Using Drosophila melanogaster to understand how microbes affect host behavior

Author
Elya, Carolyn Nicole

Publication Date
2017

Peer reviewed|Thesis/dissertation
Using *Drosophila melanogaster* to understand how microbes affect host behavior

by

Carolyn Nicole Elya

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Molecular and Cell Biology in the Graduate Division of the University of California, Berkeley

Committee in charge:

Professor Michael B. Eisen
Professor Kristin Scott
Professor Russell Vance
Professor Richard Calendar
Professor Damian Elias

Spring 2017
Using *Drosophila melanogaster* to understand how microbes affect host behavior

Copyright 2017
by
Carolyn Nicole Elya

This dissertation is licensed under the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Abstract

Using *Drosophila melanogaster* to understand how microbes affect host behavior

by

Carolyn Nicole Elya

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Michael B. Eisen, Chair

Animals live in a world teeming with microbial life. Co-existing over evolutionary time, both microbes and animals have evolved methods to influence the other to maximize their respective fitness. It has been the focus of my doctoral work to study how microbes affect animal behavior using the model species *Drosophila melanogaster*.

Like all other animals so-far encountered, the fruit fly digestive system is occupied by microbes consisting of mostly bacteria. Contrary to my expectations, I did not encounter fly behaviors that changed by varying their associated gut bacteria. In order to understand what changes were taking place in flies with manipulated gut flora, I assayed gene expression in dissected guts or in whole flies. I discovered that associating flies from the embryonic stage with zero, one or three bacterial taxa resulted in the same transcriptional profile in the adult gut, suggesting that adult guts are buffered against the bacteria that passage through them. However, associating flies with the yeast *Saccharomyces cerevisiae*, whether alone or in combination with these bacteria, was sufficient to recapitulate the transcriptional profile observed in guts of conventionally-reared flies. This suggests that yeast, not bacteria, are key in mediating the gut transcriptional response of adult flies. In contrast, transcription within the entire adult animal varied with exposure to different microbial populations, suggesting that different bacterial taxa can influence fly hosts prior to adulthood.

Yeast, *Saccharomyces cerevisiae*, is a key food source for *D. melanogaster* both in the wild and in the laboratory. Previous work by former Eisen laboratory member Kelly Schiabor had demonstrated a correlation between attraction to yeast grown under natural (sugar-replete, nitrogen-limiting) conditions (YVN) and chimeric variant in the gene encoding a component of fly olfaction, odorant receptor 22 (Or22). In conjunction with Alli Quan, I tested the hypothesis that the chimeric Or22
allele in *D. melanogaster* mediates sensitivity of flies to YVN and therefore makes these flies better suited to detect yeast in the environment. Through sequence analysis, bidirectional crosses between chimeric and non-chimeric lines and ultimately, replacement of a non-chimeric Or22 allele with a chimeric allele within a non-chimeric background, we found that Or22 alone does not mediate sensitivity to YVN. Still, the signs of selection at the Or22 locus suggest that this receptor confers some adaptive function in wild flies.

*Entomophthora muscae* is a fungal pathogen that infects, alters the behavior of, and then kills dipterans. Predominantly reported in house flies and other large Muscoidea, critically-ill flies summit, extend their proboscis, and raise their wings up and away from their dorsal abdomen in the moments prior to death, dying in an elevated position that appears to benefit fungal dispersal and therefore fitness. Despite being described over 160 years ago, we know little of *E. muscae* biology, especially the molecular means through which it alters host behavior. Serendipitously, I discovered a strain of *E. muscae* (CNE1) that infects wild *Drosophila* species, including *D. melanogaster*. I have isolated this strain in the laboratory both *in vivo*, through active propagation between healthy fruit flies, and *in vitro*, in liquid culture. By observing the isolated fungal culture with modern technology, I have been able to corroborate and add to the series of observations describing how this bizarre organism infects, grows and abandons spent hosts. Crucially, by having isolated a strain that naturally infects a model organism, I have been able to begin testing a variety of molecular hypothesis as to which host machinery is necessary for observed phenotypes, demonstrating that specific neurons and genes are not involved in mediating end-of-life behaviors. With stable *E. muscae* CNE1 culture, I have been able to show that the fungus first travels to the brain and central nervous system in infected flies before proliferating uncontrollably in the body cavity. I have also, with the aid of Michaels Bronski and Eisen, sequenced the genome of the isolate, and, as a consequence, have been able to assay gene expression of both host and fungus over the course of infection. Ultimately, this work represents only the beginning of what is possible in the *E. muscae* CNE1-*D. melanogaster* system, not just for understanding the molecular basis of host behavioral manipulation, but also for studying varied aspects of host-microbe interactions.
# Table of Contents

**Chapter 1:** Introduction........................................................................................................................................... ii

**Chapter 2:** Stable host gene expression in the gut of adult *Drosophila melanogaster* with different bacterial mono-associations .............................................. 10

**Chapter 3:** Or22 allelic variation alone cannot account for differences in discrimination of yeast-produced volatiles by *D. melanogaster* ......................... 40

**Chapter 4:** Discovery and isolation of *Entomophthora muscae* CNE1, a strain of a behavior-manipulating fungal pathogen that naturally infects *D. melanogaster* .................................................................................................................. 60

**Chapter 5:** Characteristics of *E. muscae* CNE1 and initial experiments probing its effect on fly behavior .................................................................................................................. 78

**Chapter 6:** Progression of *E. muscae* CNE1 infection in *D. melanogaster*: gene expression and histology ............................................................................................................... 102

**Chapter 7:** Progress towards sequencing the genome of *E. muscae* CNE1 .................................................................................................................................................. 124

**Chapter 8:** Concluding thoughts and future directions .......................................................... 134

Bibliography........................................................................................................................................................................... 136

Appendix................................................................................................................................................................................ 146
Acknowledgements

The dissertation sitting in front of you would never had been had it not been for some truly wonderful people (and dogs).

Mike (aka BossMike, El Jefe, future senator of California): Thanks for believing in me, for always being excited about science, and, perhaps most importantly, for putting up with six years of bad puns.

Will: Thanks for taking me under your wing. You were always there to bounce ideas off of and offer advice even when everyone in the Eisen lab thought I was nuts. (Fun fact: I am nuts). Your time and mentorship got me to where I am now. I wish you all the best.

Kristin, Russell, Damian and Richard Calendar: Thanks for listening to my crazy ideas and helping me find my way.

Richard Humber: Thank you for embracing a neophyte mycologist and for all the knowledge and enthusiasm you’ve provided.

Alli, Kelly, Bronski, Elizabeth, Stadler, Steve, Ciera, Peter, Aaron, Jackie, Ashley, Jenna, Xiaoyong, Augusto, the Matts and all of the rest of the Eisen crew who it has been my privilege to share a lab with: you guys are the best. Thanks for teaching me everything I learned in graduate school.

