GENOMIC ANALYSIS OF CHRONIC HEAT STRESS RESISTANCE IN THE NEMATODE *CAENORHABDITIS REMANEI*

by

SALLY CLARIDGE

A THESIS

Presented to the Department of Biology
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

June 2017

An Abstract of the Thesis of

Sally Claridge for the degree of Bachelor of Science
in the Department of Biology to be taken June 2017

Title: Genomic analysis of chronic heat stress resistance in the nematode *Caenorhabditis remanei*

Approved: _____

Dr. Patrick C. Phillips

Selection can drive sub-populations to diverge from each other, but gene flow can homogenize them. Migration-selection dynamics are one of the fundamental aspects of speciation and population divergence, but they have not been rigorously investigated in an experimental context. We aimed to elucidate how gene flow affects the rate of adaptation to a novel environment via the experimental evolution of the nematode *Caenorhabditis remanei* in a chronic heat stress environment. Five replicate populations of *C. remanei* were evolved to novel (31℃) and ancestral (20℃) environments and treated with 0- and 5-percent migration rates. Female fecundity information was collected to estimate the extent of adaptation in the heat-stress-evolved population. The migration treatment stunted the rate of adaptation to the novel environment, though population divergence still occurred. We anticipate there will be genomic changes in the descendant lines that lead to adaptation. Whole genome sequencing data from both the ancestral and descendant populations will be compared on a locus-by-locus basis to identify these changes. Migration is expected to reduce signatures of differentiation from weakly selected loci and to enhance signatures associated with loci of large effect.

Acknowledgements

To my late grandfather, Westbrook "Chee" Claridge, who pushed me to always know more tomorrow than I do today.

I want to sincerely thank and acknowledge Dr. Patrick Phillips, in whose lab I have worked for the past two years. His unyielding faith in me has gotten me over many a hurdle. I am also forever grateful to my mentor Christine O'Connor for her endless support, guidance, patience, and trust. She has taught me more about the scientific process than any class could. I would like to thank my other thesis committee members, Dr. Samantha Hopkins and Dr. Cristin Hulslander, for serving as representatives of the Robert D. Clark Honors College and the Biology Undergraduate Affairs Committee, respectively. A special thank you goes out to Dr. Tyler Hether who helped me develop and implement our genomic analysis pipeline. I would like to thank Paula Berry, Ali Bedbury, Alley Engelbrecht, and Audrey Hernan as past and current members of the stress selection team who helped Christine and me with the numerous transfers, hatch-offs, freezes, and assays encompassed by this project. A big thank you to all other members of the Phillips Lab Group for embracing me as a brand-new researcher and creating such a wonderful work environment. And finally, to my incredible family and supportive friends who have selflessly commiserated and joyously celebrated with me throughout this thesis process, thank you for listening to me talk about my worms for the last two years.

Table of Contents

List of Figures

List of Tables

Introduction

Life on Earth is thought to have begun 3.7 billion years ago with a single-celled prokaryote, and from that single organism a vast variety of modern species arose. This diversity is astounding, with an estimated 8.74 million different eukaryotic species inhabiting the Earth (Mora et al. 2011). In the tree of life pictured in Figure 1, each branch represents a new lineage that diverged from the ancestor and became a new species. However, there are many factors that affect the occurrence and rate of speciation, e.g. gene flow, recombination, and natural selection (Nosil and Feder 2012). Theories based upon these factors have led to predictions about how speciation occurs, and these expected patterns have been observed in nature. However, these studies have usually focused on reproductive isolation and genes important for speciation (Orr 2005; Nosil and Schluter 2011; Wu 2011). How these genes are structured in the genome and how whole genomes evolve during the speciation process have yet to be examined in an experimental, hypothesis-testing framework. With the advent of next-generation sequencing, researchers can now feasibly obtain and analyze whole genomes from many individuals. This technological advancement has opened a field of genome-wide approaches for comparing inter-population differentiation to elucidate genomic effects on patterns of species divergence.

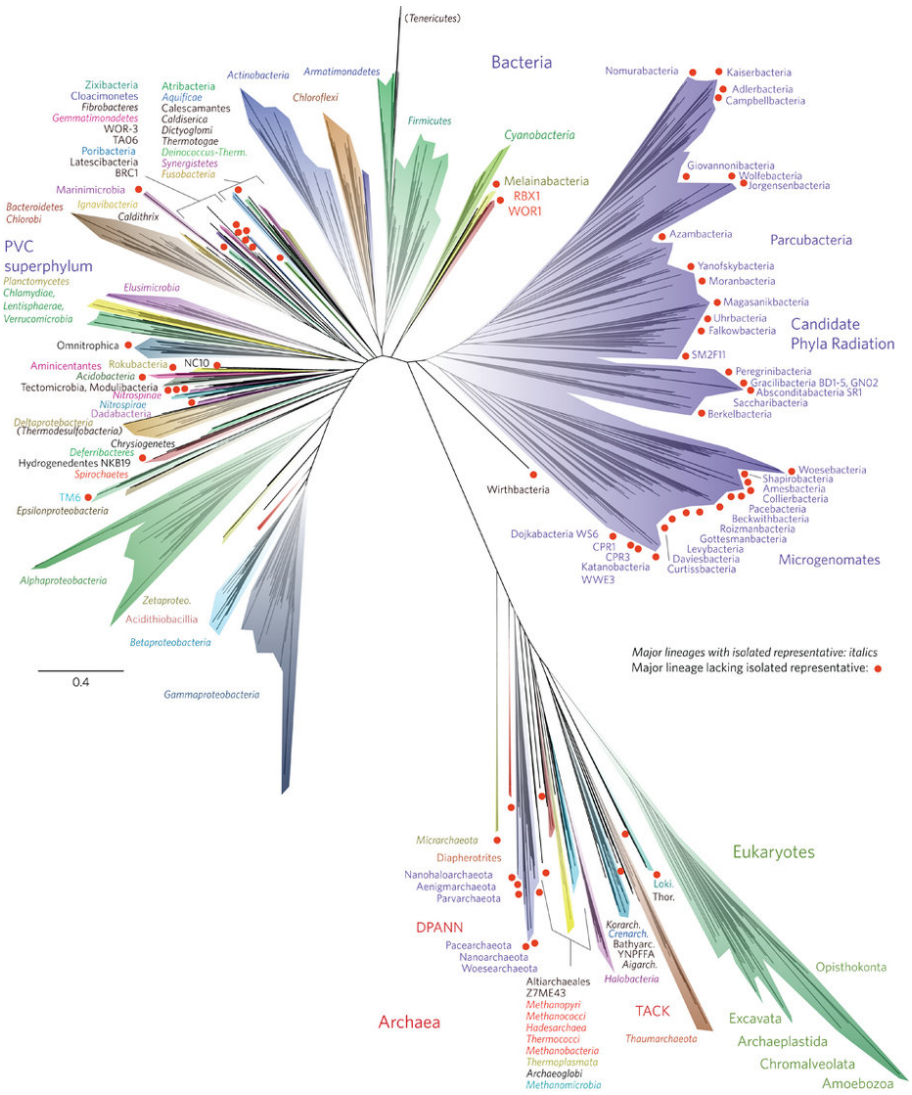

Figure 1. Most recent version of the tree of life.

(Image from Hug et al. 2016).

Migration-Selection Dynamics

Migration and selection are important and opposing forces in evolution, and interactions between the two affect how populations adapt and evolve to novel environments. Variation in the environment can yield contrasting selective pressures on populations that can lead to genetic and phenotypic divergence between subsets of the

population. Sometimes barriers completely isolate these sub-populations and prevent reproduction, i.e. allopatric speciation or vicariance. These sub-populations will independently evolve along trajectories according to the specific selective pressures acting upon them. In allopatric divergence, genomic divergence is inevitable given enough time (Nosil and Feder 2012). Alternatively, the environmental variation could yield selection that produces diverging sub-populations that remain connected by some level of gene flow, e.g. sympatric (Bolnick and Fitzpatrick 2007) and parapatric (Gavrilets et al. 2000) speciation. Gene flow introduces non-adaptive alleles to the diverging sub-populations, and recombination alters genic associations (Nosil and Feder 2012).

Traditional theories concerning speciation have posited that divergence with gene flow was an unlikely scenario (Jordan 1905; Mayr 1963; Felsenstein 1981). When individuals interbreed, their progeny harbor combinations of the parents' genetic complements. This genetic exchange would be expected to hinder any divergence between sub-populations. If the strength of migration is greater than the strength of selection, then this model predicts homogenization of the sub-populations. Counter to this expectation, multiple incidences of genetic and phenotypic divergence with gene flow have been documented: threespine stickleback (*Gasterosteus aculeatus*) across the lake-stream ecotone (Berner et al. 2009), lizards in the White Sands National Monument in light- and dark-substrate environments (Rosenblum 2006), and *Drosophila melanogaster* along the slopes of Evolution Canyon in Israel (Michalak et al. 2001). These incidences indicate that divergence can occur in the face of gene flow, and this divergence is the expected result if the strength of selection is greater than that

of migration. Recent theories and models have been proposed to describe this process and to predict the genomic outcomes of "speciation with gene flow" (Figure 2).

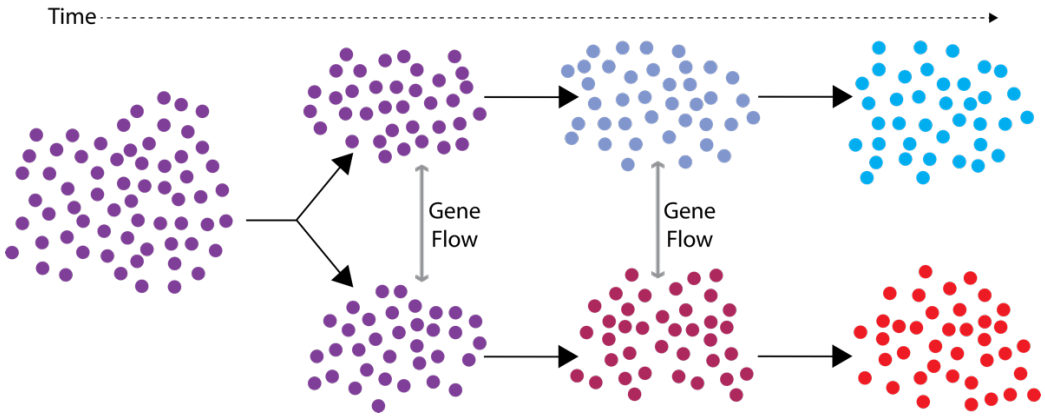

Figure 2. Speciation with gene flow schematic.

