SYNERGIZING GENETIC ENGINEERING AND EVOLUTIONARY BIOLOGY: ANIMAL-BASED LINEAGE TRACKING TO STUDY ADAPTIVE MUTATIONS

by

ZACHARY CHRISTOPHER STEVENSON

A DISSERTATION

Presented to the Department of Biology
and the Division of Graduate Studies at the University of Oregon
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

December 2023

DISSERTATION APPROVAL PAGE

Student: Zachary Christopher Stevenson

Title: Synergizing Genetic Engineering and Evolutionary Biology: Animal-Based Lineage Tracking to Study Adaptive Mutations

This dissertation has been accepted and approved in partial fulfilment of the requirements for the Doctor of Philosophy degree in the Department of Biology by:

William Cresko	Chairperson
Patrick C. Phillips	Advisor
Andrew Kern	Core Member
Diana Libuda	Core Member
Peter Ralph	Institutional Representative

and

Krista Chronister	Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded December 2023

DISSERTATION ABSTRACT

Zachary Christopher Stevenson

Doctor of Philosophy

Department of Biology

November 2023

Title: Synergizing Genetic Engineering and Evolutionary Biology: Animal-Based Lineage Tracking to Study Adaptive Mutations

The fitness effects of new mutations are one of the central drivers of evolutionary change. Mutation is the ultimate source of novel genetic diversity, yet only a small fraction of new mutations provides an adaptive advantage. Additionally, many new adaptive mutations provide only a slight advantage and are challenging to identify and quantify their selection coefficient. Furthermore, a given mutation may be advantageous in one environment and disadvantageous in another environment. Natural selection acts upon the phenotype produced by the new mutation and if adaptive, the mutation can increase in frequency. Evolution acts though the lineage–the fundamental unit of evolution–because it is the lineage which changes overtime in response to selection. Over the past 150 years a robust and comprehensive set of theory has been developed around evolutionary biology and adaptive mutations. However, comprehensive estimation of the fitness effects of new mutations has remained challenging, but recent developments in genetic engineering open new opportunities to significantly advance the field. Within this dissertation, I present several novel approaches for editing the genome using a large genomic library, which allows the implementation of barcoding approach for evolutionary lineage tracking in an animal system for the first time. I apply this methodology to measure the fitness of a known Ivermectin resistant strain of *Caenorhabditis elegans* in what is the largest animal experimental evolution to date and highlight important new directions in evolution and genomics that this new approach allows. This dissertation includes previously published and unpublished coauthored material.

Zachary Christopher Stevenson

EDUCATION

University of Oregon Eugene, OR
Ph.D. Candidate 2017-Present
Advisor: Patrick C. Phillips

University of Utah Salt Lake City, UT
B.S. Biology 2012
Advisor: Erik M. Jorgensen

RESEARCH EXPERIENCE

University of Oregon 2017-Present
Development of high throughput genome engineering technology for *Caenorhabditis elegans* for experimental evolution.

- Development of TARDIS for massive paralleled transgenesis in large populations of *C. elegans*.
- Streamlined CRISPR transgene integration via development of novel integration-specific selection.
- Experimental evolution of adaptive lineages within various anthelmintic selective conditions.

This work has resulted in three publications in eLife, Genome Biology and Evolution, and G3, along with a provisional patent. Projects are funded by The NIH training grant T32 GM007413-42, R01 AG56436, and R35 GM131838

Knudra Transgenics & AxumBio 2013-2017
Research and development of CRISPR methodologies, novel biosensors, and high-thought compound testing.

- Created in partnership with many clients over 100 unique CRISPR projects.
- Generated parasitic *daf-12* variant homologs in *C. elegans* and screened for agonist compounds.
- Implemented QUAS system for increased expression of fluorescence in biosensors.

University of Utah 2010-2013
Research and development of optimized RNAi selective strategies and gene trap technologies

- Research in RNAi inactivation via transgenic flock house virus B2 protein.
- Research and development in randomized Mos1 transposon with fluorescent for identification of promoter elements.

HONORS AND AWARDS

- NIH Predoctoral Training Fellow, University of Oregon 2019-2021
- Biology Undergraduate Research Program Award, University of Utah 2011
- Undergraduate Research Opportunities Program Award, University of Utah 2010

SERVICE

- **Graduate Evolutionary Biology and Ecology Students: Seminar Coordinator** 2019-2022
  Assisted in developing three public seminars at the University of Oregon on the topics of evolutionary biology, sciences and society, and forensic genetics, bring over 600 people in attendance across all seminars.
- Reviewer for Evolution 2023

PUBLICATIONS

Zachary C. Stevenson, Megan J. Moerdyk-Schauwecker, Stephen A. Banse, Dhaval S. Patel, Hang Lu, Patrick C. Phillips, (2023) High-Throughput Library Transgenesis in *Caenorhabditis elegans* via Transgenic Arrays Resulting in Diversity of Integrated Sequences (TARDIS) eLife12:RP84831 https://doi.org/10.7554/eLife.84831.1

Kasimatis, K. R., Sánchez-Ramírez, S., & **Stevenson, Z. C.** (2021). Sexual Dimorphism through the Lens of Genome Manipulation, Forward Genetics, and Spatiotemporal Sequencing. Genome Biology and Evolution, 13(2). https://doi.org/10.1093/gbe/evaa243

Stevenson, Z. C., Moerdyk-Schauwecker, M. J., Jamison, B., & Phillips, P. C. (2020). Rapid Self-Selecting and Clone-Free Integration of Transgenes into Engineered CRISPR Safe Harbor Locations in *Caenorhabditis elegans*. G3 Genes|Genomes|Genetics, 10(10), 3775–3782. https://doi.org/10.1534/g3.120.401400

PATENTS

Stevenson, Z. C., Banse, S. A., & Phillips, P. C. (2021). Genetic data compression and methods of use (Patent No. US20210332387A1).

Hopkins, C.E, Trisha Brock, T., Marshall, T., Colasanto, M., **Stevenson, Z.C.,** (2019) Homologous Recombination Reporter Construct and Uses Thereof (Patent No. US20210277421A1)

UNDERGRADUATE MENTORING

Graduate Mentor, Eleanor A. Laufer 2021-Present

Selective Advantage of *avr-14, avr-15* and *glc-1* Knockout in *Caenorhabditis elegans* in High Ivermectin Conditions.

- Mentored several molecular biology skills such as PCR, gel electrophoresis, plasmid cloning, Illumina sequencing, and *C. elegans* CRISPR.
- Awarded two undergraduate research scholarships.
- Awarded NIH administrative supplement for summer research.
- Awarded Vice President for Research Undergraduate Fellowship.
- Presented posters at University of Oregon Undergraduate Research Symposium.

Graduate Mentor, Tiphanie R. Pfefferle Summer 2019

Development of the PAM flexible xCas9 for CRISPR genome engineering in *Caenorhabditis elegans*.

- Mentored several molecular biology skills such as PCR, gel electrophoresis, plasmid cloning, and *C. elegans* CRISPR as part of the Summer Program for Undergraduate Research (SPUR)
- Presented poster at University of Oregon SPUR Symposium.

TEACHING EXPERIENCE

Graduate Employee, BI330/331 Microbiology Lecture & Lab Instructor Assistant Spring 2018
Instructors: Dr. Adam Saunders & Dr. Alan Kelly
University of Oregon
Provided homework assistance, office hours, and grading for microbiology lecture. For microbiology lab, taught students several microbiological assays such as microscope use, gram staining, growth curves, and general microbe culture.

Graduate Employee, BI123 Biology of Cancer Winter 2018
Discussion Sections and Lab Assistant.
Instructor: Dr. Alan Kelly
University of Oregon
Provided additional lecture on topics covered from the primary class. Taught students on several lab techniques such as microscope usage and slide preparation. Also taught hands-on experiments showing the detrimental effects of ultraviolet light on cells.

Graduate Employee, BI214 General Biology IV: Mechanisms. Fall 2017
Discussion Sections and Lab Assistant.
Instructor: Dr. Alan Kelly
University of Oregon
Provided homework assistance, office hours, and grading for lecture. For lab, lectured and taught students several molecular biology assays such as microscope use, micropipetting, gel electrophoresis, and transformation.

Teaching and Lab Assistant, Laboratory Techniques. Summer 2012
Biology Undergraduate Research Program Lab Assistant and Discussion Leader
Instructor: Dr. Rose Mary
University of Utah
Worked with a class 20 undergraduate students, teaching various molecular cloning technologies and protein analysis. Taught documentation skills and lab safety procedures.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Patrick Phillips, for his incredible support and confidence in this challenging project. Dr. Phillips provided me with an unparalleled environment to explore my interest in genetic engineering and evolutionary biology. Dr. Phillips' mentorship and guidance was absolutely essential to my success, and I am truly grateful for his confidence in my work. I would also like to thank my committee members, William Cresko, Andrew Kern, Peter Ralph, and Diana Libuda for their guidance and advice.

I would also like to thank the members of the Philips lab. Megan Moerdyk-Schauwecker has been a daily colleague and essential to every project presented here. Megan's enthusiasm and deep molecular genetics background was a constant resource in developing the projects and I am forever grateful for her willingness to help and discuss every nuanced aspect of this dissertation. I would like to thank Dr. Stephen Banse, who provided near weekly exciting and stimulating conversations on developing new and (arguably) outrageous projects, and my work has been significantly influenced by his enthusiasm for molecular genetics. I would also like to thank the many other members of the Phillips lab, Anna Coleman-Hulbert for keeping the lab functioning daily, Christine Sedore, Kristin Robinson, and many undergraduates for their help in maintaining lab supplies and assisting in several projects. I want to thank Eleanor Laufer for her dedication and great assistance in maintaining and supporting the experimental evolution projects. I would also like to thank past Phillips lab members, Dr. Katja R. Kasimatis for her mentorship, collaboration, and discussions on the experimental evolution projects, and Dr. P. Alex de Verteuli for her support and hosting me on my early days in the lab. I am also grateful for the admirative staff, Arlene Deyo, Sarah Nash, and Leah Frazier for their essential support for the Institute of Ecology and Evolution. I would also like to thank the Genomics Core (GC3F) for providing assistance in designing and running many sequencing experiments.

I would also like thank the coffee run crew current and past, Christine Sedore, Rose Al-Saadi, Alex de Verteuli, & Katja R. Kasimatis for enabling me to keep my addiction, as well as many important conversations about science. I also want to acknowledge the many, many baristas at Tailored Coffee Roaster, Coffee Plant Roasters, Farmers Union Coffee Roasters, Starbucks, and on campus coffee providers Elements Café and Hearth Café for providing an endless stream of espresso and coffee.

I would like to thank my parents, Laurie Stevenson and Chad Stevenson for allowing me to explore my interest and fostering my education in biotechnology in middle school and high school. There were several critical butterfly effect moments that without their willingness and support, I would have never entered biology. I also want to thank my grandmother, Carol Landures for also fostering my interest in the outdoors and biology as a young child and her many interesting questions over the years.

I would also like to thank my funding sources. Funding for this dissertation was provided by the National Institutes of Health grants T32 GM007413-42 awarded to Zachary C. Stevenson, R01 AG56436, R35 GM131838 to awarded to Patrick C. Phillips, and R01 AG056436 awarded to Patrick C. Phillips and Hang Lu.

Finally, I would like to thank my loving spouse, Evelin Damian Ramírez. Her support everyday has been monumental to my success, and I could not have completed this dissertation without her. Her willingness to listen to all my talks, read over my work, as well as keeping me grounded every day. She is an amazing person, and I am deeply lucky to have her has my partner.

In loving memory of Bill & Deno Landures

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

The profound diversity that we see across biological systems arises from evolutionary processes. At the center of the origin of that diversity is mutation, the ultimate source of evolutionary change. Over the last 150 years there has been several theories developed on mutation and their evolutionary consequences. During Darwin's time, DNA and the biological nature of mutations had not been discovered. While obvious facts of inheritance (children often look like their parents) was well known, the process to transfer this information to the next generation was yet to be elucidated. One prominent theory of the time was blending inheritance, which suggested that phenotypic characteristics inherited by parents would be blended, resulting in an average of the two. An essay written by Fleeming Jenkin in the North British Review (Morris, 1994) was critical of Darwin's theory of evolution, because novel traits would become diluted by blended inheritance. Darwin's remarked that Jenkin's essay was the most valuable he had received on *On The Origin of Species*. Only later with the emergence of mendelian genetics was blending inheritance abandoned, while also giving rise to the modern synthesis of genetics and evolution. With mutation and natural selection established–and central to evolutionary biology–many theories have been proposed to address the evolutionary consequences of mutations. Are most mutations deleterious? Adaptive? Or of no consequence? Prevailing views at the time considered most mutations were deleterious (Crow and Temin, 1964), primarily rooted in the observation that isolated mutations often reduced viability. However, most mutations do not provide an appreciable phenotype (Loewe and Charlesworth, 2006), and historically estimated frequencies of strong deleterious estimations are likely incorrect. To reconcile know mutation rates (observed in protein changes at the molecular level) and the lack of strong deleterious effects, neutral and nearly neutral theory was proposed. Neutral theory suggests that most mutations are not 'visible' to natural selection and will simply change in frequency by genetic drift. What remains missing is an empirical framework to quantify the selective outcomes of mutations at scale, and ground the distribution of fitness effects in data, particularly in animal systems. With modern molecular technologies, and high-throughput experiments, we can for the first-time address over 150 years of theory on mutation from the lab bench in a rigorous manner.

Natural Selection

The mid-19th century witnessed the concurrent and independent contributions of Charles Darwin and Alfred Russel Wallace, who are jointly acknowledged for expounding the principle of natural selection. This seminal idea provided the requisite for evolutionary change, the 'why' behind evolution. Both Charles Darwin and Alfred Russel Wallace were captivated by the wealth of diversity in the natural world. Their travels to various regions exposed them to a plethora of species, each uniquely adapted to their environment. These observations spurred them to develop a theory that could explain the origin and diversification of life forms. In 1859, Darwin's seminal work, "*On the Origin of Species*," was published, presenting the theory of natural selection (Darwin, 1859). He proposed that in the struggle for existence, individuals with advantageous traits are more likely to survive and reproduce, passing those traits on to their offspring. Over time, this process accumulates changes, leading to the emergence of new species. Wallace had arrived at a similar conclusion independently, and the two scientists jointly presented their findings to the scientific community (McKINNEY, 1966). The groundbreaking revelation of natural selection predates the discovery of DNA as the genetic material responsible for heredity and evolution. During the 19th century when Darwin and Wallace were formulating their ideas, the underlying molecular basis of inheritance was largely unknown. The concept of genes, let alone the intricate structure and functions of DNA, had not yet entered the scientific discourse. Natural selection is built upon the observations of variation, adaptation, and heredity within populations. Natural selection predicts favorable variations accumulate over time through the differential survival and reproduction of individuals with advantageous traits, which would require some ability of these traits to be inherited. While Gregor Mendel and Charles Darwin were contemporaries, unfortunately, the connections of heritability, genetics, and evolution would not emerge until the mid- 20th century.

The modern synthesis, a pivotal paradigm in evolutionary biology developed during the mid-20th century as a convergence between the seminal ideas of Charles Darwin's natural selection and Gregor Mendel's genetic revelations (Huxley, 1942). This conceptual unification effectively bridged the perceived chasm between Darwinian notions of gradual and continual species transformation with the insights into inheritance and variability from Mendelian genetics. At this transition, we start to see the birth of molecular evolution, the investigation of evolution at the DNA and protein level. In the 1960s, John Lee Hubby & Richard Lewontin pioneered the

study of genetic variation using biochemical markers, such as isoenzymes (Charlesworth et al., 2016; Hubby and Lewontin, 1966; Lewontin and Hubby, 1966). Their research revealed surprising levels of genetic variability within populations, challenging the prevailing notion of species as uniform entities. Isoenzymes are variations of enzymes with distinct molecular forms, serving as markers for this diversity. By comparing the frequency of different isoenzyme alleles among populations, Hubby & Lewontin gained insight into the evolutionary relationships between species. Their work illuminated the dynamic nature of genetic diversity and underscored its role in the evolutionary process. With the understanding that the natural world was diverse, not just at the macro level, but also at the molecular, questions started to emerge about the nature of this diversity and how it was maintained and generated. Are these mutations adaptive or deleterious? How was genetic diversity maintained?

Mutation, Neutral Theory & Adaptation

It is clear mutations are central to evolutionary theory, and much of population genetics has been based on off theoretical assumptions of their effects. Kimura and Ohta introduced the neutral theory and nearly neutral theory of molecular evolution in 1968 & 1969 (Kimura, 1968; Kimura and Ohta, 1969). This theory suggested that much of the genetic variation observed at the molecular level is not subject to the direct influence of natural selection. Instead, these variations are considered neutral or nearly neutral, having little impact on an organism's fitness. This idea marked a shift in our understanding, emphasizing the significance of genetic drift alongside selection in shaping populations. However, of course, some genetic mutations do provide an adaptive benefit and will increase in frequency among the population over time. In any given environment, there are possible genetic mutations which could confer a selective benefit. What are these mutations? How do we measure their fitness contribution? How do we approach these molecular evolutionary questions from a more empirical framework? Genetic mutation leads to selectable phenotypic outcomes, and modern molecular advances to edit the genome provide an unprecedented opportunity to integrate the genome, and craft a hypothesis-driven framework to address century-old evolutionary questions.

Experimental evolution is the study of evolution in a laboratory setting. It is a powerful tool for understanding the nature of mutations and generally the mechanisms of evolution, as it allows for strict control of the environment. Some of the first experiments were performed by

geneticist Hermann J. Müller in the 1920s to comprehend the genetic and evolutionary impacts of ionizing radiation (Muller, 1927). Müller subjected fruit flies (*Dropshphilla melanogaster*) to varying doses of ionizing radiation to induce mutations, followed by controlled breeding to study the impacts of the induced mutations. Müller utilized genetic markers on the X-chromosome, which was strategic, given its role in determining sex and its relatively well-defined characteristics in fruit flies. Müller recognized that utilizing genetic markers on the X-chromosome would enable him to track and identify mutations more accurately, as well as to analyze patterns of inheritance in subsequent generations. This pioneering approach not only provided important empirical evidence for the mutagenic effects of X-rays but also contributed significantly to our understanding of the link between radiation, genetic mutations, and the broader field of genetics and evolution.

Perhaps the most famous evolution experiment is the long-term experiment conducted by Dr. Richard Lenski with *Escherichia coli* (*E. coli*), which stands as a seminal investigation in experimental evolution (Lenski, 2017; Lenski et al., 1991). Started in 1988, the experiment involves the continuous cultivation of twelve *E. coli* populations under glucose-limited conditions to observe their evolutionary trajectories. Notably, the evolution of citrate utilization emerged as a focal point of the experiment. *E.coli* are unable to utilize citrate aerobically, one population (*Ara-3*) exhibited a unique adaptation, evolving the capacity to utilize citrate and exploit it as a carbon source after around ~31,000 generations. This adaptive event highlighted the role of historical context and genetic background in evolution, sparking additional investigations into the genetic basis of the novel trait. Lenski's experiment fundamentally contributes to the understanding of adaptation, selection, and mutation over the large multi-generational timescales in which evolution operates for the emergence of complex adaptations to arise. While several experiments have been performed to create a handful of mutations or observes the effects of mostly mutations of high effect, much like an iceberg, there is more beneath the surface.

The Distribution of Fitness Effects

As empirical methods emerge to measure and identify single mutations with adaptive effects, it has also become clear that most identifiable mutations are significantly deleterious and of large effect (Burch et al., 2007). Müller’s experiments and similar mutant-characterization work tend

to identify mutations of large deleterious effects which are more likely to produce profound phenotypes impacting the overall organism. Whereas experiments tailored to adaptive mutations, as in the work of Dr. Lenski's long term experimental evolution are almost exactly opposite, identifying only mutations of large positive adaptive effect. There remains a significant 'blind-spot' in our experimental capacity to identify mutations with less maximal effects. Mutations that might be slightly adaptive or deleterious are likely to be the majority of non-lethal mutations (Eyre-Walker and Keightley, 2007) and they can influence evolutionary dynamics (Figure 1.1). The distribution of fitness effects is also not stagnant. As populations continue to adapt, and new mutations arise, mutations which were once adaptive can appear deleterious as the mean fitness increases (Desai et al., 2013; Whitlock et al., 1995). Further confounding our ability to measure fitness effects, the environment must also be taken into consideration. Mutations which provide a fitness effect within one environment may be deleterious in another (Figure 1.1C). Several different environmental factors, such as temperature (Logan et al., 2014), and small molecule concentration can change the selection coefficient of new mutations (Baym et al., 2016; Lenski, 2017; Lenski et al., 1991; Levy et al., 2015). For a fully comprehensive understanding of the fitness effects of mutations, the historical (generational) and environmental contexts must be understood. Under an empirical framework, both must be integrated for a rigorous measurement of selection.

To address this challenge, advanced genetic engineering techniques have been developed which enable the large-scale manipulation and identification of genetic lineages, opening a new era of experimentation and integrating the historical and environmental context. One example utilized synthetic genetic markers to identify and follow lineages. Commonly referred to as barcoding (see Box 1.1), this methodology has an extraordinary level of precision and provides insight into the adaptive evolutionary process (Blundell and Levy, 2014). Through the utilization of barcoded lineage tracking, the experimental observation and quantification of numerous independent lineages during an experiment has become achievable. Various methods for tracking lineages have been implemented, usually involving native mutations as is the case in the Lenski experiments. However, these methodologies are inherently limited, offering a limited number of markers, and such markers are often not neutral with respect to the experiment. In contrast, barcodes present a vast array of markers for experimental utilization due to their synthetic design because they feature randomized sequences. A barcode containing just a 15-basepair sequence of

Box 1.1 Barcoding and molecular biology: The term barcode is an often-used term with various meaning and molecular characteristics–not all barcodes are the same or can provide the same utility. This can be quite confusing at first glance because they all share a common need, to identify a biological unit based on sequence. The exact biological unit, as well as the molecular design of the barcode, can be quite different. Barcodes have been experimentally adopted to study cellular lineages in development and even the study of cancer (Elsner, 2018). These barcodes often change their sequence upon cellular division or increase in number. Barcodes of this nature are not pre-engineered but will be 'programmed' by molecular methods to change sequence. Allowing for the reconstruction of the developmental lineage. Ecological definitions of barcodes are naturally occurring polymorphisms which are highly correlated with a given species, allowing for simple identification (Kress et al., 2015). In synthetic biology, barcodes are often used to mark pooled sequenced of DNA libraries or engineered expression units (genes) for quality control of the synthesis pipeline, or for downstream applications (O'Connell et al., 2023; Sidore et al., 2020). For evolutionary biology applications–which are the nature of the barcodes presented in this dissertation–are large, diverse, libraires of tiny, randomized SNPs with the intent to integrate into the genome. Once integrated, they are 'locked' within that lineage. These barcodes are designed for simple amplicon sequencing to measure their frequency for a given experimental evolution (Ba et al., 2019; Blundell and Levy, 2014; Jasinska et al., 2020; Levy et al., 2015; Stevenson et al., 2023). For a more detailed discussion of possible application of barcoding individuals, refer to Chapter 4–TARDIS as a method for creating barcoded individuals.

random nucleotides can yield over one billion distinct combinations. To date, the most expansive experiment involving barcoded evolutionary studies was performed by Dr. Sasha Levy (Levy et al., 2015). This study encompassed the evolution of over half a million distinct lineages under conditions of constrained glucose availability in the environment. While the concept and execution of this approach is straightforward, the tracking of barcoded lineages enabled Levy et al. to accurately assess the fitness of each individual lineage. Their efforts led to the identification of approximately ~18,000 adaptive lineages. By utilizing the barcoded-lineage approach, Levy et al. were able to not only measure the fitness of large effect mutations, but also much smaller ones, empirically measuring a more comprehensive view of the distribution of fitness effects. In later work, they were able to use the lineage paradigm to match mutations with their selection coefficients (Ba et al., 2019). Interestingly, we see the predictions of population genetics (Desai et al., 2013), where the mean fitness increase overtime, mutations that were increasing in frequency and historically adaptive eventually decrease in frequency as more fit mutations increase and out compete them. This unprecedented achievement marks the first time that the adaptive advantages possessed by many individual lineages within a population has been measured with a historical and environmental context. Lineage tracking is just one example that has emerged from the advancements of high-throughput genome editing and sequencing. Few examples exist of groups utilizing genetic engineering for empirical-based evolution experiments. However, there is powerful potential for such experiments for specific hypothesis testing in evolution–particularly in animal systems which are orders of magnitude more challenging to engineer compared to microbes. Selection experiments have utilized fluorescent markers to tract adaptive lineages (Hegreness et al., 2006), or to simply mark genetic background in a competition experiment (Crombie et al., 2018). Kasimatis et al. (2022) utilized ectopic selection of a toxic transgene, *peel-1*, to remove a competitor *Caenorhabditis elegans* male sperm from a population in a sexual selection experiment (Kasimatis et al., 2022). It is clear that increasing our transgenic throughput, as well as the generation of simpler genome engineering approaches, can open new experimental paradigms in evolutionary biology—a central goal of this dissertation.

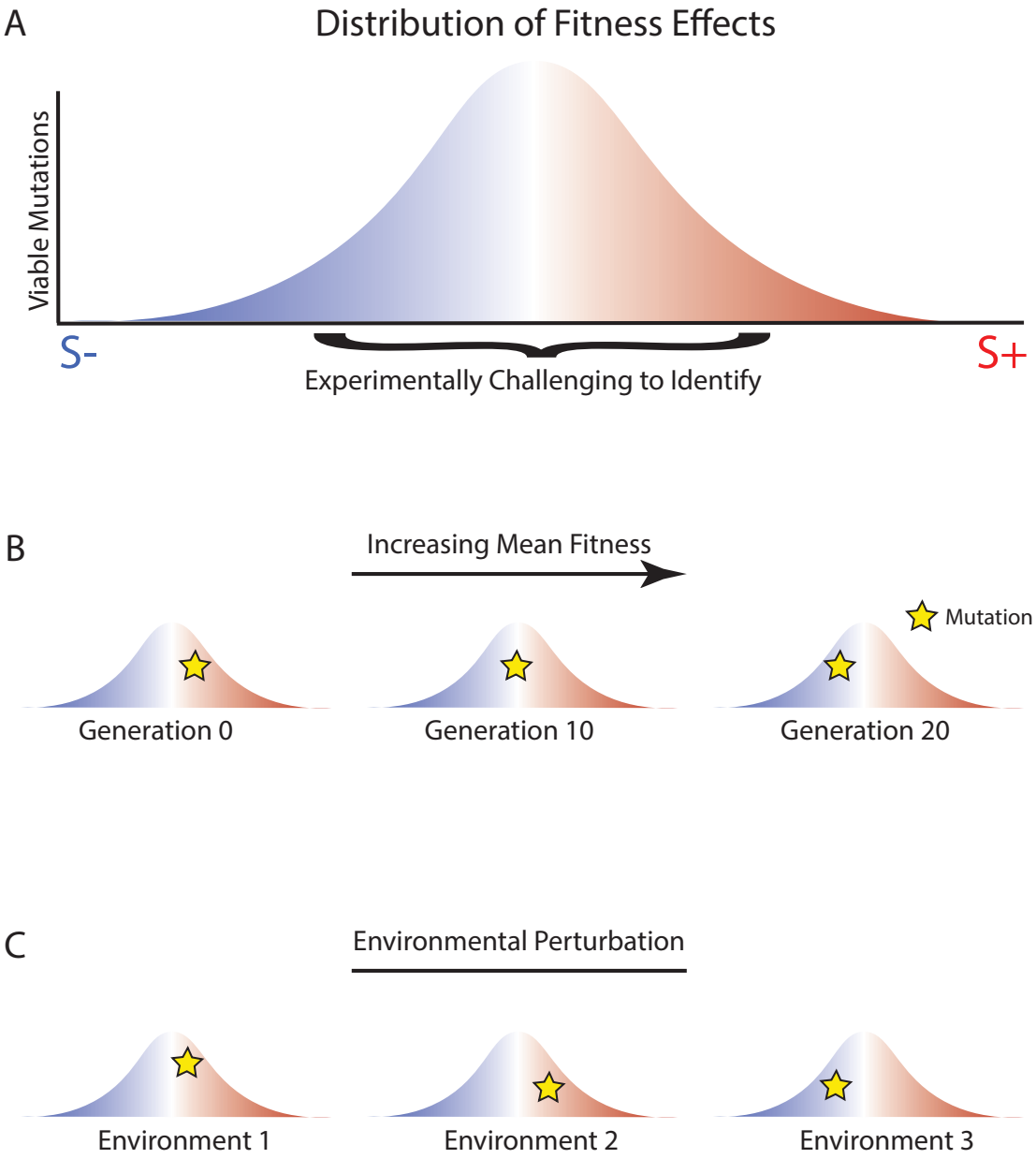


Figure 1.1 Distribution of Fitness Effects. A) Schematic overview of the general concept of the distribution considering non-lethal mutations. Mutations on the far ends are experimentally simpler to identify because of their larger effects. However, it is predicted that most mutations are of small effect, and therefore more challenging to isolate and observe. Additionally, as B) the mean fitness increases overtime as a response to selection, any specific mutation will eventually be out competed by new, more adaptive mutations resulting in a changed selection coefficient. In differential environments C) mutations which may have selective benefits in one can be deleterious in another.

Synergistics of genetic engineering and evolution

Genome engineering represents a major shift in our ability to test evolutionary hypotheses. Historically, as in the Müller experiments described prior, genomic 'manipulation' was primarily driven by random mutagens. Advancements in molecular biology have increase our precision in genome manipulation. Several toolkits have been developed which allow for the experimental manipulation of DNA, such as CRE and FLP recombinases (Gu et al., 1994; Sadowski, 1995; Sauer and Henderson, 1988), transposons (Sandoval-Villegas et al., 2021), talons, and zinc-finger nucleases (Joung and Sander, 2013; Klug, 2010). While each of these systems have their place, the development of CRISPR/Cas system as the premiere genome editing system has largely replaced them for many applications (Jinek et al., 2012; Wang and Doudna, 2023). CRISPR/Cas technology represents a revolutionary breakthrough in the field of genetics and molecular biology, offering powerful tools for precise manipulation of the genetic code. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) was first discovered in the 1980s when scientists observed an unusual repetitive DNA sequence in the genomes of certain bacteria (Ishino et al., 1987). The true potential of CRISPR/Cas technology was unveiled in the early 2010s when researchers harnessed this system to edit genes with unparalleled precision, marking a transformative moment in the world of genetic engineering. The importance of CRISPR/Cas technology in genetic engineering cannot be overstated. This revolutionary tool has provided scientists with the ability to modify the DNA of organisms easily and precisely, from bacteria to humans. By using CRISPR/Cas, researchers can target specific genes, insert, or delete genetic material, and even repair or replace faulty genes associated with diseases. This has immense implications for biomedicine, agriculture, and various other fields, including evolutionary biology, as it opens opportunities for understanding natural selection, genetic drift with an empirical framework, and fundamentally manipulating the genetic material at the heart of evolution. Beyond simple genetic manipulations with CRISPR/Cas technology, there are two main variations that facilitate the regulation of genes: CRISPR activation and CRISPR inhibition (Pickar-Oliver and Gersbach, 2019). CRISPR activation, often referred to as CRISPRa, involves the use of modified Cas proteins to amplify gene expression. Researchers can design a guide RNA (gRNA) to target a specific gene of interest and then attach a transcriptional activator to stimulate gene expression, ultimately leading to the increased production of the corresponding protein. Conversely, CRISPR inhibition, or CRISPRi, aims to reduce or silence the expression of

a particular gene. This is achieved by fusing a transcriptional repressor to the Cas protein, preventing the gene from being transcribed into RNA and, subsequently, translated into a functional protein. These variations of CRISPR/Cas technology provide a versatile toolkit for fine-tuning gene regulation and have immense potential for both basic research in evolutionary biology (Kasimatis et al., 2021) as well as therapeutic applications (Kang et al., 2022).

The capacity to manipulate the genome has ushered in unprecedented experimental possibilities in animal-based research, as well as within microbial and cell culture systems. Within the realm of single-cell research, transgenesis can be conveniently conducted in parallel through both transformation and transduction techniques, thereby facilitating a wide spectrum of experiments that would otherwise be unfeasible in animal systems. The high-throughput characteristic of transformation played a pivotal role in pioneering barcoded lineage tracking experiments, which would have been unfeasible if integration was performed one barcode at a time. In contrast, when dealing with animal systems, the ability to conduct high-throughput experiments would likewise open the door to a wide array of experimental paradigms. Unlike microbial systems, where foreign DNA can, in principle, be transferred to subsequent generations once it enters the cell, in animal systems, injections typically need to be administered to the germline to induce a heritable event. Adding complexity to this matter, injection methods must be carried out one at a time, significantly impeding the efficiency of the transgenesis process.

There is a clear gap in our experimental ability to study evolutionary biology from an empirical framework in animal systems despite the recent advancements in genome engineering. Fundamental to our understanding of evolution is the evolution of adaptation and how selection is acting historically within a population and with the environment. While this a central topic in evolutionary genetics (Jensen et al., 2019; Kern and Hahn, 2018), this dissertation aims to present fundamental technological advancements necessary to empirically study what is at the core of evolutionary biology–the distribution of adaptive fitness effects within an animal system.

Dissertation Outline

To study the distribution of fitness effects by barcoded lineage tracking within an animal model requires major advancements in our transgenic capabilities. This dissertation aims to address and

present the foundational technological advancements required to increase our throughput in animal systems. I have chosen to address this need in *Caenorhabditis elegans* for several reasons (Meneely et al., 2019). As a model for evolutionary biology, *C. elegans* is ideal among animals for its short lifespan and high reproductive output of several hundred during its reproductive period, which it can reach after just three days post fertilization (Teotónio et al., 2017). *C. elegans* also has major genetic advantages as a model, being the first animal to have its genome sequenced, with a number of innovative genetic tools already established in the system.

Chapter 2 discusses the various technological advancements in genome editing, sequencing, and experimental paradigms (Kasimatis et al., 2021). In addition to myself, Katja R. Kasimatis, and Santiago Sánchez-Ramírez were key contributors to this published work, with Katja R. Kasimatis as lead author. Sexual dimorphism is frequently observed in sexually reproducing species, often accompanied by noticeable variations in gene expression between males and females. These discrepancies in gene expression can lead to diverse evolutionary outcomes. To gain a deeper understanding of these consequences within an evolutionary context, we can employ methods derived from molecular genetics and biomedical research. This chapter delves into the molecular processes underlying the development of sexual dimorphism and introduces several gene-editing techniques, sequencing methods, and experimental approaches tailored to the study of sexual selection.

Chapter 3 discusses our approach to transgenic insertion utilizing integration-specific selection (Stevenson et al., 2020). In addition to myself, Megan Moerdyk-Schauwecker, Brennen Jamison, and Patrick C. Phillips, were key contributors to this work. Integration of transgenes is essential to our efforts to bring high throughput transgenesis to an animal model. For *C. elegans*, this remained a challenge that required several steps to go from concept to final integrated transgene. We developed a 'synthetic landing pad' which works much like antibiotic selection for transformation in bacteria. In addition, we utilized native *C. elegans* homology directed repair to 'clone in-vivo.' DNA fragments are injected which will self-assemble as a complete transgene. This process can go from concept to final transgenic animal in less than a week, greatly increasing our throughput.

Chapter 4 discusses our paralleled high-throughput transgenesis approach for an animal model we call TARDIS (Transgenic Arrays Resulting In Diversity of Integrated Sequences). In addition to myself, Megan Moerdyk-Schauwecker, Stephen A. Banse, Dhaval S. Patel, Hang Lu, and Patrick C. Phillips were key contributors to this published work. In Chapter 3, I outlined the process for integrating transgene with integration-specific selection which provided a means to increase transgenesis for a single transgene. In this chapter, I outline our approach to parallelize this process creating the first animal-library transgenesis approach. We build two separate libraries, one library based on barcodes for lineage tracking and one based on promoters for gene expression.

Chapter 5 discusses our patented approach to library transgenesis. In addition to myself, Stephen A. Banse and Patrick C. Phillips were key contributors to this provisional patent (Stevenson et al., 2021). In this provisional patent, we outline many of the various applications of the TARDIS platform and several other possible ways TARDIS could be implemented for models beyond *C. elegans*.

Chapter 6 discusses our preliminary selection experiments utilizing barcoded lineage tracking to measure the fitness of a mutant strain compared to wildtype. In addition to myself, Eleanor A. Laufer, Kristin J. Robinson, and Patrick C. Phillips contributed to this unpublished work. Utilizing TARDIS, we barcoded several unique lineages of both a known mutant strain resistant to ivermectin, and lineages of a wildtype strain which is highly sensitive to ivermectin. We found that we can measure the fitness of each lineage with high precision in a novel environmental context, which build upon our technological advancements in the prior work.

These chapters taken together outline the rational and technological development for high throughput experimental evolution, utilizing first ever experiment for barcoded lineage tracking in an animal system.

CHAPTER 2 SEXUAL DIMORPHISM THROUGH THE LENS OF GENOME MANIPULATION, FORWARD GENETICS, AND SPATIO-TEMPORAL SEQUENCING

This chapter was published in Volume 13, issue 2 of the journal Genome Biology and Evolution in 2021. Katja R. Kasimatis and Santiago Sánchez-Ramírez are co-authors on this manuscript. I co-wrote the sections related to genome editing methodology.

The citation for this publication is as follows:

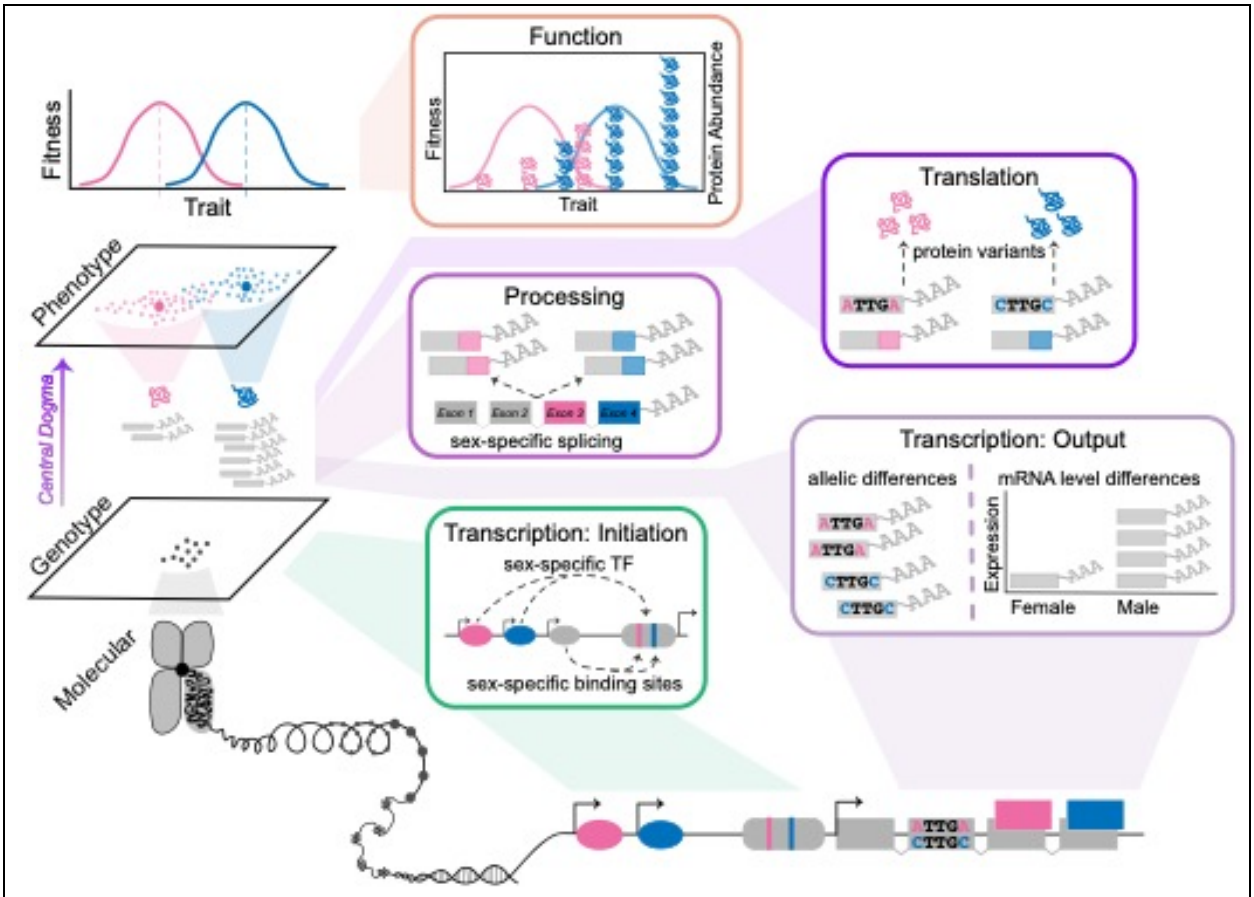
Kasimatis KR, Sánchez-Ramírez S, Stevenson ZC. 2021. Sexual Dimorphism through the Lens of Genome Manipulation, Forward Genetics, and Spatiotemporal Sequencing. *Genome Biology and Evolution* **13**. doi:10.1093/gbe/evaa243

INTRODUCTION

A central goal of evolutionary genetics is to understand the genetics of adaptation. This goal requires researchers to probe the genomic response to selection on phenotypes with a known fitness effect in nature (Barrett and Hoekstra, 2011). We can approach this challenging task by studying distinct components of the problem: mapping the genetic basis of the phenotype, measuring selection on the phenotype, and scanning the genome for signatures of this selection. Sexual dimorphism (SD) of phenotypes adds an additional layer of complexity, because the sexes maximize fitness differently and are subject to different selective pressures (Arnqvist and Rowe, 2005; Parker, 1979; Trivers, 1972). Yet, the sexes share the majority of their genetic material and, thus, SD is a function of shared, and of sex-specific and sex-biased genetic architecture, gene regulation, and gene expression. Therefore, to truly understand how sex-specific selection shapes the evolution of SD in genomes, it is essential to identify the molecular biology processes linking the genome with the phenome (Box 2.1).

Identifying the genetic variants and sex-biased networks underlying SD has proved challenging. In the last decade, research has centered on patterns of sex-biased gene expression, which has led to the identification of strong, consistent sex-biases coupled with rapid molecular evolution and genomic organization of sex-biased genes (Böhne et al., 2014; Cutter and Ward, 2005; Harrison et al., 2015; Innocenti and Morrow, 2010; Jin et al., 2001; Ranz et al., 2003; Reinke et al., 2000; Yang et al., 2006). While informative, these global patterns mask the underlying molecular mechanisms and often do not directly provide spatial resolution within the organism. These limitations hinder our understanding of how SD is cued within tissues and across developmental time. Finally, the focus on transcriptional patterns alone excludes other sources of phenotypic variation such as translation.

Combining molecular genetics with classic evolutionary approaches and genome technology, provides an opportunity to uncover the molecular mechanisms linking a sexually dimorphic phenotype with its underlying genetic basis. Such integration is feasible, efficient, and cost-effective in the emerging era of "4D genome technologies" and can provide high-resolution analyses of biological features in distinct physiological and tissue systems, and across developmental and evolutionary time. Using this integrated evolutionary framework, we can begin to address long-standing questions in the field, such as: What is the genetic architecture of sexually dimorphic traits? What are the genetic constraints on sexual dimorphism? What is the relationship between sex chromosome evolution and sexual dimorphism? and When, where, and how are sex-biased networks formed and how are they sustained across an organism's lifecycle? This perspective aims to provide a synthetic view of how 4D genome technologies integrated into evolutionary frameworks can uncover the mechanistic basis and genomic manifestation of SD with unprecedented detail. We suggest that these new paradigms will overcome an emerging recognition of limitations to existing approaches for deciphering signals of SD and sexually antagonistic selection.



Box 2.1. Molecular mechanisms that can contribute to sex differences.
To understand how sex-specific selection drives the evolution of traits in populations, we need some understanding of the underlying genetic basis of the traits, as highlighted in Lewontin's (1974) classic text. A critical component of Lewontin's genotype-phenotype map is the first transformation, which encompasses the central dogma of molecular biology: DNA to RNA to protein (as shown below). These molecular underpinnings are particularly important in the context of SD as the largely, and in some cases completely, shared genetic material is producing distinct phenotypes within each sex.

Phenotypic variation can arise through modifications to coding sequences including gene duplication, changes in gene regulation, and modifications during translation (Grath and Parsch, 2015; Khramtsova et al., 2019; King and Wilson, 1975; Levine and Tjian, 2003; Mank, 2017; Wyman et al., 2012). Sex differences can be generated by completely sex-limited

genes, often located on sex chromosomes (Mank, 2009) or by genetically encoded differences in the initiation of transcription (shown in green). Here, differences in transcription factor (TF) binding frequency between the sexes or sex-specific TF binding sites will drive sex-specific mRNA levels. Additionally, sex-biased deployment of master regulators can initiate a cascade of sexual differentiation.

Transcription can be divided into two stages: the production of mRNA transcripts followed by processing of these transcripts (shown in purple). Differences in transcriptional output between the sexes include overall differences in mRNA expression levels generated through either differential TF binding or sex-specific degradation of mRNA. Alternatively, a sexually antagonistic polymorphism can generate allelic differences in mRNA transcripts between the sexes. Ultimately, this effect is not realized unless the translated protein variants differ in form and function between the sexes (shown in dark purple). During the post-transcriptional regulation stage, sex-specific alternative splicing (Chang et al., 2011; Hartmann et al., 2011; Li et al., 2013) and small RNA regulators (Bezler et al., 2019; Warnefors et al., 2017) generate sex-specific mRNA (as shown in purple). As with a sexually antagonistic polymorphism, this effect is only realized if the sex-specific isoforms have protein variants that differ in form or function (shown in orange).

While not exhaustive, this list includes the major mechanisms that have been identified or hypothesized to contribute to sexual dimorphism. However, more work is needed to determine if one mechanism is more common than others or if the molecular mechanisms contributing to SD differ for simple versus complex traits.

APPROACHING OUTSTANDING QUESTIONS

To map the genetic basis and molecular mechanisms of a sexually dimorphic phenotype, we can manipulate selection, correlate genomic patterns with SD and genetic sex, and verify the functional importance of genes through genomic manipulation. We briefly explore five complementary experimental paradigms and highlight how they can link genotype, phenotype, and fitness across the lifecycle of each sex to provide the maximum temporal resolution of the genetic basis and molecular mechanisms of SD.

Evolve and resequence (E&R)

The E&R approach (Schlötterer et al., 2015) combines experimental evolution with whole-genome sequencing to trace allele frequencies over tens or hundreds of generations (Fig. 2.1A-C). Examining allele frequency changes permits an estimation of the strength of selection on regions of the genome that contribute to the SD of interest. E&R is a powerful approach to manipulate selection in a sex-specific manner to examine sex-biased genetic architecture and determine: if there are genomic hotspots of SD, the relative contribution of coding versus regulatory sequence, and the number of contributing loci. Experimental evolution approaches have successfully isolated sex-specific selection (Rice, 1996) and sexual selection (Chenoweth et al., 2008; Edward et al., 2010; Maklakov et al., 2009; Rostant et al., 2020; Snook et al., 2013), though few studies have examined the genomic response (Hsu et al., 2020). New transgenic technology will expand the potential of E&R to identify sex-biased elements of genetic architecture by creating high-throughput mechanisms for altering the variance in or manipulating the developmental timing of a sexually dimorphic phenotype, or isolating selection to a given sex. For example, introducing inducible knockdown technology (Supplemental Table 1) into the genome prior to E&R can provide a fine-scale experimental tool to alter gene expression in a sex-specific manner. By manipulating gene expression, the phenotypic mean can be shifted toward more or less SD, which will affect the response to selection. Inducible technology can aid in altering gene expression or the timing of gene expression, both of which will affect the sex-specific response to selection. Limiting selection to act within one sex during E&R will also be aided by tools that remove a phenotype in one sex, such as inducible sterility (Kasimatis et al., 2018) or generate progeny of a single sex (Douglas et al., 2020).

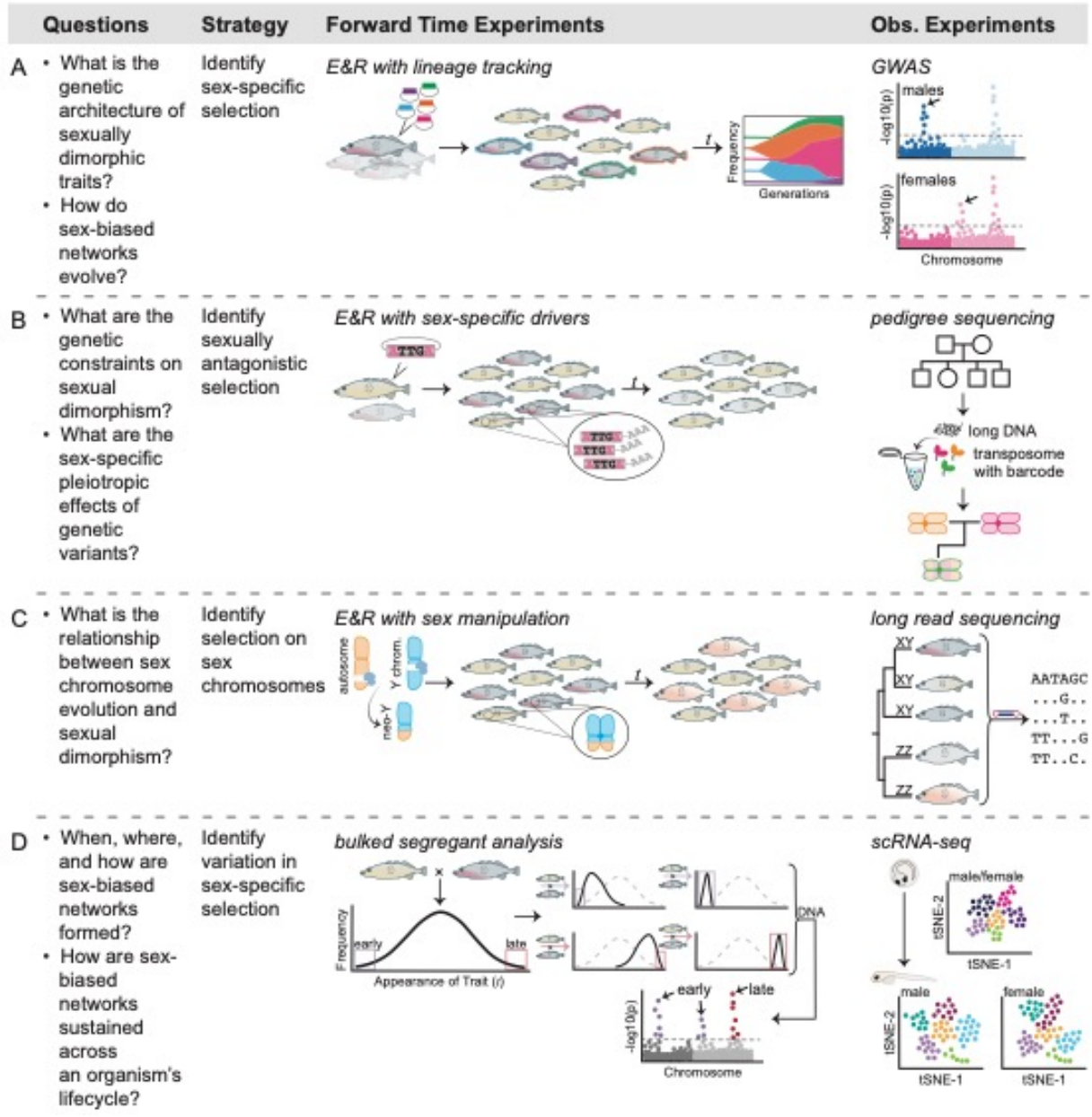
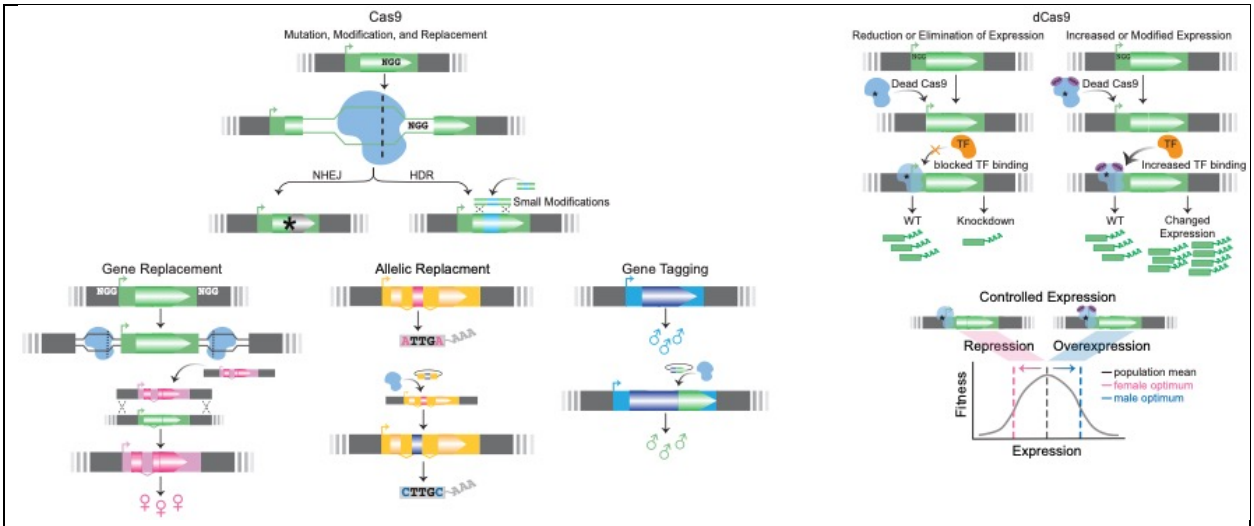


Figure 2.1 Evolutionary frameworks for identifying the genetic basis of SD. Strategies for identifying: (A) the genetic basis of SD through experimental evolution and between-sex comparative genomics, (B) sexual antagonism and SD through experimental evolution and pedigree tracing, (C) sex chromosome directed SD through experimental evolution and comparative genomics, and (D) developmental SD through bulked segregant analysis and single-cell sequencing.