Vivian, Benjamin, Ted, Josh, Thad and the rest of the Ludington lab: Thanks for letting me be part of your growing lab and share in your enthusiasm for all things related to fly guts.

My MCB 2011 classmates: Thanks for being amazing, inspiring and supportive throughout the last six years.

My family: Thanks for still talking to me, even though I do horrible things to flies.

Nora: No thanks for barfing on the morning of my qualifying exam, growling at every male member of the Eisen lab and generally being ill-behaved. But thanks for being there when the going got ruff.

Bruce: Thanks for keeping tabs on Nora. Please stop peeing on my textbooks.

Kevin: Thanks for helping me through all the ups and downs of the past six years. You don’t know how wonderful you are.
Chapter 1: Introduction

Microbes affect animal behavior

Animals live in a world dominated by microbes (1). While this fact might now be taken for granted, it is a relatively new revelation in biology, made possible by advances in sequencing technology that have allowed us to detect organisms in our environment that have previously been unobservable (e.g. are low abundance or unculturable). Though the term “microbe” has long carried connotations of animal antagonism (there’s a reason why microbes are often referred to “germs”), it is now apparent that animals have diverse relationships with the microbes with which they co-exist, ranging from adversarial to cooperative.

The microbes with which animals interact are not ignorant of their animal neighbors. Having lived in close contact through evolutionary time, microbes as well as animals have evolved strategies to interact with their hosts to improve their own fitness. Given this two-way interaction, it should come as no surprise that microbes affect various facets of animal biology, including behavior. Some of the most widely-known examples of microbes that alter animal behavior are parasites: they induce behaviors that are generally otherwise unobserved in the host and that overtly favor parasitic over host fitness. For one example, fungi within the species complex Ophiocordyceps unilateralis, known popularly as “zombie ant fungus”, infect specific species of ants, inducing infected animals to wander away from the nest, climb a nearby plant and die biting the underside of a leaf (2–4). After death, a fruiting body emerges through the ant’s head and disperses infectious spores. The exquisite positioning of death to benefit the parasite’s fitness strongly suggests manipulation by the parasite. Another example is the protozoan parasite Toxoplasma gondii, which can only sexually reproduce in its primary host, cats (5). When mice or rats (secondary hosts) become infected by consuming fecally-deposited oocytes, they exhibit a loss of innate fear to cat-derived odors, leaving them more susceptible to be predated on by a cat (6–9). This resultant behavioral change clearly serves the parasite, not the host, again strongly suggesting parasitic manipulation.

A growing body of work has shown that influencing animal behavior is not unique to pathogenic microbes, though the benefit to the microbe in many cases is less clear. In spotted hyenas (Crocuta crocuta), evidence suggests that bacteria occupying the scent pouch bacteria mediate chemical communication between individuals (10). In rodents, multiple groups have reported changes in anxiety, learning, locomotion and social behavior as a result of modifying the composition of the gut microbiome (11–20). Additional work in mice has found that the production of a volatile compound involved in intraspecies social communication, trimethylamine, is produced by gut bacteria and metabolized in a sex-dependent manner by mouse liver enzymes (21). In the Hawaiian bobtail squid (Euprymna scolopes), luminous bacteria Vibrio fischeri are required for proper development of the squid’s light organ and allow the squid to hunt using them as a light source (22). These examples all demonstrate significant impacts that microbes make on host behavior, though how these changes benefit the implicated microbes is, for the most part, uncertain.
Why study flies?

It has been the focus on my doctoral work to establish and utilize a laboratory-tractable model to uncover the molecular bases by which microbes influence animal behavior. To this end, I have employed the fruit (or more accurately, pomace) fly *Drosophila melanogaster* as a model host. *D. melanogaster* is an ideal system to study these interactions for a variety of reasons. First, flies exhibit very close relationships to microbes. In the wild, they spend the most important parts of their life (development, mating, feeding) in rotting fruit; their young are dependent on the yeast found in rotting fruit for proper development, their digestive tracts are occupied by bacteria with metabolic ties to yeast and they are susceptible for many microbial pathogens (23–29). In addition to their many natural interactions with microbes, fruit flies have a well-annotated genome and a wealth of genetic and molecular tools that can be employed to define the mechanisms underlying biological phenomena. Finally, flies exhibit a variety of interesting and quantifiable behaviors (e.g. feeding, olfactory attraction, mating) yet possess only about 100,000 neurons (in contrast to millions in vertebrates such as fish and mice), making them more tractable organisms in which to study behavioral circuits (30).

Throughout my doctoral work, I’ve used the model *D. melanogaster* to explore the relationship between microbes and the behavior of their animal hosts for three very different systems. I began with gut bacteria *Acetobacter* and *Lactobacillus*, transitioned to nutritional source *Saccharomyces cerevisiae* and most recently have studied the fungal pathogen *Entomophthora muscae*. Here, I will present an overview of each of these systems to guide the reader in the research chapters that follow.

Behavior and the gut microbiome

Every gut-bearing metazoan so far analyzed, including deuterostomes, ecdysozoans, and lophotrochozoans (31–35), possesses gut microbes1. Our new appreciation that metazoan guts support complex microbial ecosystems has led to a reevaluation of how these microbes can affect their host’s biology and revealed that gut microbes can affect animal behavior. In rodents, studies have shown that gut microbes affect behaviors ranging from anxiety to memory (Table 1.1). In humans, there is high comorbidity between psychiatric disorders (e.g. anxiety, depression, autism) and gastrointestinal pathologies (e.g. irritable bowel syndrome, Crohn’s) (36–38).

---

1 While the terms "gut microbiome" and "gut microbes" technically encompasses all microbial life associated with an organism’s digestive tract (i.e. bacteria, fungi, archaea and viruses), the term is commonly employed to describe only the bacteria components of this community as they tend to comprise its bulk. I will follow suit in my usage of the term in this introductory chapter.
Table 1.1. Rodent studies reporting behavioral changes following microbiota changes.

