QUANTITATIVE METHODS TO UNDERSTAND REPRODUCTIVE ISOLATION THAT CONTRIBUTE TO SPECIATION

A Dissertation
Presented to
The Academic Faculty

by

Lijiang Long

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Quantitative Biosciences in the
School of Biological Sciences

Georgia Institute of Technology
Dec 2020

QUANTITATIVE METHODS TO UNDERSTAND REPRODUCTIVE ISOLATION THAT CONTRIBUTE TO SPECIATION

Approved by:

Dr. Patrick T. McGrath, Advisor
School of Biological Sciences
Georgia Institute of Technology

Dr. Greg Gibson
School of Biological Sciences
Georgia Institute of Technology

Dr. Jeffery T. Streelman
School of Biological Sciences
Georgia Institute of Technology

Dr. Peng Qiu
School of Biomedical Engineering
Georgia Institute of Technology

Dr. Annalise Paaby
School of Biological Sciences
Georgia Institute of Technology

Date Approved: Oct 21, 2020

To my parents

ACKNOWLEDGEMENTS

First of all, I would really like to give my sincere gratitude to Dr. Patrick T. McGrath. He has not only been a superior advisor, but also a friend with whom I can share my life. As an advisor, he used his deep knowledge of evolutionary genetics to guide me through all the complex research problems. At the same time, he encouraged me to explore fields outside biology which will benefit my future career. As a result, I believe I was trained in the most appropriate way by the best advisor in the world.

Besides Dr. Patrick McGrath, I would also like to thank two other advisors when I did rotation in the first year: Dr. Annalise Paaby and Dr. Hang Lu. They both spent lots of time with me during the rotation and they remain important advisors I would consult with whenever I have some questions. I really like Dr. Paaby's gifts to my daughter, including a bowl of homemade chicken soup for my wife.

I would also thank the rest of my committee members: Dr. Greg Gibson, Dr. Todd Streelman and Dr. Peng Qiu. Their feedbacks at our annual meetings are always useful to improve my research and presentation skills.

I also want to thank all the current and past members of McGrath Lab. I had many collaborations with Eric Anderson, Wen Xu, Yuehui Zhao, Weipeng Zhuo, Junyu Li, Manu Tej Sharma Arrojwala, Vineeth Aljapur, Nicole Baran, Zachary Johnson and many others. All these collaborations turned out to be meaningful and I really enjoyed working with you every day.

Outside lab, I would like to thank all my friends especially Zhiqiang Lin, for the numerous dinner we had together, for the roads we traveled, and for the memories we shared.

Finally, I would like to thank my wife Ling Wang who always supports me when I was down, shares happy with me and helps taking care of our baby girl Naomi through countless nights.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
C. elegans	Caenorhabditis elegans
CI	Confidence Interval
CV	Copachromis virganialis
DBSCAN	Density-based spatial clustering of applications with noise
HMM	Hidden Markov Model
L4	Larval stage 4
MC	Mchenga conophoros
NIL	Near Isogenic Line
SLiM	Selection on Linked Mutations
SNP	Single Nucleotide Polymorphism
TA	Toxin-antidote
TI	Tramitichromis intermedius

SUMMARY

Ever since Darwin's qualitative theory of the origin of species, there is growing demand for quantitative methods to study mechanisms underlying the speciation process. One key component towards new species formation is reproductive isolation. Reproductive isolation can be either pre-mating (e.g. mating behavior difference) or post-mating (e.g. Dobzhansky-Muller sites). In this thesis, I will present novel quantitative methods developed to study two aspects of reproductive isolation: mating rituals difference in Lake Malawi cichlids and genetic incompatibilities by selfish genetic elements in *C. elegans*. In chapter 2, I will talk about cichlids bower behaviors where male fishes construct bowers to attract female mates by manipulating sand with their mouths thousands of times over the course of many days. Variations in bower type ('pits' and 'castles') is one mechanism to create nonrandom mating and maintain a large number of cichlids species in Lake Malawi. To enable quantitative comparisons of these behaviors in different species, an automatic behavior quantification pipeline was built. Specifically, pixel-based Hidden Markov Modeling was combined with density-based spatiotemporal clustering for action detection. Each action video clip was then classified into ten categories using a 3D Residual Network (3D ResNet). These ten categories distinguish spitting, scooping, fin swipes and spawning. I showed that this approach is accurate (> 76% accuracy) in distinguishing fish behaviors and animal intent can be determined from these clips, as spits and scoops performed during bower construction are classified independently from spits and scoops performed during feeding. I applied this approach to >700 hours of video recordings taken from seven independent trials encompassing multiple species and hybrid crosses, collectively

containing hundreds of thousands of independent behavioral events. In chapter 3, I use quantitative methods to measure fitness combined with population modeling to study the evolutionary origin of selfish genetic elements and their ability to spread in populations. Previous research found that a toxin-antidote element called *peel-zeel* is under balancing selection. Here, I explore different models that could cause balancing selection on this locus, which make different predictions on the fitness effect of the *peel-zeel* locus in hermaphrodites. However, pair-wise competition assays showed the loss of the toxin gene *peel-1* decreased fitness of hermaphrodites, contradicting my expectation that *peel-1* will decrease animal fitness due to its toxicity. This fitness advantage is independent of the antidote gene *zeel-1*. This work showed that toxin-antidote systems can spread through populations independent of their selfish effects and suggests linked variants for dauer pheromone response could be responsible for the balancing selection. Finally, in chapter 4, I use simulation methods to study the effect of toxin-antidote elements on linked and unlinked genetic variation in the case of admixture. While both simulation and calculation showed toxin-antidote elements are able to quickly spread in a population without toxin-antidote element, the evolution trajectories of the rest of the genome depends on the initial frequency of the toxin-antidote haplotype in the admixed population. Using calculations and simulations, I showed that unlinked neutral genetic variants will increase their frequency when the initial frequency of *peel-zeel* is higher than 1/3 and decrease when the initial frequency of *peel-zeel* is lower than 1/3. My doctoral thesis with many quantitative methods will advance our understanding of the genetic basis of species evolution and evolutional dynamics of selfish genetic elements.

CHAPTER 1. INTRODUCTION

Darwin's famous book 'On the Origin of Species' in 1859 inspired the research projects of evolutionary biology for more than one hundred years. Despite little molecular or genetic understanding of heritability, Darwin's view of speciation is still impacting how we interpret the speciation process. In his view, new species are formed in the same process as natural selection for a particular trait. Natural selection is a gradual process that happens in the population and the long-term differentiation eventually leads to new species. This qualitative description was further advanced by Fisher, Haldane, Wright and many others in the 1930s who developed population genetics models to understand the genetic mechanisms underlying the speciation process (Mayr and Provine, 1998). Another important researcher in this period is Theodosius Dobzhansky. He found that there was far more genetic variability in the field than expected and argued selection was able to drive differentiation among populations (Dobzhansky and Queal, 1938). These population genetics models and other ecological genetics research together improves our understanding of the speciation process. Now speciation is considered a continuous complex process that unify both 'microevolution' and 'macroevolution'. Both ecological features (niche differentiation, population migration) and genetic features (drifting, mutation, recombination, genetic incompatibility, etc.) should be taken into account. These two features often act synergistically to create new species. For example, niche differentiation can accelerate the evolution of reproduction isolation and reproductive isolation can further boost niche differentiation. One route to better understand this process requires us to pinpoint the specific genes responsible for the reproductive isolation. In order to do so, quantitative comparisons of different phenotypes are required. After identifying the gene, we need to figure out the physiological and/or genomic change following the mutation of that gene to understand if niche differentiation is further accelerated by reproductive isolation. To achieve this goal, accurate simulation and fitness assays can help either predict or compare the effect of a specific gene mutation. To motivate the study of reproductive isolation, this chapter gives an introduction to the relationship between reproductive isolation and speciation, what roles selfish genetic elements play and the motivations to use cichlids and *C. elegans* as model in the study of speciation.

1.1 Process of speciation

Speciation is caused by reduced gene flow between populations (Endler, 1977, Wu, 2001). There are many geographical isolations to reduce gene flow between populations. And these different modes are often classified by the extent of geographical separation and gene flow reduction. The most extreme separation and most common is allopatric speciation where two populations are strictly separated by geographical barriers and no gene flow exists between the isolated populations (Lande, 1980). Darwin's Galápagos finch is a classic example for it. All the finches are descendent of a common ancestor about two million years ago (Lamichhaney et al., 2015). Gene flow is not only restricted between the islands and the mainland, but also restricted between the islands. On the other side of the spectrum is sympatric speciation. Sympatric speciation happens when a new species evolves from the ancestral species while both continue to inhabit the same geographic region. In this case, gene flow is almost unrestricted and there is no obvious niche differentiation (Smith, 1966). Studies of the crater Lake Apoyo suggest that cichlids went through sympatric speciation (Barluenga et al., 2006). In a young and small volcanic lake in Nicaragua, phylogeographic and population-genetic studies showed that this lake is seeded by a single ancestor *Amphilophus citrinellus*. A new species Amphilophus zaliosus was evolved from the ancestral species in this small lake. These two species are sympatrically distributed. Thus, it is concluded that only sympatric speciation can account for the evolution of the new species. In the middle of allopatric and sympatric speciation is parapatric speciation where there is still gene flow between nascent species while being geographically separated. It is very difficult to observe in nature because patterns of parapatric speciation can be explained by an alternative mode of speciation.

1.1.1 Reproductive isolation

After persistent reduced gene flow between populations, RI (reproductive isolation) marks the formation of a new species. RI means the two population can no longer exchange genetic materials. The Mechanisms of RI include pre-mating isolation and post-mating isolation (Wu and Palopoli, 1994). In pre-mating isolation, the two species have evolved new phenotypes with respect to behaviors and ecological niches they occupy. Mating preference behavior can lead to assortative mating and finally speciation. For example, it was found that two butterfly species *Heliconius melpomene* and *Heliconius cydno* have evolved different color pattern mimicry. Experiments in the laboratory environment showed that these two species have preference for color pattern when they choose mates: animals with the same color patterns tends to mate with each other than those with different color patterns (Jiggins et al., 2001). As a result, assortative mating reduces the gene flow between the two species, creating a pre-mating isolation.

For post-mating isolation, if mating occurs between two species, the progenies are inviable or sterile. One commonly used model to understand is the two-locus Dobzhansky-Muller model (Johnson, 2000). They reasoned that a single gene is not able to induce incompatibility because of the inherent fitness loss. Thus, a model of two genes interaction induced incompatibility is proposed. In this model, two populations evolved mutations on different genes. The interaction of the two mutations, when animals of the two species cross, creates the hybrid incompatibility. Dobzhansky-Muller type incompatibility could have arisen due to random genetic drifting. More importantly, it builds a link between natural selection and speciation. When animals adapt to the local environment, different mutations could have been fixed in different populations and eventually lead to Dobzhansky-Muller type reproductive isolation.

An example of a Dobzhansky-Muller incompatibility was found in hybrids between *D. melanogaster* and *D. simulans* (Presgraves et al., 2003). Researches showed that the *D. simulans* gene *Nup96* is incompatible with one or more genes on the *D. melanogaster* X chromosome. *Nup96* encodes a subunit of nuclear-pore complex with orthologs in yeast, worm and humans. It is responsible for the selective control of materials transferred between nuclear and cytoplasm of cells expressing it.

1.2 The role of selfish genetic elements in speciation

Selfish genetic elements, also referred to as selfish genes, are genetic segments that gain a transmission advantage over other genetic elements of an individual's genome (Hurst and Werren, 2001, Austin et al., 2009). They can be neutral or harmful to the organisms carrying them. There are many mechanisms used by selfish genetic elements: Homing endonucleases, transposons, B-chromosomes and postsegregation distorters. For postsegregation distorters, there are two types of them based on the location of action: nuclear and cytoplasmic. Both types of postsegregation distorters are able to kill embryos that do not inherit any copies of it. *Medea* (maternal effect dominant embryonic arrest) is a classic example that is identified in flour beetles *Tribolium* (Beeman et al., 1992). Two alleles of *Medea* were reported. One that has a maternal effect toxin as well as a linked antidote. The other wild-type allele does not have these two elements. All the progenies of a heterozygous female will get a 'toxin' expressed from *Medea*. Progenies that carry Medea will have an 'antidote' expressed in the embryos while those don't carry it will be killed. In this way, Medea can quickly spread in a population. The speed of spread is frequency-dependent and is positively correctly with its frequency in the population (Wade and Beeman, 1994). The fast speed to reach fixation makes identification of these postsegregation distorters difficult since they only show their effect in hybrids. Nevertheless, postsegregation distorters have also been identified other eukaryote organisms including vertebrates(Keller and Ross, 1998) (Seidel et al., 2008) (Ben-David et al., 2017) (Noble et al., 2020) (Ben-David et al., 2020). For example, *scat* (severe combined anemia and thrombocytopenia) was found in mice and acts in a maternal fashion (Hurst, 1993).

Selfish genetic elements induced genetic conflict is not rare in creating hybrid incompatibility. One example is *D. pseudoobscura* which has two subspecies (Wang et al., 1997, Machado et al., 2002). Hybrid males between the two subspecies suffers from segregation distortion such that they only yield females. This is because one of the parental species have both the X-linked distorter gene and suppressor while the other species has none of these. Hybrid males have the distorter genes expressed and those sperms carrying the Y chromosome are killed. More molecular genetics studies have identified several other sites responsible for hybrid sterility and hybrid segregation distortion. By comparing the genetic architecture of both phenotypes, a partially overlapped pattern is observed. These findings have led to the idea that hybrid incompatibilities may be the outcome of arm-race between selfish genetic elements and their suppressors in the genome (Werren, 2011).

1.3 Using model organisms to study speciation

In theory, evolution of new species can be studied in any animal. It is also advantageous to study a diverse range of organisms such that mechanisms and/or theories can be generalized instead of being specific to a particular species. Each organism has its unique strengths and weaknesses. For example, Drosophila species are widely used for speciation studies owing to its rapid life cycle, relatively simple genetics with few chromosomes and easy to raise in a laboratory environment. Across my thesis research, two organisms are used. I will introduce the unique properties of these two organisms that make them great systems to study long-term evolution.

1.3.1 Use of cichlids to study pre-mating isolation

Cichlid fishes are famous for its great amount of species in the world. While they are distributed worldwide, they are most diverse in Africa and South America. In east Africa, there are more than 2,000 phenotypically diverse species that evolved in the past ~10 million years. Most cichlids species can be found at relatively shallow depth such as river and lakes (Keenleyside, 1991). Very few of them can be found at saltwater habitats. The lake that has most species is Lake Malawi. It is estimated that Lake Malawi have between

700 to 1000 endemic species and all these species are evolved in the last million years (Kocher, 2004). Most of the species come from a single haplochromine lineage colonizing the lake between 1-5 million years ago (Loh et al., 2013). Different species in Malawi exhibit various phenotypes including color patterning, morphology, jaw shape and behavior. This rapid speciation process is called adaptive radiation and has attracted attention of many researchers. Phylogeny studies based on mitochondria and nuclear DNA have found that riverine cichlids have hybridization events with the lake species many times in history and contributed to the genetic diversity in the lake (Joyce et al., 2011). By sequencing a large number of species in the lake, genetic change including gene duplication, non-coding genomic elements divergence and transposable element insertions have been identified to be the genetic materials for evolution to act on (Brawand et al., 2014). However, the causal genetic and molecular changes that cause a particular phenotype remain largely unexplored. The low genetic diversity between different species make genetic mapping for a particular phenotype between species possible (The average sequence divergence between different species pairs is 0.1%-0.25%).

Besides many other phenotypical differences such as craniofacial and brain patterning, there are two types of dwelling environment: rock and sand (Sylvester et al., 2010, Sylvester et al., 2013, Fraser et al., 2009). In each of the rock and sand group, there are more than 200 species. Among the sand group, there are two major types of bower building exist among these species. They are mounds (‘castles’) and depressions (‘pits’). Male cichlids construct these very large courtship structures by transferring sand for several days. A ‘castle’ male is unlikely to mate with a ‘pits’ female and vice versa. This behavior difference then creates a pre-mating isolation and will reduce gene flow between ‘castles’ and ‘pits’ animals. Comparative genome analysis showed that these two behaviors have evolved multiple times over the evolution of Lake Malawi species. When hybrid animal between ‘castles’ and ‘pits’ are created, they show an interesting behavior that is first marked by a ‘pit’ and then followed by a ‘castle’. This sequential expression of behaviors offers a unique opportunity to look into the genetics underlaying behavior evolution. When hybrid animals transit form ‘pit’ to ‘castle’, expression also transit from ‘pit’ alleles to the ‘castle’ alleles for genes that are highly enriched for neurodevelopment and neural plasticity genes. Even though the specific genes governing behavior changes have not been identified, cis-regulatory evolution seems to be important for the behavior differences.

The diverse behavior differences in Lake Malawi cichlids offer a unique opportunity to study the genetic basis of behavior evolution. At the same time, the unique properties of Lake Malawi cichlids also challenge current genotyping and phenotyping techniques. Compared with the research community for the model organism zebrafish, number of groups to research cichlids is smaller. A result of this smaller community is a not well annotated genome and little knowledge to the specific genes. Similar to many other natural behaviors, cichlids bower behaviors can only be expressed in a natural environment and has to be expressed in a very long term. This creates a challenge on methods both to keep track of the behaviors and compare the behavior differences between different animals.

1.3.2 Use of Caenorhabditis *to genetic basis of long-term evolution*

While cichlids are a great vertebrate model to study speciation, *C. elegans* and other *Caenorhabditis* nematode species have grown as a population model to study long-term evolution. To date, more than 25,000 species have been identified within nematodes and it has been speculated that there could be more than 100 million species (Blaxter, 2003). The *Caenorhabditis* genus belongs to a group of primarily bacterivorous nematodes called rhabditids. In the *Caenorhabditis* genus, the *C. elegans* species is the most famous among genetics researchers due to the work of Sidney Brenner. *C. elegans* is famous for its transparent body, short lifespan and invariant cell lineage (Hobert and Loria, 2005, Chalfie et al., 1981, Hsin and Kenyon, 1999). A large community has built tools and reagents that allow for powerful genetic analysis techniques that can be used to study a trait of interest. It was the first multicellular organism with a whole genome sequence, and information about the expression and function is updated on a daily basis to a shared knowledge database WormBase (Harris et al., 2010).

While historically used for genetic analysis, recent work has developed *C. elegans* as a model organism for the study of evolution. Because of self-fertilization, genetic changes to the individuals of *C. elegans* are trackable such that long-term experimental evolution can be used to study the genetic basis of selection. Its short growing time enables a short turnaround time for many experiments. For example, a short 10 generation fitness competition study can be done within a month, whereas this may take 2 or 3 years for an vertebrate animal (Zhao et al., 2018). Also because of its genetic tractability, mapping of the genetic basis of trait variation is possible in *C. elegans*. The past decade has seen loci accounting for life-history differences, pathogen susceptibility and behavior differences identified (Andersen et al., 2014, Bendesky et al., 2012, Large et al., 2016, Reddy et al., 2009, Lee et al., 2017, McGrath et al., 2009). These genetic mechanisms help us understand the evolutionary forces that have shaped the *C. elegans* genome. Genetic mapping in related species *C. remanei* (Reynolds and Phillips, 2013, Jovelin et al., 2003, O'Connor et al., 2019) and *C. briggsae* (Bi et al., 2015, Ren et al., 2018, Inoue et al., 2007) has identified additional variants for lifespan, stress response, chemosensory behaviors and hybrid incompatibility.

One conclusion from these studies is the importance of balancing selection in *C. elegans*. Within *C. elegans* wild strains, long term balancing selection is able to maintain ancient genetic diversity despite self-fertilization induced effective population size reduction. One of the first reported locus that is under balancing selection is *glc-1* (Ghosh et al., 2012). *glc-1* encodes the alpha-subunit of a glutamate-gated chloride channel that was identified for avermectins resistance. Avermectins are a family of drugs and pesticides that are widely used to treat parasitic worms and insect pests. There are two major haplotypes of *glc-1* that segregate among the wild isolates of *C. elegans*. One haplotype contains a four-amino-acid deletion in the ligand binding domain of GLC-1 that is responsible for the resistance differences to both synthesized avermectins and avermectin-producing bacterium *Streptomyces avermitilis*. This resistance incurs a fitness advantage when anthelmintics such as avermectins are applied in the field or they encounter avermectin-producing bacteria in the wild. It is believed that these two haplotypes have been segreaging in the population for 7.6 million generations. There is then a hypothesis that *glc-1* loss of function might impose a fitness loss to the animals in environments that lack avermectins such that these two alleles are under balancing selection. A second reported locus under balancing selection involves a toxin-antidote system that acts as a paternal effect postsegregation distorter. One haplotype contains both the toxin gene *peel-1* and the antidote *zeel-1* in close linkage. Offspring of heterozygous animals that lack the zeel-1 antidote die, which should result in the spread of *peel-zeel*. It is unknown why both haplotypes persist in the population, however, population genetics suggest that they diverged approximately 8 million generations ago (Seidel et al., 2008). A third characterized locus under balancing selection involves two chemoreceptors *srx-43* and *srx-44*. These two chemoreceptors sense a component of pheromone, called icas#9 (Greene et al., 2016a, Greene et al., 2016b). This locus affects the foraging behavior in response to icas#9 due to differences in expression level of *srx-43* and expression pattern of *srx-44*.

Besides these specific examples, resequencing studies also support the importance of balancing selection in *C. elegans* populations. Besides the laboratory strain N2, the Hawaii strain CB4856 is another widely used strain because it is one of the wild strains that have most genetic differences compared with N2. A draft genome of the CB4856 strain was created by a combination of short-read and long-read sequencing. Comparison between the N2 genome and CB4856 genome has identified highly divergent regions that are enriched for F-box, MATH, BATH, seven-transmembrane G-coupled receptors, and nuclear hormone receptors (Thompson et al., 2015). A comprehensive genome-wide comparison of all 609 wild *C. elegans* strains isolated worldwide also showed that genetic variations are enriched in punctuated hyper-divergent regions. These hyper-divergent regions only cover 20% of the genome while including most of the genetic variation in *C. elegans* (Lee et al., 2020). Also, these regions show strong signature of balancing selection. Seven-transmembrane receptor class genes such as G-protein coupled receptors are enriched in the hyper-divergent regions, implying ancient genetic variation for environment sensing and responding are actively maintained in the population.

While there are many benefits to use *C. elegans* to study speciation, there are a few obstacles for us to fully understand its evolutionary history. One major challenge is that we don't have a lot of knowledge about its ecology. *C. elegans* are mostly found in anthropogenic habitats where recent human migration is believed to contribute to the signatures of selective sweeps in the genome. The natural habitat for *C. elegans* is not well known. Extensive sampling in Hawaii islands has discovered the most genetic diversity in this species, raising a promising hypothesis that *C. elegans* are originated from the Hawaii islands. Another challenge is a lack of quantitative methods to measure the fitness of a allele in a specific environment and predict its evolutionary dynamics considering the idiosyncratic features of the special androdioecy reproductive system.

1.4 Aims

In this thesis, I seek to advance the quantitative methods in the study of species evolution by looking into two systems. The first is Lake Malawi cichlid fishes and the second is *C. elegans* (*Caenorhabditis elegans*). In cichlids, species show different social behaviors (York et al., 2015). Researching the genetic basis requires a quantitative comparison between the different behaviors. However, this is not easy because these natural behaviors are expressed over very long timescales and they only happen in a familiar environment to the animals. In *C. elegans*, long-term evolutionary forces like balancing selection are able to maintain variations for many generations. However, balancing selection for selfish genetic elements contradicts our expectation for genic drive. To solve the puzzle, quantitative modeling and competition assays to directly measure the fitness effect of these selfish genetic elements are necessary. Finally, the conflict between selfish genetic elements and the rest of the genome is an important factor leading to speciation (Austin et al., 2009). The dynamics of selfish genetic elements are often studied while ignoring the dynamics of the rest of the genome. Thus, sophisticated simulations are necessary to understand the dynamics of the whole genome. In this thesis, I want to directly address these challenges.

CHAPTER 2. AUTOMATIC CLASSIFICATION OF CICHLID BEHAVIORS USING 3D CONVOLUTIONAL RESIDUAL NETWORKS

2.1 Introduction

Animals respond to and interact with their environment using a rich repertoire of behavioral actions. A major challenge for neuroscientists is to understand how neural circuits coordinate these behaviors in response to sensory and internal stimuli. Automated identification and classification of behavioral actions will aid in this task, as most neural responses are stochastic, requiring a large number of replicates to accurately estimate relationships with complex behaviors (Egnor and Branson, 2016). Additionally, many behaviors are executed over long timescales through the accumulated actions of thousands of individual decisions (e.g. foraging, construction, and social behaviors), making manual analysis of the full course of behavior entirely impractical (Russell et al., 2017, Mouritsen, 2018, Tucker, 1981). One common approach is to manually observe snapshots of long-term behaviors over extended periods of time. This method is labor intensive and thus can severely limit the total number of animals that can be measured in a single study. Furthermore, it cannot provide a complete and detailed quantitative description of the full behavioral trajectory. An alternative approach is to design abbreviated assays that elicit the behavior of interest during a short period of time. One issue, however, is that many natural behaviors may be expressed differently over short timescales, and/or in unnatural/unfamiliar environments. Recently, deep learning approaches have revolutionized our ability to automatically analyze video and image data (Nath et al., 2019, Mathis et al., 2018, Weissbrod et al., 2013, Wild et al., 2018). Convolutional Neural Networks (CNNs) can be applied to images for the purpose of object detection, identifying all animals within an individual frame (Girshick, 2015, Ren et al., 2015, Redmon et al., 2016, Pereira et al., 2019). CNNs can also identify body parts, such as eyes, legs, or wings, allowing for the determination of an animal's posture at any specific time (Graving et al., 2019, Pereira et al., 2019, Kain et al., 2013, Petrou and Webb, 2012, Gunel et al., 2019, Andriluka et al., 2018, Nath et al., 2019). Although position and pose alone are not sufficient for defining animal behavior, postural time series can be used to define some behavioral actions. Such analysis has, for example, been used to quantitatively describe different types of stereotyped movements in Drosophila flies, such as different locomotor and grooming behaviors (Berman et al., 2014). Similar approaches have also been used successfully on other species (Stephens et al., 2008). It remains challenging, however, to translate changes in position and posture into complex behaviors that are characterized by an animal's interaction with the environment. For example, many goal-directed behaviors involve significant manipulation of the physical environment in ways that are essential for survival or reproductive success, such as mice digging burrows, birds building nests, spiders weaving webs, and bowerbirds or cichlid fishes constructing courtship bowers (Hansell, 2000, Benjamin and Zschokke, 2000, Vollrath, 1992, Collias and Collias, 2014, Dawkins, 1982, MeKaye, 2001). In such cases, information about changes to the physical environment itself is essential to fully describe the behavior.

One possible solution for analysis of these types of behaviors is to train a deep learning network that takes videos in as input, and then outputs a prediction for the corresponding behavior type. For example, 3D Residual Networks (3D ResNets) have been successfully used to classify human behaviors, distinguishing between hundreds of different action classes (Hara et al., 2018, Qiu et al., 2017). These deep-learning networks use 3D kernels with the ability to extract spatiotemporal features directly from videos. Videos are fed into these networks raw, without any individual body posture information beyond what can be learned from the training data. One major benefit of these 3D networks, when compared to networks that process each frame individually, is that they can integrate spatial and temporal information to recognize changes in the animal's environment that might indicate a particular behavior (e.g. digging, feeding, or construction behaviors). However, a significant challenge in applying action recognition to large videos is detecting actions of interest in a time and space frame, which is known as shot transition detection. This requires splitting large (up to 10 hours long) videos into small enough temporal units such that each unit contains only one action of interest and excludes as much irrelevant information as possible.

In this paper, we describe an approach for analyzing construction, feeding, and mating behaviors from hundreds of hours of videos of Lake Malawi cichlids behaving freely in naturalistic and social home tank environments. Lake Malawi is the most species-rich freshwater lake on Earth, and it contains 700-1,000 cichlid species that have rapidly evolved in the past 1-2 million years (Brawand et al., 2014). Approximately 200 of these species express long-term bower construction behaviors, in which males manipulate sand to create large courtship structures, or bowers, to attract female mates (York et al., 2015) (Figure 2.1A). Males construct pits and castles over the course of many days and make thousands of decisions about where to scoop up and then spit out mouthfuls of sand. Bower construction is therefore a useful model for understanding how goal-directed behaviors are executed over long time periods in dynamic environments.

In order to measure bower construction, we first develop an action detection algorithm that uses Hidden Markov Models (HMMs) and density-based spatial clustering to identify regions of the video where the fish has manipulated sand using its mouth, fins, or other parts of its body. Then, after generating small video clips that encompass these events, we use a 3D ResNet to classify each sand disturbance event into one of ten action categories. We demonstrate that this approach can be used to quickly, accurately, and automatically identify hundreds of thousands of behaviors across hundreds of hours of video. Through this approach, we measure the times and locations of construction, mating, and feeding behaviors expressed over the course of many days in three different species and one hybrid cross.

2.2 Results

2.2.1 Collection of video recordings of male cichlid fish constructing bowers

If provided access to an appropriate type of sand and gravid females, Lake Malawi male cichlids will typically construct species-typical bower structures in standard aquarium tanks within 1-2 weeks (York et al., 2015) (Figure 2.1B). After testing sand from different suppliers, we found that males construct most vigorously with a sand mixture composed of black and white grains. Only half of each home tank was accessible for top-down video recording, due to physical constraints such as support beams and water supply/drain lines. In order to restrict sand manipulation behaviors to the video-accessible region of each tank, we placed a custom-built acrylic tray containing sand directly within the video camera field of view. We then introduced a single male individual to four female individuals (Figure 2.1C). For each trial, we collected ten hours of video daily for approximately ten consecutive days, resulting in the collection of ~100 hours of video per trial. We analyzed two types of trials: construction trials, in which males constructed bowers in the presence of four females, and control trials, in which tanks were empty, or tanks contained four females but no male. We analyzed eight construction trials, encompassing three different species and one F1 hybrid cross: pit-digger Copadichromis virginalis, CV, n=1; castle-builder Mchenga conophoros, MC, n=3; pit-digger Tramitichromis intermedius, TI, n=2; and pit-castle hybrid MCxCV F1, n=2. Visual inspection confirmed that each male built a species-typical bower. We also analyzed three empty tank control trials, and five female only control trials. Table 2.1 summarizes all trials that were analyzed in this paper.

To initially analyze these videos, we first focused on a single MC trial, identifying three days when castle building occurred. A trained observer manually annotated all spit and scoop events for a single hour on each day. In total, 1,104 spit events and 1,575 scoop events were observed (there is not a 1:1 correspondence between spit and scoop events because males sometimes scoop sand from multiple locations before performing a single spit) (Figure 2.2). Analysis of these three hours of videos required 36 hours of human time, or a 12:1 ratio of human time to video time. Full analysis of a single trial would require 1,200 hours of human time, or approximately 30 weeks of full-time work dedicated to manual annotation. Automated analysis of these videos was thus necessary for the full characterization of bower construction even in this limited set of trials, let alone any larger scale investigation of bower behavior.

2.2.2 Automatic identification of sand disturbance events

In the process of manually annotating these videos, we noticed that spit and scoop events resulted in an enduring and visually apparent spatial rearrangement of the black and white grains of sand. Scoop and spit events were associated with a permanent transition between pixel values. We tested whether events could be identified despite the frequent occurrences of fish swimming over by inspecting local regions in which a large number of pixels underwent permanent transitions. We observed temporary changes in pixel values when fish or shadows occluded the sand and permanent changes in pixel values at the locations of sand disturbance events (Figure 2.3A-F). Consistent with this, we plotted the greyscale value of every pixel over entire 10-hour videos and found that individual pixel values were, in general, fixed around a specific mean value for extended periods of time, but showed small oscillations and large but temporary deviations about this value across video frames (Figure 2.3G, H). In addition, we identified permanent transitions in pixel value, in which the mean grayscale value of a single pixel would change to, and retain, a new value.