Alternatively, transgenic technology can expand our understanding of the genomic response during E&R through haplotype barcoding. This tool would provide a method to track the frequency of individual haplotypes (Fig. 1A; see "Genetic Manipulation"). Haplotype barcoding is ideal for tracking fitness differences between genetic variants in a competitive setting and could be applied to sexually dimorphic variation within and between the sexes. For example, haplotype barcodes could be integrated into multiple genetic backgrounds, crossed to form an ancestral population, and evolved under sex-specific selection or inter-sexual competition. Unlike traditional E&R experiments which use whole-genome sequencing of the ancestral and evolved population, haplotype barcoding allows for individuals to be sequenced in the smaller barcode at high coverage across many time points throughout evolution. This repeated sequencing provides high-resolution allele frequency traces of different haplotypes to understand the dynamics of genetic variants and how they relate to fitness changes over time. Haplotype barcoding can also be adapted to study differences in recombination rate between the sexes by integrating multiple barcodes at known positions. Differences in recombination rate between the sexes can influence sex-specific genetic architecture (Sardell and Kirkpatrick, 2020; Trivers, 1988), which could link sex-biased elements contributing to SD. Barcoding at known genomic regions will allow us to follow this process in forward-time experiments and compare recombination rates between the sexes under different environmental conditions.

E&R can also be used to address the relationship between sexually antagonistic selection and SD by generating a negative genetic correlation between female and male fitness (Bonduriansky and Chenoweth, 2009) through manipulating sexual selection (Scott Pitnick et al., 2001; S. Pitnick et al., 2001; Rice, 1996). Again, genetic manipulations introduced before E&R, such as inducible knockdowns or altered expression through CRISPRi (Supplemental Table 1; Box 2.2), will facilitate understanding the pleiotropic effects of genetic variants in a sex-specific manner. Manipulating expression during E&R can also provide information on how a gene's interactions are structured within each sex and how these interactions evolve. Additionally, genomic editing can be used to introduce a sex-biased regulatory sequence or genetic variant into the mismatched sex, which will relax the degree of sexual antagonism during E&R and thus reduce the genetic constraints on SD (Fig. 2.1B; Box 2.2). Genomic editing also provides the ability to manipulate the sex-determining region and fuse chromosomes together (Shao et al., 2019) to study neo-sex chromosome formation. When used in an E&R framework, this approach will provide insight in



Box 2.2. Genetic manipulation through CRISPR/Cas

CRISPR (clustered regularly interspaced short palindromic repeats) has become the premier method for mutating and editing the genome with precision (Doudna and Charpentier, 2014; Jinek et al., 2012; Pickar-Oliver and Gersbach, 2019). While the vast scope of CRISPR utilities cannot be represented here, we outline three mechanisms and their application to SD. Cas9, the most common nuclease associated with the CRISPR system, is a targetable nuclease, which provides experimental control over the location of the nuclease activity. Cas9 and other CRISPR-associated nucleases are 'guided' by specific single-stranded RNA encoding the sequence of interest. The PAM (Protospacer Adjacent Motif) sequence restricts the locations of Cas9 targeting. Cas9 targets guide site locations with 'NGG' PAM sequence and cleaves double-strand DNA (shown below). When Cas9 creates a break, the cell will attempt to repair the break by non-homologous end joining (NHEJ) or homology directed repair (HDR). NHEJ is error-prone and thus a simple way to create mutations in the desired gene. However, depending on the mutation's location, the protein may still fold and function properly. HDR modifies a gene in a specific manner through a 'donor template' to repair the DSB (reviewed in Doudna & Charpentier 2014). This process can use the homologous chromosome as a template or non-native DNA template sequence can be introduced to the genome. The synthetic construct (e.g., plasmid, PCR product, or single-stranded oligo) must contain sequences homologous to the genome on both the 5' and 3' sides of the desired insert to co-opt the genome repair machinery for integration. HDR has many applications in the study of SD (as shown below). For example, whole genes can be deleted to determine the gene's sex-

specific function. Alternatively, whole genes or promoter regions can be replaced with a sex-biased version to test sex-specific function. A more focused approach would directly edit alleles to verify the significance of sexually antagonistic polymorphisms. Beyond verifying the function of genetic elements in generating SD, HDR can also be used to add epitopes or fluorescent tags to aid during experimentation (see Fig. 1 and "Genetic Manipulations").

Catalytically inactive variants of Cas9, dead Cas9 (dCas9), target and bind to the sequence without cutting the genome (shown below). In doing so, the dCas9 protein physically blocks the binding of transcription factors and creates a knock-down phenotype (Larson et al., 2016; Pickar-Oliver and Gersbach, 2019). This utility, termed CRISPRi, is analogous to RNAi, and may complement the knockdown phenotype or be an adequate replacement (Stojic et al., 2018). Alternatively, dCas9 can be fused to various transcription activators, such as VP64, to change the target genes' expression level and specificity (reviewed in Pickar-Oliver & Gersbach 2019). By changing the regulatory sequences driving dCas9 expression, specific genes can be experimentally altered in a tissue-specific or cell-specific manner. Experimentally, this manipulation of expression opens many possibilities to directly manipulate genetic architecture in a sex-biased way.

While CRISPR's successes have been widespread (Supplemental Table 1), unique challenges exist for species-specific utility. Targeting the genome requires knowing the DNA sequence and the genomic location of sequences. Adapting CRISPR for new species or highly divergent strains may be complicated if this information is lacking. Additionally, gene duplications and pseudogenes which arose from duplications can also pose a challenge since they likely share much of their sequence in common. Again, high-quality genome assemblies can help to control for this problem, though off-target CRISPR effects can still occur. Finally, delivery of Cas9 and the guide RNA is unique to each species and will require optimizing the method with that species to achieve targeted genome-editing. Nevertheless, CRISPR has been widely adopted among may species and non-models with great success; recent protocols developed for non-model organisms, including firebug *Pyrrhocoris apterus* (Kotwica-Rolinska et al., 2019), malaria mosquitos (Hammond et al., 2017), and lizards (Rasys et al., 2019). This continued methodological progress, coupled with advances in sequencing technology, will expand the potential applications of CRISPR across taxa.

real-time on how neo-sex chromosomes evolve and the role of sex chromosomes in resolving sexual conflict (Fig. 2.1C).

Despite their power, E&R studies are still sensitive to population size, biological replication, and ancestral haplotype structure (Kofler and Schlötterer, 2014; Otte et al., 2021). Importantly, the molecular integration of transgenic elements homogenizes the genome. To create standing genetic variation for selection to act upon during E&R will require either crossing multiple transgenic strains together or mutagenesis. In the case of crossing, haplotype structure must be carefully considered during experimental design as it can greatly impact the sex-specific response to selection. This approach to E&R relies on manipulating the genome of particular isolates or strains rather than following the genomic response of segregating genetic variants and therefore may not represent all possible evolutionary pathways observed in natural populations. These limitations can be avoided by using transgenics tools only after E&R to verify candidate genes. For example, CRISPR (Box 2.2) can replace a haplotype in the ancestral background with the evolved haplotype (Perli et al., 2020). However, the full benefits of transgenics during E&R will be realized when used as an integrated tool. This goal can be met as transgenics become more efficient and feasible in a range of taxa (Supplemental Table 1), allowing for multiple strains to be genetically manipulated and crossed.

Bulked-segregant analysis

An alternative approach to E&R is bulked segregant analysis (Brauer et al., 2006), which uses selection on the tails of a phenotypic distribution to map the genetic basis of extreme phenotypes. By repeating over multiple rounds of selection, the variance in the trait can be reduced, which facilitates mapping. Bulked segregant analysis could be a powerful approach to mapping genetic variants of SD and understanding the role of dominance in sexually dimorphic traits, both of which will benefit from existing introgression lines between strains or species. This approach may also be particularly useful for selecting on sex-specific variation during development to map the genetic basis of when and how SD is generated (Fig. 2.1D). This approach can be coupled with CRISPR transgenics to validate the function of candidate genes (see "Genetic Manipulation"). Alternatively, RNA-sequencing and particularly single-cell sequencing can be used to identify expression differences between bulk populations (Ben-David et al., 2020).

Pedigree tracing

Pedigree and parent-offspring trio sequencing offer an alternative to E&R for studying sexual dimorphism in populations that are not conducive to experimental evolution, such as in natural populations or organisms with a long generation time (Bates et al., 2020; Johnston et al., 2017; Lucotte et al., 2020). These approaches explicitly correlate haplotype structure with genetic sex and identify recombination events within a population. Additionally, pedigree and trio sequencing approaches explicitly take into account population structure, providing an advantage over genome-wide association studies (Bates et al., 2020). Pedigree tracing has already proved powerful for identifying signatures of selection in wild populations (N. Chen et al., 2019; Johnston et al., 2017). Advances in long-read genomic sequencing and reduced sequencing costs are making these approaches more feasible across taxa. A promising emerging framework being adapted from human genomics is the use of linked-read sequencing to gain insight on phased genomes (Lutgen et al., 2020). Specifically, phased genomic information within a pedigree framework can be used to correlate haplotype structure and local genetic architecture with genetic sex. Additionally, phased genomes gained through linked-sequencing or parent-offspring trio sequencing can be used to study segregation distortion and sexually antagonistic variants within the genome (Lucotte et al., 2020).

Genome-wide association studies (GWAS) and high resolution transcriptomics

An alternative approach to manipulating selection is to observe the genomic footprint of selection in natural populations through GWAS or expression association approaches. GWAS provide a powerful approach to associate sex-specific variation in a phenotype with its underlying genetic basis (Fig. 2.1A). Taking such a sex-stratified approach will distinguish sex-specific allelic effects (Khramtsova et al., 2019). While GWAS is sensitive to population demographics, these confounding effects can be controlled for in a logistic regression framework. An association framework also can be used for gene expression data with the potential to reveal how *cis* and *trans* genomic variants influence transcription on a genome-wide scale (Sun and Hu, 2013). This framework will be especially powerful when coupled with cell or tissue-specific transcriptomes (see "4D Transcriptomics"). The human Genotype-Tissue Expression project (GTEx) is revolutionizing this area of research, identifying over one-third of genes to have a sex-biased expression profile in at least one of the 44 tissues sampled (Oliva et

al., 2020). For both data types, sample size will impact the ability to accurately detect signatures of selection and may be a limiting factor for some natural population studies.

Comparative transcriptomic studies provide network-level information about sex-specific node connectivity and redundancy. By coupling classic evolution-development frameworks, particularly during early development, with single-cell sequencing technology (see "4D Transcriptomics), we can begin to create a continuous understanding of SD through time (Fig. 1D). Spatial transcriptomics has already transformed developmental biology (Farrell et al., 2018; Zhou et al., 2019) and sex-stratified approaches with these methods will only further our knowledge.

Comparative genomics and long-read sequencing

Comparative genomic studies focusing on sex chromosomes isolate the genotype space of the genotype-phenotype-fitness map (Fig. 2.1C). The relationship between the origin of sex chromosomes and SD is a long-standing area of research (Bachtrog et al., 2011; Charlesworth, 1991; Mank, 2009; Rice, 1984), however, the quality of genome assemblies has been a major limiting factor, especially for non-model organisms. Traditional methods for studying sex chromosomes are also benefiting from new technology. Specifically, long-read PacBio and Oxford Nanopore sequencing (Amarasinghe et al., 2020) are providing chromosome length scaffolds for assembling short-read data. These methods generate high-quality assemblies that extend through repetitive regions and tandem duplications, which are problematic for short-read data, but may be common and potentially important components of sex chromosomes (Bachtrog et al., 2019; Bracewell and Bachtrog, 2020; Peichel et al., 2019). Similarly, gene duplication and sex-specific functionalization is viewed as an important mechanism leading to the resolution of sexual conflict and the evolution of SD (Connallon and Clark, 2011; Gallach et al., 2010; Gallach and Betrán, 2011; Wyman et al., 2012). Long-read sequencing can help identify and disentangle recent duplication events more accurately than standard short-read data. Finally, long-read RNA sequencing technologies, such as Iso-seq, are proving to be powerful in identifying sex-specific alternative splicing and the role of this mechanism in the development of sexually dimorphic traits (Zhao et al., 2019).

SPATIAL AND TEMPORAL PATTERNS OF SEX-BIASED EXPRESSION AND REGULATION

Recent advances in sequencing technologies, such as tomographic or spatial transcriptomics (Tomo-seq) and single-cell RNA sequencing (scRNA-seq), allow us to track transcriptomic dynamics across different cell types and tissues, and across development to provide fine-scale resolution of SD in gene expression and regulation. We discuss four methods, which can be used independently or coupled with experimental manipulation to observe the patterns of SD.

4D transcriptomics of SD: scRNA-seq and Tomo-seq

Single-cell sequencing expands the feasibility of quantitative gene expression methods across taxa and biological samples. Specifically, single-cell RNA amplification techniques coupled with cell sorting devices offer a major advantage over bulk-cell RNAseq by providing transcript-level expression for thousands or even millions of cells (Haque et al., 2017; Hashimshony et al., 2016; Islam et al., 2011; Tang et al., 2009). More recent analytical improvements are enabling post-sequencing identification of cell populations by applying advanced clustering and unsupervised learning techniques, such as t-distributed stochastic neighbor embedding (t-SNE), greatly improving the spatial resolution in scRNA-seq data (Kobak and Berens, 2019). Although these techniques have largely been used to distinguish gene expression profiles between cell populations within a single individual, scRNA-seq comparisons between the sexes in humans is beginning to unveil the mechanisms of SD (Tukiainen et al., 2017). Additionally, new analytical approaches are enabling differential single-cell expression contrasts between individuals(Becht et al., 2020; Butler et al., 2018; Ntranos et al., 2019), which will facilitate contrasts between cell populations of females and males. While complexity and expense can build up for an experiment with female and male treatments and multiple developmental time points, a cost-effective, although less high-throughput alternative, is quantitative PCR to monitor pivotal genes on specific cell populations that may have been identified in coarser scans (VanInsberghe et al., 2018). Importantly, scRNA-seq will not only provide an understanding of sex-biased differential expression through development (Fig. 2.1D), but can also be used to understand the sex determination cues from sex chromosomes (Fig. 2.1B) with spatial and cellular resolution.

Alternatively, Tomo-seq avoids the cell sorting and classification required for scRNA-seq by providing genome-wide gene expression quantification in contiguously cryo-sliced whole-

body segments (Combs and Eisen, 2013; Junker et al., 2014; Kruse et al., 2016). Organisms and developmental stages with low-dimensional bodies, such as embryos, larvae, and worms, are emerging as ideal systems to examine gene expression along anteroposterior, dorso-ventral, and lateral dimensions (Combs and Eisen, 2013; Ebbing et al., 2018; Junker et al., 2014). Recent Tomo-seq work in *Caenorhabditis elegans* comparing hermaphrodite and male expression patterns identified the location of genes with sex-biased expression outside of reproductive tissues (Ebbing et al., 2018). While size remains a limitation for larger-bodied organisms, this technique could be applied, in some cases, to distinguish spatial and functional differences between organs or other low-dimensional structures of females and males (Wu et al., 2016).

Measuring sex-specific transcription binding activity: ATAC-seq

Sex-specific regulation can arise in part from transcription factors binding to open chromatin (Box 1), yet most of the evidence we have about sex-specific regulation comes from indirect studies of *cis*- and *trans*-regulatory changes in inter-species and inter-population hybrids (Coolon et al., 2018; Meiklejohn et al., 2014; Turner et al., 2014). To directly address the role of regulation variation in SD, chromatin immunoprecipitation and sequencing (ChIP-seq) and related methods (Naqvi et al., 2019) can be used to quantify DNA-protein interactions in high-throughput manner. However, they require *a priori* knowledge of specific protein targets and large amounts of starting material (Jiang and Mortazavi, 2018). Alternatively, the assay for transposase-accessible chromatin sequencing (ATAC-seq, Supplemental Table 1) is the next iteration of genome-wide DNA-protein interaction assays and overcomes some of these shortcomings by: directly accessing open chromatin enzymatically with the hyperactive Tn5 transposase, not requiring protein-specific markers, allowing for low amounts of starting material, and being time- and cost-efficient (Buenrostro et al., 2013; Yan et al., 2020). Additionally, ATAC-seq is more sensitive, which decreases the signal-to-noise ratio seen in ChIP-seq, and can be integrated into a single-cell sequencing framework (e.g., scATAC-seq). In *C. elegans*, novel regulators have been uncovered using ATAC-seq, revealing complex regulatory dynamics across developmental stages (Daugherty et al., 2017). Overall, ATAC-seq has potential to examine broadly distributed regulatory regions across the genome, which can help disentangle sex-specific binding activity both spatially within the organism and across development.

GENETIC MANIPULATION FOR HYPOTHESIS TESTING

Many toolkits have been devised to manipulate the genetic architecture and expression of specific genes, allowing for spatio-temporal control and visualization of gene expression to manipulate SD and verify candidate genes. We discuss the feasibility and technical limitations of CRISPR gene engineering (Box 2.2) and highlight four established toolkits, which CRISPR made more accessible.

Expression control through Gal4/UAS

The Gal4/UAS system allows for spatio-temporal control of gene expression by splitting the regulation and coding sequence to independently investigate the effects of regulation versus transcription levels on gene function (Supplemental Table 1). Utilizing sex-specific promoters to drive Gal4 expression allows for feminization or masculinization of specific tissues (Fig. 2.1B). For example, sex-specific Gal4 drivers were used to investigate SD in *Drosophila* sleep behaviors (Khericha et al., 2016) and pathology (Regan et al., 2016). Extending these studies to a multi-generation framework will allow for selection to be manipulated in a sex-biased manner to understand the effect of sex-biased regulation on population fitness (Fig. 2.1). While this system provides a powerful approach to controlled gene expression, native gene expression is not strictly conserved (Wang et al., 2017) and must be considered during experimental design and interpretation.

Expression control through Cre-lox

Cre-lox allows for deletion of specific sequences (Gu et al., 1994), translocation of chromosome fragments (Deursen et al., 1995), inversion of gene orientation (Grégoire and Kmita, 2008), and integration of transgenes (Levy et al., 2015) to manipulate genetic architecture in a controlled manner. Two lox sites are integrated for genetic deletions, flanking the desired sequence to be deleted (Supplemental Table 1). The expression of Cre induces recombination of the two lox sequences, excising the intermediate stretch of DNA between them. Other utilities simply rely on changing the orientation or location of the lox sites. Under tissue-specific promoters, Cre expression can be controlled spatially and temporally to manipulate sex-specific constraints on SD or alter developmental cues (Fig. 2.1B, D). Cre-lox has been used to investigate sexually-dimorphic behavior and delete the testosterone androgen receptor in mice (Juntti et al., 2008). While Cre-lox can provide precision control over the desired genetic manipulation, several Cre drivers have transient expression and can lead to the Cre recombinase activity in undesired cells

and tissue types (Song and Palmiter, 2018). To overcome this obstacle, several 'split-Cre' systems can drive portions of the Cre recombinase protein under different drivers, allowing for higher specificity (Hirrlinger et al., 2009).

Expression control through targeted knockdowns

We can learn about the molecular function underlying SD through controlled and targeted depletion of gene products in both permanent and inducible contexts. Knockdown methodologies, such as RNA interference (RNAi, Supplemental Table 1) can be used to suppress expression, which provides a powerful tool for examining expression variation between the sexes (Fire et al., 1998). RNAi causes the knockdown of a gene by eliminating the genes' mRNA by injecting double-strand RNA or *in vivo* expression (Crotty and Pipkin, 2015; Dzitoyeva et al., 2001), and has been adopted in a wide variety of organisms, including humans (Setten et al., 2019). RNAi can be used to verify the necessity and sufficiency of candidate sexually dimorphic genes. For example, RNAi was used to identify the molecular basis of a color SD in the queenless ant, *Diacamma* sp. (Miyazaki et al., 2014), to examine the function of water strider male antennae during mating (Khila et al., 2012), and to test female and male fertility genes in *Drosophila* (D. S. Chen et al., 2019; Chen et al., 2012; VanKuren and Long, 2018). While RNAi is a powerful and widely applicable technology across taxa, the effect can be weak and non-specific degradation can occur (Boutros and Ahringer, 2008). In some cases, CRISPRi can overcome these limitations and provide a substitute for RNAi (Box 2.2).

The auxin-inducible degradation (AID, Supplemental Table 1) system has recently been utilized for targeted gene knockdown (Nishimura et al., 2009). AID uses a transgenic plant protein, TIR1, which recognizes a small specific degron tag on a protein of interest and degrades this protein in the presence of auxin. The degron tag can be added to native genes by CRISPR (Box 2.2), or transgenic integrations of genes with the degron tag can be introduced into a wildtype or mutant background. Importantly, AID has higher specificity compared to RNAi and temporal control is simpler to achieve through the addition of auxin. Despite its power, AID is sensitive to the concentration of auxin, less permeable in some tissues, and auxin-independent degradation has been observed (Papagiannakis et al., 2017; Schiksnis et al., 2020; Zhang et al., 2015). AID has successfully been used for protein depletion in cell culture and animal models (Holland et al., 2012; Kanke et al., 2011; Kasimatis et al., 2018; Zhang et al., 2015), except in

zebrafish where the current form of the AID system has a limited effect (Yamaguchi et al., 2019). To the best of our knowledge, AID has not been specifically applied to questions of SD, however, this method is ideal for manipulating sex-limited selection in an E&R framework (Fig. 2.1A).

Haplotype tracking through fluorescent reporters and barcoding

The ability to visually mark when and where a gene gets expressed is arguably the most basic and essential tool utilized by molecular genetics to investigate genetic architecture and can provide a visual context for expression differences between the sexes. Fluorescent reporters can be tagged to a native protein or act as an independent transgene (Box 2.2) and have been developed in many color variants for a wide range of utilities (Rodriguez et al., 2017), including competition experiments to identify adaptive lineages (Crombie et al., 2018; Hegreness et al., 2006) and sex-stratified experiments to parse sexually-dimorphic gene expression (Serrano-Saiz et al., 2017). However, some fluorescent reporters are very dim depending on the transcriptional activity, and translational reporters, in some cases, can disrupt protein activity, which prevents the incorporation of a fluorescent tag.

While fluorescent reporters allow for simple identification, the total number of reporters are significantly limited. High-throughput approaches that include neutral genomic-integrations – namely barcodes – have recently been adapted to study adaptive lineages in yeast and bacteria (Blundell and Levy, 2014; Jasinska et al., 2020; Levy et al., 2015). While barcoded lineage-tracking has not been explicitly adopted in animal systems, unique lineage identification has been implemented in competitive experiments utilizing reporters (Marie-Orleach et al., 2016). Expanding on fluorescent reporter marked lineages, various sex-specific lineages could be created and marked for competition experiments (Fig. 2.1A). After overcoming the technical limitation of genomic barcoding, high-throughput lineage tracking will be the next great breakthrough in experimental evolution in animal systems.

CONCLUSIONS

SD constitutes much of the diversity observed between organisms and is integrated across the genotype-phenotype-fitness map. By harnessing cutting-edge methods developed for molecular biology and biomedical research, we can design explicit experiments to address how this remarkable diversity evolved from a shared genome. With few exceptions, the technological advancements discussed here will allow us to increase the spatial, temporal, and molecular resolution of the underpinnings of SD, and expand our ability to implement molecular and genetic studies in non-model organisms. The field is poised to synergize advances in molecular biology and sequencing technology within evolutionary frameworks, promising novel insights on the creation and maintenance of SD in the near future.

Bridge

Here, we have outlined a number of potential innovative uses of genomic engineering in addressing evolutionary questions. In the next chapter, I describe our approach to simplifying additive transgenics in *Caenorhabditis elegans* utilizing CRISPR/Cas9. As mentioned above, the ability to edit the genome can provide insights into the process of evolution. Our methodology increased the speed of and facilitates a simpler transgenesis process than had been possible before.

CHAPTER 3 RAPID SELF-SELECTING AND CLONE-FREE INTEGRATION OF TRANSGENES INTO ENGINEERED CRISPR SAFE HARBOR LOCATIONS IN *Caenorhabditis elegans*

This chapter was published in volume 10 of the journal G3 in 2020. Megan J. Moerdyk-Schauwecker, Brennen Jamison, and Patrick C. Phillips are co-authors on this publication. Moerdyk-Schauwecker and I designed the technology. Megan J. Moerdyk-Schauwecker and Brennen Jamison created the various strains. Megan J. Moerdyk-Schauwecker and I analyzed the data. Megan J. Moerdyk-Schauwecker, Patrick C. Phillips, and I wrote the manuscript.

The full supplemental material for this publication can be found at:
https://doi.org/10.25387/g3.12469700

The citation for this publication is as follows:

Stevenson ZC, Moerdyk-Schauwecker MJ, Jamison B, Phillips PC. 2020. Rapid Self-Selecting and Clone-Free Integration of Transgenes into Engineered CRISPR Safe Harbor Locations in *Caenorhabditis elegans*. *G3: GenesGenomesGenet* **10**:3775–3782. doi:10.1534/g3.120.401400

INTRODUCTION

The introduction of transgenes is a staple in the molecular biologist's toolkit, with a broad range of utilities including expression of individual variants, ectopic expression of tagged native genes, and the addition of genes from other species. Injection of double-strand DNA into the *Caenorhabditis elegans* gonad arm generally results in assembly of these fragments via regions of microhomology, leading to the formation of extrachromosomal arrays (Mello et al., 1991; Stinchcomb et al., 1985). These extrachromosomal array structures can be in excess of 1 Mbp and contain up to hundreds of copies of the injected gene (Mello et al., 1991; Woglar et al., 2020). Extrachromosomal arrays are not stably inherited—either between cells within an individual or between generations—and have variable expression levels, which can be problematic depending on the biological question. To avoid the stochastic element of array

expression, it is often desirable to integrate transgenes. Historically in *C. elegans*, microparticle bombardment (Praitis et al., 2001) and ultraviolet light exposure (Evans, 2006) or gamma exposure (Mello and Fire, 1995) have been used to integrate transgenes randomly. However, these methods are less than ideal as: integration usually results in multiple copies, which can impact expression; the integration location is random which can disrupt expression of native genes or insert the transgene into regions prone to transcriptional silencing; and both methods require expensive and specialized equipment to create transgene integrations and to identify the loci of insertion. More recent approaches have utilized transposon-based integration methods such as MosSCI and miniMos (Frøkjær-Jensen et al., 2014, 2008), which use a transposon to create a double-strand break, allowing for single transgene integration into a predefined or random region respectively. Recently, CRISPR/Cas9 techniques have been adopted for transgene integration. These approaches include generalized methods for transgenic cargo insertion, such as those using a selective marker like Hygromycin B resistance (Chen et al., 2013) or a self-excising cassette (SEC) (Dickinson et al., 2015; Kasimatis et al., 2018). More specialized CRISPR strategies, such as the SKI LODGE method, facilitates tissue-specific expression by splitting the coding and promoter element (Silva-García et al., 2019). A particular advantage of this latter strategy is that it introduces modular transgene integration, allowing for a more straightforward integration into a backbone of standard genetic elements that are pre-integrated within a safe harbor location.

Regardless of whether a transposon or CRISPR/Cas9 strategy is used, integration of the transgene by homology-directed repair (HDR) is generally inefficient compared to non-homologous end joining (NHEJ) (Dickinson et al., 2015, 2013; Frøkjær-Jensen et al., 2008; Ward, 2015). Both MosSCI and CRISPR show approximately the same integration efficiencies which varies greatly depending on the transgene. CRISPR and MosSCI also require robust screening methods to identify the rare correct transgene integration. Co-CRISPR simultaneously targets a second gene to generate a visible phenotype (e.g., *dpy-10*) (Arribere et al., 2014; Kim et al., 2014), thereby allowing identification of a sub-population with active Cas9 expression and genome targeting. This enriched population must then be further screened, generally by PCR and Sanger sequencing, to identify correct integration events at the desired locus. In general, co-CRISPR is not widely used for transgene integrations, though there are exceptions (Farboud et al., 2019; Silva-García et al., 2019). In the case of selectable genes, the transgene generates a

visible phenotype displayed not only by integrants but also individuals with heritable extrachromosomal arrays because the injected donor homology contains a fully functional copy of the selectable gene. Distinguishing extrachromosomal arrays from integration events requires anti-array selection techniques such as heat shock induction of *peel-1*, a toxic transgene (Frøkjær-Jensen et al., 2012; McDiarmid et al., 2020; Seidel et al., 2011) and visual screening for loss of an additional gene used to mark the array (e.g., fluorescent protein or *rol-6*). These methods are imperfect and molecular methodologies such as genotyping PCRs must be used to verify genuine integrations. Currently, no method is available for *C. elegans* that provides integration-specific selection of transgenes. Such a screen would fit into the category of a "screen from heaven" where only the desired transgenic integrant is alive on the petri dish (Jorgensen and Mango, 2002).

In most cases, transgenes must be cloned into plasmids with homology arms matching the targeted genomic region for single-copy integration. This process requires various cloning strategies for each desired transgene. *C. elegans* can recombine fragments with microhomology and express resulting transgenes in an array (Kemp et al., 2007; Mello et al., 1991) Others have tested this strategy to create a donor homology for transgene integration. For example, Paix *et al.* (2016) and Philip *et al*. (2019) attempted to overcome the cloning obstacle by integrating transgenes with overlapping PCR fragments that, once recombined *in-situ*, should produce a functional gene. However, as with plasmid-based templates, neither method provides direct selection for the transgene integration. As such, depending on the design, array formation can provide false-positives, increasing the difficulty of identifying a correct assembly and integration—a notable complication.

Here, we present a novel transgene integration strategy that utilizes a custom-designed safe harbor location to eliminate many of the steps required to go from concept-to-integrated transgene. Our approach removes the selective advantage from the array and selects only for the integration event by splitting the coding sequence for Hygromycin B resistance. Additionally, we show the cloning stage can be bypassed in this system, utilizing the worms' native homology mediated repair to clone our transgene *in-situ*. Coupling these methods can reduce the labor and time required to produce a transgenic nematode, allowing the experimenter to go from PCR-to-integrated transgene in approximately one week.

Materials and Methods

Strains and growth conditions

Bristol N2-PD1073 (Yoshimura et al., 2019) and the derived strains PX692, PX693, PX694, PX695, PX696, and PX697 (Table S1) were maintained on NGM-agar plates seeded with OP50 or HB101 *Escherichia coli* at 15C unless otherwise noted.

Molecular biology

All plasmids unique to this publication are listed in Table S2, and all primers used in this study are listed in Table S3. All-in-one-plasmids encoding both Cas9 and the desired sgRNA were created by site-directed mutagenesis of pDD162 (Addgene #47549) (Dickinson et al., 2013) (using the Q5 site-directed mutagenesis kit (NEB) per manufacturer directions. Guide and Cas9 sequences were confirmed by Sanger sequencing. The guide targeting Chromosome II:8420188-8420210 has been previously described, and the constructed plasmid (pMS8) is equivalent to pDD122 (Addgene #47550) (Dickinson et al., 2013). Synthetic guide sites utilized in the landing pads were based on guides previously shown to be highly efficient in other species or generated based on predicted efficiency scores (Table 1). Predicted off-target effects were determined using the method of (Doench et al., 2016) while predicted on-target efficiency was calculated using Sequence Scan for CRISPR (Xu et al., 2015) and the method of (Hsu et al., 2013). Repair template plasmids were assembled from overlapping fragments using the NEBuilder HiFi Kit (NEB) per manufacturer instructions. For the landing pads, the *Cbr-unc-119* rescue gene and a portion of the homology arms containing the guide site were removed from pCFJ151(Addgene #19330) (Frøkjær-Jensen et al., 2008) and replaced with a multiple cloning site to create pMS2. The SEC from pDD285 (Addgene #66826) (Dickinson et al., 2015) was then inserted into SacI digested pMS2 to create pMS4. The C-terminal portion of the hygromycin resistance gene and the *unc-54* 3' UTR were then independently amplified from pCFJ1663 (Addgene #51484) (from the lab of Erik Jorgensen) and inserted into the SbfI site of pMS4 to create the final landing pad plasmids (pMS70-75). The six synthetic guide sites were included in the primers used to amplify the hygromycin resistance fragment (Table 3.1, Table S3.2, and Table S3.3). A complete annotated sequence of pMS74 can be found in Figure S3.1.

Table 3.1 SLP guide efficiency for insertion of *rpl-28p::mKate2*

Strain	Guide Sequence	Reference	Doench Score[a]	SSC Score[b]	Specificity Score[c]	Marker Positive Broods	HygR Resistant Broods (%)	Correct Integration Broods
PX692	GTTTGAG TAGAGCA CTCAG**AG G**	Kane et al. (2017).	66.9	0.7991	99.3	59	5 (8.5%)	3 (5.1%)
PX693	GACAGTG GACATCT AAGCGG**A GG**	Kane et al. (2017).	61.5	1.2308	100.0	60	1 (1.7%)	1 (1.7%)
PX694	GTCCAGC GGCAGAT CGGCGG**A GG**	Ge et al. (2016)	45.1	1.0511	99.7	73	7 (9.6%)	5 (6.8%)
PX695	GAGTTCT GTAATTC AGCATA**A GG**	Agudel et al. (2017).	52.8	-0.0095	99.0	74	1 (1.4%)	1 (1.4%)
PX696	GGACAGT CCTGCCG AGGTGG**A GG**	Varshney et al. (2016)	40.9	0.5977	99.6	76	6 (7.9%)	5 (6.6%)
PX697	GGGGCCT GTGAAAT ACACAG**A GG**	N.A.	84.1	0.9981	99.2	77	4 (5.2%)	4 (5.2%)

[a]Predicted guide efficiency as per (Doench *et al.* 2016)

[b]Predicted guide efficiency as per (Xu *et al.* 2015)

[c]Predicted off-target effects as per (Hsu *et al.* 2013)

To generate an *rpl-28p*::*mKate2*::*unc-54 3'UTR* reporter, *rpl-28p* amplified from pBCN39-R4R3 (Addgene #34914) (Semple et al., 2012) and the mKate2 coding sequence and *unc-54* 3'UTR amplified from pDD285 were inserted into SacI digested pMS2 to give pMS12. The reporter was then amplified from pMS12 and inserted into an intermediate construct containing: *rps-0p* and the N-terminal fragment of the hygromycin resistance gene from pCFJ1663 (amplified in two fragments to remove intron), a pUC57 backbone, a truncated 5' genomic homology arm from pMS2 and artificial sequences; to give the final insertion vector pMS81. A complete annotated sequence of pMS81 can be found in Figure S2. A second split hygromycin insertion vector, pZCS52, was made by amplifying the homology arms and split hygromycin from pMS81 by PCR and adding the *sqt-1(e1350)* gene amplified from pDD285.

To generate an additional fluorescent co-injection marker, *eft-3p* and *tbb-2* 3' UTR amplified from pDD162 and wrmScarlet amplified from pSEM89 (a gift from Thomas Boulin) (Bindels et al., 2017; Mouridi et al., 2017) were cloned into a pUC19 backbone to give pZCS16. The Cre expressing plasmid pZCS23 was made by PCR amplifying the backbone, *eft-3p* and *tbb-2* 3' UTR from pZCS16 and adding *NLS::Cre* from synthetic gBlocks (IDT).

Strain generation by CRISPR/Cas9

A mixture consisting of 50 ng/µl pMS8, 10 ng/µl of the appropriate landing pad plasmid and 2.5 ng/µl pCFJ421 (Addgene #34876) (Frøkjær-Jensen et al., 2012) was microinjected into the gonad of young adult N2 hermaphrodites. Screening and removal of the SEC were done following Dickinson *et al.* (2015). Presence of the insertion and removal of the SEC was confirmed by PCR and Sanger sequencing (Table S3). Confirmed transgenics were backcrossed once to N2 to create the final strains PX692-PX697 (Table 1, Table S1).

Quantification of synthetic guide RNA efficiency

For each landing pad strain (PX692-PX697), a mixture consisting of 50 ng/µl of all-in-one plasmid targeting the corresponding synthetic guide site and 10 ng/µl pMS81 was microinjected into the gonad of young adult hermaphrodites. Following injection, all worms were maintained at 25□C for the duration of the experiment. Injections were performed until approximately 60 broods per strain had at least one F1 progeny expressing the fluorescent donor homology, thereby marking the brood as successfully injected. Broods were screened for fluorescence at approximately 48h post-injection (hpi), and all fluorescent individuals were counted regardless

of developmental stage (Figure S3). Hygromycin B was then top spread to plates at a final concentration of 250 µg/ml and plates were then screened starting five days later for resistant progeny. Individuals from surviving broods were PCR screened to confirm correct integration.

Removal of the hygromycin selectable marker with Cre

A confirmed homozygous integrant line for *rpl-28p*::*mKate2*::*unc-54 3'UTR* was injected with 10ng/µl of Cre expression plasmid pZCS23 and 10ng/µl pCFJ421 co-injection marker. 30 co-injection marker positive F1 individuals were screened by PCR for the removal of the hygromycin gene. F2 progeny from 3 of the most promising candidates were then rescreened to confirm homozygous removal.

In-situ assembly for integrated transgenes

Two or six PCR fragments with 30bp overlaps, covering the *sqt-1(e1350)* gene, were amplified from pDD285 using Q5 polymerase (NEB) per manufacturer instructions. Homology arms with 30bp overlaps to the *sqt-1(e1350)* gene were similarly amplified from pMS81. These homology arms were then complexed with the adjoining *sqt-1(e1350)* PCR fragment through a second round of PCR.

For *in-situ* assembly and integration, a mixture consisting of 50ng/µl pMS79, 5ng/µl pZCS16, and 40fmol/µl of each of the appropriate gel purified PCR products was microinjected into the gonad of young adult PX696 worms. As a control, 10 ng/µl pZCS52 was substituted for the PCR products. Following injection, all worms were maintained at 25°C for the duration of the experiment. After 24 hours, injected adults were moved to new plates to facilitate counting. F1 individuals were screened for red fluorescence (Figure S4) and the roller phenotype at 3-4 days post-injection. Hygromycin B was then added to plates at a final concentration of 250 µg/ml. Each day for five days post-exposure, plates were scored for hygromycin resistance. Individuals resistant to hygromycin and with the roller phenotype were singled without hygromycin and screened for Mendelian inheritance of the roller phenotype to indicate an integration event. Lines with promising candidates for single copy-integration were singled until they produced homozygous *rol* progeny, which were then screened for the presence of the wrmScarlet co-injection marker, genotyped by PCR across the insert and Sanger sequenced for correct transgene assembly and integration (Table S3).

Accessibility of reagents and protocols

pMS4 (Addgene ID 154837), pMS74 (Addgene ID 154838), PMS79 (Addgene ID 154839), pMS81 (Addgene ID 154840), and pZCS16 (Addgene ID 154824) are available from Addgene (https://www.addgene.org/Patrick_Phillips/). Strain PX696 is available from the *Caenorhabditis* Genetics Center and applicable sequences are available as File S1. Other strains and plasmids are available upon request. Full protocols and all relevant sequences are available on the lab website (github.com/phillips-lab/SLP). All supplementary files associated with this manuscript are available on FigShare.

Results

Generation of synthetic split landing pad sites

Current methods of transgene insertion in *C. elegans* do not specifically select for integration and therefore require additional phenotypic or PCR screening or anti-array selection to avoid selecting heritable extrachromosomal arrays. We sought a faster and simpler method by using CRISPR/Cas9 genome engineering to custom-designed synthetic split landing pads (SLPs). A split uracil selection system has been developed for yeast, whereby a functional *URA3* gene is reconstituted by an integration event. Therefore only transgenic cells survive in the absence of uracil (Levy et al., 2015). We reasoned a similar system could be adapted to *C. elegans*. Hygromycin resistance was chosen as the selectable event as it works across *C. elegans* strains, provides more substantial selection than other antibiotic resistance genes, and does not rely on mutant backgrounds such as with *unc-119* rescue (Radman et al., 2013). An artificial guide site plus the 3' hygromycin coding sequence and transcription terminator were integrated into the target genome site as part of the SLP, whereas the promoter and 5' portion of the hygromycin coding sequence were included in the repair plasmid (Figure 1A). A central 500bp region was included in both fragments, allowing for homology-directed repair. Since a complete resistance gene is not present in either the insertion strain or extrachromosomal arrays containing the repair plasmid, only individuals with proper homology-directed repair have a functional hygromycin resistance gene and survive hygromycin exposure (Figure 1B-D).

The SLP was inserted at Chromosome II:8,420,157. Both this general region and this specific CRISPR site have been shown to be permissive for gene expression, including germline expression (Dickinson et al., 2015; Frøkjær-Jensen et al., 2012, 2008). The SLP was introduced

using the SEC selection method, and removal of the SEC left a loxP site downstream of the hygromycin resistance gene terminator (Figure 1A). By also including a loxP site upstream of the promoter in the repair template, this allows for optional removal of the HygR gene in confirmed integrants by injection of a Cre expressing plasmid (Figure S5).

Efficiency of transgene integration

Given that the SLP is entirely artificial, the guide site can be of the experimenter's choosing. As previous work has shown that the choice of guide site can influence integration efficiency (Farboud and Meyer, 2015), we made six different SLP strains, each differing only in their guide site (Table 3.1). Five of these were sites previously shown to be highly efficient in other model organisms, while the sixth was designed to maximize the predicted guide efficiency. All had very low predicted off-target effects. We developed a *rpl-28p::mKate2* reporter plasmid (pMS81), with homology arms compatible with each of the six SLPs. We then tested integration efficiency into all six SLPs. The plasmid also functions as a co-injection marker and extrachromosomal arrays should result in mKate2 positive, hygromycin sensitive animals. 1.4-9.6% of successfully injected broods (as determined by the presence of at least one mKate2 positive individual at 48 hpi) produced hygromycin resistant individuals. Overall, 79.2% of hygromycin resistant broods (19 of 24) also had perfect integrations events as determined by PCR (Table 1). Imperfect integrations are most often the result of HDR on one side of a double-strand break and incorrect integration on the opposite side, a known issue with transgenic integrations, although although rearrangements within the transgene also occur, likely as a byproduct of the array assembly process (Stinchcomb et al., 1985).

Overall, we found three guide sites to have similar relatively high efficiencies, while one was slightly less efficient and the other two were much less efficient. The observed efficiencies were not always consistent with the predicted guide site efficiencies. For the top three guide sites 7.9-9.6% of co-injection marker positive broods contained integrants (not all of which were correct) which equated to approximately 300-450 co-injection marker positive progeny per integration event. In our hands, injection of thirty total individuals from any of the four best-performing strains was nearly always sufficient to obtain at least one correct line. This is on par with other transgene integration methods directed at this region (Dickinson et al., 2015, 2013; Frøkjær-Jensen et al., 2012, 2008) which have variable integration efficiencies depending on the

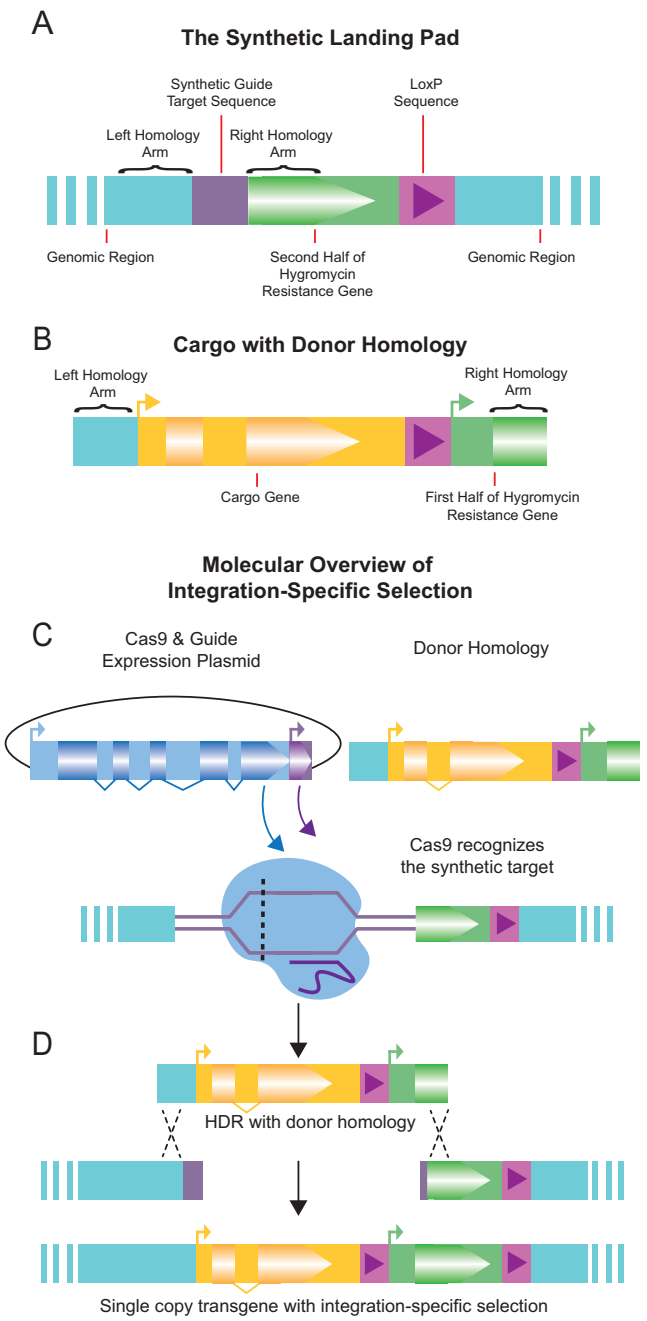


Figure 3.1 Overview of Integration Specific Selection. A) The Synthetic Landing Pad (SLP) with synthetic guide RNA target sequence, the 3′ fragment of the hygromycin resistance gene (partial coding sequence and UTR), and a single loxP sequence. B) The donor homology with cargo transgene to be inserted, a second loxP sequence, and the 5′ fragment of the hygromycin resistance gene (promoter & partial coding sequence). C) Cas9 & guide expression plasmid is injected with donor homology. Cas9 targets and creates a double-strand break at the synthetic target location. D) Once the double-strand break is made, repair with the donor homology integrates the transgenic cargo, and restores the hygromycin gene, allowing selection to occur only upon integration.

transgene, but generally range from 5-30% with a few exceptions reaching higher frequency (Frøkjær-Jensen et al., 2012). As previously observed (Paix et al., 2014), broods with larger numbers of co-injection marker positive progeny (jackpots) were the most likely to yield integrants. However, this number was not predictive of perfect versus imperfect integration (Figure S3.3).

In-situ donor assembly and integration

While plasmids offer the advantage of producing large quantities of the repair template, they require time and labor to produce. Standard cloning practices require a source of DNA, a ligation step, bacterial transformation, plasmid purification, and verification. As we sought to both simplify the process and reduce the overall time-to-integration, we attempted to bypass the cloning step and utilize the *C. elegans* native homology-directed repair to produce a transgene (Figure 2). While clone-free transgenesis has been previously demonstrated in *C. elegans* (Paix et al., 2016; Philip et al., 2019) we wanted to see if this process could synergize with the split-selection system to further improve the process as the previous work did not provide direct selection on the integration event. To test this approach, we utilized the *sqt-1(e1350*) mutation as it gives a dominant roller phenotype allowing us to assay for *in-situ* assembly. Transgene integration was most efficient for the plasmid vector, with 20% of co-injection marker positive broods containing an integrant. However, confirmed *in-situ* assembled and integrated *sqt-1* genes were also obtained using both two and six PCR fragments (Table 2, PCR Confirmed Integrations). Two parts correctly assembled and integrated more often than six parts. In some cases, hygromycin resistant individuals were observed without the *sqt-1* roller phenotype (Table 2, Hygromycin Resistant Broods). We believe these represent incorrect integration events, where at least the 5' hygromycin resistance coding fragment was integrated into the genome. As these cannot be correct integration and assembly events, we did not pursue or characterize them.

Table 3.2 In-situ assembly & integration efficiency.

		Hygromycin Resistant Broods			***In-situ* Assembly**		
	Marker Positive Broods	**All Roller**	**Mixed**	**All wt**	**Homozygous Roller Isolated**	**PCR Confirmed Integrations**	**Error Free Integrations**
Plasmid	15	3	0	0	3	3	3 (20.0%)
2pc PCR	41	6	3	0	8	6	3 (7.3%)
6pc PCR	51	0	8	5	1	1	0 (0.0%)

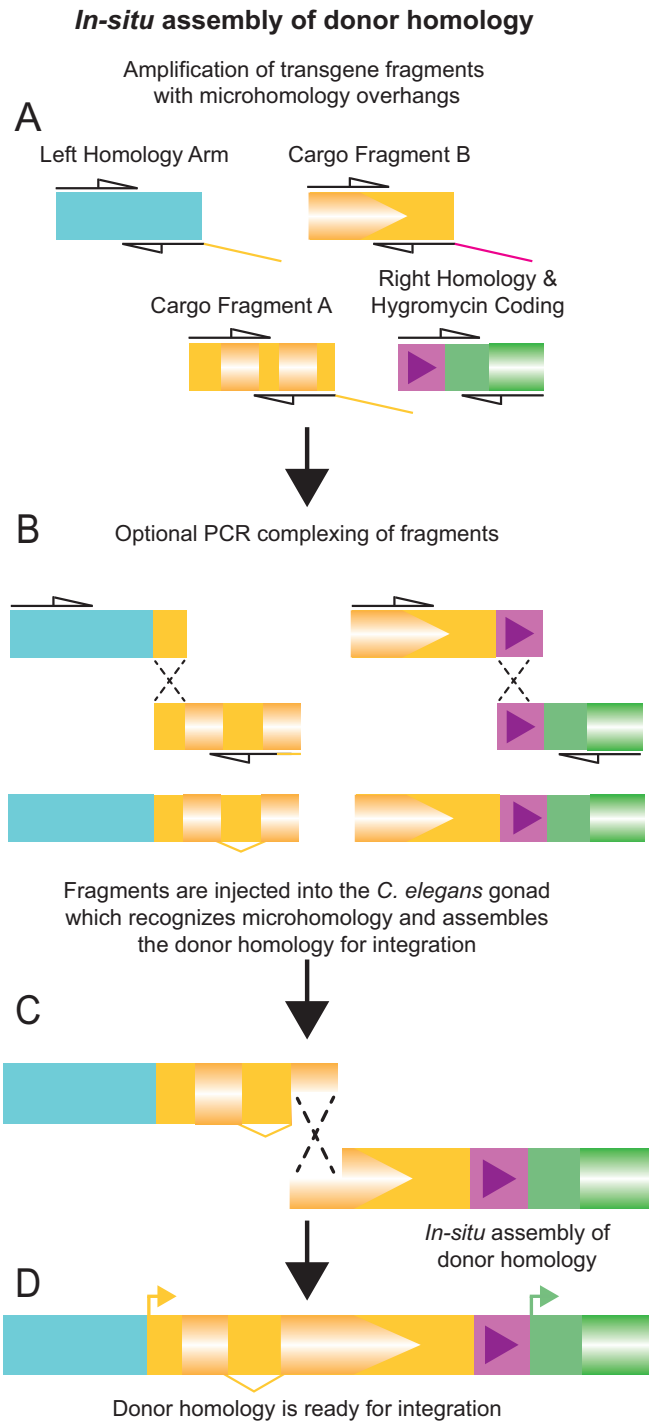


Figure 3.1 Overview of *in-situ* assembly. A) Amplification of homology arms and cargo fragments by PCR with overlaps of ~30bp B) Optional complexing by a second round of PCR reduces the number of fragments and increases the frequency of correct integration. C) Upon microinjection, PCR products are recombined by the worm using microhomology to make D) the complete donor homology ready for integration.