Dark purple dots represent a homogenous, freely interbreeding population. The forked arrow indicates a divisive event, e.g. environmental change, that separates the main population into two distinct sub-populations that remain connected by some level of gene flow. Over evolutionary time, these two sub-populations continue to diverge and eventually speciate.

"Divergence hitchhiking" is an important theory to consider. It posits that a physical connection between a locus under selection and an adjacent locus will provide a mechanism through which the adjacent locus can be maintained in a divergent population despite no fitness benefit to the organism (Smith and Haigh 1974; Via 2012; Nosil and Feder 2012). These grouped associations can be single nucleotides or even whole chromosomes that exhibit greater differentiation than expected under neutral evolution (Nosil et al. 2009). Extension of physical linkage and hitchhiking can cause these associations to grow in size and in level of divergence between populations, yielding what have been called "genomic islands of divergence" that surround the original locus under selection (Nosil and Feder 2012). Feder et al. (2012) and Nosil and Feder (2012) have even suggested that if divergent selection acts upon enough loci, the

effective gene flow can decrease to a level where the whole genome “hitchhikes” and creates “genomic continents of divergence.”

Though divergence hitchhiking has been observed in natural populations (Turner et al. 2009; Strasburg et al. 2012), these studies are limited to a singular moment in time and often must estimate the history of or make assumptions about the populations in question. Thus, they provide evidence for the expected genomic patterns but say little of the genomic changes that accompanied the process of species divergence. To address this informational gap in study design, we utilized an experimental evolution approach in which we followed the evolutionary progress of populations of known ancestry through multiple generations of adaptation to a novel, stressful environment. Multiple biological replicates of experimentally evolved populations of *C. remanei* are compared both phenotypically and at the genome level to potentially isolate loci conferring chronic heat stress resistance. This evolve-and-resequence paradigm (reviewed by Long et al. 2015) has been applied successfully in multiple systems, e.g. to elucidate patterns of whole-genome evolution in *Escherichia coli* over thousands of generations (Barrick et al. 2009), to identify single nucleotide polymorphisms (SNPs) in populations of *Saccharomyces cerevisiae* experimentally adapted to sulfate limitation (Araya et al. 2010), and to survey loci affecting body size variation in *D. melanogaster* (Turner et al. 2011). Using this experimental approach allows us to compare the phenotypic and genomic statuses of our populations at the end of forty generations and at intermediate generational time points.

Project Purpose and Hypotheses

The goal of this research project was to dissect the genomic architecture of chronic heat stress resistance phenotype in *Caenorhabditis remanei*, a nematode worm (also called a roundworm). This phenotype is a complex trait that involves expression and interaction of numerous genes and pathways to elicit an adequate response to the environment. These complicated and multivariate interactions obfuscate how these genetic networks function. Defining and characterizing the genetic architecture that underlies a complex phenotype of interest is vital to understanding some of the larger evolutionary questions that have been and continue to be investigated, such as speciation, disease vectors, adaptation, and variation between populations (Hansen 2006). Obtaining a better understanding of how the genetics of a complex trait evolve under environmental stress could help elucidate how these gene regulation networks function as a whole and how their interactions over time affect phenotypic expression. Characterization of the genetics underlying the chronic heat stress resistance phenotype may provide some answers to the overarching evolutionary biology question of, "What is the genetic basis of adaptation?"

Studying stress response as a model complex trait is advantageous because many of the significant pathways of large effect have been well-characterized and understood in their basic functions (Figure 3). Heat stress resistance is a heritable trait that is affected by selection in *C. remanei* (Reynolds and Phillips 2013). Additionally, it is highly tractable compared to other stressors. For instance, osmotic stress requires the addition of salts to the growth media, and oxidative stress involves the use of paraquat, a compound that is exceedingly toxic to humans. Creating a chronic heat stress

environment only requires a constantly-running incubator set to a specific temperature, which is more manageable than other stress conditions that could have been studied in the laboratory. Pilot female fecundity assays indicated that at 31˚C, *C. remanei* females exhibited greatly reduced fecundity, but the heat stress was not enough to induce complete sterility. Thus, 31˚C was chosen as the novel environment.

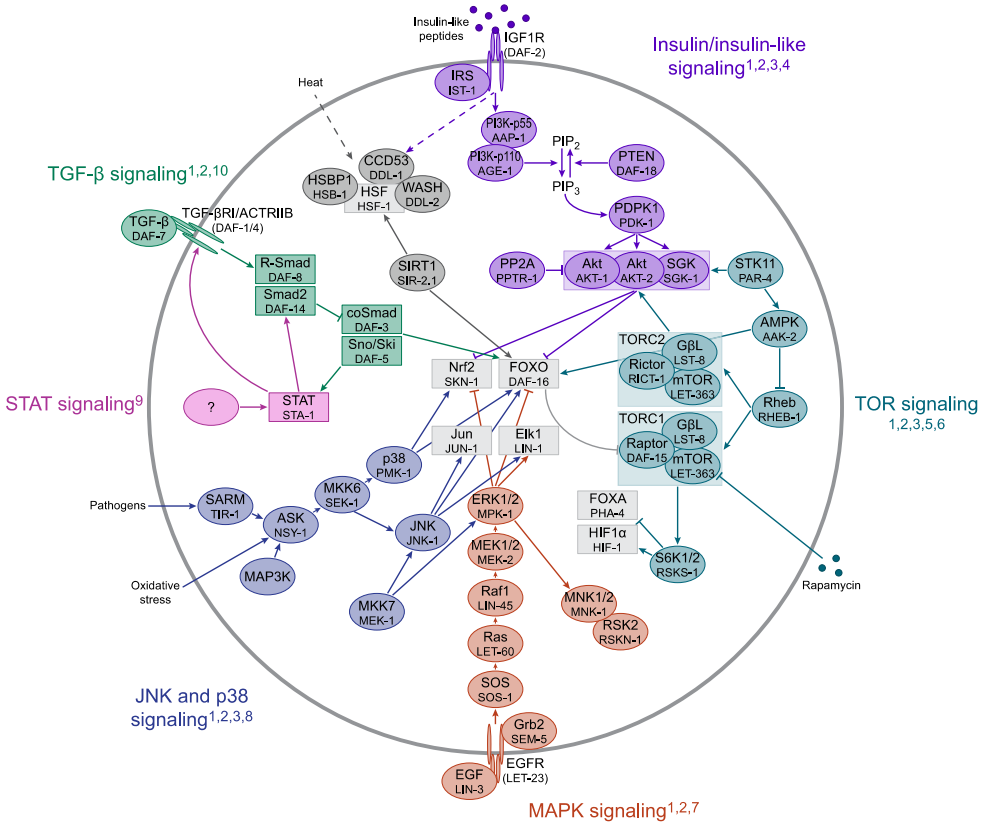

Figure 3. Various interacting stress response pathways in *C. elegans.*

Ovals indicate proteins and rectangles indicate transcription factors. Arrows indicate activating interactions, while bars indicate inhibitory interactions. (Image from Sikkink et al. 2015).

Populations of *C. remanei* were experimentally evolved in a heat-stress environment of 31˚C and a 20˚C, control laboratory environment with two levels of migration, 0 (no-migration) and 5 percent, for forty generations. The no-migration

treatment serves as a model of allopatric divergence, whereas the 5-percent treatment models parapatric, and potentially sympatric, divergence. Experimental evolution is a valuable technique because it provides large amounts of information concerning genetic divergence between populations under selective pressure while controlling for many variables that usually must be estimated in natural populations, such as population size and history. We collected female fecundity data for the evolved and ancestral populations to assess the effects of selection on chronic heat stress resistance and the interaction between migration and selection. We hypothesized that replicate populations evolved in the 31˚C environment would adapt to the novel, chronic heat-stress and would exhibit increased female fecundity in it after multiple generations.

Migration between two sub-populations is predicted to impede adaptation to a novel environment via the influx of non-adapted alleles into the gene pool. Selection will act more strongly on loci that contain alleles that are highly advantageous in the novel environment and more weakly on loci that are only mildly advantageous. This differential action of selection on various loci will affect what kinds of signals of divergence we observe. When non-adapted individuals migrate into an adapted population generation after generation, we expect loci with a small effect on selected traits to not diverge while loci with a larger effect on selected traits will exhibit divergence (Griswold 2006; Yeaman and Whitlock 2011). In other words, in populations linked by gene flow, loci that are neutral or weakly selected for will be homogenized and only loci strongly selected for will exhibit a signal of differentiation (Via 2001; Savolainen et al. 2006; Feder et al. 2012). We hypothesize that migration will constrain divergence between populations that is due to drift, and that will make

divergence between sub-populations that is due to selection more readily visible when analyzing genomic data.

Caveat

Due to time constraints on this project, we were unable to complete whole genome sequencing for the experimentally-evolved populations. Thus, to demonstrate the analytical framework that will be used to process and visualize these data at a later date, we developed a bioinformatic pipeline and applied it to an exemplar set of data from the ancestral population of *C. remanei* used for the experimental evolution (strain PX443). The methodology and results of this analysis are reported in the following chapters. The genomic analysis provided information on the genetic variation present in the ancestral population, which potentially affected the outcomes of our experimental evolution project.

For the next stage of analysis for this experimental evolution project, whole genome sequencing libraries will be created for the ancestral and descendent populations under all three migration treatments. These libraries will be processed, aligned to a reference *C. remanei* genome, and compared on a locus-by-locus basis to identify regions under selection. We predict that genetic drift and selection will lead to divergence in the absence of migration between the 20 ˚C and 31 ˚C experimental populations. Migration between the 20 ˚C and 31 ˚C populations is predicted to limit divergence to regions of the genome under selection.

Methods

Model Organism: *Caenorhabditis remanei*

Figure 4. *C. remanei* nematodes.

Female (left) and male (right). (Images courtesy of Kristin L. Sikkink).

Caenorhabditis remanei is a nematode species typically found in soil-based habitats, e.g. compost or woodlands (Kiontke and Sudhaus 2006). *C. remanei* is an ideal model organism for use in evolutionary genetics studies for a few key reasons. Firstly, unlike many of its hermaphroditic relatives, e.g. *C. elegans*, *C. remanei* is a gonochoristic species, i.e. a given individual is male (XO) or female (XX) (Figure 4). This gonochorism makes *C. remanei* an obligate outcrosser, meaning that it only reproduces sexually, which creates and maintains increased genetic diversity within a population compared to *C. elegans* and other hermaphroditic species (Graustein et al. 2002). Additionally, the *C. remanei* genome is available to the public and has been fully sequenced and annotated (Coghlan et al. 2008). This facilitates the genomic and genetic analyses that are vital for studying evolutionary trends. This nematode is approximately 1 mm in length, which allows for tractable upkeep of large experimental populations.

