MALBAC amplification (1-tier indexing). Each unique barcode is embedded in the 27mer portion of one set of NT, NG, and 27mer primers, which labels a single cell's amplicons. It will allow the pooling of the amplicons of 96 single sperm cells in one sequencing library preparation, where the 2-tier indexing (commercial standard library preparation) gets introduced.

*Generation of indexed MALBAC primers*

Preserving the nature of MALBAC, we created indexed MALBAC primers that have few requirements. The requirements are: 1) strictly exclude C nucleotide in the barcode, and A, T, and G nucleotides must be all present to maximize the randomness. 2) Two neighboring A nucleotides in the barcode must be at least three bp apart so that the triple Ts of NT primer would not bind to the 27mer. 3) Each barcode varies from the remaining barcodes by at least two nucleotides, which would allow accurate de-multiplexing in case if sequencing errors occur in the barcoded region. 4) No five or more G nucleotides in a row were allowed in the primer, as our initial test reflected that five or more stretch of G nucleotides inhibited the MALBAC. We have developed a Python computer program (MALBAC Barcodes R' Us) to select the barcodes following these parameters,

generating approximately 300 valid unique 8-bp barcodes. We synthesized 96 different sets of indexed NT, NG, and 27mers, with barcodes located between sites 18 and 25 in the 27mer sequence (Table 1).

*Testing the synthesized MALBAC primers*

We performed MALBAC using the synthesized primers on ~1 picogram of DNA (similar to the amount of DNA in a *Daphnia* sperm). The amplification with all sets of primers was successful, yielding ~1μg of DNA amplicon. The amplification coverage of these indexed primers across the genome is equivalent to the original primers. Regular PCR for 12 randomly selected loci from each chromosome using these MALBAC products as templates showed that these loci are all present in the MALBAC amplicons (Figure 2).

To further ease the MALBAC process, we tested if a common stretch of 27mer present in all the modified NG and NTs can exponentially amplify the full-amplicons. We designed oligos containing only the common 5' stretch GTG AGT GAT GGT TGA GG that is 17 nucleotides long or referred to as 17mer henceforth. The strategy was to ease the laborious task of adding individual 27mer in all the 96 samples in a plate by adding a common 17mer that can be premixed in the PCR buffer. Using 17mer, we observed that the MALBAC amplified the samples. However, compared to the original common 27mer used in the amplification stage, the 17mer generated almost half DNA content than did 27mer, and the size distribution was also different (bigger size generated by the 17mer than 27mer) (Figure 3). The lesser DNA quantity generated by the 17mer would have been the slightest problem because the DNA content was still enough for library preparation. Nevertheless, the bigger DNA size and distribution would mean that the DNA needed to be sheared for library preparation.

In contrast, the DNA size distribution for the generated DNA from 27mer is around 600 bp, desirable for library preparation (Figure 3). However, the major problem is that shearing would very much destroy the purpose of introducing the barcode during the MALBAC amplification. So, we set aside the strategy of using a common 17mer during the amplification stage.

## Discussion

With improvements in technology, DNA sequencing has reduced the cost to the point where the reagent cost of sample preparation is the limiting factor. The quality and quantity of sequence data required per sample is often less than the commercial cost of library preparation. The available library preparation kits have limited throughput, drastically increased cost when scaling to hundreds or thousands of samples. However, several published studies have presented a way to drastically reduce the cost associated with library preparation (Rohland & Reich, 2012). One strategy is to pool samples before library preparation to save funds and time, but the samples need to be indexed (barcoded) first.

We based our strategies to help examine the evolutionary forces in the evolution of recombination rate. Different theories converge on predicting that the transition to a novel environment will lead to an increased recombination rate due to novel selection pressure (Butlin, 2005). However, domesticated animals under strong directional selection seem to have no recombination rate increase (Munoz-Fuentes et al., 2015). Due to the lack of population-level recombination rate data, this hypothesis remains untested mainly in natural systems. In species with well-understood ecology like *Daphnia*, the lack of population and individual level recombination rate data is a significant challenge for understanding how ecological shift and selection affect recombination rate evolution. Despite our efforts to assess fine-scale recombination rate data on two different chromosomal segments (Neupane & Xu, 2020) being attainable, it is still laborious to scale up.