<table>
<thead>
<tr>
<th>Strain, species, sex</th>
<th>Treatment</th>
<th>Observed behavioral change</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>BALB/c mice, male</td>
<td>Oral antibiotic</td>
<td>Increased exploratory behavior</td>
<td>(11)</td>
</tr>
<tr>
<td>BALB/c mice, male</td>
<td>Oral <em>Lactobacillus rhamnosus</em></td>
<td>Reduced anxiety- and depression-like behaviors</td>
<td>(12)</td>
</tr>
<tr>
<td>NMRI mice, male</td>
<td>Germ-free</td>
<td>Increased motor activity, reduced anxiety</td>
<td>(13)</td>
</tr>
<tr>
<td>Swiss Webster mice, female</td>
<td>Germ-free</td>
<td>Anxiolytic-like behavior</td>
<td>(14)</td>
</tr>
<tr>
<td>C57BL/6N mice, male</td>
<td>Maternal immune activation (MIA) + oral <em>Bacteroides</em></td>
<td>Rescue of communication, anxiety, repetitive behaviors observed in MIA offspring</td>
<td>(15)</td>
</tr>
<tr>
<td>BALB/c mice, male</td>
<td><em>Bifidobacterium infantum</em></td>
<td>Decreased motor activity</td>
<td>(16)</td>
</tr>
<tr>
<td>BALB/c mice, male</td>
<td>Germ-free then environmental exposure</td>
<td>Decreased anxiety</td>
<td></td>
</tr>
<tr>
<td>C57BL/6 mice, male</td>
<td>Antibiotic treatment then microbiome transplantation from high-fat diet (HFD) or normal diet (ND) animals</td>
<td>Decreased exploratory, learning and stereotypical behavior in HFD microbiota animals</td>
<td>(17)</td>
</tr>
<tr>
<td>F344 rats, male</td>
<td>Germ-free</td>
<td>Increased anxiety</td>
<td>(18)</td>
</tr>
<tr>
<td>C57BL/6 mice, male</td>
<td>Germ free then gavaged with HFD or ND microbiota at 4 or 8 wks</td>
<td>ND at 8 wks or HFD at 4 or 8 wks showed reduced reciprocal social interaction, decreased sociability and preference for social novelty compared to ND at 4 wks</td>
<td>(19)*</td>
</tr>
<tr>
<td></td>
<td>Offspring of HFD mothers + <em>L. reuteri</em> (4 wks)</td>
<td>ND-like social behaviors</td>
<td></td>
</tr>
<tr>
<td>Kunming mice, male</td>
<td>Germ-free</td>
<td>Reduced anxiety, better memory, decreased depressive behavior</td>
<td>(20)</td>
</tr>
<tr>
<td></td>
<td>Microbiota from major depressive disorder patient</td>
<td>Increased depression, increased anxiety</td>
<td></td>
</tr>
</tbody>
</table>

*Wks = weeks; weaning occurred at 4 weeks

Though much of our understanding of how the microbiome impacts host biology comes from vertebrates, especially mice, flies have emerged as an alternative model system. In addition to the myriad molecular tools and relative simplicity of the *D. melanogaster* nervous system, flies offer the additional advantage of harboring a relatively sparse microbial community (24–27). There are about five taxa in laboratory-reared flies, generally *Acetobacter* and *Lactobacillus*, but even in wild flies only tens of taxa are observed (28, 39). This is a stark contrast to the communities of mammalian guts, which
tend to contain hundreds to thousands of taxa (40). The relative simple makeup of the fruit fly gut microbiome makes it much more feasible to understand the contribution of single and multiple species to the overall community. As a further benefit, the microbial species that comprise the guts of laboratory-reared flies can be isolated and cultured in the laboratory, which is not true of many of the fastidious species in the vertebrate gut.

The focus of most of the gut microbiome work in Drosophila has focused on fly physiology, especially development and nutrition. While axenic (germ-free) flies develop to adults indistinguishably from their conventionally-colonized peers on a complete diet, axenic flies show severe growth deficits compared to conventional flies when reared on a yeast-limited diet (41). Mono-associating flies with Lactobacillus plantarum restores this delayed developmental timing via a mechanism that dependent on TOR signaling and amino acid transport in the fat body (41). Similarly, another study found that germ-free flies on poor diets grow more slowly or die as first instars, but their growth is rescued by mono-associating with Acetobacter pomorum (42). This rescue depends on A. pomorum genes involved in the periplasmic pyrroloquinoline quinone-dependent alcohol dehydrogenase-dependent oxidative respiratory chain (PQQ-ADH). While PQQ-ADH mutant A. pomorum is able to colonize flies as well as wild-type A. pomorum, flies mono-associated with the mutant PQQ-ADH strain show reduced insulin-like signaling and delayed growth. Additionally, supplementing poor diet with acetic acid (a PPQ-ADH-dependent product), while alone unable to rescue the growth defect, in conjunction with mutant PQQ-ADH rescues insulin signaling and growth phenotype. Other work has demonstrated that nutritional state of hosts reared on rich media differs between axenic flies and flies associated with defined consortia of gut bacteria (26, 43, 44). For example, flies monoassociated with Acetobacter spp have lower triglyceride levels than flies associated with Lactobacillus spp glucose levels; flies associated with both Acetobacter and Lactobacillus have lower levels still (43).

While other roles of gut microbial taxa in Drosophila biology have been firmly established, at the onset of my research only one study had examined how gut microbes impacted fly behavior. Excitingly, this study reported that increased levels of Lactobacillus plantarum (a species naturally associated with laboratory-reared OreR) conferred a homotypic (like with like) mating preference in adults (45).

Given the mounting evidence that gut microbes are important for Drosophila biology and gut microbes play roles in behaviors of other animal hosts, it would be interesting to know how different microbial taxa affect the host at the molecular level and to use this information uncover additional ways that microbes influence fly behavior.

Saccharomyces cerevisiae: food and ecological partner of D. melanogaster

D. melanogaster is known as the fruit fly because it feeds, mates and develops on fermenting fruit, but the fruit flies’ association with fruit is predominantly owed to the presence of fermenting yeasts, especially Baker’s yeast, Saccharomyces cerevisiae. S. cerevisiae is a vital source of nutrition for adults and larvae. S. cerevisiae alone satisfies all of the metabolic and energetic needs of a developing fly (46). Correspondingly, mated females are attracted to and prefer to oviposit on yeast-colonized substrates (47). Still, nutrition isn’t the only advantage for co-localizing with yeast. Yeast also affords a degree of microbial protection to inchoate flies by aerobically fermenting sugar to ethanol (the
Crabtree effect), thereby killing ethanol-intolerant (i.e. most) organisms in its environment (48–50). Not only is *D. melanogaster* able to better than most at tolerating ethanol, but adults even demonstrate a preference for substrates that contain low amounts of alcohol over those that are completely ethanol-free (51).

Yeast also benefit from their interaction with flies. Sticky yeast cells cannot be dispersed by wind, so they must rely on alternative routes to reach new substrates. Yeast cells can adhere to the cuticle of feeding adults and hitch a ride to a new niche (52), or, as spores, can survive passage through the digestive tract to colonize a new substrate (53). Interestingly, evidence suggests that the latter route confers the additional benefit of increasing the rate of yeast outbreeding (54).