These populations are easily stored and maintained because 2,000 worms can live on a single 100-mm Petri plate. *C. remanei* also has a short generation time of 4 days, making long-term experimental evolution projects temporally manageable. Lastly, populations of juvenile worms can be frozen at -80˚C for an extended period of time and thawed when required, which allows for phenotypic and genomic assays to be completed on both the ancestral and evolved populations when time permits.

Worm Maintenance Protocols

The ancestral strain of *C. remanei* used for this project (PX443) was derived from a wild isolate and had been propagated in the laboratory environment for 75 generations. Each biological replicate was maintained on ten 100-mm Petri plates, five in the 20˚C incubator and five in 31˚C. We used agar with Nematode Growth Media (NGM)-lite that was seeded with *E. coli* strain OP50 as a food source (Stiernagle 2006). All Petri plates used throughout the project used this media-food combination.

Transfer Protocol

Every 4 days, worms were transferred to new 100-mm Petri plates to refresh their food supply and to prevent overcrowding and contamination. This constituted one generation. Worms were transferred via S basal, a standard liquid medium for nematode culture (Stiernagle 2006). Worms were washed from the plates with S basal and filtered through a 20-µm Nitex nylon screen. This filter size only allows L1-stage worms through, which prevents overlapping generations from being transferred to the new plates. For the no migration treatment, the L1 concentration in the filtrate was

calculated, and the appropriate volume of filtrate corresponding to 2,000 L1s was added to each fresh plate. For the 5-percent migration treatment, 1,900 L1s from the resident population and 100 L1s from the non-resident population were added to each fresh plate and incubated in the resident environment.

Hatch-off Protocol

Hatch-offs were used to age-synchronize worms by isolating eggs from a population. We used an adjusted version of the standard hatch-off protocol involving a 5-percent solution of sodium hypochlorite (bleach) and sodium hydroxide (Stiernagle 2006). Worms were washed from the plates with S basal into 15-mL conical centrifuge tubes, and the volume of worm suspension was reduced to 5 mL. We added 300 µL 4M sodium hydroxide and 600 µL 5-percent bleach and allowed the tube to stand for 3 minutes. Bleach was washed from the remaining eggs with multiple S basal rinses, and tubes were put on rotators in the appropriate incubators for 24 hours, yielding a population of age-synchronized L1-stage worms.

Freezing Protocol

We froze our experimentally evolved lines every 10 generations. To have enough worms for a large-scale freeze, we doubled our populations during the penultimate and final transfers, yielding 20 100-mm Petri plates of worms for freezing at the generation 10, 20, 30, and 40 time points. The freeze process started with a large-scale hatch-off of the 20 plates. Hatch-off concentrations were corrected to approximately 20 L1s/ µL and chilled on ice for 15 minutes. Two cryo-tube vials were prepped for liquid nitrogen storage and as many 1.5-mL cryo-tube vials as possible

were prepped for -80˚C storage. Each tube was prepped with a well-mixed combination of 500 µL agar-based glycerol freezing solution and 500 µL concentration-corrected hatch-off (approximately 10,000 L1s). Cryo-tubes were stored in either the -80˚C freezer or the liquid nitrogen tank. Test thaws (protocol outlined below) were conducted 48 hours after the initial freeze to ensure worms would be viable for later use.

Thawing Protocol

To thaw a vial of worms, we brought the vial and a labeled, 100-mm Petri plate to room temperature and transferred the vial contents onto the plate. The plate was stored in the appropriate incubator for two days. We then transferred 2,000 worms (without filtering) to a new plate to clean them of freezing solution. Three days later, we transferred the worms to a new plate following the Transfer Protocol. There was one additional transfer 4 days later. This 9-day timeline ensured adequate post-freeze recovery for the worms.

Experimental Evolution

We experimentally evolved populations of PX443 worms, which allowed for selection upon this strain's standing genetic variation. Sub-populations of *C. remanei* derived from the ancestral population were evolved in both a control laboratory (20˚C, "ancestral") environment and a chronic heat stress (31˚C, "novel") environment for forty generations. We tested the effect of migration between sub-populations in both environments with two migration rates: 0 (no migration) and 5 percent. For each environment-migration condition, five replicate sub-populations of ~10,000 individuals

were maintained, yielding a total of ~20,000 worms per biological replicate (Figure 5). Every ten generations, individuals were frozen at -80˚C for preservation, resulting in a total of four freezes for each biological replicate of both migration treatments.

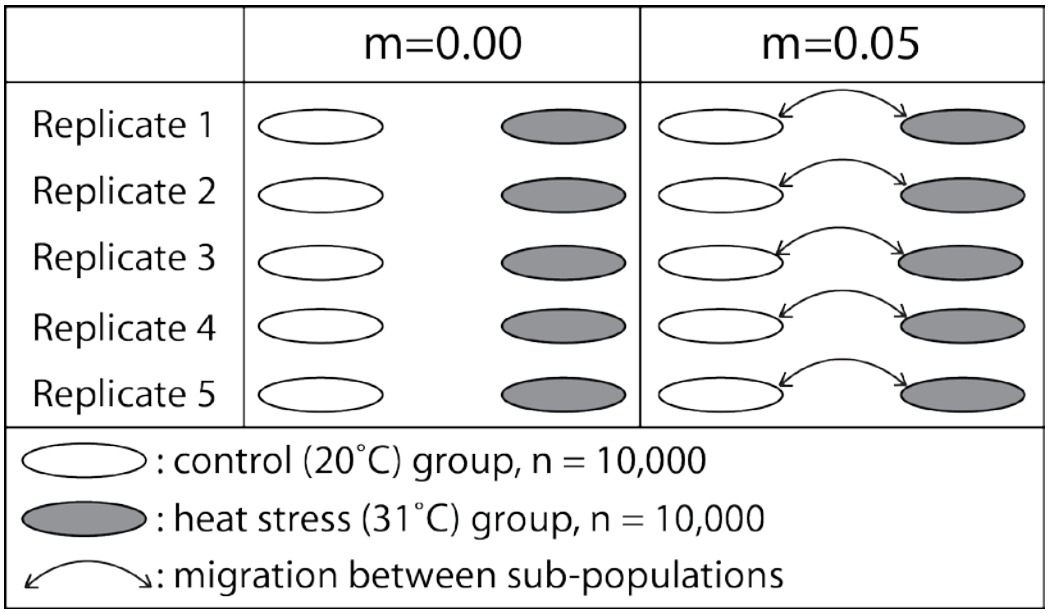

Figure 5. Tabulated experimental design.

Abbreviations: m is migration, n is number of individuals.

Female Fecundity Assays

Female fecundity assays are conducted to measure the number of viable offspring produced by a single individual over five days. Assays were conducted for our various populations in both the novel and ancestral environments, and each assay started with 25-30 replicates. Prior to conducting the assay, the population to be assayed was hatched-off to age synchronize the worms. To set up an assay, one virgin female and two male L4-stage worms from the same population were isolated on a 35-mm Petri plate and stored in the appropriate incubator. Every 24 (± 4) hours, all three worms were

transferred to a fresh 35-mm Petri plate to refresh their food supply. Successfully hatched offspring were counted approximately 48 hours after egg-lay. Female fecundity is measured for 5 days or until both males died or the female died (Figure 6). Assays were discontinued if more than one gravid female was found on a plate during a transfer. For each replicate, we calculated the mean five-day female fecundity and standard error (SE) using the summarySE() function from the *R Graphics Cookbook* (Chang 2012). The statistical significance of differences in average female fecundity between treatments was determined using a linear mixed model from the lme4 R package (Bates et al. 2015), which accounted for the intra- and inter-replicate variance in the calculation, and p-values were calculated using the Anova() function in the car R package (Fox and Weisberg 2011).

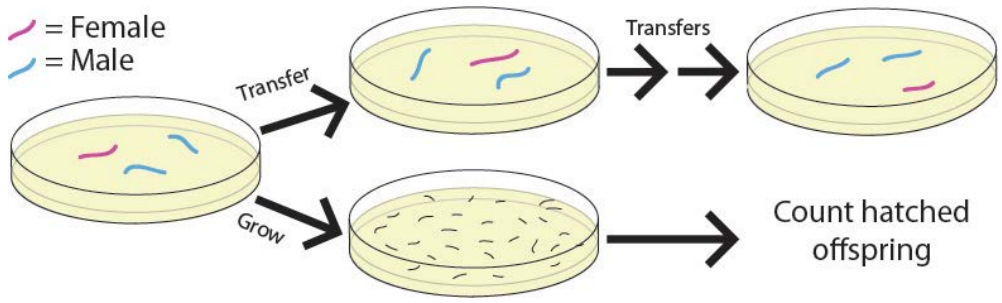

Figure 6. Female fecundity assay design schematic.

Genomic Analysis

As mentioned previously, due to time limitations, we were unable to sequence all the populations in this experimental evolution project and have not yet been able to do any genomic analysis. We developed an analytical pipeline by proxy using the ancestral *C. remanei* population (PX443) over the course of its 75 generations of

adaptation to a laboratory environment. Whole genome sequencing had been previously completed for populations at generation 0 (T0), 30 (T30), and 75 (T75). We utilized a *C. remanei* reference genome (strain PX356) previously made in the Phillips Laboratory and publicly available on WormBase (Fierst et al. in preparation). Next-generation sequencing technologies and bioinformatic techniques were used to analyze whole genome sequencing data from these three populations. DNA from multiple individuals from each population was combined prior to sequencing, a technique known as pooled population sequencing (Pool-seq) (Zhu et al. 2012; Schlötterer et al. 2014). This allowed for genomic data from hundreds of individuals in a population to be obtained and sequenced cheaply and efficiently. Reads from each of the three populations were aligned to the reference genome and compared on a locus-by-locus basis to identify regions of differential allele frequency and nucleotide diversity (π).

DNA Extraction and Next-Generation Sequencing

We extracted DNA from our pooled populations with Qiagen DNeasy Blood and Tissue Kit and measured DNA concentrations using the Qubit 2.0 Fluorometer. Sequencing libraries were prepared using the Nextera DNA Library Prep Kit protocol. All sequencing was done on the Illumina HiSeq 2000 platform at the University of Oregon Genomics and Cell Characterization Core Facility. The T0 and T30 populations were paired-end, 150-bp sequenced, and the T75 population was single-end, 100-bp sequenced.