An analysis to sequence 96 sperms for 50 individuals, making ~5,000 libraries. Some low-cost library construction methods like RIPTIDE (https://igenomx.com/product/riptide. Accessed Nov 2021) still cost about $10 per sample, summing up library construction to almost $50,000. We planned to develop a novel 2-tier combinatorial indexing single-cell sequencing approach, reducing the library costs by a hundred-fold. The current study focuses on the first-tier indexing, which is accomplished by barcoding the DNA amplicon of each cell using re-engineered MALBAC amplification (described in Results). Furthermore, the 2-tier indexing occurs through barcoded sequencing library construction using available kits (future scope). MALBAC costs ~$0.7 per sample and provides an even amplification coverage across the genome (Zong et al., 2012), hence, making it our choice.

Here we present that modifying eight nucleotides in the common 27mer can introduce a unique barcode indexed during MALBAC amplification itself; this poses a very efficient strategy to introduce barcode in hundreds of single sperm samples integrating it with the mandatory WGA step. First, when combined with indexing (introducing the first barcode during WGA), 96 single sperm cells can be grouped, pooled, and sequenced in a few runs. We tested the first step of the 2-tier indexing on ~1 pg of genomic DNA. Second, a potential concern is the differences in ligation or primer efficiency of different barcodes. We performed a screening of 96 unique barcoded primers that gave us insight into whether the primer's efficiency varies (Figure 3). It is noteworthy to mention that one of the parameters, having five or more G nucleotides in a series, would reduce the primer efficiency, later added to the primer design rules.

This approach would significantly increase throughput by streamlining the indexing and library preparation in 96-well plates, reducing the technician's time. The significant reduction of cost in multiplexing hundreds of samples to perform NGS would make it revolutionary because this

method can readily be adopted across various model organisms, owing to the sensitivity and versatility of MALBAC. Lesser financial investment in sequencing hundreds of sperm samples breaks the barrier to effectively producing genetic maps that would otherwise be financially impractical. The future work continuing to establish the whole streamlined workflow is to implement the $2^{nd}$ -level indexing. We will be performing standard barcoded Illumina library preparation eliminating the DNA fragmentation step on a pool of indexed MALBAC products of 96 single sperm cells from *Daphnia* species. These libraries are appropriate for whole-genome sequencing; each sperm will be sequenced at 5X genome-wide coverage (Xu et al., 2015). After sequencing, we will de-multiplex the Illumina barcodes and then the MALBAC barcodes into individual sperm using open-source de-multiplexing software. Our goal is to increase throughput and decrease reagent costs while building appropriate libraries for pooled sperm sequencing to generate genetic maps of multiple *Daphnia* species. However, our engineered method is still applicable and adoptable by academic laboratories to create hundreds of barcoded libraries at a hundred magnitude less cost than the commercial cost of library preparation.

Acknowledgment

We are grateful to Trung Huynh for writing a program, MALBAC Barcodes R' Us, to generate the lists of barcoded primers. Also, we would like to extend our gratitude to Thinh Pham for rigorously helping to ease the process of screening 96 primers. Similarly, many thanks to Hongjun Wang and Marelize Snyman for their constant help and moral support.

## Materials and Methods

### *Whole-Genome Amplification*

To obtain enough DNA from each sperm for genotyping, the lysed single sperm cell was used for MALBAC (multiple annealing and looping-based amplification) whole-genome amplification (Zong et al., 2012). MALBAC consists of a pre-amplification stage and a standard PCR amplification. The preamplification is initiated with random primers, each having a common 27-nucleotide sequence (5'-GTGAGTGATGGTTGAGGTAGTGTGGAG- 3') and eight variable nucleotides that can evenly hybridize to the templates. Despite the indexed primers, the MALBAC protocol was unchanged. The change introduced here is the indexed 27mer, and its respective NG and NT sequences were used as described in Table 1.