For two-part assemblies, most hygromycin resistance events were accompanied by *sqt-1* assembly and integration, as indicated by the ability to isolate homozygous roller populations. However, not all of these insertions matched the expected sequence. Two of the eight insertions could not be amplified by PCR, suggesting larger scale errors, while three had point mutations identified during Sanger sequencing, and three had no detectable errors. In no case did we detect multiple copy insertions. In the six-part experiment, all resistant plates had non-roller (incorrect) integration events, with a few having roller individuals as well (Table 2, Hygromycin Resistant Broods). In most cases, a homozygous roller line could not be isolated, suggesting these individuals were the result of correctly assembled genes in arrays paired with incorrect integrations. In one case, a homozygous roller line could be isolated, indicating multiple integration events had taken place in that brood. In this case, the roller causing integration had a correctly assembled *sqt-1(e1350*) gene but also contained a second copy of one of the homology arms which was identified by Sanger sequencing.

The inclusion of a fluorescent co-injection marker allowed for monitoring of array loss in this experiment. Since there is no selection on the transgene containing plasmid, any arrays that form should be rapidly lost. As expected, prior to the addition of hygromycin, array positive individuals could be seen. However, none of the homozygous individuals isolated for sequencing (approximately 3-5 generations after injection) contained arrays, demonstrating arrays are indeed quickly lost in this system. Even so, it remains best practice to confirm array loss through either use of an array co-injection marker or a vector specific PCR performed in conjunction with the genotyping PCR.

Discussion

We have created a fast and efficient strategy for integrating transgenes into the *C. elegans* genome that bypasses some pitfalls and laborious steps present in other methods. Combining split selection with self-assembly of repair templates takes what before would require at best two to three weeks down to as little as a week, while simultaneously reducing the required expertise and overall hands-on time (Figure 3.3). The core technology relies on integration-specific selection, made possible by SLPs at a safe harbor insertion site. These SLPs can be inserted into any *C. elegans* strain using a single set of reagents since the protocol presented does not rely on a particular genetic background. The landing pad presented is universal, and the experimenter can

choose any type of cargo to be integrated utilizing either plasmids or our CIS approach. However, the SLP design could be modified with additional elements to facilitate specific types of insertions, such as reporter constructs or allelic variants. While we see no obvious phenotypic effects from the constitutive expression of hygromycin resistance protein, the SLP includes loxP sites, allowing for a second injection of a Cre expression plasmid to remove the hygromycin gene. While removal by Cre injection is relatively efficient (Figure S3.5), if routine removal of hygromycin is desired, incorporation of inducible Cre and a marker gene into either the landing pad or insertion vector would allow for self-excision using a protocol similar to SEC (Dickinson et al., 2015). Further, additional SLPs using either the same or different guide and selective gene could be inserted into other sites in the *C. elegans* genome known to be permissive for transgene expression. This would facilitate the construction of more complex, multigene transgenic nematodes. The antibiotics neomycin (Giordano-Santini et al., 2010), puromycin (Semple et al., 2010) blasticidin (Kim et al., 2014) and nourseothricin (Obinata et al., 2018) have all been used for selection in *C. elegans,* and the coding sequences of the corresponding resistance genes could be split. When using multiple SLPs within the same genetic background, lox site variants (eg. lox511I, lox2272, loxN) should be considered to prevent intrachromosomal recombination. The SLP, conceptually adopted from yeast, does not need to be restricted to *C. elegans*. While the formation of heritable arrays is unique to *Caenorhabditis* nematodes among model systems, and thus does not complicate single-copy integration transgenesis in other models, the concept of custom-designed SLPs could provide direct readouts for integrations events, with specific, custom-built guide target sequences. For example, a split fluorescent coding sequencing could suffice to screen injected embryos for proper, site-specific integration in model vertebrates and be coupled with a non-native, experimentally chosen guide RNA to reduce off-target effects while increasing on-target cutting and HDR.

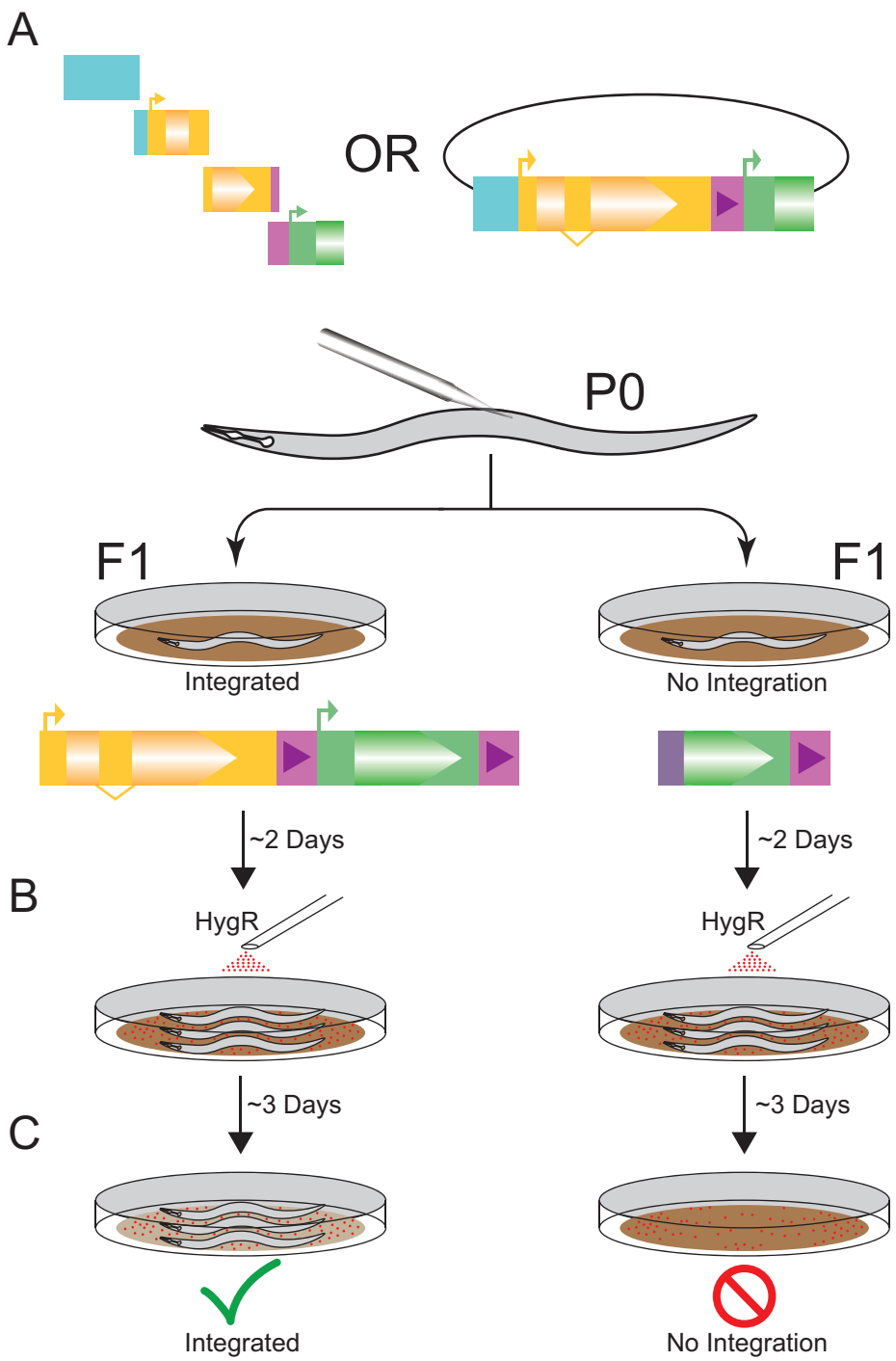


Figure 3.2 Experimental Overview. A) Injections of PCR fragment or cloned donor homology with Cas9 and guide expression plasmid. With PCR products, *in-situ* assembly forms the cargo transgene. B) Two days after injection, worms are exposed to hygromycin B. Since the array does not provide selection C) only integrated worms survive the exposure, providing the integration-specific selection.

The ability of *C. elegans* to self-assemble exogenous DNA fragments based on microhomology represents a possible alternative to plasmid cloning for insert assembly. Individuals with an assembled and integrated transgene were seen using both two or six PCR products. However, six pieces resulted in fewer correct integration and more incorrect integrations (Table 2, Hygromycin Resistant Broods), suggesting that, as expected, proper assembly occurs more often with fewer PCR products. As such, it is likely desirable to complex PCR products by overlap PCR before integration where feasible. While the use of PCR products, rather than plasmids, represents a more rapid protocol with fewer technical steps, it comes with the trade-off of a lower frequency of correct integrations requiring injection and screening of a larger number of worms. Thus, while use of PCR products results in a shorter time to integration confirmation, the total amount of hands on time is similar between the two protocols. Ultimately, at this time, choice of PCR products versus plasmid will largely come down to lab preference, although certain protocols, such as insertion of a library of similar constructs may favor the PCR approach. Importantly, SLP based integration-specific selection is fully compatible with both cloned plasmid donors and the CIS approach.

Site-specific transgene insertions rely on homology-directed repair (HDR). However, non-homologous end joining is almost always the prevalent pathway in repairing a double-stranded break (Ward, 2015; Xu, 2015). During guide efficiency testing, the top three guides all had similar insertion frequencies, suggesting we are approaching the upper limit of cutting efficiency and that further improvements will require improved rates of HDR. Furthermore, improved HDR should assist in the assembly of PCR products in worms and reduce the rate of false positives due to incorrect assembly. Low rates of HDR are not specific to *C. elegans* and impair HDR-based insertion in many model systems. As a result, multiple HDR enhancement strategies have been proposed in multiple model organisms (Beumer et al., 2008; Böttcher et al., 2014; Ward, 2015). Adaptation and advancement of one or more of these methods will likely represent the next breakthrough in genome editing efficiency in *C. elegans*.

Bridge

In the next chapter, I present TARDIS, our technological advancement to take integration-specific selection to a library-scale, unlocking experimental paradigms such as barcoded lineage tracking for experimental evolution.

CHAPTER 4 HIGH-THROUGHPUT LIBRARY TRANSGENESIS IN *CAENORHABDITIS ELEGANS* VIA TRANSGENIC ARRAYS RESULTING IN DIVERSITY OF INTEGRATED SEQUENCES (TARDIS)

This chapter was published in volume 12 of the journal eLife in 2023.

Megan J. Moerdyk-Schauwecker, Stephen A. Banse, Dhaval S. Patel, Hang Lu, and Patrick C. Phillips are co-authors on this publication. Patrick C. Phillips, Stephen A. Banse, and I designed the technology. Megan J. Moerdyk-Schauwecker, Dhaval S. Patel, Hang Lu, and I designed the various strains. Megan J. Moerdyk-Schauwecker, and I analyzed the data. Patrick C. Phillips, Stephen A. Banse, Megan J. Moerdyk-Schauwecker, and I wrote the manuscript.

The full supplemental material for this publication can be found at:

https://doi.org/10.6084/m9.figshare.c.6264162

INTRODUCTION

Transgenesis, which is the specific and heritable introduction of foreign DNA into genomes, has been a central tool for functional analysis and genetic engineering for nearly 40 years. The power of transgenesis is due in part to the wide variety of assays and techniques that are built upon controlled introduction of novel DNA sequences into a native genome. While there are many uses for transgenesis, in practice most can be grouped into those inserting a small number of known sequences (specific transgenesis) and those introducing many sequence variants from experimental libraries (exploratory transgenesis). While the ability to perform specific transgenesis has become a de facto requirement for all model organisms, exploratory transgenesis remains effectively limited to single-cell models (both prokaryotic and eukaryotic) caused by biological limitations generated by inheritance in multicellular organisms. In single-cell models, high-throughput transgenesis has been used for exploratory sampling of sequence space using protein interaction libraries (Joung et al., 2000), barcode-lineage tracking libraries (Ba et al., 2019; Levy et al., 2015), directed evolution (Packer and Liu, 2015), synthetic promoter

library screens (Wu et al., 2019), and mutagenesis screens (Bock et al., 2022; Erwood et al., 2022; Kim et al., 2022; Sánchez-Rivera et al., 2022). Despite the usefulness of such experiments in single-celled systems, either in microorganisms or in cell culture, increasing transgenic throughput in multicellular models holds the potential to expand the impact of exploratory transgenesis in functional domains, such as inter-tissue signaling, neuronal health, and animal behavior, that are dependent on multicellular interactions and therefore difficult to replicate in single-cell models.

Exploratory transgenesis in single-cell models has been facilitated by the availability of *in-vitro*-generated DNA libraries, selectable markers, plasmids, *in vivo* homologous recombination, and most importantly, the ability to massively parallelize transgenesis using microbial transformation or eukaryotic cell transfection/transduction. Currently, there is no practical means to make populations of uniquely transgenic individuals from sequence libraries at a similar scale in animal systems due to the Weismann Barrier (Weismann, 1893): the split between soma and germline. The requirement that the germline be accessible and editable has forced animal systems into a transgenic bottleneck compared to single-cell systems because it is very difficult to introduce exogenous DNA directly into the germline in a high-throughput manner, relying instead on injection, bombardment, or some other physical intervention. This low-throughput limitation in animals dramatically reduces the sequence diversity that can be sampled, effectively preventing large-scale exploratory experiments from being performed. Attempts have been made to parallelize transgenic creation in multicellular model organisms, for example, the development of Brainbow (Livet et al., 2007; Weissman and Pan, 2015), ifgMosaic analysis (Pontes-Quero et al., 2017), P[acman] libraries in *Drosophila* (Venken et al., 2009), and multiple types of transformation in plants (Ismagul et al., 2018; Xu et al., 2022). In *Caenorhabditis elegans*, CRISPR technology combined with custom engineered sites within the genome ("landing pads") has facilitated the generation of single-copy integrations (Malaiwong et al., 2023; Nonet, 2021, 2020; Silva-García et al., 2019; Stevenson et al., 2020; Vicencio et al., 2019), and attempts have been made to multiplex transgenesis using traditional integration methods in conjunction with specialized landing pad (Gilleland et al., 2015; Kaymak et al., 2016; Mouridi et al., 2022; Radman et al., 2013). While these efforts have increased throughput over standard single copy integration methods, throughput still remains too low for effective

exploratory transgenesis, and in some cases requires significant additional labor, cost, equipment and/or expertise.

Here we present "Transgenic Arrays Resulting in Diversity of Integrated Sequences" (TARDIS) (Stevenson et al., 2021), a simple yet powerful alternative to traditional single-copy transgenesis. TARDIS greatly expands throughput by explicitly separating and reordering of the conceptual steps of transgenesis (Figure 1). To increase throughput, TARDIS begins with an *in vitro* generated DNA sequence library that is introduced into germ cells via traditional low throughput methods (i.e., germline transformation, Fig 1). While traditional transgenesis typically couples the physical introduction of DNA into cells with the integration of a selected sequence from the original library, the DNA sequences in TARDIS are designed to be incorporated in large numbers into diverse, heritable sub-libraries (TARDIS libraries), rather than be directly integrated into the desired genomic locus. In addition to the sequence library, a functioning selectable marker is also included to stabilize the inheritance of the TARDIS library over generations. These TARDIS libraries function to create "metaploidy" – expanding the total number of alleles available for inheritance, essentially making the worm genetically "bigger on the inside." TARDIS library-bearing animals are then allowed to propagate under selection to generate a large population of TARDIS library carriers. After population expansion, genome integration of a single sequence unit is performed by inducing a double strand break at a genetically engineered landing pad. This landing pad is designed to both integrate a sequence unit and act as a second selectable marker. We chose *C. elegans* to validate the TARDIS approach because *C. elegans* naturally form extrachromosomal arrays that can be several megabases in size (Carlton et al., 2022; Lin et al., 2021; Mello et al., 1991; Stinchcomb et al., 1985) from injected DNA, which simplifies the generation of heritable "TARDIS library arrays" (TLA) that encompass sequence significant diversity.

We demonstrate the functionality of TARDIS for two use cases: unique animal barcoding and promoter library transgenesis. Barcoding has been widely adopted in microbial systems for evolutionary lineage tracking (Ba et al., 2019; Jahn et al., 2018; Levy et al., 2015) and for developmental lineage tracking in animals (Kebschull and Zador, 2018; McKenna et al., 2016). In microbial systems, barcode libraries have relied on highly diverse randomized oligo libraries, as compared to animal systems, which have relied on CRE recombinases or randomized Cas9-induced mutations. Here we present a novel TARDIS barcoding system for an animal model

which mimics the scope and diversity previously only possible using microbial systems. Our results show that large, heritable libraries containing thousands of barcodes can be created and maintained as extrachromosomal arrays. Individual sequences are selected and removed from the library upon experimental induction of Cas9 in a proportion consistent with the composition of the TLA with rare overrepresented sequences. We found that TARDIS is also compatible with the integration of large promoters and can be used to simultaneously integrate promoters into multiple genomic locations, providing a tool for multiple insertions at defined locations across the genome. While we demonstrate the system's advantages in *C. elegans*, in principle, the system is adaptable for any situation where the sequences for integration can be introduced with high diversity and heritability, and where a genomic site for integration can be made or is available.

Results

Generation of barcode landing pad

We designed a specific landing pad for the introduction and selection of small barcode fragments from high-diversity, multiplexed barcode libraries (Figure 2). This landing pad was designed to be targeted by Cas9 and requires perfect integration on both the 5' and 3' ends of a synthetic intron for functional hygromycin B resistance. Current split selection landing pads only provide selection on one side of the double-strand break, which can result in a small percentage of incomplete integrations (Stevenson et al., 2020). To fully test a large library approach, the requirement of genotyping to identify correct integrations must be overcome. A split-selection, hygromycin resistance (HygR) system was chosen for its simplicity and integration-specific selection. A unique synthetic CRISPR guide RNA target sequence was created by removing coding sequence on both sides of an artificial intron, resulting in a non-functional HygR gene. By removing critical coding sequence on both sides of the gene, only 'perfect' integration events will result in hygromycin resistance (Figure 2A). The synthetic landing pad was integrated at Chromosome II: 8,420,157, which has previously been shown to be permissive for germline expression (Dickinson et al., 2015; Frøkjær-Jensen et al., 2012, 2008).

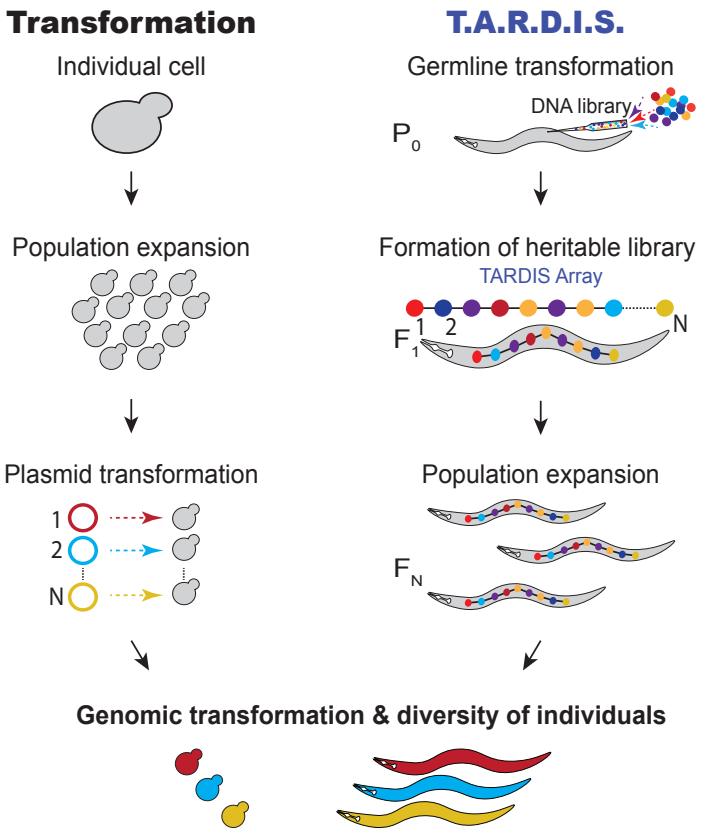


Figure 4.1 Transformation compared to TARDIS. For transformation, a large population of cells are individually transformed with a DNA library, resulting in a diverse population of individuals. TARDIS achieves a diversity of individuals by splitting transgenesis into two separate processes: (1) The introduction of a diverse library, which is formed into a TARDIS library array, passed down to future generations and thus replicated; and (2) an event that triggers the integration a sequence from the library at random, resulting in a diversity of integrated sequences.

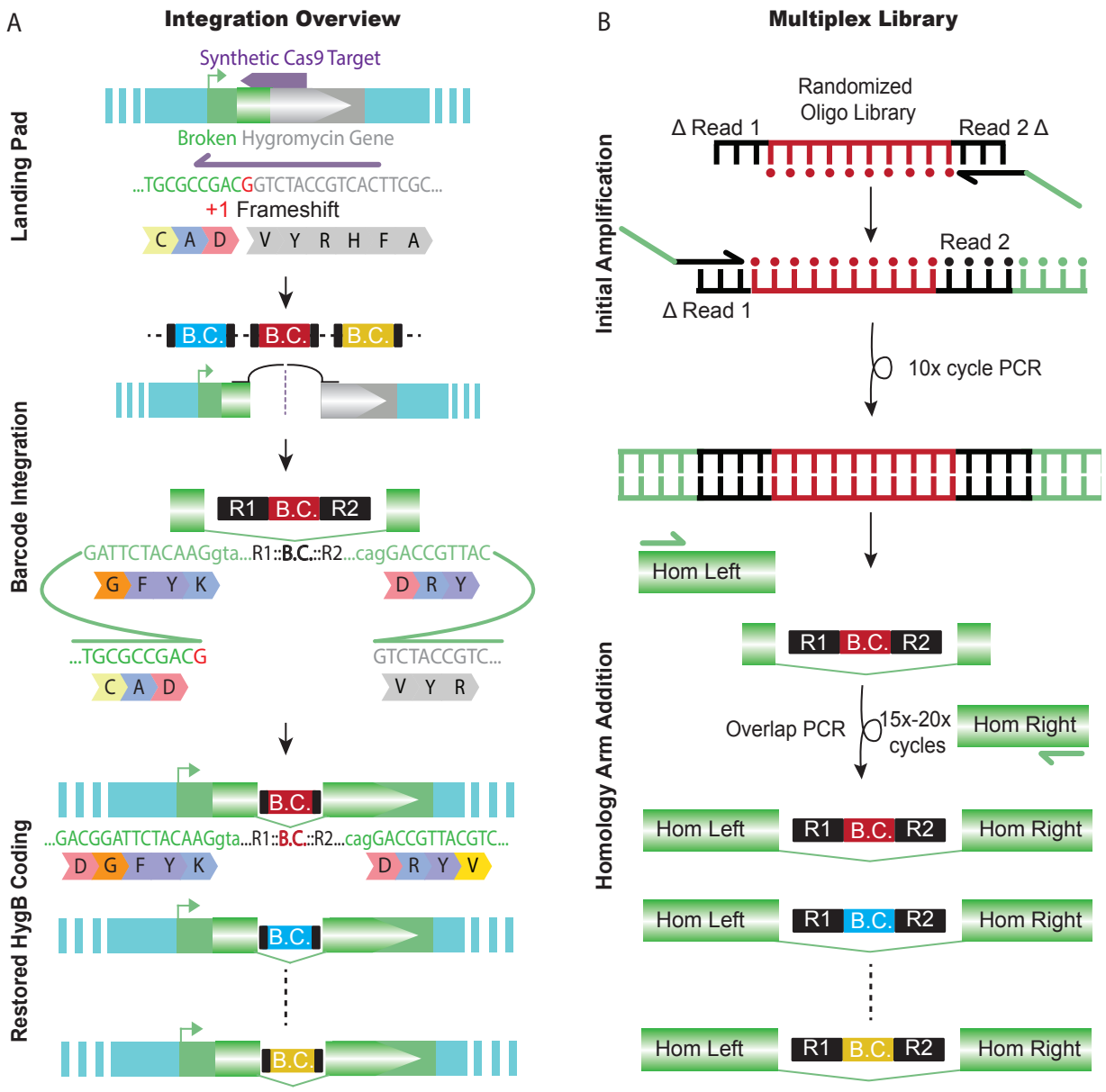


Figure 4.2 Barcode landing pad and diverse donor library. A) Schematic design for the barcode landing pad and integration. A broken hygromycin resistance gene is targeted by Cas9, which repairs off the TARDIS array, integrating a barcode and restoring the functionality of the gene. B) The TARDIS multiplex library was created from a randomized oligo library, which underwent 10 cycles of PCR to make a dsDNA template. The barcode fragment was then added into a three-fragment overlap PCR to add homology arms and make the final library for injection.

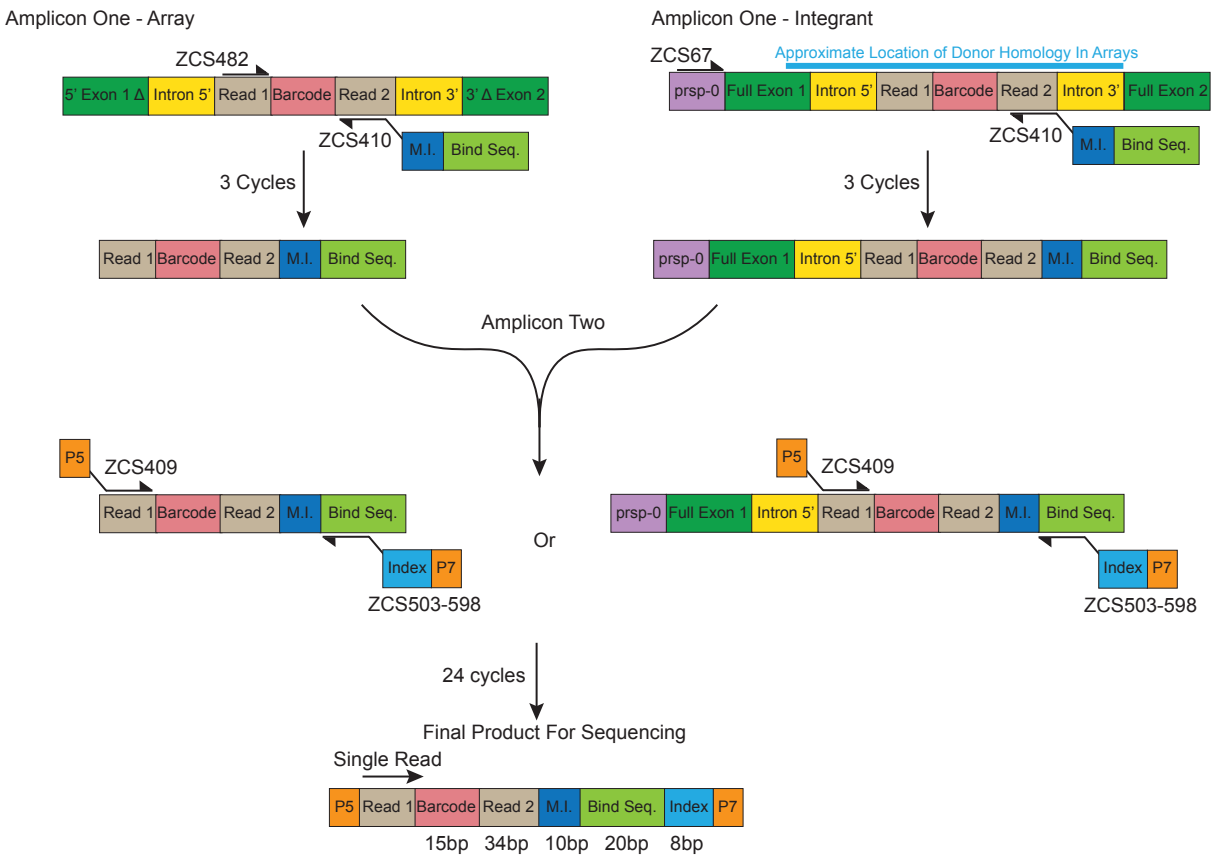


Figure 4.3 —figure 4.2 supplement 1: Schematic layout for the two separate PCR processes for identification of barcode counts in arrays (Amplicon One-Array) and integrants (Amplicon One-Integrant).

Generation of high diversity donor library and TARDIS arrays

Transgenes or DNA sequences can be cloned into plasmid vectors for injections in *C. elegans*. However, the cloning process is laborious, and the plasmid vector is unnecessary for integration into an array or the genome. We sought to provide a protocol for library generation that maximized diversity and eliminated the requirement of cloning (Figure 2B). Oligo libraries have been used for barcoding (Levy et al., 2015) and for identification of promoter elements (Boer et al., 2020) in yeast, but practical implementation of large synthetic libraries for transgenesis has never been performed in an animal system. We used randomized synthesized oligos to build a highly diverse library of barcodes, similar to the one described by Levy et. al. (2015), via complexing PCR. Given randomized bases present at the 11 nucleotide positions centrally located within the barcode, our base library can yield a theoretical maximum of approximately 4.2 million sequences. Our overlap PCR approach achieves high levels of diversity with minimal 'jackpotting' — sequences with higher representation than expected (Figure 3—figure supplement 1). With low coverage sequencing, we found almost 800,000 unique barcode sequences, providing a large pool of potential sequences that can be incorporated into TARDIS arrays. Only 472 sequences were overrepresented (counts greater than 50), accounting for approximately 6.7% of the total reads and only approximately 0.06% of the unique barcodes detected.

We injected our complexed barcodes and isolated individual TARDIS array lines, each containing a subset of the barcode library (Figure 3). Individual injected worms were singled, and we identified four arrays from three plates. Arrays 1 and 2 were identified on separate plates, and were therefore derived from independent array formation events, while array 3, profile 1 and array 3, profile 2 were both identified on the same plate. Analysis of array diversity within these lines shows, somewhat unexpectedly, that during array formation a subset of barcode sequences tended to increase in frequency (Figure 3A and B). Higher frequency barcodes in arrays tend to be independent of the jackpotted sequences of the injection mix as very few are represented in the set of high frequency barcodes from the injection mix. The high frequency barcodes also varied between arrays.

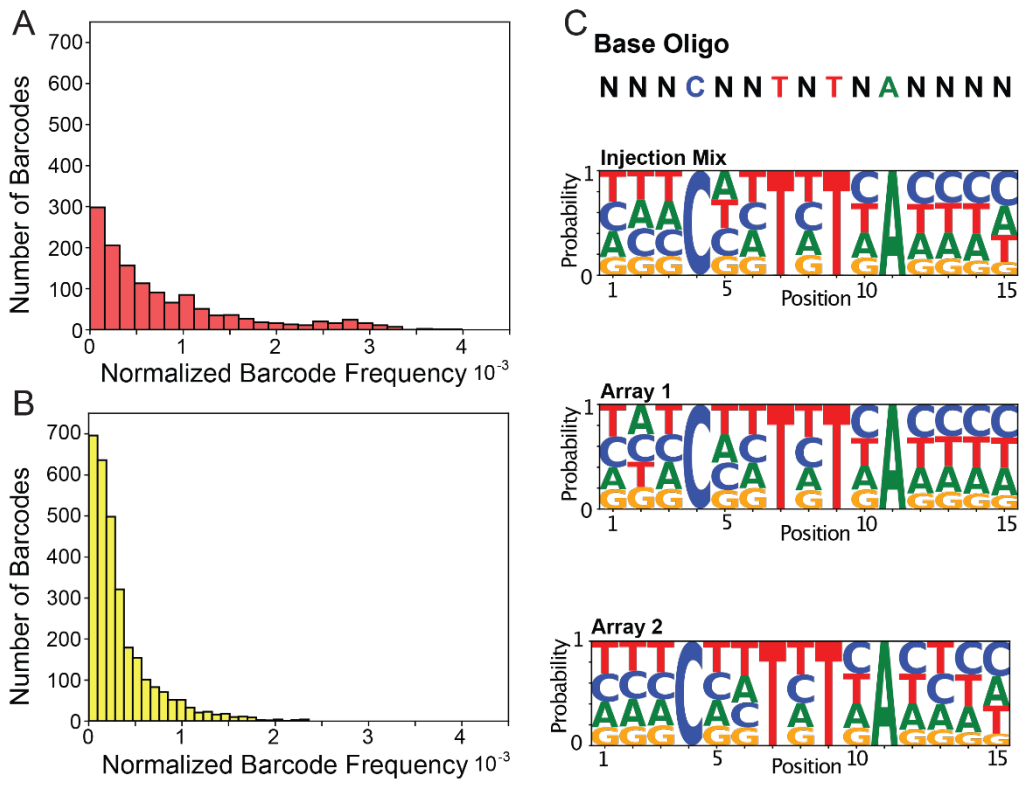


Figure 4.4 TARDIS Library Arrays can contain large barcode diversity. A) Frequency distribution of 1,319 unique barcodes in array 1 (PX816). B) Frequency distribution of the 3,001 unique barcode sequences in array 2 (PX817). C) Sequence logo probabilities of the 15 base pair positions of the barcodes in the injection mix, array 1 and array 2.

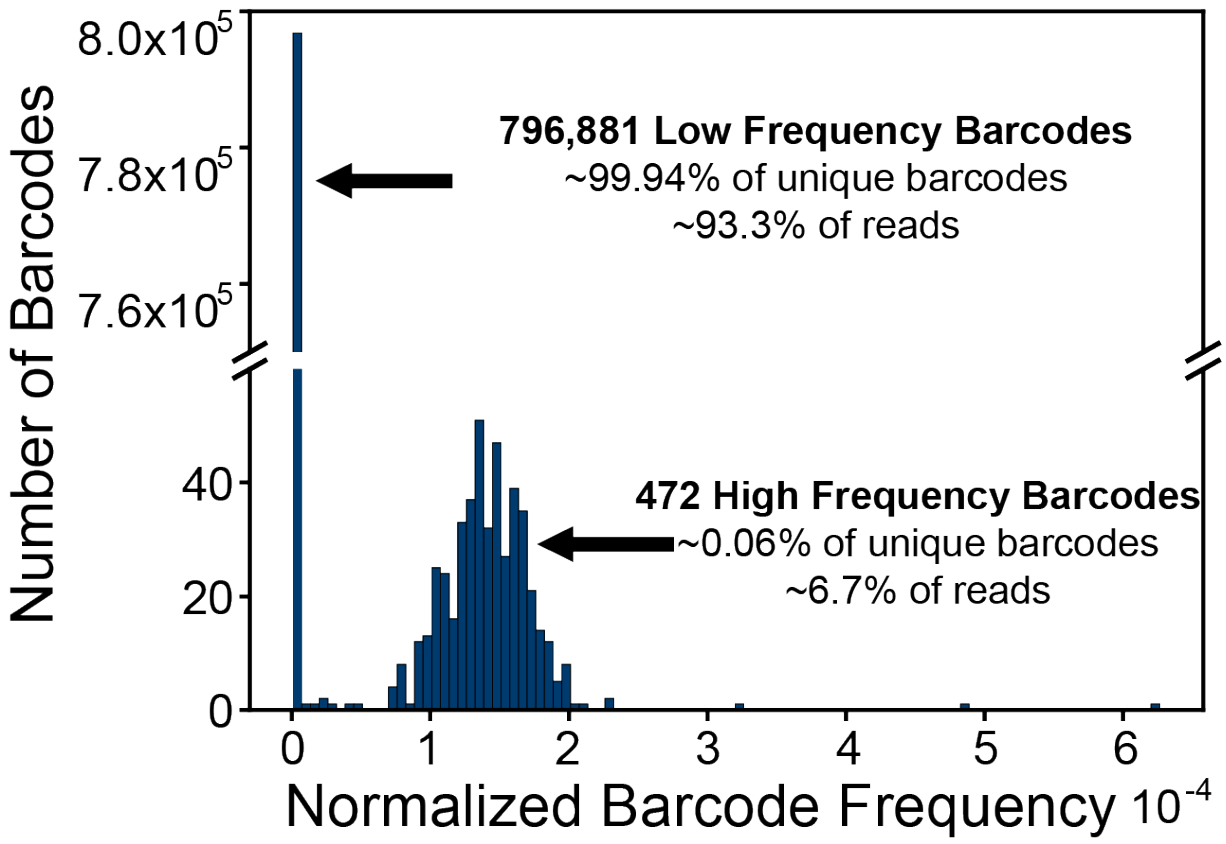


Figure 4.5—figure 4.4 supplement 1: A) Barcode frequencies for Injection mix used for arrays 1, 2, and 3. There are approximately 1 million reads represented. In total, 797353 unique sequences were identified. A few of these unique sequences were represented at a higher frequency with a count cutoff greater than 50.

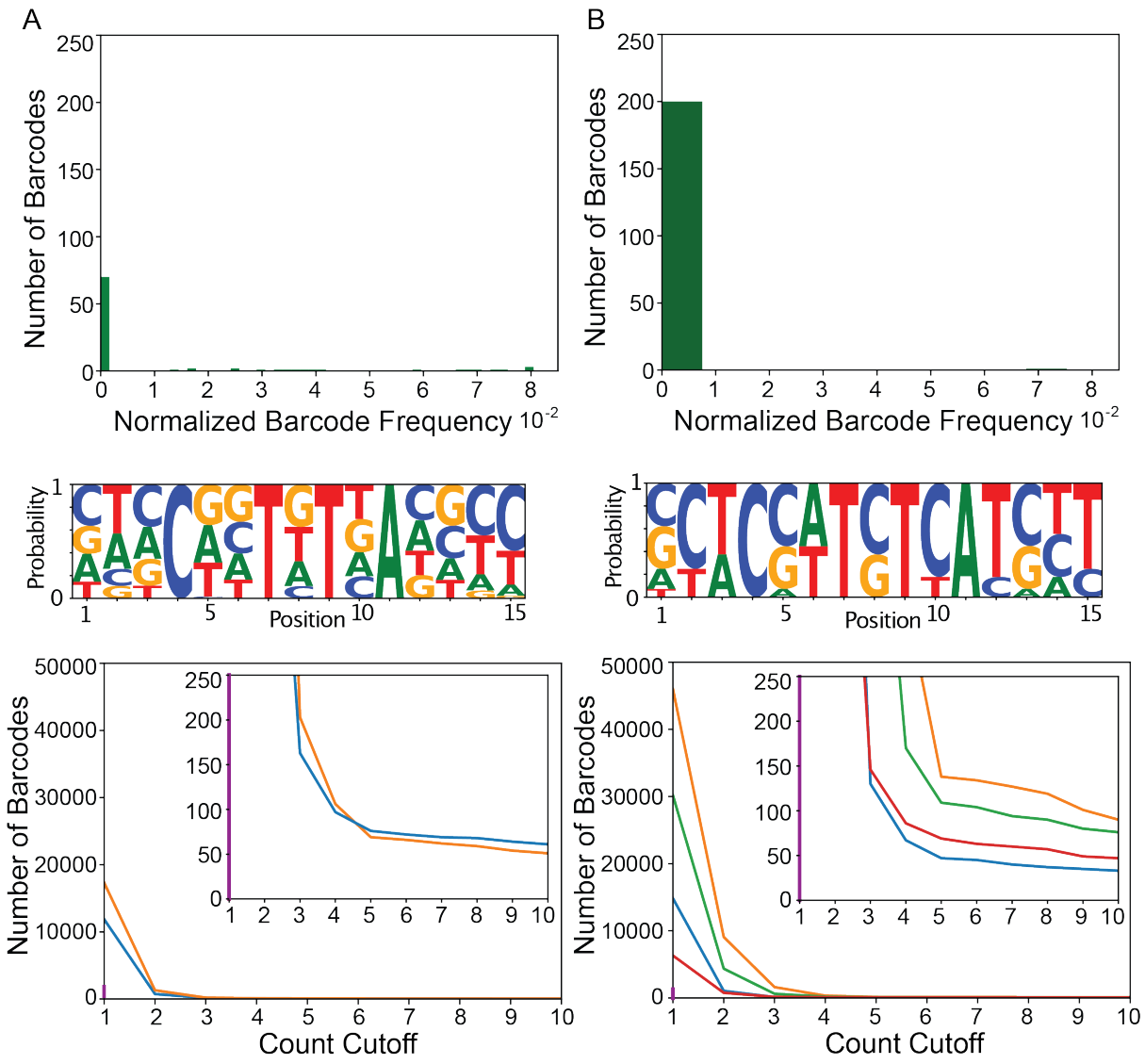


Figure 4.6—figure 4.4 supplement 2: TARDIS Array 3. Two individual arrays were isolated from the same plate. Both show considerably less diversity than TARDIS arrays 1 and 2. Distribution of unique barcode frequencies, sequence logo base pair probability, and count cut off for A) TARDIS Array 3 profile 1 and B) TARDIS Array 3 profile 2.

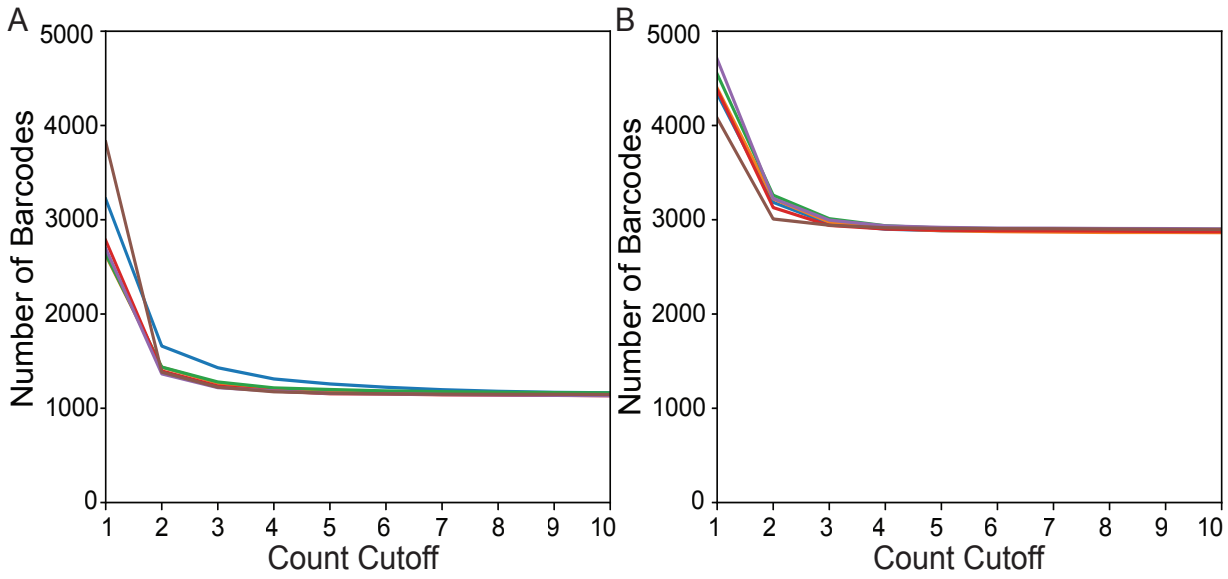


Figure 4.7 figure 4.4 supplement 3: Determination of proper count cutoff for A) TARDIS Array 1 and B) TARDIS Array 2.

We found that array formation does not seem to favor any particular barcode sequence motif (Figure 3C) and that arrays can range considerably in diversity. Array 1 had 1,319 unique barcode sequences, array 2 had 3,001 unique barcode sequences, array 3 profile 1 had 91 unique barcode sequences, and array 3 profile 2 had 204 unique barcode sequences (Figure 3—figure supplement 2). Across the four arrays, we found a total of 4,395 unique barcode sequences. When we compared the individual sequences incorporated during the three independent injections, we found little overlap. 96.5% (4395/4553) of the identified sequences were unique to one injection, 3.0% (136) were incorporated twice, and 0.5% (22) were recovered from all three injections. In contrast to the diversity between injection events, a similar comparison of the two profiles derived from a single injection for array 3 showed considerable overlap, with 68% (62/91) of the profile 1 sequences also being present in profile 2. Overall, our results suggest our complexing PCR oligo library can produce a highly diverse library and that arrays can store a large diversity of unique sequences.

The distribution of element frequency within a given array follows a clear Poisson distribution. Arrays 1 and 2 show more diversity, with barcode frequencies more similar to one another than the two profiles isolated from array 3 (Figure 3—figure supplement 2). The null assumption is that the array is formed from a simple sample of the injected barcodes in equal proportions. However, arrays have been already reported to jackpot certain sequences. For example, when Lin et. al. (2021) injected fragmented DNA, they found that larger fragments were favored in the assembly. In our case, we find some barcode sequences become jackpotted, despite being identical in size. A possible explanation is that early in formation, arrays are replicating sequences, possibly to reach a size threshold. Consistent with this hypothesis, arrays with higher barcode diversity had frequencies closer to one another, while arrays with lower diversity had wider frequency ranges.

Integration from TARDIS array to F1

Our primary motivation in developing the TARDIS method was to utilize individual sequences from the TARDIS array as integrated barcodes. To assay the integration efficiency, we performed TARDIS integration on two biological replicates from a TARDIS array line (PX786) synchronized in the presence of G-418. Out of the 100 L1's per plate initially plated on antibiotic free plates, an average of 41 worms (N=255 plates) for replicate one, and 62 worms for replicate

two (*N*=125 plates), survived to the next day. These surviving individuals contained the array, allowing them to survive early life G-418 exposure, and generally showed fluorescent co-marker expression as well. Following heat shock to induce Cas9, replicate one produced 104 plates with hygromycin resistant individuals, indicating barcode integration, and replicate two produced 71. These results suggest that approximately 200-300 worms need to be heat shocked to obtain an integrated line when using 150bp homology arms and relatively small inserts such as the barcodes. To assay the integration frequency from the array to the F1, we performed TARDIS integration on four biological replicates derived from PX786. We found that frequency of integration for barcodes in F1 individuals was strongly correlated with the barcodes' frequency in the TLA (Figure 4A), ($R \approx 0.96$, $p \approx 5.7\text{x}10^{-154}$). Notably, there are two replicated outliers across the four biological replicates. One barcode (TTAAATTATCACATG), tended to integrate more often than would be predicted by its frequency in the array, while barcode (GCTCATTCTGACGTA) integrated less frequently than expected (Figure 4—figure supplement 1). In general, however, we did not observe any noticeable bias in sequence motif selection following integration (Figure 4B). Several individual lineages were isolated from the population with hygromycin selection, validating functional restoration of the *HygR* gene, and three were randomly chosen for Sanger sequencing to confirm perfect barcode integration. As expected, these sequenced barcodes were also found amongst the barcode sequences of the array.

Generation and integration of TARDIS promoter library

For testing insertion of promoter libraries via TARDIS, two separate landing pad sites utilizing split selection were engineered in chromosome II (Figure 5A). The first contained the 3' portions of both the *mScarlet-I* and the *HygR* genes in opposite orientation to each other and separated by a previously validated synthetic Cas9 target (Stevenson et al., 2020). Similarly, the second landing pad site contained 3' portions of *mNeonGreen* and *Cbr-unc-119(+)* separated by the same synthetic Cas9 target, allowing both sites to be targeted by the same guide. These landing pads were engineered into an *unc-119(ed3)* background to allow for selection via rescue of the Uncoordinated (Unc) phenotype. A strain containing only the split *mScarlet-I*/split *HygR* landing pad was also constructed, in which case a copy of *Cbr-unc-119(+)* was retained at the landing pad site. Repair templates contained the 5' portion of the respective selective gene, a lox site allowing for optional removal of the selective gene after integration (by expression of Cre) and the chosen promotors in front of the 5' portion of the respective fluorophore. The selective gene

and fluorophore fragments contained >500bp overlaps with the landing pad to facilitate homology directed repair. Correct homology directed repair at both junctions resulted in worms that were fluorescent, hygromycin resistant and had wild type movement.

The initial promoter library tested was composed of 13 promoters targeted to a single landing pad site with split *mScarlet-I* and split *HygR* (Table 1). These promoters ranged in size from 330-5545bp (total repair template length of 2238-7453bp). Seven different array lines were generated which exhibited distinct profiles when probed by PCR as a crude measure of array composition and diversity (Figure 5—figure supplement 1A). Promoter specific PCR showed these arrays to contain 2-13 of the 13 injected promoters, with a mean of 10.7 and a median of 12 (Figure 5—figure supplement 1B). For the selected line (PX819), 12 promoters were incorporated into the TARDIS array. From this line, approximately 200 G-418 resistant L1s (ie. those containing the array) were plated onto each of 60 plates and then heat shocked as L2/L3s to initiate integration. Hygromycin resistant individuals were recovered from 59 of the 60 plates, indicating one or more integration events on each of those plates. Four individuals were singled from each of these plates, with the intent of maximizing the diversity of fluorescent profiles and analyzed by PCR to identify the integrated promoters (Figure 5—figure supplement 1B). Based on the banding patterns, 83 of these PCR products were sequenced with nine different promoters confirmed as integrated (Table 1 and Figure 5B). This included both the smallest (*aha-1p*) and the largest promoter (*nhr-67p*) in the set. Notably, two of the three promoters that were in the array but not recovered as integrants were found to be integrated in a subsequent experiment (see below), suggesting the failure to be recovered in this case was likely due to the array composition rather than any properties of these particular promoters. For approximately half of the plates, two or more promoters were identified from the four worms chosen. Of the 83 PCR products sequenced, five had incorrect sequences and/or product sizes inconsistent with the promoter identified and three failed to prime. Additionally, several samples failed to amplify or gave a non-specific banding pattern and likely also represent incorrect integrations.

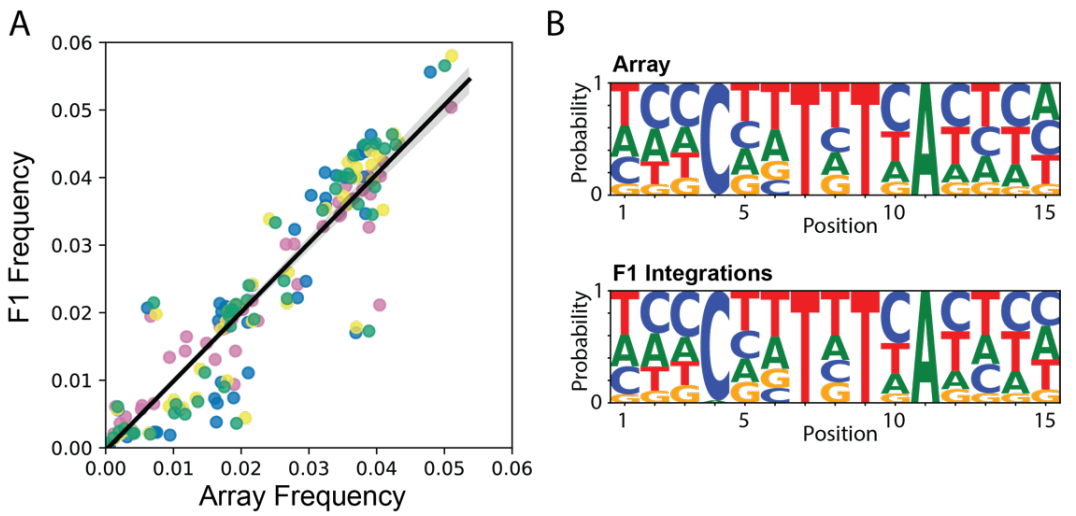


Figure 4.8 Integration frequency from TARDIS Library Array to F1. A) Frequency of integration from TARDIS Library Array to the F1, R ≈ 0.96, $p \approx 5.7 \times 10^{-154}$. Different colors represent four biological replicates. Line shading represents 95% confidence interval. B) Sequence probabilities of PX786 compared to the F1 integrations (91 unique barcodes were identified in the array and 118 in the F1s, with a five read threshold).