Bioinformatic Pipeline

The next-generation sequencing yielded raw data that needed to be processed prior to analysis. First, we examined the quality scores of the FASTQ files containing the sequence reads. We used the FASTX-Toolkit (v. 0.0.14; Hannon Lab 2014) to generate a text summary report of the quality scores for each nucleotide position and to visualize the quality score output (Figure 7). Illumina guidelines suggest that quality scores of 30 and above are sufficient for clinical research (Illumina 2011), which provided us with a valid quality score threshold for our sequencing data. It was expected that the average quality of the tail end of the reads would drop due to sequencer error, so quality scores slightly below 30 in these regions were consideres as meeting the threshold. Based on visual inspection, we verified all FASTQ files contained sequencing data of appropriate quality for filtering and alignment. The process_shortreads component of an open-access sequence analysis package called Stacks (v. 1.42; Catchen et al. 2011, 2013) was used to remove adapter sequences leftover from sequencing and to quality filter the reads. Reads were discarded if they had an uncalled base or had quality score values below 5. Single-end reads were cleaned for adapter_1, and paired-end reads were cleaned for both adapter_1 and adapter_2 (see Glossary).

We aligned our processed sequence reads to the reference genome using three commonly-used alignment programs: Bowtie2 (v. 2.2.9; Langmead and Salzberg 2012), Genomic Short-read Nucleotide Alignment Program (GSNAP; Wu and Nacu 2010), and BBMap (v. 34.41; Bushnell 2015). For all alignment programs, the open-source SAMtools package (v. 1.3.1; Li et al. 2009) was used for sorting, indexing, and

computing mean coverage. The alignment from the program with the best mean coverage and standard deviation was used for the statistical analyses. All alignments were stored as Sequence Alignment/Map (SAM) files, and these SAM files were converted to binary SAM (BAM) files, sorted, indexed, and converted to mpileup using SAMtools.

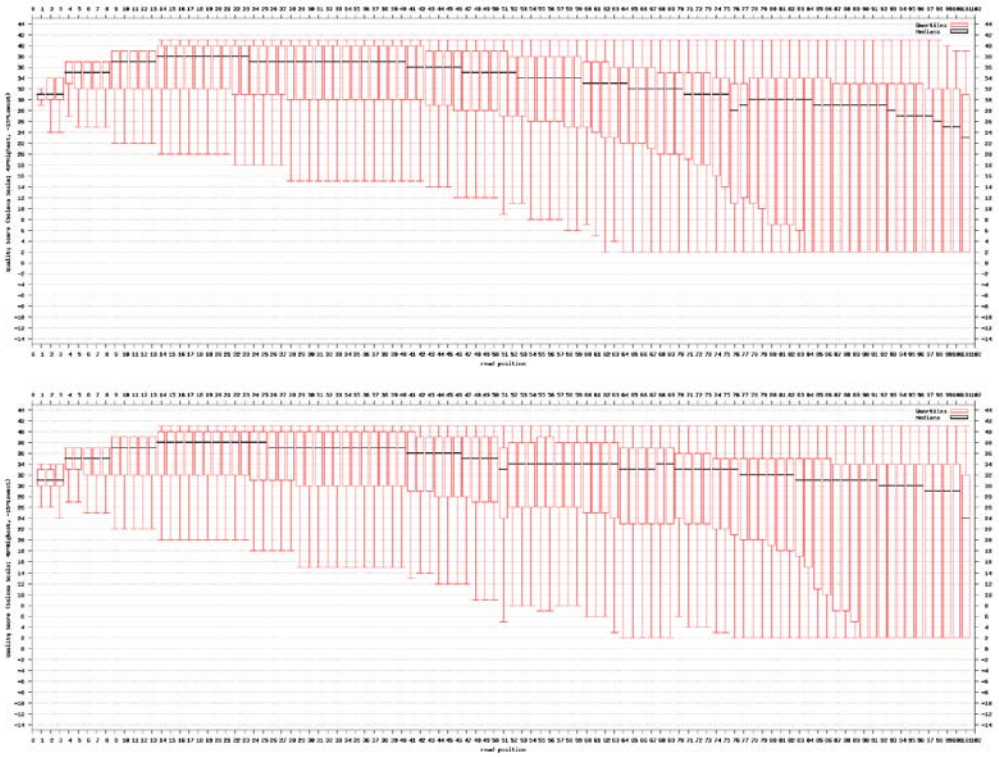

Figure 7. Quality scores for paired-end 150-bp reads of the T0 ancestor population. Horizontal, black lines indicate median quality for the associated read position; red rectangles represent the interquartile range (IQR) between quartiles 1 (bottom value) and 3 (top value); whiskers show maximum 1.5 IQR.

We used the PoPoolation (Kofler et al. 2011a) software package for calculating π, which is a single-population statistic. We filtered the mpileup files to even out the coverage using a minimum mapping quality value of 20, maximum allowed coverage of 500×, target coverage of 50×, and a without-replacement method for subsampling (note

that in PoPoolation for FASTQ type, the sanger modification is used for phred33 encoding). We calculated π with a sliding window of 5,000 bp with an overlapping step size of 3,500 bp. We required a minimum of 2 alleles for calling a SNP and a chromosome pool size of 1,000.

The PoPoolation2 (Kofler et al. 2011b) software package was used for pairwise population analyses. The SAMtools mpileup feature was used to merge the BAM files from the T0, T30, and T75 populations and to convert the resulting merged file to a synchronized mpileup file with PoPoolation2's helper jar, mpileup2sync.jar, set with a minimum quality of 20 and 12 threads. Pairwise Fisher's exact test was calculated per SNP. The coverage maximum was set to top 2 percent of sequencing depth distribution, which correlated with 291×, 137×, and 186× for the T0, T30, and T75 data, respectively. Coverage depth is an important variable to consider when conducting statistical analyses on sequencing data. Low coverage can result in artificially low estimations of nucleotide diversity (Lynch 2008) and can lead to the propagation of sequencing errors in downstream analyses (Sims et al. 2014). Taking these risks into account, minimum coverage was set to 50×, which yielded sufficient power for determining significant disparities between populations.

Results

Female Fecundity Assays

We measured female fecundity as a phenotypic indicator of selection and a proxy for fitness using the assay described in the Methods chapter. Assays were conducted in both 31˚C and 20˚C phenotyping environments. The number of assays we completed as of 21 April 2017 are listed in Table 1.

		Assayed at 31˚C				Assayed at 20˚C			
		m=0.00		m=0.05		m=0.00		m=0.05	
Gen.	Rep.	Heat	Control	Heat	Control	Heat	Control	Heat	Control
30	1	15	33	25	25	-	-	-	-
	2	25	26	28	28	-	-	-	-
	3	25	12	26	24	-	-	-	-
	4	19	29	28	30	-	-	-	-
	5	29	29	23	30	-	-	-	-
40	1	-	-	30	30	-	-	30	30
	2	30	-	30	30	29	-	-	-
	3	-	-	-	-	-	-	-	-
	4	30	-	-	-	-	-	-	-
	5	29	32	30	30	20	30	27	43
Ancestor		56				69			

Table 1. Number and distribution of female fecundity assays completed.
Gen. = generation; Rep. = replicate; m = migration; Heat = 31˚C-evolved; Control = 20˚C-evolved; a "-" indicates that assays have not been conducted for the replicate.

Generation 30 in the Novel Environment (31˚C)

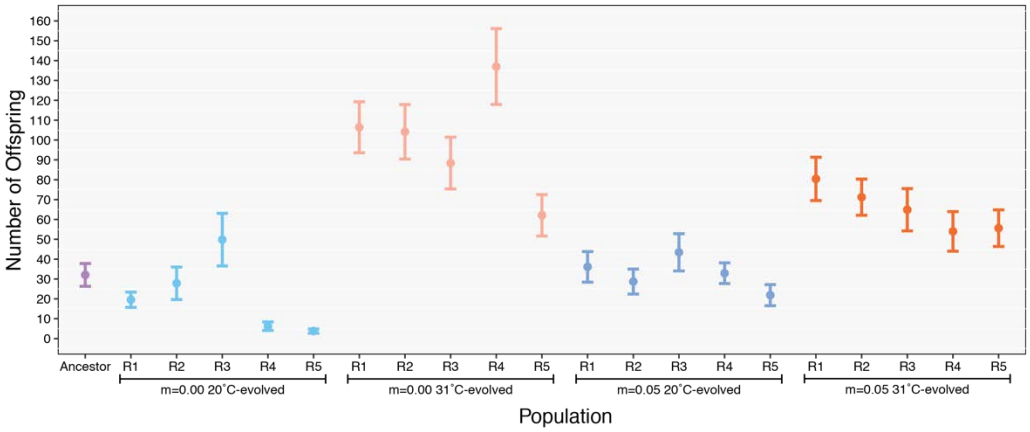

Figure 8. Female fecundity at 31˚C (ancestor and generation 30).

Female fecundity assay results for the ancestor and descendant populations of the no-migration (m=0.00) and 5-percent (m=0.05) migration treatments assayed at generation 30 in the novel, 31˚C phenotyping environment. R = replicate; midpoint represents mean female fecundity; error bars indicate ±1 SE.

In the novel environment, assays were completed for the ancestor and all biological replicates of the descendant populations with both the no-migration and 5-percent treatments at the generation 30 time-point (Figure 8). All replicates of the no-migration 31˚C-evolved population exhibited significantly greater average female fecundity than all replicates of the no-migration 20˚C-evolved population (p-value = 4×10^{-14}), which indicates there was a significant effect of selection in the chronic heat stress environment. We do not observe an independent effect of migration, evidenced by the lack of significant difference in average female fecundity between the replicates of the no-migration and 5-percent 20˚C-evolved populations (p-value = 0.08). Considering the compounding effects of the 5-percent migration treatment and the heat stress environment, we see a significant interaction between migration and selection (p-value = 5×10^{-4}).

Generation 30 and 40 in the Novel Environment (31˚C)

In addition to the female fecundity assays completed for the generation 30 time point in the 31˚C phenotyping environment, select replicates of each migration treatment have also been assayed at generation 40 (Figure 9). Note the differences in the scale of the axes between the no-migration and 5-percent populations. Qualitatively, it seems that each replicate's mean fecundity is relatively similar between generation 30 to 40 in the no-migration population and increases slightly between generation 30 to 40 in the 5-percent population. However, the significance of these observations has yet to be determined. It is important to note that the number of assays completed for populations at generation 30 is much greater than the number completed at generation 40, making these observations solely speculative at this point.