The close relationship between flies and yeast is contingent on flies being able to find yeast in a large and complex environment. Since humans began imbibing fermented beverages, it has been known that yeast produce an array of aromatic volatile compounds in addition to alcohol during fermentation. As it turns out, those odors are not just pleasant to us. Flies are able to sense many of these byproducts of yeast fermentation and some of these compounds, including isoamyl acetate (banana), ethyl hexanoate (pineapple, banana), ethyl octanoate (banana, apricot, apple) and ethyl acetate (pear), are attractive to flies (55).

Like mammals, flies sense smells through their olfactory system which is located in the antennae and maxillary palp. These structures are comprised of sensory hairs, or sensilla, each of which is innervated by one (or sometimes two) olfactory neurons (56). Each olfactory neuron, in turn, expresses up to four different ligand-specific odorant receptors (ORs), each of which forms a heteromeric complex with a highly-conserved co-receptor (Orco, formerly Or83b) to confer function (57). Olfaction occurs when an odorant diffuses into a sensillum, binds to its recognizing odorant receptor complex and activates the olfactory neuron in which that complex is expressed (58). This results in the relaying of a signal to the brain's antennal lobe that is processed as the sensation of odor (58).

Flies have a total of 62 unique odorant receptor genes (59). Odorants recognized by each of these genes has been determined by ectopic expression in an otherwise vacant neuron then presenting a panel of odorants and measuring the electrophysiological response to each (59–61). In this way, the timing, and magnitude of the response of each OR for a series of compounds has been determined, as well as their sensitivity to compounds at different concentrations.

The production of these pleasant-smelling esters during fermentation is energetically costly and regulated by yeast, suggesting that they are not just non-consequential byproducts (62). Indeed, recent work demonstrated that ATF1, which encodes alcohol acetyl transferase, a gene involved in synthesizing esters, mediates attraction of yeast to fruit flies and promotes dispersal by insect vectors (63). However, not all yeasts are equally attractive: volatiles that yeast produce vary depend on substrate, growth conditions and strain (55, 64). The volatiles produced can provide information that the fly uses to discriminate between yeasts; unsurprisingly, some fly lines do demonstrate robust preferences for particular yeast strains over others (55). Through the efforts to define the molecular logic of the fly olfactory system, we know which odorant receptors respond to particular odorants and where these receptors are expressed. Still, we don’t understand which components of sensation that flies use in a natural context to differentiate between similar volatile bouquets and to choose one substrate over another.
Entomophthora muscae: The mind-controlling fly destroyer

Entomophthora, from the Greek “Entomo” for insect and “phthora” for destroyer, is an aptly-named genus consisting of fungi that infect, feed on and ultimately kill their insect hosts. *E. muscae* (originally *Empusa muscae*) was first described by Cohn in 1855\(^2\) and is a species that specializes in parasitizing dipterans (65). Infection by *E. muscae* is striking, not only because the fungus covertly consumes its host from within, but because it causes drastic behavioral manipulations, the most obvious of which occur at the end of the fly’s life (Figure 1.1).

A fly that is critically infected with *E. muscae* proceeds through a stereotyped series of actions prior to its untimely death. These behaviors begin a few hours before sunset, where the fly first begins to climb a nearby substrate, exhibiting what is known as summit disease (66). Eventually the fly loses the ability to walk and extends its proboscis, which adheres to the substrate via fungal holdfasts (67). The fly then lifts its wings up and away from its dorsal abdomen stopping when they are about 90 degrees from the body axis (68). The fly dies in this position adhered to its substrate, and in so doing ideally positions the fungus inside to continue its cycle of destruction.

Over the course of the next few hours, the fungus within the dead fly differentiates into structures called conidiophores that emerge through the weakest points in the fly's cuticle. The fungus usually grows through the intersegmental membranes of the dorsal abdomen, giving the cadavers a distinct banding pattern (69). A primary conidium (aka “spore”) forms at the tip of each conidiophore; once mature these conidia are forcibly ejected into the surrounding environment with the goal of landing on a susceptible fly host (70, 71). Launched primary conidia are polynucleated, campanulate (bell-shaped) and are surrounded by a sticky “halo” (72). The discharge mechanism for primary conidia is still debated: while some maintain that conidia are launched via a “fungal canon” mechanism and the “halo” is comprised of co-ejected cytoplasm (72), a study looking at primary conidia of several *Entomophthora* spp claimed that the “halo” is comprised of material from rupture of the conidium’s outer cell wall and is evidence for an alternative ejection method (73).

---

\(^2\)Though 1855 is the first time *E. muscae* was formerly reported, Ferdinand Cohn’s assertion that “Dass diese sonderbare Todesweise der Fliegen, die jedem Kinde bekannt ist, allein den Naturforschern entgangen sei, ist nicht anzunehmen” (Google translation: That this strange means of death of the flies, which is known to every child, has escaped the only naturalists, cannot be accepted) suggests that *E. muscae* was well known to the people of this area, albeit not reported in the scientific literature until Cohn’s publication in 1855.
If the primary conidium misses its target or fails to germinate upon landing on the host (77), it can sporulate anew to generate a smaller secondary conidium (70). One study has claimed that primaries form secondaries regardless of whether they hit their target (78). Evidence suggests that, for some strains, secondary conidia are more infectious than primary conidia (76, 78). Once the spore hits its target, the conidium will use mechanical and enzymatic force to bore through the cuticle and into the fly's hemolymph (69, 79). Now inside the fly, the fungus replicates using the fat body as an energy source (69). When available resources are exhausted, the fungus elicits the end of life behaviors and the destructive fungal life cycle begins again.

Cohn made his initial observations of E. muscae in the humble Stubenfliegen (house fly, Musca domestica); ever since, house flies have been the predominant host in which E. muscae has been studied. These fungi tend to be fastidious to culture both in vitro and in vivo: as a result, what we know about E. muscae relies heavily on field work following epizootic outbreaks in wild fly populations and to a lesser extent on laboratory-based experiments with the fungus. E. muscae has been observed in the wild in North America, South America, Europe and India (71, 80, 81) generally in farms and fields. For temperate regions, its typically reported in the spring or fall and can occur in wet or dry conditions (71, 82, 83). Frequency of infection appears to scale proportionately with population density of hosts and older hosts are observed to form resting spores rather than infectious conidia to propagate vegetative infection (71, 84). Time from exposure to death increases with host size and decreases with the initial exposure: males tend to die before females, but there appears to be no distinction between sexes when weight is controlled for (85).
Younger hosts appear more susceptible to infection and are more likely to produce infectious cadavers (85).