In order to automatically identify permanent transitions in pixel color, we used an HMM to calculate a hidden state for each pixel that represented its current color. In general, the HMM partitioned each pixel into a small set of values (~10-50) over the course of each 10 hour video (Figure 2.3G, H). In testing the speed and accuracy of the HMM on raw data, we observed that large temporary deviations greatly impacted the accuracy of the HMM calls, potentially due to their violation of the assumption of a Gaussian distribution. Therefore, these large deviations were excluded using a rolling mean filter. We also found a speed/resolution tradeoff to the frame rate. Since we wanted to perform this analysis for the entire 1296x972 video, we needed to analyze over one million pixels through time. We found that one frame per second, sampled from the 30 frames per second in the raw video, was a reasonable tradeoff – an entire video could be analyzed in approximately two hours on a 24-core machine. After setting appropriate parameters, manual inspection of the HMM fits showed reasonable agreement with our expectations: enduring transitions in pixel values were associated with sand disturbance by fish, and three orders of magnitude more of these transitions were detected in tanks with fish versus empty tanks (Figure 2.3I). In addition to identifying when and where a sand disturbance occurs, this approach also allowed us to create background images that excluded fish and shadows from each frame (Figure 2.3J).

2.2.3 Clustering of HMM changes to detect sand disturbance events

When a fish scoops up or spits out sand, HMM transitions occur in hundreds to thousands of nearby pixels. In order to group individual pixel transitions together, we applied density-based spatial clustering to the x, y, and time coordinate of each individual HMM transition (Ester et al., 1996) (Figure 2.4A). To find the best parameters, we used exploratory data analysis and parameter grid search (Figure 2.5). This approach allowed us to determine the spatial and temporal location of thousands of individual fish-mediated sand disturbances on each day of each bower trial. We also calculated distributions for the width, height, temporal span, and number of HMM transitions for every cluster (Figure 2.6). These data further demonstrated that clusters were associated with the presence of fish, as empty tanks showed a minimal number of clusters (Figure 2.4B). Inspection of a subset of these clusters confirmed that the clusters were associated with regions of sand that were undergoing lasting change (Figure 2.4C).

To further validate these clusters, we generated 200x200 pixel, four second video clips for a randomly sampled subset of clusters centered over their mean spatial and temporal position (~2,000 per trial for seven of the eight bower construction trials). Manual review of these video clips revealed that the majority (>90%, 13,288/14,234 analyzed events) of clusters were true sand change events caused by fish behaviors, with the remaining portion including reflections of events in the glass, shadows caused by stationary or slow-moving fish, or small bits of food, feces, or other debris settling on the sand surface.

To ensure this procedure recovered most sand disturbance events, we leveraged our manually annotated spit/scoop dataset (Figure 2.2) and performed HMM and density-based spatial clustering. For each event, the spatial and temporal center of the event was compared. The differences in time between human annotations and machine annotations follow a gaussian distribution (Figure 2.7A). 95.6% (725/758) human annotations have at least one machine annotation in the (-1s, +1s) interval. Of these 725 events, 93.7% (679/725) also had a machine annotation event within 70 pixels (approximately 3.5 cm), which is ~ 1/3 of the oval fish length (Figure 2.7B). By these criteria, 89.6% (679/758) of human annotated events can be retrieved by the automated machine HMM/clustering process.

2.2.4 Automatic classification of cichlid behaviors with 3D residual networks

On average, ~1,000 clusters were identified in each hour of video (Figure 2.4B). We aimed to automatically identify the subset of these events corresponding to bower construction behaviors in each video. However, scooping and spitting sand during bower construction represents only a subset of behaviors that cause sand change in our paradigm. For example, feeding behaviors performed by both males and females also involve scooping and spitting sand, and are expressed frequently throughout trials. Quivering and spawning behavior, in which a male rapidly circles and displays his fins for a gravid female, are less frequent but also cause significant sand change. Identifying bower construction behaviors therefore requires identifying other behaviors that cause sand change. To achieve this, we turned to a deep learning approach and assessed whether 3D Residual Networks (3D ResNets), which have been recently shown to accurately classify human actions from video data (Qiu et al., 2017), could accurately distinguish fish behaviors that cause sand change in our paradigm.

To create a training set for the 3D ResNet, we first generated a video clip for each cluster based on its location in space and time. This process narrowed down each event to a four second, 200x200 pixel video clip from the original 1080p video. To create a training set for the 3D ResNet, a trained observer manually classified 14,172 of these video clips (~2,000 per trial) into one of ten categories (bower scoop, bower spit, bower multiple, feed scoop, feed spit, feed multiple, drop sand, quivering, other, and shadow/reflection). Feeding was the most frequently observed behavior, accounting for nearly half of all clips (47.0%, 6,659/14,172 annotated clips; feeding scoops, 15.2%; feeding spits, 11.6%; multiple feeding events, 20.2%). Bower construction behaviors were the next most common (19.5%; bower scoops, 9.4%; bower spits, 8.1%; multiple bower construction events, 1.9%). Quivering and spawning events were the least frequently observed, accounting for just 2.6% of all clips. The remainder of sand change events were annotated as either sand dropping behavior (5.6%), “other” behaviors (e.g. brushing the sand surface with the fins or the body; 18.7%), or shadows/reflections (6.6%) (Figure 2.8).

A 3D ResNet was then trained on 80% (~11,200 clips) of the data, and the remaining 20% of the data was reserved for validation (2,752 clips) (Table 2.2 and Figure 2.9A, B). To place the ResNet predictions in the context of human performance, we also measured the accuracy of a previously naive human observer that underwent 12 hours of training and then manually annotated a test set of 3,039 clips from three trials and all ten behavior categories (Table 2.3). The 3D ResNet achieved ~76% accuracy on the validation set, which was better than the accuracy of the newly trained human observer (~73.9% accuracy, 2,246/3,039 clips). Confidence for 3D ResNet predictions on the validation set ranged from 22.1-100%, and confidence tended to be greater for correct predictions (mean confidence 92.93±0.279%) than for incorrect predictions (mean confidence 78.28±0.074%) (Figure 2.10A, C). We found an imbalance in the distribution of incorrect predictions across categories (Table 2.2). For some categories, such as "build multiple", "feed multiple", and "fish other", video clips could contain behaviors that also fit into other categories. For example, a "feed multiple" clip by definition contains multiple feeding scoop and/or feeding spit events, a "bower multiple" clip contains multiple bower scoops and/or bower spits, and a "fish other" clip may contain a bower scoop and a fin swipe (or some other combination of behaviors). We found that erroneous "within building" category predictions for build multiple, "within feeding" predictions for feed multiple, and "fish other" predictions accounted for ~82% of all incorrect predictions. We further found that the area under the Precision Recall Curve was 0.91 (Figure 2.10B). By setting a confidence threshold of 90%, most (~62%) incorrect predictions were excluded while most (70%) correct predictions were included. At the same time, ~86% of correct bower scoop predictions and ~88% of correct bower spit predictions were included. 69% of all predictions were above the 90% confidence threshold, and overall accuracy for these high-confidence predictions was ~87% (Figure 2.10C).

We tested how many annotated clips are required to achieve a similar accuracy by inputting 5-50% of the original labeled clips into the training dataset. A similar accuracy was achieved using just 50% of the clips, but a significant decrease in accuracy (75% to 62%) was observed when the number of training clips was decreased from 50% to 30% of the original set (Figure 2.9B).

Since there was an uneven distribution of the different sand perturbation events, we were curious if our model might be biased towards the more frequent events and show a lower accuracy than a training set with equal distribution of different events. To test this hypothesis, we built a model from an equal sampling of all ten categories. First, we created an ‘equal sampling’ set where each category in the training dataset has the same number of clips. Because the category ‘build multiple’ has the least number of clips (277 clips out of 14172 annotated video clips), for each category, we used 220 clips for training and 50 clips for validation. As a ‘random sampling’ control, we randomly sampled an equal number of clips for training and validation as the uniform set. Specifically, there are 2200 randomly sampled clips for training and 500 for testing. Overall, the accuracy was comparable, although we did note some large differences in accuracy based upon category (Figure 2.10F). For example, build multiple, a very rare category, was much less accurate in the randomly sampled set (18% vs. 75%). When we compared the overall accuracy of each model, we noted that the higher performing model depended on the validation set. When an equal sampling of validation clips was used, the ‘equal sampling’ model performed better (64.0% to 60.5%). When a random sampling of validation clips was used, the ‘random sampling’ method was more accurate (62.2% to 57.7%).

The spatiotemporal dimension of each video clip is a very important hyperparameter to tune to achieve the best classification accuracy. The spatial scale is optimal when the whole behavior is captured, and when irrelevant pixels are excluded from the frame as much as possible. The same is true for temporal scale. Therefore, we decided to test the amount of spatial and temporal data necessary to accurately classify each clip. We found that the classification accuracy initially increased with frame size and peaked at 120x120 pixels per frame (Figure 2.10D). The accuracy decreased from 76% to 74% when the entire 200x200 pixels frame was used as input. This decrease in accuracy further affirmed our choice to crop the clips before feeding them to the ResNet, as it eliminates information unrelated to action in the frame. Next, we tested the relationship between the number of frames per clip and the classification accuracy. We tuned two parameters, including the total number of frames sampled (n=16, 32, 48) and the interval between each sampled frame (i=0, 1, 2, 3, 4). Classification accuracy steadily increased from 67% (n=16) to 72% (n=48) when more frames were used. When the total number of frames was fixed, accuracy was greater when the sampling interval (i) was larger (Figure 2.10E).

2.2.5 Distinguishing between feeding and bower construction events

The differences between construction behaviors and feeding behaviors are subtle and are often indistinguishable to inexperienced observers (Table 2.3). Feeding and construction both involve scooping sand into the mouth, swimming, and then spitting sand from the mouth. Feeding behaviors are also typically performed more frequently relative to bower construction (Figure 2.8). We were therefore concerned that the 3D ResNet would be unable to accurately distinguish between feeding versus construction behaviors, which would prevent accurate measurement of spatial patterns associated with bower construction.

Much to our surprise, the ResNet reliably distinguished between feeding and construction events; for construction and feeding scoops and spits, the model achieved F1 scores of 0.74, 0.86, 0.72, and 0.73 (for build scoops, build spits, feeding scoop, and feeding spits, respectively), with balanced precision and recall scores. These F1 scores were comparable to the overall averaged F1 score (or accuracy) of the model, 0.76. Remarkably, the model outperformed a newly trained human observer in distinguishing between construction and feeding: the average F1 score for the model across these categories was 0.76, while the newly trained observer's average F1 score was 0.71. The difference in performance was further evident when we quantified the proportion of build-feed false positives (build scoops mis-classified as feed scoops, build spits mis-classified as feed spits, and vice versa); these erroneous predictions were 2.5x more frequent in the newly trained observer's annotations compared to the model's (201/1,469 predictions in these categories, or 13.7% for the human; 68/1,211 predictions in these categories, or 5.6% for the model). Differences in the spatial position and relative timing were readily observed among bower construction, feeding, and spawning events (Figure 2.9C, D).

2.2.6 Accuracy of model on new trials

The trained model showed high validation accuracy on the seven trials. However, our ultimate goal is to measure bower construction in hundreds of independent trials. To that end, we tested how generalizable this model is when applied to previously unobserved individuals. To test this, we retrained models using six of the seven trials and tested their ability to classify video clips from the remaining trial. In all seven cases, we saw a significant drop in accuracy, from 76% to between 48% and 62% depending on the trial (Figure 2.11). We could, however, recapture much of this accuracy loss by including a small subset of videos from the excluded trial (~100-400), suggesting that including a limited number of labeled videos from new trials could dramatically increase the accuracy. Interestingly, trials that had the largest imbalance in the frequency of different categories compared to the mean also showed the largest decrease in accuracy (Figure 2.12).

Figure 2.1 Measurement of Lake Malawi cichlid bower behaviors in laboratory.

Approximately 200 species of Lake Malawi cichlids exhibit bower behaviors. In these species, sociosexual cues trigger reproductive adult males to construct large courtship structures by manipulating sand with their mouths. The geometric structure of the bower is species-specific. A. Example of a castle structure built in Lake Malawi. B. Example of a castle structure built in a standard aquatics facility aquarium. C. Top down view of acrylic tray used to constrain bower building to a third of the tank. Video recordings using this view were used to characterize bower building behaviors throughout this paper. Scale bar is 10cm. Photo credit to Dr. Ad Konings (A).

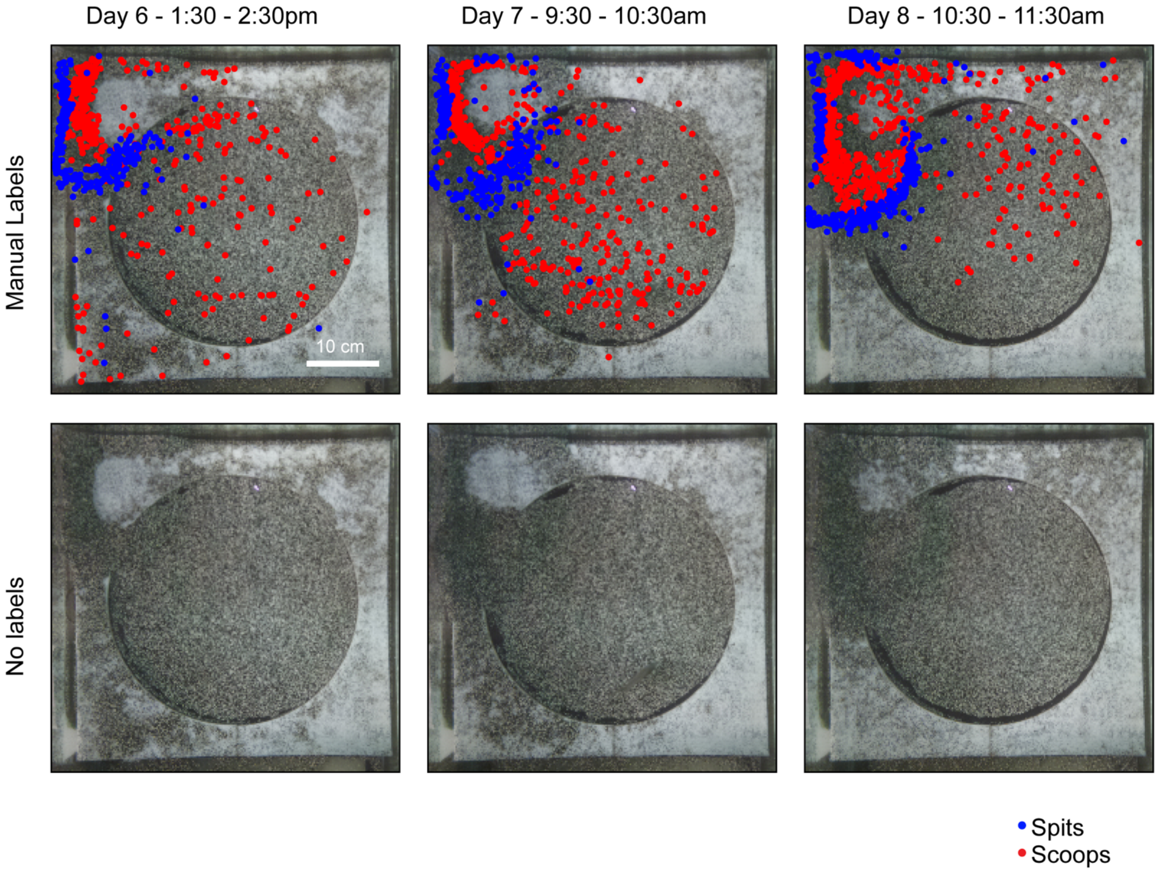


Figure 2.2 Example images on three different days during bower building by a Mchenga conophoros male.

Top panel shows a top down view with all building scoops (red) and building spits (blue) during a one hour period identified by manual inspection. Bottom panel shows same images without scoops and spit events. A castle bower is being built on the middle left portion of the tray. Scale bar is 10cm.

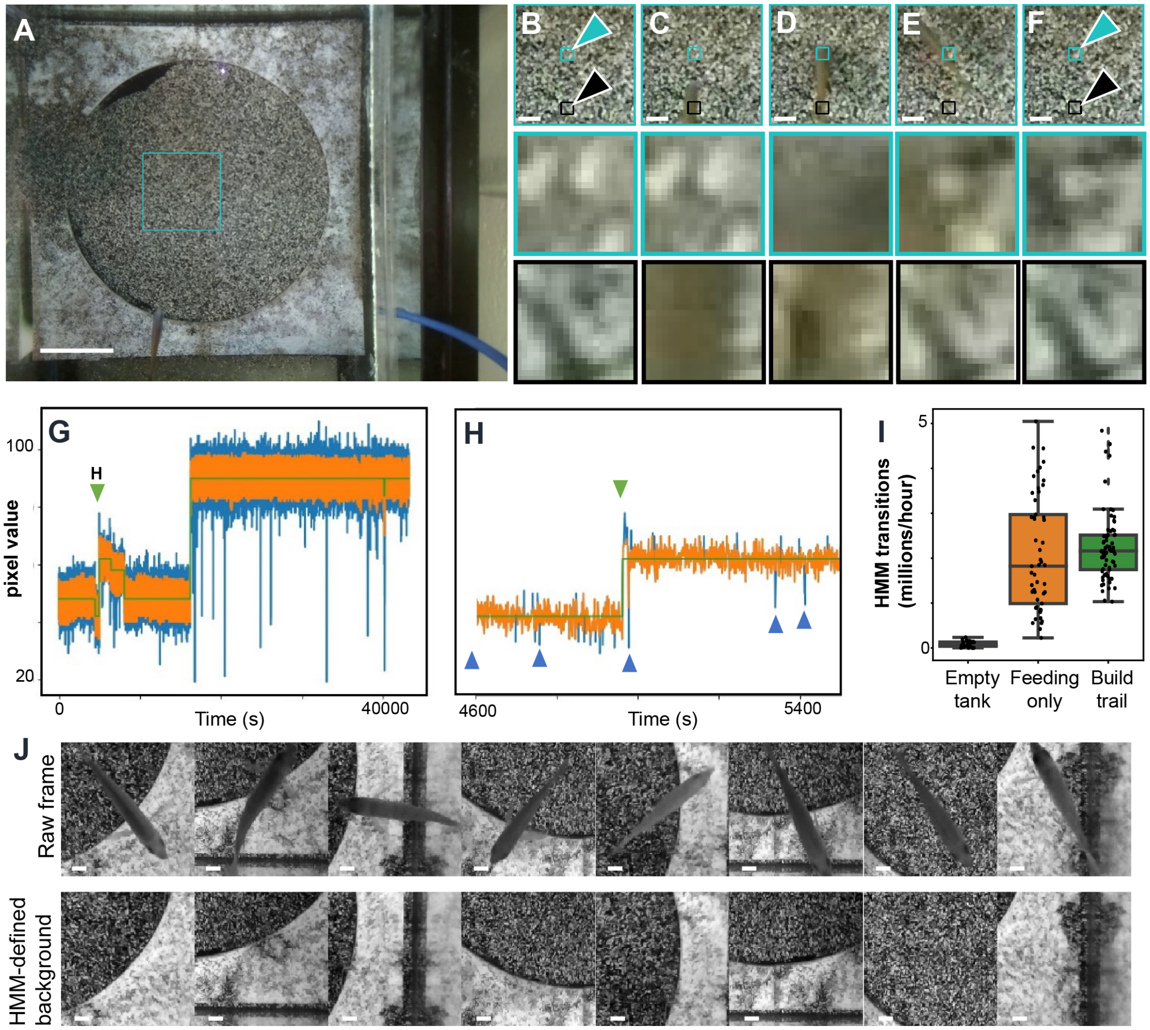


Figure 2.3 Automated detection of sand change from video data.

The sand in this behavioral paradigm is composed of black and white grains (e.g. as seen in panel **A**), and therefore sand manipulation events during bower construction cause permanent rearrangement of the black and white grains at specific locations. We aimed to detect these events by processing whole video frames (**A**, with turquoise box indicating an example region of interest. Scale bar is 10cm) sampled once per second, and tracking the values of individual pixels throughout whole trials. Fish swimming over sand cause transient changes in pixel values (e.g. **B-F**, black arrows indicate an example location of a fish swimming over the sand; the bottom row depicts a zoomed in 20x20 pixel view of a location that the fish swims over, sampled from representative frames across four seconds). In contrast, sand manipulation behaviors cause enduring changes in pixel values (e.g. **B-F**, turquoise arrow indicate an example location of a fish scooping sand. Scale bar is 2cm; the middle row contains a zoomed in 20x20 pixel view of a location where the fish scoops sand). We used a custom Hidden Markov Model to identify all enduring state changes for each pixel throughout entire videos (**G**, green line indicates HMM-predicted state, orange line indicates raw grayscale pixel value, and blue lines indicate transient fluctuations beyond the pixel's typical range of values likely caused by fish swimming or shadows). Because fish swim over the sand frequently, a large number of transient changes are ignored (e.g. pixel value fluctuations indicated by blue arrows in panel **H**), while enduring changes are identified (e.g. pixel value change indicated by green arrow in Panel **H**). **I**. Number of HMM transitions identified per hour based upon trial type. "Feeding only" are trials containing four females. "Build trial" contain four females and one male that builds a bower. The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. **J**. The HMM could be used to calculate a background image at a given timepoint, resulting in removal of the fish and the associated shadow from the image. Scale bar is 1cm.

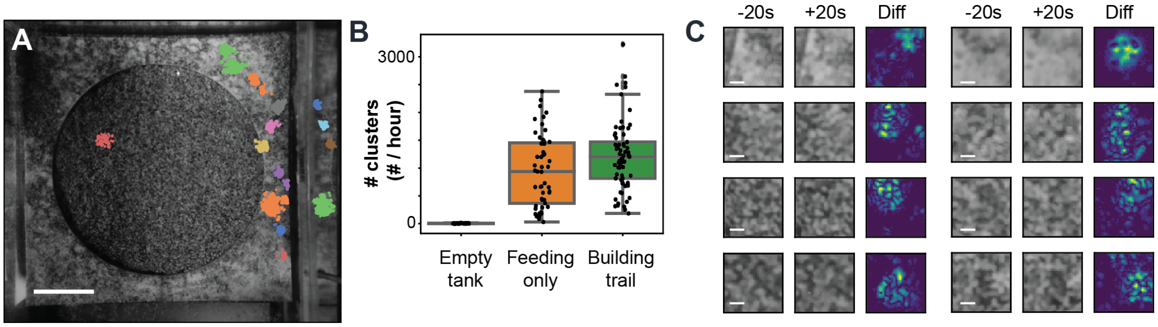
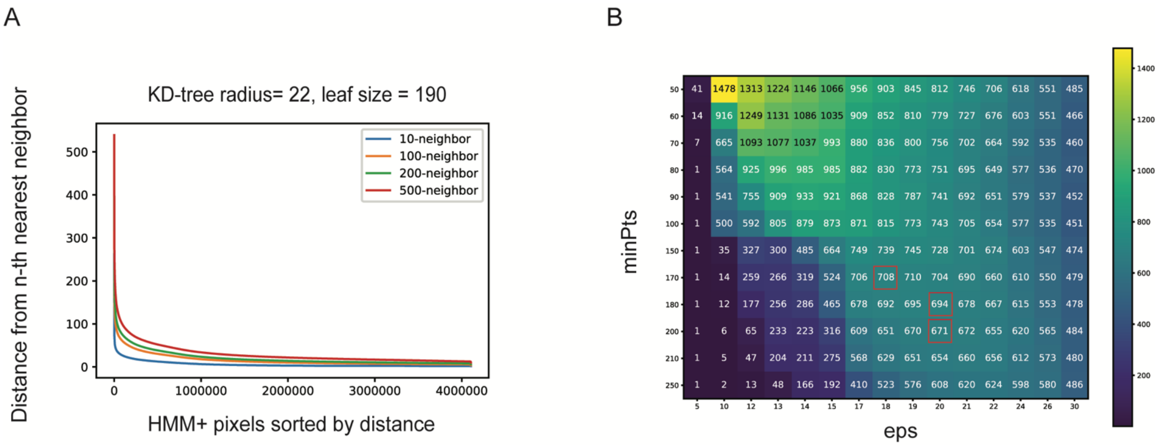


Figure 2.4 Clustering HMM events identifies sand disturbance events.

A. Example of clusters identified in a 60 second period. HMM transitions are color coded based upon their cluster membership. Scale bar is 10cm. B. Number of clusters identified per hour based upon trial type. “Feeding only” are trials containing four females. “Build trial” contain four females and one male that builds a bower. The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. C. Before and after images (20 seconds) for eight example clusters. Left and middle panel show raw grayscale images. Right panel shows heatmap displaying pixel value differences in the left and middle frame. Yellow indicates large changes. Blue indicates no change. Scale bar is 1cm.



	5	10	12	13	14	15	17	18	19	20	21	22	24	26	30
50	41	1478	1313	1224	1146	1066	956	903	845	812	746	706	618	551	485
60	14	916	1249	1131	1086	1035	909	852	810	779	727	676	603	551	466
70	7	665	1093	1077	1037	993	880	836	800	756	702	664	592	535	460
80	1	564	925	996	985	985	882	830	773	751	695	649	577	536	470
90	1	541	755	909	933	921	868	828	787	741	692	651	579	537	452
100	1	500	592	805	879	873	871	815	773	743	705	654	577	535	451
150	1	35	327	300	485	664	749	739	745	728	701	674	603	547	474
170	1	14	259	266	319	524	706	708	710	704	690	660	610	550	479
180	1	12	177	256	286	465	678	692	695	694	678	667	615	553	478
200	1	6	65	233	223	316	609	651	670	671	672	655	620	565	484
210	1	5	47	204	211	275	568	629	651	654	660	656	612	573	480
250	1	2	13	48	166	192	410	523	576	608	620	624	598	580	486

Figure 2.5 HMM pixel distance exploratory analysis and parameter search for DBSCAN

A. HMM+ pixels sorted by distance from n-th nearest neighbor. This plot was used to visualize the distribution of n-th nearest neighbor distances across HMM+ pixels. The knee point of the k-dist graph was used to estimate the optimal values for parameter eps to be 20-30. B. Number of identified clusters under different values for minPts and eps. This plot shows the number of identified clusters from segment of video data using different values for minPts and eps. Red boxes indicate values at which trained observers reviewed video clips of sand change clusters to identify optimal values for minPts and eps. See methods for further details.

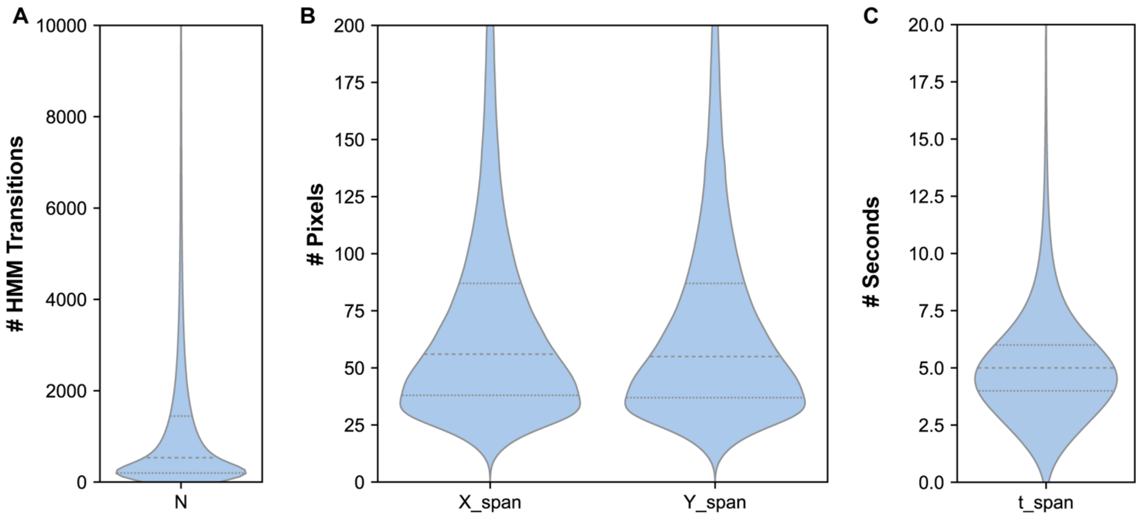


Figure 2.6 Violin plots showing distribution of cluster features.

Dashed lines within the violins indicate the 1st quartile, median, and 3rd quartile values. A. Distribution of cluster sizes, using number of HMM transitions assigned to each cluster. B. Distribution of width and heights of each cluster. C. Distribution of time-length of each cluster. Since HMM transitions were only calculated at one second intervals, the number time-length must be an integer number.

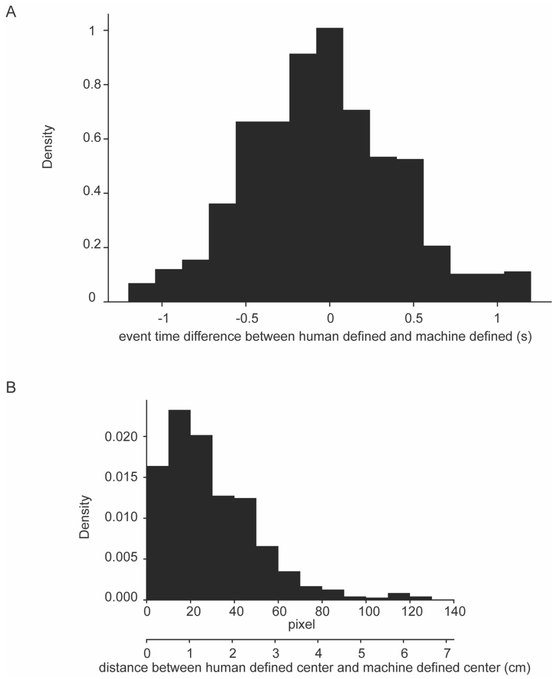


Figure 2.7 Spatial and temporal difference between human defined center and machine defined center.

A. The distribution of time differences between human defined center and machine defined center. B. The distribution of distance between human defined center and machine defined center.

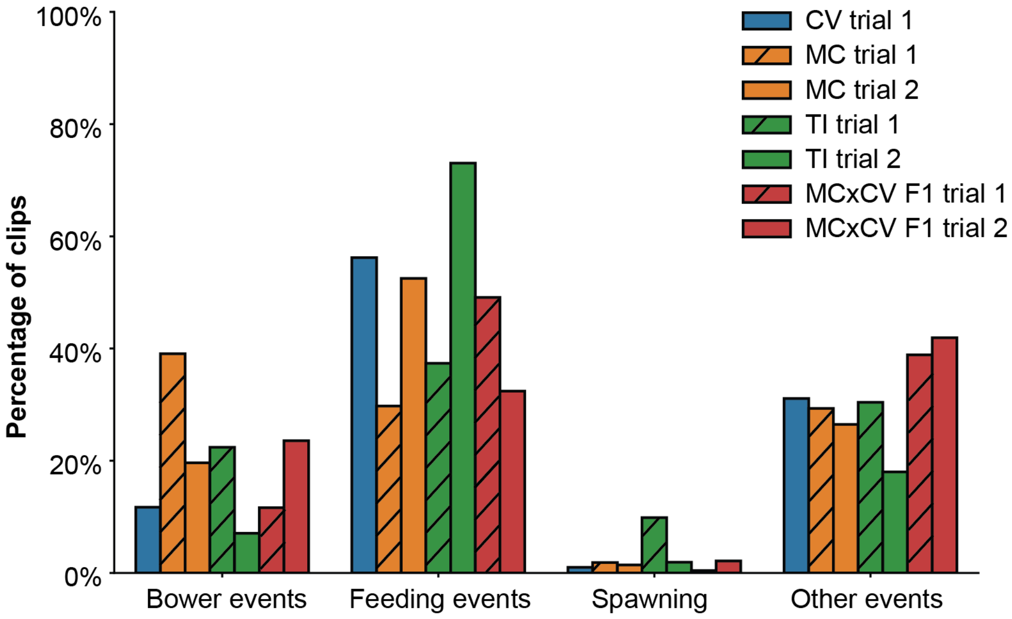


Figure 2.8 Distribution of different sand perturbation events.