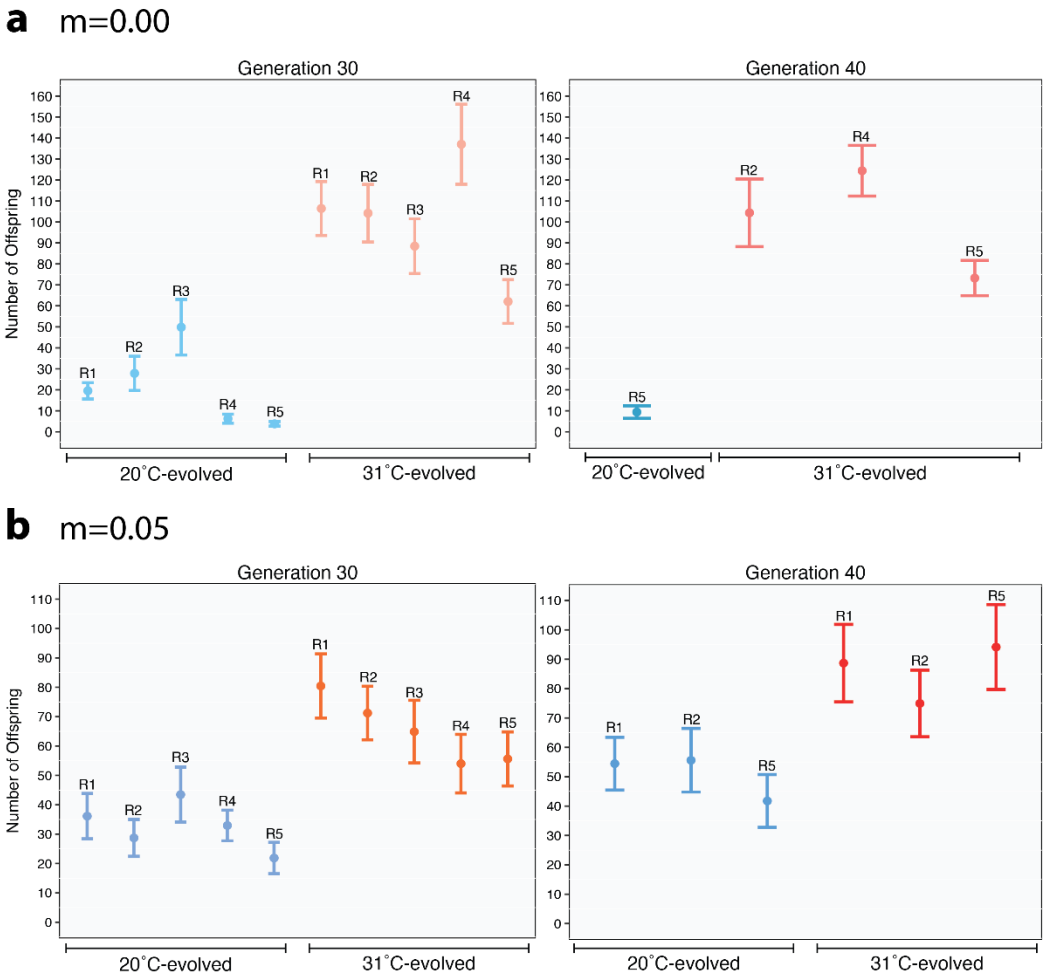

Figure 9. Female fecundity at 31˚C (generation 30 and 40).

Female fecundity assay results for the experimentally-evolved populations of the (A) no-migration (m=0.00) and (b) 5-percent (m=0.05) migration treatments assayed at generation 30 and 40 in the 31˚C phenotyping environment. R = replicate; midpoint represents mean female fecundity; error bars indicate ±1 SE.

Generation 40 in the Ancestral Environment

Fecundity assays were also conducted for the ancestor and select populations at the generation 40 time-point in the ancestral environment, the 20˚C laboratory standard for *C. remanei* (Figure 10).

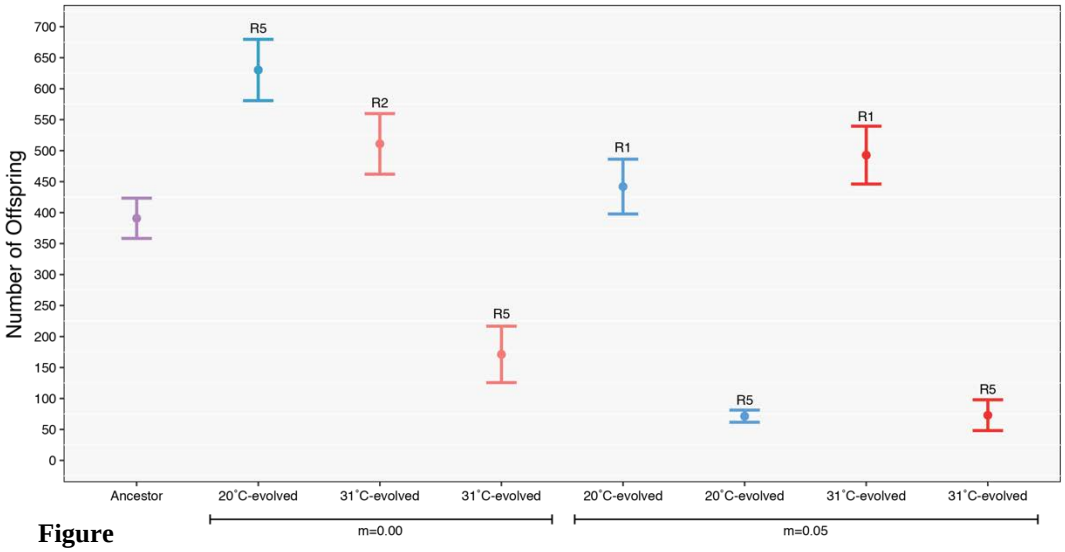

Figure 10. Female fecundity at 20˚C.

Female fecundity assay results for the ancestor and descendant populations at the generation 40 time-point in the 20˚C phenotyping environment. R = replicate; midpoint represents mean female fecundity; error bars indicate ±1 SE.

Genomic Analysis

We developed a bioinformatic pipeline for aligning and analyzing whole genome sequencing data. A representative study using this pipeline was conducted for three time points during the laboratory adaptation of the ancestral *C. remanei* population from which the experimental evolution populations were derived: 0, 30, and 75 generations. We calculated average π for overlapping sliding windows from each population and used Fisher's exact test to find regions of significant divergence between pairs of the populations. These statistics were plotted using the Circos (Krzywinski et al 2009) software package, and only the ten largest contigs are shown.

Alignment Program Selection

We wanted to determine which alignment program yielded the greatest mean coverage for the three populations. We aligned our sequence reads to the reference genome using three different programs that utilized slightly different algorithms: Bowtie2, GSNAP, and BBMap. Using the SAMtools depth command, we calculated the average and standard deviation of the depth of coverage per position (Table 2). The highest mean coverage resulted from using GSNAP, so we used the GSNAP alignment for all subsequent analyses with the PoPoolation and PoPoolation2 software packages.

Population	Bowtie2		GSNAP		BBMap	
	Average	SD	Average	SD	Average	SD
T0	103.189	134.899	136.73	167.844	132.146	168.061
T30	47.8095	88.0653	59.7012	104.809	58.8994	105.949
T75	46.7113	86.6736	62.3245	109.895	54.6579	101.099

Table 2. Coverage statistics for Bowtie2, GSNAP, and BBMap alignments.

SD = standard deviation.

Nucleotide Diversity (π)

Nucleotide diversity (π) is a statistical measure of how much polymorphism exists within a single population (Nei and Li 1979). It is defined as the per site average of nucleotide differences between any two DNA sequences from the population in question. It serves as an indicator of genetic variation within a population and can be used to analyze the extent of divergence between multiple populations. Using the PoPoolation software package, we calculated π with a sliding window of 5,000 bp with an overlapping step size of 3,500 bp (Figure 11). The genome-wide average π (π_{Tn}) was calculated for each population: π_{T0} = 0.0195819, π_{T30} = 0.013675, and π_{T75} = 0.013391.

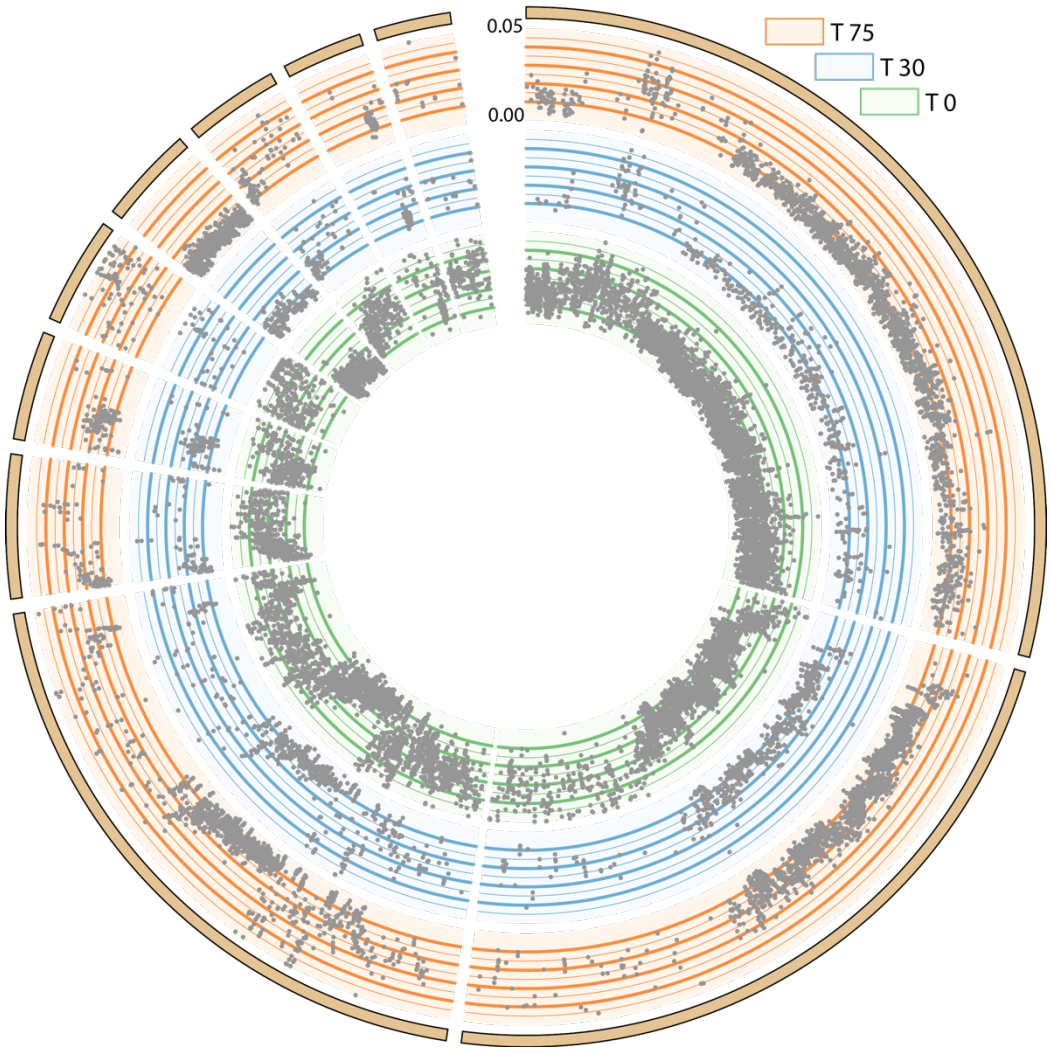

Figure 11. Laboratory adaptation analysis: Nucleotide diversity.