There are several varieties of E. muscae, which have different morphologies of primary conidia and which tend to infect particular species of flies. In the absence of molecular data to resolve taxonomy, all of these fungi are classified as members of the E. muscae species complex, and keys to identification have been developed based primarily on conidia morphology and fly host (80, 86). Curiously, morphology of a given E. muscae strain has been shown to vary between different hosts, though the significance of this is not understood (87). A range of fly species and even non-dipterans can be infected and killed by E. muscae in a laboratory setting, though not all infected hosts manifest the stereotypical end of life behaviors (88, 89). Still, while experimentally E. muscae has been shown to have a wide range of hosts, it’s difficult to conclude if this potential generalism actually matters in an ecological context. For example, though a given strain of E. muscae could kill both a house fly and a fruit fly, we don’t know if a cadaver of one of these species would actually come into contact with a healthy individual of the other species. Given this issue and a paucity of molecular data, we have a very poor understanding of the diversity within this group. There is no published genome for any Entomophthora species and their ploidy is unknown. Just recently, work surveying tandem epizootics in one host (Delia radicum) then another (Coenosia tigrina) showed that these outbreaks were caused by two distinct subpopulations of E. muscae that were undergoing low rates of genic exchange, providing the first evidence for recombination in Entomophthora (83).

In the laboratory, house flies have been used to study the ethology of end of life behaviors induced by E. muscae infection (i.e. summitting, proboscis extension and wing raising). These behaviors occur reliably in this order and take on average 75 minutes to execute (68). This same study also examined timing of death of infected flies, demonstrating that flies which die from patent E. muscae infection (i.e. following this stereotyped set of behaviors) do so only around sunset each day. Any animals that survive past sunset will not die until the following day, confirming that death by E. muscae is a gated phenomenon.

In addition to a stereotyped death ritual, house flies infected with E. muscae have been shown to exhibit atypical temperature preferences. Early on the infection (approximately within the first half of fungal incubation), flies exhibit “behavioral fever”, exhibiting a preference for warm environments (35-40°C), (74, 75). Behavioral fevering improves chances of survival from infection; the earlier on in infection the fly fevers, the better the outcome (74). Behavioral fevering, however, is not a specific response to E. muscae but rather a general strategy of insect hosts for fighting pathogens (90). Similar to fevering in endotherms, behavioral fevering raises the internal temperature of the insect body with the goals of stalling pathogen growth and decreasing pathogen virulence to extend or improve the chances of host survival (90). Interestingly, later on during E. muscae infection (generally the latter half of fungal incubation), infected flies demonstrate a preference for cooler temperatures (74, 75). Given that the fungus grows optimally in cool conditions, this late-manifesting preference is hypothesized to be dictated by the fungus (74). Unlike behavioral fevering, this preference for cool environments in the latter half of infection appears to be specific for E. muscae infected animals.

It has also been reported that healthy male house flies are more sexually attracted to infected female cadavers than they are to uninfected dead females (91). Unsurprisingly, healthy males that attempt to copulate with infected female cadavers have significantly higher chances of getting infected and killed by E. muscae, which would be of obvious
benefit to the fungus (91). A follow-up study assayed the attractiveness of male flies to infected male or female cadavers, finding that males are more attracted to female cadavers (92). However, analysis of hydrocarbons of infected and uninfected female cadavers showed that infected cadavers actually produce attractive pheromones in lower abundance than their uninfected counterparts. It’s possible that non-pheromonal cues account for the attractiveness of infected females to males. However, as the initial finding has never been reproduced and house flies have been reported to be interested in dead house flies regardless of infection status (82), it would be prudent to be skeptical of this result until it is corroborated by additional independent studies.

Despite inspiring curiosity and intrigue for over a century, how *E. muscae* achieves control of its host remains poorly understood, with essentially no information as to what’s occurring at the molecular level. Being able to study this fungus in a host with established molecular and genetic tools, such as *D. melanogaster*, would allow us to begin to understand the means by which *E. muscae* manipulates host behavior.

**Contents of this dissertation**

I’ll begin by describing how *D. melanogaster* gene expression changes in response to associations with different bacteria species (Chapter 2). Then, I’ll detail my efforts to understand the role that a particular odorant receptor, Or22, plays in the attraction of *D. melanogaster* to *S. cerevisiae* grown under natural and laboratory conditions (Chapter 3). Finally, I will discuss the work that I’ve done with a wild, *Drosophila*-infecting isolate of *E. muscae* that I discovered in late 2015 (Chapters 4-7). First, I’ll describe how I brought this fungus into the laboratory to study (Chapter 4) then detail what I’ve so far learned about the fungal life cycle in the host *Drosophila melanogaster* and the basis of manipulated host behaviors (Chapter 5). Subsequently, I’ll present a systematic morphological and molecular description of the progress of infection (Chapter 6) followed by a description of the efforts made to sequence and assemble the genome of the *E. muscae* isolate (Chapter 7). Finally, I will present my perspective on the field of microbes and behavior and suggest where to focus future efforts (Chapter 8).
Chapter 2: Stable host gene expression in the gut of adult Drosophila melanogaster with different bacterial mono-associations

Note the contents of this chapter, with some minor modifications, were previously published as:


Abstract

There is growing evidence that the microbes found in the digestive tracts of animals influence host biology, but we still do not understand how they accomplish this. Here, we evaluated how different microbial species commonly associated with laboratory-reared Drosophila melanogaster impact host biology at the level of gene expression in the dissected adult gut and in the entire adult organism. We observed that guts from animals associated from the embryonic stage with either zero, one or three bacterial species demonstrated indistinguishable transcriptional profiles. Additionally, we found that the gut transcriptional profiles of animals reared in the presence of the yeast Saccharomyces cerevisiae alone or in combination with bacteria could recapitulate those of conventionally-reared animals. In contrast, we found whole body transcriptional profiles of conventionally-reared animals were distinct from all of the treatments tested. Our data suggest that adult flies are insensitive to the ingestion of the bacteria found in their gut, but that prior to adulthood, different microbes impact the host in ways that lead to global transcriptional differences observable across the whole adult body.

Introduction

The digestive tracts of virtually all animals examined to date are inhabited by microbes and there is increasing evidence that interactions between gut microbes and their animal hosts influence a wide range of host phenotypes (1). However, the molecular basis for these effects is poorly understood.

Drosophila melanogaster has become a model for the study of host-microbe interactions in the gut owing to the relative simplicity of the tissue and its microbial community. The Drosophila gut microbe community is comprised of about 50 bacterial species in wild flies (28, 93–95) and about ten species in laboratory-reared flies (24–26, 28, 93). In laboratory-reared animals, the majority of these bacteria are lactic acid and acetic acid producers (members of family Lactobacillaceae and Acetobacteraceae, respectively) that can be cultured outside of the host. Drosophila embryos can be stripped of their endogenous microflora using household bleach (96) and either maintained under sterile conditions to generate axenic (germ-free) animals or treated with a defined set of microbes to generate gnotobiotic (of known microbial content) animals (43).