A human observer manually classified 14,234 video clips into one of ten categories. Bower events are scoops, spits, or multiple scoop/spit events associated with bower construction. Feeding events are scoops, spits, or multiple scoop/spit events associated with feeding behaviors. Spawning events involve male fish quivering to attract females. Other events include sand perturbations caused by fins or the body, reflections of events in the aquarium glass, or sand dropped from above.

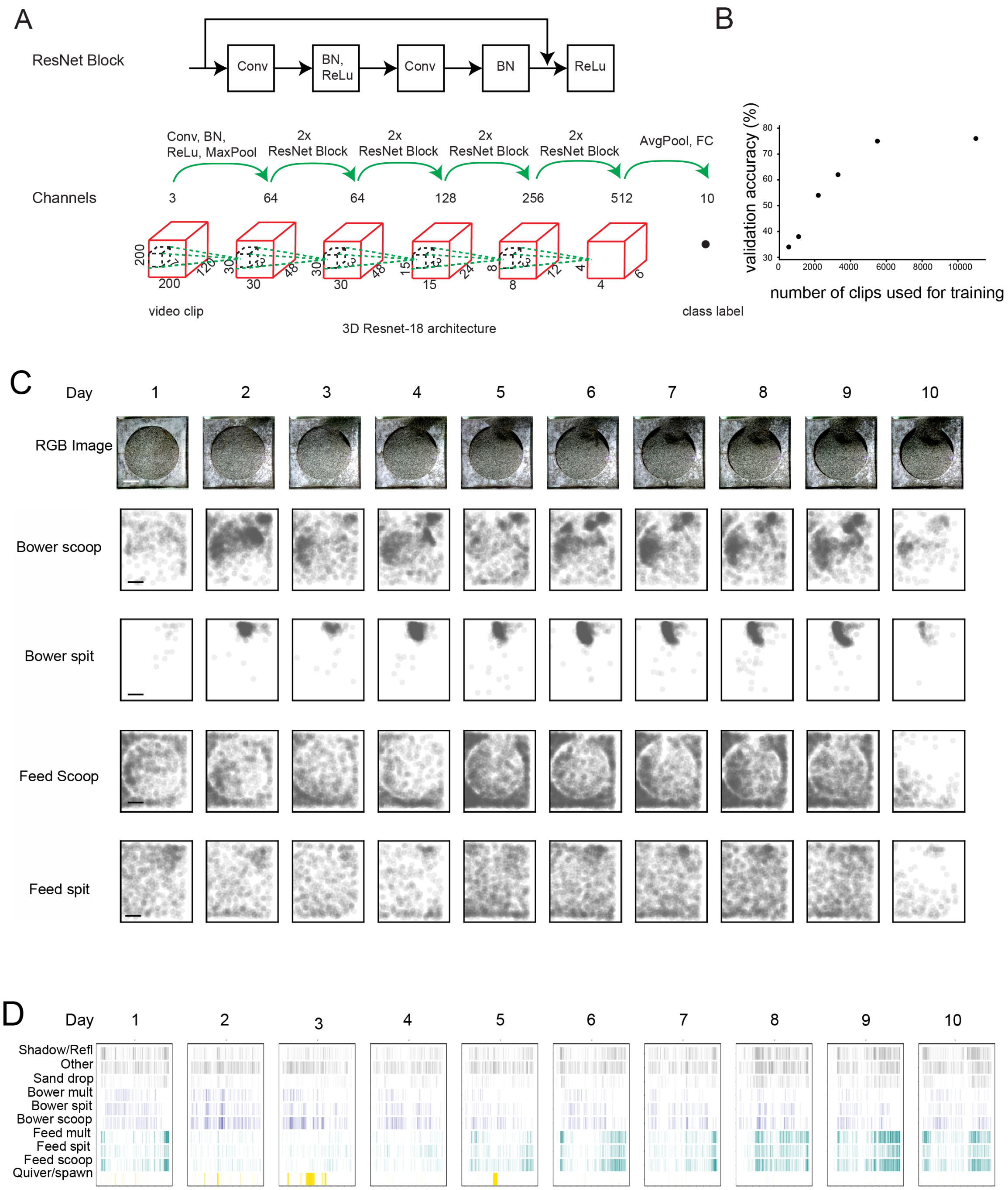


Figure 2.9 Automated classification of sand disturbance events using a 3D residual network.

A. Schematic of 3D residual network. B. Validation accuracy of machine-learning labels as a function of number of videos used to train the model. C. Spatial position of four categories of sand manipulation events over ten days of building by a Mchenga conophoros. A castle structure was built in the top middle of the field-of-view. Scale bar is 10cm. D. Raster plot of time of each sand disturbance event by classification. C and D show that building events occur at different spatial and temporal positions than feeding events.

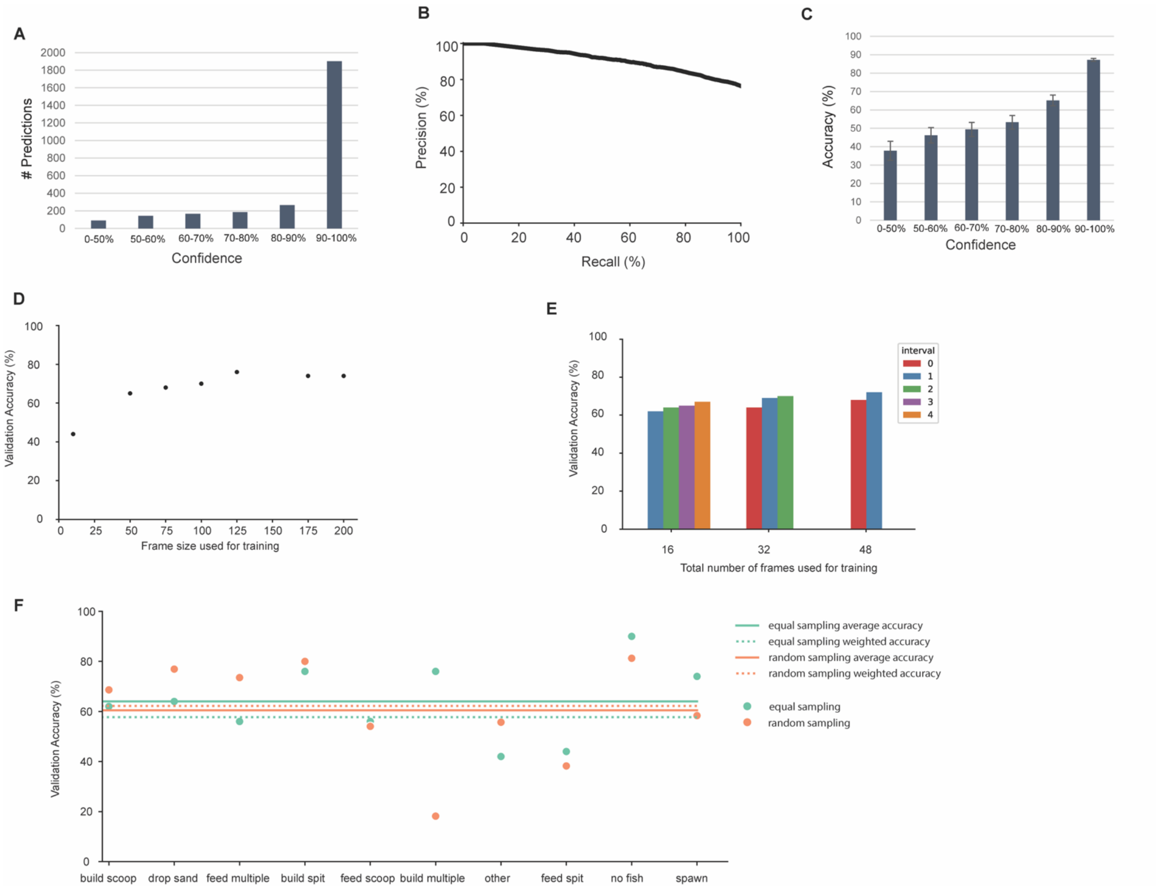


Figure 2.10 Ablation studies and performance analysis for the 3D ResNet.

A. Number of predictions in validation set by confidence. B. Precision Recall Curve for the action classifier. C Accuracy (mean ± standard error) of predictions grouped by confidence scores. D. Model validation accuracy by frame size used for training. E. Model validation accuracy by total number of frames used for training. F. Per category accuracy by data sampling method.

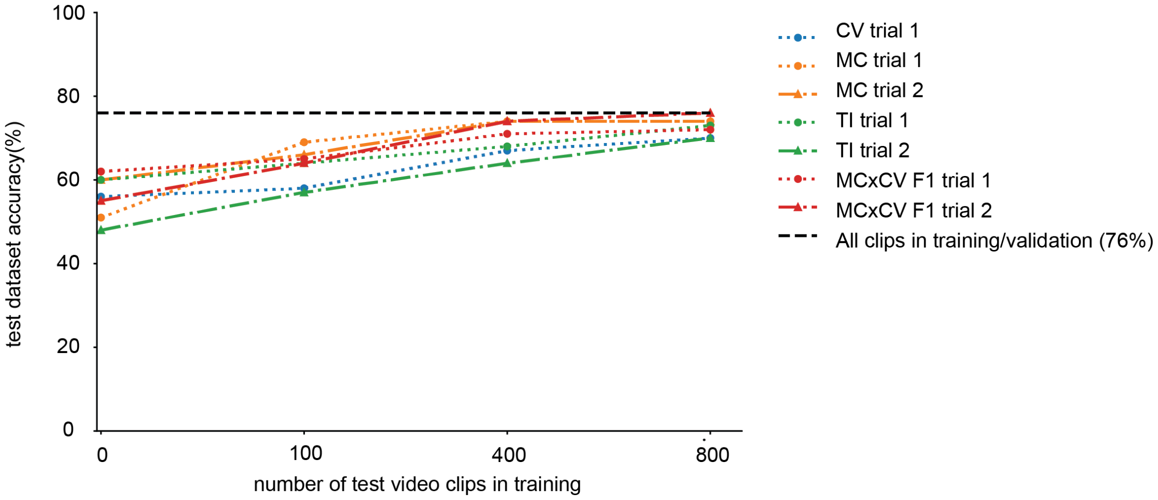


Figure 2.11 Application of the model on new trials.

Accuracy of model to trials using 0, 100, 400, or 800 training videos. We restricted training data for one of the seven and then tested the accuracy of the model on video clips specific to that trial. Models that used zero video clips for training showed a decrease in accuracy.

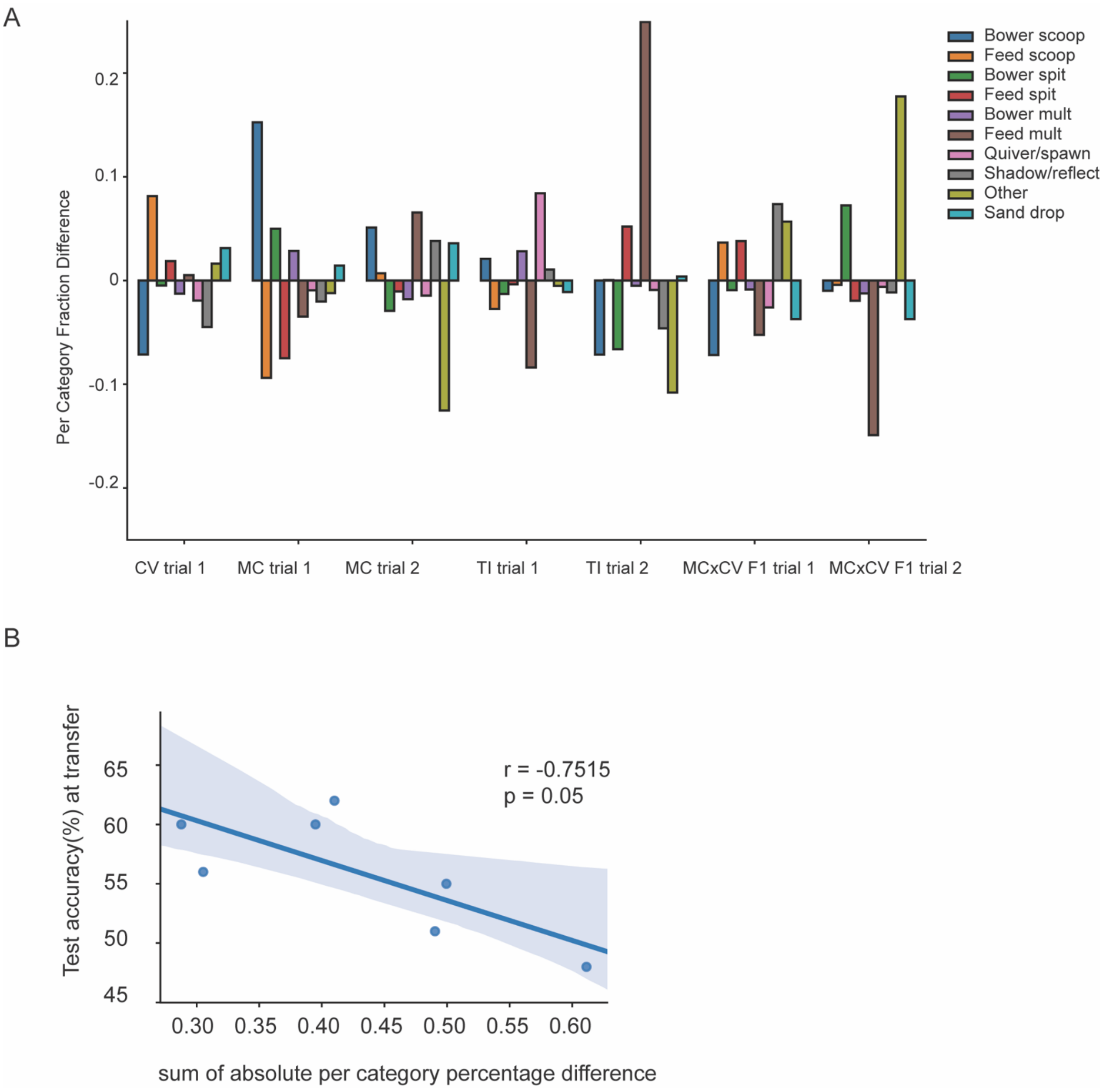


Figure 2.12 Relationship between category distribution difference and the 3D ResNet performance.

A. Category distribution difference for each trial versus distribution of all other trials. B. Test trial accuracy versus its distribution difference from the training dataset. For each trial, the sum of absolute per category difference (from panel A) is used as x-axis and its accuracy when used as test dataset while the rest 6 trials as training dataset is used as the y-axis. A 90% confidence interval is shown as shaded blue.

Table 2.1 Summary of video recordings used in this report.

Species[A]	Type	n	Bower Shape	Training set	Description
-	Empty	3	-	No	No fish in tank
CV	Feeding only	2	-	No	Four female fish
MC	Feeding only	2	-	No	Four female fish
TI	Feeding only	1	-	No	Four female fish
CV	Building	1	Pit	Yes	One male and four female fish.
MC	Building	3	Castle	Yes[B]	One male and four female fish.
TI	Building	2	Pit	Yes	One male and four female fish.
MC/CV F1	Building	2	Pit/Castle[C]	Yes	One male and four female fish.

[A] CV is Copachromis virganialis; MC = Mchenga conophoros; TI = Tramitichromis intermedius

[B] Only two of the three MC trials was manually labeled for training

[C] F1 hybrids display codominant phenotype. Males initially build a pit structure and then transition to build a castle structure nearby the original pit.

Table 2.2 Confusion matrix for sand disturbance events classified with a 3D residual network

	Predicted Label													
Label	**Category**	b	f	p	t	b	m	s	x	o	d	Total	Percent	Acc
	Build scoop (b)	**187**	25	1	0	2	1	0	0	26	2	244	9%	77%
	Feed scoop (f)	23	**303**	0	3	0	48	0	1	28	0	406	15%	75%
	Build spit (p)	1	0	**194**	5	6	1	1	0	10	2	220	8%	88%
	Feed spit (t)	1	6	15	**225**	0	38	2	1	27	26	341	12%	66%
	Build mult (b)	8	0	9	0	**29**	1	1	0	7	0	55	2%	53%
	Feed mult (m)	16	73	4	22	2	**402**	0	0	17	1	537	20	75%
	Spawn (s)	0	0	0	0	1	0	**56**	0	15	1	73	3%	77%
	Shadow/Reflect(x)	0	0	1	1	0	1	0	**169**	11	1	184	7%	92%
	Other (o)	26	26	4	10	10	11	8	9	**388**	20	512	19%	76%
	Sand drop (d)	1	0	1	12	0	1	0	2	14	**149**	180	7%	83%
	Total	263	433	229	278	50	504	68	182	543	202	2752	100%	76%

Table 2.3 Confusion matrix for sand disturbance events classified by a newly-trained human observer.

	Newly-trained label												
label	**Category**	b	f	p	t	b	m	s	x	o	d	Total	Percent	Acc
	Build scoop (b)	**277**	63	88	14	3	4	0	1	15	0	465	15.3%	59.6%
	Feed scoop (f)	58	**356**	5	20	1	7	0	2	8	2	459	15.1%	77.6%
	Build spit (p)	3	0	**196**	16	3	4	0	3	4	1	230	7.6%	85.2%
	Feed spit (t)	3	4	64	**209**	0	6	1	0	2	26	315	10.4%	66.3%
	Build mult (b)	4	0	10	3	**12**	4	0	1	6	0	40	1.3%	30.0%
	Feed mult (m)	14	43	8	34	36	**498**	0	2	7	2	644	21.2%	77.3%
	Spawn (s)	0	0	0	0	0	0	**66**	1	10	0	77	2.5%	85.7%
	Shadow/Reflect (x)	0	1	1	3	0	0	0	**219**	6	1	231	7.6%	94.8%
	Other (o)	19	15	5	11	2	5	3	13	**273**	2	348	11.5%	78.4%
	Sand drop (d)	0	0	67	10	0	1	1	0	11	**140**	230	7.6%	60.9%
	Total	378	482	444	320	57	529	71	242	342	174	3039	100.0%	73.9%

Discussion

Automated classification of behavior is an important goal for many areas of basic, applied, and translational research. Advances in hardware have made collection of large amounts of video data cost effective, either using small, battery-operated cameras (e.g. Go Pro), mobile phones, or small microcomputers (e.g. Raspberry Pi). The small size and inexpensive nature of these hardware systems makes it possible to collect large volumes of data from many individuals. For example, here we collected hundreds of hours of video from seven home tanks in standard aquatics housing facilities. Inexpensive cloud data storage systems additionally allowed for the transfer and archiving of this data. We used Dropbox to store all video data for these experiments, and our long-term goal is to collect data from hundreds of trials in many species.

In this paradigm, our recording system collects >300 gigabytes of video data for each behavioral trial. A major challenge is thus designing a pipeline to identify behaviors of interest through large amounts of background noise. Here we demonstrate one possible solution that involves first identifying environmental disturbances and then classifying the behavioral causes of those disturbances. This approach is rooted in two main advances: first, an action detection algorithm for identifying times and locations of sand disturbances, allowing small video clips containing behaviors of interest to be generated on a large scale; and second, custom-trained 3D Residual Networks to classify each clip into one of ten possible categories.

For action detection, we tailored an HMM to recognize permanent changes in pixel color that occur whenever fish alter the sand. The core approach involves identification of lasting changes in the sand, which manifest as permanent changes in pixel color within specific subfields of view. This approach may be useful for studying many other animal behaviors that are defined by manipulation of the environment, such as nest construction, burrow digging, or web weaving. Many animals also disturb the environment as they forage and feed, for example they may disturb or dig through ground substrate or ingest physical components of the environment such as leaves, fruits, berries, or algae. Thus, this approach may facilitate measurement of many behaviors in more complex and naturalistic environments. One limitation to this approach is that individual pixels must be focused on a specific region of the background. In our paradigm, we accomplished this by fixing the camera, but slightly moving cameras could also be used if an accurate means of registration is available to align frames at different time points.

Following identification of sand disturbance events, we used action recognition to classify events into ethologically meaningful behavioral categories. Previous machine learning strategies have relied on positional tracking and/or pose estimation data to classify animal behaviors (Hong et al., 2015, Anderson and Perona, 2014, Robie et al., 2017). In contrast, our paradigm was poorly suited to these methods, due to a lack of conspicuous stereotypical joint movement from top-down video, and to an abundance of stereotypical interactions between subjects and their environment which were critical for defining the behaviors of interest. We found that a 3D ResNet classified video clips of animal behavior into ten categories more accurately than a newly trained human observer, demonstrating that these networks can effectively quantify many different behaviors of interest from large volumes of raw video data. This result suggests that 3D ResNets may be a powerful tool for measuring animal behavior in naturalistic settings and may drastically increase the scale of experimental designs in systems that, historically, have been constrained by the amount of human observation time required to measure behaviors of interest.

Feeding and construction behaviors have different underlying goals and are critical for survival and reproduction in the wild, but their physical execution is often indistinguishable to inexperienced observers. Separating these behaviors is critically important in our paradigm, where there is a high risk that decisions to scoop and spit during bower construction will be undetectable through a high volume of feeding actions. Remarkably, the 3D ResNet was able to distinguish between feeding and construction behaviors more accurately than a newly trained human observer. The ability to accurately identify subtle differences in behavior may have important implications for better understanding the progression of neurological diseases characterized by subtle changes in locomotor function over extended periods of time, or monitoring efficacy of different treatment strategies for improving locomotor function.

Striking differences in the relative spatial positions and timing of predicted feeding versus construction behaviors were readily apparent across whole trials. For example, in castle-building MC males, bower spits were more spatially concentrated than bower scoops, feeding spits, and feeding scoops, consistent with the idea that castle construction is driven by scooping sand from dispersed locations and spitting into a concentrated region. Similarly, feeding behaviors and bower construction behaviors were expressed daily but at different times, whereas spawning events occurred infrequently in punctuated bursts. These data show that our system can be used to map tens of thousands of behavioral events in time and space, allowing future studies to unravel how these different complex behaviors are expressed in dynamic environments over extended time periods. It will be important to place these sand manipulation events with information about the actual structure being built by the fish. Development of an approach to characterize the shape of the bower at any given time will be necessary for connecting building behaviors with the structure.

The 3D ResNet accurately classified behaviors across three different species and one interspecies hybrid cross that all differ in morphology and color patterning. For example, Copadichromis virginalis males exhibit black body coloration, yellow heads and dorsal fins, and narrower jaws relative to Tramitichromis intermedius males, which exhibit yellow and red body coloration, blue heads, and a relatively wide jaw apparatus. Bower construction in particular has evolved in hundreds of species, and the evolution of feeding morphology and behavior is thought to be central to the explosive radiation of cichlids into thousands of species. Our results suggest that our system will likely be effective for measuring natural variation in these behaviors among hundreds of species. Lake Malawi cichlids can also be hybridized across species boundaries, enabling powerful genetic mapping approaches (e.g. quantitative trait loci mapping) to be applied in subsequent hybrid generations to identify genetic variants that influence complex traits. Our data shows that our system is effective for phenotyping hybrid individuals, allowing future studies to identify specific regions of the genome that are responsible for pit vs. castle building. Lastly, high prediction accuracy for quivering, a conserved and stereotyped sexual behavior expressed by many fish, supports that action recognition may be useful for analyzing mating behaviors in many fish species. More broadly, our results suggest that 3D ResNets may be effective tools for measuring complex behaviors in other systems even when individuals vary substantially in physical traits.

2.3 Materials and Methods

2.3.1 Animals and husbandry

2.3.1.1 Subjects

Lake Malawi bower-building species (Copadichromis virginalis, Tramitichromis intermedius, Mchenga conophoros) derived from wild-caught stock populations, as well as genetically hybrid individuals derived from two of these species (described below), were housed in social communities (20-30 individuals) in 190 liter glass aquaria (90.2 cm long x 44.8 cm wide x 41.9 cm tall) into adulthood (>180 days). Aquaria were maintained under conditions reflective of the Lake Malawi environment: pH=8.2, 26.7ºC water, and a 12 h:12 h light:dark cycle with 60-minute transitional dim light periods. Fish were fed twice daily with dried spirulina flakes (Pentair Aquatic Eco-Systems).

2.3.1.2 In vitro hybridization

Reproductively active males and females were visually identified based on abdominal distension (females), nuptial coloration (males), and expression of classic courtship behaviors (e.g. chasing/leading and quivering). A hybrid cross was created between Mchenga conophoros (female) and Copadichromis virginalis (male). To cross-fertilize, a petri dish was filled with water from the home tank, and eggs were collected into the dish by applying gentle pressure between the pectoral region and the anal pore of the female. Eggs remained fully submerged while the male's sperm was extracted into the same dish by applying gentle pressure to both sides of the abdomen. The mixture was immediately and gently agitated and then eggs were gently rinsed twice with fresh aquarium water to reduce polyspermy. Eggs were transferred into a beaker containing a fresh oxygen tube, fresh aquarium water, and a drop of methylene blue to minimize risk of fungal infection.

Water replacement was performed at least once daily until hatching (approximately 5-6 days post-fertilization).

2.3.2 Behavioral trials

Bower building occurs for multiple days. Animal care guidelines required that testing over such extended time periods had to be done in the home tank (as opposed to external testing arenas). In our facilities, home tanks are supported on tank racks with built-in piping and support beams that partially occlude top-down fields of view (FOVs). Additionally, all tanks have a central support crossbeam that partially occludes top-down FOVs. We found that a ~36 cm diameter sand tray placed on one half of the home tank provided a sufficient volume of sand for males to construct bowers, and was small enough to fit into an unobstructed top-down FOV. We designed a custom acrylic platform to surround the sand tray to prevent subjects from spitting sand over the edge of the tray onto the bottom of the aquarium. Thus, in this design, subject males and females could freely enter and exit the sand tray region throughout the trial.

For all behavioral experiments, a single reproductive adult male and four reproductive adult stimulus females of the same species or hybrid background were introduced into designated home tanks equipped with additional LED strip lighting (10 h:14 h light:dark cycle synced with full lights on), and a custom-designed hollow acrylic case (43.1 cm long x 43.1 cm wide x 10.2 cm tall, with a 35.6 cm diameter circular opening) surrounding a circular plastic tray (35.6 cm diameter x 6.4 cm deep, and elevated 3.8 cm above the aquarium bottom) filled with sand (Carib Sea; ACS00222). Sand trays were positioned approximately 30 cm directly below a custom-designed transparent acrylic tank cover (38.1 cm long x 38.1 cm wide x 4.4 cm tall) that contacted the water surface to eliminate rippling for top-down video recordings. Subjects and stimulus females were allowed to freely interact throughout the entirety of the recording trial.

2.3.3 Recording and monitoring hardware

We used a Raspberry Pi 3 Model B (RASPBERRYPI3-MODB-1GB; Raspberry Pi Foundation) connected a Raspberry Pi camera v2 (RPI 8MP CAMERA BOARD; Raspberry Pi Foundation) and a 1 TB external hard drive (WDBUZG0010BBK-WESN; Western Digital) to collect video clips for each trial. Data was stored locally on the hard drive until the end of the trial and then transferred to Dropbox through an Ethernet connection for analysis. The Raspberry Pi camera was placed approximately 58 cm above the sand tray.

2.3.4 Data collection and analysis software

2.3.4.1 Video Collection

Upon start of a trial, an automated recording protocol was initiated collecting RGB video data during full light hours (08:00 to 18:00 EST) for 7-10 days. h264-encoded videos were collected at a 30 frame per second frame rate and a resolution of 1296x972 using custom Python scripts that used the picamera package (https://picamera.readthedocs.io). h264 videos were encoded into mp4 videos using ffmpeg (https://www.ffmpeg.org/). Data was transferred to a laboratory Dropbox account using rclone (https://rclone.org/).

2.3.4.2 Identification of HMM state changes

All code for running action detection is available on github at https://github.com/ptmcgrat/CichlidActionDetection. This repository included code that accomplishes the following: (i) uses a Hidden Markov Model (HMM) algorithm to detect changes in pixel values through time by sampling one frame per second, (ii) uses a density-based clustering algorithm to identify clusters of HMM+ pixels, or putative sand change events, and (iii) creates video clips and frames for manual labeling and machine learning classification.

To calculate HMM-predicted pixel values, we used the opencv and numpy packages for Python to decompress color mp4 videos into gray scale numpy arrays (0-255) to access data for each pixel across the entire video. To filter out short-term changes caused by fish, we calculated two rolling mean values for each pixel across a 120 second window either before and after each time point. Pixel values that were 7.5 units above or below either rolling mean value were removed and then interpolated using the ‘numpy’ package. Enduring changes in pixel values were identified using the ‘hmmlearn’ package for Python. Initial testing indicated that the time to calculate the HMM for the entire video would be on the order of days. Since the time to calculate the HMM states is on the order of O(N2), where N is the number of hidden states, we performed two additional pre-processing states to reduce the number of hidden states for each pixel. First, we only considered mean state values (i.e 0,2,..252,254). Second, since each pixel did not explore the entire range of possible values, we also calculated the values for each pixel that were found 10 or more times in the entire video. By using these heuristics, we could reduce the number of states to ~10-20.

We also found that changes in lighting over the course of the day created small changes in mean pixel value that resulted in HMM state transitions. To limit the number of these small HMM changes, we also prevented transitions less than 4 units by modifying the transition probability matrix.

2.3.4.3 Clustering of HMM state changes together into sand-manipulation events

In order to group HMM state changes together that were caused by the same fish-mediated event, we used density-based spatial clustering of applications with noise (DBSCAN) within the Python package ‘scikit-learn’ to identify clusters of HMM change in the presence of noise. DBSCAN analyzes the region surrounding each HMM pixel in time and space, determines if the neighboring region contains a minimal number of HMM+ pixels using a KD-tree, expands on dense groups of points, and repeats. DBSCAN parameters were set based on the observed size of sand change events and based on a k-dist graph. This enabled us to identify spatiotemporal clusters of HMM+ pixels, representing putative sand change events. Detailed discussion of different aspects of clustering follows.

Occasionally there were large changes in lighting that resulted in state changes for the majority of pixels. We filtered out HMM+ changes from times when 1% or more of the pixels changed.

Parameters for density-based clustering

minPts: observers reviewed several hundred putative sand perturbance events and estimated the minimum size of a true sand change cluster to be 10 pixels x 10 pixels x 3 frames, and HMM+ pixels change to cover at least 15% of the putative sand change region.

Based on these estimates we calculated the range for the minimum number of pixels in a sand change event to be between 50-250 pixels. For this paper we used 90 points for the minPts parameter.

eps: For a given k we defined a function k-dist from the database D into the non-negative real numbers, mapping each point to the distance from its k-th nearest neighbor. After sorting the points in the database in descending order based on their k-dist values, the graph of this function suggested a density distribution in the database. This graph is called the sorted k-dist graph, as described previously (Ester et al., 1996). We then fit a nearest neighbor tree to all points and used the k neighbors query to find the minPtsth nearest neighbor for each point, and the k-dist graphs for minPts = 200. We found that most of the points were close to each other; and most points had at least 200 points within 40 units.

We used the knee point of the first k-dist graph (at minPts = 200; Figure S2) to estimate the optimal values for eps to be 20-30. We then ran DBSCAN on a grid of parameters and quantified the number of clusters labeled under each set of parameters. Three observers then annotated three sets of clips corresponding to minPts and eps values (Figure S2). After comprehensive review, we found the eps = 18 and minPts = 90 to best reflect true sand change clusters.

Nearest Neighbor KD-tree treeR/neighborR and leaf size

treeR and neighborR are equivalent parameters for constructing KD-trees (Pedregosa et al., 2011). Within a radius around each point, all distances between this point and other points are calculated. DBSCAN queries the distances within eps (eps=18 in our analysis) for each point, so the treeR/neighborR ⩾ eps. We set this parameter to 22 to prepare the distance matrix for DBSCAN with eps <= 22.

leaf_size: this parameter is a threshold below which the calculation switches from traversing tree to brute-force. For small data sets (N less than 30 or so), brute force algorithms can be more efficient than a tree-based approach. Changing leaf_size will not affect the results of a query, but can significantly impact the speed of a query and the memory required to store the constructed tree as seen in (Pedregosa et al., 2011) and here: https://jakevdp.github.io/blog/2013/04/29/benchmarking-nearest-neighbor-searches-in-python/#Scaling-with-Leaf-Size. We set leaf_size above minPts 90 (leaf_size=190).