Nucleotide diversity (π) is plotted for each of the three populations in concentric plots. The outer beige ring is broken into contigs. Each gray dot indicates π for one of the overlapping 5,000-bp windows. Values of π range from 0.00 to 0.05 for each ring. Thick and thin solid lines through each ring indicate 0.005 increases in scale.

Fisher's Exact Test

Fisher's exact test can determine whether allele frequency differences between two populations are significant (Fisher 1922). Statistical analysis of genome sequencing data is often limited by low coverage since sampling effects can greatly alter absolute allele frequencies and pairwise F_{ST} calculations. Fisher's exact test yields more accurate

calculations than the chi-squared test and G-test of independence when sample sizes are small. We conducted Fisher's exact test for each pairwise combination of populations: T30 and T75, T0 and T75, and T0 and T30 (Figure 12).

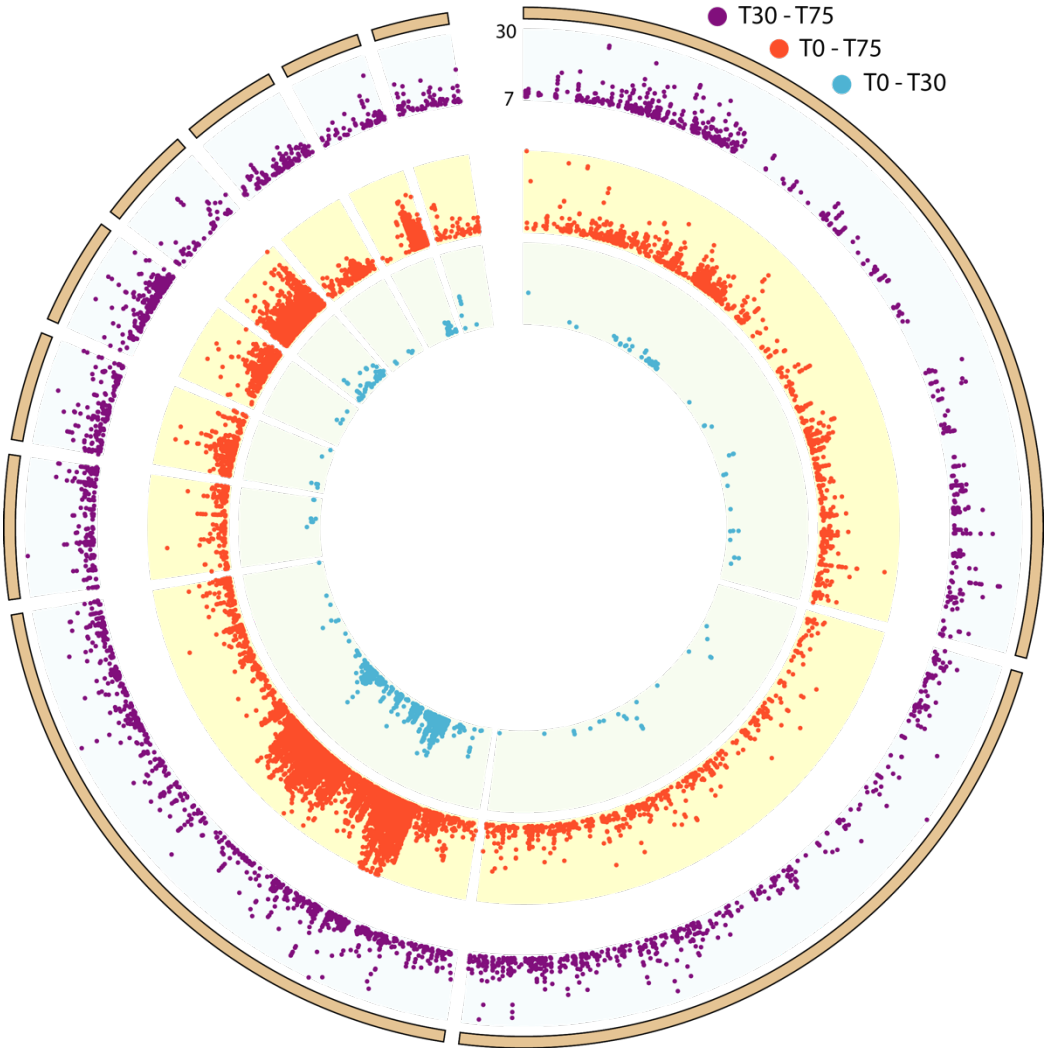

Figure 12. Laboratory adaptation analysis: Fisher's exact test.

Results from pairwise Fisher's exact tests are plotted for each of the three population combinations in concentric plots. The outer beige ring is broken into the same contigs as seen in Figures 11 and 12. Each dot indicates the $-\log_{10}$(p-value) for a single site. 6,248 sites are plotted for T30-T75; 29,251 sites are plotted for T0-T75; and 1,730 sites are plotted for T0-T30. The scale is from 7 to 30, which correlates with a p-value range of 1×10^{-7} to 1×10^{-30}.

Discussion

We employed an evolve-and-resequence approach to address hypotheses regarding genome-wide patterns that evolve under the co-influence of migration and selection. Female fecundity was used as a phenotypic indicator of the effect of selection and as a proxy for fitness. Data supports our hypothesis that populations of *C. remanei* would adapt to the novel, 31˚C environment. Our data also supports the claim that populations can diverge in the face of gene flow. In the second stage of this project, we developed a bioinformatic pipeline for future analysis of whole genome sequencing information from our experimentally evolved populations. We analyzed laboratory-adaptation data from the ancestral *C. remanei* population to test its utility and investigate levels of genetic variation in our ancestral population.

Female Fecundity

Female fecundity is a metric that can be used to estimate the degree of adaptation to the novel environment and indicate if there is a cost to chronic heat stress resistance and/or if populations are in a recovery period after heat stress exposure. We conducted female fecundity assays for multiple replicates of all our populations in both the novel (31˚C) and ancestral (20˚C) environments to determine the effects of adaptation on female reproduction.

Generation 30 in the Novel Environment (31˚C)

For worms at generation 30 in the 31˚C phenotyping environment, females from all five replicates of the 31˚C-evolved population had a significantly greater average

fecundity than females from all five replicates of the 20˚C-evolved population under both migration treatments. The no-migration 31˚C-evolved females exhibited a mean fecundity approximately 5.47-fold greater than the 20˚C-evolved, which demonstrates that there was a significant effect of selection. Combining fecundity counts for all the 5-percent 31˚C-evolved females, their mean fecundity was 2.03-fold than the 5-percent 20˚C-evolved females. This difference in fold-change suggests that the effect of selection on female fecundity was dependent on the effect of migration, which we determined was a significant interaction. These results also show that our populations diverged even in the face of a high rate of gene flow.

We can conclude that the females' increased fecundity is not attributable as byproducts of laboratory adaptation or genetic drift, for in comparison, the 20˚C-evolved worms that served as a control for selection show significantly lower mean fecundity. Since the 20˚C-evolved worms experienced identical laboratory conditions except for chronic temperature environment, it is reasonable to conclude that the increased fecundity of the 31˚C-evolved worms is due to adaptation to their chronic heat stress environment.

For 31˚C-evolved generation 30 worms assayed in the novel environment, 5-percent females exhibited a mean fecundity that was 68.2% of the mean female fecundity of the no-migration females. These results suggest that the alleles entering the 31˚C-evolved population from the 20˚C-evolved population had a negative effect on female fecundity at 31˚C, which supports our hypothesis that migration would slow the rate of adaptation to the novel heat stress environment. In contrast to the 31˚C-evolved females, the mean female fecundity of all the 5-percent 20˚C-evolved worms was

183.2% that of all the no-migration 20˚C-evolved worms. However, when the variance between replicates is considered, there was no significant difference in mean female fecundity between the no-migration and 5-percent 20˚C-evolved worms in the novel environment. This indicates that there was not an independent effect of migration.

Generation 30 and 40 in the Novel Environment (31˚C)

We noted that for the no-migration populations assayed in the novel environment, there seemed to be little difference in mean fecundity between generation 30 and 40 for both the 20˚C- and 31˚C-evolved worms. This phenotypic trajectory could indicate possible temporal effects on phenotype as a result of adaptation and could be indicative of a potential upper limit for female fecundity in the heat stress environment. Such a limit could result from a loss of available genetic variation in the populations. However, between generation 30 and 40 of the 5-percent populations at 31˚C, mean female fecundity appears to increase. This suggests that these populations are still adapting to their environments.

Generation 40 in the Ancestral Environment

When assayed in the ancestral environment, all replicates of the no-migration 20˚C-evolved females seemed to exhibit a higher mean fecundity than the ancestor, but the significance of this relationship has yet to be determined. The ancestor had already been laboratory-adapted for 75 generations prior to the commencement of this project, so this result suggests that 40 more generations of adaptation to the standard laboratory environment could have had a positive effect on female fecundity. However, we have not conducted female fecundity assays for generation 30 in the ancestral environment,

and we have not assayed all replicates at generation 40. Therefore, this conclusion is purely speculative and requires greater statistical rigor.

The female fecundity assays conducted at in the ancestral environment revealed an interesting trend regarding migration. The no-migration replicate 5 females exhibited very different mean fecundity depending on whether they were 20˚C- or 31˚C-evolved; the mean fecundity of the replicate 5 20˚C-evolved females was approximately 3.5-fold higher than that of the combined-replicate 31˚C-evolved females. Such a large disparity is not observed in the 5-percent migration populations, and the 31˚C-evolved worms demonstrated higher mean fecundity. For replicate 1 of the 5-percent population, the 31˚C-evolved females exhibited a mean fecundity 1.12-fold greater than the 20˚C-evolved females, and for replicate 5, the difference was even smaller with the mean fecundity of the 31˚C-evolved females being 1.03-fold greater than that of the 20˚C-evolved females. While these observations would support our hypothesis that gene flow would homogenize diverging populations, the number of fecundity assays done in the 20˚C phenotyping environment are few and we have not determined the significance of these relationships. Therefore, we cannot make any definitive conclusions.