A variety of phenotypes have been associated with the presence and/or specific composition of microbes in the Drosophila melanogaster gut including nutrition and
metabolism (26, 43, 97), intestinal cell growth (98–100), development (41, 42), lifespan (25, 101–103) and a variety of different behaviors, including social attraction in larvae and assortative mating in adults of two laboratory-reared strains (45, 104, 105).

We screened a set of behaviors in conventional flies (olfactory preference) or flies with modified microbiota (mating, food preference) looking for robust differences that could be due to influence by gut microbes (data not shown). In the absence of an obvious behavioral phenotype, it seemed prudent to more generally consider how different gut microbe taxa impacted host biology before continuing forward.

The presence of microbes in the environment and gut has a significant effect on physiology, morphology and gene expression in the adult *D. melanogaster* gut (97, 103, 106), however, the effect of individual microbe species on the gene expression of adult guts has not been examined. Given the known differential phenotypic effects of different bacterial taxa, we hypothesized that the epithelial cells of the digestive tract might respond to different bacteria in different ways, and that these responses might provide clues to the molecular mechanisms that underlie the effects of specific microbes of *Drosophila* physiology and behavior.

Here we describe the results of a series of experiments using mRNA sequencing to analyze gene expression in the guts of a laboratory line of *D. melanogaster* reared from embryogenesis to early adulthood with different combinations of bacteria and yeast species commonly found in laboratory populations.

**Results**

**Limited variation in transcription in adult *D. melanogaster* gut in response to mono-association with different bacterial species**

To elucidate the effects of specific bacterial species on host gut gene expression in the adult fruit fly, we implemented previously published protocols (43) to mono-associate laboratory stocks of *D. melanogaster* with three bacterial species previously shown to associate with *D. melanogaster* in the laboratory: *Acetobacter pasteurianus*, *Lactobacillus brevis* and *Lactobacillus plantarum*. *A. pasteurianus* and *L. brevis* were selected based on their high abundance in a survey of wild-type female (CantonS) flies reared on standard media in our laboratory (Figure 2.S1) as well as their ubiquity in other gut microbe surveys of *D. melanogaster* (24–26, 28, 93). *L. plantarum* was chosen because it has been observed in several other gut microbe surveys and has been implicated in mediating specific aspects of fly development and behavior (41, 45).

We opted to use well-characterized type strains for the two *Lactobacillus* species (*L. plantarum* from ATCC, *L. brevis* from NBRC) for our initial experiments, even though neither was derived from flies, as both had fully sequenced genomes that we anticipated would be of use in downstream analyses. At the time, a well-characterized strain was lacking for *A. pasteurianus* so we opted to use a fly-derived strain.

Mono-associated animals were prepared on three different occasions. Our approach was to inoculate sterile medium with our microbe of choice before introducing germ-free embryos. Embryos were reared on this once-inoculated medium until the point of sampling (five days post-eclosion) so as to maintain a constant association between fly and the species of interest. For each of our three preparations, we dissected and sequenced mRNA
libraries from three young adult females (five days post-eclosion) to produce three biological replicates for each treatment, giving us a total of 27 samples.

We visually examined variability in gene expression within and between experiments and treatments across samples organized by either experimental date (Figure 2.1A) or bacterial mono-association (Figure 2.1B). Contrary to our expectation, we did not observe any significant differences in gene expression associated with the different bacterial treatments.

**Figure 2.1. Limited variation in gut gene expression with bacterial mono-association.** Expression data from guts dissected from five-day post-eclosion, Wolbachia-free, mated female CantonS *D. melanogaster* individuals mono-associated with one of three bacteria (Ap = A. pasteurianus, Lbrev = L. brevis, Lp = L. plantarum) were clustered by gene (average linkage, uncentered correlation) after first filtering out genes that lacked three
instances of FPKM greater than two (Gene Cluster 3.0). FPKM values for each gene were normalized to range from -1 to 1 before plotting. A) Samples arranged by bacterial treatment. B) Samples arranged by date of experiment. Scale bars for each heatmap are given to the right of the plot.

To confirm this visual impression, we used a simple statistical test (ANOVA) to identify individual genes differentially expressed between treatments. After applying a Bonferroni correction for multiple testing, the expression levels of only two genes were significantly associated with treatment: CAH2, a carbonic anhydrase, and CG17574, a gene of unknown function (Figure 2.2). Even relaxing our significance threshold, and considering the fifty genes that demonstrate the most significant differences in expression within this set, it is clear that there are minimal differences in gene expression between the three different bacterial mono-associations (Figure 2.2). Instead, it appears that each bacterial species alone affected host gene expression in the same way.

Figure 2.2. Genes showing greatest difference in expression values from dissected adult guts as determined by one-way ANOVA. A) Scatterplot of log10-transformed FPKM values for each bacteria mono-associated gut replicate (Ap = A. pasteurianus, Lbrevis = L. brevis, Lp = L. plantarum). Genes are ordered from lowest ANOVA p-value (top) to highest (bottom). P-values have undergone a Bonferroni correction for multiple testing. B) Data from A presented as a heatmap. FPKM values for each gene are linearly normalized to range from -1 to 1 before plotting.
Host transcription in the gut is markedly different between conventional and yeast mono-association compared to axenic and bacteria mono-association treatments

After observing a lack of differences in our initial set of samples, we decided to expand our dataset to include other treatments that might shed light on this unexpected result. To this end, we measured gene expression in guts from conventionally reared flies, axenic flies and flies mono-associated with the yeast *Saccharomyces cerevisiae* (which is known to be associated with flies in the wild (107) and can alone provide complete nutrition for the developing fly (46)) rather than bacteria. Additionally, we tested if providing a simplified multi-species microbial community (either all three of *A. pasteurianus*, *L. brevis* and *L. plantarum* with or without *S. cerevisiae*) would yield a transcriptional program more similar to that of conventional than to mono-associated samples.

Based on studies demonstrating that microbes play critical roles in animal development, we anticipated that the expression pattern of guts from axenic flies would be distinct from those of mono-associated, poly-associated (bearing a simplified microbial community) or conventional animals (97, 106). We also expected expression in mono-associated and poly-associated flies to either closely resemble that of conventionally-reared flies or lie somewhere in between the axenic and conventional samples.

To test these predictions, we dissected and sequenced mRNA from wild-type, 5-day-old, female CantonS flies that were either axenic, conventional, mono-associated with *S. cerevisiae* or poly-associated with an equal amounts of *A. pasteurianus*, *L. brevis* and *L. plantarum* with or without *S. cerevisiae*.