Timescale:

Since DBSCAN uses one radius to search clusters in all dimensions, we scaled the time dimension so that the temporal lengths of events were similar in magnitude to their spatial width, such that events were, in general, roughly spherical in 3D. By manually reviewing hundreds of events, we determined that the duration of sand change events was < 5 seconds, and the spatial widths were typically < 50 pixels. Based on this, the time dimension (on frame/second) was scaled by 10x.

2.3.5 Creation of video clips for each cluster

Finally, to create video clips for machine learning, we used the "opencv" package for Python to create small video clips around the center of each cluster. The width, height and length of these videos were 200 pixels, 200 pixels, and 120 frames (4 seconds, respectively.

For manual labeling, we also included cluster information on the pixels that underwent HMM transitions during the 4 second time window.

2.3.6 Machine Learning

2.3.6.1 Behavioral definitions for manual annotation

The following 10 categories were used to categorize each of the manually-labeled video clips.

Bower scoop: subject male collects sand into its mouth during bower construction.

Bower spit: subject male expels sand from its mouth during bower construction.

Bower multiple: multiple bower scoops and/or spits are expressed by the subject male within the same video clip.

Feeding scoop: fish collects sand into its mouth during feeding.

Feeding spit: fish expels sand into its mouth during feeding.

Feeding multiple: multiple feeding scoops and/or spits are expressed by a fish within the same video clip.

Spawn/quiver: the subject male rapidly vibrates his body left to right while simultaneously circling, often but not necessarily with a female in frame. The male's body is typically arched left to right, with his anal fin (egg spots) displayed directly in front of the female. The female may also be present and circling in immediate proximity with the male.

Sand dropping: A fish expels or releases sand from the mouth either while high in the water (after which the sand sprinkles down through the water before settling), or release of sand upon initiation of a rapid burst of swimming (typically chasing or being chased). A rarer subset of sand dropping events includes filtering sand through the operculum while swimming, typically during feeding.

Other: Changes to the sand caused by any other fish activity not described above, often as a result of swiping of the fin or rubbing of the ventral surface of the body along the sand during performance of other behaviors. More rare cases included instances in which two fish both perform behaviors in the same clip but the sand change was designated as a single cluster.

Shadow/reflection

Other changes that are not caused by fish manipulating or changing sand, most commonly reflections of activity in the aquarium glass and shadows cast by a stationary or very slow-moving fish, or in rare instances food, feces, or other debris settling on the sand surface.

All labeled data can be found at:

https://data.mendeley.com/datasets/3hspb73m79/draft?a=b72c1f6d-505a-431a-ba3d-824cd148c01e

2.3.7 Deep learning of cichlid behaviors

All code for running behavior classification of available on github at https://github.com/ptmcgrat/CichlidActionRecognition. A trained observer manually classified 14,234 video clips randomly selected from representative days across seven trials, spanning seven subjects, three species, and one hybrid cross. Each clip was classified into one of the ten categories listed above. We randomly selected 80% of manually annotated clips for training an 18 convolutional layer 3D ResNet, and the remaining 20% of clips were used for validation. Briefly, 3D ResNets are 3D convolutional neural networks (CNNs) that incorporate features of Residual Networks (ResNets), in which signals are bypassed across convolutional layers during training. This approach allows 3D ResNets to be deeper and more accurate than traditional 3D CNNs for action classification tasks. For training, validation, and prediction we used a 18- layer Resnet3D model as previously described (Qiu et al., 2017). The architecture, including the shape of each layer, of the neural network is shown in Figure 5A. Each ResNet Block consists of 2 convolutional layers, each has a kernel size of 3x3x3. The first convolutional layer has a kernel size of 7x7x7. ReLu is used as the activation function across the neural network. Prior to training, each video clip was randomly cropped at a 120x120 frame. Each training video clip was also randomly cropped temporally down to 96 continuous frames. Random horizontal and vertical flip, at a rate of 0.5, was also used for data augmentation. Finally, each channel was then normalized based on the mean value for that channel. Validation and test video clips were always cropped in the center spatially and temporally. For training, stochastic gradient descent was used to optimize the parameters of the neural network. Specifically, the learning rate was set to 0.1 (and set to decrease after 10 consecutive epochs of no change in validation loss), momentum was set to 0.9, dampening was set to 0.9, and weight decay was set to 1.0x10-5. The network was only initialized at the start and was trained for 100 epochs with a batch size of 8 per epoch. The final accuracy is calculated as the average the last five epochs when the accuracy reached plateau by visual inspection.

2.3.8 Testing model generalizability

In order to test the generalizability of the learned model, we used 6 out of the 7 projects for training/validation and the other one for testing. We tested all 7 combinations of these 7 projects. Sample split, network architecture and data augmentation were the same as above and test accuracy is calculated as the average the last five epochs when the test accuracy reached plateau. In order to test if label some of the test video clips could help the model to generalize to the new dataset, we randomly put 100,400,800 randomly selected test clips in training and the remaining test clips in test. After this, the new training/validation dataset went through the data augmentation and network computation. The new test accuracy is calculated on the remaining test video clips.

To figure out why accuracy decrease in test trial, we compared the distribution of the categories for each test trial and the that of the training trials. For each trial, we first calculated the percentage for each category. This distribution is subtracted by that of the training trials to get the distribution difference. Finally, the accuracy when this trial is used as test dataset is regressed on the sum of absolute per category difference. The p value is calculated from Pearson's Correlation Coefficient.

CHAPTER 3. PEEL-1/ZEEL-1 IS AN UNSELFISH SELFISH GENETIC ELEMENT

3.1 Introduction

Selfish genetic elements are genetic elements that can enhance their own transmission at the expense of other genetic elements in the genome. For example, transposable elements represent one class of selfish genetic elements that can drive their own replication in the genome, often at the expense of the fitness of the organism. Conflict of transmission patterns between selfish genetic elements and the rest of the genome could drive speciation events. It has been speculated that P-element-mediated hybrid dysgenesis is thought to create reproductive isolation between different species of *Drosophila melanogaster* (Ginzburg et al., 1984). The hybrid progenies of crosses between males that have the P-elements and females without those elements suffer from gonadal dysfunction and elevated deleterious mutation rates. Cytoplasmic incompatibility induced by Wolbachia infection is another kind of selfish genetic element that also contribute to speciation. The cytoplasmic incompatibility can be shown by the death of embryos sired by an infected father (Werren, 1997). Study of the Wolbachia infected species *Drosophila recens* and uninfected sister species *Drosophila subquinaria* showed that reproductive isolation is caused by cytoplasmic incompatibility (Shoemaker et al., 1999).

Not only are selfish genetic elements important to evolution of new species, they are also now being hijacked by bioengineers to cure human diseases. Transposable elements, such as Piggybac, have been used to induce pluripotent stem cells (Rajabpour et al., 2014).

Inspired by toxin-antidote elements and homing nuclease, genetically engineered malaria resistant mosquitos are designed to carry selfish genetic elements in close linkage with the genetic modification in order to spread quickly in wild populations (Champer et al., 2020, Windbichler et al., 2011). Despite their importance in speciation and human health, the evolutionary forces that act on these selfish genetic elements are not fully understood.

Toxin-antidote elements (TAs) are a special type of selfish genetic elements. They are typically composed of two closely linked genes that express toxin and antidotes, respectively. Based on which parent deliver the toxin, toxin-antidote elements can be further classified as paternal effect and maternal effect TAs. In most cases, the toxin gene and the antidote gene are closely linked on the same chromosome. When animal carrying the toxin-antidote element crosses with the animal without these elements, a quarter of the F2 will be killed because all of F2s will get the toxin from the heterozygous parents while a quarter of the progenies will not get the antidote expressing in the zygote. In this way, the frequency of TA in F2s is higher than 50%. Due to the strong genic drive, toxin-antidote elements can spread in a population even if it imposes a fitness cost to the animals carrying it (Wade and Beeman, 1994). One challenge to understand the evolutionary force acting on toxin-antidote elements is a lack of quantitative methods to compare the fitness effect of toxin-antidote elements in different environments. Studies in bacteria often use fitness-related phenotypes such as stress tolerance, persistence or biofilm formation (Muthuramalingam et al., 2019, Goormaghtigh et al., 2018). These phenotypes, even though closely linked to fitness, only account for part of the factors leading to fitness change. Recently, competition assays were also used to systematically study the fitness effect of all 13 toxin-antidote elements in *Pseudomonas putida* (Rosendahl et al., 2020).

Deletion of all these toxin-antidote elements did not appreciably reduce the fitness of bacteria cells. These results in bacteria, even though helpful to understand the selection of TA in bacteria, are not directly transferable to other eukaryotic systems because prokaryotes use different developmental pathways and reproductive system. Also, toxin-antidote element in bacteria does not function in a transgenerational approach like that in eukaryotes. Thus, study of toxin-antidote elements in a eukaryotic model will help us fully understand the spread of these elements.

Finding a toxin-antidote element in multicellular organisms is not easy. This is because most animals are dioecious, where generation of offspring requires separate males and females. In these kinds of systems, toxin-antidote elements are expected to quickly fix in a population. Thus, toxin-antidote elements can only be revealed by interspecific crosses or crosses between distant population. For example, *Medea* was found by crossing animals from different continents: one from Asia (Singapore) and the other one in North America (United States) (Beeman et al., 1992).

Caenorhabditis have become a promising model for the identification and study of toxin-antidote elements identified. The first pair of toxin-antidote identified in *C. elegans* is a pair of paternal effect toxin antidote elements: *peel-1* (paternal effect epistatic embryonic lethal) and *zeel-1* (zygotic eistatic embryonic lethal). *peel-zeel* were first reported in 2008 as an unexpected finding in the process of generating RILs (recombinant inbred lines) (Seidel et al., 2008). *peel-1* encodes a toxin that is delivered via sperm specific vesicles and *zeel-1* encodes an antidote that is transiently expressed in the embryo (Seidel et al., 2011). PEEL-1 and ZEEL-1 are both transmembrane proteins. Without ZEEL-1, PEEL-1 can stop embryo development from the embryonic two-fold stage. Though the detailed molecular mechanism of *peel-1* toxicity remains unclear, *peel-zeel* has been developed as a reagent for many applications such as making integrated transgenic animals (Frøkjær-Jensen et al., 2012, Muñoz-Jiménez et al., 2017, Joseph et al., 2018).

A second example of a toxin-antidote element identified in *C. elegans* is a pair of maternal effect toxin-antidote genes: *pha-1* and *sup-35 (Ben-David et al., 2017)*. *pha-1* was originally identified as an essential devolopment gene in genetic screens (Kuzmanov et al., 2014). Howerver, work by Ben-David and colleagues demonstrated that *pha-1* encodes a zygotic antidote, without which animals will show developental defects induced by the maternal effect toxin gene *sup-35*.

Recently studies have expaned analysis of TAs in a second species of *Caenorhabditis*: *Caenorhabditis tropicalis*. *C. tropicalis* also reproduces primarily as a hermaphrodite, although the evolution of this mating system was unique to *C. elegans*. Work by two groups have identified at least five more TAs (Noble et al., 2020, Ben-David et al., 2020). These are maternal effect TAs and most of them do not kill embryos but instead cause larval arrest or developmental delay.

One interesting phenomenon that was observed in these toxin-antidote elements in both species is the large amount of genetic diversity between the TA haplotype and the corresponding non-TA haplotype. For example, across all 330 unique wild isolates of *C. elegans* sampled around the world, there are two major haplotypes at the *peel-zeel* loci: the Bristol haplotype and the Hawaii haplotype. The Bristol haplotype has the functional *peel-1* and *zeel-1* while the Hawaii haplotype has these two genes deleted (Figure 3.2A and B). In the genomic region close to *peel-zeel*, the average divergence between the two haplotypes in the coding region is 5%, which is 50 times higher than the genome wide average (Seidel et al., 2008). Population genetics analysis identifies a strong signature of balancing selection by analyzing the protein sequence of the nearby gene *srbc-64*. Among all the wild isolates identified so far, 465 wild isolates encompassing 185 distinct isotypes have the Bristol haplotype and 301 wild isolates encompassing 145 distinct isotypes have the Hawaii haplotype (Figure 3.2B). These two haplotypes are distributed worldwide and can be found as close as in a single orchard (Richaud et al., 2018). The high divergence implies the two haplotypes diverged roughly 8 million generations ago, possibly before the evolution of androdioecy (males and self-fertilizing hermaphrodites). However, the long history does not agree with the expectation of genic drive where the driver Bristol haplotype can be quickly fixed within outcrossing populations. Even if the idiosyncratic features of *Caenorhabditis* androdioecy were taken into account in simulations, self-fertilization would only slow down the fixation of TA elements (Noble et al., 2020). For *pha-1/sup-35*, there is also high genetic divergence between the alleles with and without the *pha-1/sup-35.* The haplotype without *pha-1/sup-35* has a 20kb deletion relative to the haplotype with *pha-1/sup-35*. There is also a large inversion in this region (Ben-David et al., 2017). Studies in *C. tropicalis* have found antagonizing TAs, where two toxin-antidote elements are found as pairs on the same chromosome (Ben-David et al., 2020). Simulations in one study have shown antagonizing TAs impose a strong heterozygote advantage such that these two toxin-antidote elements are under balancing selection. Whereas the other research simulating the same scenario with antagonizing TAs showed positive-frequency dependent selection which bias the fixation of the most frequency toxin-antidote element in the population (Noble et al., 2020). These inconsistencies of simulations are not only caused by different simulation programs, but also by the fact that little is known about the evolution of toxin-antidote elements by the whole research community. In this chapter, we seek to understand why *peel-zeel* is under balancing selection.

There are many mechanisms that two alleles/haplotypes can be balancing selected in the population including negative frequency-dependent selection, spatial or temporal habitat heterogeneity, and heterozygote advantage (Brisson, 2018). Negative frequency-dependent selection happens when the relative fitness of a variant increase as its frequency decrease in the population. Spatial or temporal habitat heterogeneity happens when the fitness of the variant depends on the habitat (Levene, 1953, Ravigné et al., 2004). This often happens when the variant is responsible for environment sensing and responding. In addition to these traditional models, theoretical analysis specific to TAs also showed a stable equilibrium point between the two haplotypes when the TAs induce a female fecundity loss (Wade and Beeman, 1994).

In this chapter, we developed and tested different models by which this genetic element could be under balancing selection. We used a family-based population genetic method to estimate how much fitness loss has to be incurred by the *peel-zeel* element in order to reach an equilibrium in an androdioecy population. We showed that a high outcrossing rate was required to maintain *peel-zeel* in a population. The amount of fitness loss depended on the outcrossing rate. Second, while the outcrossing rate is variable across different strains and geographical locations, we set out to test if there is any fitness loss associated with the *peel-zeel* region using quantitative pairwise competition assays. To our surprise, the Bristol haplotype actually has a relatively higher fitness and fecundity than the Hawaii haplotype. We further found that this elevated fitness and fecundity was caused by the toxin gene *peel-*

1 which acts in a *zeel-1* independent manner. Finally, we tested if linked variants of *peel-zeel* have functions over environment sensing. While we found a dauer pheromone response difference for linked variants, two candidate genes *srbc-64* and *ugt-31* are not responsible for the phenotypic change. These results suggested that toxin-antidote elements can spread through the population independent of their selfish effects. Toxin-antidote elements can also provide fitness benefits, either during their initial evolution or after being coopted by the animals following their spread in the population. Linked variants for dauer pheromone response could be responsible for the balancing selection.

3.2 Results

3.2.1 Equilibrium of peel-zeel requires high outcrossing rate and a cost to hermaphrodite fitness

By delivering toxin to the sperm and killing the zygote that does not carry the antidote, the Bristol haplotype is able to eliminate the Hawaii haplotype in about 100 generations (Wade and Beeman, 1994). This genic drive, similar to positive selection, is expected to reduce local genetic diversity and skew the site frequency spectrum to rare. This expectation does not agree with the observed high genetic diversity between these two haplotypes which are predicted to have diverged 8 million generations ago (Seidel et al., 2008). Allele frequencies in this region show strong signature of balancing selection using the Tajima's D test. In order to explore how balancing selection could act on this locus, we considered three models by which *peel-zeel* can be under balancing selection. For the first two models, we assumed that since *peel-1* codes for a toxic protein, the *peel-zeel* haplotype would incur a fitness cost to the hermaphroditic animals carrying it due to an inability of the antidote to completely counteract the toxin. For the first model, genic drive would be counteracted by the fitness cost such that an equilibrium point exists for *peel-zeel* in a single population (**Figure 3.1A**). In the second model, we considered how different environments could induce different outcrossing rate and fitness cost for the *peel-zeel* locus. It is known that environmental factors influence the rate that males are generated, and different environments could have differences in the number of unique animals that colonize a rotting fruit or other material. Therefore, it is reasonable to think that outcrossing rate could vary. In this way, some environments could have *peel-zeel* fixed while others have *peel-zeel* purged (**Figure 3.1B**). Finally, the hyper-divergent region around *peel-zeel* contains several other genes important to the lifestyle of animals. We hypothesized linked genetic polymorphism is driving the evolution of this haplotype, such that *peel-zeel* can be under balancing selection (**Figure 3.1C**).

To find a theoretical estimate of how much fitness cost is necessary to maintain both haplotypes in a single population, we adopted a family-based population genetics model (Wade and Beeman, 1994) (**Table 3.1** and **Figure 3.2D**). This model was originally used to model Medea in flour beetles. Parameter h was used to model the dominance of the fecundity cost and parameter t was used model the lethality. To adapt this model to the unique mating system in *C. elegans* and further advance the model, we made three changes. First, *C. elegans* has a different reproductive system. While males can still cross with hermaphrodites to create outcrossing progenies, hermaphrodites themselves were able to self-fertilize. Thus, we expanded the table by adding three self-fertilization families and an additional parameter k was used to model the outcrossing rate. Second, we changed the offspring fitness in some families to mimic paternal effect lethality.

Third, we modeled hermaphrodite fitness cost instead of female fecundity loss to directly account for both fecundity loss and/or slower growth rate that affect transmission probability. We only modeled hermaphrodite fitness cost (parameter *s*) because hermaphrodite expresses toxins in its germline, and we can directly measure hermaphrodite fitness in the laboratory using pairwise competition assays. Fitness cost for hermaphrodites in outcrossing might be different from that in self-fertilization, thus we used *s'* to denote fitness cost in self-fertilization. To simplify calculation and simulation, we assume equal *s* and *s'* in the following analysis.

To understand how fitness cost affects the dynamics of *peel-zeel*, we calculated the frequencies of *peel-zeel* in a population for each generation using **Table 3.1** under different fitness cost (**Figure 3.2D**). When we set fitness cost s to 0.5, a fully outcrossing population started with either 95% *peel-zeel* (10% heterozygotes and 90% Bristol homozygotes) or 5% *peel-zeel* (10% heterozygotes and 90% Hawaii homozygotes) converged to 80% in the population in about 40 generations (**Figure 3.2E**). This agreed with the calculation that there was always a convergent equilibrium point at $(2-s)/2$ (Wade and Beeman, 1994). When there was no fitness cost ($s=0$), *peel-zeel* will be fixed in the population in about 100 generations due to its genic drive.

To understand how outcrossing rate affect the dynamics of *peel-zeel*, we calculated the frequencies of *peel-zeel* under different outcrossing rates. The outcomes (fixed, purged or balanced) of *peel-zeel* depends on a complex interaction between fitness cost and outcrossing rate. *peel-zeel* can either be balanced or purged in the population with a fitness cost $s > 0.25$ (**Figure 3.2F**). When outcrossing rate is high, genic drive is able to counteract the fitness cost and thus these two haplotypes can be balanced. The outcrossing rate in this case only changed the speed to reach equilibrium, with a lower outcrossing rate slows down the speed to equilibrium. On the other hand, when s < 0.25, changing the outcrossing rate can fix, eliminate, or balance *peel-zeel* (**Figure 3.2G**). *peel-zeel* (s=0.05) can be balanced when the outcrossing rate was high (*k=1*), whereas it will be eliminated in the population when outcrossing rate was low (*k=0.8*). When the outcrossing rate was intermediate (*k=0.9*), *peel-zeel* will reach fixation in around 80 generations. There are two reasons to explain this phenomenon. First, the equilibrium frequency *(2-s)/2* was negatively correlated with parameter s such that a lower *s* increased the equilibrium frequency. Second, when *peel-zeel* is at very high frequency (close to 1), almost all the Hawaii haplotypes were in heterozygotes. While the individuals that outcross can keep Hawaii and Bristol haplotypes in an equilibrium frequency, individuals that self-fertilize can drive Bristol haplotype to an even higher frequency. This is because the frequency of Bristol haplotype in the progeny of heterozygotes is 2/3. When s is small, this increased frequency by selfing heterozygotes can overcome the reduced fitness for Bristol homozygotes such that Bristol allele will get fixed in the population.

These results showed a complex interaction between outcrossing rate and fitness cost in determining the evolutional dynamics of *peel-zeel*. To find the parameter space where *peel-1/zeel-1* can be balanced, fixed or purged from the population, we did a grid search and found the parameter space was divided to three parts. (**Figure 3.2H**). In general, a high outcrossing rate is required for *peel-zeel* reach an equilibrium point. When the outcrossing rate is lower than 0.7, the *peel-zeel* loci will either get fixed (small s) or purged (large s) from the population regardless of the degree of hermaphrodite fitness cost. In a fully outcrossing population, the *peel-zeel* loci can be balanced whenever $s > 0$. The parameter space for *peel-zeel* to reach equilibrium is even smaller when the parameter h, the dominance of *peel-zeel* in its effect on animal fitness, is higher (**Figure 3.3**). When the fitness effect is close to additive ($h=0.4$), fitness cost has to be between 0 and 0.2.

C. elegans in the wild has a outcrossing rate close to 1% (Barrière and Félix, 2005). At this outcrossing rate, genic drive is small. Thus, *peel-zeel* will be fixed if $s < 0.0005$ and purged if $s > 0.0005$ (**Figure 3.2H**). Model 1 can be rejected by this result because equilibrium point in a population can neither be established in a population with very low outcrossing rate. If some populations have outcrossing rate even lower, almost any fitness cost to *peel-zeel* will get it purged from the population. This result favors model 2 where *peel-zeel* can dominant some populations but not others because fitness cost is likely to be environment dependent. In summary, these results showed that fitness cost and genic drive can generate an equilibrium frequency for *peel-zeel*. However, reaching this equilibrium requires a high outcrossing rate that is unlikely the case for *C. elegans.*

3.2.2 Linked variants of peel-zeel region increased hermaphrodite fitness

We need data about the fitness cost of *peel-zeel* locus in hermaphrodite animals to test model 2 and model 3. For model 2, we predict that *peel-zeel* should carry a fitness cost on the animals. For the third model, we predict that closely linked genetic variants would encode functional variation that could be under balancing selection. To test this hypothesis, we leveraged a previously-constructed nearly isogenic line (NIL) QX1198 that carries the

Hawaii haplotype (140-370kb) of *peel-zeel* in the N2 background. The fitness of N2 can then be compared directly to the NIL strain using a previously described pairwise competition assay (Large et al., 2016). Specifically, QX1198 was competed against PTM229, a N2 strain that also contains a silent mutation in the *dpy-10* gene. This silent mutation is neutral and provides a common genetic variant for assaying the relative proportion of each strain using digital droplet PCR. Contrary to model 2, the strain PTM229 has a higher fitness than QX1198, with a relative fitness (w) of 1.18 (1.15-1.21, 95% CI) (**Figure 3.4A**). The fitness difference is similar to that of the positive control strain CX12311 (w = 1.19±0.09, 95% CI). CX12311 was a near isogenic line that carries ancestral *npr-1* and *glb-5* alleles and was shown previously to have a significant fitness effect in laboratory environment (Zhao et al., 2018).

To understand the physiological change that was responsible for the fitness difference, we measured two metrics that were directly linked to fitness: fecundity and growth rate. We found QX1198 has both reduced fecundity and slower growth rate. Specifically, N2 hermaphrodites on average lay 351 eggs (337-366, 95% CI) throughout lifetime while QX1198 lay 321 eggs (312-331, 95% CI), 9% lower than that of N2 (**Figure 3.4B**). After growing 72 hours at 20°C, N2 animals were 9% larger in length than QX1198 (**Figure 3.4C**).

With linked genetic variants change phenotypes affecting fitness, balancing selection can act on these additional variants. Thus, this result is consistent with the third model. While we detected an increased fitness for the Bristol allele in the laboratory conditions, its fitness effect can be negative in some natural conditions. We have previously shown animals adapt to the unique laboratory environment using a set of genetic changes (Large et al., 2016) (Zhao et al., 2018). Fitness effect of linked genetic variants could also change by different environments.

The consequence of this increased fitness on the dynamics of *peel-zeel*, on top of the gene selection, was an accelerated rate to fixation (**Figure 3.4D**). The difference in time for *peel-zeel* with and without increase fitness (s=-0.18) to reach 50% of the population from a starting frequency of 5% is small in a fully outcrossing population (21 vs 23 generations). This is because genic drive dominants the dynamic in a population with high outcrossing rate. The time difference, however, is large (23 vs 44 generations) when the population only has intermediate outcrossing rate (k = 0.5). In summary, these results showed linked variants of *peel-zeel* had a positive fitness effect in laboratory environment and this positive fitness will accelerate the fixation of *peel-zeel* in a population.

3.2.3 peel-1 incurs a direct fitness advantage

Besides *peel-zeel* polymorphism, the NIL strain had a large introgression that might have other polymorphisms influencing fitness. To directly test the fitness effect of *peel-zeel*, we used CRISPR/Cas9 to make a *peel-1* knockout strain (**Figure 3.5A**). This mutation created a frameshift at the 44th amino acid of *peel-1* gene and this frameshift generated a stop codon at the 46th amino acid of *peel-1*. We confirmed this was an loss of function allele by crossing *peel-1*$^{-/-}$ with the Hawaii strain CB4856 and demonstrating that there was no embryo lethality (0%) in F2, whereas N2 crossed with CB4856 yields 24.5% embryo lethality as expected (**Figure 3.5B**). We then used this strain to compete with N2, either with the *dpy-10* barcode on N2 or on *peel-1*$^{-/-}$ strain. Both reciprocal competition assays showed that the functional *peel-1* gene has a small but significant relative fitness advantage of 1.06 (1.04-1.07, 95% CI) over *peel-1*$^{-/-}$ animals (**Figure 3.5C**). This fitness difference accounts for 32% of the fitness difference between NIL strain QX1198 and N2.

To find the physiological change responsible for the fitness difference, we assayed the brood size and growth rate for *peel-1*$^{-/-}$. We found *peel-1*$^{-/-}$ hermaphrodites lay 331 (321-340, 95% CI) eggs through lifespan (**Figure 3.5D**). This is 6% lower than N2. After growing at 20°C for 72 hours, N2 grows 7% faster than *peel-1*$^{-/-}$ (**Figure 3.5E**). These results showed functional *peel-1,* even though acting as a toxin, was favored in a population than the loss-of-function allele in laboratory environment. This added fitness could help *peel-zeel* to spread in the population when their frequency was low and genic drive small. When the starting percentage of *peel-zeel* was 1% in a population with high outcrossing rate ($k = 0.7$), it took a neutral *peel-zeel* 150 generations to reach 40% (**Figure 3.6**). On the other hand, with a beneficial *peel-zeel* ($s = 0.06$ and $h = 0$), it only took 100 generations to reach 40%. This difference was even stronger in a population with lower outcrossing rate ($k = 0.7$). It took a neutral *peel-zeel* ~340 generations to reach 40% whereas it only took a beneficial *peel-zeel* ($s = 0.06$ and $h = 0$) ~80 generations to reach 40%. These results showed that the fitness advantage of *peel-zeel* is able to help its spread in the population where their initial frequency is low.

In the Bristol haplotype, *peel-1* is expressed in the sperm and delivered to the embryo. The ability to kill embryo depends on the presence of *zeel-1* that is transiently expressed in the embryo (Seidel et al., 2011). We hypothesized the fitness advantage of *peel-1* is also *zeel-1* dependent since these two genes act closely together. To test this hypothesis, we used CRISPR-Cas9 to introduce a stop codon into *zeel-1* in the *peel-1*$^{-/-}$ background. When we compete this double loss-of-function against *peel-1* single loss-of-function either directly or use N2 as a shared control, we found no fitness differences between them (**Figure 3.5F**). This result suggests the fitness advantage conferred by *peel-1* is *zeel-1* independent.

3.2.4 srbc-64 and ugt-31 are not responsible for the fitness difference

With *peel-zeel* only accounts for 32% of the fitness difference for the NIL strain, we hypothesized additional polymorphisms in the hyper-divergent region might contribute to the rest of the fitness difference. In the hyper-divergent region of the Hawaii haplotype, *peel-1* and *zeel-1* were deleted. Four other protein coding genes were either partially (*nekl-1*) or entirely (*srbc-64, ugt-31, Y39G10AR.16*) included in the hyper-divergent region (**Figure 3.8A**). *srbc-64, ugt-31 and Y39G10AR.16* have amino acid identity less than 95% while nekl-1 have amino acid identity greater than 99%. *srbc-64* was shown previously to be a G protein-coupled receptor that regulates dauer larval development (Kim et al., 2009). Dauer larvae were central to the *C. elegans* lifestyle (Fielenbach and Antebi, 2008). Also, previous genome wide studies of high Tajima' D region showed an enrichment for chemoreceptor genes (Greene et al., 2016b). We believe it was a very promising candidate gene for the fitness difference. *ugt-31* is a less explored gene and it is predicted to have UDP-glycosyltransferase activity (Kim et al., 2018). Human UDP-glucuronosyltransferase (UGT) play a key role in the removal of non-polar drugs. The major detoxification organ liver have abundant UGT proteins expressing (Rowland et al., 2013). Previously, we performed RNA-seq on N2-CB4856 hybrid F1s and *ugt-31* has shown significant allele specific expression difference (**Figure 3.7A**). Across the SNPs on the exons of *ugt-31*, the expression of Bristol allele is about twice as much as the Hawaii allele, implying *cis*-regulatory elements difference between the Bristol and Hawaii alleles of *ugt-31*. Hybrid F1

RNA-seq also showed allele specific expression for *Y39G10AR.16.* Across the SNPs on the exons, the expression of the Bristol allele is about 8 times more than that of Hawaii allele (**Figure 3.7B**). However, this gene has no predicted homology in vertebrate genomes. For this reason, we excluded this gene from further analysis. Based on this information, we decided to focus on *srbc-64* and *ugt-31* to test its fitness effect.

Using CRISPR/Cas9, we introduced stop codons to both haplotypes of *srbc-64* and *ugt-31* (**Figure 3.8B, C**). Pairwise competition assays showed that *srbc-64* or *ugt-31* loss-of-function in the Bristol haplotype does not change animal fitness significantly (**Figure 3.8C**). This is also the case in the Hawaii haplotype. We further confirmed this result by competing single gene loss-of-function in both haplotypes. If a certain gene was responsible for the fitness difference between these two haplotypes, creating loss-of-function mutants in both haplotypes for the same gene should be able to eliminate the fitness difference. Again, we did not see a significant fitness change, confirming that there was no fitness difference between the two alleles of *srbc-64* or *ugt-31*. In summary, these results showed that *srbc-64* or *ugt-31* do not contribute to the fitness difference between the Bristol and Hawaii haplotypes.