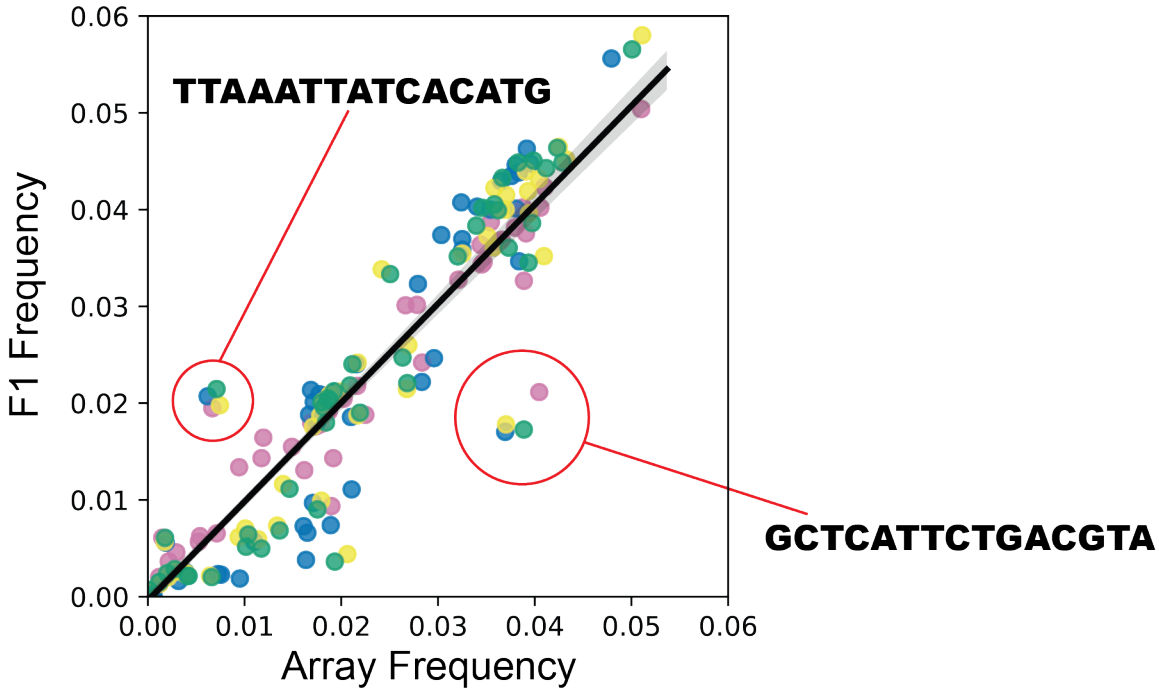


Figure 4.9—figure 4.8 supplement 1: F1 integration events followed a consistent pattern, with replicated outlier barcode sequences. Generally, the same barcodes integrated at approximately the same frequency across the four replicates.

To test if TARDIS could be used to target multiple sites simultaneously, a second promoter library containing seven promoters targeted to each site *(ahr-1p*, *ceh-10p*, *ceh-20p*, *ceh-40p*, *ceh-43p*, *hlh-16p*, *mdl-1p*) was injected into worms containing both landing pad sites. Five plates of mixed stage worms were heat shocked, and worms that were both hygromycin resistant and had wild-type movement were found on three of those plates. Worms that were hygromycin resistant but retained the Unc phenotype were also observed on some plates, representing individuals with integrations at a single site. For two of the plates a single pair of integrations was observed, in both cases being *ahr-1p::mScarlet* plus *hlh-16p::mNeonGreen*. For the third plate, two different combinations were recovered: *ahr-1p::mScarlet* plus *mdl-1p::mNeonGreen* and *ceh-40p::mScarlet* plus *ceh-10p::mNeonGreen* (Figure 5C). While multi-site CRISPR is known to be possible (Arribere et al., 2014), these results suggest that TARDIS provides a unique way to engineer multiple locations using a single injection.

When transcriptional reporter lines were examined by fluorescent microscopy, expression of the fluorophores was concentrated in but not exclusive to the nucleus, consistent with the presence of nuclear localization signals (NLS) on the fluorophores. For all promoters, expression was seen in at least one previously reported tissue (Table 1) but was absent in one or more tissues for several of the promoters. Expression of single copy reporters is frequently more spatially restricted than that from integrated or extrachromosomal arrays (Aljohani et al., 2020). The differences in expression pattern may also reflect differences in the region used as the promotor or the fact that only a single developmental stage (late L4/early adult) was examined. Overall, we find that TARDIS can be used to screen functional libraries, either individually or in combination.

Discussion

Here we present the first implementation of a practical approach to large-scale library transgenesis in an animal system (Figure 1). Building on over a half century of advancements in *C. elegans* genetics, we can now make thousands of specific, independent genomic integrations from single microinjection events that traditionally yield at most a small handful of transgenic individuals. Increasing transgenesis throughput has long been desired, and in *C. elegans* several attempts have been made to multiplex transgenic protocols. Library mosSCI and RMCE, which

both introduce a multiplexed injection mixture and do indeed achieve multiple integrations (Kaymak et al., 2016; Nonet, 2020). However, just as in the case of standard mosSCI or single donor injections for RMCE, anti-array screening, genotyping, and the direct integration of the process limit the multiplex potential substantially. One group has adopted arrays with small pools of guides coupled with heatshock inducible Cas9 to produce randomized mutations at targeted locations (Froehlich et al., 2021). (Mouridi et al., 2022) built on the utility of heatshock Cas9 and integrated three individual sequences from an array. While these prior multiplexed methods made substantial contributions in improving the efficiency of specific transgenesis, none have yet demonstrated multiplexing beyond tens of unique sequences—orders of magnitude below what would be needed for exploratory transgenesis. TARDIS therefore provides the first true library-based approach for multiplexing transgenesis in *C. elegans*. This protocol shares similarities with TARDIS, in that diverse arrays are coupled with inducible Cas9. However, the focus of that technology was to produce randomized genomic edits, and it does not produce precise, library integrations into the genome. Recently, another group

TARDIS as a method for creating barcoded individuals

Genetic barcode libraries have been applied to many high-throughput investigations to reduce sequencing costs and achieve a higher resolution within complex pools of individuals. By focusing the sequencing reads on a small section of the genome, a larger number of individual variants can be identified or experimentally followed. This critical advancement has led to the widespread use of barcoding for evolutionary lineage tracking in microbial systems (Ba et al., 2019; Blundell and Levy, 2014; Kasimatis et al., 2021; Levy et al., 2015; Levy, 2016; Venkataram et al., 2016)–uncovering the fitness effects of thousands of individual lineages without requiring large coverage depth of the whole genome. In addition to this application, using barcoded individuals can be used to facilitate any application that involves screening a large pool of diverse individuals within a shared environment. For example, barcodes have been used in microbial studies investigating pharmaceutical efficacy (Smith et al., 2011) and barcoded variant screening (Emanuel et al., 2017). The TARDIS-based system presented here provides an approximately 1,000X fold increase in barcoding throughput in *C. elegans*, making it a unique resource among multicellular models that allows the large diversity pool and design logic of microbial systems to be adapted to animal models.

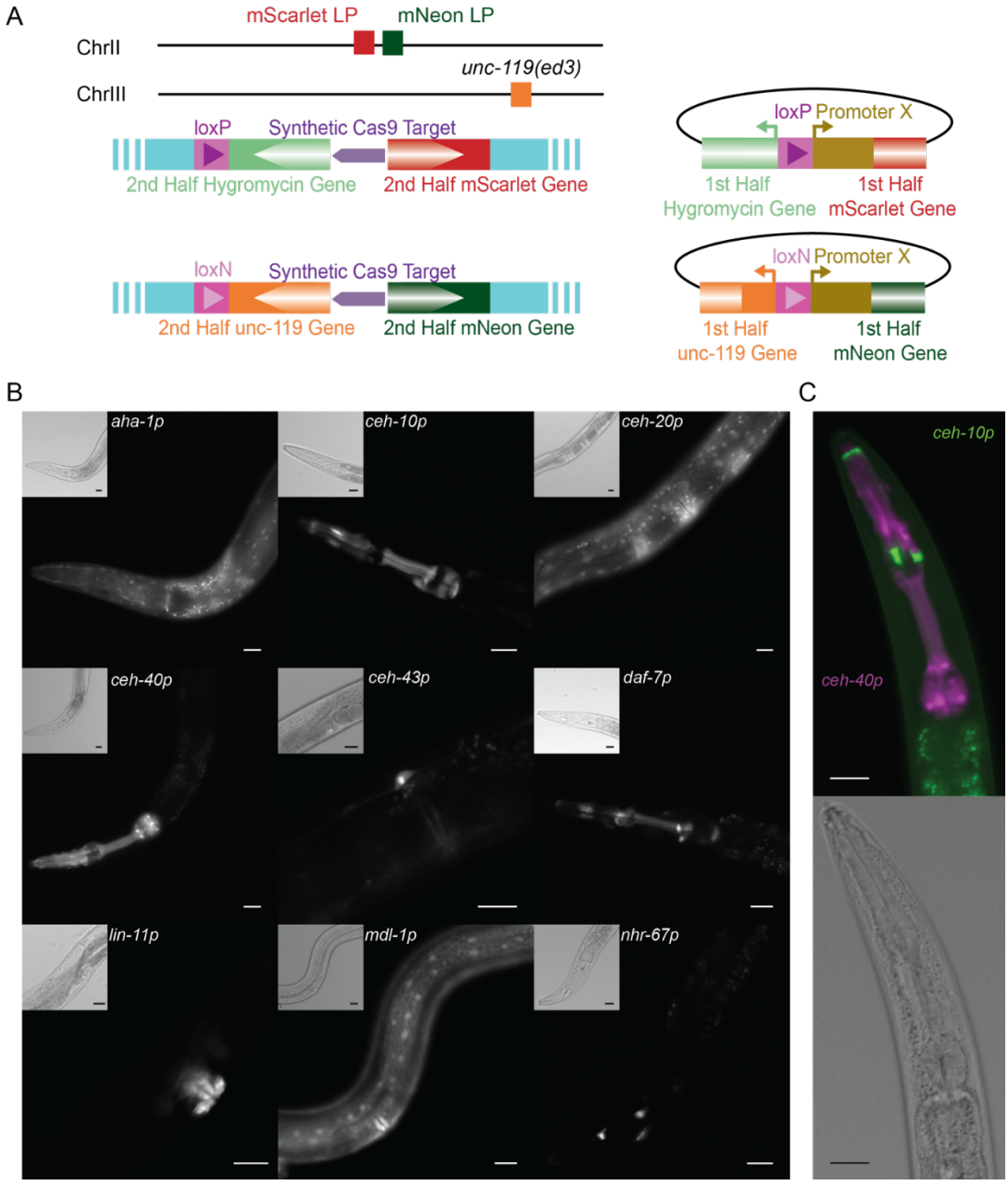
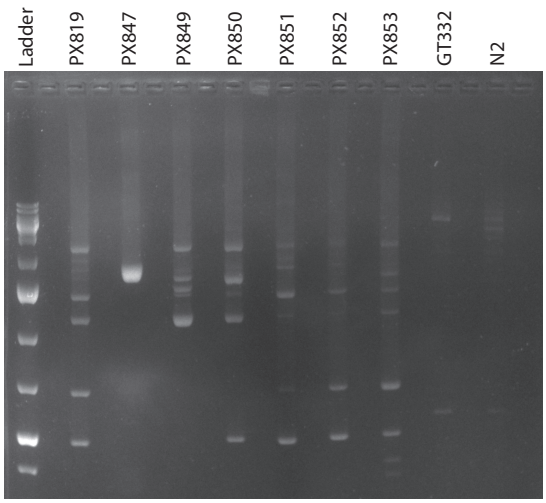


Figure 4.10 TARDIS promoter library. A) Overview of the two split landing pads and their associated promoter insertion vectors. Both the selective marker and the fluorophore expression are restored upon correct integration. B) Transcriptional reporters for nine genes were recovered from a single heatshock of a single TARDIS array line (PX819). Integration was into the single *mScarlet-I*/HygR landing pad. Main images show mScarlet-I expression for the indicated reporter while insets show polarized image of the same region. C) Example simultaneous, dual integration from a single TARDIS array into the double landing pad strain with PEST. *ceh-10p*::*mNeonGreen*::*PEST* is false colored green and *ceh-40p*::*mScarlet-I*::*PEST* is false colored magenta. All scale bars represent 20µm

A.



B.

	aha-1p	*hlh-16p*	*ceh-40p*	*ceh-10p*	*ahr-1p*	*mdl-1p*	*egl-43p*	*ceh-20p*	*ceh-43p*	*daf-7p*	*lin-11p*	*egl-46p*	*nhr-67p*	
PX819	x	x	x	x	x	x	x	x	x	x	x		x	12/13
PX847							x						x	2/13
PX849			x	x	x	x	x		x	x	x	x	x	10/13
PX850	x		x	x	x	x	x	x	x	x	x	x	x	12/13
PX851	x	x	x	x	x	x	x	x	x	x	x	x	x	13/13
PX852	x	x	x	x	x	x	x	x	x	x	x	x	x	13/13
PX853	x	x	x	x	x	x	x	x	x	x	x	x	x	13/13
	5/7	4/7	6/7	6/7	6/7	6/7	7/7	5/7	6/7	6/7	6/7	5/7	7/7	

C.

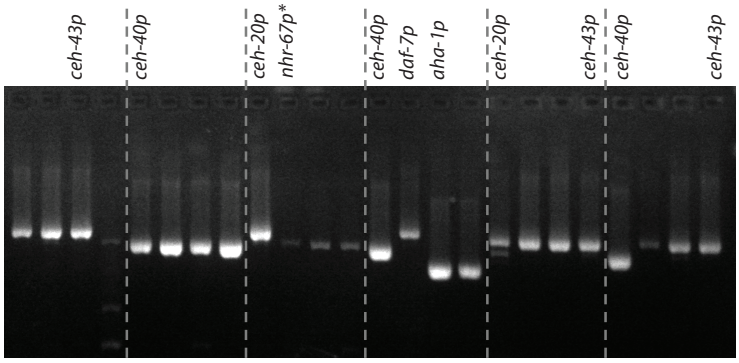


Figure 4.11—figure 4.10 supplement 1: (A) Promoters from seven array bearing lines were amplified using universal primers and show distinct profiles. Note that not all promoters found in the arrays were detected in this screen. (B) Promoters contained in each array bearing line as determined by promoter specific PCR (C) Line PX819 was heat shocked to trigger integration and four hygromycin resistant progeny were singled from each of the 59/60 plates with hygromycin resistant individuals. Singled worms were screened by PCR and select promoters were chosen for sequencing based on their size profile. PCR and sequencing result are shown for representative plates. *=Due to their size of the *egl-46* and *nhr-67* promoters do not reliably amplify with the universal primers. Therefore, samples with no or weak amplification were rescreened with primers specific to these two promoters (not shown).

Table 4.1 Characteristics of injected promoters and presence in tested array line (PX819) and integrated lines derived from that array.

Promoter	**Promoter size (bp)**	**Expected expression location**	**Array**	**Integrated**
aha-1	330	neurons, hypodermis, intestine, pharynx (Jiang et al., 2001)	Y	Y
hlh-16	514	head neurons (Bertrand et al., 2011)	Y	N
ceh-40	965	dopaminergic neurons (Sarov et al., 2012)	Y	Y
ceh-10	1172	neurons, seam cells (Reece-Hoyes et al., 2007)	Y	Y
ahr-1	1387	ALM and RME neurons (Huang et al., 2004)	Y	N
mdl-1	2000	neurons, body wall, pharynx (Reece-Hoyes et al., 2007)	Y	Y
egl-43	2001	neurons, gonad (Hwang et al., 2007)	Y	N
ceh-20	2015	neurons, seam cells, vulva (Reece-Hoyes et al., 2007)	Y	Y
ceh-43	2096	neurons, anterior hypodermis (Reece-Hoyes et al., 2007)	Y	Y
daf-7	2524	head neurons, coelemocytes, pharynx (Klabonski et al., 2016)	Y	Y
lin-11	2857	neurons, uterus, vulva, head muscle (Gupta et al., 2003)	Y	Y
egl-46	4477	Neurons (Wu et al., 2001)	N	N
nhr-67	5545	neurons, excretory cell, rectal valve cell, vulva (Fernandes and Sternberg, 2007)	Y	Y

While we designed our barcode sequence units for the purpose of barcoding individuals, this approach could also prove useful in future optimization and functional understanding of array-based processes. In particular, the high sequence diversity but identical physical design of the synthetic barcode library may provide a unique window into extrachromosomal array biology that would be helpful in optimizing sequence units for incorporation into heritable TLAs. For example, an unexpected result of the barcoding experiment was the discovery that a small minority of sequences were overrepresented, or 'jackpotted,' in the TLA relative to their frequency in the injection mix (Figure 3 and Figure 3—figure supplement 1). Our expectation was that arrays would form in an equal molar fashion proportional to the injection mix based on the model that arrays are formed by physical ligation of the injected DNA fragments (Mello et al., 1991). Deviations from random array incorporation have been observed before, and a bias for incorporating larger fragments has been proposed as an explanatory mechanism (Lin et al., 2021). Our results suggest that the ultimate array composition is not directly proportional to the molarity of the injected fragments, or strictly weighted towards the size of the fragment as has been suggested. In contrast, we propose that array size affects the maintenance of extrachromosomal arrays. As such, selection can act to increase the rate of recovery for arrays that have increased their size through random amplification of some sequences by an unknown process early in the formation of the array or by expansion of similar sequences by DNA polymerase slippage during replication, as has been well documented for native chromosomes (Levinson and A., 1987). These hypotheses would be consistent with observations of Lin et. al. (2021), if the underlying mechanism for their observation is that inclusion of larger fragments tends to be positively correlated with ultimate array size, and therefore likelihood of maintenance.

TARDIS as a method for the introduction of promoters and other large constructs

While the barcode approach demonstrates the potential for using TARDIS to integrate large numbers of 433bp PCR products, previous work using CRISPR/Cas9 initiated homology directed repair has suggested that integration efficiencies decrease with the size of the insert (Dickinson and Goldstein, 2016). We therefore implemented TARDIS for integrating promoters cloned into a vector backbone and ranging in size from 330bp to 5.5kb, to determine TARDIS

functionality under a physically different use case directed specifically at functional analysis. We found that promoter libraries could be integrated into either single sites or two sites simultaneously. Unsurprisingly, the frequency at which various promoters were recovered varied from array to array (for example *ahr-1p* was never recovered in the single site integration experiment, despite being present in the array, while it was the most common promoter recovered in the two-site integration experiment) and likely reflects the same relationship between integration frequency and prevalence in the array, as was seen with the distribution of insert abundance for the barcodes. While we showed that plasmid donors can be used in the TARDIS pipeline, neither of our two arrays contained all 13 plasmids. Given that the estimated 1-13MB size of arrays (Carlton et al., 2022) would be adequate to hold copies of each of the plasmids, as well as the extreme diversity obtained when using smaller DNA fragments, differential presence of a given promotor fragment was somewhat unexpected. This may reflect a preferential use of linear fragments in the *in-situ* assembly of arrays. Future use of linear fragments where feasible may increase incorporation and overall diversity (Priyadarshini et al., 2022).

For both the one and two site promoter library integrations, transgenic individuals were readily detected, suggesting the TARDIS method for integration was highly efficient. It has long been understood that successful CRISPR editing at one site significantly increases the chances of successful editing at a second site. This is the premise behind commonly used co-conversion screening strategies (also referred to as co-CRISPR), such as the *dpy-10* screen commonly used in *C. elegans* (Arribere et al., 2014; Ward, 2015). Here we show that same type of co-conversion also occurs when using only "large" (>1kb), plasmid-based repair templates containing gene-sized repair constructs. Additionally, we have simultaneously targeted the same two landing pads presented here using standard CRISPR techniques and find that approximately half of hygromycin resistant individuals also have rescue of the Unc phenotype (i.e. editing has occurred at both sites; data not shown). Given the high rate of co-conversion, this work demonstrates multiplex integrations are possible not only by targeting multiple repair templates to a single site but also by simultaneously utilizing multiple insertion sites.

In order to recover individual edits most efficiently, given the high frequency of integration using TARDIS, we recommend to either heat shock small cohorts of array bearing

individuals, such that most cohorts only yield one edited individual or to screen multiple individuals per cohort. Additionally, while split-selection methods allow for direct verification of integration, depending on the downstream use case, integrations should be confirmed by sequencing as errors can still occur, including internal deletions within the insert.

Expansion of TARDIS to other multicellular systems

Unlocking the investigative potential of transgenesis in animal systems would enable exploratory experiments normally restricted to single-cell models. For example, alanine scanning libraries and protein-protein interactions (Cunningham and Wells, 1989; Matthews, 1996; Wells, 1991), CRISPR library screening (Bock et al., 2022), and promoter library generation (Delvigne et al., 2015; Zaslaver et al., 2006). While we demonstrate the use of TARDIS in *C. elegans* here, the intellectual underpinnings of the approach are agnostic to the research model used. Conceptually, TARDIS facilitates high-throughput transgenesis by using two engineered components: a heritable TARDIS Library containing multiplexed transgene-units and a genomic split selection landing pad that facilitates integration of single sequence units from the library. To generate the first TARDIS libraries, we capitalized on the endogenous capacity of *C. elegans* to assemble experimentally provided DNA into heritable extrachromosomal arrays. Extrachromosomal arrays are formed from exogenous DNA, are megabases in size (Lin et al., 2021; Woglar et al., 2020), do not require specific sequences to form and replicate, and can be maintained in a heritable manner via selection (Mello et al., 1991). These qualities make them suitable for use as a heritable library upon which TARDIS can be based. To adopt TLAs in systems beyond *C. elegans*, methods must be adopted to introduce large heritable libraries into the germline, as most systems do not maintain extrachromosomal arrays. In mice, the locus *H11* has been used for large transgenic insertions (Liu et al., 2022), while in *Drosophila*, the use of PhiC31-mediated transgenesis coupled with bacterial artificial chromosomes (BACs) have allowed for many approximately 10kb+ sized fragments to be integrated into their respective genomes (Venken et al., 2006). Each of these large integration strategies can provide a vehicle for stable inheritance of a TLA.

The second component of the TARDIS integration system is a pre-integrated landing pad sequence. We have generally favored split selectable landing pads (SSLPs) that use HygR for its effectiveness (Mouridi et al., 2022; Stevenson et al., 2021, 2020). The SSLPs are engineered to accept experiment-specific units from the array. For example, here we used SSLPs designed to

accept barcodes for experimental lineage tracking and promoters for generation of transcriptional reporters. To translate TARDIS to other systems, a genomic site needs to be engineered to act as a landing pad that can utilize sequence units from the TLA and can be customized to the specific system and use. Because TLAs allow the experimenter to design the library of interest and the landing pad to recapitulate the strengths of single-cell systems, adoption of TARDIS in multicellular animal experiments can leverage the high-resolution, high-diversity exploratory space of DNA synthesis. In addition to adapting assays currently restricted to single-cell models, TARDIS also opens the door to animal-specific uses, such as developmental biology, neurobiology, endocrinology, and cancer research.

In developmental genetics, the lack of large-library transgenesis has resulted in 'barcode' libraries in a different form, utilizing randomized CRISPR-induced mutations to form a unique indel. For example, GESTALT (McKenna et al., 2016) creates a diversity of barcodes *in-vivo* via random indel formation at a synthetic target location. LINNAEUS (Spanjaard et al., 2018) similarly utilizes randomized targeting of multiple RFP transgenes to create indels, allowing for cells to be barcoded for single cell sequencing. TARDIS barcodes do not rely on randomized indel generation and thus can be much simpler to implement with sequencing approaches outlined above.

In-vivo cancer models have also adopted the high-resolution, high-variant detection of barcodes for the study of tumor growth and evolution. Rogers at al., developed Tuba-seq (Rogers et al., 2017; Winslow, 2022), a pipeline that takes advantage of small barcodes allowing for *in-vivo* quantification of tumor size. In Tuba-seq, barcodes are introduced via lentiviral infection, leading to the barcoding of individual tumors. TARDIS brings the multiplexed library into the animal context without requiring viral vectors or intermediates, thereby allowing large *in-vivo* library utilization and maintenance. Capitalizing on the large sequence diversity possible within synthesized DNA libraries with a novel application in multicellular systems generates new opportunities for experimental investigation in animal systems heretofore only possible within microbial models.

Summary

In summary, here we have presented Transgenic Arrays Resulting in Diversity of Integrated Sequences (TARDIS), a simple yet powerful approach to transgenesis that overcomes the

limitations of multicellular systems. TARDIS uses synthesized sequence libraries and inducible extraction and integration of individual sequences from these heritable libraries into engineered genomic sites to increase transgenesis throughput up to 1000-fold. While we demonstrate the utility of TARDIS using *C. elegans*, the process is adaptable to any system where experimentally generated genomic loci landing pads and diverse, heritable DNA elements can be generated.

Materials and Methods

Table 4.2 Key Reagents

Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**	**Additional information**
Gene Promoter (aha-1)	*aha-1p*	wormbase.org	WBGene00000095	
Gene Promoter (hlh-16)	*hlh-16p*	wormbase.org	WBGene00001960	
Gene Promoter (ceh-40)	*ceh-40p*	wormbase.org	WBGene00000461	
Gene Promoter (ceh-10)	*ceh-10p*	wormbase.org	WBGene00000435	
Gene Promoter (ahr-1)	*ahr-1p*	wormbase.org	WBGene00000096	
Gene Promoter (mdl-1)	*mdl-1p*	wormbase.org	WBGene00003163	
Gene Promoter (egl-43)	*egl-43p*	wormbase.org	WBGene00001207	
Gene Promoter (ceh-20)	*ceh-20p*	wormbase.org	WBGene00000443	
Gene Promoter (ceh-43)	*ceh-43p*	wormbase.org	WBGene00000463	
Gene Promoter (daf-7)	*daf-7p*	wormbase.org	WBGene00000903	
Gene Promoter (lin-11)	*lin-11p*	wormbase.org	WBGene00003000	
Gene Promoter (egl-46)	*egl-46p*	wormbase.org	WBGene00001210	

Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**	**Additional information**
Gene Promoter (nhr-67)	*nhr-67p*	wormbase.org	WBGene00003657	
Strain (*Caenorhabditis elegans*)	*N2*	Caenorhabditis Genetics Center		
Strain (*C. elegans*)	*N2-PD1073*	doi: 10.17912/micropub.biology.000518		Available from the Caenorhabditis Intervention Testing Program- upon request (https://citp.squarespace.com/)-
Strain (*C. elegans*)	PX740	This paper		N2-PD1073 fxIs47 [*rsp-0p:: 5' ΔHygR:: GCGAAGTGACGGTAGACCGT :: 3' ΔHygR::unc-54 3'::loxP*]
Strain (*C. elegans*)	GT331	This paper		aSi9[*lox2272 Cbr-unc-119(+) lox2272 + loxP 3'3' ΔHygR + 3' ΔmScarlet-I::PEST*]; *unc-119(ed3)*
Strain (*C. elegans*)	GT332	This paper		aSi10[*lox2272 Cbr-unc-119(+) lox2272 + loxP 3' ΔHygR + 3' ΔmScarlet-I*]; *unc-119(ed3)*
Strain (*C. elegans*)	GT336	This paper		aSi12*[lox2272 rps-0p::HygR + hsp-16.41p::Cre::tbb-2 3'UTR + sqt-1(e1350) lox2272 + loxN 3' ΔCbr-unc-119(+)::tjp2a_guide:: 3' ΔmNeonGreen::PEST::egl-13nls::tbb-2 3'UTR]* aSi9*[lox2272 Cbr-unc-119(+) lox2272 + loxP 3'ΔHygR::tjp2a guide::3'ΔmScarlet-I::PEST::egl-13nls::tbb-2 3'UTR] II; unc-119(ed3) III*
Strain (*C. elegans*)	GT337	This paper		aSi13[*lox2272 + loxN 3' ΔCbr-unc-119(+) + 3' ΔmNeonGreen::PEST]* aSi14*[lox2272 + loxP 3' ΔHygR + 3' ΔmScarlet-I::PEST*]; *unc-119(ed3),*
Strain (*C. elegans*)	QL74	Gift from QueeLim Ch'ng		oxEx1578 [*eft-3p::GFP + Cbr-unc-119(+)*] 6x outcross EG4322
Strain (*C. elegans*)	PX786	This paper		fxEx23 [TARDIS #5 *5' ΔHygR::Intron5'::Read1::NNNCNNTNTNANNNN::Read2::Intron3'':: 3' ΔHygR* (89 Unique Sequences) *hsp-16.41p::piOptCas9::tbb-2 34' UTR + rsp-27p::NeoR::unc-54 3' UTR + U6p:: GCGAAGTGACGGTAGACCGT*]; fxSi47[*rsp-0p:: 5' ΔHygR:: GCGAAGTGACGGTAGACCGT :: 3' ΔHygR::unc-54 3'::loxP*

Table 4.2 Cont.

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Strain (*C. elegans*)	PX816	This paper		fxEx25 [TARDIS #1 *5′ ΔHygR::Intron5′::Read1::NNNCNNTNTNANNNN::Read2::Intron3′:: 3′ ΔHygR* (1,319 Unique Sequences) *hsp-16.41p::piOptCas9::tbb-2 34′ UTR + rsp-27p::NeoR::unc-54 3′ UTR + U6p:: GCGAAGTGACGGTAGACCGT*]; fxSi47[*rsp-0p:: 5′ ΔHygR:: GCGAAGTGACGGTAGACCGT :: 3′ ΔHygR::unc-54 3’::loxP*
Strain (*C. elegans*)	PX817	This paper		fxEx26 [TARDIS #2 *5′ ΔHygR::Intron5′::Read1::NNNCNNTNTNANNNN::Read2::Intron3′:: 3′ ΔHygR* (3,001 Unique Sequences) *hsp-16.41p::piOptCas9::tbb-2 34′ UTR + rsp-27p::NeoR::unc-54 3′ UTR + U6p:: GCGAAGTGACGGTAGACCGT*]; fxSi47[*rsp-0p:: 5′ ΔHygR:: GCGAAGTGACGGTAGACCGT :: 3′ ΔHygR::unc-54 3’::loxP*
Strain (*C. elegans*)	PX818 profile 1	This paper		fxEx27 [TARDIS #3 *5′ ΔHygR::Intron5′::Read1::NNNCNNTNTNANNNN::Read2::Intron3′:: 3′ ΔHygR* (91 Unique Sequences) *hsp-16.41p::piOptCas9::tbb-2 34′ UTR + rsp-27p::NeoR::unc-54 3′ UTR + U6p:: GCGAAGTGACGGTAGACCGT*]; fxSi47[*rsp-0p:: 5′ ΔHygR:: GCGAAGTGACGGTAGACCGT :: 3′ ΔHygR::unc-54 3’::loxP*
Strain (*C. elegans*)	PX818 profile 2	This paper		fxEx28 [TARDIS #4 *5′ ΔHygR::Intron5′::Read1::NNNCNNTNTNANNNN::Read2::Intron3′:: 3′ ΔHygR* (204 Unique Sequences) *hsp-16.41p::piOptCas9::tbb-2 34′ UTR + rsp-27p::NeoR::unc-54 3′ UTR + U6p:: GCGAAGTGACGGTAGACCGT*]; fxSi47[*rsp-0p:: 5′ ΔHygR:: GCGAAGTGACGGTAGACCGT :: 3′ ΔHygR’::unc-54 3’::loxP*
Strain (*C. elegans*)	PX819	This paper		N2 fxEx24 [(*rps-0p:: 5’ ΔHygR +loxP + aha-1p::SV40 NLS:: 5’ ΔmScarlet-I) + (rps-0p:: 5’ ΔHygR +loxP + ahr-1p::SV40 NLS::5’ ΔmScarlet-I) + (rps-0p:: 5’ ΔHygR +loxP + ceh-10-1p::SV40 NLS::5’ ΔmScarlet-I) + (rps-0p:: 5’ ΔHygR*

Table 4.2 Cont.

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
				*+loxP + ceh-20p::SV40 NLS::5' ΔmScarlet-I) + (rps-0p:: 5' ΔHygR +loxP + ceh-40p::SV40 NLS::5' ΔmScarlet-I) + (rps-0p:: ΔHygR +loxP + ceh-43p::SV40 NLS::5' ΔmScarlet-I) + (rps-0p:: 5' ΔHygR +loxP + daf-7p::SV40 NLS::5' ΔmScarlet-I) + (rps-0p:: ΔHygR +loxP + egl-43p::SV40 NLS::5' ΔmScarlet-I) + (rps-0p:: 5' ΔHygR +loxP + hlh-16p::SV40 NLS::5' ΔmScarlet-I) + (rps-0p:: 5' ΔHygR +loxP + lin-11p::SV40 NLS::5' ΔmScarlet-I) + (rps-0p:: 5' ΔHygR +loxP + mdl-1p::SV40 NLS::5' ΔmScarlet-I) + (rps-0p:: 5' ΔHygR +loxP + nhr-67p::SV40 NLS::5' ΔmScarlet-I) + hsp-16.41p::piOptCas9::tbb-2 34' UTR + prsp-27::NeoR::unc-54 3' UTR]; aSi10[lox2272 + Cbr-unc-119(+) + lox2272 + loxP + 5' ΔHygR::unc-54 3' UTR + 5' ΔmScarlet-I::egl-13 NLS::tbb-2 3' UTR, II:*8420157]*; unc-119(ed3) III*
Strain (*C. elegans*)	EG4322	doi.org/10.1038/ng.248; Caenorhabditis Genetics Center		
Strain (*Escherichia coli*)	PXKR1	This paper		NA22 transformed with pUC19
Recombinant DNA	Plasmid pDSP15	This paper	193853 (Addgene)	*5' ΔHygR::loxP::MCS::5' Δ mScarlet-I*
Recombinant DNA	Plasmid pDSP16	This paper	193854 (Addgene)	*5' ΔCbr-unc-119(+)::loxN::MCS::5' Δ 5'mNeonGreen*
Recombinant DNA	Plasmid pMS84	This paper	193852 (Addgene)	*U6p*::GGACAGTCCTGCCGAGGTGG
Recombinant DNA	Plasmid pZCS36	This paper	193048 (Addgene)	*hsp16.41p::Cas9(dpiRNA)::tbb-2 '3UTR*
Recombinant DNA	Plasmid pZCS38	This paper	193049 (Addgene)	*rsp-27p::NeoR::unc-54 3' UTR*
Recombinant DNA	Plasmid pZCS41	This paper	193050 (Addgene)	*U6p*::GCGAAGTGACGGTAGACCGT
Sequence-based reagent	ZCS422	This paper		Design and construction of barcode donor library
commercial kit	DNA Clean and Concentrator	Zymo Research	Cat No.: D4004	

Table 4.2 Cont.

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
commercial kit	Genomic DNA Clean and Concentrator	Zymo Research	Cat. No.: D4011	
commercial kit	Zymoclean Gel DNA Recovery Kit	Zymo Research	Cat. No.: D4008	
commercial kit	Zyppy Plasmid Miniprep Kit	Zymo Research	Cat. No.: D4019	
software	Cutadept	doi.org/10.14806/ej.17.1.200	Version 4.1	
software	AmpUMI	doi.org/10.1093/bioinformatics/bty264	Version 1.2	
software	Starcode	doi.org/10.1093/bioinformatics/btv053	Version 1.4	
software	Google colab	colab.research.google.com		
software	python (version)	Guido van Rossum, 1991	Version 3.7.13	
software	Juypter notebook (IPython)	doi:10.3233/978-1-61499-649-1-87	Version 7.9.0	
software	matplotlib	doi.org/10.5281/zenodo.3898017	Version 3.7.13	
software	Fiji	imagej.net/software/fiji/	Version 2.9.011.53t	
drug	G-418	GoldBio (CAS Number 108321-42-2)	Cat No.: G-418-5	
drug	Hygromycin B	GoldBio (CAS Number 31282-04-9)	Cat No.: H-270-10-1	

Table 4.2 Cont.

General TARDIS reagents

Strains generated for this publication along with key plasmids and reagents are listed in the Key Resources Table. A full list of all plasmids is given in Supplemental Table 1. All plasmids were cloned by Gibson Assembly following the standard NEB Builder HiFi DNA Assembly master mix protocol [New England Bio Labs (NEB), Massachusetts, USA], unless otherwise indicated. All plasmids have been confirmed by restriction digest, Sanger sequencing, and/or full plasmid sequencing. All primers used in the construction and validation of plasmids are listed in Supplemental Table 2.

To generate our heatshock inducible Cas9, *hsp16.41p::Cas9dpiRNA::tbb-2 '3UTR,* the *hsp16.41* promoter was amplified from pMA122 (Addgene ID34873) (Frøkjær-Jensen et al., 2012). The germline licensed Cas9 and tbb-2 3' UTR were amplified from pCFJ150-Cas9 (dpiRNA) (Addgene ID107940) (Zhang et al., 2018). All fragments were assembled into PCR linearized pUC19 vector (NEB) to give the final plasmid pZCS36.

To generate a standard empty guide vector, *U6p*::(empty)gRNA, the U6p and gRNA scaffold from pDD162 (Addgene ID47549) (Dickinson et al., 2015) was amplified and assembled into PCR linearized pUC19 to generate pZCS11.

To generate *rsp-27p::NeoR::unc-54 3' UTR,* the full resistance cassette was amplified from pCFJ910 (Addgene ID44481) and assembled into PCR linearized pUC19 vector to give pZCS38.

Genomic DNA isolation for array and integrant characterization

For processing large populations of worms, a widely used bulk lysis protocol was adapted (Fire Lab 1997 Vector Supplement, February 1997). In brief, 450µl of worm lysis buffer (0.1M Tris-Cl pH8.0, 0.1M NaCl, 50mM EDTA pH8.0, and 1% SDS) and 20µl 20mg/ml proteinase K were added to approximately 50µl of concentrated worm pellet. Samples were inverted several times to mix and incubated at 62C° for 2 hours. After incubation, samples were checked under the microscope to ensure no visible worm bodies were left in the solution. Chip DNA binding buffer (Zymo, California, USA) was added in a 2:1 ratio and gently inverted to mix. Samples were then purified with Zymo -Spin IIC-XLR columns following manufacture protocol. Samples were eluted in 50µl of water. Each sample was then digested with 10mg/ml RNase A (ThermoFisher Scientific Massachussets, USA, Cat. No. EN0531) at 42C° for 2 hours. Genomic DNA was then reisolated by adding a 2:1 ratio of Chip DNA binding buffer and purifying with Zymo-Spin IIC-XLR columns. Final genomic samples were quantified by Nanodrop.

For individual worm lysis, individual array bearing worms were isolated and lysed in 4µl of EB (Zymo, Cat. No.: D3004-4-16) buffer with 1mg/ml proteinase K (NEB). Each sample was rapidly frozen in liquid nitrogen and then thawed to disrupt the cuticle and then incubated at 58C° for 1 hour, with a subsequent incubation at 95C° for 20 minutes to inactivate the proteinase K.

TARDIS integration- general protocol

On Day 0, TARDIS array bearing *C. elegans* grown to a high density of gravid adults were hypochlorite synchronized in NGM buffer (Leung et al., 2011) and grown overnight in 15ml NGM with G-418 (1.56mg/ml) at 15C° with nutation. On Day 1, L1s were washed three times with NGM buffer to remove G-418, plated onto media without selective agent and continued to be grown at 15°C. On Day 2, L2/L3s were heat-shocked at 35.5°C for one hour. After heat shock, worms were grown at 25°C until gravid adults when Hygromycin B was top spread on plates at a final concentration of 250µg/ml.

Construction of landing pad for barcodes

To create the barcode landing pad, an intermediate Chr. II insertion vector, pZCS30 was built from pMS4 by using PCR to remove the *let-858* terminator. pZCS30 served as the vector backbone for pZCS32. To assist in cloning, the backbone was split into two PCR fragments. The broken *HygR* gene was amplified in two parts, *rsp-0p::5'ΔHygR* and *3'ΔHygR::unc-54* 3' UTR, from pCFJ1663 (Addgene ID51484). Overlapping PCR was used to fuse both *HygR* fragments. The resulting broken *HygR* cassette removed the intron found in pCFJ1663 as well as four codons from exon one and three codons from exon two, while also creating +1 frameshift and a reverse orientation guide RNA target for pZCS41. A second overlapping PCR was performed to fuse the broken *HygR* cassette to backbone fragment two. The resulting two-part clone was then assembled to give pZCS32.

The barcode landing pad TARDIS strain, PX740 was created by injecting a mixture of 10ng/µl pZCS32, 50ng/µl pMS8, and 3ng/µl pZCS16 (Addgene ID154824) (Stevenson et al., 2020) into the gonad of young adult N2-PD1073 (Teterina et al., 2022) hermaphrodites. Screening and removal of the SEC were performed following (Dickinson et al., 2015). Presence of the correct insertion was confirmed by Sanger sequencing using the primers listed in Supplemental Table 3.

To create the barcode landing pad targeting guide RNA, *U6p*:: GCGAAGTGACGGTAGACCGT, the guide sequence GCGAAGTGACGGTAGACCGT was added by overlapping primers to the vector pZCS11 to give the final construct pZCS41.

Design and construction of barcode donor library

Oligo ZCS422 was ordered with 11 randomized N's (hand mixed bases) [Integrated DNA Technologies (IDT), Iowa, USA] and has the following sequence: CTACACGACGCTCTTCCGATCTNNNCNNTNTNANNNNAGATCGGAAGAGCACACGT CTG. Four 'hard-coded' base pairs were included within the randomized sequence. ZCS422 was used as the core for the generation of two separate complexing PCR barcode homologies referred to as "barcode-15X" and "barcode-20X" to denote the number of complexing cycles (Figure 2). All PCRs were performed using the high-fidelity Q5 polymerase (NEB) as per manufacture instructions. All primers used for barcode synthesis can be found in Supplemental Table 4. For both "barcode-15X" and "barcode-20X, the left and right homology arms were generated separately by PCR and purified by gel extraction. An initial 10 cycle PCR was performed to convert the oligo into a 201bp double stranded product which was gel extracted with Zymo clean Gel DNA Recovery Kit (Cat. No.: D4008) following manufacture protocol. The low cycle number was done to retain diversity and to minimize effects of PCR jackpotting.

For "barcode-15X", to generate the complete donor homology, the double stranded barcode template was combined with both the left and right homology arms for a three-fragment overlap PCR. To maximize diversity, high concentrations of the individual templates were used. The reaction contained 52 fmol/µl of barcode template and 22 fmol/µl of left right arms in a single 50µL Q5 reaction. A total of 15 cycles were performed. The lower cycle was again done to reduce PCR jackpotting. The single product was gel extracted as a 433bp fragment. The final donor fragment is referred to as 'barcode-15X.'

To generate "barcode-20X" a similar three-fragment overlap PCR was used. 4.3fmol/µl of barcode template, 15.33fmol/µl of left arm, and 3.3fmols/µl of right arm were combined across six Q5 50ul reactions and a total of 20 cycles were performed. The right arm concentration was lower caused by low concentration extraction. The single product was gel extracted as a 433bp fragment. The final donor fragment is referred to as 'barcode-20X.'

Generation of barcode TLA lines

The TARDIS array bearing line PX786 was created by injecting 50ng/µl of barcode-15X, 10ng/µl pZCS38, 15ng/µl pZCS41, 5ng/µl pZCS16, and 20ng/µl pZCS36 into young adult

PX740 hermaphrodites. Individual injected worms were grown at 15C° for four days and then treated with G-418 (1.56mg /ml). A single stable array line was isolated and designated PX786.

The TARDIS array bearing lines PX816, PX817, PX818 profile 1 and PX818 profile 2 were created by injecting 100ng/µl of barcode-20X, 10ng/µl pZCS38, 15ng/µl pZCS41 and 20ng/µl pZCS36. Individual injections were grown at 15°C for four days and then treated with G-418 (1.56mg/ml). Full genotypes are provided in Supplemental Document 1 as the full genotypes cannot be contained within a table.

Estimation of barcode integration frequency population sample preparation

PX786 was grown to gravid adults in the presence of G-418 with concentrated NA22 transformed with pUC19 for ampicillin resistance as a food source (designated PXKR1). Once gravid, the strain was hypochlorite synchronized and grown overnight in 15ml NGM buffer with G-418 at 15°C with nutation. For each of the four replicates, a synchronized L1 population was divided in half. The first half was pelleted by centrifugation (2,400xg for two minutes) and frozen (-20°C) until processed. These samples represented the array-bearing samples. Another sample of approximately 150,000 L1s was plated to large NGM and subjected to the standard TARDIS heat shock and grown until the population was primarily gravid adults. Then, this population was hypochlorite synchronized and grown in NGM buffer at 15°C with Hygromycin B (250 µg/ml). These entire samples were pelleted and frozen, representing the F1 samples.

PCR for barcode quantification

Several different PCRs were performed depending on the intended downstream sequencing quantification. See Figure 2-figure supplement 1 for a schematic layout of the different PCR steps. The primers used for barcode quantification are given in Supplemental Table 5. To quantify the diversity of arrays from either a bulk population or individual worms, two separate PCRs were performed to quantify the diversity of arrays.

The first PCR (Amplicon one-array) was performed for three cycles to add Unique Molecular Identifiers (UMI), allowing for downstream de-duplication. For each sample either 100ng of genomic DNA (bulk samples) or the entirety of the worm lysate (single worms) was used as the template. PCR samples were then purified using the Zymo DNA Clean and Concentrator-5 Kit (Cat. No.: D4004) following manufacture protocol and eluted with 24µl water. Samples were not quantified prior to the next step as most DNA will not be from the

target PCR product. A second PCR (Amplicon two) using the entire 24µl of the extract from the previous step was performed for 24 cycles to added indices. In some cases, a smaller, non-specific product was also formed, so each sample was run on a 2% agarose gel and extracted for the 169bp size product.

Two separate PCRs were performed to quantify the diversity of integrated barcode sequences. The first PCR (Amplicon one-integrant) was performed for three cycles to add UMI sequences. For each sample, 100ng of genomic DNA was used as the template. PCR products were then purified as described above and followed the Amplicon two protocol. Each product was quantified on a Synergy H1 plate reader using software Gen 5 3.11. Samples were mixed at an equal molar ratio for a 20nM final concentration for Illumina Sequencing.

Illumina sequencing and data processing for barcode characterization

To quantify the diversity of barcodes in each sample, PCR products were sequenced on either a single NextSeq 500 lane or NovaSeq SP, with single read protocols performed by the Genomics and Cell Characterization Facility (GC3F) at the University of Oregon. Compressed fastq files were processed with cutadept 4.1 (Martin, 2011) to remove low quality reads (quality score < 30, max expected error=1, presence of 'N' within the read) and trimmed to 87bp. For the NextSeq lane the specific nextseq trim=30 command was used. The sequences were then demultiplexed using cutadept. For duplicate removal, AmpUMI (Clement et al., 2018) in "processing mode" was used with umi regex "CACIIIIIIIIIGAC" for individual index files. Deduplicated reads were then trimmed to 15 base pairs with cutadapt for each file. Starcode (Zorita et al., 2015) was then used for mutation correction and counting of each barcode sequence. Each unique sequence was only kept if its final length was 15 base pairs. For the injection mix, each unique barcode was kept regardless of total reads. For all TARDIS arrays and F1 integrations, we used the observed plateau in the number of observed unique barcodes for various count cutoffs to establish a conservative threshold of five or more reads for true barcode sequence (Figure 3—figure supplement 3). Visualizations were created with Python 3.7.13 (Rossum, 1991) and matplotlib 3.5.2 (Hunter, 2007). Sequence logos were created with Logomaker (Tareen and Kinney, 2020) Correlation and p-values generated by scipy input stats.pearsonr (Virtanen et al., 2020). This statistical test was chosen because the relationship from array to integration is approximately linear. All data were processed in Jupyter Notebooks (Kluyver et al., 2016)

utilizing Google Colaboratory (colab.research.google.com). All python code is available on Figshare.

Design of landing pads for transcriptional reporters

The utilized fluorophores, mScarlet-I (Bindels et al., 2017) and mNeonGreen (Shaner et al., 2013) were synthesized with the desired modifications as genes incorporated into the pUCIDT-KAN plasmid (IDT). First, a SV40 nuclear localization sequence (NLS) was added after the 13th codon of the *mScarlet-I* gene. This same 66bp sequence was also used in place of the first four codons of the *mNeonGreen* gene. Secondly, a PEST domain (Li et al., 1998) flanked by MluI restriction endonuclease sites and an additional NLS from the *egl-13* gene (Lyssenko et al., 2007) were added to the 3' end of the gene. The *C. elegans* Codon Adapter (https://worm.mpi-cbg.de/codons/cgi-bin/optimize.py) (Redemann et al., 2011) was used to codon optimize both modified fluorophore sequences and to identify locations for three synthetic introns. The first two introns contained 10-base pair periodic A_n/T_n-clusters (PATCs), which have been shown to reduce rates of transgene silencing (Frøkjær-Jensen et al., 2016), while the third was a standard synthetic intron. Finally, the 3' UTR of the *tbb-2* gene, which is permissive for germline expression (Merritt et al., 2008), was added to the end of fluorophore genes. The modified *mScarlet-I* and *mNeonGreen* genes were PCR amplified and assembled into NotI and SnaBI linearized pDSP1, a standard backbone vector derived from pUCIDT-KAN. The resulting *mScarlet-I*-containing plasmid was designated pDSP6 and the *mNeonGreen* containing plasmid was designated pDSP7. In addition, pDSP9, a version of mScarlet-I lacking the PEST destabilization sequence was generated by PCR amplifying the shared *egl-13* NLS and *tbb-2* 3' UTR sequence from pDSP6 and then assembling this fragment into MluI and SnaBI linearized pDSP6.

Landing pads were built using a modification of our previous split landing pad strategy (Stevenson et al., 2020). Each landing pad contained the 3' portion of a selectable marker followed by a validated guide sequence and the 3' portion of a fluorophore. The guide sequence (GGACAGTCCTGCCGAGGTGGAGG) has no homology in the *C. elegans* genome and has been previously shown to allow for efficient editing (Stevenson et al., 2020). This sequence was targeted by the plasmid pMS84 which was made from pZCS2, a plasmid made in the same manner as pZCS11 but which is missing a segment of the plasmid backbone, using the Q5 site-

directed mutagenesis kit (NEB). *mScarlet-I* was paired with a *HygR* marker (Dickinson et al., 2013) while *mNeonGreen* gene was paired with the *Cbr-unc-119(+)* rescue cassette (Frøkjær-Jensen et al., 2008).

Construction of split HygR /mScarlet-I landing pads

The split *HygR*/*mScarlet-I* landing pad was inserted into the well-characterized *ttTi5605* Mos1 site on Chromosome II (Frokjaer-Jensen *et al.*, 2008). pQL222 (a gift from Dr. QueeLim Ch'ng), a modified version of the pCFJ350 (Frøkjær-Jensen et al., 2012) in which the original resistance marker was changed to a kanamycin and zeocin cassette, was digested with BsrGI to provide a linear vector backbone. The *Cbr-unc-119* gene, with a lox2272 sequence added to the 5' end, and a multiple cloning site (MCS) with a lox2272 site added to the 3' end were PCR amplified from pQL222. These two fragments were assembled into the linearized backbone to yield pDSP2.