As above, we first examined the aggregate gene expression data visually (Figure 2.3A). The expression data clearly showed that guts from gnotobiotic animals whose treatment included yeast gave transcription patterns most similar to those of guts from conventional animals. We were surprised to observe that rather than being distinct from all other samples as we had expected, transcriptional patterns of guts from axenic animals most closely resembled that of animals either mono- or poly-associated with bacteria. Taken together, samples appeared to fall within one of two transcriptional regimes: “conventional-like” (conventional and yeast-mono and poly-associations) and “bacteria-like” (axenic, bacteria mono- and poly-associations).
Figure 2.3. Yeast drives genome-wide difference in gut gene expression. A) Average linkage hierarchical clustering was performed in Gene Cluster 3.0 across all genes that are expressed at least at two FPKM in at least two out of 11 samples. Bacteria mono-association data has been averaged across each treatment to collapse down into a single column. FPKM values for each gene are normalized to range from -1 to 1 before plotting. Abbreviations: Ap avg = average for *A. pasteurianus*-mono-associated samples; Lbrev avg = average for *L. brevis*-mono-associated samples, Lp avg = average for *L. plantarum*-mono-associated samples, 3bac = poly-associated (without yeast), Ax = axenic, Conv = conventional, Yeast = *S. cerevisiae*-mono-associated, 4mic = poly-associated (with yeast). Scale bar is shown at bottom right. B) Top) heatmap of 579 genes that are overexpressed in axenic, bacteria-mono-associated and poly-associated (without yeast) guts compared to other gut samples (Bonferroni p-value>0.05, ANOVA). Bottom) Results from Panther GO-Slim biological function enrichment test (108) for gene set above compared to reference set of all genes observed across all gut datasets (556 were identified by Panther and used for analysis out of 579) C) Top) Heatmap of 1728 genes that are overexpressed in conventional, yeast mono-associated and poly-associated (with yeast) compared to other gut samples (Bonferroni p-value>0.05, ANOVA). Results from Panther GO-Slim biological...
processes enrichment test with 1728 (1663 identified) genes compared to reference set of all genes observed across all gut datasets. Note for B) and C): all individual sample values were used for ANOVA analysis, not the average value as plotted in A).

To understand which genes most distinguished these two transcriptional groups, we again performed a gene-by-gene ANOVA. This analysis revealed more than 2,000 genes with significant expression differences between the two groups (p-value under 0.05 after Bonferroni correction). Genes that were more highly expressed in “conventional-like” samples were enriched for annotations for a variety of metabolic processes, most notably involving lipids and amino acids (2.3B). Genes more highly expressed in “bacterial-like” samples were also enriched in metabolism annotations, although this enrichment was less pronounced and more general than for the former set (Figure 2.3C).

Overall, these gut expression data are consistent with a model in which the presence of live yeast shapes the transcriptional program of the gut in mated, young adult females. Specifically, these results argue that yeast induces metabolic changes that are not evoked by association with any one species of bacteria associated with laboratory-reared D. melanogaster or a simplified community comprised of the all three bacterial species considered herein.

Global host gene expression in conventional animals is distinct among treatments

As our expectation that host transcription within the gut (i.e. proximal transcription) would change in response to mono-association with different bacterial taxa did not hold true, we were curious if the same would hold true for whole animals. We therefore extracted and sequenced mRNA from individual whole, 5-day post-eclosion, mated female CantonS flies that were prepared with the same microbial treatments as our gut samples (i.e. mono-associated with either A. pasteurianus, L. brevis, L. plantarum or S. cerevisiae, poly-associated with A. pasteurianus, L. brevis and L. plantarum, poly-associated with A. pasteurianus, L. brevis, L. plantarum and S. cerevisiae, axenic or conventional).

From our observation that gene expression in the gut tracked with exposure to yeast, we hypothesized that global transcription would also be dependent on the presence of S. cerevisiae. Based on other gene-expression studies examining axenic and conventional animals, we anticipated that the majority of the affected genes still would be those expressed in the gut (97, 106).

Data from whole animals (Figure 2.4A) show that conventionally-reared flies have a transcription profile distinct from all of the samples tested. ANOVA comparing conventional samples to all other samples showed that ~1700 genes display different patterns of expression between these two groups after applying a strict Bonferroni correction for multiple testing. Many biological processes are enriched among the genes expressed at higher levels in the conventionally reared samples (Figure 2.4B), while genes expressed at lower levels show a marked enrichment for processes involving protein folding and biogenesis (Figure 2.4C).
Figure 2.4. Analysis of gene expression trends from gnotobiotic whole flies. A) Transcriptome-wide heatmap from axenic, conventional, yeast-mono-associated, bacteria-mono-associated and poly-associated whole flies clustered by gene expression. Average linkage hierarchical clustering using an uncentered correlation similarity metric was performed in Gene Cluster 3.0 across all genes that are expressed at least at two FPKM across two out of eleven samples. Abbreviations: Ap = A. pasteurianus-mono-associated; Lbrev = L. brevis-mono-associated, Lp = L. plantarum-mono-associated, 3bac = poly-associated without yeast, Ax = axenic, Conv = conventional, Yeast = S. cerevisiae-mono-associated, 4mic = poly-associated with yeast. Scale bar is shown at bottom right. B) Top) heatmap of 1159 of 1385 genes that are overexpressed in conventional whole flies.
compared to other whole fly samples (Bonferroni p-value > 0.05, ANOVA). Genes absent in heatmap did not pass filtering criteria. Bottom) Results from Panther GO-SLIM biological function enrichment test (108) for gene set above (1278 genes were identified of 1385) compared to reference set of all genes observed across all whole fly datasets. C) Top) Heatmap 351 that are overexpressed in all non-conventional whole-fly samples compared to conventional whole flies (Bonferroni p-value > 0.05, ANOVA). Results from Panther GO-Slim biological processes enrichment test with gene set above (348 genes were identified out of 351) compared to reference set of all genes observed across all whole fly datasets.

Although the global analysis clearly showed that conventional samples exhibit a unique transcription pattern, we reasoned that there could still be a subset of genes that maintain a high similarity between yeast-containing and conventional samples. Gene-by-gene analysis demonstrated that there was only a small overlap in transcription between conventional and yeast-containing samples (~170 genes showed similar patterns of expression between samples, Figure 2.S2) and so better supports a model in which none of the gnotobiotic treatments tested in this study could recapitulate conventional levels of gene expression at the level of the whole animal.

To determine if any differences manifested between different bacterial mono-associated animals at the level of the entire organism that were not apparent in dissected guts, we completed a gene-by-gene analysis as per Figure 2.2. As in our guts, there are few significant differences between whole flies that are mono-associated with different bacteria once a Bonferroni correction is applied (Figure 2.S3), though the data do demonstrate more variance than what was observed in dissected guts alone.