Two very promising candidate genes *srbc-64* and *ugt-31* did not contribute to the fitness difference between the two haplotypes was quite unexpected, especially consider *srbc-64* was shown previously to have a strong effect in dauer pheromone. Dauer pheromone response has been shown previously to have a strong effect on animal fitness (Greene et al., 2016a, Large et al., 2016). To test if the fitness difference between N2 and NIL strain QX1198 was also attributable to dauer pheromone response difference, we tested N2, QX1198 (*peel-zeel* NIL) and MT13649 (*nurf-1*) to three synthesized components of the dauer pheromone cocktail, ωC3 (ascr#5), C6-MK (ascr#2), and ΔC9 (ascr#3) (Butcher et al., 2007, Butcher et al., 2008). MT13649 animals has been shown previously to have defects in entering dauer in response to C6-MK and ΔC9 but not to ωC3 and was used as a control strain (Large et al., 2016). Similar to MT13649 (*nurf-1*), QX1198 animals entered dauer in response to ωC3 but were resistant to enter dauer in the presence of C6-MK and ΔC9 (**Figure 3.9A**). This dauer pheromone response difference was also very similar to the reported phenotype of *srbc-64 (Tm1946)* mutant (Kim et al., 2009). This result then created a paradox. If *srbc-64* was responsible for the dauer pheromone response difference, why deleting *srbc-64* in the Bristol haplotype did not change animal fitness? We first hypothesized our CRISPR stop codon mutant did not create a loss-of-function allele for *srbc-64*. To test this hypothesis, we put N2, *srbc-64*(*kah190*), which was *srbc-64* CRISPR mutant, and QX1198 in 1500nM C6-MK plates. While QX1198 still showed reduced dauer percentage on the assay plates (30.8%%±11.7%, 95% CI) than N2 (72.3%±4.3%, 95% CI), there is not a significant change for *srbc-64* CRISPR mutant in Bristol allele(72%%±7.8%, 95% CI) (**Figure 3.9B**). This is also the case for Hawaii allele (**Figure 3.9C**). As a control, we also compared C6-MK response of *srbc-64* (*Tm1946*) allele that has the second exon and flanking introns deleted and was previously shown to be defective in C6-MK response (Kim et al., 2009). To our surprise, dauer formation was again not affected by this allele (**Figure 3.9D**). This means in our experiment system, *srbc-64* was not responsible for dauer pheromone response. To find the gene that was responsible for the pheromone response difference between QX1198 and N2, we tested *ugt-31, peel-1* and *zeel-1* CRISPR mutants. None of these show a significant difference (**Figure 3.9B, C, D, E**). In summary, these results showed the NIL strain QX1198 contains polymorphism for dauer pheromone response and the polymorphism was not in *srbc-64*, *ugt-31*, *peel-1 or zeel-1*.

Table 3.1 Family based population dynamics model of *peel-zeel*

Family	Mating types		Frequency	hermaphrodite fitness	Offspring genotype		
	Sire	Dam			PP	P+	++
1	PP	PP	$X_{pp}X_{pp}k$	1-s	1		
2	P+	PP	$X_{p+}X_{pp}k$	1-s	0.5	0.5	
3	++	PP	$X_{++}X_{pp}k$	1-s		1	
4	PP	P+	$X_{pp}X_{p+}k$	1-hs	0.5	0.5	
5	P+	P+	$X_{p+}X_{p+}k$	1-hs	0.25	0.5	0.25(1-t)
6	++	P+	$X_{++}X_{p+}k$	1-hs		0.5	0.5
7	PP	++	$X_{pp}X_{++}k$	1		1	
8	P+	++	$X_{p+}X_{++}k$	1		0.5	0.5(1-t)
9	++	++	$X_{++}X_{++}k$	1			1
10	PP selfing		$X_{pp}(1-k)$	1-s’	1		
11	P+ selfing		$X_{p+}(1-k)$	1-h’s’	0.25	0.5	0.25(1-t)
12	++ selfing		$X_{++}(1-k)$	1			1

Symbol P in Genotype denotes the paternal effect toxin-antidote system while ‘+’ denotes the haplotype without the toxin-antidote. Symbol X denotes the ratio of a certain genotype in a population. The parameter k specifies the outcrossing rate. When k = 1, there is completely outcrossing, and partial outcrossing is given by 0<k<1. The parameter s is the degree *peel-zeel* reduce hermaphrodite fitness in outcrossing. Parameter s’ denotes the fitness cost in self-fertilization. Dominance of the fitness cost is defined by h. The parameter t is modeling the paternal effect lethality. In the *peel-zeel* case, t is known to be close to 1.

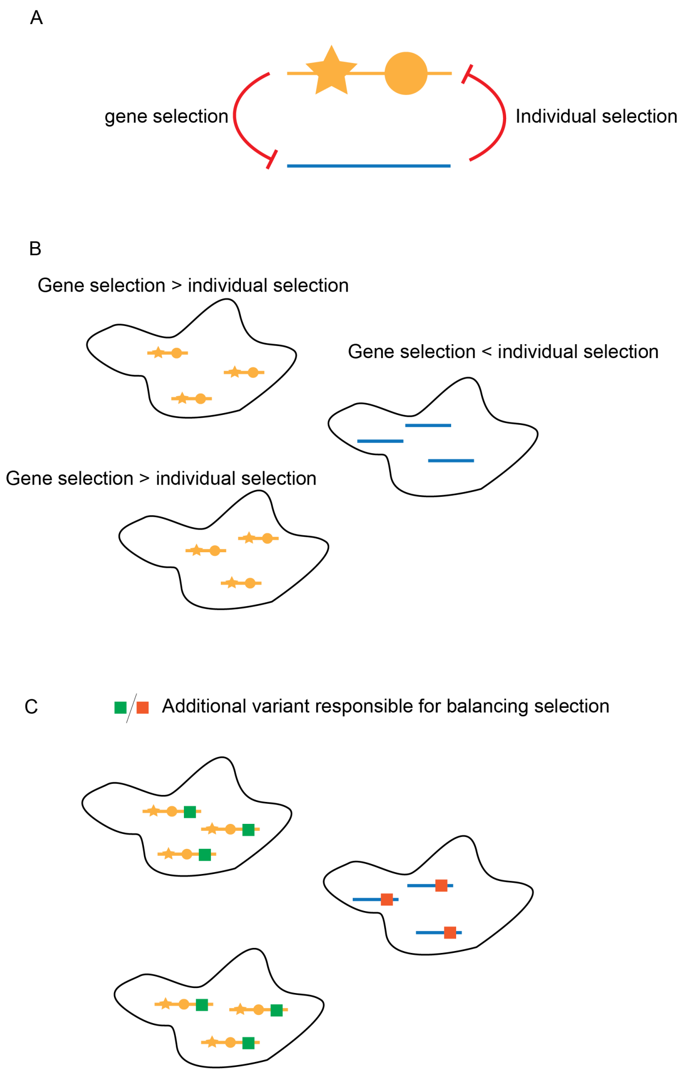


Figure 3.1 Hypothesized models for balancing selection of *peel-zeel*

A. The genic drive is counteracted by a reduced fitness in the population. B. Outcrossing rate and fitness cost is different to each niche, such that *peel-zeel* can be fixed in some locations while purged in others. C. Another linked mutation of *peel-zeel* drive balancing selection.

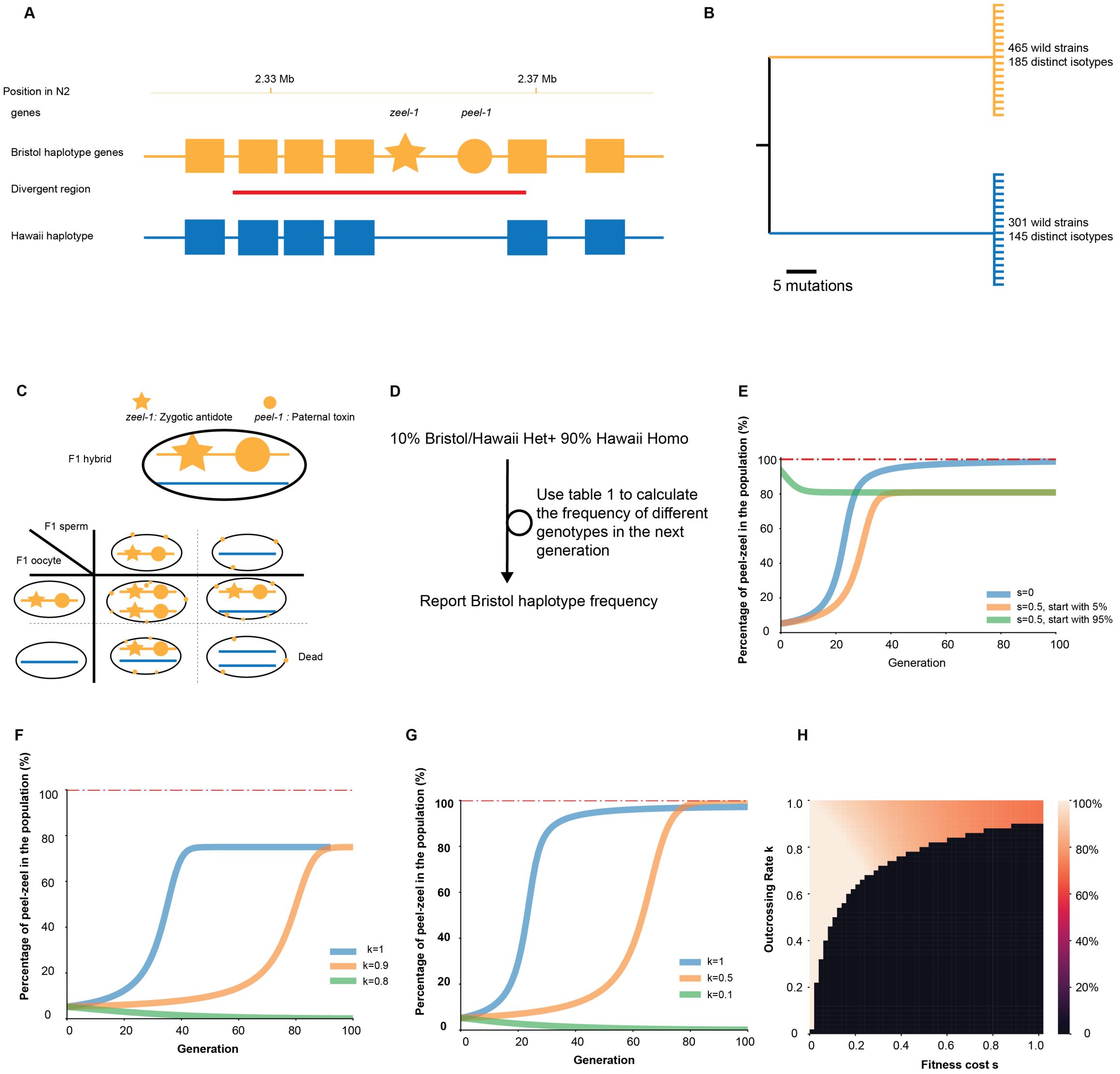


Figure 3.2 ***peel-zeel*** **can be balanced in a population if they incur fitness cost to the animals carrying it.**

A. A schematic graph showing the *peel-zeel* locus and the two major haplotypes. Two major haplotypes exist in *C. elegans* wild isolates genomes: the Bristol haplotype (yellow) and the Hawaii haplotype (blue). There is a divergent region (red line) that marks more than 5% divergence in the protein coding regions. It covers the full coding region of *srbc-64* and part of *nekl-1*. *peel-1* (yellow circle) is 2.5kb from *zeel-1* (yellow star) on the Bristol haplotype and is deleted on the Hawaii haplotype. **B.** Schematic of the phylogeny of the *peel-zeel* locus (regenerated from (Seidel et al., 2008)). Two major branches distinguish these two haplotypes and the number of wild isolates and isotypes are labeled on each branch. Scale bar is 5 mutations for the *srbc-64* gene. **C**. A model for driver activity of *peel-zeel*. In the self-fertilization progenies of a F1 hybrid between Bristol haplotype and Hawaii haplotype, progenies on the bottom left corner were killed because toxin *PEEL-1* (yellow circle) is expressed in the germline and deposited into sperms, whereas antidote *ZEEL-1* (yellow star) is expressed zygotically and moved to cell membrane. **D.** Schematic of the modeling. **E.** Model predicts an equilibrium frequency for *peel-zeel* if they incur fitness disadvantage to the animals carrying it. Without fitness disadvantage (blue curve, s=0, h=0), *peel-zeel* will be fixed in a fully outcrossing population in about 100 generations. With fitness disadvantage (yellow and green curve, s=0.5, h=0), *peel-zeel* frequency will converge to an equilibrium point at 80% regardless of its starting frequency in the population. **F.** Outcrossing rate only affect the rate of spread for *peel-zeel* when its fitness disadvantage is high. In highly outcrossing populations (k = 1 or k= 0.9), *peel-zeel* reaches the same equilibrium frequency when it imposes strong fitness disadvantage to the animals carrying it (s = 0.75, h = 0). **G.** Outcrossing rate affect both the rate of spread and the outcome for *peel-zeel* when its fitness disadvantage is low. When outcrossing rate is high (k=1), *peel-zeel* can be balanced in a population (blue curve, s=0.05, h=0). When outcrossing rate is intermediate, *peel-zeel* will be fixed in the population whereas in a population with low outcrossing rate, *peel-zeel* will be eliminated from the population. **G.** *peel-zeel* can be balanced in the highly outcrossing population when it imposes fitness disadvantage to the animals carrying it. In the heatmap, each block is showing the percentage of Bristol haplotype in the population after evolving 1,000 generations given the combination of two parameters, outcrossing rate k and fitness disadvantage s.

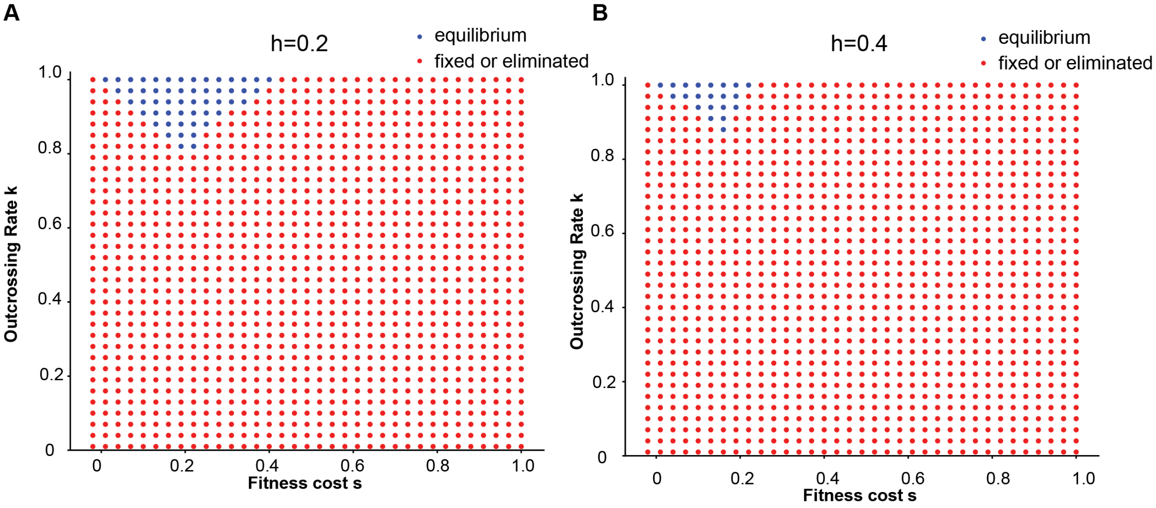


Figure 3.3 Increasing fitness dominance parameter h decreases parameter space for peel-zeel to be balanced.

A. Parameter space for peel-zeel to be balanced. Combination of two parameters, outcrossing rate k and fitness disadvantage s, determines whether *peel-zeel* can be balanced (blue dots) or fixed or eliminated (red dots) in the population (h=0.2). **B.** Parameter space for peel-zeel to be balanced. Combination of two parameters, outcrossing rate k and fitness disadvantage s, determines whether *peel-zeel* can be balanced (blue dots) or fixed or eliminated (red dots) in the population (h=0.4).

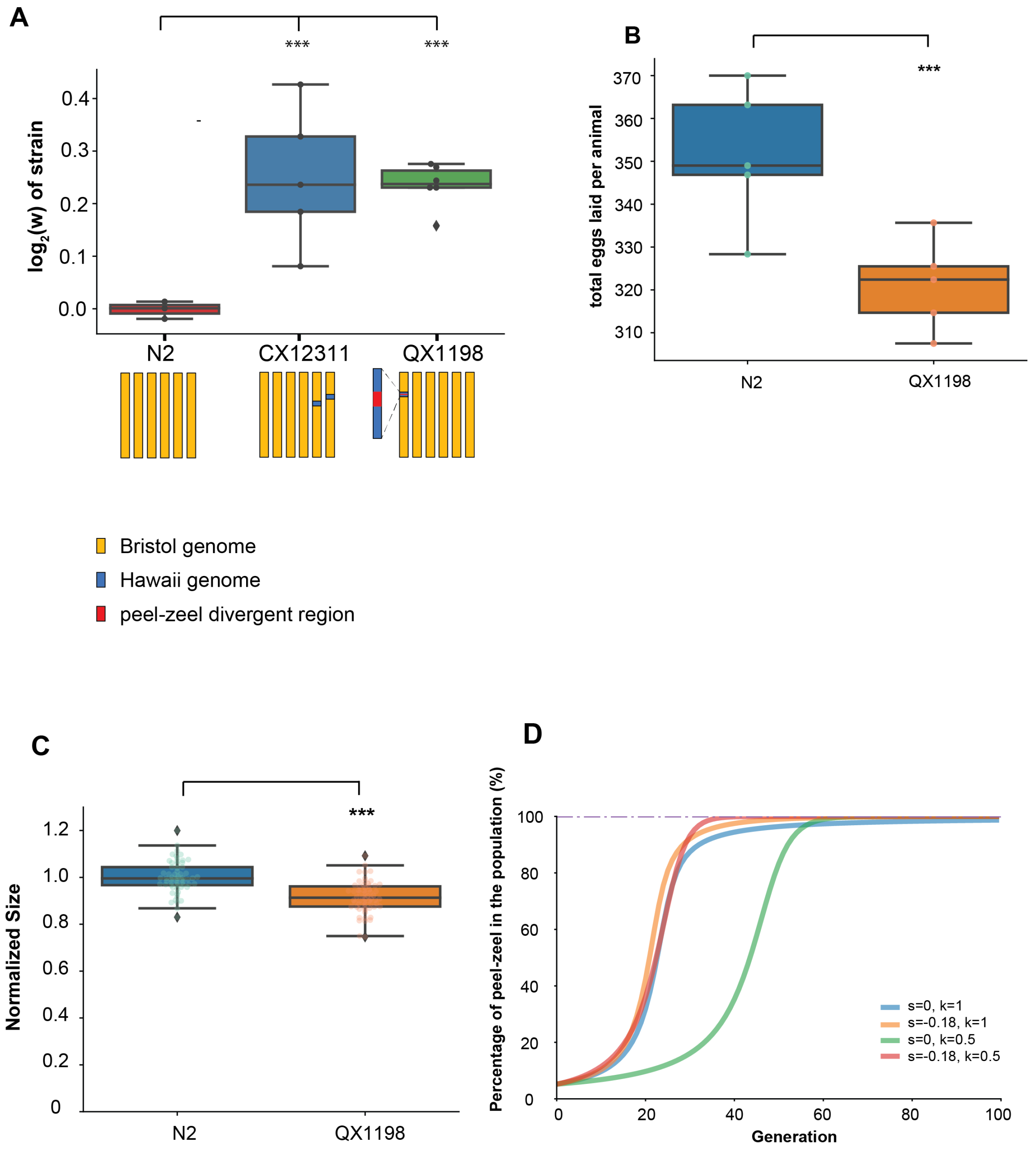


Figure 3.4 The *peel-zeel* is linked to genetic variation that increases laboratory fitness in hermaphrodites.

A. Competition assay showing the relative fitness of PTM229 to strains on the x-axis. PTM229 has the barcoded allele of *dpy-10.* This barcode allele does not change fitness and can be used as a marker to distinguish the two strains in the competition assay. CX12311 has ancestral alleles of *npr-1* and *glb-5* introgressed from Hawaii genome (blue box). This strain has been shown previously to be less fit than N2 in laboratory conditions. QX1198 has a 140kb-370kb Hawaii genome introgression (blue box) in an otherwise N2 genome, covering the hyper-divergent region (40kb, red box) between Bristol haplotype and Hawaii haplotype. Positive values indicate PTM229 is more fit; negative values indicate strain on the x-axis is more fit. The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. ***p<0.001 by two-tailed t-test with equal variance. **B.** Total number of eggs laid by each animal of N2 and NIL strain QX1198. The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. ***p<0.001 by two-tailed t-test. **C.** Growth analysis of NIL strain QX1198. Animals' size was normalized to the average size of N2 when grow at 20°C for 72 hours. The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. ***p<0.001 by two-tailed t-test. **D.** Modeling *peel-zeel* dynamics in a population when linked variants are beneficial. The population is started with 10% *peel-zeel* heterozygotes and 90% non-*peel-zeel* homozygotes. Note that linked beneficial variants (s=-0.18, h=0) only slightly accelerated fixation rate of *peel-zeel* in a fully outcrossing population (k=1, blue curve vs yellow curve), whereas fixation rate is greatly accelerated in a partial outcrossing population (k=0.5, green curve vs red curve).

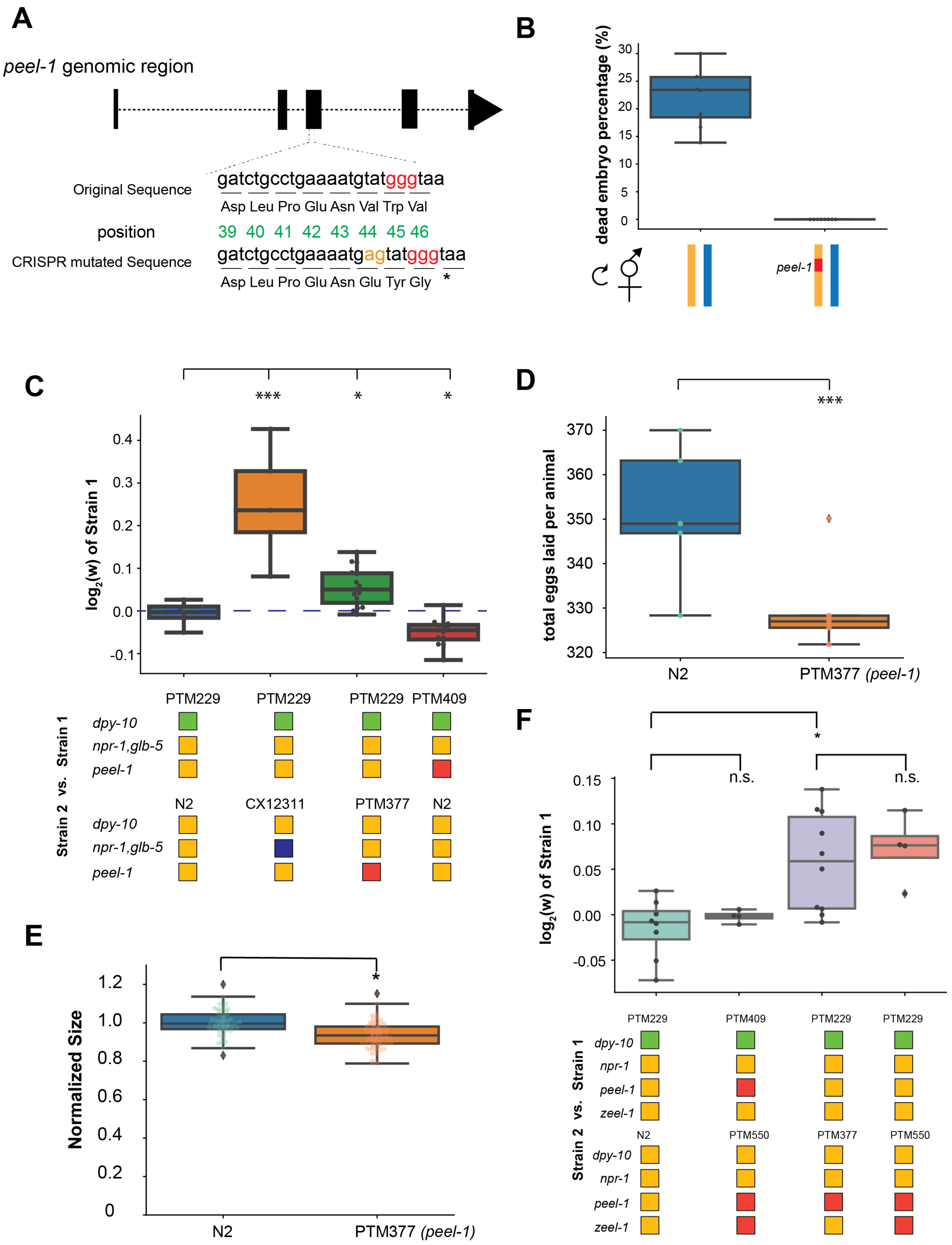


Figure 3.5 Loss of the toxin gene *peel-1* decreased laboratory fitness in hermaphrodites.

A. Schematic of the *peel-1* loss-of-function allele created by CRISPR/Cas9. GGG(marked in red) is used as the PAM site and two additional nucleotides (marked in yellow) is inserted into the third exon of *peel-1* gene through homologous recombination. This insertion creates a frameshift and a stop codon (marked by *) is introduced two amino acids 3' to the insertion site. The resulting protein sequence is 46 amino acids long. (*peel-1* is 174 amino acids long). The green numbers denote the amino acid position of the peel-1 protein sequence. **B**. Percentage of dead embryos from selfing N2(yellow line)xCB4856(blue line) F1 and PTM377(*peel-1* mutant, yellow line with red block)xCB4856(blue line) F1. N2 has the Bristol haplotype and CB4856 has the Hawaii haplotype of *peel-zeel*. The box plot shows quartiles of proportions of embryos that fail to hatch while the whiskers show the rest of the distribution unless the point is an outlier. **C.** Competition assay showing the relative fitness between pairs of strains in laboratory environment. PTM229 has the barcoded allele of *dpy-10* (green box). This barcode allele does not change fitness and can be used as a marker to distinguish the two strains in the competition assay. CX12311 has ancestral alleles of *npr-1* and *glb-5* introgressed from Hawaii genome (blue box). This strain has been shown previously to be less fit than N2 in laboratory conditions. PTM377 has the *peel-1* CRISPR mutant in the N2 background. PTM409 has both the dpy-10 barcode and *peel-1* mutant. Positive values indicate Strain 1 is more fit; negative values indicate Strain 2 is more fit. The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. ***$p<0.001$, *$p<0.05$ by two-tailed t-test with equal variance. **D.** Total number of eggs laid by each animal of N2 and PTM377 (*peel-1*). The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. ***$p<0.001$ by two-tailed t-test. **E.** Growth analysis of NIL strain PTM377 (*peel-1*). Animals' size was normalized to the average size of N2 when grow at 20°C for 72 hours. The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. *$p<0.05$ by two-tailed t-test. **F.** Competition assay showing the relative fitness between pairs of strains in laboratory environment. PTM229 has the barcoded allele of *dpy-10* (green box). This barcode allele does not change fitness and can be used as a marker to distinguish the two strains in the competition assay. PTM377 has the *peel-1* CRISPR mutant in the N2 background. PTM409 has both the dpy-10 barcode and *peel-1* mutant. PTM550 has both *peel-1* and *zeel-1* CRISPR mutant in the N2 background. Positive values indicate Strain 1 is more fit; negative values indicate Strain 2 is more fit. The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. *$p<0.05$ by two-tailed t-test. n.s. not significant.

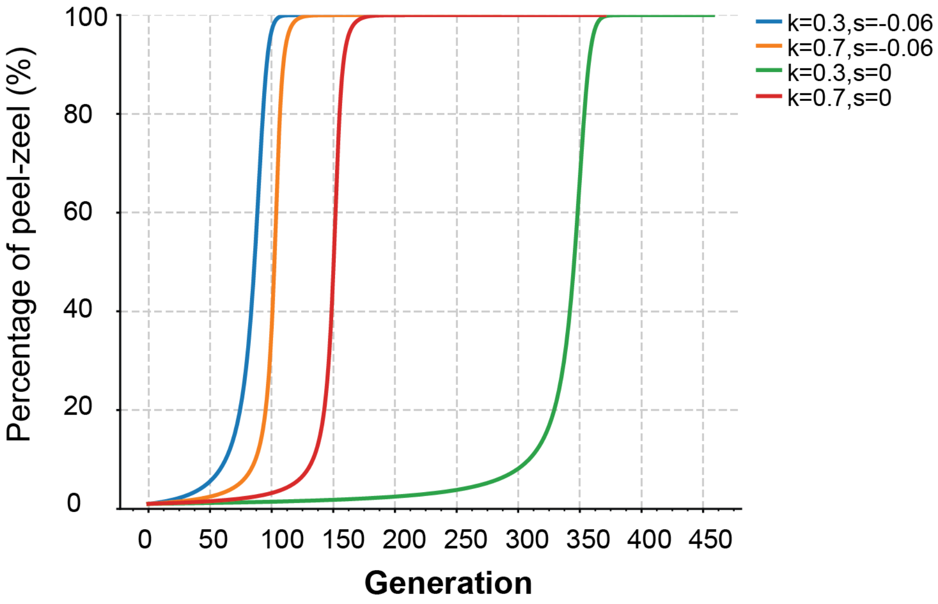


Figure 3.6 Fitness advantage of *peel-zeel* accelerates transmission when starting frequency is low.

y-axis is the percentage of *peel-zeel* in a population at a particular generation in the x-axis. Outcrossing rate is modeled by parameter k and fitness advantage is modeled by parameter s. In this simulation, dominance parameter h=0 and the population is initiated with 1% peel-zeel.

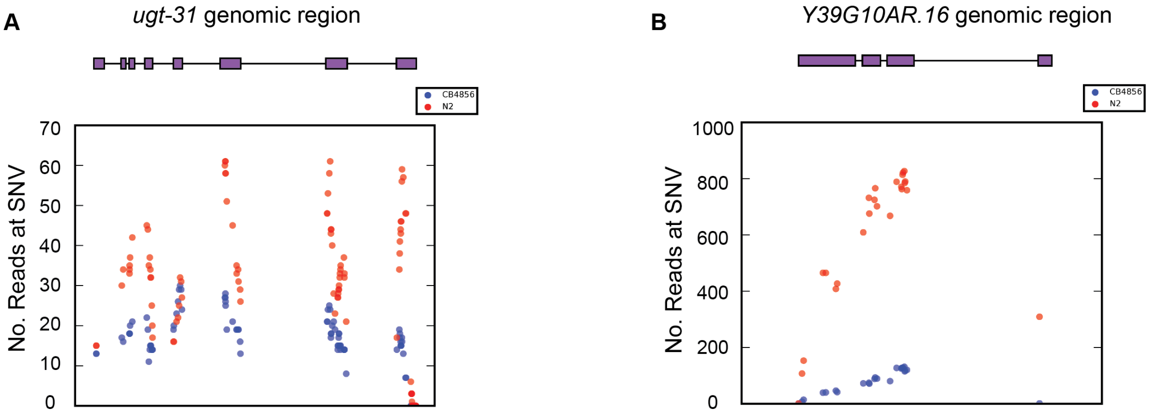


Figure 3.7 F1 hybrid RNA-seq reveals allele specific expression for *ugt-31* and *Y39G10AR.16*.

A. Allele specific expression for *ugt-31*. Above the box shows the *ugt-31* gene structure. It aligns with the x-axis of the plot in the box. y-axis shows the number of allele specific reads at a particular single nucleotide variant (SNV). Red dots denote N2 specific reads and blue dots denote Hawaii strain CB4856 specific reads. **B.** Allele specific expression for *Y39G10AR.16*. Above the box shows the *Y39G10AR.16* gene structure. It aligns with the x-axis of the plot in the box. y-axis shows the number of allele specific reads at a particular single nucleotide variant (SNV). Red dots denote N2 specific reads and blue dots denote Hawaii strain CB4856 specific reads.