Next, the 3' 949 bases of the *HygR* marker were amplified along with the *unc-54* 3'UTR from pDD282 (Dickinson et al., 2015). The primers used were designed to invert the loxP sequence at the 3' end of *unc-54* 3'UTR from its original orientation in pDD282 and to add the guide sequence to the 5' end of the *HygR* fragment. The 3' 821 bases of the *mScarlet-I* gene along with the tbb-2 3' UTR were amplified from pDSP6. These these two amplicons were assembled into a SbfI/SnaBI digested pDSP2 vector to yield pDSP61. Similarly, the *mScarlet-I* gene was amplified from pDSP9 and assembled into pDSP2 along with the *HygR* fragment to give pDSP62, a PEST-less version of the landing pad construct. Both the PEST-containing and PEST-less versions of the split *HygR*/*mScarlet-I* landing pads were integrated into QL74, a 6x outcross of EG4322 (Frøkjær-Jensen et al., 2008), using the standard MosSCI technique (Frøkjær-Jensen et al., 2012) to yield strains GT331 and GT332.

Construction of Split Cbr-unc-119(+)/mNeonGreen landing pad

To construct the *Cbr-unc-119(+)*/*mNeonGreen* landing pad, we wanted to find a genomic safe harbor site permissive to germline expression of transgenes. The *oxTi179* universal MosSCI site on Chromosome II permits germline expression but interrupts *arrd-5*, an endogenous *C. elegans* gene. Therefore, CRISPR-mediated genome editing was used to place the landing pad between *ZK938.12* and *ZK938.3*, two genes adjacent to *arrd-5* whose 3' UTRs face each other. The genomic sequence CATGGTATAAAGTGAATCAAGG was targeted by the plasmid pDSP45

which was made from pDD162 (Dickinson et al., 2013) using the Q5 site-directed mutagenesis kit (NEB).

Chromosomal regions II:9830799-9831548 and II:9831573-9832322 were amplified from genomic DNA for use as homology arms. The self-excising cassette (SEC) was PCR amplified from pDD282 such that the loxP sites were replaced by lox2272 sites. A MCS was amplified from pDSP2 while a linear vector backbone fragment was amplified from pDSP1. All five of these PCR fragments were assembled into a circular plasmid, which was immediately used as a template for seven synonymous single-nucleotide substitutions into the terminal 21bp of the *ZK938.12* gene fragment by Q5 site-directed mutagenesis kit (NEB). The resultant plasmid was named pDSP47.

The 3' 846 based of the *Cbr-unc-119(+)* rescue cassette plus the 3' UTR were amplified from pDSP2 such that the lox2272 sequence after the 3' UTR was replaced with a loxN site and the guide site GGACAGTCCTGCCGAGGTGGAGG was added upstream of the coding sequence. The 3' 818 bases of *mNeonGreen* plus the tbb-2 3' UTR were amplified from pDSP7. These two amplicons were assembled into StuI/AvrII digested pDSP47 to yield pDSP63.

Following the protocol from Dickinson et al. (2015), the landing pad from pDSP63 was integrated into the GT331 strain using pDSP45 as the guide plasmid. Upon integration, this yielded strain GT336. Activation of the Cre recombinase within the SEC by heat shock caused both the removal of the SEC from the *mNeonGreen* landing pad and the *Cbr-unc-119(+)* cassette from the *mScarlet-I* landing pad. The combined effect of this double excision event was to yield strain GT337 which has an Unc phenotype and no longer has the hygromycin resistance and Rol phenotypes.

Design and construction of promoter library

Targeting vectors were constructed to provide the 5' portions of each split gene pairing. Both targeting vectors had the same multiple cloning site, allowing promoter amplicons to be assembled into either vector using the same set of primers. In addition, each selectable marker gene is flanked by a lox site that matches the sequence and orientation of the lox site flanking the 3' portion of the marker in the genomic landing pad, allowing for the optional post-integration removal of the selectable marker gene using Cre recombinase.

To construct the split *HygR*/*mScarlet-I* targeting vector, the *rps-0* promoter plus the 5' 627 bases of the *HygR* gene were amplified from pDD282 such that a loxP site was added in front of the promoter sequence The MCS was amplified from pDSP2 and the 5' 803 bases of the *mScarlet-I* gene were amplified from pDSP6. All three of these amplicons were assembled into NotI/SnaBI digested pDSP1 to yield pDSP15.

To construct the split *Cbr-unc-19(+)*/*mNeonGreen* targeting vector, the promoter and the 5' 515 bases of the *Cbr-unc-19(+)* were amplified from pDSP2 such that a loxN site was added prior to the promoter. The MCS was amplified from pDSP2 and the 5' 830 bases of the *mNeonGreen* gene were amplified from pDSP7. All three of these amplicons were assembled into NotI/SnaBI digested pDSP1 to yield pDSP16.

The entire intergenic region was used for *aha-1p*, *ahr-1p*, *ceh-20p*, *ceh-40p*, *egl-46p*, *hlh-16p* and *nhr-67p*. For *ceh-43p* the 2096bp upstream of the *ceh-43* start codon was used. For *mdl-1p*, *egl-43p* and *ceh-10p,* the promoters describe in (Reece-Hoyes et al., 2013) were used. For *daf-7p* and *lin-11p* the promoters described in (Entchev et al., 2015) and (Marri and Gupta, 2009) respectively were used. Promoters were amplified from N2 genomic DNA using primers designed to add the appropriate homology to the targeting vector and assembled into PCR linearized pDSP15 or pDSP16 for split *HygR*/*mScarlet-I* and split *Cbr-unc-19(+)*/*mNeonGreen* respectively.

Insertion of promoter libraries by TARDIS

For integration of a promoter library into a single landing pad site, a mixture consisting of 15ng/ml guide plasmid (pMS84), 20ng/µl *hsp16.41::Cas9* plasmid (pZCS36), 10ng/µl neomycin resistance plasmid (pZCS38) and 0.45fmol/µl of each of the 13 repair template plasmids (Table 1) was microinjected into the gonad arms of young adult GT332 hermaphrodites. Individuals were incubated at 20°C and after three days treated with 1.56mg/ml G-418 to select for array bearing individuals. Once stable array lines were obtained, integration was done using the standard TARDIS protocol, using a density of approximately 200 L1s per plate.

For integration of a promoter library into two landing pad site, a mixture consisting of 15ng/ml guide plasmid (pMS84), 20ng/µl *hsp16.41::Cas9* plasmid (pZCS36), 0.5ng/µl neomycin resistance plasmid (pZCS38) and 0.45fmol/µl of each of the 14 repair template plasmids (seven targeted to each site) was microinjected into the gonad arms of young adult

GT337 hermaphrodites. Individuals were incubated at 20°C and after three days treated with 1.56mg/ml G-418 to select for array bearing individuals. Once a stable array line was obtained, plates of mixed stage worms were transferred to plates without drug, heat shocked at 35.5°C for 1.5hrs and returned to 20°C. Three days after heat shock, hygromycin was added at a final concentration of 250µg/ml.

For both scenarios, candidate worms (those which had both hygromycin resistance and wild-type movement) were singled and screened by PCR. The identity of the integrated promoters was determined by Sanger sequencing of the PCR product. Primers used for genotyping are given in Supplemental Table 4.6.

Microscopy

Individual late L4/young adults were mounted on 2% agarose pads and immobilized with 0.5M levamisole. Imaging was performed on a DeltaVision Ultra microscope (Cytiva, Massachusetts, USA) using the 20x objective and Acquire Ultra software version 1.2.1. Fluorescent images were acquired using the orange (542/32nm) and green (525/48nm) filter sets for mScarlet-I and mNeonGreen respectively. Light images were captured at 5% transmission and a 0.01 second exposure. Fluorescent images were captured at 5% transmission and a 2sec (*aha-1p*), 1sec (*ceh-40p*, *ceh-43p*, *nhr-67p*, *ceh-10p::mNeonGreen*), 0.5sec (*ceh-10p::mScarlet-I*, *ceh-20p*, *daf-7p*) or 0.2sec (*lin-11p*, *mdl-1p*) exposure. Images were processed in Fiji (ImageJ) version 2.9.0/1.53t.

Accessibility of reagents, data, code, and protocols

The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, figures, and tables. Plasmids pDSP15 (Addgene ID 193853), pDSP16 (Addgene ID193854), pMS84 (Addgene ID 193852), pZCS36 (Addgene ID 193048), pZCS38 (Addgene ID193049), and pZCS41 (Addgene ID 193050), are available through Addgene and can be freely viewed and edited in ApE (Davis and Jorgensen, 2022) and other compatible programs. Strains PX740, GT332 and GT337 are available from the *Caenorhabditis* Genetics Center (cgc.umn.edu). Strains and plasmids not available at a public repository are available upon request. Illumina sequencing data are available at BioProject ID: PRJNA893002. All other data, code, plasmid and landing sequences and original microscopy images are available on Figshare (Stevenson et al., 2022). We plan to continue to develop TARDIS technology and

provided descriptions of updated libraries and advancements at: https://github.com/phillips-lab/TARDIS.

Bridge

In the next chapter, we present the patented version of TARDIS, which outlines our claims and ideas for brining large library transgenesis to other systems.

CHAPTER 5 GENETIC DATA COMPRESSION AND METHODS OF USE

CROSS-REFERENCE TO RELATED APPLICATION

This application claims priority to U.S. Provisional Application No. 63/013,365, filed April 21, 2020, which is incorporated herein by reference in its entirety.

FIELD

This disclosure relates to systems and methods for producing genetically modified cells or organisms, particularly utilizing genetic data compression to generate a plurality of genetically modified cells or organisms.

ACKNOWLEDGMENT OF GOVERNMENT SUPPORT

This invention was made with government support under grant numbers AG056436 and GM131838 awarded by the National Institutes of Health. The government has certain rights in the invention.

BACKGROUND

A fundamental tool in molecular biology is the generation of transgenic individuals having experimentally provided DNA integrated into their genomes. Technological advances in transgenesis are highly impactful, as illustrated by the role of transgenesis in the foundation of modern molecular biology and the recent changes brought about by the development of targetable nucleases and related genome engineering techniques that allow the creation of non-random, custom-designed genetic modifications. Targetable nucleases such as transposons, zinc-finger nucleases, and Cas proteins have all been widely adopted for such precision genetic engineering. As precision has increased, throughput in transgenesis and DNA synthesis has grown exponentially.

However, all transformation technologies to date, including CRISPR, are limited by the need to move nucleic acids across cell boundaries, *e.g.*, requiring a single nucleic acid transit

event per designed genomic modification. This limitation is typically overcome in microbial and cell culture systems by massively parallelizing these transit events, using ease of movement across cell membranes and availability of large population sizes, making library transgenesis possible through transformation and transfection of DNA plasmids. In non-microbial models (*e.g.*, *C. elegans, D. melanogaster*, and mammalian systems), such pre-designed, high-throughput, parallelized techniques are not available. Similar throughput in transgenesis has yet to be realized because of species-specific limitations. Broadly speaking, microinjections (or other low throughput DNA transportation) must be performed at precise stages in development to modify individual genomes. These protocols are typically done with the experimental goal of creating a single transgenic modification. Therefore, creating precision transgenics on the scale of tens to thousands from individual injections has generally yet to be realized. Thus, there is a barrier in non-microbial models to performing experiments such as lineage tracking or DNA transformation-based genetic screening that require generation of large populations of individuals with unique genetic modifications.

SUMMARY

Provided herein are genetic engineering systems that allow creation of large numbers of pre-designed, single-copy lineages simultaneously. The method is in some examples termed Transgenic Arrays Resulting in Diversity of Integrated Sequences (TARDIS). In some examples, the TARDIS system massively increases transformation throughput by supplying the organism with a "database" of DNA sequences in the form of a heritable DNA element (such as an array or artificial chromosome). The DNA element represents a "compressed" state including a plurality of DNA sequences that can be integrated at a future time point. This compression allows individual "transit events" to represent many simultaneous DNA sequences that can later be integrated individually. To integrate the large number of compressed sequences, the TARDIS system uses the heritability of the compressed DNA to enlarge the population carrying the compressed sequences. Integration is then facilitated at genetically engineered sites (*e.g.*, a synthetic landing pad), which are engineered to recombine with the elements of the compressed DNA in a defined way. The recombination event is a "decompression" step, allowing for a functional output.

Some embodiments herein provide methods of producing a plurality of genetically modified cells, which include introducing a nucleic acid molecule including a plurality of index sequences into a cell including a genomic polynucleotide (*e.g.*, an intron or exon of a gene, or a promoter element) that includes a synthetic landing pad, wherein each of the plurality of index sequences includes a first portion of a nucleotide sequence and the synthetic landing pad includes a second portion of the nucleotide sequence to produce a cell including the synthetic landing pad and the nucleic acid molecule including the plurality of index sequences. In some examples, the cell may be a eukaryotic cell (for example, a yeast cell, a mammalian cell, a *Caenorhabditis elegans* cell, or a *Drosophila* cell), or a bacterial cell. In some examples, introducing the nucleic acid molecule including the plurality of index sequences into the cell including the genomic polynucleotide that includes the synthetic landing pad includes injecting the nucleic acid molecule into an animal including the cell. The methods also include generating a plurality of progeny cells including the genomic polynucleotide that includes the synthetic landing pad and the nucleic acid molecule including the plurality of index sequences, integrating a single index sequence into the synthetic landing pad in each of the plurality of progeny cells, thereby linking the first and second portions of the nucleotide sequence, and selecting progeny cells including integrated index sequences based on presence or activity of the linked first and second portions of the nucleotide sequence, thereby producing a plurality of genetically modified cells. In some examples, the lineage of the cell is traced by detecting an index sequence in progeny of at least one of the plurality of genetically modified cells.

In some embodiments, the nucleic acid molecule including the plurality of index sequences is an extrachromosomal array, a plasmid, or an artificial chromosome. In some examples, the nucleic acid molecule including the plurality of index sequences includes about 500-3,000 index sequences. In particular examples, each of the plurality of index sequences includes a homologous fragment of the genomic polynucleotide, wherein each of the plurality of index sequences are different. In these and other examples, the first portion and the second portion of the nucleotide sequence may reconstitute a functional gene (for example, a selectable marker or reporter gene) when linked. In some examples, each of the plurality of index sequences includes a sequence variant of a reference coding sequence, a reference non-coding sequence, a library sequence, a randomized sequence, or a promoter element. In particular examples, the method further includes selecting a single sequence variant of the reference coding

sequence by selecting a genetically modified cell including the reference coding sequence variant, selecting a single sequence variant of the reference non-coding sequence by selecting a genetically modified cell including the reference non-coding sequence variant, selecting a single library sequence by selecting a genetically modified cell including the library sequence, selecting a single randomized sequence by selecting a genetically modified cell including the randomized sequence, or selecting a single promoter element by selecting a genetically modified cell including a screenable marker or reporter gene operably linked to the promoter element in the genomic polynucleotide.

In some embodiments, the synthetic landing pad further includes a site-specific nuclease (SSN) recognition site and homology arms flanking the SSN recognition site, and each of the plurality of index sequences is flanked by the homology arms in the nucleic acid molecule including the plurality of index sequences. In particular examples, each of the homology arms is about 150-500 nucleotides in length. In some examples, the SSN is a caspase (for example, Cas9), zinc-finger nuclease, or TALEN.

In some embodiments, the method further includes selecting a genetically modified cell including an index sequence by an assay phenotype, or by expression of a selectable marker or reporter, and generating variants of the index sequence. In such embodiments, the method further includes introducing a nucleic acid molecule including the variants of the index sequence into a cell including a genomic polynucleotide comprising a synthetic landing pad, wherein each of the variants of the index sequence includes a first portion of a nucleotide sequence and the synthetic landing pad includes a second portion of the nucleotide sequence to produce a cell including the synthetic landing pad and the nucleic acid molecule including the variants of the index sequence. Such methods further include generating a plurality of progeny cells including the genomic polynucleotide including the synthetic landing pad and the nucleic acid molecule including the variants of the index sequence, integrating a single variant of the index sequence into the synthetic landing pad in each of the plurality of progeny cells, thereby linking the first and second portions of the nucleotide sequence, and selecting progeny cells including integrated variants of the index sequence based on presence or activity of the linked first and second portions of the nucleotide sequence.

Further embodiments herein provide genetically modified cells including an extrachromosomal array including a plurality of index sequences, and a genomic polynucleotide including one of the plurality of index sequences integrated at a synthetic landing pad, wherein the integrated index sequence includes a first portion of a nucleotide sequence and the synthetic landing pad includes a second portion of the nucleotide sequence, and wherein the first and second portions of the nucleotide sequence are operably linked in the genomic polynucleotide. Particular embodiments provide a multicellular organism including such a plurality of genetically modified cells, wherein the genetically modified cells include different index sequences.

The foregoing and other features of the disclosed subject matter will become more apparent from the following detailed description, which proceeds with reference to the accompanying figures.

BRIEF DESCRIPTION OF THE DRAWINGS

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

FIG. 5.1 is an overview of *in situ* assembly. Amplification of homology arms and cargo fragments by PCR with overlaps of about 30 bp (FIG. 5.1A). Optional complexing by a second round of PCR reduces the number of fragments and increases the frequency of correct integration (FIG. 5.1B). Upon microinjection, PCR products are recombined by the organism (*e.g.*, a worm) using microhomology (FIG. 5.1C) to make the complete donor homology ready for integration (FIG. 5.1D). FIG. 5.2 is a schematic diagram of an embodiment of the technology showing production of *C. elegans* including an array (*e.g.*, an artificial chromosome or *in situ* array) of gene fragments in the F1 generation and resulting independent diversity in progeny organisms (top). A single worm is injected with a plurality of sequences, which do not integrate immediately, but produce the array which is inherited to later generations. This population is expanded (advanced generations (A.G.)), and then CRISPR is induced when desired. Once induced, each individual worm takes a random individual sequence. For example, if 1,000 lines are desired, 1,000 unique sequences are injected to form the array. The population is then expanded until it is somewhere significantly over 1,000, and then CRISPR is induced. Each

FIG. 1

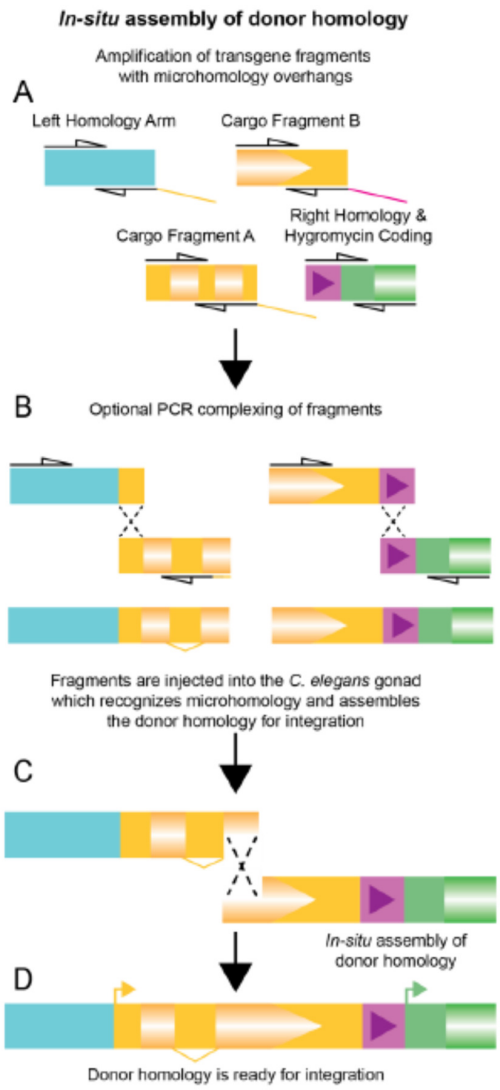


Figure 5.1 In-situ assembly of donor homology

individual worm in the population will try and integrate only 1 of the 1,000 sequences. At the genomic level (bottom), the *C. elegans* genome includes a "broken gene" or transcription control element (*e.g.*, a promoter), which may be a target for creating variation or a selectable marker with a cargo. This "synthetic landing pad" includes a Cas9 recognition site, such that CRISPR creates a break in the genome and unique sequences from the array of gene fragments are integrated in each progeny worm.

FIGS. 5.3A and 5.3B show exemplary schematics of index directed modulation of the genome. FIG. 3A shows a genomic synthetic landing pad designed to accept indexes via homology directed repair. The index carries a portion of a selectable DNA sequence of a Hyg Resistance (*HygR*) cassette. Upon integration, the *HygR* gene (or other selectable sequence) is reconstituted. FIG. 3B shows that the injected mix of indexes generates complex DNA arrays in the animal via the TARDIS technology. The array acts as a "data compression" technology, with 1,000+ indexes being stitched together to form a single selectable array (*e.g.*, neomycin selectable).

FIG. 5.4 illustrates index directed "directional" genomic screening. Two versions of the synthetic landing pads enable controlled, directional regulation of distant genomic loci. In Version A (top), after insertion, the index becomes the targeting sequence in a CRISPR gRNA-tracrRNA hybrid RNA molecule that is tethered to a transcriptional activator allowing positive regulation of genes with the appropriate indexed sequence. In Version B (bottom), the index is embedded in a bidirectionally transcribed location. The resultant dsRNA then feeds into the RNAi pathway to negatively regulate distant genes that feature the indexed sequence.

FIG. 5.5 is a schematic diagram illustrating an embodiment method for detecting protein-protein interactions. For example, the landing pad (*e.g.*, "TARDIS insertion site") includes gene regulatory sequences and the coding sequence for a peptide that can be used as a readout for protein-protein interactions. Upon integration of a protein coding sequence (prey) from the array, the landing pad site produces a hybrid protein that contains both the prey and the peptide used to detect protein-protein interactions. In addition, a second protein sequence is expressed (from the same or a different genomic location) that is hybrid for the test sequence used to test for protein-protein interactions (bait) and a second peptide sequence with functional relevance.

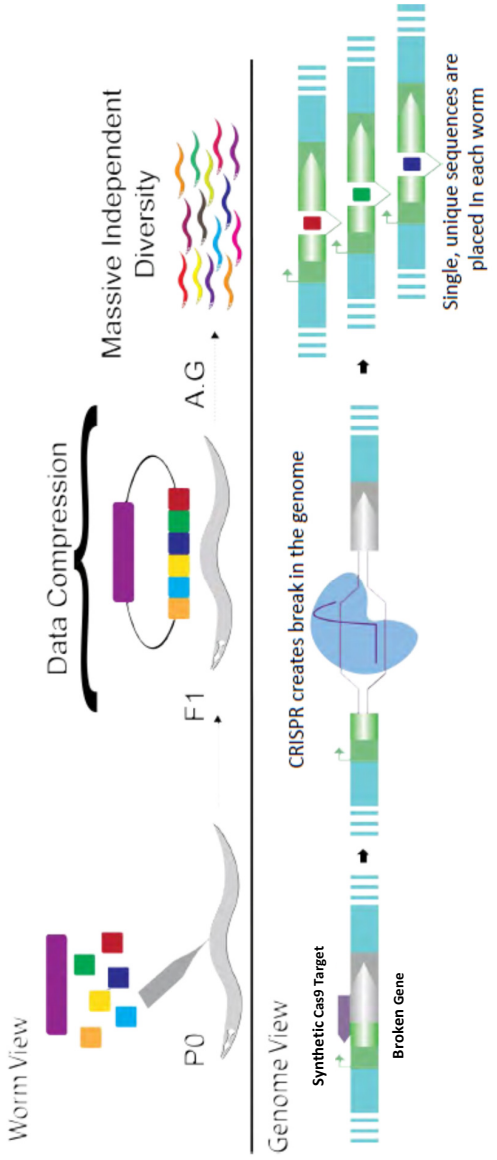


Figure 5.2 Overview of TARDIS integration

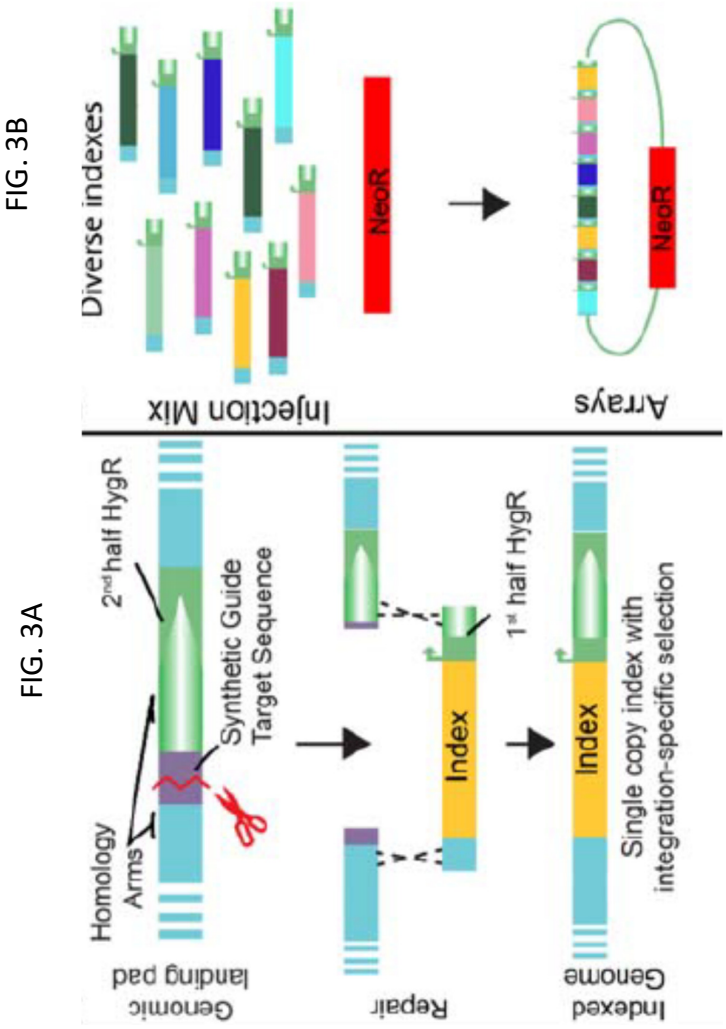


Figure 5.3 Index integration and decompression

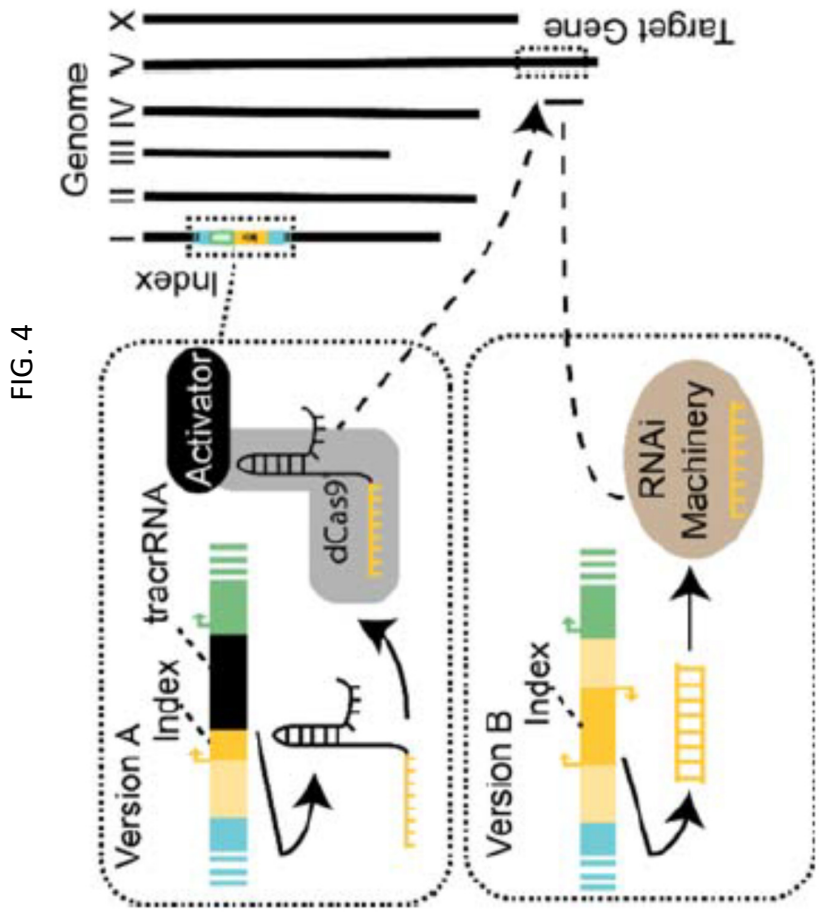


Figure 5.4 Index-directed genome screening

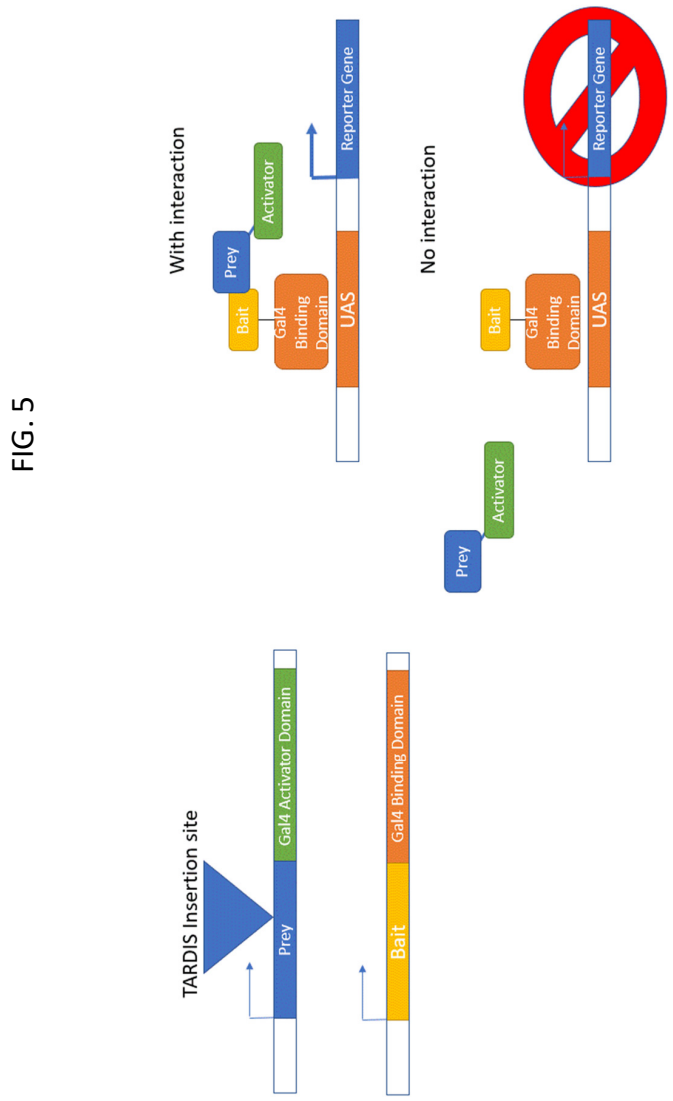


Figure 5.5 Detecting protein-protein interactions

If the bait and prey interact, then the two functional peptides are brought into contact and result in a functional readout.

FIG. 5.6 includes a diagram showing the creation and marking of individuals utilizing a TARDIS landing pad design. In this example, an organism's genome contains a landing pad characterized by a genomic polynucleotide (positive sense strand DNA (SEQ ID NO: 10); negative sense strand DNA (SEQ ID NO: 11); translation (SEQ ID NO: 12)) including a Cas9 recognition site ("synthetic Cas9 target") followed by a 3' fragment of a reporter gene (*mScarlet*) (left). A Cas9 induces a DNA break at the recognition site (positive sense strand DNA (SEQ ID NO: 13); negative sense strand DNA (SEQ ID NO: 14); translation (SEQ ID NO: 15)), which is then repaired by homology directed repair (HDR) to integrate a single, random index sequence (*e.g.*, a random barcode) and the 5' fragment of the reporter gene from an array of index sequences (middle). Integration of the index sequence (positive sense strand DNA (SEQ ID NO: 16); negative sense strand DNA (SEQ ID NO: 17); translation (SEQ ID NO: 18)) therefore reconstitutes the functional *mScarlet* reporter, allowing selection of an individual with an integrated index sequence (right). By selecting a plurality of cells descended from the cell containing the array, a library of individual cells that each contain a single index sequence from the array is generated.

FIG. 5.7 illustrates an exemplary method for detecting individuals using single-cell sequencing of individuals marked on the mRNA level, as compared to existing methods for single-cell sequencing, in a process of studying aging. In existing methods (top), the lineage of particular cells from transformed worms is not able to be identified and tracked. In such methods, it is impossible to discern from a population of cells descended from transformants which cells in different stages of growth are descended from particular transformants, because there is no mRNA "barcode" identifying individual transformants. According to methods disclosed herein (middle), an mRNA barcode corresponding to particular index sequences identifies individual transformed cells, which can then be used to identify their progeny cells at different stages of growth. The "regulated aging" hypothesis postulates that the lineages of different barcoded individuals will be directed to particular cell types during aging (bottom left). On the other hand, the "dysregulated aging" null hypothesis is that particular cell types during aging will contain cells descended from random individual transformants (bottom right). Using

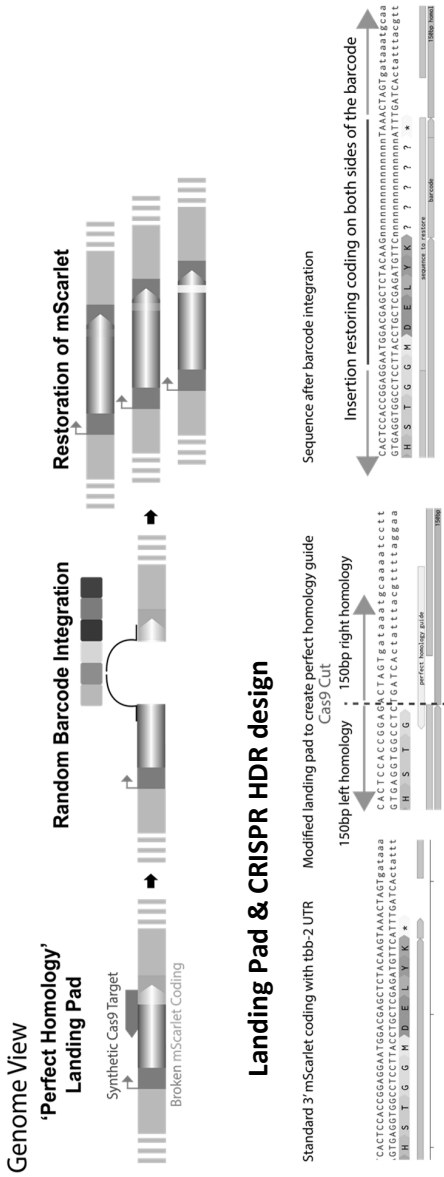


Figure 5.6 Barcode Integration

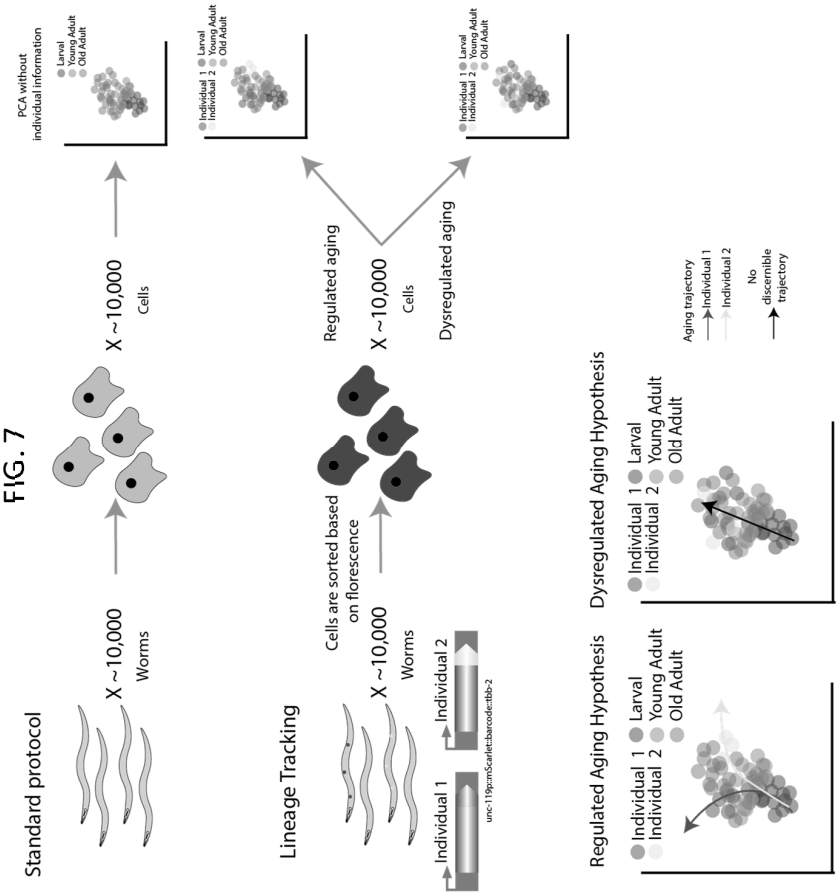


Figure 5.7 Barcoded single cell sequencing

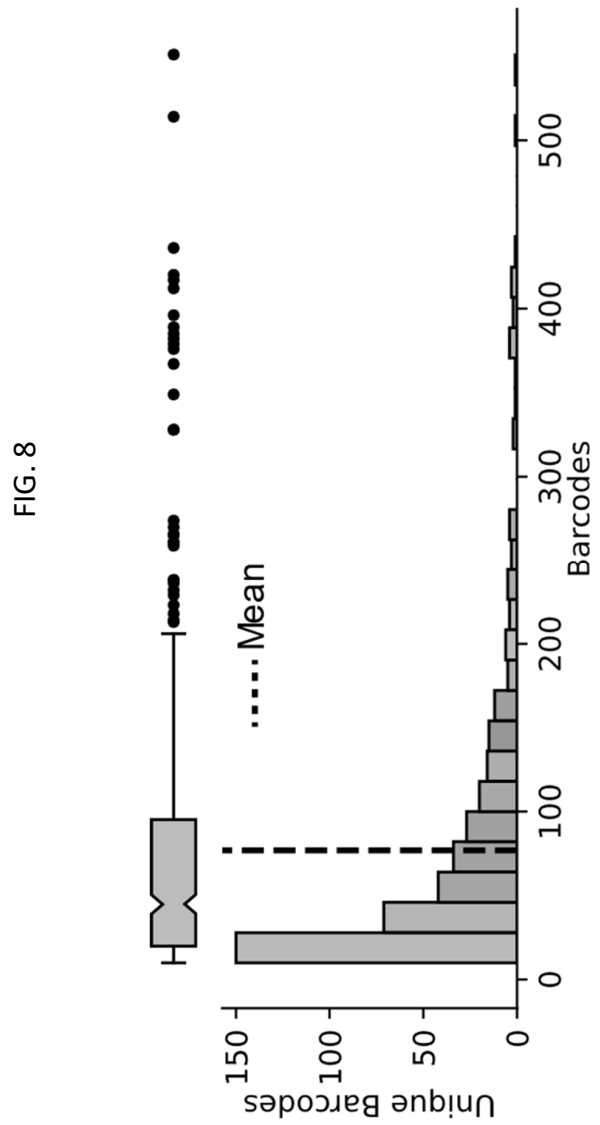


Figure 5.8 TARDIS array diversity

FIG. 5.8 is a histogram showing the distributions of reads/counts from PX742 sequencing. The X axis is the number of counts a unique index sequence had, and the Y-axis is the number of unique index sequences with that count. the methods provided herein, it is possible to test the regulated aging hypothesis in complex multicellular organisms.

FIG. 5.9 is a histogram showing the number of unique index sequences ("barcodes") as a function of the change in frequency observed from the PX742 array to integrated population.

FIG. 5.10 shows line charts of lineage frequency as a function of transfer number observed in three independent replicates.

FIG. 5.11 illustrates histograms showing the change in frequency across unique lineages from the PX742 array to the integrated population observed in three independent replicates. Marked with arrows are the four lineages which increased the most from the P0.

SEQUENCE LISTING

Any nucleic acid and amino acid sequences listed herein or in the accompanying Sequence Listing are shown using standard letter abbreviations for nucleotide bases and amino acids, as defined in 37 C.F.R. § 1.822. In at least some cases, only one strand of each nucleic acid sequence is shown, but the complementary strand is understood as included by any reference to the displayed strand.

The Sequence Listing is submitted as an ASCII text file in the form of the file named Sequence_Listing.txt, which was created on April 12, 2021, and is 9,000 bytes, which is incorporated by reference herein.

SEQ ID NO: 1 is the nucleic acid sequence of Ultramer ZCS133, a long ssDNA oligo which serves as a template for PCR amplification. PCR amplification produces a large diverse double-stranded DNA mixture with random sequences in the place of the Ns, for example for generating barcodes for barcoded lineage tracking.

SEQ ID NO: 2 is the nucleic acid sequence of ZCS134, a forward primer used to amplify off of the template ZCS133. It also provides additional sequence to add 35 bases of homology to the genome on the left hand side, while providing the intron/exon boundary for restoration of the Hygromycin Gene.

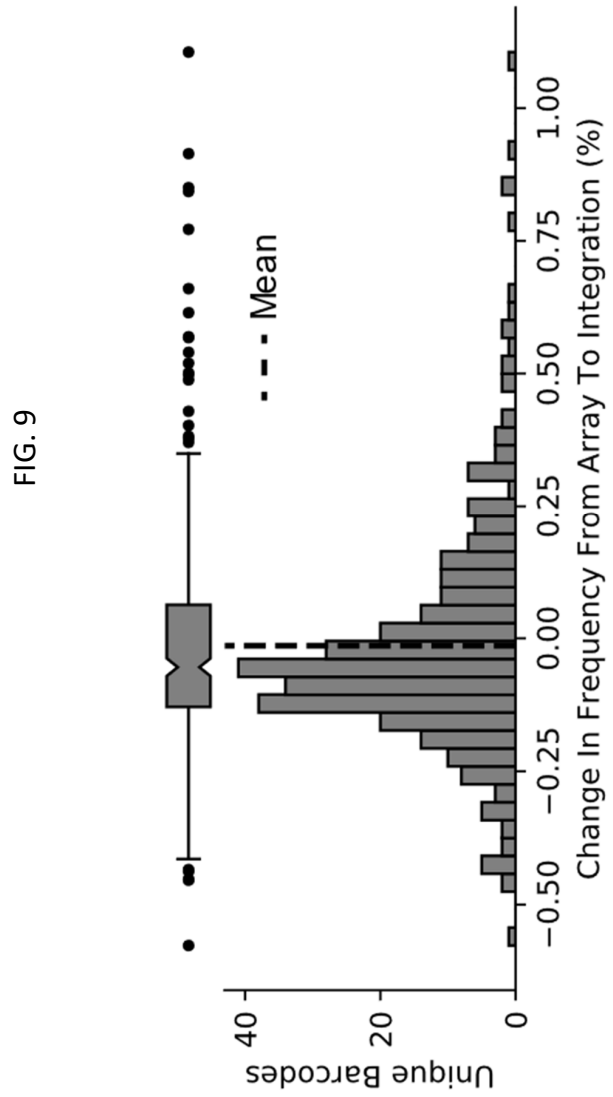


Figure 5.9 TARDIS array integration to F1

FIG. 10

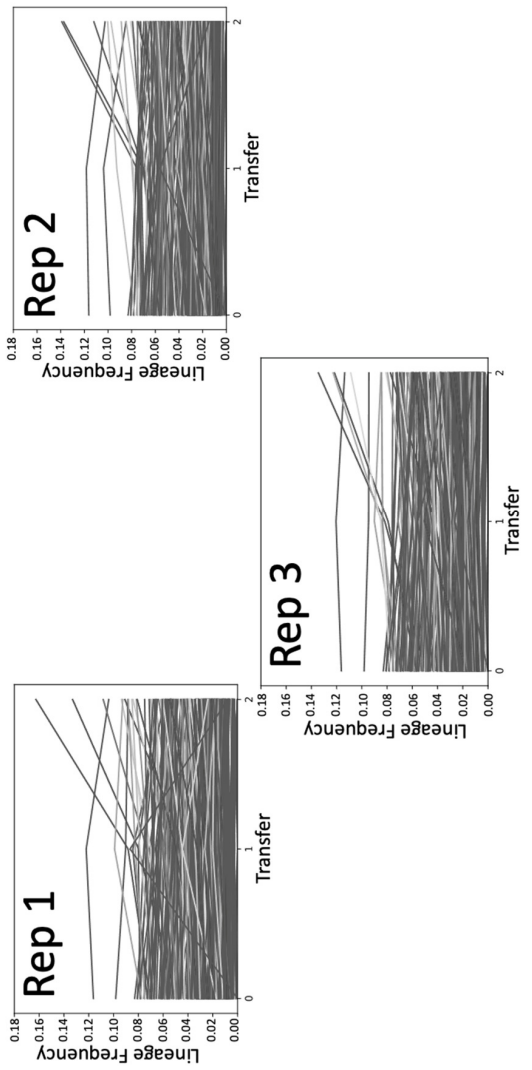


Figure 5.10 Barcoded lineage tracking

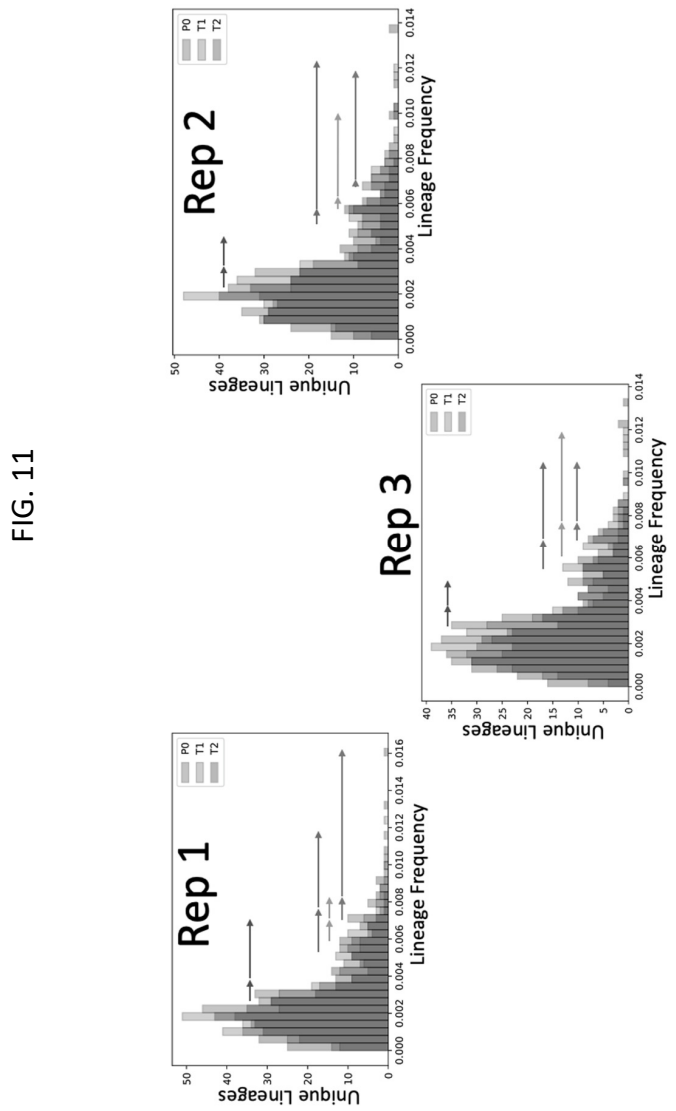


Figure 5.11 Change in frequency of lineages

FIG. 12

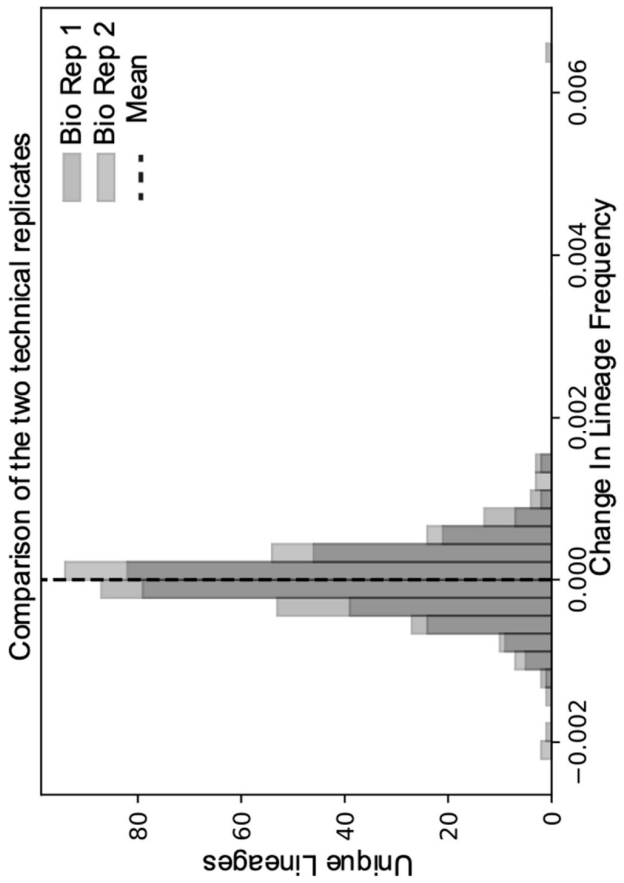


Figure 5.12 Technical replication of amplicon sequencing

FIG. 5.12 shows a lineage histogram showing the change in frequency across unique lineages for two independent PCR replicates, compared for a difference in frequency

SEQ ID NO: 3 is the nucleic acid sequence of ZCS135, a reverse primer used to amplify off of the template ZCS133. Also provides additional sequence to add 35 bases of homology to the genome on the right hand side, while providing the intron/exon boundary for restoration of the Hygromycin Gene.

SEQ ID NO: 4 is the nucleic acid sequence of ZCS273, a forward primer used to amplify off of a standard Hygromycin template to provide 500 bases of homology to the genome on the left hand side.

SEQ ID NO: 5 is the nucleic acid sequence of ZCS275, a reverse primer used to amplify off of a standard Hygromycin template to provide 500 bases of homology to the genome on the left hand side.

SEQ ID NO: 6 is the nucleic acid sequence of ZCS276, a forward primer used to amplify off of a standard Hygromycin template to provide 500 bases of homology to the genome on the right hand side.

SEQ ID NO: 7 is the nucleic acid sequence of ZCS278, a reverse primer used to amplify off of a standard Hygromycin template to provide 500 bases of homology to the genome on the right hand side.

SEQ ID NO: 8 is the nucleic acid sequence of pZCS41, an exemplary U6::guideRNA plasmid for targeting the TARDIS SLP for barcoded lineage tracking.

aacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgctttttgcacaacatgggggatcatgtaa
ctcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggca
acaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagtt
gcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtat
cattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacg
aaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgattt
aaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccact
gagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaacc

accgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagatac
caaatactgttcttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctg
ttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggt
cgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagct
atgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgca
cgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatg
ctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacat
gttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgacc
gagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcatt
aatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagctcactcattagg
caccccaggctttacactttatgcttccggctcgtatgttgtgtggaattgtgagcggataacaatttcacacaggaaacagctatga
ccatgattacgccaagcttgcatgcctgcaggtcgactctagaggatcccattatacatagttgataattcactggccgtcgttttaca
acgtcgtgactgggaaaaccaaaaaaaaactagcaataaaggaataaaaaactgtacaccttaaaggcgcacactctgttttgcaa
attttatttttagttgtgaattttctgctgagacctgaaaatagcaactttagtactactataatttgtcaacctttcaaaaaaagcatgca
atttttgagaaactcttataaaagctattattaaaaaaacacctttttttccaaaattattccacaaaaaatatgttatgaaatgcctacacc
ctctcacacacactctttatactactctgtcaaactcacgagatgtctgccgcctcttgtgttgcccctatataaacacctcctattgcg
agatgtcttGGCGAAGTGACGGTAGACCGTGTTTTAGAGCTAGAAATAGCAAGTTA
AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTT
TTGTGAAATTTctggcgtaatagcgaagaggcccgcaccgatcgcccttcccaacagttgcgcagcctgaatggcga
atcgggtaccgagctcgaattcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttg
cagcacatccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttcccaacagttgcgcagcctgaatg
gcgaatggcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatggtgcactctcagtacaatctgctct
gatgccgcatagttaagccagccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgc
ttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaaggg
cctcgtgatacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcggggaaatgtgcgcgg
aacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaa
ggaagagtatgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgtttttgctcacccagaaacgctg
gtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgaga
gttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggca
agagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggca
tgacagtaagagaattatgcagtgctgccataaccatgagtgat

SEQ ID NO: 9 is the nucleotide sequence of a synthetic SSN recognition site.

