ADVANCING SPACE LIFE SCIENCE RESEARCH USING *DROSOPHILA MELANOGASTER*

A National Academies Decadal Survey

Focused Topic Whitepaper

Trudy Mackay, PhD.
Clemson University
tmackay@clemson.edu

Janani Iyer, PhD.
USRA/ KBR Wyle
NASA Ames Research Center
janani.s.iyer@nasa.gov

Siddhita D. Mhatre, PhD.
COSMIAC Univ. New Mexico/ KBR Wyle
NASA Ames Reserch Center
siddhita.d.mhatre@nasa.gov

This white paper highlights *Drosophila* as a spaceflight model and discusses how *Drosophila* based quantitative genomics can be applied to understand the biological impacts of spaceflight environment.

Introduction

Spaceflight conditions provide a unique and dynamic environment with multiple stressors. These stressors include microgravity/reduced gravity, ionizing radiation, altered temperature, altered gas composition, altered day/night light cycles, isolation, and reduced atmospheric pressures. Spaceflight habitats are engineered to operate within these stressors, yet still have unique physical parameters, for example, there is 10 times more CO_2 in the closed environment on the International Space Station (ISS) than normally encountered on Earth. It is important to understand the effects of this environment on the whole organism to facilitate future long-duration crewed missions to the Moon and Mars.

The closed environment of spaceflight presents challenges to life science research in space. Spaceflight missions have volume/mass constraints contributing towards flight logistics. Some of these constraints include up-mass and down-mass limitations, "sharing the ride" with other payloads, and limited space on the ISS, making execution of experiments with large animals/plants complicated. Another limitation is crew time and resources, as astronaut time is divided between conducting experiments and other ISS operations. Spaceflight experiments also require specialized hardware; favorable hardware should leave small footprints while accommodating a large sample size. Data downlink constraints are important considerations when planning a spaceflight experiment.

Drosophila melanogaster as a spaceflight model

Drosophila melanogaster (the fruit fly) is a well-characterized model organism to study basic biological processes and has functionally important genes that are highly conserved between flies and humans. For example, over 75% of the human disease-causing genes have a close homolog in *Drosophila*. The entire fly genome is now sequenced, showing 70% homology overall to the human genome and with less genetic redundancy. *Drosophila* is genetically tractable, allowing for basic, synthetic, and translational research and development. Ground studies involving *Drosophila* as a model organism have advanced our understanding of genetics, molecular biology, systems biology, behavior, and development; and this will continue in space.

Fruit flies were the first animals to be launched in space in July 1947, and have been historically used as a model organism in early space missions (reviewed in Miquel and Souza, 1991; Harrington, 2014). *Drosophila* has multiple advantages for spaceflight studies: (1) a short generation time and average lifespan of 60-80 days; (2) ability to propagate large population sizes within the limited volume and mass requirements in space, thus providing high statistical power; (3) low cost of breeding and ease of fly maintenance; (4) ability to perform genetic manipulations; (5) availability of a large repertoire of genetic tools, including genetic mutations, the *UAS-GAL4* bipartite system for spatial and temporal control of gene expression, including RNA interference (RNAi), and inbred, sequenced wild derived lines for quantitative trait locus (QTL) mapping; and (6) suitability for short and long-duration multi-generational experiments. Multi-generational experiments using *Drosophila* would further our understanding of the effects of spaceflight on reproduction, epigenetic signatures, and genetic effects on future generations.

Previous experimentation onboard the ISS demonstrates that *Drosophila* can be accommodated in existing hardware using minimal crew time to allow for neurobehavioral, neuroanatomical, immunity/pathogenesis, cardiovascular, and developmental studies. Some of the studies conducted on the ISS using *Drosophila* have demonstrated that spaceflight alters innate immune responses (Marcu et al., 2011), cardiac functioning (Walls et al., 2020), mutations in germ cells (Ikenaga et al., 1997), and neuronal and metabolic alterations (Mhatre and Iyer et al., under review). Additionally, *Drosophila* has been successfully used in ground-based hypergravity and radiation studies (Hosamani et al., 2016, Hateley et al., 2016, Tanaka and Furuta, 2021). Experiments involving space radiation exposure have demonstrated that not all tissues are affected equally and that space radiation can induce mutations in the germline and affect successive generations. It has been observed that changes in gravity impact gene and protein expression in both developing and adult organisms in a pathway-specific manner and that *Drosophila* can also be used to understand host/pathogen interactions in space (Gilbert et al., 2020).