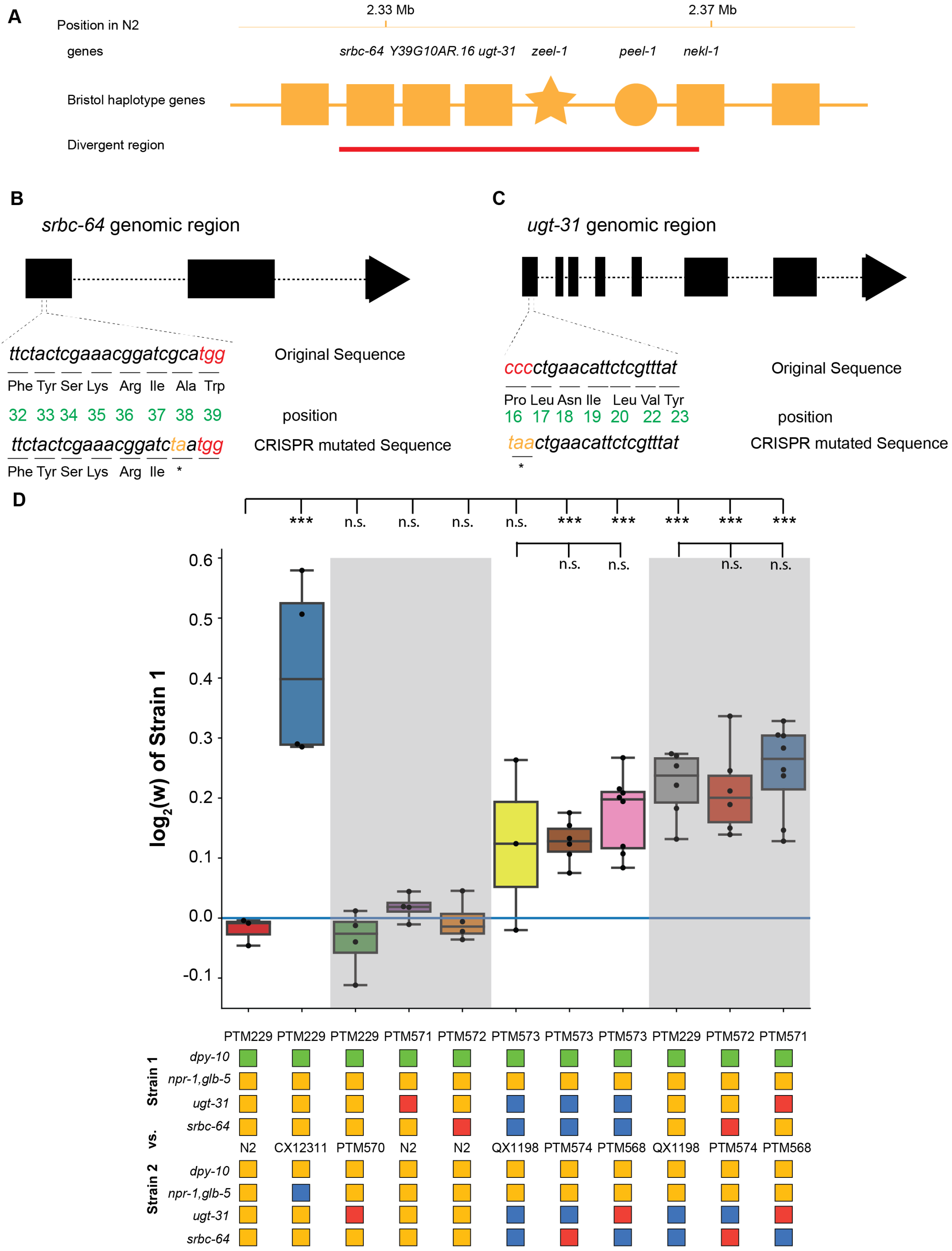


Figure 3.8 ***srbc-64*** **or** ***ugt-31*** **does not change laboratory fitness in hermaphrodites.**

A. A schematic graph showing the *peel-zeel* locus. There is a divergent region (red line) that marks more than 5% divergence in the protein coding regions. It covers the full coding region of *srbc-64* and part of *nekl-1*. **B**. Schematic of the *srbc-64* loss-of-function allele created by CRISPR/Cas9. TGG (marked in red) is used as the PAM site and two nucleotides (marked in yellow) is replaced in the first exon of *srbc-64* gene through homologous recombination. Amino acid This mutation creates a stop codon (marked by *). The resulting protein sequence is 37 amino acids long (*srbc-64* is 289 amino acids long). This mutation is introduced to Bristol haplotype and Hawaii haplotype independently. **C**. Schematic of the *ugt-31* loss-of-function allele created by CRISPR/Cas9. Ccc (marked in red) is used as the PAM site and three nucleotides (marked in yellow) is replaced in the first exon of *ugt-31* gene through homologous recombination. This mutation creates a stop codon (marked by *). The resulting protein sequence is 15 amino acids long (*ugt-31* is 521 amino acids long). This mutation is introduced to Bristol haplotype and Hawaii haplotype independently. **D.** Competition assay showing the relative fitness between pairs of strains in laboratory environment. PTM229 has the barcoded allele of *dpy-10* (green box) in an otherwise N2 background. This barcode allele does not change fitness and can be used as a marker to distinguish the two strains in the competition assay. Blue box denotes introgression from the Hawaii haplotype. Red box denotes CRISPR mutant. Note that PTM574 and PTM568 have *srbc-64* or *ugt-31* mutated in the Hawaii haplotype, whereas PTM571 and PTM572 have *srbc-64* or *ugt-31* mutated in the Bristol haplotype. Positive values indicate Strain 1 is more fit; negative values indicate Strain 2 is more fit. The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. ***$p<0.001$, *$p<0.05$ by two-tailed t-test with equal variance. n.s. not significant.

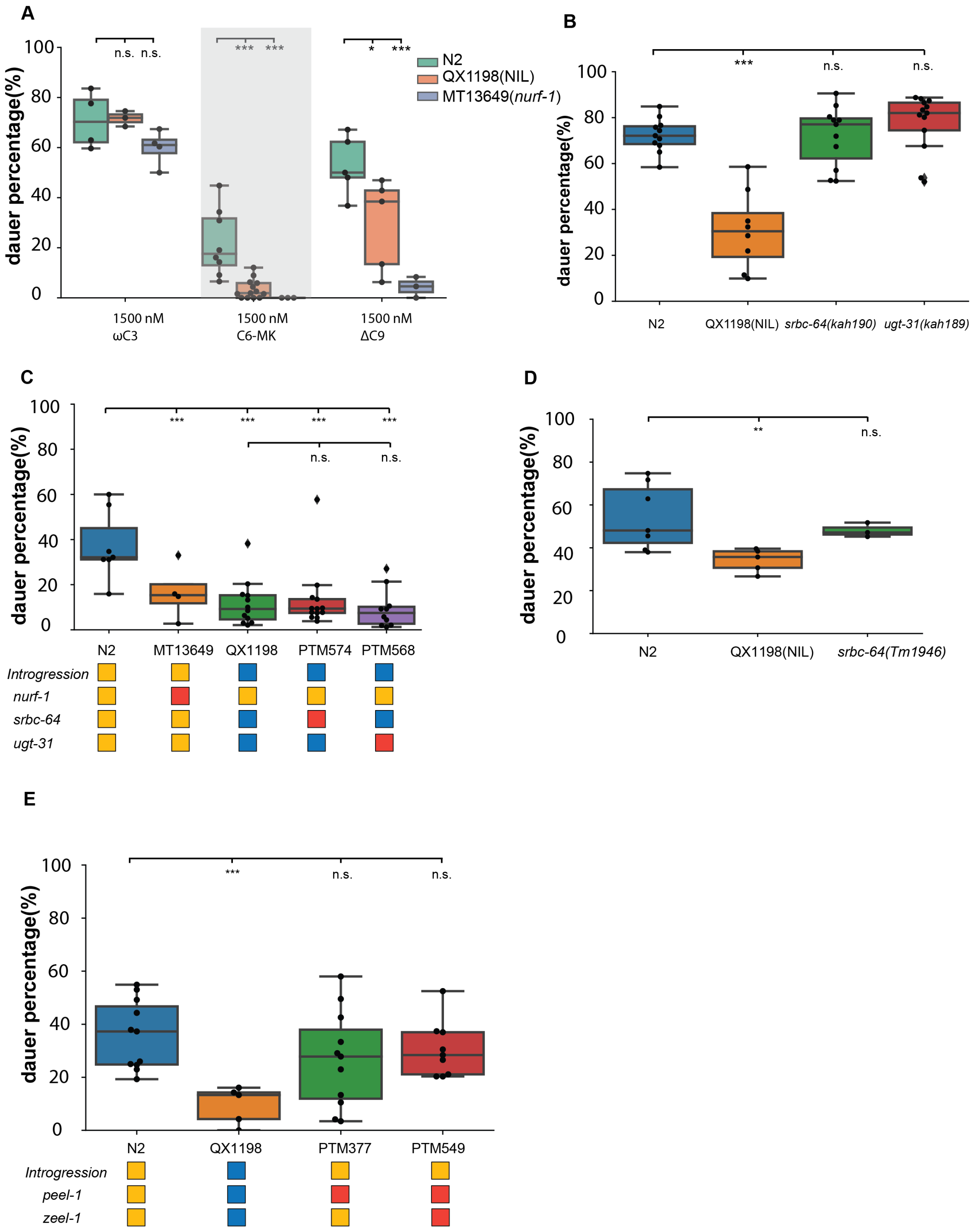


Figure 3.9 *srbc-64*, ugt-31, *peel-1* and *zeel-1* are not responsible for the dauer pheromone response difference between N2 and QX1198.

A. Dauer formation of the N2, QX1198 (*peel-zeel* NIL strain) and MT13649 (*nurf-1*) in response to synthesized pheromone components. y-axis indicates the percentage of animals that enter dauer in response to the pheromone concentrations listed on the x-axis. The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. ***p<0.001, *p<0.05 by two-tailed t-test with equal variance. n.s. not significant. **B.** Dauer formation of the N2, QX1198 (peel-zeel NIL strain) and *srbc-64*(*kah190*) and *ugt-31(kah-189)* in response to 1500 nM C6-MK. *srbc-64*(*kah190*) and *ugt-31(kah-189)* are CRISPR mutants created in N2. y-axis indicates the percentage of animals that enter dauer. The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. ***p<0.001, *p<0.05 by two-tailed t-test with equal variance. n.s. not significant. **C.** Dauer formation of the N2, MT13649 (*nurf-1*), QX1198 (peel-zeel NIL strain), PTM574 (*srbc-64*) and PTM568(*ugt-31*) in response to 1500 nM C6-MK. Yellow box denotes the genotype is Bristol like whereas blue box denotes the genotype is Hawaii like. Red box denotes a mutation in that gene. y-axis indicates the percentage of animals that enter dauer. The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. ***p<0.001 by two-tailed t-test with equal variance. n.s. not significant. **D.** Dauer formation of the N2, QX1198 (*peel-zeel* NIL strain) and *srbc-64*(*Tm1946*) in response to 1500 nM C6-MK. y-axis indicates the percentage of animals that enter dauer. The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. **p<0.01 by two-tailed t-test with equal variance. n.s. not significant. **E.** Dauer formation of the N2, QX1198 (*peel-zeel* NIL strain), PTM377(*peel-1* CRISPR mutant) and PTM549(*peel-1*,*zeel-1* double CRISPR mutant).

Yellow box denotes the genotype is Bristol like whereas blue box denotes the genotype is Hawaii like. Red box denotes a mutation in that gene. y-axis indicates the percentage of animals that enter dauer. The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. ***p<0.001 by two-tailed t-test with equal variance. n.s. not significant.

3.3 Discussions

In this study, we tested three models that can induce balancing selection on *peel-zeel*. While quantitative modeling clearly rules out model 1, model 2 and model 3 cannot be completely rejected. Existing evidence that also support model 2 is that wild strains of *C. elegans* differ in their male competence and this difference has a genetic basis (Gimond et al., 2018). Outcrossing rate in the wild is determined by male mating ability. As a result, a variation of outcrossing rate in different environments is expected. Also, when animals adapt to new environments, males in a population increased from 1% to 40% (Anderson et al., 2010). As *C. elegans* in the wild has a boom-and-burst lifestyle, adapting to new environment should not be uncommon. If *peel-zeel* incurs a fitness cost to animals in the wild, this variation in outcrossing rate alone is able to fix *peel-zeel* in some populations and purge them from others. One caveat here is the assumption that the fitness effect of *peel-zeel* will change in wild environments. To completely reject model 2, we need to confirm there is also a fitness advantage for *peel-zeel* in natural environments. Evidence supporting model 3 is that linked variants of *peel-zeel* changed animal fitness. This suggests balancing selection can act on these linked variants. Simulations in chapter 4 showed that *peel-zeel* will act like a neutral genetic variant when outcrossing rate is 1%. So balancing selection does not have to counteract the genic drive in *C. elegans*, though it might have to interact with the fitness effect of *peel-zeel.* At the same time, model 2 and model 3 are not completely exclusive. For example, if animals in a particular environment has very low outcrossing rate, a fitness cost to the animals expressing *peel-zeel* as well as a lower fitness for linked variants on the Bristol haplotype, then both model 2 and model 3 predicts fixation for the Hawaii haplotype.

A toxic gene incurs a fitness advantage to the animal is quite unexpected. We don't know yet if this advantage is acquired during their initial evolution or being coopted by the animals following their initial spread. Different time points of this gained fitness advantage can affect how we interpret the evolution origin and dynamics of *peel-zeel.* When the toxin-antidote element first evolves from its precursor in a population, its frequency is very low. The family-based population genetics model shows that the genic drive is negligible when toxin-antidote element is rare in the population (rate of spread is on the order of p^2). We have shown that this added fitness advantage can greatly accelerate its fixation in the population. Thus, a toxin-antidote element with a fitness advantage will be 'selected' among those without fitness advantage. On the other hand, if this fitness advantage is gained after it spread in the population, this fitness advantage will eliminate those 'cheaters' who only express the antidote and have the toxin eliminated from the genome.

peel-1 functions independent of *zeel-1* implies other molecular pathways are used by *peel-1. C. elegans* follow metapopulation dynamics where competition between groups is a major evolutionary force. If a population has *peel-zeel* segregating, some of the progenies will be killed either at the embryonic stage or larval stage. As a result, this population will grow slower than a population without *peel-zeel* segregating. competition between groups will then tend to eliminate *peel-zeel*, creating a conflict between *peel-zeel* and the rest of the genome. *peel-1,* however, seems to corporate with other genomic elements to improve its fitness. What are the evolutionary force generating the corporation between *peel-1* and the rest of the genome is an interesting question to explore in chapter 4.

3.4 Materials and Methods

3.4.1 Strains used in this chapter

Table 3.2 strains

Strain	Genotype	Comments
N2	Wild type reference	Isolated in Bristol, UK
CB4856	Wild isolate	Isolated from Honolulu, Hawaii
QX1198	qqIr5[niDf9,CB4856>N2] I.	qqIr5 contains a 140-370kb introgression from CB4856 into N2.
CX12311	kyIR1[CB4856>N2] V; qgIR1 [CB4856>N2] X	kyIR1 (V, CB4856>N2) is an introgression of the region surrounding glb-5 from CB4856 into N2 qgIR1 (X, CB4856>N2) is an introgression of the region surrounding npr-1 from CB4856 into N2. Left breakpoint between 4,753,766 and 4,762,579. Right breakpoint between 4,882,488 and 4,885,498
PTM229	dpy-10 (kah82) II	dpy-10 Thr 90: acc -> act.

Table 3.2 Continued

PTM377	peel-1 (kah126) I	Original peel-1 sequence: ATCTGCCTGAAAATGTATGGGTAAAT Mutated peel-1 sequence: ATCTGCCTGAAAATGAGTATGGGTAAAT
PTM409	peel-1(kah126) I; dpy-10 (kah82) II	PTM377 crossed with PTM229 to create this strain.
PTM550	zeel-1(kah181) I;peel-1(kah126) I	zeel-1(kah181) was created by CRISPR/Cas9 original zeel-1 sequence: gccagaccttgaggaggcaaatggtaa Mutated zeel-1 sequence: gccagaccttgagTAAgcaaatggtaa
PTM568	*ugt-31*(kah187) I, qqIr5[niDf9,CB4856>N2] I.	*ugt-31* mutant created using CRISPR/Cas9 in the QX1198 background. Original ugt-31 sequence: ggcctccgagcccctgaacatt Mutated ugt-31 sequence: ggcctccgagTAActgaacatt
PTM570	*ugt-31*(kah189) I	*ugt-31* mutant created using CRISPR/Cas9 in the N2 background. Original ugt-31 sequence: ggcctccgagcccctgaacatt Mutated ugt-31 sequence: ggcctccgagTAActgaacatt Another name: N2-49-13
PTM571	*ugt-31*(kah189) I, dpy-10 (kah82) II	PTM570 crossed with PTM229. Progenies screened for both ugt-31 and dpy-10. Another name: N2-49-13-10b

Table 3.2 Continued

PTM572	*srbc-64*(kah 190) I, dpy-10(kah82) II	*srbc-64* mutant created using CRISPR/Cas9 in the N2 background. Original srbc-64 sequence: gaaacggatcgcatggaaat Mutated srbc-64 sequence: gaaacggatctaatggaaat This strain is then crossed with PTM229 to get dpy-10 barcode. Another name: N2-48-99-10b
PTM573	qqIr5[niDf9,CB4856>N2] I, dpy-10(kah82) II	QX1198 crossed with PTM229 to introduce the dpy-10 barcode. Another name: QX1198-10-b
PTM574	*srbc-64*(kah 191) I, qqIr5[niDf9,CB4856>N2] I	*srbc-64* mutant created using CRISPR/Cas9 in the QX1198 background. Original srbc-64 sequence: gaaacggatcgcatggaaat Mutated srbc-64 sequence: gaaacggatctaatggaaat Another name: QX1198-51-67

3.4.2 Population dynamics prediction

All codes to change population dynamics parameters and then plot the trajectories were stored at https://github.com/LijiangLong/Individual_based_simulation.

3.4.3 Competition assay to measure organism fitness

All pairwise competition assays were done on the 9cm NGM plates. These plates were seeded with OP50 bacteria and stored at 4°C until one day before using them. At the beginning of the experiment, 10 L4 worms of each strain were transferred onto the same plate. This plate was then incubated at 20°C for 5 days until a 1cm×1cm agar was transferred onto a new 9cm NGM plate. The old plate was then washed with 1ml M9 buffer to collect the worms and stored in -80°C for future use. After this point, the plates were transferred every three days for a total of 7 transfers. The genomic DNA of 1st, 3rd, 5th and 7th transfer were isolated using Zymo 96-well DNA isolation kit (cat # D4071). Isolated genomic DNA were fragmentated using EcoRI-HF by incubation at 37°C for 4 hours. These fragmentated DNA were purified using Zymo 96-well DNA purification kit (cat # D4024) using provided protocol. After purification, DNA concentrations were measured using Qubit DNA HS assay and adjusted to 1ng/uL. To be able to distinguish those two strains on a plate, a Taqman probe was designed for the genetic marker on one of the strains using the online tool. After this, the DNA and Taqman probe were mixed with the digital plate PCR (ddPCR) mix and went through standard ddPCR procedure. The fractions of each strain were quantified using the BioRad QX200 machine with standard absolute quantification protocol. After we get the sequential fraction of each strain, a linear regression is applied to fit this data using the following equation with the assumption of one generation per transfer.

$$\log\left(\frac{\frac{p(a)_0}{p(a)_t} - p(a)_0}{1 - p(a)_0}\right) = (log(\frac{W_{aa}}{W_{AA}}))t$$

3.4.4 *Fecundity assay*

All fecundity assays were done at 20°C on 3cm NGM plate. These plates were seeded with 50ul OP50 bacteria with OD_{600} of 2.0. The plates were allowed to try overnight and then transferred to cold room until one day before use. At the beginning of the assay, six fourth larval stage (L4) worms were transferred to each assay plate. The worms were allowed to grow 24 hours before they are transferred onto a new plate. Number of eggs laid were counted using a standard dissecting microscope. This process is repeated every 12 hours thereafter until 100 hours or there is no egg on the new plate. The fecundity is then calculated by first summing over the different time points and then divided by the total number of worms in a single assay plate.

3.4.5 Growth rate assay

Each assay plate is a standard NGM plate seeded with OP50 bacteria. At the beginning of the assay, 10-20 adult worms were transferred onto the assay plate to lay eggs. After 2 hours, they were transferred out of the plate. The eggs were allowed to grow for 72 hours at 20°C. The assay plate is then mounted on a video tracking camera and was recorded for one minute. This video clip was then analysed using a customized MATLAB script that keeps track of each animal and calculate the average size of each worm. The average size from each plate was then normalized by the average size of three N2 plates.

3.4.6 Creating CB4856 and N2 hybrid F1s

10 CB4856 males and 20 N2 (*fog-2*) hermaphrodites were transferred onto each standard 3cm NGM plate seeded with OP50 bacteria. A total of 6 replicates were created. These animals were allowed to mate overnight. Then these N2 (*fog-2*) hermaphrodites were transferred onto a new NGM plate to lay eggs for 3 hours until they were transferred out

of the plate in order to synchronize the progenies. After 60 hours at 20°C, the worms on each assay plate were washed off into a 1.5mL Eppendorf tube using M9 buffer. Finally, 1mL Trizol was added and the tube was kept in -80°C until RNA were extracted.

3.4.7 RNA extraction using Trizol

Prior to RNA extraction, working space was cleaned and then the tubes were freeze and thaw using liquid nitrogen for 6 times. When Trizol was completely thawed, 200 µL chloroform were added into each tube. After mixing through gentle inverting, the tubes were incubated on ice for 5 minutes. Then the tubes were centrifuged at 12,000 g for 15 minutes at 4°C. The upper aqueous phase were transferred into a new Rnase-free tube and 500 µL isopropanol were added. After mixing through gentle inverting, the tubes were incubated on ice for 15 minutes. This was followed by 12,000 g centrifuge for 10 minutes at 4°C. Finally, the RNA pellets were washed twice with 1 mL cold 75% ethanol and then resuspended in 20 µL RNase-free water.

3.4.8 Allele specific expression analysis

RNA libraries were created using Illumina TruSeq Standard mRNA kit (20020595) following its standard protocol. These libraries were sequenced using an Illumina NextSeq 500 Platform. The codes used to analyse the sequencing reads were available at https://github.com/LijiangLong/RNASeq_Nematode. The reads were first quality controlled using FastQC. In order to distinguish N2 and CB4856 allele specific reads and avoid bias introduced by genome alignment, the reads were mapped both to N2 genome and CB4856 genome separately using bwa. For each sample, two bam files were created corresponding to using either N2 or CB4856 as the reference genome.

The number of reads for each allele were then counted at each SNP position. Specifically, the number of N2 specific reads were counted from the N2 genome mapped bam file and the number of CB4856 specific reads were counted from the BC4856 genome mapped bam file. After this, a customized python script was used to plot the gene and the allele specific expression level on each SNP.

3.4.9 CRISPR/Cas9 enabled genome editing

CRISPR/Cas9 was used following a previously published co-conversion method to edit the target gene and *dpy-10* gene at the same time (Arribere et al., 2014).

3.4.10 Synthetic pheromone response assay

For every 250ml NGM solution, 0.25ml 5mg/ml Cholesterol, 0.75g NaCl and 5.5g Noble agar are added in a 1L flask and total volume is adjusted to 250ml by water. After sterilization, NGM solution is cooled down to 60°C while stirring. 0.25 mL 1M CaCl2, 0.25 mL1M MgSO4, and 6.25 mL 1M KPO4 pH 6.0 (freshly prepared = 1.32 mL 1M K2HPO4 + 8.68 mL KH2PO4) are added to the flask along with the desired amount of synthetic pheromone. NGM solution is then pour into separate 3cm plates. These plates are allowed to cool down at 20°C for one hour and then they were wrapped with parafilm and stored at 4°C. At storage, they are not allowed to stay on top of one another to avoid uneven water evaporation. OP50 are heat killed and are also stored at 4°C one day before use. L4 hermaphrodites of each strain are picked one day before use. On the first day of the assay,

assay plates are first dried with lid open for one hour. 50µl OP50 are added to the center of plate and allow to dry for another 30min. 6 hermaphrodites of each strain are transferred to the assay plate to lay eggs for 3 hours at 25°C. These assay plates are then allowed to grow another 72 hours until the number of dauer animals are scored.

CHAPTER 4. SELFISH OR UNSELFISH IS INITIAL FREQUENCY DEPENDENT

4.1 Introduction

In chapter 3, we identified two ways that *peel-zeel* can spread in a population: selfish gene drive and fitness benefits to the hermaphrodites carrying it. *C. elegans* in the wild has a boom-and-burst lifestyle (Frézal and Félix, 2015). Usually, a population is founded by less than ten animals and the population grows several generations, up to 10,000 animals, until food resources are depleted in that habitat. Dauers, a special non-feeding stage, are developed and attach to other animals in order to get transferred to other locations. While most fail, the ones successfully find a new habitat will start another cycle. Consider the two haplotypes of *peel-zeel* are distributed globally and can be sampled from a single orchard, it is very likely a new population can be formed by an admixture of *peel-zeel* and *non-peel-zeel* animals. Also, migration between established populations is also common in *C. elegans* (Barrière and Félix, 2005). This will also likely to bring *peel-zeel* animals to *non-peel-zeel* populations. In the new admixture population, *peel-zeel* has a different transmission pattern than the rest of the genome because of its selfish nature. What will happen to the rest of genome is not explored yet. Answer this question will not only help us understand the evolutionary force that has shaped *C. elegans* genome, but also answer a broader question about selfish genetic elements: is the transmission of selfish genetic elements always at the cost of other genetic elements in the genome?

Evolutionary forces that can change the frequency of a variant include mutation, natural selection, gene flow, genetic drift and genetic hitchhiking. New mutations give rise to the genetic materials for phenotypical variations and thus selection can act on them. Gene flow between population or sometimes between species can also give organisms the genetic materials to adapt to local environment (Jones et al., 2018). Due to the limited size of populations, neutral genetic variants do not always stay in the population. Instead, genetic drifting will change allele frequencies until that allele is either fixed or purged from the population, especially in a small population. Finally, natural selection or being linked to variants under natural selection (genetic draft or hitchhike) will deviate the variant from drifting.

To understand the evolutionary force on the linked and unlinked variants of toxin-antidote elements, one approach is to have experimental populations in the laboratory and keep track of all the allele frequencies of all genetic variants. However, this approach is time-consuming and may not reflect what will happen in the field. Laboratory animals often adapt to the unique environment in the laboratory (Zhao et al., 2018, Large et al., 2016, McGrath et al., 2009). These adaptive changes may have a complex epistatic interaction with the selfish genetic element and other genome regions, such that the dynamics in the laboratory might not predict that in the field. Also, in order to study the evolution of other genetic elements, there has to be neutral genetic differences between the animals carrying or not carrying the toxin-antidote element in order to keep track and compare the dynamics. This is difficult because of two reasons. First, while there are many ways to introduce mutations into the genome, it remains difficult to test if a mutation is truly neutral. Even if the mutation is mapped to a non-coding region, it may or may not influence the expression of another gene, thus affecting the fitness of the organism. Recently, a pairwise competition assay was developed in order to assay the fitness impact of any variant (Large et al., 2016). Such experiments, though accurate, take months of work to eventually confirm if a variant is truly neutral. Second, it is expensive to keep track of multiple genetic variants in a population. Traditional methods such as PCR or restriction enzyme enabled counting is qualitative. Quantitative methods such as digital PCR, however, remains low throughput. High throughput quantitative methods such as while genome sequencing is expensive to both prepare the libraries and get the libraries sequenced. In summary, evaluating the dynamics of both selfish genetic elements and other genomic elements is not practical in laboratory populations.

Another way to predict study evolution is through modeling and simulation. The dynamics of toxin-antidote elements have been well studied using a family based population genetics model (Wade and Beeman, 1994). In this model, only the toxin-antidote elements dynamics is modeled. Both linked and unlinked genetic variants are difficult to model because linkage has to be considered. This then requires expanding the model to a much more complex one that is hard to get analytical solutions and useful conclusions.

Recently, population simulation tools such as SLiM (Selection on linked mutations) are emerging as a powerful and efficient tool to predict evolutionary dynamics of toxin-antidote elements (Haller and Messer, 2019). Researches in Caenorhabditis species have identified five maternal effect toxin-antidote elements in *C. tropicalis*, one in *C. briggsae* and one in *C. elegans*. These elements appear as pairs on the same chromosome, and thus are called antagonizing toxin-antidote elements (Ben-David et al., 2020). In order to study the dynamics of these unique antagonizing elements, they used SLiM to simulate the evolution and found that there is a strong heterozygous advantage for antagonizing toxin-antidote elements such that they can be balancing selected in the population. This matched the theoretical prediction and implies a genetic barrier between animals having antagonizing toxin-antidote elements. SLiM has also been used in other studies for selfish genetic elements, showing great promise to study both toxin-antidote element and other genetic elements at the same time (Champer et al., 2019, Champer et al., 2020).

In this study, we used SLiM to simulate a scenario where homozygous animals carry the toxin-antidote element are mixed with homozygous animals without the toxin-antidote element. In order to keep track of the genome, neutral genetic markers are added and tracked through several generation. We found that the fitness of those neutral genetic markers is initial frequency dependent. A low starting frequency of toxin-antidote elements in the population will reduce the fitness of neutral genetic markers. My results suggest that the transmission conflict between toxin-antidote element and the rest of the genome is initial frequency dependent. When we want to introduce genetically modified mosquitos into the wild, we might want to start with very low frequency such that the genetic diversity of wild strains is not impacted.

4.2 Results and Discussion

4.2.1 Toxin-antidote elements reduce fitness for unlinked genomic elements

One quarter of the progeny from two hybrid parents will be killed or have dramatically reduced fitness (**Figure 4.1A**). This creates a strong gene selection for the toxin-antidote elements and these elements can be fixed in around 100 generations in a fully outcrossing population (**Figure 4.1 C**). To understand the dynamics of both toxin-antidote elements and the rest of the genome, we simulated an admixture process using software SLiM (**Figure 4.1B**). A group of homozygous toxin-antidote animals were mixed with another group of homozygous non-toxin-antidote animals to create a new group of animals and SLiM is simulating the evolution of this new population for 200 generations. To keep track of both toxin-antidote elements and the other genomic elements, we placed the toxin-antidote element at the left tip of the chromosome (position 1 in a 10,000 bp chromosome) and added five neutral genetic markers that were 1kb, 2kb, 3kb, 5kb and 10kb away from the Toxin-antidote element respectively. In our simulations, recombination rate is set to 10^{-4} per base pair per generation, meaning that these five neutral markers are 9, 16, 23, 32, 43 cM away from the toxin-antidote element respectively. Over 200 generations, no new mutations were allowed such that we can accurately keep track of those neutral genetic markers. To prove that we have correctly modeled the behavior of toxin-antidote elements, we compared the dynamics of the toxin-antidote element either using SLiM simulations (population size 10,000) or using a family based population genetic model (Wade and Beeman, 1994). Both are started with 5% toxin-antidote homozygotes and 95% non-toxin-antidote homozygotes. These two curves highly overlap, confirming we had the correct simulation (**Figure 4.1C**). While the toxin-antidote element drives to fixation, neutral genetic markers originally associated with toxin-antidote element also change their frequency in the population (**Figure 4.1D, E**). Depending on the distance to the toxin-antidote element, markers have different dynamics. The marker at 1kb increased its frequency from 5% to 13% after 200 generations because of genetic hitchhiking. The marker at 2kb did not change its frequency, whereas marker at 3kb, 5kb and 10kb all decreased their frequency in the population significantly (5% to 3%). The decrease of frequency is progressive (**Figure 4.1F**), depending on linkage to the toxin-antidote element. The average frequency for most distant marker at 10kb was 4.9% at 8th generation, 4.3% at 32nd generation and 2.9% at 128th generation.