### *Standard PCR amplification*

The standard touchdown PCR was used to evaluate the presence of different microsatellite markers. The thermal cycling program for microsatellite amplification consisted of 3min at 95° C, ten cycles of 35 s at 95° C, 35 s at 56° C (the temperature increased by 1° C for each cycle) and 45 s at 72° C, 30 cycles of 35 s at 95° C, 35 s at 48° C, 45 s at 72° C and a final 10 min at 72° C.

References

Adey, A. C. (2021). *Tagmentation-based single-cell genomics*. 1693–1705. https://doi.org/10.1101/gr.275223.121.31

Amini, S., Pushkarev, D., Christiansen, L., Kostem, E., Royce, T., Turk, C., Pignatelli, N., Adey, A., Kitzman, J. O., Vijayan, K., Ronaghi, M., Shendure, J., Gunderson, K. L., & Steemers, F. J. (2014). Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing. *Nature Genetics*, *46*(12), 1343–1349. https://doi.org/10.1038/ng.3119

Bagnell, C. R. (2005). Laser capture microdissection. *Molecular Diagnostics: For the Clinical Laboratorian*, *274*(November), 219–224. https://doi.org/10.1385/1-59259-928-1:219

Blainey, P. C. (2013). The future is now: Single-cell genomics of bacteria and archaea. *FEMS Microbiology Reviews*, *37*(3), 407–427. https://doi.org/10.1111/1574-6976.12015

Butlin, R. K. (2005). Recombination and speciation. *Molecular Ecology*, *14*(9), 2621–2635. https://doi.org/10.1111/j.1365-294X.2005.02617.x

Di, L., Fu, Y., Sun, Y., Li, J., Liu, L., Yao, J., Wang, G., Wu, Y., Lao, K., Lee, R. W., Zheng, G., Xu, J., Oh, J., Wang, D., Sunney Xie, X., Huang, Y., & Wang, J. (2020). RNA sequencing by direct tagmentation of RNA/DNA hybrids. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(6), 2886–2893. https://doi.org/10.1073/pnas.1919800117

Gawad, C., Koh, W., & Quake, S. R. (2016). Single-cell genome sequencing: Current state of the science. *Nature Reviews Genetics*, *17*(3), 175–188. https://doi.org/10.1038/nrg.2015.16

Hughes, S., Arneson, N., Done, S., & Squire, J. (2005). The use of whole genome amplification in the study of human disease. *Progress in Biophysics and Molecular Biology*, *88*(1), 173–189. https://doi.org/10.1016/j.pbiomolbio.2004.01.007

Lasken, R. S. (2013). Single-cell sequencing in its prime. *Nature Materials*, *12*(4), 367–376. https://doi.org/10.1038/nmat3550

Munoz-Fuentes, V., Marcet-Ortega, M., Alkorta-Aranburu, G., Forsberg, C. L., Morrell, J. M., Manzano-Piedras, E., Soderberg, A., Daniel, K., Villalba, A., Toth, A., Di Rienzo, A., Roig, I., & Vila, C. (2015). Strong artificial selection in domestic mammals did not result in an increased recombination rate. *Molecular Biology and Evolution*, *32*(2), 510–523. https://doi.org/10.1093/molbev/msu322

Navin, N. E. (2014). Cancer genomics: one cell at a time. *Genome Biology*, *15*(8), 452. https://doi.org/10.1186/s13059-014-0452-9

Neupane, S., & Xu, S. (2020). Adaptive Divergence of Meiotic Recombination Rate in Ecological Speciation. *Genome Biology and Evolution*, *12*(10), 1869–1881. https://doi.org/10.1093/gbe/evaa182

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, *22*(5), 939–946. https://doi.org/10.1101/gr.128124.111

Shapiro, E., Biezuner, T., & Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nature Reviews Genetics*, *14*(9), 618–630. https://doi.org/10.1038/nrg3542