GCGAAGTGACGGTAGACCGT

DETAILED DESCRIPTION

Unless otherwise noted, technical terms are used according to conventional usage. Definitions of common terms in molecular biology may be found in *Lewin's Genes X*, ed. Krebs *et al.*, Jones and Bartlett Publishers, 2009 (ISBN 0763766321); Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Publishers, 1994 (ISBN 0632021829); Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by Wiley, John & Sons, Inc., 1995 (ISBN 0471186341); and George P. Rédei, E*ncyclopedic Dictionary of Genetics, Genomics, Proteomics and Informatics,* 3rd Edition, Springer, 2008 (ISBN: 1402067534), and other similar references.

Unless otherwise explained, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. The singular terms "a," "an," and "the" include plural referents unless the context clearly indicates otherwise. "Comprising A or B" means including A, or B, or A and B. It is further to be understood that all base sizes or amino acid sizes, and all molecular weight or molecular mass values, given for nucleic acids or polypeptides are approximate, and are provided for description.

Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present disclosure, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety, as are the GenBank Accession numbers (for the sequences present in GenBank on March 26, 2020). In case of conflict, the present specification, including explanations of terms, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Overview of Several Embodiments

Provided herein are compressed nucleic acid sequences (*e.g.*, arrays) and target genomic loci (*e.g.*, landing pads) that can be utilized for genomic analysis and manipulation, among other uses. This system, referred to in some embodiments as Transgenic Arrays Resulting in Diversity of Integrated Sequences ("TARDIS") is applicable in any biological system in which genetically diverse, heritable DNA elements can be generated. When utilized in animal systems, including bacteria, yeast, *C. elegans*, fish, and mammalian systems, TARDIS provides a library-based transgenic approach that rivals the throughput of microbial and cell culture methods. TARDIS facilitates high throughput transgenesis in animals by using two engineered components, a heritable DNA array that carries a library of index sequences for integration, and a genomic landing pad that facilitates integration of single sequence units from the library. An advantage of the disclosed TARDIS data compression system is the ability to generate large numbers of novel genetic variants without the need to create and integrate individual sequences one at a time, as in traditional transgenesis systems.

The features of the disclosed systems include "compressed" DNA sequences. For example, in some embodiments applying TARDIS to *Caenorhabditis elegans*, TARDIS utilizes artificial chromosomes, or arrays, as a "compressed" nucleic acid molecular library comparable to synthetic libraries used in microbial systems. *C. elegans* arrays are known to be large in size, and generally are 1-2 MB (Woglar *et al*. (2020) *PLOS Biology* 18(8), e3000817). Unlike plasmids, artificial chromosomes do not require specific sequences to form and replicate, which allows for experimental flexibility in the composition. Arrays can be inherited, but do not follow typical Mendelian inheritance. Because experimental composition is highly flexible, selectable genes can be added to the array to select for only progeny that inherit the array.

The second component of the TARDIS integration system relies on a pre-integrated "landing pad" sequence. Landing pads are engineered locations in the genome, which facilitate future integrations of single sequence units from the heritable library/array. In some embodiments, the landing pad contains components needed to express the "compressed" sequence. These features, combined with a library of compressed sequences (*e.g.*, within an artificial chromosome or within a native chromosomal location) allows for experimental diversity not previously possible with other transgenic methodologies.

In some embodiments, the compressed index sequences include a gene or promoter fragment that is not functional until combined with a second fragment in the landing pad by integration of the index sequence. For example, an index sequence and a landing pad may each comprise half of a split selectable gene, for example, a gene that encodes antibiotic (*e.g.*, Hygromycin B) resistance. The landing pad can generally be targetable from any nuclease or recombinase that either creates a double-strand break, or recombines synthetic sequence into the genome at a recognition site. In some embodiments, the landing pad is targeted by Cas9.

In some embodiments, provided herein are methods of producing a plurality of genetically modified cells, which include introducing a nucleic acid molecule comprising a plurality of index sequences into a cell comprising a synthetic landing pad, wherein each of the plurality of index sequences comprises a first portion of a sequence and the synthetic landing pad comprises a second portion of the sequence, to produce a cell comprising the synthetic landing pad and the nucleic acid molecule comprising the plurality of index sequences. The method also includes generating a plurality of cells comprising the synthetic landing pad and the nucleic acid molecule comprising the plurality of index sequences (for example by allowing the cells to proliferate or expanding the cell population), integrating one of the plurality of index sequences into the synthetic landing pad in each of the plurality of cells, thereby linking the first and second portions of the sequence, and selecting cells comprising the integrated index sequence based on presence or activity of the linked first and second portions of the sequence, thereby producing a plurality of genetically modified cells.

In particular examples, the sequence is a non-functional gene and the first portion and the second portion reconstitute a functional gene when linked. The non-functional gene can be any gene that provides a detectable readout. In some examples, the non-functional gene is an antibiotic resistance gene (*e.g.*, hygromycin resistance), such that when one of the plurality of index sequences is integrated into the landing pad, the cell expresses the antibiotic resistance gene and cells with a correct integration can be identified or selected based on antibiotic resistance. In other examples, the non-functional gene is a reporter gene, such as a fluorescent protein (*e.g.*, a green fluorescent protein, a red fluorescent protein, or a cyan fluorescent protein, including but not limited to mScarlet or GFPnovo2), such that when one of the plurality of index sequences is integrated into the landing pad, the cell expresses the reporter protein and cells with

a correct integration can be identified or selected based on detection of fluorescence. Additional reporters can include LacZ or CAT. In other examples, one portion of the non-functional gene may include a regulatory element (such as a promoter) and the other portion of the non-functional gene may be a coding sequence or portion thereof. Other examples are described in the particular embodiments below.

The plurality of index sequences can be any nucleic acid sequence of interest. In some embodiments, the index sequences are from about 1-15,000 nucleotides in length (*e.g.*, about 1-50 nucleotides, about 10-100 nucleotides, about 25-250 nucleotides, about 50-500 nucleotides, about 200-1000 nucleotides, about 500-2500 nucleotides, about 1000-5000 nucleotides, about 2500-10,000 nucleotides, or about 5000-15,000 nucleotides long). In other examples, the index sequences are greater than 10,000 nucleotides, such as greater than 15,000 nucleotides, greater than 20,000 nucleotides, greater than 25,000 nucleotides, greater than 50,000 nucleotides, greater than 75,000 nucleotides, or greater than 100,000 nucleotides. In some examples, the index sequences are gene coding sequences or portions and/or fragments thereof (such as a gene library), non-coding sequences, promoter elements or portions and/or fragments thereof, amplicon products, randomized genomic sequences, or barcodes. The synthetic landing pads may include gene regulatory elements or portions and/or fragments thereof, a reporter sequence (such as a fluorescent protein or antibiotic resistance gene), or portions of a gene coding sequence. Specific examples of index sequences and synthetic landing pads are discussed in further detail below.

In some embodiments, site-specific integration of an index sequence is accomplished by utilizing a site-specific nuclease (SSN) that recognizing and binding to particular nucleotide sequences, for example, in the genome of a host organism. A DNA sequence that is recognized by the SSN may be referred to as a SSN recognition site. Polypeptide domains that are capable of recognizing and binding to DNA in a site-specific manner generally fold correctly and function independently to bind DNA in a site-specific manner, even when expressed in a polypeptide other than the protein from which the domain was originally isolated. Similarly, SSN recognition sites for recognition and binding by DNA-binding polypeptides are generally able to be recognized and bound by such polypeptides, even when present in large DNA

structures (*e.g.*, a chromosome), particularly when the site where the target sequence is located is one known to be accessible to soluble cellular proteins (*e.g.*, a gene).

While DNA-binding polypeptides identified from proteins that exist in nature typically bind to a discrete nucleotide sequence or motif (*e.g.*, a consensus recognition sequence), methods exist and are known in the art for modifying many such DNA-binding polypeptides to recognize a different site. DNA-binding polypeptides include, for example and without limitation: zinc finger DNA-binding domains, leucine zippers, UPA DNA-binding domains, GAL4, TAL, LexA, RNA-guided CRISPR-Cas9, Tet repressors, LacR, and steroid hormone receptors.

In some embodiments, methods provided herein utilize a SSN that includes a DNA-binding polypeptide to recognize and bind a recognition site for the SSN, create a double-strand break targeted to a synthetic landing pad, resulting in integration of an index sequence into the synthetic landing pad by homology-directed repair (HDR). In particular embodiments, the SSN comprises a DNA-binding polypeptide selected from the group consisting of a zinc finger, TAL, and RNA-guided CRISPR/Cas-derived DNA-binding polypeptide. In particular examples, the SSN is a CRISPR endonuclease, such as Cas9, xCas9, SpRYCas9, or Cas12a. In other examples, the endonuclease is a Zinc finger nuclease (ZFN) or a TAL Effector nuclease (TALEN). In still further examples, a transposon-based method such as mos1-mediated single copy insertion (mosSCI,) or Tol2 is used, or a very rare cutting endonuclease is used, for which a cutting site could be engineered into the landing pad (*e.g.*, srfI in *C. elegans*). The double strand break is then repaired with one of the plurality of index sequences by homology directed recombination. In some examples, the synthetic landing pad and each of the plurality of index sequences further comprise flanking homology arms which are the same (*e.g.*, the 5' ("left") homology arm in the landing pad and the index sequences are the same and the 3' ("right") homology arm in the landing pad and the index sequence are the same). In particular examples, a homology arm is about 100-1,000 nucleotides in length (*e.g.*, about 150-750 nucleotides, about 150-600 nucleotides, about 150-500 nucleotides, about 250-600 nucleotides, or about 500 nucleotides long).

In specific examples, the methods utilize Cas9, and the synthetic landing pad includes a Cas9 guide RNA and flanking homology arms and each of the plurality of index sequences

further comprise the same flanking homology arms. In some examples, the cell is transgenic for Cas9, or a recombinant Cas9 protein is introduced into the cell.

The disclosed methods can be performed in a eukaryotic cell (such as a yeast, *Caenorhabditis elegans*, or mammalian cell) or a bacterial cell (such as a *E. coli* cell). In some examples, the methods are carried out *in vitro*. In other examples, the methods are carried out *in vivo*, for example in *C. elegans*. Thus, in some examples, an organism is transgenic for the synthetic landing pad and is administered the plurality of index sequences. In one non-limiting example, the plurality of index sequences is administered to the germline of *C. elegans* (see, *e.g.*, FIG. 2).

In some examples, the plurality of index sequences that are introduced into the cell are in an extrachromosomal array of nucleic acids, a plasmid, or an artificial chromosome (such as a bacterial artificial chromosome or a yeast artificial chromosome).

In particular examples, the plurality of index sequences are in an extrachromosomal array. In one example, an extrachromosomal array of nucleic acid molecules is produced in *C. elegans* by injecting a plurality of nucleic acid molecules into the gonad arm of the *C. elegans*, thereby forming the extrachromosomal array of nucleic acids comprising the nucleic acid molecules. In some examples, the extrachromosomal array includes about 1-65,000 index sequences, for example, about 1-1000 about 10-100, about 50-250, about 200-600, about 500-1000, about 750-1500, about 500-3000, about 1000-5000, about 2500-10,000, about 5000-15,000, about 10,000-25,000, about 30,000-50,000, or about 40,000 to 65,000 index sequences. In one non-limiting example, the array includes about 600 index sequences. The size of the extrachromosomal array is in some examples about 1 kb to about 2 Mb, for example, about 1-500 kb, about 100-200 kb, about 250-750 kb, about 500-1000 kb, about 750-1500 kb, or about 1000-2000 kb.

Exemplary Method Embodiments

The disclosed methods can be utilized in a variety of methods for analysis of gene expression, function, and evolution. Exemplary embodiments are provided below. While at least some of the embodiments discussed below refer to CRISPR/Cas9 methods, it should be understood that other CRISPR systems or other methods for integration of index sequences (such as those discussed above) can also be utilized in these methods.

Integration of gene libraries

In this embodiment, the index sequences (*e.g.*, a disclosed array) include gene libraries, which may be native or non-native to the cell or organism utilized. The landing pad includes regulatory elements for genes within the library. Upon recombination, the single genes within the library integrate and are expressed uniformly across independent strains. Thus, large gene libraries can be compared within a single experiment. In some embodiments, the first portion of the sequence in each of the plurality of index sequences is an element from a gene library, and the second portion of the sequence in the synthetic landing pad is a regulatory sequence for the elements from the gene library, wherein each of the plurality of genetically modified cells comprises a single element from the gene library.

Identification of gene expression patterns

In this embodiment, the index sequences (*e.g.*, a disclosed array) include a library of promoter elements, and the landing pad includes an inactive reporter protein coding sequence. Integration of a single promoter element restores the reporter protein coding sequence, showing the expression pattern of the gene from which the promoter element is taken. If the library contains ~1000 variants, integration of each promoter for a genome of ~20,000 genes would reduce the required labor by ~1000-fold compared to current individual analysis methods.

In some embodiments of the method, the first portion of the sequence in each of the plurality of index sequences is a different promoter element, and the second portion of the sequence in the synthetic landing pad encodes a reporter. In some examples the reporter encodes a fluorescent protein (*e.g.*, a green fluorescent protein, a red fluorescent protein, or a cyan fluorescent protein, including but not limited to mScarlet or GFPnovo2). Additional reporters include *LacZ* or *CAT*. In particular examples, each of the plurality of genetically modified cells comprises a single promoter element linked to the reporter sequence and the reporter. The method further includes detecting a signal from the reporter, such as detecting fluorescence if the reporter is a fluorescent protein. The signal from the reporter can provide qualitative and/or quantitative information regarding the expression of the gene from which the promoter element is taken, including but not limited to spatial and/or temporal expression patterns.

Evolution of novel proteins

In this embodiment, the index sequences (*e.g.*, a disclosed array) are formed from amplicon products. Several organisms (*e.g.* yeast and *C. elegans*) are capable of recombining amplicons in a directed way to result in a functional gene or portion thereof. These variable amplicon products are recombined in the array (*e.g.*, artificial TARDIS chromosome) to produce several combinations. The landing pad includes elements which are not varied, and depend on the protein that is being evolved. Once simultaneously integrated, variants are selected based on an assay phenotype.

In some embodiments of the method, each of the plurality of index sequences comprises a first portion of a reference sequence of interest (*e.g.*, a coding or non-coding sequence) comprising at least one variation from the reference sequence, and the synthetic landing pad comprises a second portion of the reference sequence of interest, wherein each of the plurality of genetically modified cells comprises a reconstituted sequence of interest comprising the at least one variation. In particular examples, the reference sequence is a native form of the sequence of interest. In some examples, the plurality of variants of the sequence of interest is generated by mutagenic PCR.

The method may also further include detecting an assay phenotype resulting from the reconstituted variant sequence and/or selecting a variant of interest based on the assay phenotype. Exemplary assay phenotypes include, but are not limited to membrane transport activity (for example for new molecular ligand activity), neuronal longevity (for example, for a neuropeptide variant), or enzyme activity (for example, temperature sensitivity). An appropriate assay can be selected based on the sequence being evolved.

Down-regulation of genes

In this embodiment, the index sequences (*e.g.*, a disclosed array) include gene promoter fragments or fragments of gene coding sequence. In some embodiments, the array includes promoter fragments, and the landing pad includes an inactive CRISPR guide RNA, which becomes active upon recombination with the promoter sequence and targets an inactive version of Cas9 to the native gene promoter, decreasing or inhibiting gene expression. Thus, in one example, the first portion of the sequence in each of the plurality of index sequences is a different promoter element, and the second portion of the sequence in the synthetic landing pad

further includes a CRISPR guide RNA. The cell also includes a catalytically inactive Cas9 (*e.g.*, the cell is transgenic for a *Cas9*, such as *dCas9*). Expression of the reconstituted sequence is down-regulated upon activation of Cas9.

In other embodiments, the index sequences (*e.g.*, a disclosed array) includes gene coding sequences and the landing pad includes a mechanism to produce an RNAi response to decrease gene expression. In some examples, the landing pad includes dual and opposite orienting promoter elements to produce an RNAi response, thereby decreasing or inhibiting gene expression. In one example, the first portion of the sequence in each of the plurality of index sequences comprises a different gene coding sequence or portion thereof, and the second portion of the sequence in the synthetic landing pad comprises dual and opposing promoter elements. Expression of the reconstituted sequence is then down-regulated by an RNA interference process. In other examples, the landing pad includes a constant region, which would be transcribed into the mRNA. That constant region is targeted by a constitutive RNAi process, which is seeded by dsRNA generated from another location in the genome or experimentally introduced into the animal/cells, thus down-regulating the reconstituted sequence by RNAi.

In some examples, expression (*e.g.*, amount) of the sequence is detected by observation the amount of a reporter, such as a fluorescent protein. In another example, tissue samples are obtained and the quantity of the protein (*e.g.*, if the protein has a tag, such as a His tag) or mRNA is measured. Down-regulation of expression can be determined by comparison of expression of the sequence compared to a control.

Up-regulation of genes

In this embodiment, the index sequences (*e.g.*, a disclosed array) includes gene promoter fragments and the landing pad includes an inactive CRISPR guide RNA, which becomes active upon recombination with the promoter sequence. This targets an inactive *Cas9* coupled to a transcription factor to the native promoter and increases gene expression. Thus, in some examples of the methods, the first portion of the sequence in each of the plurality of index sequences is a different promoter element, and the second portion of the sequence in the synthetic landing pad further comprises a CRISPR guide RNA. The cell further also includes a catalytically inactive *Cas9* linked to a transcription activator, wherein expression of the reconstituted sequence is up-regulated upon activation of *Cas9*.

In some examples, expression of the sequence is detected by observation the amount of a reporter, such as a fluorescent protein. In another example, tissue samples are obtained and the quantity of the protein (*e.g.*, if the protein has a tag, such as a His tag) or mRNA is measured. Up-regulation of expression can be determined by comparison of expression of the sequence compared to a control.

Mutagenesis of genes

In this embodiment, the index sequences (*e.g.*, a disclosed array) include randomized genomic sequences extracted from the organism. The landing pad includes an inactive CRISPR guide RNA, which becomes active upon recombination and targets an active version of Cas9 to the corresponding native sequence. This causes repeated cutting of the locus from the Cas9 until the native site is mutated. This method allows random mutagenesis, while being able to identify the targeted gene without whole genome sequencing. For example, because the location of the landing pad is known, that region can be directly sequenced (*e.g.*, by Sanger sequencing) and the targeted gene can be identified. For example, if a mutant phenotype is observed, the synthetic landing pad is sequenced, and the index sequence will encode a fragment of gene X. Gene X can then be sequenced at its native location and the mutation is identified.

In one example, the first portion of the sequence in each of the plurality of index sequences is a randomized genomic fragment from an organism of interest, and the second portion of the sequence in the synthetic landing pad further comprises a CRISPR guide RNA, and wherein the cell further comprises a Cas9 protein. The reconstituted sequence is modified by activated *Cas9* to produce a mutagenized sequence. In particular examples, the mutagenized sequence is identified by sequencing. In another example, the amplicon product from the synthetic landing pad is hybridized to an array (such as a gene array), in order to identify the mutagenized gene.

Analysis of evolution of adaptation

In this embodiment, the index sequences (*e.g.*, a disclosed array) include small, randomized DNA sequences (*e.g.* barcodes). In some examples, the barcode is about 4-20 nucleotides in length (such as about 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides long). In other examples, the barcode is greater than 20 nucleotides in length, such as greater than 25, greater than 30, greater than 40, or greater than 50 nucleotides long. Once integrated in

a synthetic landing pad, the sequences are locked to the genome and can be observed or traced through a lineage, for example, for analyzing evolution.

In some examples, the first portion of the sequence in each of the plurality of index sequences is a unique barcode sequence, and expression of the reconstituted sequence is monitored by detection of the barcode. In particular examples, the method is performed *in vivo* (*e.g.*, in *C. elegans*) and the lineage of a cell can be traced by detecting the barcode.

Discovery of functional residues

In this embodiment, the index sequences (*e.g.*, a disclosed array) include variants of single gene fragments which have been systematically or randomly varied, to change residues upon translation. The landing pad includes a portion of the gene coding sequences that does not vary, including the gene's regulatory sequences. Once integrated, these variants are expressed and can be analyzed (*e.g.* by a phenotype assay). Variants with a selected phenotype can be determined by sequencing and correlated with their effects. In one example, the array is an alanine scanning library, which allows variation of each residue.

Protein-protein interactions

In this embodiment, the index sequences (*e.g.*, a disclosed array) include potential protein coding sequences (*e.g.*, synthesized or obtained through genomic fragmentation). A schematic of an exemplary embodiment is shown in FIG. 5. For example, the landing pad (*e.g.*, "TARDIS insertion site") includes gene regulatory sequences and the coding sequence for a peptide that can be used as a readout for protein-protein interactions. Upon integration of a protein coding sequence (prey) from the array, the landing pad site produces a hybrid protein that contains both the prey and the peptide used to detect protein-protein interactions. In addition, a second protein sequence is expressed (from the same or a different genomic location) that is hybrid for the test sequence used to test for protein-protein interactions (bait) and a second peptide sequence with functional relevance. If the bait and prey interact, then the two functional peptides are brought into contact and result in a functional readout. In some examples, this assay could reconstitute a functional transcriptional regulator where transcription of a reporter is modulated, for example, reconstitution of a fluorescent reporter so that fluorescence indicates interaction, bringing together two fluorophores whose interaction can be monitored (*e.g.*, by FRET), or reconstitution of a functional phenotype, such as antibiotic resistance. In one non-limiting example, the

peptides that provide a functional readout are a Gal4 activator domain and a Gal4 binding domain and the reporter is operably linked to an upstream activator sequence (UAS) for *Gal4*. If the bait and prey interact, the reporter gene is expressed, providing information about the protein-protein interaction. If the bait and prey do not interact, then the reporter gene is not expressed.

Changing native localization of genes

In this embodiment, the index sequences (*e.g.*, a disclosed array) includes gene promoter fragments and the landing pad includes an inactive CRISPR guide RNA, which becomes active upon recombination with the promoter sequence and targets an inactive version of *Cas9* coupled to a tissue-specific transcription factor to the native promoter. This results in altered tissue expression of the gene. Thus in some examples, the first portion of the sequence in each of the plurality of index sequences is a different promoter element, and the second portion of the sequence in the synthetic landing pad further comprises a CRISPR guide RNA, and wherein the cell further comprises a catalytically inactive *Cas9* linked to a tissue-specific factor. Tissue expression of the reconstituted sequence is altered upon activation of *Cas9*. Detecting changed expression of the native gene can be done in several ways, including qPCR to broadly assay changes in mRNA, single cell sequencing to detect cell-specific changes, *in situ* hybridization, or a phenotypic readout (including, but not limited to longevity). Alternatively, if the native gene has a fluorescent tag or other reporter, the location change of the expression can be determined by detecting the reporter.

EXAMPLES

The following examples are provided to illustrate certain particular features and/or embodiments. These examples should not be construed to limit the disclosure to the particular features or embodiments described.

Example 1

***In situ* Donor Assembly and Integration**

Two or six PCR fragments with 30 bp overlaps, covering the *sqt-1(e1350)* gene, were amplified from pDD285 using Q5 polymerase (NEB) in accordance with manufacturer instructions. Homology arms with 30 bp overlaps with the *sqt-1(e1350)* gene were similarly

amplified from pMS81. These homology arms were then complexed with the adjoining *sqt-1(e1350)* PCR fragment through a second round of PCR.

For in-situ assembly and integration, a mixture consisting of 50 ng/µL pMS79, 5 ng/µL pZCS16 and 40 fmol/µL of each of the appropriate PCR products was microinjected into the gonad of young adult PX696 worms. As a control, 10 ng/µL pZCS52 as substituted for the PCR products. Following injection, all worms were maintained at 25°C for the duration of the experiment. After 24 hours, injected adults were moved to new plates to facilitate counting. F1 individuals were screened for red fluorescence and the roller phenotype at 3-4 days post injection. Hygromycin B was then added to plates at a final concentration of 250 µg/mL. Each day for five days post exposure, plates were scored for hygromycin resistance. Individuals resistant to hygromycin and with the roller phenotype were singled to new plates without hygromycin and screened for Mendelian inheritance of the roller phenotype to indicate an integration event. Lines with promising candidates were singled until they produced homozygous progeny, which were then screened by PCR and Sanger sequencing for correct transgene assembly and integration.

While plasmids offer the advantage of producing large quantities of the repair template, they can be laborious to produce. Standard cloning practices require a source of DNA, a ligation step, bacterial transformation, plasmid purification, and verification. This process can be costly in terms of time and funds while requiring technical expertise and lab equipment. As we sought to reduce the overall time-to-integration, we attempted to bypass the cloning step and utilize the *C. elegans* native homology directed repair to produce a transgene (FIG. 5.1). To test this approach, we utilized the *sqt-1(e1350)* mutant as it has a dominant roller phenotype allowing us to easily assay for in situ assembly. Confirmed correct in situ assembled and integrated *sqt-1* gene was obtained using both two and six PCR fragments (Table 3.2). As expected, two parts were easier to correctly assemble and integrate compared to six parts. In some cases, hygromycin resistant individuals were observed without the *sqt-1* roller phenotype. We believe these represent incorrect integration events, where at least the 5' hygromycin resistance coding fragment was integrated into the genome. As these cannot represent correct integration and assembly events, we did not pursue or characterize them.

For two-part assemblies, most hygromycin integration events were accompanied by *sqt-1* assembly and integration, as indicated by the ability to isolate homozygous roller populations. However, not all of these insertions matched the expected sequence. Two of the 8 insertions could not be amplified by PCR, suggesting large insertions or deletions, while three had point mutations identified during sequencing and three had no detectable errors. In contrast, in the six-part experiment, all resistant plates had non-roller (incorrect) integration events, with a few having roller individuals as well (Table 5.1). In most cases, a homozygous roller line could not be isolated, suggesting these individuals were the result of correctly assembled genes in arrays paired with incorrect integrations. In one case a homozygous roller line could be isolated, indicating multiple integration events had taken place in that brood. In this case the roller causing integration had a correctly assembled *sqt-1(e1350)* gene but also contained a second copy of one of the homology arms.

Example 2

Generation of Barcoded *C. elegans*

PCR products for donor homology and Injection mixture

Amplicon 1 (barcode amplicon) was prepared by PCR amplification from ultramer ZCS133 with oligos ZCS134 and ZCS135 (Table 2) for ~30 cycles with Q5 polymerase (New England Biolabs). Amplicon 1 (134 bp) was gel extracted.

Amplicon 2 (left homology arm) was PCR amplified off pZCS32 with oligos ZCS273 and ZCS275 (Table 2) for ~30 cycles with Q5 polymerase and gel extracted (500 bp).

Amplicon 3 (right homology arm) was PCR amplified off pZCS32 with oligos ZCS236 and ZCS278 (Table 2) for ~30 cycles and gel extracted (500 bp).

Amplicon complex (compressed barcodes for array formation) was prepared by complexing PCR to fuse Amplicons 1, 2, and 3. All three amplicons were mixed at ~1 ng each, along with oligos ZCS273 and ZCS278 and PCR amplified for ~30 cycles with Q5 polymerase. The final amplicon product was gel extracted.

An injection mixture was prepared by mixing 10 ng/µL pZCS36 (pUC19-hsp-16.41::piOptCas9::tbb-2u), 0.5 ng/µL pZCS38 pUC19- prsp-27::NeoR::unc-54utr;), 15 ng/µL pZCS41 (SEQ ID NO: 8) (perfect homology sgRNA targeting to SLP 32.1PC), with ~45 ng/µL

amplicon complex. The injection mix must be above 40 ng/µL, however, a toxic or upper limit has not yet been identified. The injection mixture was frozen at -20 °C until needed.

Table 5.1 PCR product barcode library from ZCS133

Name	**Sequence***	**SEQ ID NO:**
Ultramer ZCS133	CCGACGGATTCTACAAGgtaagtttaaacatatNNNNNNNNNNNNNNNtatttaaattttcagGACCGTTACGTCTACC	1
ZCS134	GATACGTCCTCCGTGTCAACTCCTGCGCCGACGGATTCTACAAGgtaagtttaaacata	2
ZCS135	ATTGGGAGGGCGGCGGAGGCGAAGTGACGGTAGACGTAACGGTCctgaaaatttaaata	3
ZCS273	CGCGTCTCTTCCGTGCG	4
ZCS275	CGTCGGCGCAGGAGTT	5
ZCS276	GTCTACCGTCACTTCGCCTCC	6
ZCS278	GACCTCGTATTGGGAGTCTCCG	7

*Uppercase signifies coding sequence and lowercase signifies non-coding sequence.

Injection of C. elegans

Day 0: Single L4 staged strain 32.1PC (Synthetic Landing Pad strain for Barcode Integration) were placed into a small NGM plate with food at 15 °C overnight.

Day 1: Injection mixture was spun in microcentrifuge at maximum speed for 10 minutes. About 30 young adult worms were injected and placed on individual plates at 20 °C.

Day 4: Each injected worm plate was top spread with 1.56 mg/mL G418. If the plate was approaching starvation, 50-100 µL concentrated HB151 or OP50 was added. Water was added if necessary, to ensure enough liquid volume to cover the top of the plate. Plates were dried briefly on the benchtop, then placed back at 20 °C.

Day 11: Plates were screened for those worms that survived G418 selection, indicating presence of TARDIS array. Lines were maintained on G418 to retain TARDIS array and each line was frozen down.

Day 11+: Several large plates were grown with G418 and concentrated food until the population was saturated and gravid. A bleaching event was performed to isolate the eggs. Briefly, worms and bacteria were pelleted in a 15 mL falcon tube. After pouring off supernatant, 5 mL M9 medium, 600 µL bleach, and 300 µL NaOH were added, and vortexed to mix briefly. Worms were allowed to dissolve in bleach for approximately 5 minutes, checking every minute afterwards for completely dissolved worms and free-floating eggs.

Once free-floating eggs were isolated, they were buffer extracted with M9 to wash away the bleach 3X times by pelleting the eggs, pouring off the supernatant, and repeating with M9. On the last wash, the falcon tube was filled with M9 and G418 added at 1.56 mg/mL. The eggs were allowed to hatch in M9 overnight in a nutator at 20 °C. Only worms with the TARDIS array will hatch due to the Neomycin selection.

L1 worms were pelleted and washed with M9 2X. L1 worms were plated to desired density on a large plate without G418 and allowed to grow at 20 °C overnight. The L2/L3 population was heat shocked at 35.5 °C for 1 hour, inducing the heatshock *Cas9*, targeting the SLP and creating a break.

The worms were allowed to reach adulthood and start laying eggs. Then 250 µg/mL Hygromycin B was added to select for only barcode integrated worms. If the barcode integrated into the chromosome, the worms have a restored hygromycin resistance gene. If not, the worms will die.

Depending on how many worms were heat shocked and how many barcodes were in the TARDIS array, the population will be highly diverse. Around 1 in 10 worms heat shocked exhibited an integration event, and the TARDIS arrays have around 600 barcodes per independent array.

Example 3

Index Directed Genomic Screening

While forward genetic screens are invaluable in biological study, they suffer from three major weaknesses: (1) Gain-of-function mutations are exceedingly difficult to obtain; (2) Mutations in many genes relevant to the phenotype of interest are difficult or impossible to obtain due to epistasis; and (3) Tissue- or developmental stage-specific mutations are either

exceedingly difficult or impossible to recover. These issues are addressed by the Index Directed Genomic Screening (IDGS) system that enables (1) controllable, "directional" mutations, (2) ~1,000 fold increase in transformation throughput, (3) recovery of mutations for target genes/phenotypes unobtainable by traditional mutagenesis (*e.g.* phenotypes that exhibit post-reproductively or genes that cannot be targeted due to epistasis), and (4) the reduction of months of mapping work to a simple Sanger sequencing reaction. The IDGS system enables the creation transgenic animals with unique index insertions targeting random genes specifically desired spatiotemporally controlled genetic changes. The IDGS technology represents a transformational change in genetic screening across all biological disciplines.

A genomic index integration system

Genetically engineered genomic "synthetic landing pads" (SLPs) are used that can carry a portion of a selectable sequence (*e.g.*, a hygromycin antibiotic resistance cassette (*HygR*)), sequences for facilitating integration of new DNA sequences (*e.g.*, a synthetic *Cas9* guide target, and homology arms for homology directed repair (HDR) (FIG. 3A)). Integration of index sequences reconstitute the selectable sequence (*e.g.,* reconstitution of the functional *HygR*) by the native HDR machinery. Appropriate integrations are selected using the appropriate selection for the landing pad used. The throughput of integration can be increased using the "data compression" of the TARDIS system, in which injection of massively diverse index mixes results in neomycin selectable intermediate arrays that carry >1,000 indexes each (FIG. 5.3B). One array-carrying individual can give rise to large array carrying populations in which integration is induced (FIG. 5.3A). Once induced, only single indices are integrated, and subsequently expressed across the population, allowing the full library of indices to be expressed in single lineages. This approach holds many benefits over traditional mutagenesis in *C. elegans*. For example, inserted indexes can be cloned by simply sequencing a PCR product, a >1000-fold decrease in labor versus traditional mutation mapping. This allows identification of indexes in post-reproductive, developmental arrest, or sterility situations. While the synthetic landing pads make identification of the genetic alteration fast and cheap, the synthetic landing pads themselves can be engineered to couple the index with genetic regulatory systems for screens (see below).

Directional regulatory index synthetic landing pads

To perform genetic screens using index sequences the index sequence is coupled with regulatory machinery that will ultimately alter gene function in a sequence-specific fashion. Two example synthetic landing pad configurations for genetic screens are provided here, one specific for gene up-regulation and one for down-regulation (FIG. 5.4). The up-regulation synthetic landing pad embeds the index integration site into a CRISPR guide RNA expression cassette such that the index becomes the target sequence of the gRNA. This enables the index to guide Cas9 proteins to distant genomic loci in a sequence specific fashion. When coupled with a DNAse-deficient *Cas9* (*dCas9*) variant that is tethered to a transcriptional activator for upregulation of genes, the indexed-gRNAs enable a high-throughput, forward genetic screen specific for upregulation of genes in a model animal. The second example synthetic landing pad is a dual-promoter system in which the index is transcribed in bi-directionally. The resultant dsRNA triggers an RNAi response for genes that contain the index sequence. In a genetic background without systemic RNAi, the loss of gene function is restricted to the tissues in which the index promoters are expressed.

For both synthetic landing pad versions, usage of spatially and temporally restricted promoters gives a unique, high-throughput forward genetic screening system with spatiotemporal control for altered gene function. This enables screening for mutational effects with tissue and developmental restriction not previously available. While two index mediated screening approaches are described in this Example, there are many other ways to use this highly adaptable system. Additional non-limiting exemplary applications are provided in Table 5.3.

Table 5.2 Exemplary applications of IDGS technology

Index Synthetic Landing Pad	**Effector**	**Uses**
		Genetic Screening
CRISPR Guide RNA	Cas9	Targeted somatic mutation with temporal-spatial control from *Cas9* and guide RNA promoters
	dCas9 with tethered positive transcriptional regulator	Targeted up-regulation of native genes with temporal-spatial control from *dCas9* and guide RNA promoters

Index Synthetic Landing Pad	**Effector**	**Uses**
	dCas9 with tethered negative transcriptional regulator	Targeted down-regulation of native genes with temporal-spatial control from *dCas9* and guide RNA promoters
Bi-directional promoter	Native RNAi machinery	Bidirectional transcription of index to create dsRNA for RNAi. In a *sid-1(-)* background the promoters used would give spatiotemporal control of gene inactivation for screening
Expressed mRNA	Transitive RNAi	Transcription of index as part of an mRNA which is targeting for RNAi. Transitive RNAi would induce silencing of genes with the index sequence
		Screening Protein Function
Protein fusion used as prey in two-hybrid system	Another expressed protein which acts as the bait, which induces a phenotype when it physically interacts with the prey protein.	Protein-protein interactions in an animal where post-translational regulation, intercellular signaling, and tissue differences are available. Use of a synthetic landing pad attaching the index sequence to a split transcription factor enables organism level two-hybrid assays

Table 5.2 Cont.

Example 4

Adaptive Barcoded Lineage Tracking

TARDIS was originally developed for adaptive barcoded lineage tracking. Lineage tracking aims to identify and quantify the selective advantage of unique lineages during experimental evolution. This was experimentally measured by sequencing barcodes. Barcodes were integrated in the genome of independent lineages, marking each with a unique sequence. Lineage tracking requires minimally tens of thousands of individual lineages competing with one another. For microbial systems, this is achieved with library transgenesis by transformation. For animal systems, the TARDIS system was developed.

Methods

Strains and Growth Conditions.

Bristol N2-PD1073 and the derived strains PX740, PX742 and PX743 were maintained using standard *C. elegans* protocols (Stiernagle, T. (2006). Maintenance of *C. elegans*. *WormBook*, *1999*, 1–11). In brief, strains were maintained on NGM-agar plates seeded with OP50 or HB101 E*scherichia coli* at 15 ºC unless otherwise noted.

Molecular Biology

All plasmids were cloned by Gibson Ligation following standard HiFi protocol from New England Bio Labs. pUC19 vector was PCR amplified with ZCS149 and ZCS150 to open the MSC for all cloning. To generate heatshock inducible *Cas9*, *hsp16.41* was derived from pMA122. The germline licensed piRNA depleted *Cas9* and *tbb-2* 3' UTR was derived from pCFJ150-Cas9(dpiRNA). The final construct *hsp16.41p*::C*as9dpiRNA*::*tbb-2* '3UTR was Gibson cloned into pUC19 to create pZCS36. To generate an empty gRNA vector, the *U6* promoter and gRNA scaffold sequence was amplified from pDD162. The final construct *U6p*::(empty)gRNA::U6 Terminator was Gibson cloned into pUC19 to create pZCS11.

To generate the synthetic landing pad targeting plasmid pZCS41, the synthetic target sequence GCGAAGTGACGGTAGACCGT (SEQ ID NO: 9) was added with primer overlaps and Gibson cloned into vector pZCS11. To generate the synthetic landing pad sequence pZCS32, the 5605 SEC homology vector pMS4 was PCR amplified with ZCS139 & ZCS138 to remove the *let-858* terminator to create pZCS30. pZCS30 served as an intermediate to pZCS32. The broken hygromycin landing pad was constructed by removing the intron within the hygromycin resistance gene in pCFJ1663. The perfect homology targeting site was created by fusing exon 1 and exon 2, while also removing three codons from both exons. The vector backbone was amplified with ZCS140 and ZCS145 from pZCS30. The *prsp-0p*::HYGRΔ5' was amplified with ZCS141 and ZCS154 from pCFJ1663. The HYGRΔ3'::*unc-54* 3' UTR was amplified with ZCS155 and ZCS144. The final plasmid was confirmed by Sanger Sequencing. The Neomycin Resistance co-marker pZCS38 was cloned by PCR amplifying *prsp-27*::NeoR::*unc-54* 3' UTR from pCFJ910 with primers ZCS164 and ZCS165, and was cloned into pUC19. To generate an additional fluorescent co-injection marker, *eft-3p* and *tbb-2* 3′ UTR (amplified from pDD162) and wrmScarlet (amplified from pSEM89; a gift from Thomas Boulin) were cloned into a pUC19 backbone to give pZCS16.

Base Landing Pad Strain Generation

Strain PX740, which serves as the base TARDIS strains for Array Quantification and Lineage Tracking, was made by injecting pZCS32 (10 ng/µL) along with pMS8 (50 ng/µL). pMS8 is a *Cas9*, guide RNA vector for targeting Chr. II. Screening for integrations was performed following Dickenson *et. al*. (2015) Genet. 200(4):1035-49. A single post-Cre candidate was isolated and designated PX740.

For a separate application, independent of lineage tracking, TARDIS was applied to build a library of promoters to identify the expression pattern of thousands of genes. Three separate strains, GT300, GT331, and GT332, all three have split hygromycin selection landing pads similar to PX740. However, instead of being split within an intron, half of the coding sequence is integrated into the genome. On the opposite side of the split Hygromycin resistance gene is a partial coding sequence for mScarlet. The donor plasmid is constructed with a promoter, lacking partial mScarlet coding sequence, and the hygromycin resistance promoter, also lacking coding sequence. This arrangement enables easy identification of correctly inserted sequence units from the library because only correct integrants will exhibit hygromycin resistance and an observable fluorescent protein.

There are minor differences between the three strains: GT300 contains a co-marker, which is removed upon integration. GT300 is also an *unc-119* mutant, causing a paralysis phenotype that is rescued with array-expression of *cbr-unc-119*. GT331 and GT332 are different in that one of these strains has a pest tag on mScarlet to degrade mScarlet quicker, preventing the buildup of the protein and possibly provide a finer-scale resolution of when the expression is occurring.

TARDIS Libraries for Lineage Tracking

The first TARDIS library constructed and utilized was the "500 bp homology library." The 500 bp homology library was constructed by PCR amplifying ZCS133 with ZCS134 and ZCS135. ZCS133 contains a set of 15 randomized sequences (NNN's) directly in the middle. The amplicon product served as a base for further PCRs. The Left Homology Arm was amplified from pZCS32 with primers ZCS273 and ZCS275 to produce a 501 bp product. The Right Homology Arm was amplified from pZCS32 with primers ZCS276 and ZCS278 to produce a 501 bp product. The final library was generated by complexing PCR with the initial product and

the two homology arms in a single PCR reaction with primers ZCS273 and ZCS278 to produce a 1068 bp product.

This library was gel extracted to remove any templates and incorrect amplicon products. The library product was injected along with pZCS36, pZCS41, ZCS38, and pZCS16 to create PX742.

The second TARDIS library constructed and utilized was the "150 bp homology library." The primary focus of this library was to increase the number of possible donors present in the array. This was achieved by focusing on the quantity of DNA, and the PCR cycle number. Two separate Q5 PCR reactions were performed with 2.5 µL 0.5 µM ZCS357, 1.25 µL 10 µM ZCS134, 1.25 µL 10 µM ZCS135, 20 µL water, and 25 µL 2x Q5 Master Mix™. Four cycles of PCR were performed with 45 seconds of extension time at 72ºC annealing to produce the barcode base library. The product was column purified to retain as much product as possible. The 150 bp Right Homology Arm was amplified with ZCS276 and ZCS286 from pZCS32, the 150 bp Left Homology Arm was amplified with ZCS285 and ZCS275 from pZCS32. Both homology arm products were gel extracted. Four secondary complexing PCRs were performed with 1 barcode donor fragment:10 left homology arm:10 right homology arms, for 15 cycles with ZCS285 and ZCS286. The final library was gel extracted to avoid incomplete and template products. The final library was injected at 61.75 ng/µL with pZCS36 at 20 ng/µL, pZCS41 at 15 ng/µL, pZCS38 at 0.25 ng/µL, and pZCS16 at 3 ng/µL for a total concentration of 100 ng/µL to create PX743.

TARDIS Single Promoter

Injections into GT300 were done with 25 ng/µL pMS84 (gRNA targeting GT300 synthetic site). 7ng/µL heatshock *Cas9*(dpiRNA) PCR product (analogous to pZCS36), 10 ng/µL pNU1495 (*cbr-unc-119* rescue cassette; analogous to neomycin resistance in the lineage tracking TARDIS libraries), 10 ng/µL pDSP18 (*daf-7* promoter with split mScarlet and split Hygromycin resistance). Candidate library strains were selected by rescue of the paralysis phenotype.

Multiple Promoter Libraries

For GT331 and GT332, the same injection mix was used: pMS84 at 15 ng/µL, pZCS36 at 20 ng/µL, pZCS38 at 0.5 ng/µL, pNU681 (an *eft-3p*::*GFP* co-marker for array identification)

at 2 ng/µL. The promoter library was injected with the following plasmids and their corresponding promoters at 0.45 fmol/µL to facilitate approximately equal composition in the array library: pEA1(*ceh-20*), pEA2 (*ceh-23*), pEA3 (*ceh-40*), pEA4 (*egl-46*), pEA5 (*hlh-16*), pEA6 (*nhr-67*), pEA9 (*ceh-43*), pEA10 (*mdl-1*), pEA11 (*egl-43*), pEA12 (*aha-1*), pEA13 (*ceh-10*), pEA14 (*ahr-1*), pEA15 (*lin-11*), and pDSP18 (*daf-7*).

Several TARDIS array lines were generated from the injections. Two were selected for testing, one from GT331 and one from GT332.

Dual Integrations

One possible benefit of the TARDIS system is the integration of multiple sequences into different sites. To address this possibility, a single strain was generated, GT344, with two unique landing pads. One landing site is similar to GT332, with split Hygromycin B selection and split mScarlet. The second landing pad is engineered with *unc-119* split selection, which is similar in protocol to GT300. GT344 was injected with 50 ng/µL pMS79 (*Cas9* targeted plasmid with guide RNA for both landing pads), 10 ng/µL pDSP57 (split mScarlet donor homology) and 10 ng/µL pDSP58 (split *unc-119* donor homology). After nine days, Hygromycin B was top-spread at a final concentration of 250 µg/mL. Survivors were screened for rescue of *unc-119* indicating dual integrations.

TARDIS Integration Protocol for Barcoded Lineage Tracking

TARDIS array-bearing strains are grown at 15ºC to reduce expression of heatshock *Cas9*. Strains were grown with G418 at 1.56 mg/mL, which provides selection to maintain the array, unless otherwise specified. Large plates were grown until a sizable population of array bearing strains were gravid. Large gravid populations were bleached to isolate eggs following standard protocols. Synchronized L1 individuals were hatched overnight in M9 solution with G418 at 1.56 mg/mL. This ensures all progeny that hatch bear the TARDIS array library. Worms were centrifuged into a pellet and the supernatant was discarded. Worms were washed 2X with M9 to remove remaining G418. L1s were then plated at desired densities and allowed to grow at 15ºC overnight. The next day, L2 worms were heatshocked at 35.5ºC for 2 hours. Worms were then placed at a desired rearing temperature. Once the population reached Day One Adults, Hygromycin B was added at 250 µg/mL to select for barcode integration.

TARDIS Integration for Promoter Libraries

For the single promoter GT300 based library, stable array lines were isolated to large NGM plates and allowed to lay progeny for 4 days at 15ºC. Plates were heatshocked at 34ºC for 1 hour. Plates were then incubated at 25ºC. After 2 days, Hygromycin B was added at a final concentration of 250 μg/mL.

For mixed promoter libraries injected into GT331 and GT332, an identical integration protocol used for Lineage Tracking was used.

Array Quantification

Individual array line PX742 was lysed and amplified with ZCS287 and ZCS288, and sent for Illumina sequencing. Reads were acquired and quality filtered using custom Python code for only reads with quality scores of 30 or higher for the barcode region and barcodes having greater than 10 reads.

Individual array line PX743 was lysed and amplified with New England BioLabs (NEB) Next Ultra II™ kit Universal Primer and Index Primer 2 with 30 cycles. Reads were acquired and quality filtered using custom Python code for only reads with quality scores of 30 or higher for the barcode region and barcodes having greater than 10 reads.

Integration Efficiency

PX742: PX742 was grown, synchronized, and heatshocked, following the protocol above. After heatshock, 10 L3 individual were isolated to 20 plates. Once they reached Day One adulthood, Hygromycin B was top spread at 250 μg/mL. Plates were allowed to grow for several days before screening for survivors.

GT300: To address the efficiency of the GT300 landing pad, 60 individual array-bearing individuals were singled and allowed to lay progeny for four days at 15ºC. Plates were heatshocked at 3 ºC for 1 hour, and then incubated at 25ºC. Two days later, a final concentration 250 μg/mL Hygromycin B was top spread.

To further address the developmental stage in which integration was occurring, several individuals were heatshocked at the L2, L3, L4, and young adult stages.

GT331 & GT332: These libraries were tested for integration efficiency, following the integration protocol listed below. To establish the efficiency of integration, population of 50 individuals/plate, 100 individuals/plate, 200 individuals/plate, and 1,000 worms/plate were synchronized. This was repeated with two separate TARDIS library arrays, one with the GT331 background and the other with GT332. After Hygromycin exposure, plates were screened for survival.

Integration Bias Checking Methods

PX742 counts and relative frequencies were compared to a population of individual worms after TARDIS-based integration. A large integrated population of synchronized L1 worms was heatshocked. The population was allowed to recover for one generation before being lysed and amplified. Two sets of amplification had to be performed to avoid the possibility of amplifying the array. First, four rounds of PCR were performed with ZCS305 and ZCS304. ZCS304 introduced a new specific primer binding site for the second round of amplification. The PCR product was column purified and the full product was amplified with ZCS306 and ZCS307. This final product was sent for Illumina sequencing. Identical Python code and criteria for array quantification was used to filter reads. Counts were then normalized to the total counts for filtered sequencing runs for both the PX742 array and for the integration population. Barcodes which did not integrate from the array were excluded from the analysis.

Lineage Trajectory Methods

A single parental lineage population was grown to starvation on a large 150 mm plate seeded with concentrated HB101. The parental population was split into 3 separate replicates. Each replicate was also grown on large 150 mm plates seeded with concentrated HB101 and allowed to grow until starvation (~2 generations). At starvation, half of the population was transferred, and half was lysed, for barcode amplification. A total of three transfers and three time points were checked. This approximates to six generations. Barcode amplification followed the protocol set forth above. Data was also analyzed following the same criteria. Normalized data was plotted as both lineage trajectories and as histograms.

PCR jackpotting has the potential to significantly increase the lineage measurement error. To address this, two separate transfers in replicates one and two were chosen to be PCR amplified twice following identical protocols and analysis.

Results

Array Quantification

Approximately 33,137 reads were acquired for PX742 (FIG. 8). After quality filtering, 431 unique sequences were identified from the array line. As expected, barcodes were not injected in perfect equal molar quantities due to their PCR-based construction, and PCR amplification to generate the sequence libraries introduced additional bias. There were many barcodes with fewer reads than expected, with a long tail indicating the result of PCR jackpotting.

For PX743, 771,115 reads were obtained after quality filtering. Because the protocol was modified to reduce the donor homology size, and lower initial amplification cycles for the donor library, more barcodes were expected to be present in the array. 2,965 unique donor sequences were present. Lower cycles and reduced donor homology length increases the total number of barcodes that can be present in an array.

Integration Efficiency

Lineage Tracking-based Libraries

PX742 is the 500 bp large donor homology TARDIS line, with a considerably small insert. Of the 20 plates, 17 had surviving resistant progeny, indicating a proper integration event. Five plates were randomly selected for Sanger Sequence to confirm the integration. Of the five, two showed two separate integration events, indicating that more than one worm from the ten on the plates had an integration event. Therefore, the integration efficiency was approximately one in ten heatshocked individuals integrated a barcode.

Promoter-based Libraries

GT300-based integrations: For GT300, eight days after dosing with hygromycin all five plates had live worms. Three individuals from each plate were selected for confirmation by PCR (MS376/MS248/ZCS84 or MS376/MS121). Two plates had correct insertions by PCR, two plates had incorrect insertions by PCR, and one plates was inconclusive. The incorrect insertions were attributed to co-marker integration into the landing pad, and this was removed for future builds (GT331 and GT332).