Genomic Analysis

Laboratory Adaptation Project

Our analysis of whole-genome sequencing data from the generation 0, 30, and 75 time points of our ancestral *C. remanei* population has demonstrated the utility of our bioinformatic pipeline. It also allowed us to explore the trajectory of laboratory

adaptation in our ancestral population and investigate how this adaptation affected genetic variation. Additionally, we were able to make preliminary estimates of the pre-existing genetic variation upon which our experimental evolution selected. *C. remanei* is considered a hyperdiverse species (Cutter et al. 2013), since wild populations of *C. remanei* exhibit a mean neutral nucleotide diversity (π_{neu}) that is greater than 5 percent (Cutter et al. 2006; Dey et al. 2012). Once a sample from a wild population is brought into in the lab, the population size greatly decreases, as is the case with our ancestral strain, PX443.

We expect that genetic variation in this strain decreased over its 75 generations of laboratory adaptation because of genetic drift, which is exacerbated by the decrease in population size going from the wild into the laboratory. Additionally, genomic regions with decreased genetic variation could correspond to loci within the regions that were important for adaptation to the laboratory environment. Since we only have a single replicate of this laboratory adaptation, we cannot formally differentiate between the effects of genetic drift and selection on π. With multiple replicates, we could potentially identify sites under selection in the laboratory environment.

Alignments of both the T30 and T75 populations had relatively low average coverage in comparison to the T0 population, which could limit the power of inter-population comparisons of π. We see a trend of decreasing average π over time: π_{T0} (0.0195819) > π_{T30} (0.013675) > π_{T75} (0.013391). This is an expected result in populations affected by genetic drift. With additional sequencing information from the experimentally evolved populations, we will be able to explore population divergence from the perspective of overall genetic variation.

Projections for Experimental Evolution Project

The analytical pipeline developed for the laboratory adaptation data will be applied to our experimentally evolved populations as well. Comparing generational time points of our three migration-treated populations may elucidate any temporal effects of gene flow on genomic architecture over an extended period of adaptation. Additionally, inter-treatment comparisons could provide insight into how varying levels of gene flow affect whole genome evolution over time. These experimental data have the potential to either verify or refute proposed evolutionary models that attempt to explain observations in wild populations.

If regions are under selection in a novel environment, we would expect adaptive alleles to be swept to higher frequencies than non-adaptive alleles. This reduces the genetic variation at those loci. In our no-migration populations, regions of decreased π in the 31˚C-evolved population not present in the 20˚C-evolved population could contain genes important for adaptation to the novel environment. We would also expect these regions to appear in the 5-percent 31˚C-evolved population. However, with the influx of non-adaptive alleles from the 20˚-evolved population, the pattern of genetic variation will be slightly different. In the no-migration populations, divergence is expected to be widespread across the genome due to the combined effects of drift and selection. The migration treatment is expected to homogenize loci of small effect and loci that diverged because of drift. This will yield strong signatures of differentiation around loci of large effect.

Future Directions

Female Fecundity Assays

A major limitation for this project was time. Our timeline prevented us from completing female fecundity assays for all replicates of all populations by the time of this writing, which has lessened the power of our observations, especially for any comparisons between worms at generation 30 and 40. We intend to complete fecundity assays for all remaining replicates of both the no-migration and 5-percent treated populations at generation 30 and 40 to allow for more robust analyses and interpretations.

Sterility Assays

For migration to affect the genetic composition of the receiving populations, non-resident worms must be able to successfully interbreed with resident worms. If any of our experimentally-evolved worms were sterile, our theoretical migration rate would have been greater than the realized migration rate. We will do four-way combinations of worms in both the ancestral and novel environments (20˚C and 31˚C, respectively) for each evolved population assayed: resident male/resident female, resident male/non-resident female, non-resident male/resident female, non-resident male/non-resident female, where residency is dictated by assay environment. For these assays, one virgin female and one male will be isolated on 35-mm Petri plates and allowed 24 hours to mate. Worms will be reported as being either sterile or fecund. “Fecund” will be defined

as exceeding the replacement rate, i.e. producing ≥2 successfully hatched offspring on the plate, and "sterile" will be defined as producing <2 successfully hatched offspring.

Experimental Evolution

At the end of 30 generations of experimental evolution, we observed a significant effect of selection and a significant interaction between migration and selection in the chronic heat stress environment. Initially, we expected a 5-percent migration rate to be on the high end and that it had the potential to completely counter the effects of selection. However, our results indicate that selection in the heat stress environment was strong enough to yield an adaptive response in the worm populations, albeit to a lesser degree than without migration. To exaggerate the effects of migration on adaptation to the novel heat stress environment, we have experimentally evolved a third set of worms in the chronic 31˚C and 20˚C environments for 40 generations with a 20-percent migration treatment. The next steps will be to collect female fecundity, sterility, and whole genome sequencing data from the five biological replicates of these populations. With such a high migration rate, we might predict the presence of extensive homogenization between the populations. We expect the genomic signals between the low (5-percent) and high (20-percent) migration rates to differ in number and relative size. The signals present in populations treated with a 20-percent migration rate should correlate with genes under strong selection in the novel, chronic heat stress environment.

Genomic Analysis

We tested the utility of bioinformatic pipeline outlined in the Methods chapter by analyzing whole genome sequencing data from the ancestral *C. remanei* population (PX443) at three different time points during its propagation in the laboratory environment: generations 0, 30, and 75. One of the final stages of the experimental evolution project will be to collect whole genome sequencing data for all our evolved populations of various migration-selection combinations and to analyze them using this pipeline. Our genomic analysis could end with a comprehensive overview of the number and location of putative loci that are under selection. We could also attempt to describe the specific targets of selection. If we calculate Tajima's D (Tajima 1989) for each population, we could identify regions of the genome that are likely to have evolved due to non-neutral processes. Genomic regions with low Tajima's D values could allow us to identify possible genes likely to be under selection in the chronic heat stress environment. Filtering for low Tajima's D values isolates regions that are likely to have evolved in a non-neutral manner. Within these regions, genes containing SNPs deemed divergent by Fisher's exact test become putative targets of selection. These genes can then be matched to molecular functions and potential biological roles and importance. Deitz et al. (2016) employed a similar paradigm for a gene ontology analysis in *Anopheles melas* that utilized statistics calculated with the PoPoolation2 software, which indicates its feasibility.

Broader Implications

Current research has found that stressful conditions of varying temporality and severity induce diverse physiologic changes in animals that promote cell protection and preservation, which reduces the metabolic cost to the organisms, decelerates aging, and extends their lifespans (Kenyon 2010). In *C. elegans*, one pathway that produces this response is the insulin/insulin-like growth factor-1 signaling pathway. It promotes the activation of *daf*-16, a transcription factor that up-regulates the transcription of numerous genes involved with heat shock proteins (HSPs) (Kenyon 2010), and heat-shock factor-1 (HSF-1), which controls HSPs to prevent cell damage (Epel et al. 2004). Expression of HSF-1 and HSPs is induced under conditions of heat stress.

Theory on the effects of stress claims that stressful environments can accelerate aging and decrease lifespan due to deleterious effects on cell function and structure (Parsons 1995). Thus, a trait or a set of traits that confer stress resistance could halt lifespan reduction or positively affect longevity. Stress resistance is an evolutionarily conserved adaptive mechanism that has been found to extend lifespans of multiple organisms (Rodriguez et al. 2013), but research on the stress response pathway's effect in humans and their longevity is ongoing. For example, a study on humans by Epel et al. (2004) concludes that chronically stressful environments and perceived stress leads to decreased telomere length, higher oxidative stress, and reduced telomere activity, all of which contribute to cell senescence, i.e. deterioration with age, and are detrimental to longevity. Because stress response pathways have been linked to aging and homology exists between them and human disease pathways (Rodriguez et al. 2013), determining

these relationships could greatly benefit and advance research into human aging, neurodegeneration, and cancer (Table 3).

Stress Experiment	*C. elegans* gene	Human homolog/ortholog	Disease
Heat-shock	daf-16	FOXO3A	Cancer
Hypoxia	egl-9	EGLN	Cancer
Oxidative	sod-1, -2, -3	SOD1, 2, 3	Amyotrophic lateral sclerosis
Oxidative	pink-1	PINK1	Parkinson's

Table 3. Types of stress experiments to examine genes in *C. elegans*, their human homologs, and associated diseases.

Information is modified from Rodriguez et al. (2013) Table 1.

In addition to applications to human health, understanding these genetic networks could reveal insight into adaptive changes that various species undergo. In light of global warming and climate change, understanding how species respond to stressful environments will be vital in preserving and maintaining populations and communities as their ecological surroundings change. Climate changes have been cited as key abiotic stressors driving evolution (Nevo 2012). Genetic architecture greatly affects how organisms evolve, so dissection of these networks could give researchers a relative measure of a species' evolutionary potential. Hansen (2006) defines this potential as the "ability of the genetic system to produce and maintain potentially adaptive genetic variants." This evolvability dictates whether or not species will be able to respond to their changing surroundings and survive in novel environments.

Conclusion

A major challenge facing evolutionary biologists today is to understand the trajectories of speciation under the influence of various processes. We experimentally tested the relative effects of migration and selection on adaptation to a novel, heat-stress environment. We observed that a 5-percent migration rate between adapted and non-adapted populations slowed the rate of adaptation to this novel environment but did not completely homogenize the two populations. Our results demonstrate that populations can diverge, even in the face of gene flow, which runs counter to many traditional models of population divergence and speciation.

Organisms experience a wide range of stressful environments throughout their lifetime. Elucidating the genetic architecture of stress response is important because it helps researchers evaluate how a phenotype of interest may respond to selection. By identifying genes and loci that are under selection in the chronic heat stress environment, this project will contribute to the growing base of information about how complex phenotypes change and evolve under selective pressure. Understanding adaptation and phenotypic evolution is vital to addressing major environmental concerns as continual global change threatens to upset the current ecological equilibria. Additionally, because of homology between *C. remanei* and humans, translation of genetic discoveries in nematodes could greatly benefit medical research. Further elucidation of how stress response networks function could contribute to human-based research into age-related illnesses and diseases.