Wang, Y., Waters, J., Leung, M. L., Unruh, A., Roh, W., Shi, X., Chen, K., Scheet, P., Vattathil, S., Liang, H., Multani, A., Zhang, H., Zhao, R., Michor, F., Meric-Bernstam, F., & Navin, N. E. (2014). Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature*, *512*(7513), 155–160. https://doi.org/10.1038/nature13600

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*(1), 31–38. https://doi.org/10.1534/genetics.115.179028

Zong, C., Lu, S., Chapman, A. R., & Xie, X. (2012). Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science*, *338*(6114), 1622–1626. https://doi.org/10.1126/science.1227764

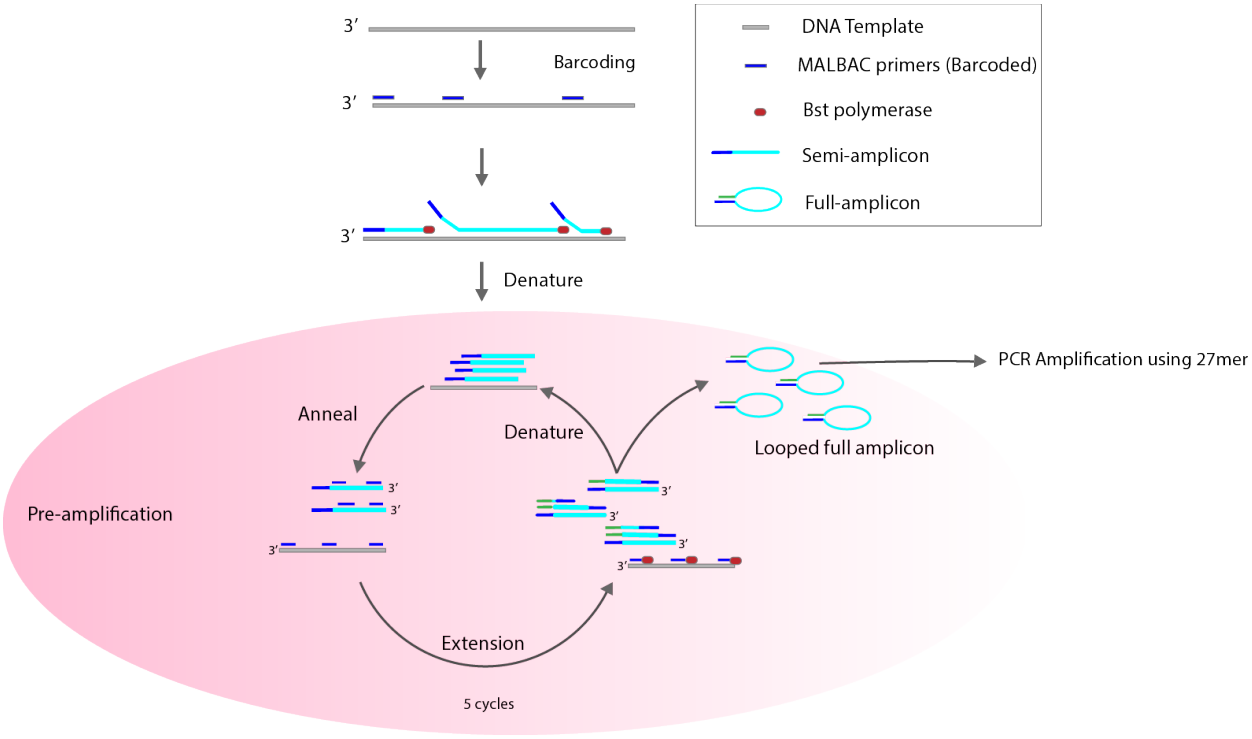


**Figure 1**: MALBAC mechanism and workflow. Barcoded primers (27mer) are shown in a dark blue bar.

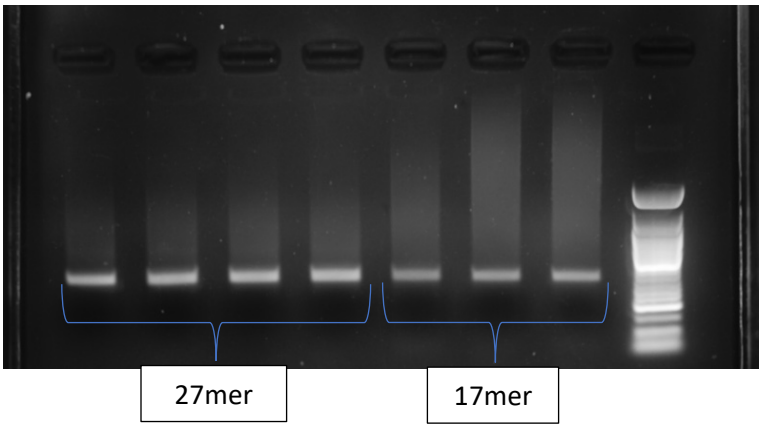