Drosophila hardware for life science research in spaceflight

Drosophila hardware currently available for use on orbit of the ISS consists of:

1. NASA Ventilated Fly Box (VFB): Each VFB holds 15 standard fly vials and has a temperature logger. Multiple VFBs can be used for an experiment. A total of 6 VFBs can support 8,000-10,000 adults, and tens of thousands of eggs and larvae.
2. NASA Fly Cassette: This hardware fits a total of 12 cassettes in an experimental run, has centrifuge capabilities, and allows for multigenerational fly growth in space.
3. Multi-use Variable-g Platform (MVP): This hardware is fitted to accommodate two independent centrifuges. It has real-time temperature and relative humidity controlling capabilities, CO_2 and O_2 monitoring, and cycles fresh cabin air into the habitat. The hardware supports the separation of fly generations with automated food change-outs. It also has video recording and lighting capabilities, enabling assessment of fly behavior in space while maintaining the circadian rhythm of flies. In the MVP platform, there are two centrifuges in total, and each has 12 modules, the 12 modules can support 3 consecutive generations of *Drosophila*. Each MVP can support thousands of adults and tens of thousands of eggs and larvae.

Quantitative genomics research with _Drosophila_

Quantitative genomics aims to understand the genetic architecture of complex traits using genotype to phenotype mapping. The goal for each trait is to understand the following: (1) What are the genes that act together to affect the variation of a trait? (2) What are the genetic variants in these genes that affect the phenotype? (3) What is the distribution of variant effects on the phenotype(s) of interest? (4) Do the effect sizes vary between males and female, different environments, and genetic backgrounds? Such context-dependent effects are very relevant to the space environment, where variation may be affected by sexual dimorphism, genotype by space environment interaction, variant by variant interaction (additive or epistatic), and have pleiotropic effects not only on quantitative phenotypes but also intermediate molecular phenotypes, such as gene co-expression networks.

***Drosophila melanogaster* Genetic Reference Panel (DGRP): A powerful resource for genome wide association studies (GWAS) in quantitative genomics**

The Mackay lab has generated a powerful resource, the DGRP (Mackay et al., 2012; Huang et al. 2014). The original DGRP consisted of 200 inbred lines with full genome sequences; this resource currently consists of 1200 fully sequenced inbred lines, with an average of 40X coverage. This allows for GWAS with all known genetic variants in the DGRP. Inbreeding allows the same genotype to be assessed in multiple environments, and for multiple individuals to be assessed per genotype, which increases the fraction of phenotypic variance due to genetics. Most QTLs have small effects, and as a result, large populations are needed to detect them. Most studies utilizing the DGRP performed GWAS for phenotypes measured on the inbred DGRP lines to identify genetic variants associated with the phenotype of interest. However, the traditional approach involves rearing and phenotyping many thousands of flies.

An alternative approach, which might be more relevant to the space environment, is to use an advanced intercross population (AIP) (Huang et al., 2012; 2020). With the AIP method, an Earth population can be constructed by randomly mating multiple DGRP lines (for example: 40) for several generations (200) with large population sizes (N=800/generation). This outbred population is then subjected to single and multigenerational natural selection in a stressful environment. If survival is the trait of interest, 10-15% of the flies are randomly chosen before exposure to the stressful environment, and the longest surviving 10-15% of flies are selected after exposure to the environment. The experiment is repeated several times to average over random environmental differences between trials. Whole genome sequencing (WGS) is then performed on the selected pools of flies to identify variants with significantly different allele frequencies between the two pools. These variants are either QTLs causing increased survival in the stressful environment, or are in linkage disequilibrium with the true causal variant. The observation from such terrestrial studies is that there is substantial and highly polygenic genetic variation for every quantitative trait studied to date. In addition, we can gain some biological insights into the genetic basis of these traits. The data output is a list of variants in multiple genes, of which most are computationally predicted (novel) without an associated phenotype, and many are known to act in another context (e.g., a gene previously implicated in development that has a novel effect on ethanol resistance). Most natural variants associated with quantitative traits are often in intergenic/intronic regions and presumably affect the phenotype via regulation of gene expression. Because many gene/gene interactions have been previously described in *Drosophila*, variants associated with quantitative phenotypes can often be mapped to known gene interaction networks to indicate biological context. Mutations, RNAi or gene editing can then be used to validate the effect of the gene or variant on the trait (~ 70% variants can be validated in this manner). Results from these studies show that most variants have small effects, so large numbers of individuals are needed to identify them. Additionally, context-dependent effects are the norm, not the exception. For example, when the lifespan of DGRP and AIP flies is determined when animals are reared in three temperature environments (18°, 25°, 28°), we find genetic variation common to all environments as well as genetic variation specific to males and females (genotype by sex interaction) and the three environments (genotype by environment interaction) (Huang et al., 2020).