To address the integration efficiency, 60 individuals were screened for integrated progeny. Of the 60, 25 of 59 plates had Hygromycin B resistant progeny. One plate had very

few live worms, so it was excluded from analysis. Candidates were not genotyped for proper vs improper integrations.

60 L4 array-bearing individuals were heatshocked and screened for hygromycin survival. Two of the 60 plates produced resistant progeny.

There was a difference in the developmental stage and the number of integrations; younger stages and young adults seemed to work better compared to L3/L4 staged individuals (Table 5.3).

Table 5.3 Developmental stages and integration efficiencies.

Developmental Stage	**Individuals**	**Integrations**	**Percentage**
L2	20	6	30%
L3/L4	19	2	10.5%
L4	60	2	3.3%
Young Adult	19	6	31.5%

GT331 and GT332-based integrations: There was a noticeable difference in the integration efficiency of the two strains (Table 5.4).

Table 5.4 Integration efficiencies for GT331 and GT332.

Survivors		
	GT331 (surviving plates/total plates)	**GT332 (surviving plates/total plates)**
50 worms/plate	1/5	0/5
100 worms/plate	5/5	1/5
200 worms/plate	4/5	2/4
1,000 worms/plate	5/5	2/4

A clear bias was seen in the integration for GT332 compared to GT331. There are several variables involved, such as the time in which worms' eggs are exposed to bleach during the synchronization process, the total number of worms present in the synchronization, the

stability of the array and copy number per cell, and variations in the exact composition from the same injection mixture.

Several promoters were isolated from the library. In some cases, multiple were isolated and noted within (): *ceh-10*(3), *ceh-40*(5), *hlh-16*(7), *ahr-1*(2), *aha-1*(2), *mdl-1*(2).

Dual Promoter integrations: One benefit of the TARDIS system is integration into multiple landing sites within a single strain. The strain GT334 was created to have two separate selectable landing pads with different fluorescent reporters. A correct integration was isolated from this injection.

Integration Bias Checking Results

There was an approximately normal distribution of reads from the array to the integrated population (FIG. 9). In total, 331 independent lineages were identified. The majority of reads indicated a less than 0.5% change in frequency from the array. In this preliminary experiment, no designs were present to remove PCR jackpotting, thus any change in frequency from the array is likely an overestimate of bias. As a result, it appears that the process of integration exhibits little to no bias.

Lineage Trajectory Results

Of the three replicates, each of them behaved similarly. Evolutionary drift would predict that lineages without strongly positive or negative selective pressure would change in frequency at random. Of the 331 lineages traced, four showed increased frequency, suggestive of positive selection (FIG. 10), while the majority of lineages decreased in frequency. When looking at the distribution of lineages frequencies, an increase in the spread of lineage frequencies was seen from the P0 distribution to the first transfer T1 across the three transfers and consistently across the three replicates, and again an increase in spread to T2, consistent with evolutionary drift (FIG. 11). Notably, the trajectories are not identical, and there is variation across the selective benefit of the lineages. With the best performing lineages varying as much as a few percentages from replicate to replicate.

PCR jackpotting may significantly and artificially alter the frequency of a lineage. Molecular identifiers can solve this problem: Two independent PCR replicates were compared for a difference in frequency (FIG. 12). Again, a spread in frequencies increased from the P0

distribution to the first transfer T1, and again an increase in spread to T2, consistent with evolutionary drift. On average the variation between replicates was strikingly low, with the majority being less than 0.2% different (FIG. 12). Therefore, the changes in lineage frequencies are most likely not due to artificial increases or decreased from the PCR protocol.

In view of the many possible embodiments to which the principles of the disclosure may be applied, it should be recognized that the illustrated embodiments are only examples and should not be taken as limiting the scope of the invention. Rather, the scope of the invention is defined by the following claims. We therefore claim as our invention all that comes within the scope and spirit of these claims.

Claims

We claim:

1. A method of producing a plurality of genetically modified cells, comprising:

introducing a nucleic acid molecule comprising a plurality of index sequences into a cell comprising a genomic polynucleotide comprising a synthetic landing pad, wherein each of the plurality of index sequences comprises a first portion of a nucleotide sequence and the synthetic landing pad comprises a second portion of the nucleotide sequence to produce a cell comprising the synthetic landing pad and the nucleic acid molecule comprising the plurality of index sequences;

generating a plurality of progeny cells comprising the genomic polynucleotide comprising the synthetic landing pad and the nucleic acid molecule comprising the plurality of index sequences;

integrating a single index sequence into the synthetic landing pad in each of the plurality of progeny cells, thereby linking the first and second portions of the nucleotide sequence; and

selecting progeny cells comprising integrated index sequences based on presence or activity of the linked first and second portions of the nucleotide sequence, thereby producing a plurality of genetically modified cells.

2. The method of claim 1, wherein the nucleic acid molecule comprising the plurality of index sequences comprises 500-3,000 index sequences.

3. The method of claim 1, wherein the first portion and the second portion of the nucleotide sequence reconstitute a functional gene when linked.

4. The method of claim 3, wherein the functional gene is a selectable marker or reporter gene.

5. The method of claim 3, wherein the synthetic landing pad further comprises a site-specific nuclease (SSN) recognition site and homology arms flanking the SSN recognition site, and each of the plurality of index sequences is flanked by the homology arms in the nucleic acid molecule comprising the plurality of index sequences.

6. The method of claim 5, wherein each of the homology arms is 150-500 nucleotides in length.

7. The method of claim 5, wherein integrating the single index sequence into the synthetic landing pad comprises introducing a DNA break at the SSN recognition site utilizing the SSN, and site-specific integration of the index sequence into the synthetic landing pad.

8. The method of claim 7, wherein the SSN is a caspase, zinc-finger nuclease, or TALEN.

9. The method of claim 1, wherein the nucleic acid molecule comprising the plurality of index sequences is an extrachromosomal array, a plasmid, or an artificial chromosome.

10. The method of claim 1, wherein each of the plurality of index sequences comprises a homologous fragment of the genomic polynucleotide, and wherein each of the plurality of index sequences are different.

11. The method of claim 10, wherein the genomic polynucleotide is an intron or exon of a gene, or a promoter element.

12. The method of claim 1, wherein the cell is a eukaryotic cell or bacterial cell.

13. The method of claim 12, wherein the cell is a yeast cell, mammalian cell, a *Caenorhabditis elegans* cell, or a *Drosophila* cell.

14. The method of claim 12, wherein the lineage of the cell is traced by detecting an index sequence in progeny of at least one of the plurality of genetically modified cells.

15. The method of claim 1, wherein introducing the nucleic acid molecule comprising the plurality of index sequences into the cell comprises injecting the nucleic acid molecule into an animal comprising the cell.

16. The method of claim 1, wherein each of the plurality of index sequences comprises a sequence variant of a reference coding sequence, a sequence variant of a reference non-coding sequence, a library sequence, a randomized sequence, or a promoter element.

17. The method of claim 16, further comprising: selecting a single sequence variant of the reference coding sequence by selecting a genetically modified cell comprising the reference coding sequence variant; selecting a single sequence variant of the reference non-coding sequence by selecting a genetically modified cell comprising the reference non-coding sequence variant; selecting a single library sequence by selecting a genetically modified cell comprising the library sequence; selecting a single randomized sequence by selecting a genetically modified cell comprising the randomized sequence; or selecting a single promoter element by selecting a genetically modified cell comprising a screenable marker or reporter gene operably linked to the promoter element in the genomic polynucleotide.

18. The method of claim 1, further comprising:

selecting a genetically modified cell comprising an index sequence by an assay phenotype, or by expression of a selectable marker or reporter; generating variants of the index sequence; introducing a nucleic acid molecule comprising the variants of the index sequence into a cell comprising a genomic polynucleotide comprising a synthetic landing pad, wherein each of the variants of the index sequence comprises a first portion of a nucleotide sequence and the synthetic landing pad comprises a second portion of the nucleotide sequence to produce a cell comprising the synthetic landing pad and the nucleic acid molecule comprising the variants of the index sequence; generating a plurality of progeny cells comprising the genomic polynucleotide comprising the synthetic landing pad and the nucleic acid molecule comprising the variants of the index sequence; integrating a single variant of the index sequence into the synthetic landing pad in each of the plurality of progeny cells, thereby linking the first and

second portions of the nucleotide sequence; and selecting progeny cells comprising integrated variants of the index sequence based on presence or activity of the linked first and second portions of the nucleotide sequence.

19. A genetically modified cell comprising: an extrachromosomal array comprising a plurality of index sequences; and a genomic polynucleotide comprising one of the plurality of index sequences integrated at a synthetic landing pad, wherein the integrated index sequence comprises a first portion of a nucleotide sequence and the synthetic landing pad comprises a second portion of the nucleotide sequence, and wherein the first and second portions of the nucleotide sequence are operably linked in the genomic polynucleotide.

20. A multicellular organism comprising a plurality of the genetically modified cells of claim 19, wherein the genetically modified cells comprise different index sequences.

ABSTRACT OF THE DISCLOSURE

Provided herein are methods of producing a plurality of genetically modified cells that include introducing a nucleic acid molecule including a plurality of index sequences into a cell comprising a synthetic landing pad, wherein each of the plurality of index sequences includes a first portion of a sequence and the synthetic landing pad includes a second portion of the sequence. The method further includes generating a plurality of cells that include the synthetic landing pad and the nucleic acid molecule including the plurality of index sequences and integrating one of the plurality of index sequences into the synthetic landing pad in each of the cells, thereby linking the first and second portions of the sequence. The linked first and second portions of the sequence result in a functional gene and cells including the integrated index sequence are selected based on presence or activity of the functional gene.

Bridge

In the next chapter, I outline utilizing TARDIS barcoded lineages for experimental evolution. I combined our newly created barcoding system along with experimental evolution to track the lineages of a mutant barcoded strains competing with wildtype individual to measure fitness.

CHAPTER 6 QUANTIFICATION OF ENVIRONMENTALLY-DEPENDENT SELECTION VIA BARCODED ANIMAL LINEAGE TRACKING

In preparation as Stevenson ZC, Laufer E, Robinson K, Phillips PC, Quantitative High-Throughput Barcoded Lineage Tracking in an Animal System. I have designed the experimental protocols, data collection and analysis. Laufer E and Robinson K also performed data collection.

Introduction

Mutation serves as the fundamental engine of diversity in the context of evolution, underpinning the dynamic process of genetic variation within populations. At the molecular level, mutations are spontaneous alterations in the DNA sequence, encompassing substitutions, insertions, deletions, and other genomic rearrangements. These stochastic events introduce novel genetic material into a population's gene pool, upon which natural selection acts, leading to adaptation to the environment. The process of natural selection is inherently contingent upon the variation in environmental conditions, as it acts as the selective filter determining which traits confer a fitness advantage in a given context. Environmental variation, both temporal and spatial, plays a pivotal role in shaping the direction and intensity of selection. Fluctuations in environmental factors, such as temperature (Logan et al., 2014), resource availability (Pekkonen et al., 2013), predation pressure (Bijleveld et al., 2015), and small molecule concentration can lead to shifting selection pressures (Baym et al., 2016; Lenski, 2017; Lenski et al., 1991; Levy et al., 2015), influencing the maintenance or divergence of genetic diversity within populations.

In the lab, we have an opportunity to modulate the environment and test adaptive hypotheses for given genetic variation with experimental replication and precision in the environmental control. Several examples exist for antibiotics, small molecules, and more complex traits such as stress tolerance (Rudman et al., 2022). However quantitatively measuring the fitness and the precise genetic architecture of the adaptive phenotype remain challenging. The distribution and selectively advantageous adaptive mutations do not remain stagnant while populations adapt, an upper limit of clonal interference is met until new adaptive genetics arises from either recombination of standing genetic variation, or as a result of new mutations (Ba et al., 2019; Blundell and Levy, 2014; Good et al., 2017). As a result, the selective coeffects of

mutations is historically dependent. New technological advancements which provide historical context are needed.

Barcodes–small, engineered polymorphisms–provide a simple and eloquent way to observe the dynamics of several lineages within an experiment (Blundell and Levy, 2014). By simply observing the change in lineage frequency overtime, monitored by periodic amplicon sequencing, the selection coefficients can be directly measured. Barcodes have become widely adopted for single-cell systems to study evolution and adopted for development lineage tracking within an animal context (Chen et al., 2022; McKenna et al., 2016). TARDIS (Transgenic Arrays Resulting in Diversity of Integrated Sequences) was recently published by our group as a methodology to perform the first large-scale chromosomal barcoding within an animal system, specifically *C. elegans* (Stevenson et al., 2023, 2021). Animal-based experimental evolutions remain particularly challenging, large population sizes are often needed, and multiple generations are required to fully observe the dynamics. The small nematode *C. elegans* provides an ideal metazoan for experimental evolutions (Teotónio et al., 2017), with large brood sizes (upwards of 300 individuals in optimal conditions) and short generation times (approximately 4 days to reach reproductive adult at 20C). In addition, much of the genetics in *C. elegans* is well known, with a fully sequenced genome, and many characterized mutants.

Here, we report the first-ever barcoded evolutionary lineage tracking experiment performed within an animal system, which allows replicated measurements of selection coefficients of a known mutant within a well-defined environmental context. To grow large population sizes, liquid culture has been previously established for *C. elegans* populations (Stiernagle, 2006). We present a modified protocol for growing several multi-million population sizes in parallel, making this experiment to our knowledge the largest animal experimental evolution study conducted to date.

Establishment of a controlled and simple selective environment was achieved by utilizing various concentrations of ivermectin (IVM), a widely-utilized anthelmintic drug. IVM is known to cause paralysis in *C. elegans* by hyperpolarization of the plasma membrane (Dent et al., 2000). In natural populations, resistance to IVM has evolved several times and is a known problem for industrial agriculture as well as in healthcare (Geerts and Gryseels, 2000). Several known extracellular glutamate-gated chloride channels are known to interact with IVM in the

wildtype background, leading to paralysis. Three combined loss of function mutations in *avr-14*, *avr-15*, and *glc-1* are known to confer over ~4,000x resistance to IVM in standard pharynx pumping assays (Dent et al., 2000). We hypothesized that the triple mutant, JD608 (*avr-14*$^{-}$ (ad1302), *avr-15*$^{-}$(ad1051), and *glc-1*$^{-}$(pk54)) would be adaptive in environments higher in IVM concentration, compared to the sensitive wildtype strain. Our results show that selection can be modulated by changing the concentration of IVM in the liquid environment, and even flip the adaptive advantage or disadvantage of specific lineages. Our project stands as an exemplar of lineage tracking within an animal context and can be applied generally to study lineage dynamics in experimental populations.

Results

Generation of mixed, barcoded populations in a novel liquid environment

Mixed populations of unique barcoded individuals for both wildtype and mutant populations were grown individually on plates and then mixed to generating starting populations for serial liquid culture (Fig 6.1). Liquid culture provides an ideal medium for large-scale population growth as a 'three-dimensional' media. We designed a novel liquid culture protocol utilizing simple magnetic stir bars to maintain mixed cultures (Fig 6.2). While this method also is highly cost-effective, the constant stir ensures populations are mixed well and prevents heterogeneity within the environment, ideal for small-molecule selection experiments. Worms were grown with multiple antibiotics and concentrated bacterial food within NGM buffer (see methods) in addition to experimental concentrations of IVM. Populations generally peaked around several hundred thousand–several times breaching over a million– before a serial transfer (Fig 6.3). At such high population sizes, we would expect genetic drift would be insignificant to measured lineage frequencies.

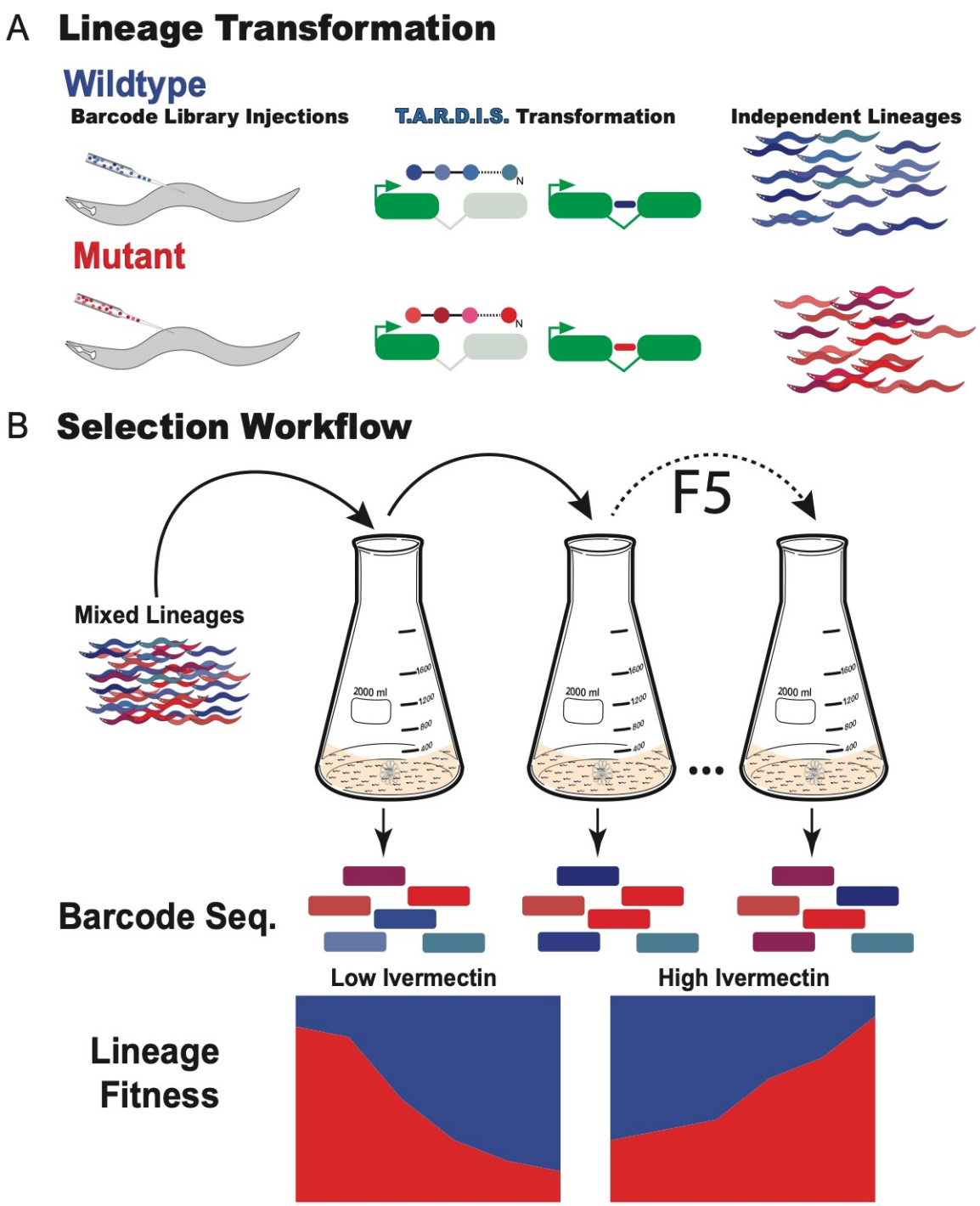


Figure 6.1 Experimental overview. A) Lineage are transformed by TARDIS transgenesis with unique barcode pools to designate the wildtype and mutant background. B) Serial cultivation of the experimental populations for five transfers, with each population being sampled for their barcode frequency, which was used to plot their trajectories and measure fitness.

Figure 6.2 Picture of the liqud culture experimental environment. Several replicates are spun with stirbars in a tempature controlled environment.

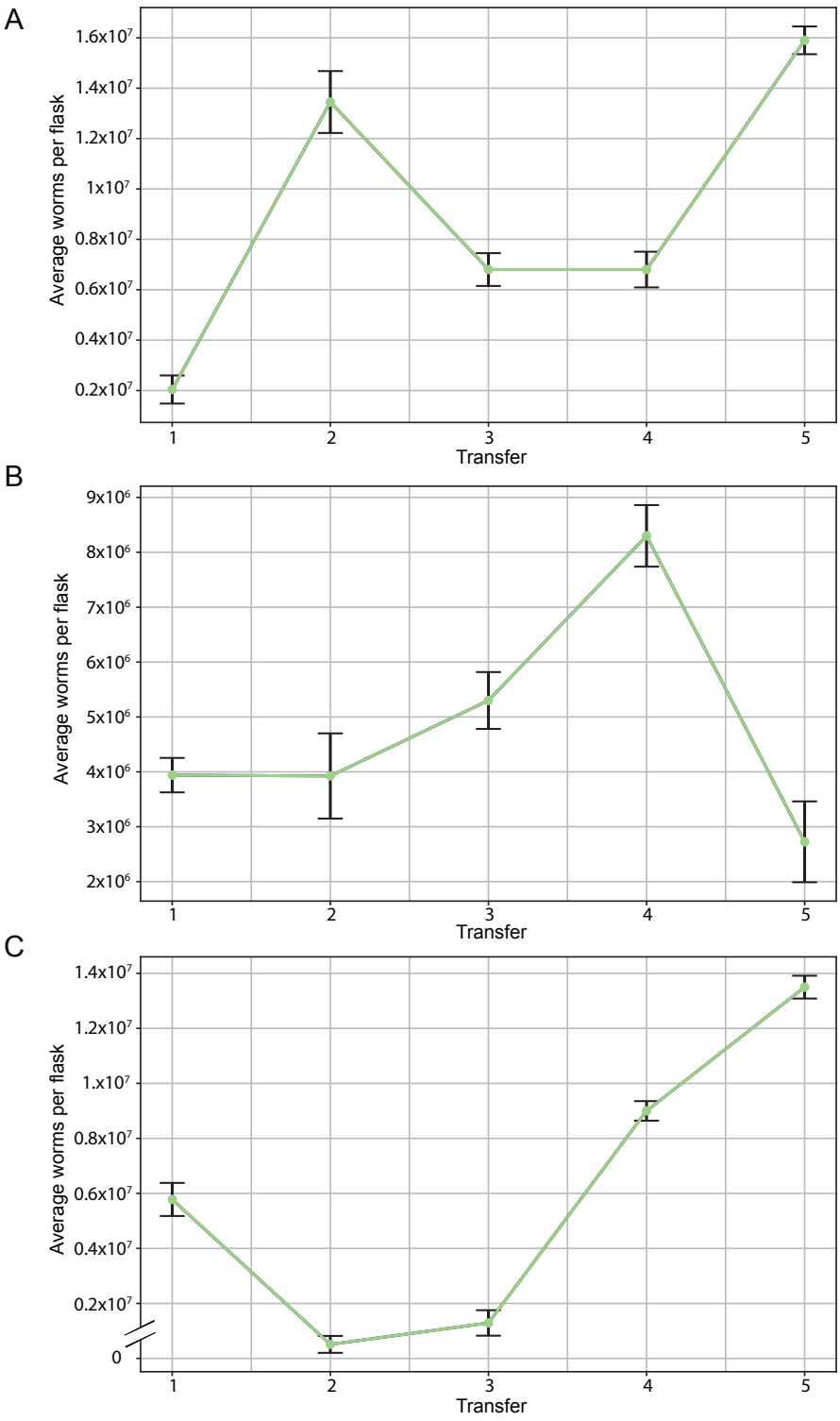


Figure 6.3 Population estimates on each transfer for A) 0nM-control, B) 1.5nM and C) 2.5nM. Error bars represent standard error.

Ivermectin exposure reverses the direction of selection

We tested two IVM selection environments and a control, 0nM (control), 1.5nM (medium concentration) and 2.5nM (high concentrations). We found that there is a clear signal of selection within each condition (Fig 6.4). In our control we find that the wildtype background was favored, showing a significant deleterious cost associated with the triple mutant background. Whereas in contrast to medium and high concentrations, the selection coefficients shifted in favor of the triple mutant. Interestingly, selection was not constant, and some cases a shift within transfers, clearly suggesting that multiple generations are required to accurately measure the selection coefficients (See 1.5nM, transfer one to three, Fig 6.4). We also observed a clear trend with the mean fitness towards increasing fitness of the mutant background with IVM (Table 6.1).

Table 6.1 Calculated mean fitness from FitSeq.

Condition (Ivermectin Molarity)	Favored Background	Approximate Mean fitness
0nM	Wildtype	0.6
1.5nM	Mutant	0.2
2.5nM	Mutant	1

Lineages provide a replicated metric for experimental quality

Our primary objective was to measure the selective advantage of the wildtype and triple mutant strains. However, while we would not expect additional background mutations of large adaptive effect, such mutations are still possible, especially within a completely novel environment such as liquid culture. By creating mixed populations of genetically identical lineages, we expect them to behave approximately identically across our selective experiments. Within preliminary experiments (data not shown), we noticed ‘divergent’ lineages–lineages showing large adaptive advantage compared to their sibling lineages. Lineages provide a higher resolution of the selective population, while also allowing for any lineage dynamics which diverge from the background genotype to be easily detected. In such cases, we can identify replicates which have acquired novel adaptive mutations which no longer represent the initial competitive genotypes and therefore could be discarded.

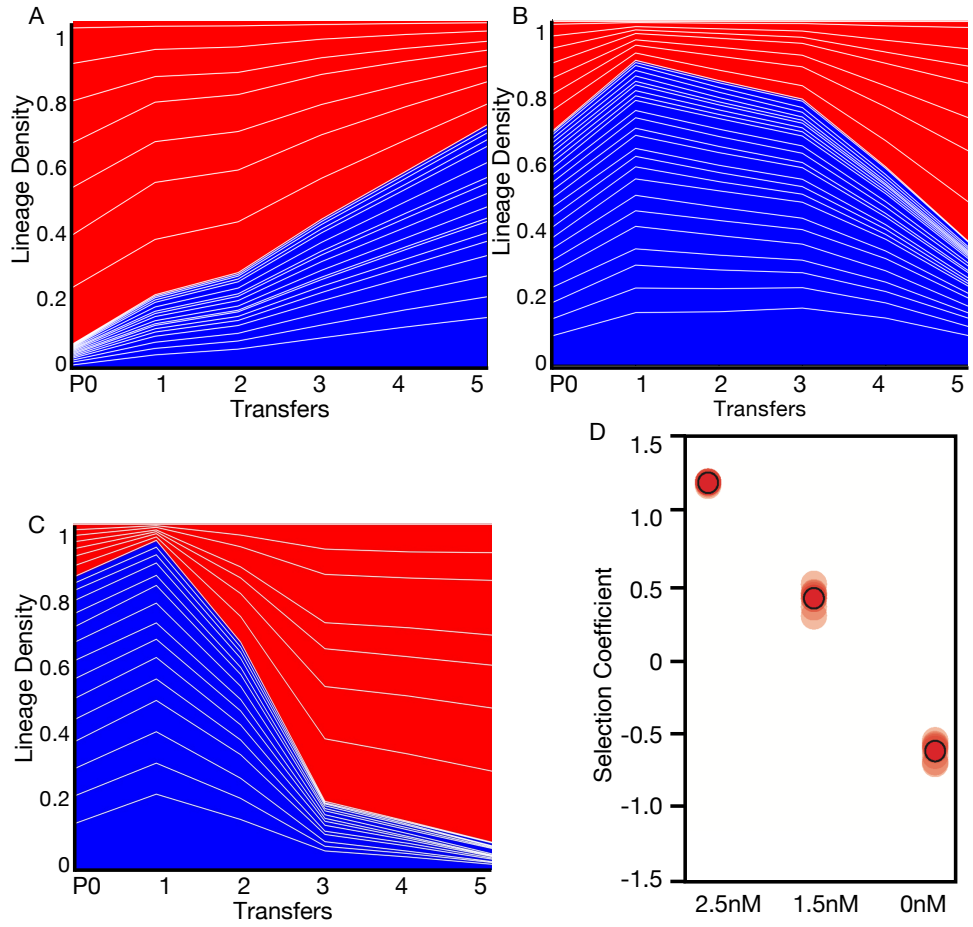


Figure 6.4 Lineage trajectories and fitness for A) 0nM- control, B) 1.5nM and C) 2.5nM. We see a clear trend towards increasing mutant frequency across the Ivermectin concentrations which results in D) an increase mutant selection coefficient normalized to the average wildtype selection coefficient.

Discussion

Here we present the first barcoded evolutionary lineage tracking experiment performed within an animal system. Our results show we can measure selection within environmental context and modulate the selection coefficients with simple changes in IVM concentration–flipping the beneficial alleles. Utilizing our liquid culture approach, we were able to grow populations in the millions, creating the largest animal experimental evolution to date.

Barcoded lineage tracking for distribution of fitness effects

While this project stands as a simple exemplar, within microbial systems similar approaches have been adopted for studying the distribution of fitness effects (Ba et al., 2019; Blundell and Levy, 2014; Levy et al., 2015). Adoption of several, highly diverse barcode TARDIS libraries could reasonably result in several hundred thousand unique lineages (Stevenson et al., 2023), comparable to even the largest barcoded experiments performed in microbes. While population sizes within the billions would be unrealistic, populations within the several hundred million is possible by simply scaling the liquid culture system described here. To compensate for lower total population sizes and therefore lower natural mutations, chemical mutagens could be adopted.

Epistatic experimental lineage tracking

Epistatic contributions to adaptive phenotypes are of clear interest to the community. In this study, we are utilizing a triple genetic mutant which has an adaptive phenotype. While we have not directly measured the fitness of the individual alleles *avr-14*$^{-}$, *avr-15*$^{-}$, and *glc-1*$^{-}$, published evidence shows the large increased resistance to IVM is unique to the triple mutant. However, double-combinations of *avr-14*$^{-}$, *avr-15*$^{-}$, and *glc-1*$^{-}$ show some levels of IVM resistance in pharynx pumping assays (Dent et al., 2000). Barcodes provide a simple independent way to measure the frequency of a genotype without directly sequencing the mutation. Epistatic, multiple mutation combination such as *avr-14*$^{-}$, *avr-15*$^{-}$, and *glc-1*$^{-}$, are impossible to competitively measure without using such an indirect measure simply because within a mixed genomic pool there is no way to discern if an individual mutation came from which combined mixed background. In future work while we plan to investigate this directly with *avr-14*$^{-}$,*avr-15*$^{-}$, and *glc-1*$^{-}$, barcodes generally can be adopted in principle for any genetic combination or phenotype.

Conclusions

In conclusion, we have presented the first barcoded lineage tracking animal experiment evolution, in what is also the largest animal experimental evolution to date. We created a simple experimental design to quantitatively measure selective contributions within an environmental context, and showed we can experimentally modulate the strength of selection, even changing the adaptive background by changing the concentration of a simple small molecule drug IVM.

Materials and Methods

Generation of barcoded lineages

The mutant strain JD608 (*avr-14*$^{-}$(ad1302), *avr-15*$^{-}$(ad1051), and *glc-1*$^{-}$(pk54)) was crossed with PX740 N2-PD1073 fxIs47 [prsp-0p:: 5' ΔHYGR:: GCGAAGTGACGGTAGACCGT :: 3' ΔHYGR::unc-54 3'::LoxP, II:8420157] to create PX776 fxIs47 [prsp-0p:: 5' ΔHYGR:: GCGAAGTGACGGTAGACCGT :: 3' ΔHYGR::unc-54 3'::LoxP, II:8420157] (*avr-14*$^{-}$(ad1302) + *avr-15*$^{-}$(ad1051) + *glc-1*$^{-}$(pk54)) containing all three IVM resistance mutations and the split hygR landing pad. PX776 was injected with TARDIS barcodes following the protocols of Stevenson et al., 2023, with a unique barcode sequence 'NANNNTNTNNCNNNN' to facility correct identification by sequencing. For the wildtype barcoded TARDIS array, PX786 was used and described in Stevenson et al., 2023. Standard TARDIS-integrated was followed and several lineages were isolated. Several unique lineages were identified by Sanger sequencing, which founded the individual lineages.

Liquid culture, selection, and sample collection

To create our liquid environment, we used NGM buffer as our base (Leung et al., 2011), in addition, we used 100µg/ml Carbenicillin, 5µg/ml Cholesterol, 125µg/ml Hygromycin B, 10µg/ml Nyastatin, 10µM Ivermectin (diluted as needed for experimental molarity), and DMSO for the control, and $4x10^9$ PXKR1 (NA22 transformed with pUC19 for Carbenicillin resistance). Bacteria was grown in several large batches and measured for cell concentration before being frozen at -80C° until needed. Serial cultures were grown in 300ml volumes mixed with magnetic stir bars and transferred 10% of the population by volume every five days. Cultures were maintained at a constant 20C° in a temperature-controlled room. Population densities were estimated by counting six individual drops ranging from 2-5ul on a glass slide. In some cases, a 10X dilution was made to simplify the counts. Several 3ml samples were taken on the day of

transfer and frozen at -20C°. Samples were then processed for genomic DNA and barcode frequency as described in Stevenson et al., 2023.

Fitness Estimations with FitSeq and data analysis.

Barcode frequencies derived from Illumina sequencing (see Stevenson et al., 2023) were provided to PyFitSeq–a python implementation of FitSeq (Li et al., 2018). Briefly, FitSeq requires the user provide the approximate generation time per transfer, which was approximately one generation per transfer, along with estimated population sizes. Only lineages which survived to the end of the experiment were counted. Mutation selection coefficients were normalized to the average wildtype fitness. Lineage frequencies and population count trajectories were visualized with matplotlib 3.5.2 and data was analyzed with Python 3.7.13.

CHAPTER 7 CONCLUSION

Despite long-standing questions on the nature and distribution of adaptive mutations, our experimental ability to measure and quantify them remains challenging. However, the synergistic possibilities combining fundamental theories in evolutionary biology with high-throughput and high-precision genetic modifications is already providing paradigm-shifting experiments within the lab. We are now able to dissect the evolutionary process with unprecedented resolution and watch as the historical context of the lineage–the fundamental unit of evolution–is changing overtime. It is clear that with new advancements in our ability to modify the genome, we will continue to see novel and insightful experiments testing over 150 years of theory.

In Chapter 2, with my co-authors I reviewed the current state of experimental evolution with a specific focus on sexual dimorphism. We discussed several experimental categories that are now becoming possible combining newly developed sequencing and genetic engineering methodologies.

In Chapter 3, with my co-authors we created the first-ever CRISPR synthetic landing pad for transgenes for *C. elegans* with a particular focus on integration-specific selection. Prior technological methods have tried on anti-array selection techniques to provide confirmation of properly integrated transgenes. However, these methodologies are imperfect, relied on specialized equipment such as fluorescent microscopes, and were prone to false-positives. Our approach provided the first integration-specific selection of transgenes within an animal system. By flipping the transgene integration paradigm from trying to eliminate and screen away from incorrect candidates to selecting for correct candidates. Our approach significantly reduced the technical skill required to find transgenic candidates.

In Chapter 4 & 5, I presented along with my co-authors TARDIS (Transgenic Arrays Resulting in Diversity of Integrated Sequences), the first-ever animal based large-library transgenesis technology. TARDIS built upon our integration-specific selection first presented in Chapter 3. We took advantage of the *C. elegans* native ability to form large, heritable extra-chromosomal DNA and utilized it as an in-vivo DNA library. We build specialized synthetic landing pads to accept both barcodes and promoters for transcriptional reporter experiments. In this chapter, we also presented the first-ever animal based barcoding methodology similarly

adopted in microbial systems. We also presented in Chapter 5 our patent of the TARDIS technology, along with several possible future applications.

In Chapter 6, I presented along with my co-authors the first-ever barcoded animal experimental evolution. Utilizing TARDIS, we created several barcoded individuals in a wildtype and mutant backgrounds. Our mutant is a known ivermectin (IVM) resistant strain, which is known to have approximately over ~4000X resistance to IVM. We competed several lineages of both wildtype and mutants in various concentrations of IVM and used the barcodes to measure their lineage trajectories and fitness.

The research presented in this dissertation provides advancements in our technical capabilities to perform genome engineering and applied it to study novel evolutionary dynamics within the lab. The combination of both genetic engineering and evolutionary biology promises to provide novel insights into the evolutionary process with fundamental theories now empirically testable.

References Cited

Aljohani MD, Mouridi SE, Priyadarshini M, Vargas-Velazquez AM, Frøkjær-Jensen C. 2020. Engineering rules that minimize germline silencing of transgenes in simple extrachromosomal arrays in *C. elegans*. *Nature Communications* **11**:6300. doi:10.1038/s41467-020-19898-0

Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. 2020. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* **21**:30. doi:10.1186/s13059-020-1935-5

Arnqvist G, Rowe L. 2005. Sexual Conflict. Princeton (NY): Princeton University Press.

Arribere JA, Bell RT, Fu BXH, Artiles KL, Hartman PS, Fire AZ. 2014. Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in *Caenorhabditis elegans*. *Genetics* **198**:837–846. doi:10.1534/genetics.114.169730

Ba ANN, Cvijović I, Echenique JIR, Lawrence KR, Rego-Costa A, Liu X, Levy SF, Desai MM. 2019. High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature* **575**:494–499. doi:10.1038/s41586-019-1749-3

Bachtrog D, Kirkpatrick M, Mank JE, McDaniel SF, Pires JC, Rice W, Valenzuela N. 2011. Are all sex chromosomes created equal? *Trends Genet* **27**:350–357. doi:10.1016/j.tig.2011.05.005

Bachtrog D, Mahajan S, Bracewell R. 2019. Massive gene amplification on a recently formed *Drosophila* Y chromosome. *bioRxiv* 451005. doi:10.1101/451005

Barrett RDH, Hoekstra HE. 2011. Molecular spandrels: tests of adaptation at the genetic level. *Nat Rev Genet* **12**:767–780. doi:10.1038/nrg3015

Bates S, Sesia M, Sabatti C, Candès E. 2020. Causal inference in genetic trio studies. *Proc Natl Acad Sci* **117**:24117–24126. doi:10.1073/pnas.2007743117

Baym M, Lieberman TD, Kelsic ED, Chait R, Gross R, Yelin I, Kishony R. 2016. Spatiotemporal microbial evolution on antibiotic landscapes. *Science* **353**:1147–1151. doi:10.1126/science.aag0822

Becht E, Zhao E, Amezquita R, Gottardo R. 2020. Aggregating transcript-level analyses for single-cell differential gene expression. *Nat Methods* **17**:583–585. doi:10.1038/s41592-020-0854-4

Ben-David E, Boocock J, Guo L, Zdraljevic S, Bloom JS, Kruglyak L. 2020. Whole-organism eQTL mapping at cellular resolution with single-cell sequencing. *bioRxiv* 2020.08.23.263798. doi:10.1101/2020.08.23.263798

Beumer KJ, Trautman JK, Bozas A, Liu J-L, Rutter J, Gall JG, Carroll D. 2008. Efficient gene targeting in *Drosophila* by direct embryo injection with zinc-finger nucleases. *Proc National Acad Sci* **105**:19821–19826. doi:10.1073/pnas.0810475105

Bezler A, Braukmann F, West SM, Duplan A, Conconi R, Schütz F, Gönczy P, Piano F, Gunsalus K, Miska EA, Keller L. 2019. Tissue- and sex-specific small RNAomes reveal sex differences in response to the environment. *PLoS Genet* **15**:e1007905. doi:10.1371/journal.pgen.1007905

Bijleveld AI, Twietmeyer S, Piechocki J, Gils JA van, Piersma T. 2015. Natural selection by pulsed predation: survival of the thickest. *Ecology* **96**:1943–1956. doi:10.1890/14-1845.1

Bindels DS, Haarbosch L, Weeren L van, Postma M, Wiese KE, Mastop M, Aumonier S, Gotthard G, Royant A, Hink MA, Gadella TWJ. 2017. mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nat Methods* **14**:53–56. doi:10.1038/nmeth.4074

Blundell JR, Levy SF. 2014. Beyond genome sequencing: Lineage tracking with barcodes to study the dynamics of evolution, infection, and cancer. *Genomics* **104**:1–14. doi:10.1016/j.ygeno.2014.09.005

Bock C, Datlinger P, Chardon F, Coelho MA, Dong MB, Lawson KA, Lu T, Maroc L, Norman TM, Song B, Stanley G, Chen S, Garnett M, Li W, Moffat J, Qi LS, Shapiro RS, Shendure J, Weissman JS, Zhuang X. 2022. High-content CRISPR screening. *Nature Reviews Methods Primers* **2**:8. doi:10.1038/s43586-021-00093-4

Boer CG de, Vaishnav ED, Sadeh R, Abeyta EL, Friedman N, Regev A. 2020. Author Correction: Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nature Biotechnology* **38**:1211–1211. doi:10.1038/s41587-020-0665-2

Böhne A, Sengstag T, Salzburger W. 2014. Comparative transcriptomics in East African Cichlids reveals sex- and species-specific expression and new candidates for sex differentiation in fishes. *Genome Biol Evol* **6**:2567–2585. doi:10.1093/gbe/evu200

Bonduriansky R, Chenoweth SF. 2009. Intralocus sexual conflict. *Trends Ecol Evol* **24**:280–288. doi:10.1016/j.tree.2008.12.005

Böttcher R, Hollmann M, Merk K, Nitschko V, Obermaier C, Philippou-Massier J, Wieland I, Gaul U, Förstemann K. 2014. Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured *Drosophila* cells. *Nucleic Acids Res* **42**:e89–e89. doi:10.1093/nar/gku289

Boutros M, Ahringer J. 2008. The art and design of genetic screens: RNA interference. *Nat Rev Genet* **9**:554–566. doi:10.1038/nrg2364

Bracewell R, Bachtrog D. 2020. Complex evolutionary history of the Y chromosome in flies of the *Drosophila obscura* species group. *Genome Biol Evol* **12**:494–505. doi:10.1093/gbe/evaa051

Brauer MJ, Christianson CM, Pai DA, Dunham MJ. 2006. Mapping novel traits by array-assisted bulk segregant analysis in *Saccharomyces cerevisiae*. *Genetics* **173**:1813–1816. doi:10.1534/genetics.106.057927

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**:1213–1218. doi:10.1038/nmeth.2688

Burch CL, Guyader S, Samarov D, Shen H. 2007. Experimental estimate of the abundance and effects of nearly neutral mutations in the RNA virus ϕ6. *Genetics* **176**:467–476. doi:10.1534/genetics.106.067199

Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**:411–420. doi:10.1038/nbt.4096

Carlton PM, Davis RE, Ahmed S. 2022. Nematode chromosomes. *Genetics* **221**. doi:10.1093/genetics/iyac014

Chang PL, Dunham JP, Nuzhdin SV, Arbeitman MN. 2011. Somatic sex-specific transcriptome differences in *Drosophila* revealed by whole transcriptome sequencing. *BMC Genom* **12**:364. doi:10.1186/1471-2164-12-364

Charlesworth B. 1991. The evolution of sex chromosomes. *Science* **251**:1030–1033. doi:10.1126/science.1998119

Charlesworth B, Charlesworth D, Coyne JA, Langley CH. 2016. Hubby and Lewontin on protein variation in natural populations: When molecular genetics came to the rescue of population genetics. *Genetics* **203**:1497–1503. doi:10.1534/genetics.115.185975

Chen C, Fenk LA, Bono M de. 2013. Efficient genome editing in *Caenorhabditis elegans* by CRISPR-targeted homologous recombination. *Nucleic Acids Res* **41**:e193–e193. doi:10.1093/nar/gkt805

Chen C, Liao Y, Peng G. 2022. Connecting past and present: single-cell lineage tracing. *Protein Cell* **13**:790–807. doi:10.1007/s13238-022-00913-7

Chen DS, Delbare SYN, White SL, Sitnik J, Chatterjee M, DoBell E, Weiss O, Clark AG, Wolfner MF. 2019. Female genetic contributions to sperm competition in *Drosophila melanogaster*. *Genetics* **212**:789–800. doi:10.1534/genetics.119.302284

Chen N, Juric I, Cosgrove EJ, Bowman R, Fitzpatrick JW, Schoech SJ, Clark AG, Coop G. 2019. Allele frequency dynamics in a pedigreed natural population. *Proc Natl Acad Sci* **116**:2158–2164. doi:10.1073/pnas.1813852116

Chen S, Ni X, Krinsky BH, Zhang YE, Vibranovski MD, White KP, Long M. 2012. Reshaping of global gene expression networks and sex-biased gene expression by integration of a young gene. *EMBO J* **31**:2798–2809. doi:10.1038/emboj.2012.108

Chenoweth SF, Rundle HD, Blows MW. 2008. Genetic constraints and the evolution of display trait sexual dimorphism by natural and sexual selection. *Am Nat* **171**:22–34. doi:10.1086/523946

Clement K, Farouni R, Bauer DE, Pinello L. 2018. AmpUMI: design and analysis of unique molecular identifiers for deep amplicon sequencing. *Bioinformatics* **34**:i202–i210. doi:10.1093/bioinformatics/bty264

Combs PA, Eisen MB. 2013. Sequencing mRNA from cryo-sliced *Drosophila* embryos to determine genome-wide spatial patterns of gene expression. *PLoS ONE* **8**:e71820. doi:10.1371/journal.pone.0071820

Connallon T, Clark AG. 2011. The resolution of sexual antagonism by gene duplication. *Genetics* **187**:919–937. doi:10.1534/genetics.110.123729

Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ. 2018. Corrigendum: Tempo and mode of regulatory evolution in *Drosophila*. *Genome Res* **28**:1766–1766. doi:10.1101/gr.244087.118

Crombie TA, Saber S, Saxena AS, Egan R, Baer CF. 2018. Head-to-head comparison of three experimental methods of quantifying competitive fitness in *C. elegans*. *PLoS ONE* **13**:e0201507. doi:10.1371/journal.pone.0201507

Crotty S, Pipkin ME. 2015. In vivo RNAi screens: concepts and applications. *Trends Immunol* **36**:315–322. doi:10.1016/j.it.2015.03.007

Crow JF, Temin RG. 1964. Evidence for the partial dominance of recessive lethal genes in natural populations of drosophila. *Am Nat* **98**:21–33. doi:10.1086/282298

Cunningham BC, Wells JA. 1989. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* **244**:1081–1085. doi:10.1126/science.2471267

Cutter AD, Ward S. 2005. Sexual and temporal dynamics of molecular evolution in *C. elegans* development. *Mol Biol Evol* **22**:178–188. doi:10.1093/molbev/msh267

Darwin C. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. London: J.Murray.

Daugherty AC, Yeo RW, Buenrostro JD, Greenleaf WJ, Kundaje A, Brunet A. 2017. Chromatin accessibility dynamics reveal novel functional enhancers in *C. elegans*. *Genome Res* **27**:2096–2107. doi:10.1101/gr.226233.117

Delvigne F, Pêcheux H, Tarayre C. 2015. Fluorescent reporter libraries as useful tools for optimizing microbial cell factories: A review of the current methods and applications. *Frontiers in Bioengineering and Biotechnology* **3**. doi:10.3389/fbioe.2015.00147

Dent JA, Smith MM, Vassilatis DK, Avery L. 2000. The genetics of ivermectin resistance in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences* **97**:2674–2679. doi:10.1073/pnas.97.6.2674

Desai MM, Walczak AM, Fisher DS. 2013. Genetic diversity and the structure of genealogies in rapidly adapting populations. *Genetics* **193**:565–585. doi:10.1534/genetics.112.147157

Deursen JV, Fornerod M, Rees BV, Grosveld G. 1995. Cre-mediated site-specific translocation between nonhomologous mouse chromosomes. *Proc Natl Acad Sci* **92**:7376–7380. doi:10.1073/pnas.92.16.7376

Dickinson DJ, Goldstein B. 2016. CRISPR-based methods for *Caenorhabditis elegans* genome engineering. *Genetics* **202**:885–901. doi:10.1534/genetics.115.182162

Dickinson DJ, Pani AM, Heppert JK, Higgins CD, Goldstein B. 2015. Streamlined genome engineering with a self-excising drug selection cassette. *Genetics* **200**:1035–1049. doi:10.1534/genetics.115.178335

Dickinson DJ, Ward JD, Reiner DJ, Goldstein B. 2013. Engineering the *Caenorhabditis elegans* genome using Cas9-triggered homologous recombination. *Nat Methods* **10**:1028–1034. doi:10.1038/nmeth.2641

Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, Virgin HW, Listgarten J, Root DE. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**:184–191. doi:10.1038/nbt.3437

Doudna JA, Charpentier E. 2014. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**:1258096. doi:10.1126/science.1258096

Douglas C, Maciulyte V, Zohren J, Snell DM, Ojarikre OA, Ellis PJI, Turner JMA. 2020. Generating single-sex litters: development of CRISPR-Cas9 genetic tools to produce all-male offspring. *bioRxiv* 2020.09.07.285536. doi:10.1101/2020.09.07.285536

Dzitoyeva S, Dimitrijevic N, Manev H. 2001. Intra-abdominal injection of double-stranded RNA into anesthetized adult *Drosophila* triggers RNA interference in the central nervous system. *Mol Psychiatry* **6**:665–670. doi:10.1038/sj.mp.4000955

Ebbing A, Vértesy Á, Betist MC, Spanjaard B, Junker JP, Berezikov E, Oudenaarden A van, Korswagen HC. 2018. Spatial transcriptomics of *C. elegans* males and hermaphrodites identifies sex-specific differences in gene expression patterns. *Dev Cell* **47**:801-813.e6. doi:10.1016/j.devcel.2018.10.016

Edward DA, Fricke C, Chapman T. 2010. Adaptations to sexual selection and sexual conflict: insights from experimental evolution and artificial selection. *Philos Trans R Soc B: Biol Sci* **365**:2541–2548. doi:10.1098/rstb.2010.0027

Elsner M. 2018. Barcodes galore for developmental biology. *Nat Biotechnol* **36**:936–936. doi:10.1038/nbt.4269

Emanuel G, Moffitt JR, Zhuang X. 2017. High-throughput, image-based screening of pooled genetic-variant libraries. *Nature Methods* **14**:1159–1162. doi:10.1038/nmeth.4495

Erwood S, Bily TMI, Lequyer J, Yan J, Gulati N, Brewer RA, Zhou L, Pelletier L, Ivakine EA, Cohn RD. 2022. Saturation variant interpretation using CRISPR prime editing. *Nature Biotechnology* **40**:885–895. doi:10.1038/s41587-021-01201-1

Evans T. 2006. Transformation and microinjection. *WormBook*. doi:10.1895/wormbook.1.108.1

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet* **8**:610–618. doi:10.1038/nrg2146

Farboud B, Meyer BJ. 2015. Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design. *Genetics* **199**:959–971. doi:10.1534/genetics.115.175166

Farboud B, Severson AF, Meyer BJ. 2019. Strategies for Efficient Genome Editing Using CRISPR-Cas9. *Genetics* **211**:431–457. doi:10.1534/genetics.118.301775

Farrell JA, Wang Y, Riesenfeld SJ, Shekhar K, Regev A, Schier AF. 2018. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**. doi:10.1126/science.aar3131

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**:806–811. doi:10.1038/35888

Froehlich JJ, Uyar B, Herzog M, Theil K, Glažar P, Akalin A, Rajewsky N. 2021. Parallel genetics of regulatory sequences using scalable genome editing in vivo. *Cell Reports* **35**:108988. doi:10.1016/j.celrep.2021.108988

Frøkjær-Jensen C, Davis MW, Ailion M, Jorgensen EM. 2012. Improved Mos1-mediated transgenesis in *C. elegans*. *Nat Methods* **9**:117–118. doi:10.1038/nmeth.1865

Frøkjær-Jensen C, Davis MW, Hopkins CE, Newman BJ, Thummel JM, Olesen S-P, Grunnet M, Jorgensen EM. 2008. Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nat Genet* **40**:1375–1383. doi:10.1038/ng.248

Frøkjær-Jensen C, Davis MW, Sarov M, Taylor J, Flibotte S, LaBella M, Pozniakovsky A, Moerman DG, Jorgensen EM. 2014. Random and targeted transgene insertion in *Caenorhabditis elegans* using a modified Mos1 transposon. *Nature Methods* **11**:529–534. doi:10.1038/nmeth.2889

Gallach M, Betrán E. 2011. Intralocus sexual conflict resolved through gene duplication. *Trends Ecol Evol* **26**:222–228. doi:10.1016/j.tree.2011.02.004

Gallach M, Chandrasekaran C, Betrán E. 2010. Analyses of nuclearly encoded mitochondrial genes suggest gene duplication as a mechanism for resolving intralocus sexually antagonistic conflict in *Drosophila*. *Genome Biol Evol* **2**:835–850. doi:10.1093/gbe/evq069

Geerts S, Gryseels B. 2000. Drug resistance in human helminths: Current situation and lessons from livestock. *Clin Microbiol Rev* **13**:207–222. doi:10.1128/cmr.13.2.207

Gilleland CL, Falls AT, Noraky J, Heiman MG, Yanik MF. 2015. Computer-assisted transgenesis of *Caenorhabditis elegans* for deep phenotyping. *Genetics* **201**:39–46. doi:10.1534/genetics.115.179648

Giordano-Santini R, Milstein S, Svrzikapa N, Tu D, Johnsen R, Baillie D, Vidal M, Dupuy D. 2010. An antibiotic selection marker for nematode transgenesis. *Nat Methods* **7**:721–723. doi:10.1038/nmeth.1494

Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. 2017. The dynamics of molecular evolution over 60,000 generations. *Nature* **551**:45–50. doi:10.1038/nature24287

Grath S, Parsch J. 2015. Sex-biased gene expression. *Annu Rev Genet* **50**:1–16. doi:10.1146/annurev-genet-120215-035429

Grégoire D, Kmita M. 2008. Recombination between inverted loxP sites is cytotoxic for proliferating cells and provides a simple tool for conditional cell ablation. *Proc Natl Acad Sci* **105**:14492–14496. doi:10.1073/pnas.0807484105

Gu H, Marth JD, Orban PC, Mossmann H, Rajewsky K. 1994. Deletion of a DNA polymerase β gene segment in T cells using cell type-specific gene targeting. *Science* **265**:103–106. doi:10.1126/science.8016642

Hammond AM, Kyrou K, Bruttini M, North A, Galizi R, Karlsson X, Kranjc N, Carpi FM, D'Aurizio R, Crisanti A, Nolan T. 2017. The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLoS Genet* **13**:e1007039. doi:10.1371/journal.pgen.1007039

Haque A, Engel J, Teichmann SA, Lönnberg T. 2017. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med* **9**:75. doi:10.1186/s13073-017-0467-4

Harrison PW, Wright AE, Zimmer F, Dean R, Montgomery SH, Pointer MA, Mank JE. 2015. Sexual selection drives evolution and rapid turnover of male gene expression. *Proc Natl Acad Sci* **112**:4393–4398. doi:10.1073/pnas.1501339112

Hartmann B, Castelo R, Miñana B, Peden E, Blanchette M, Rio DC, Singh R, Valcárcel J. 2011. Distinct regulatory programs establish widespread sex-specific alternative splicing in *Drosophila melanogaster*. *RNA* **17**:453–468. doi:10.1261/rna.2460411

Hashimshony T, Senderovich N, Avital G, Klochendler A, Leeuw Y de, Anavy L, Gennert D, Li S, Livak KJ, Rozenblatt-Rosen O, Dor Y, Regev A, Yanai I. 2016. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol* **17**:77. doi:10.1186/s13059-016-0938-8

Hegreness M, Shoresh N, Hartl D, Kishony R. 2006. An equivalence principle for the incorporation of favorable mutations in asexual populations. *Science* **311**:1615–1617. doi:10.1126/science.1122469

Hirrlinger J, Scheller A, Hirrlinger PG, Kellert B, Tang W, Wehr MC, Goebbels S, Reichenbach A, Sprengel R, Rossner MJ, Kirchhoff F. 2009. Split-Cre complementation indicates coincident activity of different genes in vivo. *PLoS ONE* **4**:e4286. doi:10.1371/journal.pone.0004286

Holland AJ, Fachinetti D, Han JS, Cleveland DW. 2012. Inducible, reversible system for the rapid and complete degradation of proteins in mammalian cells. *Proc Natl Acad Sci* **109**:E3350–E3357. doi:10.1073/pnas.1216880109

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G, Zhang F. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**:827–832. doi:10.1038/nbt.2647

Hsu S-K, Jakšić AM, Nolte V, Lirakis M, Kofler R, Barghi N, Versace E, Schlötterer C. 2020. Rapid sex-specific adaptation to high temperature in *Drosophila*. *eLife* **9**:e53237. doi:10.7554/elife.53237

Hubby JL, Lewontin RC. 1966. A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* **54**:577–594. doi:10.1093/genetics/54.2.577

Hunter JD. 2007. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* **9**:90–95. doi:10.1109/mcse.2007.55

Huxley J. 1942. Evolution: The modern synthesis. New York ; London: Harper & brothers.