**Figure 2**. Microsatellite marker PCR results on MALBAC were performed with 27mer vs. 17mer.

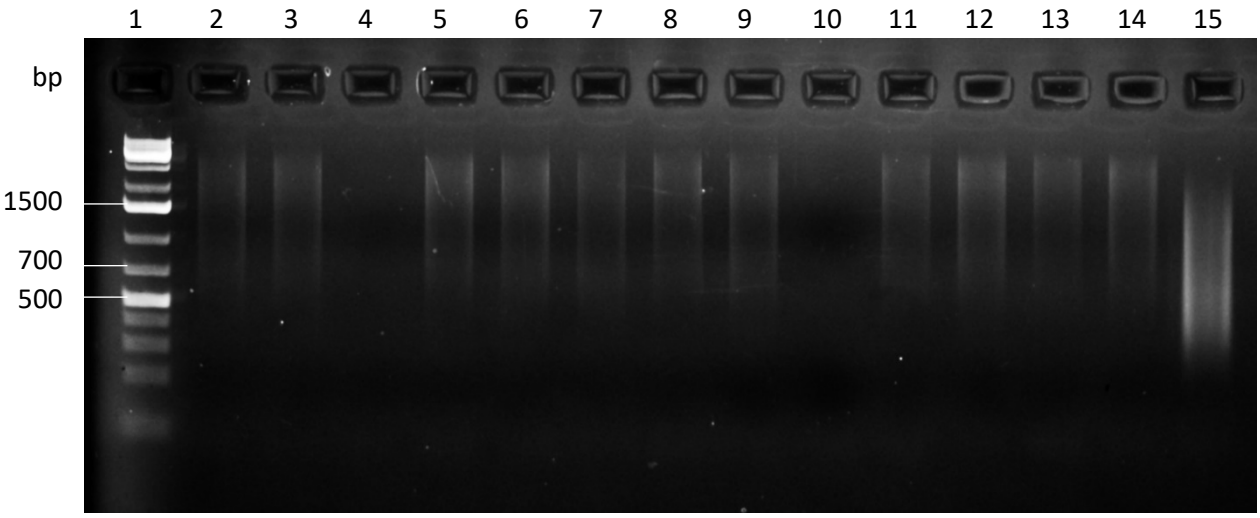


**Figure 3**. Agarose Gel electrophoresis on modified MALBAC primers with 17mer during final amplification. Lanes 4 and 10 had primers with 5 G nucleotides in a series (not working). Lanes 14 and 15 show DNA distribution when using 17mer vs. 27mer during the final amplification stage.

**Table 1**. The barcodes in each MALBAC primer set, capital letters are the barcode stretch.

| | |
|---|---|
| NG | 5’- g t g a g t g a t g g t t g a g g **T T G T A G T T** a g -3’ |
| NT | 5’- g t g a g t g a t g g t t g a g g **T T G T A G T T** a g -3’ |
| 27mer | 5’- g t g a g t g a t g g t t g a g g **T T G T A G T T** a g -3’ |