Glossary

adaptation: a change in an organism's function or underlying structure that results from natural selection and that confers survival and/or reproductive advantages to the individual

adapter_1: oligonucleotide index adapter used during Illumina library preparation; the sequence of this adapter is CTGTCTCTTATACACATCTGACGCTGCCGACGA

adapter_2: oligonucleotide index adapter used during Illumina library preparation; the sequence of this adapter is CTGTCTCTTATACACATCTCCGAGCCCACGAGAC

allele: a version of a gene that is found at a particular locus

complex trait: a trait that does not abide by a regular pattern of inheritance, is probably influenced by several genes, and can be expressed as a wide variety of phenotypes

conserved: to remain unchanged throughout the course of evolution, indicating essentiality and uniqueness

contig: an organizational unit of alignments of genome sequencing data; a continuous DNA sequence whose order of bases is known with a high level of confidence

divergence: the evolution of differences, e.g. genetic or morphological, between related species as they adapt to their particular environments

fecundity: the total number of successfully-hatched offspring that a single female produces over a period of time

genetic architecture: the framework of genetic effects that create and regulate a phenotype of interest and its variation that includes, but is not limited to, numbers and locations of loci, mutational and allelic effects, and genetic interactions (Hansen 2006)

genetic drift: the random fluctuations in allele frequencies within a population

genome: the complete set of DNA in a cell or an organism

hermaphrodite: an individual that has both male and female reproductive organs and is able to reproduce without mating with another individual

library (genomic): a collection of the genomic data from a single individual or group of individuals in the case of pooled-population sequencing

locus (pl. loci): the location of a gene on a chromosome

next-generation DNA sequencing: e.g. Illumina sequencing; a descriptor term for technologies recently developed that sequence DNA and RNA faster, more efficiently, and at a lower cost than previous methods, e.g. Sanger sequencing

phenotype: a set of observable traits of an organism that is determined by its environment and its genotype, i.e. its genes

pooled population sequencing (Pool-seq): a laboratory technique that allows for whole genome sequencing of multiple individuals in a pool

PX356: a strain of *Caenorhabditis remanei* used to build the reference genome available to the laboratory

PX443: a strain of *Caenorhabditis remanei* derived from a wild isolate and laboratory-adapted for 75 generations

whole genome sequencing: a laboratory technique that sequences the entirety of an organism's genome

Bibliography

Araya CL, Payen C, Dunham MJ, Fields S. 2010. Whole-genome sequencing of a laboratory-evolved yeast strain. BMC Genomics 11.

Barrick J, Yu D, Yoon S, Jeong H, Oh T, Schneider D, Lenski R, Kim J. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature 461:1243–1247.

Bates D, Maechler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Berner D, Grandchamp A, Hendry AP. 2009. Variable progress towards ecological speciation in parapatry: Stickleback across eight lake-stream transitions. Evolution (N. Y). 63:1740–1753.

Bolnick DI, Fitzpatrick BM. 2007. Sympatric Speciation: models and empirical evidence. Annu. Rev. Ecol. Syst. 38:459–487.

Bushnell B. 2015. BBMap. <sourceforge.net/projects/bbmap/>. Accessed 14 Feb 2017.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: Building and Genotyping Loci De Novo from Short-Read Sequences. Genes|Genomes|Genetics 1:171–182.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: An analysis tool set for population genomics. Mol. Ecol. 22:3124–3140.

Chang, W. 2012. R Graphics Cookbook: Practical Recipes for Visualizing Data. Sebastopol, CA: O'Reilly Media. p 361-364.

Coghlan A, Fiedler TJ, McKay SJ, Flicek P, Harris TW, Blasiar D, Stein LD. 2008. nGASP—the nematode genome annotation assessment project. BMC Bioinformatics 9:549.

Cutter AD, Baird SE, Charlesworth D. 2006. High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of *Caenorhabditis remanei*. Genetics 174:901–913.

Cutter AD, Jovelin R, Dey A. 2013. Molecular hyperdiversity and evolution in very large populations. Mol. Ecol. 22:2074–2095.

Deitz KC, Athrey GA, Jawara M, Overgaard HJ, Matias A, Slotman MA. 2016. Genome-Wide Divergence in the West-African Malaria Vector *Anopheles melas.* Genes|Genomes|Genetics 6:2867–2879.

Dey A, Jeon Y, Wang GX, Cutter AD. 2012. Global population genetic structure of *Caenorhabditis remanei* reveals incipient speciation. Genetics 191:1257–1269.

Epel ES, Blackburn EH, Lin J, Dhabhar FS, Adler NE, Morrow JD, Cawthon RM. 2004. Accelerated telomere shortening in response to life stress. Proc.Natl.Acad.Sci.U.S.A 101:17312–17315.

Feder JL, Egan SP, Nosil P. 2012. The genomics of speciation-with-gene-flow. Trends Genet. 28:342–350.

Feder JL, Flaxman SM, Egan SP, Comeault A a., Nosil P. 2013. Geographic Mode of Speciation and Genomic Divergence. Annu. Rev. Ecol. Evol. Syst. 44:73–97.

Felsenstein J. 1981. Skepticism Towards Santa Rosalia, or Why are There So Few Kinds of Animals? Evolution (N. Y). 35:124–138.

Fisher RA. 1922. On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P. J. R. Stat. Soc. 85:87–94.

Fox J and Weisberg S. 2011. An {R} Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage. <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>.

Graustein A, Caspar JM, Walters JR, Palopoli MF. 2002. Levels of DNA polymorphism vary with mating system in the nematode genus *Caenorhabditis*. Genetics 161:99–107.

Griswold CK. 2006. Gene flow's effect on the genetic architecture of a local adaptation and its consequences for QTL analyses. Heredity (Edinb). 96:445–453.

Hannon Lab. 2014. FASTX-Toolkit. <http://hannonlab.cshl.edu/fastx_toolkit/>. Accessed 3 Jan 2017.

Hansen TF. 2006. The evolution of genetic architecture. Annu. Rev. Ecol. Evol. Syst. 37:123–157.

Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, et al. 2016. A new view of the tree of life. Nat. Microbiol. 1:16048.

Illumina. 2011. Quality scores for next-generation sequencing: Assessing sequencing accuracy using Phred quality scoring. <https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/technote_Q-Scores.pdf>. Accessed 19 Apr 2017.

Jordan DS. 1905. The origin of species through isolation. Science. 22:545–562.

Kenyon CJ. 2010. The genetics of ageing. Nature 464:504–512.

Kiontke K and Sudhaus W. 2006. Ecology of *Caenorhabditis* species. *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.37.1, http://www.wormbook.org. Accessed 31 Apr 2017.

Kofler R, Orozco-terWengel P, de Maio N, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlötterer C. 2011. Popoolation: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PLoS One 6.

Kofler R, Pandey RV, Schlötterer C. 2011. PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics 27:3435–3436.

Krzywinski M, Schein J, Birol İ, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19:1639–1645.

Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357-359.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Long A, Liti G, Luptak A, Tenaillon O. 2015. Elucidating the molecular architecture of adaptation via evolve and resequence experiments. Nat. Rev. Genet. 16:567–582.

Lynch M. 2008. Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects. Mol. Biol. Evol. 25:2409–2419.

Mayr E. 1963. Animal species and evolution. Cambridge, MA: Belknap Pr of Harvard Univ Pr. p 424-515.

Michalak P, Minkov I, Helin A, Lerman DN, Brian R, Feder ME, Korol AB, Nevo E, Minkovt I, Lerr DN, et al. 2001. Genetic evidence for adaptation-driven incipient speciation of *Drosophila melanogaster* along a microclimatic contrast in “Evolution Canyon,” Israel. Proc. Natl. Acad. Sci. U. S. A. 98:13195–13200.

Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. 2011. How many species are there on earth and in the ocean? PLoS Biol. 9:1–8.

Nei M, Li W-H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. U. S. A. 76:5269–5273.

Nevo E. 2012. “Evolution Canyon,” a potential microscale monitor of global warming across life. Proc. Natl. Acad. Sci. 109:2960–2965.

Nosil P, Feder JL. 2012. Genomic divergence during speciation: causes and consequences. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 367:332–42.

Nosil P, Funk DJ, Ortiz-Barrientos D. 2009. Divergent selection and heterogeneous genomic divergence. Mol. Ecol. 18:375–402.

Nosil P, Schluter D. 2011. The genes underlying the process of speciation. Trends Ecol. Evol. 26:160–167.

Parsons PA. 1995. Inherited stress resistance and longevity: a stress theory of ageing. Heredity (Edinb). 75 (Pt 2):216–221.

Reynolds RM, Phillips PC. 2013. Natural Variation for Lifespan and Stress Response in the Nematode *Caenorhabditis remanei*. PLoS One 8.

Rodriguez M, Basten Snoek L, De Bono M, Kammenga JE. 2013. Worms under stress: *C. elegans* stress response and its relevance to complex human disease and aging. Trends Genet. 29:367–374.

Rosenblum EB. 2006. Convergent Evolution and Divergent Selection: Lizards at the White Sands Ecotone. Am. Nat. 167:1–15.

Savolainen V, Anstett M-C, Lexer C, Hutton I, Clarkson JJ, Norup M V, Powell MP, Springate D, Salamin N, Baker WJ. 2006. Sympatric speciation in palms on an oceanic island. Nature 441:210–213.

Schlötterer C, Tobler R, Kofler R, Nolte V. 2014. Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. Nat. Rev. Genet. 15:749–763.

Sikkink KL, Reynolds RM, Cresko WA, Phillips PC. 2015. Environmentally induced changes in correlated responses to selection reveal variable pleiotropy across a complex genetic network. Evolution (N. Y). 69:1128–1142.

Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. 2014. Sequencing depth and coverage: key considerations in genomic analyses. Nat. Rev. Genet. 15:121–32.

Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. Genet. Res. Camb. 23:23–35.

Stiernagle T. 2006. Maintenance of *C. elegans*. *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.101.1, http://www.wormbook.org. Accessed 31 Apr 2017.

Strasburg JL, Sherman N a, Wright KM, Moyle LC, Willis JH, Rieseberg LH. 2012. What can patterns of differentiation across plant genomes tell us about adaptation and speciation? Philos. Trans. R. Soc. Lond. B. Biol. Sci. 367:364–73.

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

Turner TL, Hahn MW, Nuzhdin S V. 2005. Genomic islands of speciation in *Anopheles gambiae*. PLoS Biol. 3:1572–1578.

Turner TL, Stewart AD, Fields AT, Rice WR, Tarone AM. 2011. Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster*. PLoS Genet. 7.

Via S. 2001. Sympatric speciation in animals:the ugly duckling grows up. Trends Ecol. Evol. 16:381–390.

Via S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. Philos. Trans. Biol. Sci. 367:451–460.

Wu C-I. 2001. The genic view of the process of speciation. J. Evol. Biol. 14:851–865.

Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26:873–881.

Yeaman S, Whitlock MC. 2011. The genetic architecture of adaptation under migration-selection balance. Evolution (N. Y). 65:1897–1911.

Zhu Y, Bergland AO, González J, Petrov DA. 2012. Empirical Validation of Pooled Whole Genome Population Re-Sequencing in *Drosophila melanogaster*. PLoS One 7:1–7.