These decreased frequency for neutral genetic markers can be contrasted with a different simulation where the toxin-antidote element was replaced by a mutation with a fitness advantage ($w = 1.3$) (**Figure 4.2**). The trajectories of the toxin-antidote element and the beneficial mutant were quite similar, both increasing from 5% to 90% in ~50 generations (**Figure 4.2A**). However, closely linked neutral marker (1kb) to the beneficial mutant increased from 5% to 27%, much higher than that with the toxin-antidote element (**Figure 4.2B, C**). The rest of the genome for the beneficial mutant also increase frequencies for the first 20 generations until the linkages were broken by recombination. The marker that is most distant to the beneficial mutant (10kb) still increased from 5% to 10%. While genetic hitchhiking can explain all the increased frequencies, it remains elusive why unlinked markers for toxin-antidote element decreased frequency.

While SLiM was powerful and well accepted in the community for evolution simulation, there could still be bugs in the code. That could lead to the unexpected decrease of frequency for neutral variants. To make sure the observed pattern is not due to errors in the codes, we built a customized python script to mimic the evolution of toxin-antidote element and the rest of the genome. This code was less than 100 lines long because very simple and few rules were enforced. These rules included Wright-Fisher model, toxin-antidote induced progeny kill and recombination between TA element and another marker on the genome. Using this script, we were able to replicate what has been observed in SLiM simulations (**Figure 4.3A**). Because this script took more time, we used parallel computing for small populations (N = 8000) while the SLiM was simulating a larger population (N = 10,000). These simulations were more variable but followed similar trajectories. The rest of the genome were simulated by a marker that had varying genetic distance (10cM, 20cM, 30cM, 40cM and 50cM) to the TA element. After simulating 100 generations, the frequency of partially linked marker (10cM) was higher than its initial frequency (13% vs 5%) (**Figure 4.3B**). For the makers that were 20cM and further away from the TA element, the medium frequencies were always lower than their initial frequency. Again, these results confirmed the simulations generated by SLiM and showed unlinked loci to toxin-antidote element was selected against in the population.

4.2.2 Self-fertilization exaggerates toxin-antidote induced negative fitness to unlinked genomic elements

To better understand how simulation parameters such as population size and outcrossing rate affect the dynamics of neutral genetic markers, we tuned these two parameters and looked at the frequency of the marker that is 10kb (43cM) away from the toxin-antidote element. When population size was reduced to 100, 96 out of 100 simulation had the marker eliminated from the population at 200 generations (**Figure 4.4A**). 53 out of 100 simulations had the marker eliminated when population size was 1,000, whereas this only happen once when the population size was 10,000. When the population size is small, the less frequent allele is likely to loss in the population due to genetic drifting.

Outcrossing rate was another important factor that controls the gene selection for the toxin antidote elements. To understand how outcrossing rate affect toxin-antidote elements and other neutral genetic markers, we simulated four populations with different outcrossing rate (100%, 10%, 1% and 0.1%). For the toxin-antidote element, decreasing the outcrossing rate reduced the genic drive (**Figure 4.4B**). In a fully outcrossing population, almost all (99/100) simulations had the toxin-antidote elements fixed in the population. The probability of fixation was reduced to zero when outcrossing rate is 1%. In this case, toxin-antidote behaved like a neutral variant and drifting in the population. On the other hand, the frequency of neutral genetic marker (10kb, 43cM) decreased most (5% to 0.7%) when outcrossing rate was 10% (**Figure 4.4C**). When outcrossing rate was 1% where the toxin-antidote element was acting neutral, marker frequency also decreased from 5% to 2.2%. This suggests that reduced fitness for unlinked genetic elements were independent of toxin-antidote dynamics. In summary, these results showed that unlinked genetic elements were more likely to be eliminated in population with smaller population size and lower outcrossing rate.

4.2.3 Allele frequency in dead embryos is not always its frequency in the population

To understand the mechanisms underlying the decreased frequency for unlinked genomic elements, we calculated the frequencies of toxin-antidote and other genomic elements in the first three generations (**Figure 4.5A**). Unlinked genomic elements were represented by a second chromosome that freely recombines with the toxin-antidote chromosome. For simplicity, we called this chromosome either ‘TA unlinked chromosome’ or ‘nonTA unlinked chromosome’ depending on where it was in the first generation. After fully outcrossing, animals heterozygous for both the TA chromosome and the marker chromosome were generated and had a frequency of $2pq$ in the population. In the third generation, sperm either with or without toxin were generated by the parents and animals without the TA chromosome would be killed as they lack the antidote. We noticed that in the dead embryos, the frequency of the TA unlinked chromosome was (1+p)/4 instead of its original frequency p in the population. Because of this inequity, the frequency of the TA unlinked chromosome in the surviving animals in the third generation is lower than p if $p < 1/3$ and larger than p is $p > 1/3$ (**Figure 4.5B**). This result suggests a diverging point at 1/3 and predicts that when toxin-antidote animals started with a frequency higher than 1/3, unlinked genomic elements will have increased fitness.

4.2.4 Unlinked genetic elements to toxin-antidote elements have higher fitness if they dominant the population in the first generation

To validate our prediction, we used SLiM to simulate varying starting frequencies of toxin-antidote element and looked at the frequency of unlinked genetic elements after 200 generations (**Figure 4**). In the simulation, population size is set to 1,000 and outcrossing rate is 10%. We noticed that unlinked genetic elements decrease frequency when initial TA elements were less than 1/3 and increase frequency when TA elements were more frequent than 1/3. Specifically, when TA elements started at 10%, marker (10kb) decreased from 10% to 4%. On the other hand, marker increased from 50% to 59.5%, 70% to 69.7% and 90% to 93.9%. The smaller gap at 90% than 50% can be explained by the smaller gap between survived unlinked chromosome and p when p is 90% than p is 50% (**Figure 4.5B**). These patterns can be contrasted with a population without the toxin-antidote element. In this case, marker frequency is usually drifting. When markers start at 10%, 25 out of those 100 simulations have the neutral marker eliminated from the population. Because toxin-antidote elements reduced neutral marker fitness, 68 out of those 100 simulations have the neutral marker eliminated. On the other hand, when neutral marker started at 70% without toxin-antidote element, only 1 of the 100 simulations had the marker fixed in the population. Because toxin-antidote element increased marker fitness, 9 of the simulations had the marker fixed. These results confirmed our prediction for the dynamics of genomic elements and suggests the fitness of other genomic elements are initial frequency dependent.

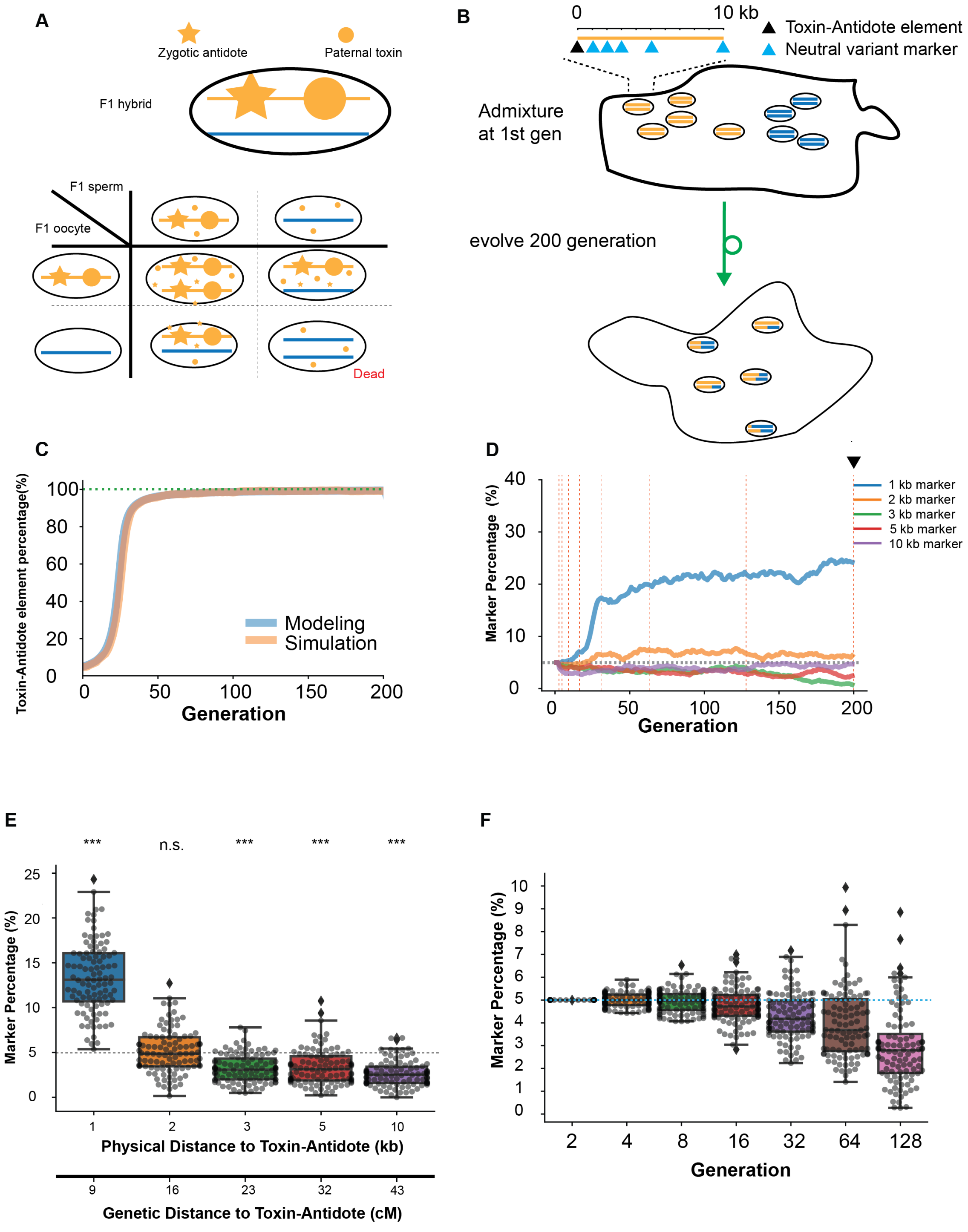


Figure 4.1 Effect of toxin-antidote element on other genetic variation.

A. Schematic of a paternal-effect toxin antidote element. One haplotype (yellow) has both the toxin gene (yellow circle) and the antidote gene (yellow star) while the other haplotype (blue) has neither of them. In the self-fertilization progenies of a F1 hybrid between these two haplotypes, progenies on the bottom left corner were killed because toxin is expressed in the germline and deposited into sperms, whereas the antidote is expressed zygotically. **B.** Schematic of the simulation using SLiM. Chromosomes (yellow and blue line) of 10 kb are simulated. At position 1 (black triangle) of the chromosome with toxin-antidote system (yellow line), there is a toxin-antidote element. Five neutral genetic variants (blue triangle) are introduced at position 1kb, 2kb, 3kb, 5kb, 10kb of the chromosome as genetic marker. Green arrows mark the simulation process. In the first generation, *p* toxin-antidote element homozygotes (upper left, yellow lines in circle) are mixed with (1-*p*) homozygotes without toxin-antidote element (upper right, blue lines in circle). This new population then propagates for another 200 generations with a fixed population size. **C.** Dynamics of the toxin-antidote system in a fully outcrossing population using either a population dynamics model (Blue line) or with SLiM simulation (population size 10,000, Yellow line). The population is initiated with 5% toxin-antidote homozygotes and 95% homozygotes without toxin-antidote. y-axis shows the percentage of the toxin-antidote element in the population at a particular generation on the x-axis. **D.** Dynamics of the neutral genetic markers at different positions relative to the toxin-antidote system (see also **B**) using SLiM simulation (population size 10,000). The population is initiated with 5% toxin-antidote homozygotes and 95% homozygotes without toxin-antidote. y-axis shows the percentage of the marker in the population at a particular generation on the x-axis. Gray dotted line shows their initial frequency in the population at 5%. The frequencies of these neutral genetic markers are recorded at generation 2,4,8,16,32,64,128 and 200 for each simulation (black arrowhead and orange dotted line, see also **E** and **F**). **E.** Neutral genetic marker frequencies after evolving 200 generations with an initial frequency at 5% (n=100) using SLiM simulation (population size 10,000). The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. *** p< 0.001 with two-tailed one sample t-test comparing to the initial frequency at 5%.n.s not significant. **F.** Genetic marker (10kb) frequencies after evolving a number of generations on the x-axis.The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier.

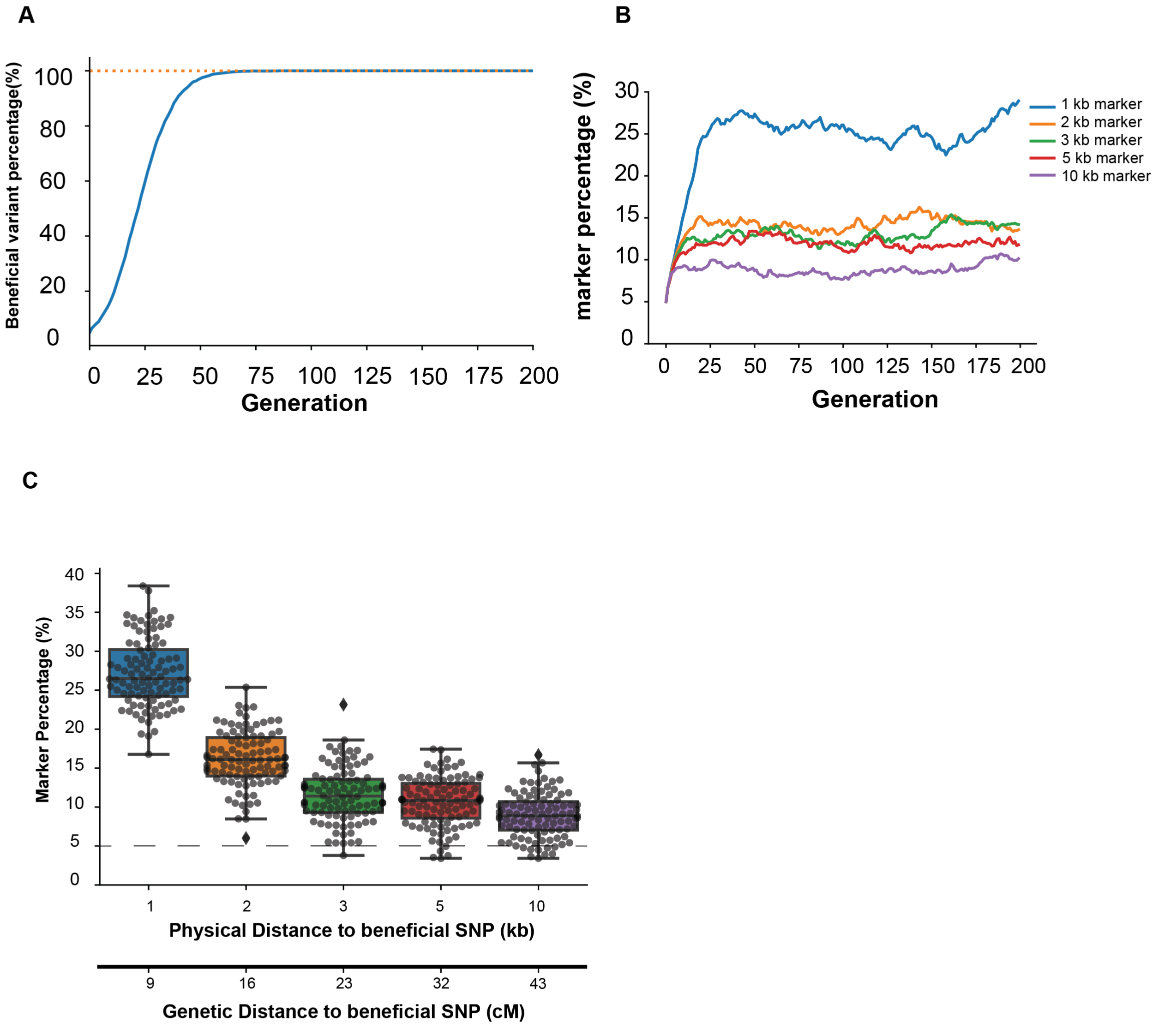


Figure 4.2 Effect of beneficial allele on other genetic variation

A. Dynamics of a beneficial allele in a fully outcrossing population using SLiM (population size 10,000). **B.** Dynamics of the neutral genetic markers at different positions relative to the beneficial allele using SLiM simulation (population size 10,000). The population is initiated with 5% toxin-antidote homozygotes and 95% homozygotes without toxin-antidote. y-axis shows the percentage of the marker in the population at a particular generation on the x-axis. Gray dotted line shows their initial frequency in the population at 5%. The frequencies of these neutral genetic markers are recorded at generation 200 for each simulation. The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier.

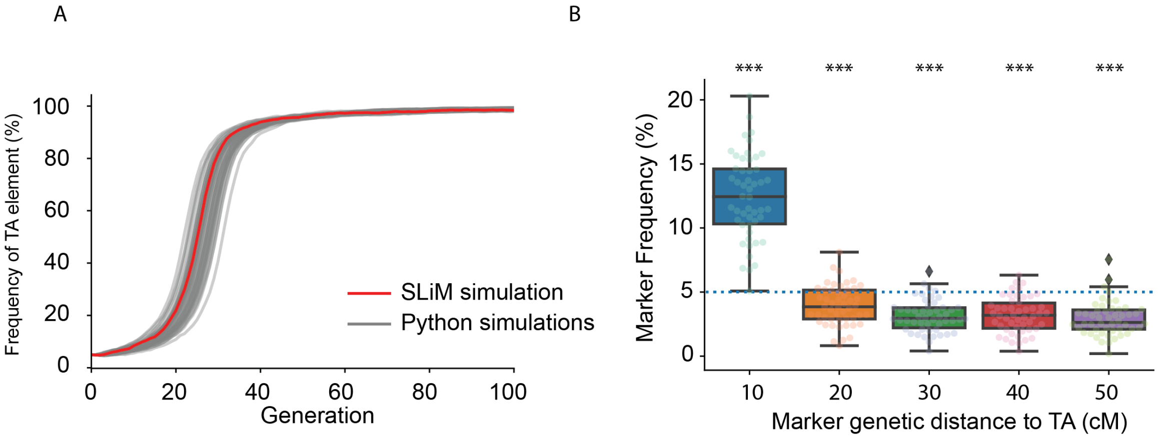


Figure 4.3 Customized python script replicates SLiM simulation results.

A. Dynamics of the toxin-antidote system in a fully outcrossing population using either a customized python script (population size 8,000, gray lines) or with SLiM simulation (population size 10,000, Yellow line). y-axis shows the percentage of the toxin-antidote element in the population at a particular generation on the x-axis. **B.** Frequencies of neutral markers after 100 generations using the customized python script. y-axis shows the percentage of neutral genetic marker in the population (n=100 for each marker). 5% toxin-antidote homozygotes are mixed with 95% homozygotes without toxin-antidote in generation 1 following the simulation scheme shown in Figure 4.1. x-axis shows the genetic distance of the neutral marker to the toxin-antidote element. This population is fixed with a population size of 8,000 and is evolved 100 generations. The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. *** $p < 0.001$ with two-tailed one sample t-test comparing to the initial frequency at 5%.

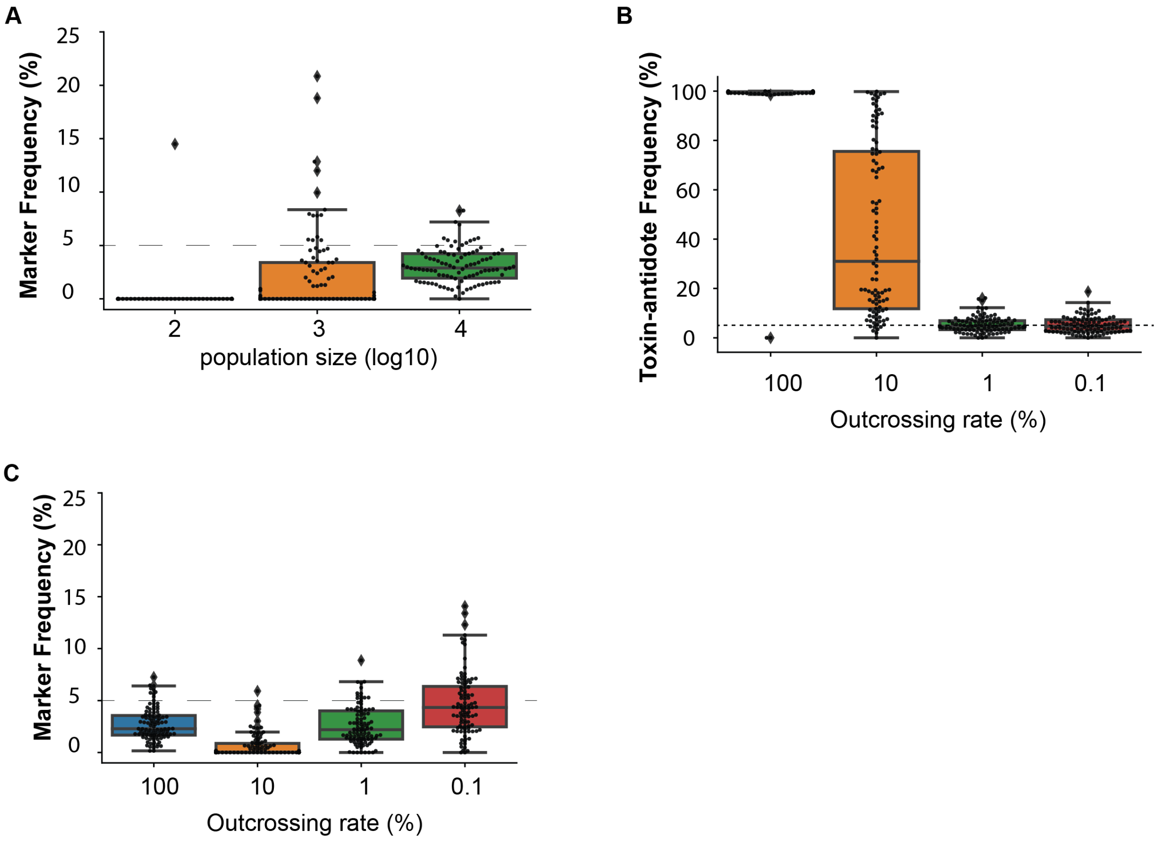


Figure 4.4 Smaller population size and lower outcrossing rate is more likely to eliminate neutral genetic markers.

A. The frequencies of the neutral genetic marker (10kb, 43cM from the toxin-antidote element) after evolving 200 generations with a fixed population size on the x-axis (100 replicates, outcrossing rate 100%). In SLiM simulations, 5% toxin-antidote homozygotes are mixed with 95% homozygotes without toxin-antidote in generation 1. The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. **B.** The frequencies of toxin-antidote element after evolving 200 generations with different outcrossing rate on the x-axis (100 replicates, population size

10,000). The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. **C.** The frequencies of the neutral genetic marker (10kb, 43cM from the toxin-antidote element) after evolving 200 generations with different outcrossing rate on the x-axis (100 replicates, population size 10,000). In SLiM simulations, 5% toxin-antidote homozygotes are mixed with 95% homozygotes without toxin-antidote in generation 1. The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier.

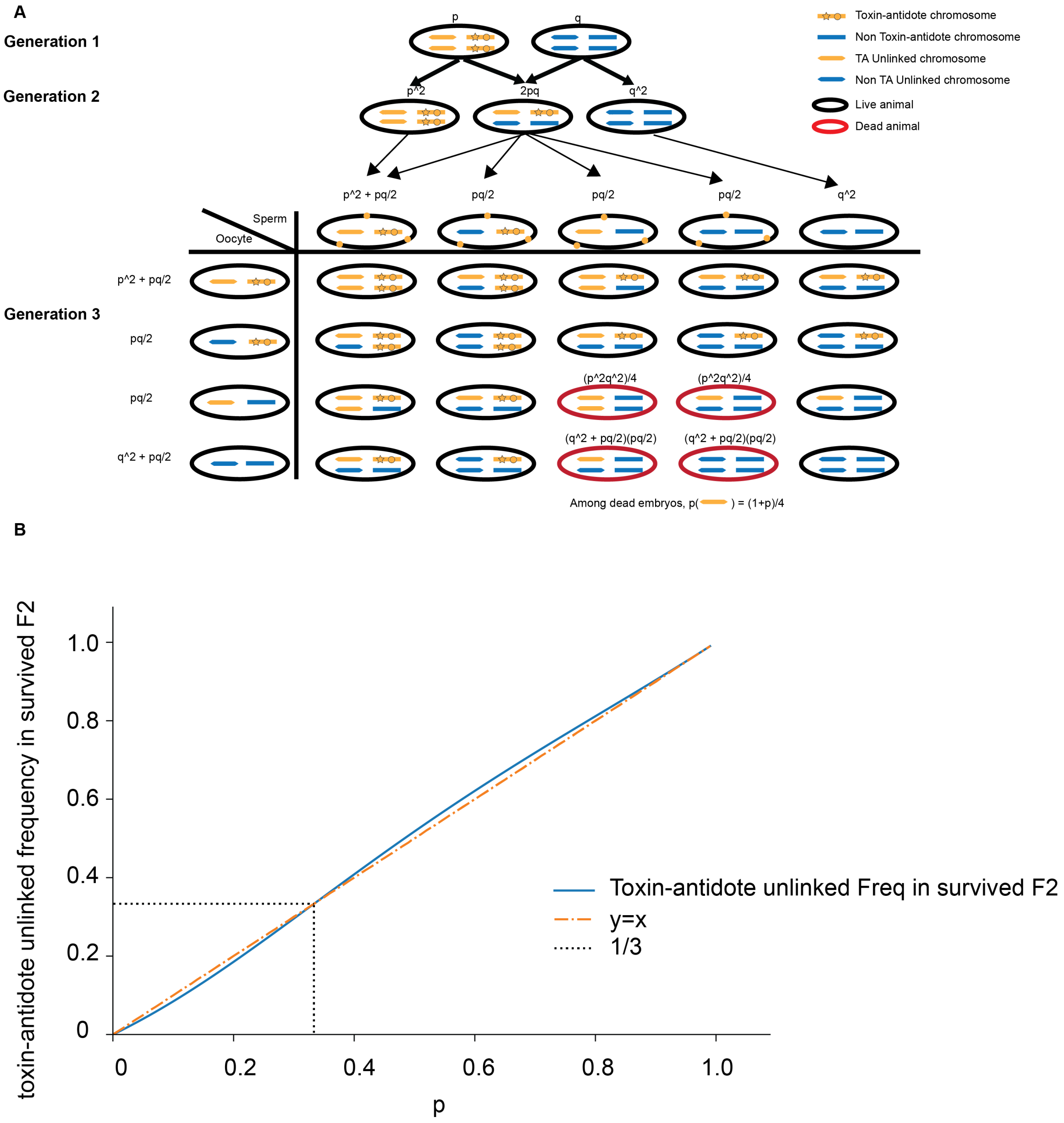


Figure 4.5 The frequency of unlinked polymorphisms to toxin-antidote element in the dead embryos is not its frequency in the population.

A. Schematic of the first three generations of the population. The population is initiated with p diploid individuals who have the chromosomes (yellow lines) carrying the toxin (circle) and the antidote (star) at the first generation. These individuals also carry another pair of neutral effect chromosome (yellow lines with pointed ends). In this population, there are 1-p diploid individuals carrying the non-toxin-antidote chromosomes (blue lines). These non-toxin-antidote individuals also carry another neutral effect chromosome (blue lines with pointed ends). These neutral effect chromosomes serve as marker for unlinked polymorphisms. In the third generation, a Punnett square of the genotypes possible in this generation is shown. Sperms with toxin (in black box) and without toxin are combining freely with oocytes to create F2s. Embryos created by a sperm carrying toxicity and homozygous for the non-toxin-antidote chromosome is dead (red circles) and eliminated from the population. The frequency of each genotype is shown above or beside each circle. Finally, the frequency of unlinked chromosome in the dead embryos is shown at the bottom right corner. **B.** Frequency of toxin-antidote unlinked chromosome in the survived animals of generation 3. y-axis is the frequency of survived animals corresponding to different starting frequency p.

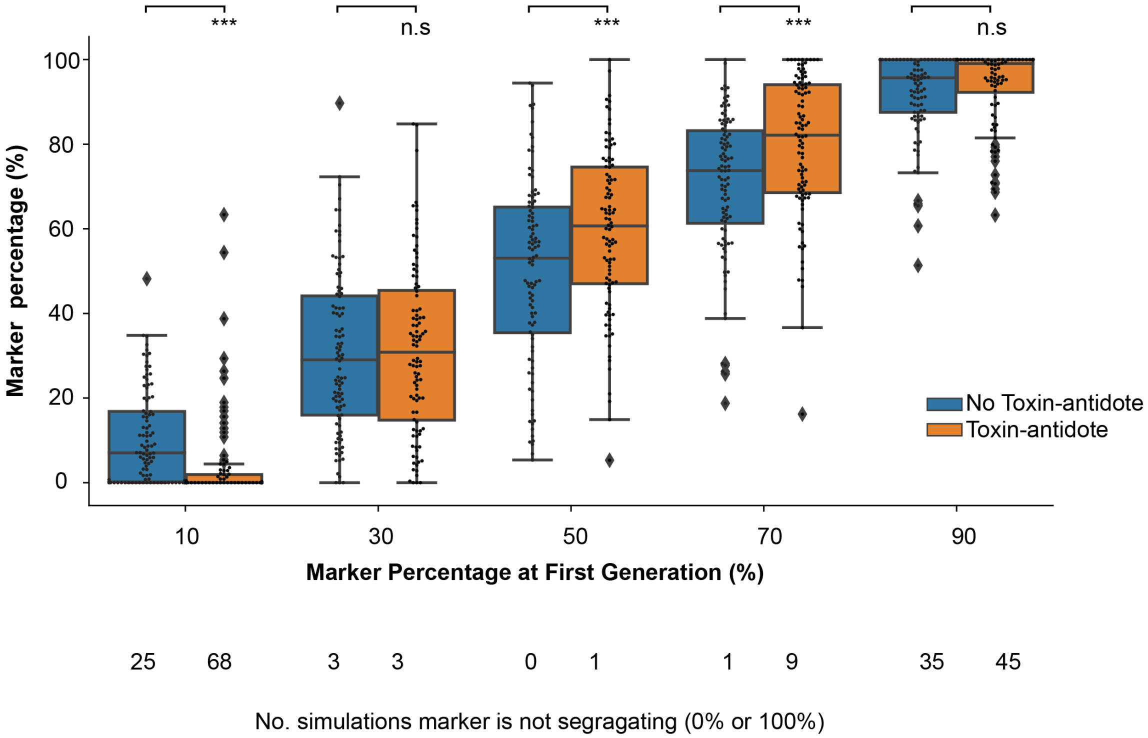


Figure 4.6 The fitness of unlinked genetic elements of toxin-antidote element is initial-frequency dependent.

The percentage of genetic marker in the population after evolving 200 generations in the population with different starting frequency with (yellow) or without(blue) the toxin-antidote element(n=100 replicates for each starting frequency). The y-axis shows the percentage of the neutral genetic marker in the population following the simulation scheme in Figure 4.1 by SLiM. The population size is 1000 and outcrossing rate is 10%. This genetic marker is 10kb away from the toxin-antidote element. The percentage of homozygous toxin-antidote and neutral marker in the first generation is shown on the x-axis. The number of simulations that fix or eliminate the marker in the population is shown at the bottom. The box plot shows quartiles of the dataset while the whiskers show the rest of the distribution unless the point is an outlier. *** $p < 0.001$, * $p < 0.05$, n.s. not significant by two-tailed one sample t-test comparing to the initial frequency respectively.

4.3 Methods

4.3.1 SLiM simulations

Scripts for the SLiM simulations are available at GitHub: https://github.com/LijiangLong/Individual_based_simulation

CHAPTER 5. CONCLUSIONS AND FUTURE STUDIES

My doctoral dissertation includes designing and implementation of novel quantitative methods to study the evolution of speciation. These quantitative methods are useful in large scale behavior comparative analysis, mapping the genetic basis of fitness change and predicting genome evolution.