Innocenti P, Morrow EH. 2010. The sexually antagonistic genes of *Drosophila melanogaster*. *PLoS Biol* **8**:e1000335. doi:10.1371/journal.pbio.1000335

Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol* **169**:5429–5433. doi:10.1128/jb.169.12.5429-5433.1987

Islam S, Kjällquist U, Moliner A, Zajac P, Fan J-B, Lönnerberg P, Linnarsson S. 2011. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res* **21**:1160–1167. doi:10.1101/gr.110882.110

Ismagul A, Yang N, Maltseva E, Iskakova G, Mazonka I, Skiba Y, Bi H, Eliby S, Jatayev S, Shavrukov Y, Borisjuk N, Langridge P. 2018. A biolistic method for high-throughput production of transgenic wheat plants with single gene insertions. *BMC Plant Biology* **18**:135. doi:10.1186/s12870-018-1326-1

Jahn LJ, Porse A, Munck C, Simon D, Volkova S, Sommer MOA. 2018. Chromosomal barcoding as a tool for multiplexed phenotypic characterization of laboratory evolved lineages. *Scientific Reports* **8**:6961. doi:10.1038/s41598-018-25201-5

Jasinska W, Manhart M, Lerner J, Gauthier L, Serohijos AWR, Bershtein S. 2020. Chromosomal barcoding of *E. coli* populations reveals lineage diversity dynamics at high resolution. *Nat Ecol Evol* 1–16. doi:10.1038/s41559-020-1103-z

Jensen JD, Payseur BA, Stephan W, Aquadro CF, Lynch M, Charlesworth D, Charlesworth B. 2019. The importance of the Neutral Theory in 1968 and 50 years on: A response to Kern and Hahn 2018. *Evolution* **73**:111–114. doi:10.1111/evo.13650

Jiang S, Mortazavi A. 2018. Integrating ChIP-seq with other functional genomics data. *Brief Funct Genom* **17**:104–115. doi:10.1093/bfgp/ely002

Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G. 2001. The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat Genet* **29**:389–395. doi:10.1038/ng766

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**:816–821. doi:10.1126/science.1225829

Johnston SE, Huisman J, Ellis PA, Pemberton JM. 2017. A high-density linkage map reveals sexual dimorphism in recombination landscapes in Red Deer (*Cervus elaphus*). *G3: GenesGenomesGenet* **7**:2859–2870. doi:10.1534/g3.117.044198

Jorgensen EM, Mango SE. 2002. The art and design of genetic screens: *Caenorhabditis elegans*. *Nat Rev Genet* **3**:356–369. doi:10.1038/nrg794

Joung JK, Ramm EI, Pabo CO. 2000. A bacterial two-hybrid selection system for studying protein–DNA and protein–protein interactions. *Proceedings of the National Academy of Sciences* **97**:7382–7387. doi:10.1073/pnas.110149297

Joung JK, Sander JD. 2013. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* **14**:49–55. doi:10.1038/nrm3486

Junker JP, Noël ES, Guryev V, Peterson KA, Shah G, Huisken J, McMahon AP, Berezikov E, Bakkers J, van Oudenaarden A. 2014. Genome-wide RNA tomography in the zebrafish embryo. *Cell* **159**:662–675. doi:10.1016/j.cell.2014.09.038

Juntti SA, Coats JK, Shah NM. 2008. A genetic approach to dissect sexually dimorphic behaviors. *Horm Behav* **53**:627–637. doi:10.1016/j.yhbeh.2007.12.012

Kang K, Song Y, Kim I, Kim T-J. 2022. Therapeutic applications of the CRISPR-Cas system. *Bioengineering* **9**:477. doi:10.3390/bioengineering9090477

Kanke M, Nishimura K, Kanemaki M, Kakimoto T, Takahashi TS, Nakagawa T, Masukata H. 2011. Auxin-inducible protein depletion system in fission yeast. *BMC Cell Biol* **12**:8. doi:10.1186/1471-2121-12-8

Kasimatis KR, Moerdyk-Schauwecker MJ, Lancaster R, Smith A, Willis JH, Phillips PC. 2022. Post-insemination selection dominates pre-insemination selection in driving rapid evolution of male competitive ability. *PLoS Genet* **18**:e1010063. doi:10.1371/journal.pgen.1010063

Kasimatis KR, Moerdyk-Schauwecker MJ, Phillips PC. 2018. Auxin-mediated sterility induction system for longevity and mating studies in *Caenorhabditis elegans*. *G3: GenesGenomesGenet* **8**:2655–2662. doi:10.1534/g3.118.200278

Kasimatis KR, Sánchez-Ramírez S, Stevenson ZC. 2021. Sexual dimorphism through the lens of genome manipulation, forward genetics, and spatiotemporal sequencing. *Genome Biology and Evolution* **13**. doi:10.1093/gbe/evaa243

Kaymak E, Farley BM, Hay SA, Li C, Ho S, Hartman DJ, Ryder SP. 2016. Efficient generation of transgenic reporter strains and analysis of expression patterns in *Caenorhabditis elegans* using library MosSCI. *Developmental Dynamics* **245**:925–936. doi:10.1002/dvdy.24426

Kebschull JM, Zador AM. 2018. Cellular barcoding: lineage tracing, screening and beyond. *Nature Methods* **15**:871–879. doi:10.1038/s41592-018-0185-x

Kemp BJ, Hatzold J, Sternick LA, Cornman-Homonoff J, Whitaker JM, Tieu PJ, Lambie EJ. 2007. In vivo construction of recombinant molecules within the *Caenorhabditis elegans* germ line using short regions of terminal homology. *Nucleic Acids Res* **35**:e133–e133. doi:10.1093/nar/gkm857

Kern AD, Hahn MW. 2018. The neutral theory in light of natural selection. *Molecular Biology and Evolution* **35**:1366–1371. doi:10.1093/molbev/msy092

Khericha M, Kolenchery JB, Tauber E. 2016. Neural and non-neural contributions to sexual dimorphism of mid-day sleep in *Drosophila melanogaster*: A pilot study. *Physiol Èntomol* **41**:327–334. doi:10.1111/phen.12134

Khila A, Abouheif E, Rowe L. 2012. Function, developmental genetics, and fitness consequences of a sexually antagonistic trait. *Science* **336**:585–589. doi:10.1126/science.1217258

Khramtsova EA, Davis LK, Stranger BE. 2019. The role of sex in the genomics of human complex traits. *Nat Rev Genet* **20**:173–190. doi:10.1038/s41576-018-0083-1

Kim H, Ishidate T, Ghanta KS, Seth M, Conte D, Shirayama M, Mello CC. 2014. A so-CRISPR strategy for efficient genome editing in *Caenorhabditis elegans*. *Genetics* **197**:1069–1080. doi:10.1534/genetics.114.166389

Kim Y, Lee S, Cho S, Park J, Chae D, Park T, Minna JD, Kim HH. 2022. High-throughput functional evaluation of human cancer-associated mutations using base editors. *Nature Biotechnology* **40**:874–884. doi:10.1038/s41587-022-01276-4

Kimura M. 1968. Evolutionary rate at the molecular level. *Nature* **217**:624–6.

Kimura M, Ohta T. 1969. The average number of generations until extinction of an individual mutant gene in a finite population. *Genetics* **63**:701–709. doi:10.1093/genetics/63.3.701

King M-C, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* **188**:107–116. doi:10.1126/science.1090005

Klug A. 2010. The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu Rev Biochem* **79**:213–231. doi:10.1146/annurev-biochem-010909-095056

Kobak D, Berens P. 2019. The art of using t-SNE for single-cell transcriptomics. *Nat Commun* **10**:5416. doi:10.1038/s41467-019-13056-x

Kofler R, Schlötterer C. 2014. A guide for the design of evolve and resequencing studies. *Mol Biol Evol* **31**:474–483. doi:10.1093/molbev/mst221

Kotwica-Rolinska J, Chodakova L, Chvalova D, Kristofova L, Fenclova I, Provaznik J, Bertolutti M, Wu BC-H, Dolezel D. 2019. CRISPR/Cas9 genome editing introduction and optimization in the non-model insect *Pyrrhocoris apterus*. *Front Physiol* **10**:891. doi:10.3389/fphys.2019.00891

Kress WJ, García-Robledo C, Uriarte M, Erickson DL. 2015. DNA barcodes for ecology, evolution, and conservation. *Trends Ecol Evol* **30**:25–35. doi:10.1016/j.tree.2014.10.008

Kruse F, Junker JP, Oudenaarden A van, Bakkers J. 2016. Chapter 15 Tomo-seq A method to obtain genome-wide expression data with spatial resolution. *Methods Cell Biol* **135**:299–307. doi:10.1016/bs.mcb.2016.01.006

Larson EL, Vanderpool D, Keeble S, Zhou M, Sarver BAJ, Smith AD, Dean MD, Good JM. 2016. Contrasting levels of molecular evolution on the mouse X chromosome. *Genetics* **203**:1841–1857. doi:10.1534/genetics.116.186825

Lenski RE. 2017. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *The ISME Journal* **11**. doi:10.1038/ismej.2017.69

Lenski RE, Rose MR, Simpson SC, Tadler SC. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am Nat* **138**:1315–1341. doi:10.1086/285289

Leung CK, Deonarine A, Strange K, Choe KP. 2011. High-throughput screening and biosensing with fluorescent *C. elegans* strains. *Journal of Visualized Experiments*. doi:10.3791/2745

Levine M, Tjian R. 2003. Transcription regulation and animal diversity. *Nature* **424**:147–151. doi:10.1038/nature01763

Levinson G, A. GG. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution*. doi:10.1093/oxfordjournals.molbev.a040442

Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS, Sherlock G. 2015. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **519**:181–186. doi:10.1038/nature14279

Levy SL Xianan. 2016. Genomic combinatorial screening platform [international patent]. WO2017075529A1.

Lewontin RC, Hubby JL. 1966. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* **54**:595–609.

Li F, Salit ML, Levy SF. 2018. Unbiased fitness estimation of pooled barcode or amplicon sequencing studies. *Cell Systems* **7**. doi:10.1016/j.cels.2018.09.004

Li F, Vensko SP, Belikoff EJ, Scott MJ. 2013. Conservation and sex-specific splicing of the transformer gene in the *Calliphorids Cochliomyia hominivorax*, *Cochliomyia macellaria* and *Lucilia sericata*. *PLoS ONE* **8**:e56303. doi:10.1371/journal.pone.0056303

Lin Z, Xie Y, Nong W, Ren X, Li R, Zhao Z, Hui JHL, Yuen KWY. 2021. Formation of artificial chromosomes in *Caenorhabditis elegans* and analyses of their segregation in mitosis, DNA sequence composition and holocentromere organization. *Nucleic Acids Research* **49**:9174–9193. doi:10.1093/nar/gkab690

Liu B, Jing Z, Zhang X, Chen Y, Mao S, Kaundal R, Zou Y, Wei G, Zang Y, Wang X, Lin W, Di M, Sun Y, Chen Q, Li Y, Xia J, Sun J, Lin C-P, Huang X, Chi T. 2022. Large-scale multiplexed mosaic CRISPR perturbation in the whole organism. *Cell* **185**:3008-3024.e16. doi:10.1016/j.cell.2022.06.039

Livet J, Weissman TA, Kang H, Draft RW, Lu J, Bennis RA, Sanes JR, Lichtman JW. 2007. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**:56–62. doi:10.1038/nature06293

Loewe L, Charlesworth B. 2006. Inferring the distribution of mutational effects on fitness in *Drosophila*. *Biol Lett* **2**:426–430. doi:10.1098/rsbl.2006.0481

Logan ML, Cox RM, Calsbeek R. 2014. Natural selection on thermal performance in a novel thermal environment. *Proc Natl Acad Sci* **111**:14165–14169. doi:10.1073/pnas.1404885111

Lucotte EA, Albiñana C, Laurent R, Bhérer C, Consortium G of the N, Bataillon T, Toupance B. 2020. Detection of sexually antagonistic transmission distortions in trio datasets. *bioRxiv* 2020.09.11.293191. doi:10.1101/2020.09.11.293191

Lutgen D, Ritter R, Olsen R, Schielzeth H, Gruselius J, Ewels P, García JT, Shirihai H, Schweizer M, Suh A, Burri R. 2020. Linked-read sequencing enables haplotype-resolved resequencing at population scale. *Mol Ecol Resour* **20**:1311–1322. doi:10.1111/1755-0998.13192

Maklakov AA, Bonduriansky R, Brooks RC. 2009. Sex differences, sexual selection, and ageing: An experimental evolution approach. *Evolution* **63**:2491–2503. doi:10.1111/j.1558-5646.2009.00750.x

Malaiwong N, Porta-de-la-Riva M, Krieg M. 2023. FLInt: Single shot safe harbor transgene Integration via fluorescent landmark interference. *G3 Genes|Genomes|Genetics*. doi:10.1093/g3journal/jkad041

Mank JE. 2017. The transcriptional architecture of phenotypic dimorphism. *Nat Ecol Evol* **1**:0006. doi:10.1038/s41559-016-0006

Mank JE. 2009. Sex chromosomes and the evolution of sexual dimorphism: Lessons from the genome. *Am Nat* **173**:141–150. doi:10.1086/595754

Marie-Orleach L, Janicke T, Vizoso DB, David P, Schärer L. 2016. Quantifying episodes of sexual selection: Insights from a transparent worm with fluorescent sperm. *Evolution* **70**:314–328. doi:10.1111/evo.12861

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**:10. doi:10.14806/ej.17.1.200

Matthews BW. 1996. Structural and genetic analysis of the folding and function of T4 lysozyme. *The FASEB Journal* **10**:35–41. doi:10.1096/fasebj.10.1.8566545

McDiarmid TA, Au V, Moerman DG, Rankin CH. 2020. Peel-1 negative selection promotes screening-free CRISPR-Cas9 genome editing in *Caenorhabditis elegans*. *PLOS ONE* **15**:e0238950. doi:10.1371/journal.pone.0238950

McKenna A, Findlay GM, Gagnon JA, Horwitz MS, Schier AF, Shendure J. 2016. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**. doi:10.1126/science.aaf7907

McKINNEY HL. 1966. Alfred Russel Wallace and the discovery of natural selection. *J Hist Med Allied Sci* **XXI**:333–357. doi:10.1093/jhmas/xxi.4.333

Meiklejohn CD, Coolon JD, Hartl DL, Wittkopp PJ. 2014. The roles of cis- and trans-regulation in the evolution of regulatory incompatibilities and sexually dimorphic gene expression. *Genome Res* **24**:84–95. doi:10.1101/gr.156414.113

Mello C, Fire A. 1995. Chapter 19 DNA Transformation Methods in Cell Biology. pp. 451–482. doi:10.1016/s0091-679x(08)61399-0

Mello C, Kramer J, Stinchcomb D, Ambros V. 1991. Efficient gene transfer in *C.elegans*: Extrachromosomal maintenance and integration of transforming sequences. *The EMBO journal* **10**:3959–70.

Meneely PM, Dahlberg CL, Rose JK. 2019. Working with worms: *Caenorhabditis elegans* as a model organism. *Curr Protoc Essent Lab Tech* **19**. doi:10.1002/cpet.35

Miyazaki S, Okada Y, Miyakawa H, Tokuda G, Cornette R, Koshikawa S, Maekawa K, Miura T. 2014. Sexually dimorphic body color Is regulated by sex-specific expression of yellow gene in ponerine ant, *diacamma* sp. *PLoS ONE* **9**:e92875. doi:10.1371/journal.pone.0092875

Morris SW. 1994. Fleeming Jenkin and the origin of species: a reassessment. *Br J Hist Sci* **27**:313–343. doi:10.1017/s0007087400032209

Mouridi E, Mouridi SE, Alkhaldi F, Frøkjaer-Jensen C. 2022. Modular safe-harbor transgene insertion (MosTI) for targeted single-copy and extrachromosomal array integration in *C. elegans*. doi:10.1101/2022.04.19.488726

Mouridi S, Lecroisey C, Tardy P, Mercier M, Leclercq-Blondel A, Zariohi N, Boulin T. 2017. Reliable CRISPR/Cas9 genome engineering in *Caenorhabditis elegans* using a single efficient sgRNA and an easily recognizable phenotype. *G3: Genes, Genomes, Genetics* **7**:1429–1437. doi:10.1534/g3.117.040824

Muller HJ. 1927. Artificial transmutation of the gene. *Science* **66**:84–87. doi:10.1126/science.66.1699.84

Naqvi S, Godfrey AK, Hughes JF, Goodheart ML, Mitchell RN, Page DC. 2019. Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science* **365**. doi:10.1126/science.aaw7317

Nishimura K, Fukagawa T, Takisawa H, Kakimoto T, Kanemaki M. 2009. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods* **6**:917–922. doi:10.1038/nmeth.1401

Nonet M. 2021. Additional landing sites for recombination-mediated cassette exchange in *C. elegans*. *microPublication biology* **2021**. doi:10.17912/micropub.biology.000503

Nonet ML. 2020. Efficient Transgenesis in *Caenorhabditis elegans* Using Flp Recombinase-Mediated Cassette Exchange. *Genetics* **215**:903–921. doi:10.1534/genetics.120.303388

Ntranos V, Yi L, Melsted P, Pachter L. 2019. A discriminative learning approach to differential expression analysis for single-cell RNA-seq. *Nat Methods* **16**:163–166. doi:10.1038/s41592-018-0303-9

Obinata H, Sugimoto A, Niwa S. 2018. Streptothricin acetyl transferase 2 (Sat2): A dominant selection marker for *Caenorhabditis elegans* genome editing. *Plos One* **13**:e0197128. doi:10.1371/journal.pone.0197128

O'Connell RW, Rai K, Piepergerdes TC, Wang Y, Samra KD, Wilson JA, Lin S, Zhang TH, Ramos EM, Sun A, Kille B, Curry KD, Rocks JW, Treangen TJ, Mehta P, Bashor CJ. 2023. Ultra-high throughput mapping of genetic design space. *bioRxiv* 2023.03.16.532704. doi:10.1101/2023.03.16.532704

Oliva M, Muñoz-Aguirre M, Kim-Hellmuth S, Wucher V, Gewirtz ADH, Cotter DJ, Parsana P, Kasela S, Balliu B, Viñuela A, Castel SE, Mohammadi P, Aguet F, Zou Y, Khramtsova EA, Skol AD, Garrido-Martín D, Reverter F, Brown A, Evans P, Gamazon ER, Payne A, Bonazzola R, Barbeira AN, Hamel AR, Martinez-Perez A, Soria JM, Consortium Gte, Pierce BL, Stephens M, Eskin E, Dermitzakis ET, Segrè AV, Im HK, Engelhardt BE, Ardlie KG, Montgomery SB, Battle AJ, Lappalainen T, Guigó R, Stranger BE, Aguet F, Anand S, Ardlie KG, Gabriel S, Getz GA, Graubert A, Hadley K, Handsaker RE, Huang KH, Kashin S, Li Xiao, MacArthur DG, Meier SR, Nedzel JL, Nguyen DT, Segrè AV, Todres E, Balliu B, Barbeira AN, Battle A, Bonazzola R, Brown A, Brown CD, Castel SE, Conrad DF, Cotter DJ, Cox N, Das S, Goede OM de, Dermitzakis ET, Einson J, Engelhardt BE, Eskin E, Eulalio TY, Ferraro NM, Flynn ED, Fresard L, Gamazon ER, Garrido-Martín D, Gay NR, Gloudemans MJ, Guigó R, Hame AR, He Y, Hoffman PJ, Hormozdiari F, Hou L, Im HK, Jo B, Kasela S, Kellis M, Kim-Hellmuth S, Kwong A, Lappalainen T, Li Xin, Liang Y, Mangul S, Mohammadi P, Montgomery SB, Muñoz-Aguirre M, Nachun DC, Nobel AB, Oliva M, Park YoSon, Park Yongjin, Parsana P, Rao AS, Reverter F, Rouhana JM, Sabatti C, Saha A, Stephens M, Stranger BE, Strober BJ, Teran NA, Viñuela A, Wang G, Wen X, Wright F, Wucher V, Zou Y, Ferreira PG, Li G, Melé M, Yeger-Lotem E, Barcus ME, Bradbury D, Krubit T, McLean JA, Qi L, Robinson K, Roche NV, Smith AM, Sobin L, Tabor DE, Undale A, Bridge J, Brigham LE, Foster BA, Gillard BM, Hasz R, Hunter M, Johns C, Johnson M, Karasik E, Kopen G, Leinweber WF, McDonald A, Moser MT, Myer K, Ramsey KD, Roe B, Shad S, Thomas JA, Walters G, Washington M, Wheeler J, Jewell SD, Rohrer DC, Valley DR, Davis DA, Mash DC, Branton PA, Barker LK, Gardiner HM, Mosavel M, Siminoff LA, Flicek P, Haeussler M, Juettemann T, Kent WJ, Lee CM, Powell CC, Rosenbloom KR, Ruffier M, Sheppard D, Taylor K, Trevanion SJ, Zerbino DR, Abell NS, Akey J, Chen L, Demanelis K, Doherty JA, Feinberg AP, Hansen KD, Hickey PF, Jasmine F, Jiang L, Kaul R, Kibriya MG, Li JB, Li Q, Lin S, Linder SE, Pierce BL, Rizzardi LF, Skol AD, Smith KS, Snyder M, Stamatoyannopoulos J, Tang H, Wang M, Carithers LJ, Guan P, Koester SE, Little AR, Moore HM, Nierras CR, Rao AK, Vaught JB, Volpi S. 2020. The impact of sex on gene expression across human tissues. *Science* **369**:eaba3066. doi:10.1126/science.aba3066

Otte KA, Nolte V, Mallard F, Schlötterer C. 2021. The genetic architecture of temperature adaptation is shaped by population ancestry and not by selection regime. *Genome Biol* **22**:211. doi:10.1186/s13059-021-02425-9

Packer MS, Liu DR. 2015. Methods for the directed evolution of proteins. *Nature Reviews Genetics* **16**:379–394. doi:10.1038/nrg3927

Paix A, Schmidt H, Seydoux G. 2016. Cas9-assisted recombineering in *C. elegans*: genome editing using in vivo assembly of linear DNAs. *Nucleic Acids Res* **44**:gkw502. doi:10.1093/nar/gkw502

Paix A, Wang Y, Smith HE, Lee C-YS, Calidas D, Lu T, Smith J, Schmidt H, Krause MW, Seydoux G. 2014. Scalable and versatile genome editing using linear DNAs with microhomology to Cas9 sites in *Caenorhabditis elegans*. *Genetics* **198**:1347–1356. doi:10.1534/genetics.114.170423

Papagiannakis A, Jonge JJ de, Zhang Z, Heinemann M. 2017. Quantitative characterization of the auxin-inducible degron: a guide for dynamic protein depletion in single yeast cells. *Sci Rep* **7**:4704. doi:10.1038/s41598-017-04791-6

Parker GA. 1979. Sexual selection and sexual conflict. Academic Press: New York.

Peichel CL, McCann SR, Ross JA, Naftaly AFS, Urton JR, Cech JN, Grimwood J, Schmutz J, Myers RM, Kingsley DM, White MA. 2019. Assembly of a young vertebrate Y chromosome reveals convergent signatures of sex chromosome evolution. *bioRxiv* 2019.12.12.874701. doi:10.1101/2019.12.12.874701

Pekkonen M, Ketola T, Laakso JT. 2013. Resource availability and competition shape the evolution of survival and growth ability in a bacterial community. *PLoS ONE* **8**:e76471. doi:10.1371/journal.pone.0076471

Perli T, Moonen DPI, Broek M van den, Pronk JT, Daran J-M. 2020. Adaptive laboratory evolution and reverse engineering of single-vitamin prototrophies in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* **86**:e00388-20. doi:10.1128/aem.00388-20

Philip NS, Escobedo F, Bahr LL, Berry BJ, Wojtovich AP. 2019. Mos1 element-mediated CRISPR integration of transgenes in *Caenorhabditis elegans*. *G3 Genes Genomes Genetics* **9**:2629–2635. doi:10.1534/g3.119.400399

Pickar-Oliver A, Gersbach CA. 2019. The next generation of CRISPR–Cas technologies and applications. *Nat Rev Mol Cell Bio* **20**:490–507. doi:10.1038/s41580-019-0131-5

Pitnick Scott, Brown WD, Miller GT. 2001. Evolution of female remating behaviour following experimental removal of sexual selection. *Proc R Soc Lond Ser B: Biol Sci* **268**:557–563. doi:10.1098/rspb.2000.1400

Pitnick S., Miller GT, Reagan J, Holland B. 2001. Males' evolutionary responses to experimental removal of sexual selection. *Proc R Soc Lond Ser B: Biol Sci* **268**:1071–1080. doi:10.1098/rspb.2001.1621

Pontes-Quero S, Heredia L, Casquero-García V, Fernández-Chacón M, Luo W, Hermoso A, Bansal M, Garcia-Gonzalez I, Sanchez-Muñoz MS, Perea JR, Galiana-Simal A, Rodriguez-Arabaolaza I, Olmo-Cabrera SD, Rocha SF, Criado-Rodriguez LM, Giovinazzo G, Benedito R. 2017. Dual ifgMosaic: A versatile method for multispectral and combinatorial mosaic gene-function analysis. *Cell* **170**:800-814.e18. doi:10.1016/j.cell.2017.07.031

Praitis V, Casey E, Collar D, Austin J. 2001. Creation of low-copy integrated transgenic lines in *Caenorhabditis elegans*. *Genetics* **157**:1217–26. doi:10.1093/genetics/157.3.1217

Priyadarshini M, Ni JZ, Vargas-Velazquez AM, Gu SG, Frøkjær-Jensen C. 2022. Reprogramming the piRNA pathway for multiplexed and transgenerational gene silencing in *C. elegans*. *Nature Methods* **19**:187–194. doi:10.1038/s41592-021-01369-z

Radman I, Greiss S, Chin JW. 2013. Efficient and rapid *C. elegans* transgenesis by bombardment and hygromycin B selection. *Plos One* **8**:e76019. doi:10.1371/journal.pone.0076019

Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. 2003. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* **300**:1742–1745. doi:10.1126/science.1085881

Rasys AM, Park S, Ball RE, Alcala AJ, Lauderdale JD, Menke DB. 2019. CRISPR-Cas9 gene editing in lizards through microinjection of unfertilized oocytes. *Cell Rep* **28**:2288-2292.e3. doi:10.1016/j.celrep.2019.07.089

Regan JC, Khericha M, Dobson AJ, Bolukbasi E, Rattanavirotkul N, Partridge L. 2016. Sex difference in pathology of the ageing gut mediates the greater response of female lifespan to dietary restriction. *eLife* **5**:e10956. doi:10.7554/elife.10956

Reinke V, Smith HE, Nance J, Wang J, Doren CV, Begley R, Jones SJM, Davis EB, Scherer S, Ward S, Kim SK. 2000. A global profile of germline gene expression in *C. elegans*. *Mol Cell* **6**:605–616. doi:10.1016/s1097-2765(00)00059-9

Rice WR. 1996. Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. *Nature* **381**:232–234. doi:10.1038/381232a0

Rice WR. 1984. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* **38**:735. doi:10.2307/2408385

Rodriguez EA, Campbell RE, Lin JY, Lin MZ, Miyawaki A, Palmer AE, Shu X, Zhang J, Tsien RY. 2017. The growing and glowing toolbox of fluorescent and photoactive proteins. *Trends Biochem Sci* **42**:111–129. doi:10.1016/j.tibs.2016.09.010

Rogers ZN, McFarland CD, Winters IP, Naranjo S, Chuang C-H, Petrov D, Winslow MM. 2017. A quantitative and multiplexed approach to uncover the fitness landscape of tumor suppression in vivo. *Nature Methods* **14**:737–742. doi:10.1038/nmeth.4297

Rossum G van. 1991. Interactively testing remote servers using the Python programming language. *CWI quarterly* 283–303.

Rostant WG, Mason JS, Coriolis J, Chapman T. 2020. Resource-dependent evolution of female resistance responses to sexual conflict. *Evol Lett* **4**:54–64. doi:10.1002/evl3.153

Rudman SM, Greenblum SI, Rajpurohit S, Betancourt NJ, Hanna J, Tilk S, Yokoyama T, Petrov DA, Schmidt P. 2022. Direct observation of adaptive tracking on ecological time scales in *Drosophila*. *Science* **375**:eabj7484. doi:10.1126/science.abj7484

Sadowski PD. 1995. The Flp recombinase of the 2-microns plasmid of *Saccharomyces cerevisiae*. *Prog nucleic acid Res Mol Biol* **51**:53–91.

Sánchez-Rivera FJ, Diaz BJ, Kastenhuber ER, Schmidt H, Katti A, Kennedy M, Tem V, Ho Y-J, Leibold J, Paffenholz SV, Barriga FM, Chu K, Goswami S, Wuest AN, Simon JM, Tsanov KM, Chakravarty D, Zhang H, Leslie CS, Lowe SW, Dow LE. 2022. Base editing sensor libraries for high-throughput engineering and functional analysis of cancer-associated single nucleotide variants. *Nature Biotechnology* **40**:862–873. doi:10.1038/s41587-021-01172-3

Sandoval-Villegas N, Nurieva W, Amberger M, Ivics Z. 2021. Contemporary transposon tools: A review and guide through mechanisms and applications of sleeping beauty, piggyBac and Tol2 for genome engineering. *Int J Mol Sci* **22**:5084. doi:10.3390/ijms22105084

Sardell JM, Kirkpatrick M. 2020. Sex differences in the recombination landscape. *Am Nat* **195**:361–379. doi:10.1086/704943

Sauer B, Henderson N. 1988. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc Natl Acad Sci* **85**:5166–5170. doi:10.1073/pnas.85.14.5166

Schiksnis EC, Nicholson AL, Modena MS, Pule MN, Arribere JA, Pasquinelli AE. 2020. Auxin-independent depletion of degron-tagged proteins by TIR1. *microPublication Biol* **2020**:10.17912/micropub.biology.000213. doi:10.17912/micropub.biology.000213

Schlötterer C, Kofler R, Versace E, Tobler R, Franssen SU. 2015. Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. *Heredity* **114**:431–440. doi:10.1038/hdy.2014.86

Seidel HS, Ailion M, Li J, Oudenaarden A van, Rockman MV, Kruglyak L. 2011. A novel sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in *C. elegans*. *Plos Biol* **9**:e1001115. doi:10.1371/journal.pbio.1001115

Semple JI, Biondini L, Lehner B. 2012. Generating transgenic nematodes by bombardment and antibiotic selection. *Nat Methods* **9**:118–119. doi:10.1038/nmeth.1864

Semple JI, Garcia-Verdugo R, Lehner B. 2010. Rapid selection of transgenic *C. elegans* using antibiotic resistance. *Nat Methods* **7**:725–727. doi:10.1038/nmeth.1495

Serrano-Saiz E, Oren-Suissa M, Bayer EA, Hobert O. 2017. Sexually dimorphic differentiation of a *C. elegans* hub neuron Is cell autonomously controlled by a conserved transcription factor. *Curr Biol* **27**:199–209. doi:10.1016/j.cub.2016.11.045

Setten RL, Rossi JJ, Han S. 2019. The current state and future directions of RNAi-based therapeutics. *Nat Rev Drug Discov* **18**:421–446. doi:10.1038/s41573-019-0017-4

Shao Y, Lu N, Xue X, Qin Z. 2019. Creating functional chromosome fusions in yeast with CRISPR–Cas9. *Nat Protoc* **14**:2521–2545. doi:10.1038/s41596-019-0192-0

Sidore AM, Plesa C, Samson JA, Lubock NB, Kosuri S. 2020. DropSynth 2.0: high-fidelity multiplexed gene synthesis in emulsions. *Nucleic Acids Res* **48**:gkaa600-. doi:10.1093/nar/gkaa600

Silva-García CG, Lanjuin A, Heintz C, Dutta S, Clark NM, Mair WB. 2019. Single-copy knock-in loci for defined gene expression in *Caenorhabditis elegans*. *G3 Genes Genomes Genetics* **9**:2195–2198. doi:10.1534/g3.119.400314

Smith AM, Durbic T, Oh J, Urbanus M, Proctor M, Heisler LE, Giaever G, Nislow C. 2011. Competitive genomic screens of barcoded yeast libraries. *Journal of Visualized Experiments*. doi:10.3791/2864

Snook RR, Gidaszewski NA, Chapman T, Simmons LW. 2013. Sexual selection and the evolution of secondary sexual traits: sex comb evolution in *Drosophila*. *J Evol Biol* **26**:912–918. doi:10.1111/jeb.12105

Song AJ, Palmiter RD. 2018. Detecting and avoiding problems when using the Cre–lox system. *Trends Genet* **34**:333–340. doi:10.1016/j.tig.2017.12.008

Spanjaard B, Hu B, Mitic N, Olivares-Chauvet P, Janjuha S, Ninov N, Junker JP. 2018. Simultaneous lineage tracing and cell-type identification using CRISPR–Cas9-induced genetic scars. *Nature Biotechnology* **36**:469–473. doi:10.1038/nbt.4124

Stevenson ZC, Banse SA, Phillips PC. 2021. Genetic data compression and methods of use [international patent]. US20210332387A1.

Stevenson ZC, Moerdyk-Schauwecker MJ, Banse SA, Patel DS, Lu H, Phillips PC. 2023. High-throughput library transgenesis in *Caenorhabditis elegans* via transgenic arrays resulting in diversity of integrated sequences (TARDIS). *eLife* **12**. doi:10.7554/elife.84831.3

Stevenson ZC, Moerdyk-Schauwecker MJ, Jamison B, Phillips PC. 2020. Rapid self-selecting and clone-free integration of transgenes into engineered CRISPR safe harbor locations in *Caenorhabditis elegans*. *G3: GenesGenomesGenet* **10**:3775–3782. doi:10.1534/g3.120.401400

Stiernagle T. 2006. Maintenance of *C. elegans*. *WormBook* 1–11. doi:10.1895/wormbook.1.101.1

Stinchcomb DT, Shaw JE, Carr SH, Hirsh D. 1985. Extrachromosomal DNA transformation of *Caenorhabditis elegans*. *Molecular and Cellular Biology* **5**:3484–3496. doi:10.1128/mcb.5.12.3484

Stojic L, Lun ATL, Mangei J, Mascalchi P, Quarantotti V, Barr AR, Bakal C, Marioni JC, Gergely F, Odom DT. 2018. Specificity of RNAi, LNA and CRISPRi as loss-of-function methods in transcriptional analysis. *Nucleic Acids Res* **46**:gky437-. doi:10.1093/nar/gky437

Sun W, Hu Y. 2013. eQTL mapping using RNA-seq data. *Stat Biosci* **5**:198–219. doi:10.1007/s12561-012-9068-3

Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* **6**:377–382. doi:10.1038/nmeth.1315

Tareen A, Kinney JB. 2020. Logomaker: beautiful sequence logos in Python. *Bioinformatics* **36**:2272–2274. doi:10.1093/bioinformatics/btz921

Teotónio H, Estes S, Phillips PC, Baer CF. 2017. Experimental evolution with *Caenorhabditis* nematodes. *Genetics* **206**:691–716. doi:10.1534/genetics.115.186288

Teterina AA, Coleman-Hulbert AL, Banse SA, Willis JH, Perez VI, Lithgow GJ, Driscoll M, Phillips PC. 2022. Genetic diversity estimates for the *Caenorhabditis* intervention testing program screening panel. *microPublication biology* **2022**. doi:10.17912/micropub.biology.000518

Trivers R. 1988. Sex differences in rates of recombination and sexual selection. Sunderland, MA: Sinauer Associates.

Trivers RL. 1972. Sexual selection and parental investment. Aldine Publishing Company: Chicago.

Tukiainen T, Villani A-C, Yen A, Rivas MA, Marshall JL, Satija R, Aguirre M, Gauthier L, Fleharty M, Kirby A, Cummings BB, Castel SE, Karczewski KJ, Aguet F, Byrnes A, Aguet F, Ardlie KG, Cummings BB, Gelfand ET, Getz G, Hadley K, Handsaker RE, Huang KH, Kashin S, Karczewski KJ, Lek M, Li Xiao, MacArthur DG, Nedzel JL, Nguyen DT, Noble MS, Segrè AV, Trowbridge CA, Tukiainen T, Abell NS, Balliu B, Barshir R, Basha O, Battle A, Bogu GK, Brown A, Brown CD, Castel SE, Chen LS, Chiang C, Conrad DF, Cox NJ, Damani FN, Davis JR, Delaneau O, Dermitzakis ET, Engelhardt BE, Eskin E, Ferreira PG, Frésard L, Gamazon ER, Garrido-Martín D, Gewirtz ADH, Gliner G, Gloudemans MJ, Guigo R, Hall IM, Han B, He Y, Hormozdiari F, Howald C, Im HK, Jo B, Kang EY, Kim Y, Kim-Hellmuth S, Lappalainen T, Li G, Li Xin, Liu B, Mangul S, McCarthy MI, McDowell IC, Mohammadi P, Monlong J, Montgomery SB, Muñoz-Aguirre M, Ndungu AW, Nicolae DL, Nobel AB, Oliva M, Ongen H, Palowitch JJ, Panousis N, Papasaikas P, Park YoSon, Parsana P, Payne AJ, Peterson CB, Quan J, Reverter F, Sabatti C, Saha A, Sammeth M, Scott AJ, Shabalin AA, Sodaei R, Stephens M, Stranger BE, Strober BJ, Sul JH, Tsang EK, Urbut S, Bunt M van de, Wang G, Wen X, Wright FA, Xi HS, Yeger-Lotem E, Zappala Z, Zaugg JB, Zhou Y-H, Akey JM, Bates D, Chan J, Chen LS, Claussnitzer M, Demanelis K, Diegel M, Doherty JA, Feinberg AP, Fernando MS, Halow J, Hansen KD, Haugen E, Hickey PF, Hou L, Jasmine F, Jian R, Jiang L, Johnson A, Kaul R, Kellis M, Kibriya MG, Lee K, Li JB, Li Q, Li Xiao, Lin J, Lin S, Linder S, Linke C, Liu Y, Maurano MT, Molinie B, Montgomery SB, Nelson J, Neri FJ, Oliva M, Park Yongjin, Pierce BL, Rinaldi NJ, Rizzardi LF, Sandstrom R, Skol A, Smith KS, Snyder MP, Stamatoyannopoulos J, Stranger BE, Tang H, Tsang EK, Wang L, Wang M, Wittenberghe NV, Wu F, Zhang R, Nierras CR, Branton PA, Carithers LJ,

Guan P, Moore HM, Rao A, Vaught JB, Gould SE, Lockart NC, Martin C, Struewing JP, Volpi S, Addington AM, Koester SE, Little AR, Brigham LE, Hasz R, Hunter M, Johns C, Johnson M, Kopen G, Leinweber WF, Lonsdale JT, McDonald A, Mestichelli B, Myer K, Roe B, Salvatore M, Shad S, Thomas JA, Walters G, Washington M, Wheeler J, Bridge J, Foster BA, Gillard BM, Karasik E, Kumar R, Miklos M, Moser MT, Jewell SD, Montroy RG, Rohrer DC, Valley DR, Davis DA, Mash DC, Undale AH, Smith AM, Tabor DE, Roche NV, McLean JA, Vatanian N, Robinson KL, Sobin L, Barcus ME, Valentino KM, Qi L, Hunter S, Hariharan P, Singh S, Um KS, Matose T, Tomaszewski MM, Barker LK, Mosavel M, Siminoff LA, Traino HM, Flicek P, Juettemann T, Ruffier M, Sheppard D, Taylor K, Trevanion SJ, Zerbino DR, Craft B, Goldman M, Haeussler M, Kent WJ, Lee CM, Paten B, Rosenbloom KR, Vivian J, Zhu J, Lappalainen T, Regev A, Ardlie KG, Hacohen N, MacArthur DG. 2017. Landscape of X chromosome inactivation across human tissues. *Nature* **550**:244–248. doi:10.1038/nature24265

Turner LM, White MA, Tautz D, Payseur BA. 2014. Genomic networks of hybrid sterility. *PLoS Genet* **10**:e1004162. doi:10.1371/journal.pgen.1004162

VanInsberghe M, Zahn H, White AK, Petriv OI, Hansen CL. 2018. Highly multiplexed single-cell quantitative PCR. *PLoS ONE* **13**:e0191601. doi:10.1371/journal.pone.0191601

VanKuren NW, Long M. 2018. Gene duplicates resolving sexual conflict rapidly evolved essential gametogenesis functions. *Nat Ecol Evol* **2**:705–712. doi:10.1038/s41559-018-0471-0

Venkataram S, Dunn B, Li Y, Agarwala A, Chang J, Ebel ER, Geiler-Samerotte K, Hérissant L, Blundell JR, Levy SF, Fisher DS, Sherlock G, Petrov DA. 2016. Development of a comprehensive genotype-to-fitness map of adaptation-driving mutations in yeast. *Cell* **166**:1585-1596.e22. doi:10.1016/j.cell.2016.08.002

Venken KJT, Carlson JW, Schulze KL, Pan H, He Y, Spokony R, Wan KH, Koriabine M, Jong PJ de, White KP, Bellen HJ, Hoskins RA. 2009. Versatile P[acman] BAC libraries for transgenesis studies in *Drosophila melanogaster*. *Nature Methods* **6**:431–434. doi:10.1038/nmeth.1331

Venken KJT, He Y, Hoskins RA, Bellen HJ. 2006. P[acman]: A BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science* **314**:1747–1751. doi:10.1126/science.1134426

Vicencio J, Martínez-Fernández C, Serrat X, Cerón J. 2019. Efficient generation of endogenous fluorescent reporters by nested CRISPR in *Caenorhabditis elegans*. *Genetics* **211**:1143–1154. doi:10.1534/genetics.119.301965

Wang H, Liu J, Gharib S, Chai CM, Schwarz EM, Pokala N, Sternberg PW. 2017. cGAL, a temperature-robust GAL4–UAS system for *Caenorhabditis elegans*. *Nat Methods* **14**:145–148. doi:10.1038/nmeth.4109

Wang JY, Doudna JA. 2023. CRISPR technology: A decade of genome editing is only the beginning. *Science* **379**:eadd8643. doi:10.1126/science.add8643

Ward JD. 2015. Rapid and precise engineering of the *Caenorhabditis elegans* genome with lethal mutation co-conversion and inactivation of NHEJ repair. *Genetics* **199**:363–377. doi:10.1534/genetics.114.172361

Warnefors M, Mössinger K, Halbert J, Studer T, VandeBerg JL, Lindgren I, Fallahshahroudi A, Jensen P, Kaessmann H. 2017. Sex-biased microRNA expression in mammals and birds reveals underlying regulatory mechanisms and a role in dosage compensation. *Genome Res* **27**:1961–1973. doi:10.1101/gr.225391.117

Weismann A. 1893. The Germ-plasm: A theory of heredity. Translated by W. Newton Parker and Harriet Rönnfeldt. New York: Scribner. doi:10.5962/bhl.title.25196

Weissman TA, Pan YA. 2015. Brainbow: New resources and emerging biological applications for multicolor genetic labeling and analysis. *Genetics* **199**:293–306. doi:10.1534/genetics.114.172510

Wells JA. 1991. Systematic mutational analyses of protein-protein interfaces. pp. 390–411. doi:10.1016/0076-6879(91)02020-a

Whitlock MC, Phillips PC, Moore FB, Tonsor SJ. 1995. Multiple fitness peaks and epistasis. *Annu Rev Ecol Syst* **26**:601–629. doi:10.1146/annurev.es.26.110195.003125

Winslow MP Dmitri Winters, Ian Mcfarland, Christopher Rogers, Zoe. 2022. Compositions and methods for multiplexed quantitative analysis of cell lineages [international patent]. 17/281919.

Woglar A, Yamaya K, Roelens B, Boettiger A, Köhler S, Villeneuve AM. 2020. Quantitative cytogenetics reveals molecular stoichiometry and longitudinal organization of meiotic chromosome axes and loops. *PLOS Biology* **18**:e3000817. doi:10.1371/journal.pbio.3000817

Wu C-C, Kruse F, Vasudevarao MD, Junker JP, Zebrowski DC, Fischer K, Noël ES, Grün D, Berezikov E, Engel FB, Oudenaarden A van, Weidinger G, Bakkers J. 2016. Spatially resolved genome-wide transcriptional profiling identifies BMP signaling as essential regulator of zebrafish cardiomyocyte regeneration. *Dev Cell* **36**:36–49. doi:10.1016/j.devcel.2015.12.010

Wu M-R, Nissim L, Stupp D, Pery E, Binder-Nissim A, Weisinger K, Enghuus C, Palacios SR, Humphrey M, Zhang Z, Novoa EM, Kellis M, Weiss R, Rabkin SD, Tabach Y, Lu TK. 2019. A high-throughput screening and computation platform for identifying synthetic promoters with enhanced cell-state specificity (SPECS). *Nature Communications* **10**:2880. doi:10.1038/s41467-019-10912-8

Wyman MJ, Cutter AD, Rowe L. 2012. Gene duplication in the evolution of sexual dimorphism. *Evolution* **66**:1556–1566. doi:10.1111/j.1558-5646.2011.01525.x

Xu H, Guo Y, Qiu L, Ran Y. 2022. Progress in soybean genetic transformation over the last decade. *Frontiers in Plant Science* **13**. doi:10.3389/fpls.2022.900318

Xu H, Xiao T, Chen C-H, Li W, Meyer CA, Wu Q, Wu D, Cong L, Zhang F, Liu JS, Brown M, Liu XS. 2015. Sequence determinants of improved CRISPR sgRNA design. *Genome Res* **25**:1147–1157. doi:10.1101/gr.191452.115

Xu S. 2015. The application of CRISPR-Cas9 genome editing in *Caenorhabditis elegans*. *J Genet Genomics* **42**:413–421. doi:10.1016/j.jgg.2015.06.005

Yamaguchi N, Colak-Champollion T, Knaut H. 2019. zGrad is a nanobody-based degron system that inactivates proteins in zebrafish. *eLife* **8**:e43125. doi:10.7554/elife.43125

Yan F, Powell DR, Curtis DJ, Wong NC. 2020. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biol* **21**:22. doi:10.1186/s13059-020-1929-3

Yang X, Schadt EE, Wang S, Wang H, Arnold AP, Ingram-Drake L, Drake TA, Lusis AJ. 2006. Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res* **16**:995–1004. doi:10.1101/gr.5217506

Zaslaver A, Bren A, Ronen M, Itzkovitz S, Kikoin I, Shavit S, Liebermeister W, Surette MG, Alon U. 2006. A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nature Methods* **3**:623–628. doi:10.1038/nmeth895

Zhang D, Tu S, Stubna M, Wu W-S, Huang W-C, Weng Z, Lee H-C. 2018. The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes. *Science (New York, NY)* **359**:587–592. doi:10.1126/science.aao2840

Zhang L, Ward JD, Cheng Z, Dernburg AF. 2015. The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in *C. elegans*. *Development* **142**:4374–4384. doi:10.1242/dev.129635

Zhao Q, Zhong W, He W, Li Yiying, Li Yaqing, Li T, Vasseur L, You M. 2019. Genome-wide profiling of the alternative splicing provides insights into development in *Plutella xylostella*. *BMC Genom* **20**:463. doi:10.1186/s12864-019-5838-3

Zhou Q, Wang T, Leng L, Zheng W, Huang J, Fang F, Yang L, Chen F, Lin G, Wang W, Kristiansen K. 2019. Single-cell RNA-seq reveals distinct dynamic behavior of sex chromosomes during early human embryogenesis. *Mol Reprod Dev* **86**:871–882. doi:10.1002/mrd.23162

Zorita E, Cuscó P, Filion GJ. 2015. Starcode: sequence clustering based on all-pairs search. *Bioinformatics* **31**:1913–1919. doi:10.1093/bioinformatics/btv053