Epistasis: Effects of one locus depend on allele frequency of interacting loci

Epistasis is another context-dependent effect, where the effect of a focal locus, A, depends on the genotype at another locus, B. With additive gene action, the effect of locus A is independent of the genotype at locus B. However, with epistasis the effect of locus A may be either decreased or increased, or even change direction, depending on the B locus genotype. A consequence of epistasis is that the effect of locus A will be different in populations with different allele frequencies at locus B. We can exploit this phenomenon by creating populations with different allele frequencies (*e.g.*, the whole DGRP and AIPs derived from a subset of DGRP lines) and map QTLs for the same quantitative trait in both populations (Huang et al., 2012). If all effects are additive, we expect a great deal of overlap of the effects of the same alleles in the two populations. However, if the effects are largely different, there may be epistasis and we can determine candidate interacting variants by a statistical analysis that uses all variants that were significant in either the DGRP or AIP analyses, and use each as a focal allele in a GWAS in the DGRP that evaluated the main effect of the focal allele, all other alleles in the DGRP, and the genetic interactions. This procedure can be used to derive a trait-specific interaction network that shows the significant variants in different populations are either connected to each other directly or via one or few intermediates, such that the genetic architecture of traits can be population-specific, but the network that they impinge is common.

Pleiotropy and regulation of gene expression in the DGRP

Alleles that affect one quantitaitve trait often have pleiotropic effects on other quantitative traits. For example, alleles that affect an organismal quantitative phenotype can also affect gene co-expression networks, which gives insight into to biological underpinnings of the trait(s) of interest. These gene-co-expression networks have been derived by performing RNA sequencing of 200 DGRP lines (Everett et al., 2020). GWAS was performed for each of the > 10,000 genetically variable transcripts in males and females to map expression QTLs (eQTLs). The eQTLs were classified as *cis* if they were within 1 kb of the body of the gene encoding the transcript, and *trans* otherwise. These data were used to build sex-specific pleiotropic *cis-trans* transcriptional regulatory networks that include a large number of novel transcripts that are likely long non-coding RNAs.

Future approaches for Quantitative Genomics (QG) using *Drosophila* in the spaceflight environment

It is possible to use the AIP population to identify genes and genetic variants that confer better survival in the space environments, in single generation and multi-generation experiments.

1) Single generation analysis of adult survival. The control samples are three independent pools of 100 females and 100 males from the same generation of the AIP as the experimental flies. The experimental flies consist of triplicate samples of 1000 males and 1000 females from the AIP population that are reared simultaneously on Earth and in the ISS environment, and transferred to new vials every three days. When ~100 remain in each replicate sample in the ISS environment, they will be pooled and frozen at -80°C for DNA and RNA sequencing on Earth. At the same time, 100 flies from each of the Earth samples will be frozen at -80°C for DNA and RNA sequencing. A GWAS of

significant allele frequency changes between the control and experimental flies from the ISS and Earth samples maps variants and genes associated with adult survival on Earth and on the ISS; differences between the Earth and ISS environment indicate variants and genes with a genotype by space environment interaction. Gene expression analyses using RNA sequencing will also indicate genes whose expression changes with age in both environments and genotype by space environment interactions for gene expression.

2) Multiple generation analysis of response to natural selection in the space environment. Ten replicate samples from the AIP population, each initially consisting of 500 males and 500 females at generation 0, will be reared simultaneously on Earth (control populations) and on the ISS (selected populations). Flies will be maintained on a two-week generation interval in both environments; after each generation has reproduced, 200 male and 200 female parents will be frozen at -80°C for DNA, RNA and ATAC sequencing on Earth. At the end of the mission, the latest generation of larvae will be transported to Earth and resulting adults will be assessed for a battery of quantitative traits simultaneously with the adults that were reared continuously on Earth. Examples of relevant quantitative traits include aspects of stress resistance (survival after heat shock, time to recover from a chill-induced coma, resistance to oxidative stress); behaviors (locomotor activity, sleep, circadian rhythm); and fitness (lifespan, reproduction). These data will inform on the genetic, genomic and epigenetic response to natural selection in the space environment.

3) Modifications of the single- and multi-generation designs above could include rearing the AIP flies under simulated 1g gravity on the ISS in order to assess the effects of radiation on the ISS; and titrating the gravity to simulate Lunar and/or Martian gravity.

4) The analysis of allele frequency differences is straightforward. We sequence DNA from each pool of frozen flies to 2-4X per fly, map to the genome, call variants, and perform statistical tests to assess signficance of difference of allele frequency for each variant between the pools. An FDR of < 0.05 is used to indicate significant differences in allele frequency. Combining these data with differential gene expression from RNA sequencing will enable us to assess whether the same genes that are associated with increased survival (single generation) and fitness (multi-generation) are also associated with changes in expression.

Conclusion

Space and ground-based experiments using *D. melanogaster* have elucidated the spaceflight-associated health risks. *Drosophila* quantitative genomics will aid in the identification of genetic variants capable of enduring harsh space environments. Further, the hardware systems with automatic tracking to monitor fly behavior in real- time in both single- and multi-generation experiments in space environments will enable effective and efficient science. Also, the future establishment of a science lab on the lunar surface similar to that on ISS will help us understand the mechanisms relevant to human biology in deep space environment.