The work in chapter 2 built a pipeline that is able to automatically detect actions of interest and then classify different actions. Having ten categories in the classification output enabled us to compare the spatial and temporal patterns of different behaviors. For example, in a top-down view, having building scoops in the center and building spits surrounding them means the fish is building a ‘pit’ bower. On the other hand, having building spits in the center and building scoops surrounding them means the fish is building a ‘castle’ bower. I showed that feeding behaviors and bower construction behaviors were expressed throughout the day but at different times, whereas spawning events occurred infrequently in punctuated bursts. These data showed this pipeline can be used to map long term behavioral events in time and space both accurately and with a high recall rate, allowing future studies to unravel the genetic basis of complex behaviors. There are many directions to further extend this work. One of the most important directions is to enable this classification network on animals/environments it has not been trained on. In theory, neural networks are similar to other supervised learning techniques where the ability to generalize is often times impaired by ‘dataset shift’, namely the joint distribution of inputs and outputs differ between the training and test stages. There are two components to the joint distribution difference: 1. Concept shift $P_{train}(y|x) \neq P_{test}(y|x)$ 2. Covariate shift

$P_{train}(x) \neq P_{test}(x)$. Both might decrease the performance when we want to use the same model in the test dataset. Concept shift should be small in our dataset because no matter which animal the model is looking at, the action of the fish given the video clip is very clear. There might be concept shift to categories like 'no fish other' because lighting is different and caused different reflections. On the other hand, there could be a strong covariate shift because different fish have different distribution of actions. Even the same fish could do epochs of feeding and building in a single day. To solve the 'dataset shift' problem, I showed in chapter 2 that label additional video clips for the new animals will increase accuracy. However, this step will take a lot of time and thus is not suitable for large scale behavior study. A future direction could then be applying other machine learning techniques to reduce the 'dataset shift'. One potential approach is to use domain adversarial training. In this approach, there is an additional classifier to the neutral network. This new classifier is designed to tell if a video clip is from the training dataset or a new dataset. The 3D ResNet then tries to fool this new classifier. In this way, the 3D ResNet is forced to learn dataset invariant features. This approach, if properly implemented, is able to streamline the computational pipeline and further boost its capacity and power in behavior studies.

Chapter 3 is about identifying the genetic variant responsible for balancing selection signature around the toxin-antidote element in *C. elegans*. Even though the exact polymorphism/mechanism is not fully identified, an unexpected fitness effect was found for the paternal effect toxin-antidote element *peel-zeel*. In standard laboratory conditions, loss of *peel-1* decreases animal fitness through reduced fecundity and growth rate. Having a positive fitness while producing a toxic protein is unexpected. To my knowledge, this is the first reported toxin-antidote system that impose a beneficial fitness by itself (Rosendahl et al., 2020). This fitness advantage is independent of the antidote gene *zeel-1*, raising a question about the molecular signaling pathways leading to the increased fitness. Even though expression pattern analysis of *peel-1* shows that it is predominantly expressing in the germline, it still remains elusive whether *peel-1* can be expressed in other tissues. Thus, whether *peel-1* functions transgenerational remains an interesting question to ask in future researches. Second, the *peel-zeel* fitness only explains part of the fitness difference between the NIL strain QX1198 and N2, raising another question about what polymorphisms accounting for the rest of the fitness difference are. One potential clue is the fact that this NIL strain shows defective response to dauer pheromone components that is not attributable to a well-known chemoreceptor gene *srbc-64*. Since the NIL strain QX1198 contains a large introgression encompassing tens of genes, a finer genetic mapping strategy is required to find the exact polymorphism accounting for these phenotypical differences. These studies will help us to better understand the evolutionary force that has shaped the *C. elegans* genome at this locus. The positive fitness incurred by the toxin also gives one example on the destinations of toxin-antidote elements that might be introduced to the field artificially.

Finally, in chapter 4, I used simulations to understand the dynamics of other genomic elements when individuals carrying and not carrying toxin-antidote system are admixed. Calculations of genotype frequency in the first three generations showed that the frequency of unlinked genomic elements of the toxin-antidote element in the survived embryos is not its frequency in the population. It also predicts a diverging point at 1/3, above which unlinked genomic elements will increase their frequency in the population. This effect is replicated using two independent simulation tools: SLiM and a customized python script.

Selfish genetic elements are sometimes defined as ‘genetic segments that can enhance their own transmission at the expense of other genes in the genome’ (Ågren and Clark, 2018). Here I show toxin-antidote elements can enhance their transmission at the benefit of other genes in the genome. The simulation results also showed that if there’s any fixed change between the two ancestral populations, toxin-antidote elements will eliminate them or push them to fixation. Thus toxin-antidote elements can reduce genetic diversity for unlinked genomic elements. This relates back to the original intention of this chapter which is to answer an unresolved question in chapter 3: why does *peel-zeel* region show strong signature of balancing selection? Results in this chapter led us to a promising hypothesis: the high divergence observed is not just by long-term balancing selection. It is also caused by reduced genetic diversity in the rest of the genome. To test this hypothesis, more work needs to be done. For example, simulations following the special lifestyle of *C. elegans* will help to mimic what happened in the wild. A colony of *C. elegans* in the field is often found by only 3-5 animals and group selection is extensive (Richaud et al., 2018). Frequent admixture between population with and without toxin-antidote elements happens in the wild. Whether these frequent admixtures can reduce genomic variance remains an interesting and important question to answer.

My doctoral thesis research demonstrates that quantitative methods are useful to study the genetic basis of long-term evolution. Comparative behavior studies are enabled by combining a set of machine learning and deep learning techniques. Identification of the causal genetic variants for fitness change is empowered by targeted genome editing with

CRISPR/Cas9 and quantitative competition experiments. And predication the conflicted evolution of toxin-antidote elements and other genomic elements is supported by quantitative modeling and simulation.

REFERENCES

ÅGREN, J. A. & CLARK, A. G. 2018. Selfish genetic elements. *PLoS genetics,* 14**,** e1007700.

ANDERSEN, E. C., BLOOM, J. S., GERKE, J. P. & KRUGLYAK, L. 2014. A variant in the neuropeptide receptor npr-1 is a major determinant of Caenorhabditis elegans growth and physiology. *PLoS Genet,* 10**,** e1004156.

ANDERSON, D. J. & PERONA, P. 2014. Toward a science of computational ethology. *Neuron,* 84**,** 18-31.

ANDERSON, J. L., MORRAN, L. T. & PHILLIPS, P. C. 2010. Outcrossing and the maintenance of males within C. elegans populations. *Journal of heredity,* 101**,** S62-S74.

ANDRILUKA, M., IQBAL, U., INSAFUTDINOV, E., PISHCHULIN, L., MILAN, A., GALL, J. & SCHIELE, B. Posetrack: A benchmark for human pose estimation and tracking. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018. 5167-5176.

ARRIBERE, J. A., BELL, R. T., FU, B. X., ARTILES, K. L., HARTMAN, P. S. & FIRE, A. Z. 2014. Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in Caenorhabditis elegans. *Genetics,* 198**,** 837-846.

AUSTIN, B., TRIVERS, R. & BURT, A. 2009. *Genes in conflict: the biology of selfish genetic elements*, Harvard University Press.

BARLUENGA, M., STÖLTING, K. N., SALZBURGER, W., MUSCHICK, M. & MEYER, A. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature,* 439**,** 719-723.

BARRIÈRE, A. & FÉLIX, M.-A. 2005. High local genetic diversity and low outcrossing rate in Caenorhabditis elegans natural populations. *Current Biology,* 15**,** 1176-1184.

BEEMAN, R., FRIESEN, K. & DENELL, R. 1992. Maternal-effect selfish genes in flour beetles. *Science,* 256**,** 89-92.

BEN-DAVID, E., BURGA, A. & KRUGLYAK, L. 2017. A maternal-effect selfish genetic element in Caenorhabditis elegans. *Science,* 356**,** 1051-1055.

BEN-DAVID, E., PLIOTA, P., WIDEN, S. A., KORESHOVA, A., LEMUS-VERGARA, T., VERPUKHOVSKIY, P., MANDALI, S., BRAENDLE, C., BURGA, A. & KRUGLYAK, L. 2020. Ubiquitous selfish toxin-antidote elements in Caenorhabditis species. *bioRxiv.*

BENDESKY, A., PITTS, J., ROCKMAN, M. V., CHEN, W. C., TAN, M.-W., KRUGLYAK, L. & BARGMANN, C. I. 2012. Long-range regulatory polymorphisms affecting a GABA receptor constitute a quantitative trait locus (QTL) for social behavior in Caenorhabditis elegans. *PLoS Genet,* 8**,** e1003157.

BENJAMIN, S. P. & ZSCHOKKE, S. 2000. A computerised method to observe spider web building behaviour in a semi-natural light environment. *European arachnology***,** 117-122.

BERMAN, G. J., CHOI, D. M., BIALEK, W. & SHAEVITZ, J. W. 2014. Mapping the stereotyped behaviour of freely moving fruit flies. *J R Soc Interface,* 11.

BI, Y., REN, X., YAN, C., SHAO, J., XIE, D. & ZHAO, Z. 2015. A Genome-wide hybrid incompatibility landscape between Caenorhabditis briggsae and C. nigoni. *PLoS Genet,* 11**,** e1004993.

BLAXTER, M. L. 2003. Nematoda: genes, genomes and the evolution of parasitism. *Adv Parasitol,* 54**,** 101-195.

BRAWAND, D., WAGNER, C. E., LI, Y. I., MALINSKY, M., KELLER, I., FAN, S., SIMAKOV, O., NG, A. Y., LIM, Z. W. & BEZAULT, E. 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature,* 513**,** 375-381.

BRISSON, D. 2018. Negative frequency-dependent selection is frequently confounding. *Frontiers in Ecology and Evolution,* 6**,** 10.

BUTCHER, R. A., FUJITA, M., SCHROEDER, F. C. & CLARDY, J. 2007. Small-molecule pheromones that control dauer development in Caenorhabditis elegans. *Nature chemical biology,* 3**,** 420-422.

BUTCHER, R. A., RAGAINS, J. R., KIM, E. & CLARDY, J. 2008. A potent dauer pheromone component in Caenorhabditis elegans that acts synergistically with other components. *Proceedings of the National Academy of Sciences,* 105**,** 14288-14292.

CHALFIE, M., HORVITZ, H. R. & SULSTON, J. E. 1981. Mutations that lead to reiterations in the cell lineages of C. elegans. *Cell,* 24**,** 59-69.

CHAMPER, J., CHAMPER, S. E., KIM, I., CLARK, A. G. & MESSER, P. W. 2019. Design and analysis of CRISPR-based underdominance toxin-antidote gene drives. *bioRxiv***,** 861435.

CHAMPER, J., KIM, I. K., CHAMPER, S. E., CLARK, A. G. & MESSER, P. W. 2020. Performance analysis of novel toxin-antidote CRISPR gene drive systems. *BMC biology,* 18**,** 1-17.

COLLIAS, N. E. & COLLIAS, E. C. 2014. *Nest building and bird behavior*, Princeton University Press.

DAWKINS, R. 1982. *The extended phenotype*, Oxford University Press Oxford.

DOBZHANSKY, T. & QUEAL, M. 1938. Genetics of natural populations. I. Chromosome variation in populations of Drosophila pseudoobscura inhabiting isolated mountain ranges. *Genetics,* 23**,** 239.

EGNOR, S. E. & BRANSON, K. 2016. Computational Analysis of Behavior. *Annu Rev Neurosci,* 39**,** 217-36.

ENDLER, J. A. 1977. *Geographic variation, speciation, and clines*, Princeton University Press.

ESTER, M., KRIEGEL, H.-P., SANDER, J. & XU, X. A density-based algorithm for discovering clusters in large spatial databases with noise. Kdd, 1996. 226-231.

FIELENBACH, N. & ANTEBI, A. 2008. C. elegans dauer formation and the molecular basis of plasticity. *Genes & development,* 22**,** 2149-2165.

FRASER, G. J., HULSEY, C. D., BLOOMQUIST, R. F., UYESUGI, K., MANLEY, N. R. & STREELMAN, J. T. 2009. An ancient gene network is co-opted for teeth on old and new jaws. *PLoS Biol,* 7**,** e1000031.

FRÉZAL, L. & FÉLIX, M.-A. 2015. The natural history of model organisms: C. elegans outside the Petri dish. *Elife,* 4**,** e05849.

FRØKJÆR-JENSEN, C., DAVIS, M. W., AILION, M. & JORGENSEN, E. M. 2012. Improved Mos1-mediated transgenesis in C. elegans. *Nature methods,* 9**,** 117-118.

GHOSH, R., ANDERSEN, E. C., SHAPIRO, J. A., GERKE, J. P. & KRUGLYAK, L. 2012. Natural variation in a chloride channel subunit confers avermectin resistance in C. elegans. *Science,* 335**,** 574-578.

GIMOND, C., VIELLE, A., SILVA-SOARES, N., ZDRALJEVIC, S., MCGRATH, P. T., ANDERSEN, E. C. & BRAENDLE, C. 2018. Evolution of sperm competition: Natural variation and genetic determinants of Caenorhabditis elegans sperm size. *bioRxiv***,** 501486.

GINZBURG, L. R., BINGHAM, P. M. & YOO, S. 1984. On the theory of speciation induced by transposable elements. *Genetics,* 107**,** 331-341.

GIRSHICK, R. Fast r-cnn. Proceedings of the IEEE international conference on computer vision, 2015. 1440-1448.

GOORMAGHTIGH, F., FRAIKIN, N., PUTRINŠ, M., HALLAERT, T., HAURYLIUK, V., GARCIA-PINO, A., SJÖDIN, A., KASVANDIK, S., UDEKWU, K. & TENSON, T. 2018. Reassessing the role of type II toxin-antitoxin systems in formation of Escherichia coli type II persister cells. *MBio,* 9.

GRAVING, J. M., CHAE, D., NAIK, H., LI, L., KOGER, B., COSTELLOE, B. R. & COUZIN, I. D. 2019. DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *Elife,* 8.

GREENE, J. S., BROWN, M., DOBOSIEWICZ, M., ISHIDA, I. G., MACOSKO, E. Z., ZHANG, X., BUTCHER, R. A., CLINE, D. J., MCGRATH, P. T. & BARGMANN, C. I. 2016a. Balancing selection shapes density-dependent foraging behaviour. *Nature,* 539**,** 254-258.

GREENE, J. S., DOBOSIEWICZ, M., BUTCHER, R. A., MCGRATH, P. T. & BARGMANN, C. I. 2016b. Regulatory changes in two chemoreceptor genes contribute to a Caenorhabditis elegans QTL for foraging behavior. *Elife,* 5**,** e21454.

GUNEL, S., RHODIN, H., MORALES, D., CAMPAGNOLO, J., RAMDYA, P. & FUA, P. 2019. DeepFly3D, a deep learning-based approach for 3D limb and appendage tracking in tethered, adult Drosophila. *Elife,* 8.

HALLER, B. C. & MESSER, P. W. 2019. SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. *Molecular biology and evolution,* 36**,** 632-637.

HANSELL, M. 2000. *Bird nests and construction behaviour*, Cambridge University Press.

HARA, K., KATAOKA, H. & SATOH, Y. Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2018. 6546-6555.

HARRIS, T. W., ANTOSHECHKIN, I., BIERI, T., BLASIAR, D., CHAN, J., CHEN, W. J., DE LA CRUZ, N., DAVIS, P., DUESBURY, M. & FANG, R. 2010. WormBase: a comprehensive resource for nematode research. *Nucleic acids research,* 38**,** D463-D467.

HOBERT, O. & LORIA, P. 2005. Uses of GFP in Caenorhabditis elegans. *Green Fluorescent Protein: Properties, Applications, and Protocols,* 47**,** 203-226.

HONG, W., KENNEDY, A., BURGOS-ARTIZZU, X. P., ZELIKOWSKY, M., NAVONNE, S. G., PERONA, P. & ANDERSON, D. J. 2015. Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning. *Proceedings of the National Academy of Sciences,* 112**,** E5351-E5360.

HSIN, H. & KENYON, C. 1999. Signals from the reproductive system regulate the lifespan of C. elegans. *Nature,* 399**,** 362-366.

HURST, G. D. & WERREN, J. H. 2001. The role of selfish genetic elements in eukaryotic evolution. *Nature Reviews Genetics,* 2**,** 597-606.

HURST, L. D. 1993. scat+ is a selfish gene analogous to Medea of tribolum castaneum. *Cell,* 75**,** 407-408.

INOUE, T., AILION, M., POON, S., KIM, H. K., THOMAS, J. H. & STERNBERG, P. W. 2007. Genetic analysis of dauer formation in Caenorhabditis briggsae. *Genetics,* 177**,** 809-818.

JIGGINS, C. D., NAISBIT, R. E., COE, R. L. & MALLET, J. 2001. Reproductive isolation caused by colour pattern mimicry. *Nature,* 411**,** 302-305.

JOHNSON, N. A. 2000. Speciation: Dobzhansky–Muller incompatibilities, dominance and gene interactions. *Trends in ecology & evolution,* 15**,** 480-482.

JONES, M. R., MILLS, L. S., ALVES, P. C., CALLAHAN, C. M., ALVES, J. M., LAFFERTY, D. J., JIGGINS, F. M., JENSEN, J. D., MELO-FERREIRA, J. & GOOD, J. M. 2018. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science,* 360**,** 1355-1358.

JOSEPH, B. B., BLOUIN, N. A. & FAY, D. S. 2018. Use of a sibling subtraction method for identifying causal mutations in Caenorhabditis elegans by whole-genome sequencing. *G3: Genes, Genomes, Genetics,* 8**,** 669-678.

JOVELIN, R., AJIE, B. & PHILLIPS, P. 2003. Molecular evolution and quantitative variation for chemosensory behaviour in the nematode genus Caenorhabditis. *Molecular ecology,* 12**,** 1325-1337.

JOYCE, D. A., LUNT, D. H., GENNER, M. J., TURNER, G. F., BILLS, R. & SEEHAUSEN, O. 2011. Repeated colonization and hybridization in Lake Malawi cichlids. *Current Biology,* 21**,** R108-R109.

KAIN, J., STOKES, C., GAUDRY, Q., SONG, X., FOLEY, J., WILSON, R. & DE BIVORT, B. 2013. Leg-tracking and automated behavioural classification in Drosophila. *Nat Commun,* 4**,** 1910.

KEENLEYSIDE, M. H. 1991. *Cichlid fishes: behaviour, ecology and evolution*, Springer Science & Business Media.

KELLER, L. & ROSS, K. G. 1998. Selfish genes: a green beard in the red fire ant. *Nature,* 394**,** 573-575.

KIM, K., SATO, K., SHIBUYA, M., ZEIGER, D. M., BUTCHER, R. A., RAGAINS, J. R., CLARDY, J., TOUHARA, K. & SENGUPTA, P. 2009. Two chemoreceptors mediate developmental effects of dauer pheromone in C. elegans. *Science,* 326**,** 994-998.

KIM, W., UNDERWOOD, R. S., GREENWALD, I. & SHAYE, D. D. 2018. OrthoList 2: a new comparative genomic analysis of human and Caenorhabditis elegans genes. *Genetics,* 210**,** 445-461.

KOCHER, T. D. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews Genetics,* 5**,** 288-298.

KUZMANOV, A., YOCHEM, J. & FAY, D. S. 2014. Analysis of PHA-1 reveals a limited role in pharyngeal development and novel functions in other tissues. *Genetics,* 198**,** 259-268.

LAMICHHANEY, S., BERGLUND, J., ALMÉN, M. S., MAQBOOL, K., GRABHERR, M., MARTINEZ-BARRIO, A., PROMEROVÁ, M., RUBIN, C.-J., WANG, C. & ZAMANI, N. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature,* 518**,** 371-375.

LANDE, R. 1980. Genetic variation and phenotypic evolution during allopatric speciation. *The American Naturalist,* 116**,** 463-479.

LARGE, E. E., XU, W., ZHAO, Y., BRADY, S. C., LONG, L., BUTCHER, R. A., ANDERSEN, E. C. & MCGRATH, P. T. 2016. Selection on a subunit of the NURF chromatin remodeler modifies life history traits in a domesticated strain of Caenorhabditis elegans. *PLoS genetics,* 12**,** e1006219.

LEE, D., YANG, H., KIM, J., BRADY, S., ZDRALJEVIC, S., ZAMANIAN, M., KIM, H., PAIK, Y.-K., KRUGLYAK, L. & ANDERSEN, E. C. 2017. The genetic basis of natural variation in a phoretic behavior. *Nature communications,* 8**,** 1-7.

LEE, D., ZDRALJEVIC, S., STEVENS, L., WANG, Y., TANNY, R. E., CROMBIE, T. A., COOK, D. E., WEBSTER, A. K., CHIRAKAR, R. & BAUGH, L. R. 2020. Balancing selection maintains ancient genetic diversity in C. elegans. *BioRxiv*.

LEVENE, H. 1953. Genetic equilibrium when more than one ecological niche is available. *The American Naturalist,* 87**,** 331-333.

LOH, Y.-H. E., BEZAULT, E., MUENZEL, F. M., ROBERTS, R. B., SWOFFORD, R., BARLUENGA, M., KIDD, C. E., HOWE, A. E., DI PALMA, F. & LINDBLAD-TOH, K. 2013. Origins of shared genetic variation in African cichlids. *Molecular biology and evolution,* 30**,** 906-917.

MACHADO, C. A., KLIMAN, R. M., MARKERT, J. A. & HEY, J. 2002. Inferring the history of speciation from multilocus DNA sequence data: the case of Drosophila pseudoobscura and close relatives. *Molecular biology and evolution,* 19**,** 472-488.

MATHIS, A., MAMIDANNA, P., CURY, K. M., ABE, T., MURTHY, V. N., MATHIS, M. W. & BETHGE, M. 2018. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci,* 21**,** 1281-1289.

MAYR, E. & PROVINE, W. B. 1998. *The evolutionary synthesis: perspectives on the unification of biology*, Harvard University Press.

MCGRATH, P. T., ROCKMAN, M. V., ZIMMER, M., JANG, H., MACOSKO, E. Z., KRUGLYAK, L. & BARGMANN, C. I. 2009. Quantitative mapping of a digenic behavioral trait implicates globin variation in C. elegans sensory behaviors. *Neuron,* 61**,** 692-699.

MEKAYE, K. 2001. Fishes, as well as birds, build bowers.

MOURITSEN, H. 2018. Long-distance navigation and magnetoreception in migratory animals. *Nature,* 558**,** 50-59.

MUÑOZ-JIMÉNEZ, C., AYUSO, C., DOBRZYNSKA, A., TORRES-MENDÉZ, A., DE LA CRUZ RUIZ, P. & ASKJAER, P. 2017. An efficient FLP-based toolkit for spatiotemporal control of gene expression in Caenorhabditis elegans. *Genetics,* 206**,** 1763-1778.

MUTHURAMALINGAM, M., WHITE, J. C., MURPHY, T., AMES, J. R. & BOURNE, C. R. 2019. The toxin from a ParDE toxin ‐ antitoxin system found in Pseudomonas aeruginosa offers protection to cells challenged with anti ‐ gyrase antibiotics. *Molecular microbiology,* 111**,** 441-454.

NATH, T., MATHIS, A., CHEN, A. C., PATEL, A., BETHGE, M. & MATHIS, M. W. 2019. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat Protoc,* 14**,** 2152-2176.

NOBLE, L. M., YUEN, J., STEVENS, L., MOYA, N., PERSAUD, R. S., MOSCATELLI, M. A., JACKSON, J. L., BRAENDLE, C., ANDERSEN, E. C. & SEIDEL, H. S. 2020. Selfing is the safest sex for Caenorhabditis tropicalis. *BioRxiv*.

O'CONNOR, C. H., SIKKINK, K. L., NELSON, T. C., FIERST, J. L., CRESKO, W. A. & PHILLIPS, P. C. 2019. Complex pleiotropic genetic architecture of evolved heat stress and oxidative stress resistance in the nematode Caenorhabditis remanei. *bioRxiv***,** 515320.

PEREIRA, T. D., ALDARONDO, D. E., WILLMORE, L., KISLIN, M., WANG, S. S., MURTHY, M. & SHAEVITZ, J. W. 2019. Fast animal pose estimation using deep neural networks. *Nat Methods,* 16**,** 117-125.

PETROU, G. & WEBB, B. 2012. Detailed tracking of body and leg movements of a freely walking female cricket during phonotaxis. *J Neurosci Methods,* 203**,** 56-68.

PRESGRAVES, D. C., BALAGOPALAN, L., ABMAYR, S. M. & ORR, H. A. 2003. Adaptive evolution drives divergence of a hybrid inviability gene between two species of Drosophila. *Nature,* 423**,** 715-719.

QIU, Z., YAO, T. & MEI, T. Learning spatio-temporal representation with pseudo-3d residual networks. proceedings of the IEEE International Conference on Computer Vision, 2017. 5533-5541.

RAJABPOUR, F. V., RAOOFIAN, R., HABIBI, L., AKRAMI, S. M. & TABRIZI, M. 2014. Novel trends in genetics: transposable elements and their application in medicine. *Archives of Iranian medicine,* 17**,** 0-0.

RAVIGNÉ, V., OLIVIERI, I. & DIECKMANN, U. 2004. Implications of habitat choice for protected polymorphysms.

REDDY, K. C., ANDERSEN, E. C., KRUGLYAK, L. & KIM, D. H. 2009. A polymorphism in npr-1 is a behavioral determinant of pathogen susceptibility in C. elegans. *Science,* 323**,** 382-384.

REDMON, J., DIVVALA, S., GIRSHICK, R. & FARHADI, A. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. 779-788.

REN, S., HE, K., GIRSHICK, R. & SUN, J. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 2015. 91-99.

REN, X., LI, R., WEI, X., BI, Y., HO, V. W. S., DING, Q., XU, Z., ZHANG, Z., HSIEH, C.-L. & YOUNG, A. 2018. Genomic basis of recombination suppression in the hybrid between Caenorhabditis briggsae and C. nigoni. *Nucleic acids research,* 46**,** 1295-1307.

REYNOLDS, R. M. & PHILLIPS, P. C. 2013. Natural variation for lifespan and stress response in the nematode Caenorhabditis remanei. *PLoS One,* 8**,** e58212.

RICHAUD, A., ZHANG, G., LEE, D., LEE, J. & FÉLIX, M.-A. 2018. The local coexistence pattern of selfing genotypes in Caenorhabditis elegans natural metapopulations. *Genetics,* 208**,** 807-821.

ROBIE, A. A., SEAGRAVES, K. M., EGNOR, S. R. & BRANSON, K. 2017. Machine vision methods for analyzing social interactions. *Journal of Experimental Biology,* 220**,** 25-34.

ROSENDAHL, S., TAMMAN, H., BRAUER, A., REMM, M. & HÕRAK, R. 2020. Chromosomal toxin-antitoxin systems in Pseudomonas putida are rather selfish than beneficial. *Scientific Reports,* 10**,** 1-13.

ROWLAND, A., MINERS, J. O. & MACKENZIE, P. I. 2013. The UDP-glucuronosyltransferases: their role in drug metabolism and detoxification. *The international journal of biochemistry & cell biology,* 45**,** 1121-1132.

RUSSELL, A. L., MORRISON, S. J., MOSCHONAS, E. H. & PAPAJ, D. R. 2017. Patterns of pollen and nectar foraging specialization by bumblebees over multiple timescales using RFID. *Scientific reports,* 7**,** 1-13.

SEIDEL, H. S., AILION, M., LI, J., VAN OUDENAARDEN, A., ROCKMAN, M. V. & KRUGLYAK, L. 2011. A novel sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in C. elegans. *PLoS Biol,* 9**,** e1001115.

SEIDEL, H. S., ROCKMAN, M. V. & KRUGLYAK, L. 2008. Widespread genetic incompatibility in C. elegans maintained by balancing selection. *science,* 319**,** 589-594.

SHOEMAKER, D. D., KATJU, V. & JAENIKE, J. 1999. Wolbachia and the evolution of reproductive isolation between Drosophila recens and Drosophila subquinaria. *Evolution,* 53**,** 1157-1164.

SMITH, J. M. 1966. Sympatric speciation. *The American Naturalist,* 100**,** 637-650.

STEPHENS, G. J., JOHNSON-KERNER, B., BIALEK, W. & RYU, W. S. 2008. Dimensionality and dynamics in the behavior of C. elegans. *PLoS Comput Biol,* 4**,** e1000028.

SYLVESTER, J. B., RICH, C. A., LOH, Y.-H. E., VAN STAADEN, M. J., FRASER, G. J. & STREELMAN, J. T. 2010. Brain diversity evolves via differences in patterning. *Proceedings of the National Academy of Sciences,* 107**,** 9718-9723.

SYLVESTER, J. B., RICH, C. A., YI, C., PERES, J. N., HOUART, C. & STREELMAN, J. T. 2013. Competing signals drive telencephalon diversity. *Nature Communications,* 4**,** 1-7.

THOMPSON, O. A., SNOEK, L. B., NIJVEEN, H., STERKEN, M. G., VOLKERS, R. J., BRENCHLEY, R., VAN'T HOF, A., BEVERS, R. P., COSSINS, A. R. & YANAI, I. 2015. Remarkably divergent regions punctuate the genome assembly of the Caenorhabditis elegans Hawaiian strain CB4856. *Genetics,* 200**,** 975-989.

TUCKER, R. 1981. The digging behavior and skin differentiations in Heterocephalus glaber. *Journal of Morphology,* 168**,** 51-71.

VOLLRATH, F. 1992. Analysis and interpretation of orb spider exploration and web-building behavior. *Advances in the Study of Behavior,* 21**,** 147-199.

WADE, M. J. & BEEMAN, R. W. 1994. The population dynamics of maternal-effect selfish genes. *Genetics,* 138**,** 1309-1314.

WANG, R. L., WAKELEY, J. & HEY, J. 1997. Gene flow and natural selection in the origin of Drosophila pseudoobscura and close relatives. *Genetics,* 147**,** 1091-1106.

WEISSBROD, A., SHAPIRO, A., VASSERMAN, G., EDRY, L., DAYAN, M., YITZHAKY, A., HERTZBERG, L., FEINERMAN, O. & KIMCHI, T. 2013. Automated long-term tracking and social behavioural phenotyping of animal colonies within a semi-natural environment. *Nat Commun,* 4**,** 2018.

WERREN, J. H. 1997. Biology of wolbachia. *Annual review of entomology,* 42**,** 587-609.

WERREN, J. H. 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proceedings of the National Academy of Sciences,* 108**,** 10863-10870.

WILD, B., SIXT, L. & LANDGRAF, T. 2018. Automatic localization and decoding of honeybee markers using deep convolutional neural networks. *arXiv preprint arXiv:1802.04557*.

WINDBICHLER, N., MENICHELLI, M., PAPATHANOS, P. A., THYME, S. B., LI, H., ULGE, U. Y., HOVDE, B. T., BAKER, D., MONNAT, R. J. & BURT, A. 2011. A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature,* 473**,** 212-215.

WU, C.-I. & PALOPOLI, M. F. 1994. Genetics of postmating reproductive isolation in animals. *Annual review of genetics,* 28**,** 283-308.

WU, C. I. 2001. The genic view of the process of speciation. *Journal of evolutionary biology,* 14**,** 851-865.

YORK, R. A., PATIL, C., HULSEY, C. D., ANORUO, O., STREELMAN, J. T. & FERNALD, R. D. 2015. Evolution of bower building in Lake Malawi cichlid fish: phylogeny, morphology, and behavior. *Frontiers in Ecology and Evolution,* 3**,** 18.

ZHAO, Y., LONG, L., XU, W., CAMPBELL, R. F., LARGE, E. E., GREENE, J. S. & MCGRATH, P. T. 2018. Changes to social feeding behaviors are not sufficient for fitness gains of the Caenorhabditis elegans N2 reference strain. *Elife,* 7.