Overall, the whole animal data suggest that, while yeast alone may suffice to recapitulate conventional host gene expression in the gut, yeast alone or in a simplified mock community are not enough to generate an animal that demonstrates an overall conventional-like transcriptional program.

Microbial load is highly variable between laboratory-reared flies raised under identical gnotobiotic conditions

Prior to conducting the experiments described above, we had been concerned about inter-animal variation in microbial load producing highly variable gene expression patterns. However, the relative lack of variation in gene expression within treatments led us to wonder if our gnotobiotic protocol generated flies with less variability in microbial load than we had anticipated.

We attempted to determine the total bacterial load of the animals we sequenced by counting 16S rRNAs in our sequencing reads. Although our sequencing library preparation protocol involves a polyA selection step to enrich for eukaryotic mRNAs, we nonetheless sequence many Drosophila rRNAs, which are not polyadenylated. We therefore expected to have sequenced some microbial rRNAs. However, we did not detect any and, in retrospect, believe this is due to a failure to lyse bacterial cells.

Instead, to determine the likely range of microbial loads that our sequenced animals possessed we repeated the preparation of mono-associated animals five times, and determined the microbial loads of multiple animals per treatment per replicate for each of these preparations. Briefly, animals were surface-sterilized, individually homogenized and plated in two dilutions on appropriate media to determine the number of colony forming units (CFUs) they contained (Figure 2.5A).
Figure 2.5. Microbial load of female *D. melanogaster* individuals. A) Log10 transformed average number of colony forming units (CFU) from plating individual gnotobiotic and conventional, laboratory-reared CantonS, Wolbachia-free, mated, 5-day post-eclosion females on two separate plates. B) Log10 transformed average number of CFU and estimated microbial cells (yeast and bacteria combined) by qPCR for individual, female, wild *D. melanogaster* raised from embryos (ranging from three to ten days post-eclosion) or caught as adults (of unknown age). The mean for each group is plotted as a horizontal line.
Consistent with previous work (106), we observed large variability in the microbial load of gnotobiotic flies, despite being reared under nominally identical conditions within a given treatment. The average load among mono-associated animals varies with bacterial species, with L. plantarum colonizing more densely than L. brevis, which in turn colonizes more densely than A. pasteurianus. Conventional animals are on average about 10 - 100x more densely colonized than mono-associated animals. Notably, animals mono-associated with S. cerevisiae on average contain only about ten viable yeast cells.

Given these observations we think it is almost certain that the animals we sequenced were effectively colonized with the relevant bacterial species, but that the amount of bacteria they contained was likely highly variable. Despite this presumed variability in the animals we sampled, we did not observe appreciable variation within and between treatments in our transcriptional data. This suggests that variability in microbial load does not play as much of a role in defining transcriptional response as we had anticipated. Still, we cannot completely exclude the possibility that the expression of a few genes did vary in response to microbial load and that we were unable to distinguish this true biological variation from experimental variation.

Wild-caught flies have comparable microbial loads to laboratory-reared gnotobiotic flies

We recognized throughout this experiment that our method of preparing gnotobiotic animals is one of many possible methods. In this study, we wanted to study young adult flies that had been stably associated with a microbe or set of microbes of interest throughout their entire life history. To do this, we inoculated an initially sterile medium with a total five million microbial cells, deposited embryos on this medium and maintained animals on this same medium until the point of sampling (five days post-eclosion). We felt that inoculating the medium once at the beginning of the experiment would most closely recapitulate the experience of flies in the wild. That is, embryos (internally sterile until hatching) and pupae (who lose their association with microbes upon remodeling of the gut and are repopulated upon eclosion (41, 109)) would acquire microbes present in their environment and and coexist with these microbes throughout their life history.

Other studies have used higher inocula for preparing gnotobiotic animals (41, 42, 110), or inoculating axenic flies upon eclosion rather than associating from the embryo (98, 103). We wanted to understand how closely our experimentally-manipulated animals reflected their wild counterparts in terms of gut microbial density. We therefore sampled wild flies from Berkeley, CA and compared their microbial load to that of animals mono-associated in the laboratory. In doing so we had to address several potential complications. It has been well established that microbial load in laboratory D. melanogaster positively correlates with age (24, 25, 41), and thus we wanted to sample animals that would be comparable in age to those used in our sequencing experiments. However, we did not want to capture newly-eclosed flies and rear them to the appropriate age in non-natural conditions, as this would deprive them of the substrate needed to replenish their natural gut microbiome and influence their microbial load (109).

Our devised solution was to first establish a stable food source (organic watermelon in a clean dish pan, referred to henceforth as the “fendel”) to continuously attract wild flies. We then began to leave various oviposition substrates (pieces of organic fruit) near the bait for 24 hours to allow wild females to deposit embryos. After 24 hours, oviposition
substrates were transferred to clean vials and wild embryos were reared to adulthood for collection and sampling. Since our rearing method necessarily involved using a finite number of baits, and we did not know how these baits could impact the final microbial load, we also collected wild flies of unknown age that came to feed at the fendel. Culture-dependent and -independent estimations of total microbial loads from wild *D. melanogaster* ranging from 3-10 days and flies of unknown age are shown in Figure 2.5B.

Our microbial load data showed that five-day old flies are associated with fewer bacteria than wild-caught adults of indeterminate age (presumably older than our young flies and exposed to more substrates as an adult than our reared flies). The qPCR estimates of microbial load was generally higher than our culturing estimates, which is not surprising given the greater diversity of microbial species associated with wild *D. melanogaster* and the expectation that these different species will not necessarily thrive under our chosen culture conditions (28, 93). The number of yeast detected by qPCR was two to three orders of magnitude lower than that for bacteria, consistent with our observation for gnotobiotic animals (Figure 2.S4). As with the gnotobiotic, laboratory-reared flies, wild fly microbial loads were highly variable even within flies of relatively controlled age. Nonetheless, the association estimates of gnotobiotic and wild flies showed that, with the exception of yeast mono-associated flies, our laboratory-reared flies bear microbial loads similar to wild flies.

**Discussion**

Our results are broadly consistent with previous studies showing that the presence of microbes in the environment and food has a significant effect on gene expression in the adult *D. melanogaster* gut, and we add several new observations to our understanding of the effects of microbes on host gut gene expression. Raising flies in the presence of three bacterial species associated with laboratory stocks of *D. melanogaster* (two of which are the highly abundant in laboratory-reared flies), either individually or in combination, has minimal effects on adult gut gene expression compared to flies raised axenically. There are, correspondingly, few differences in gut gene expression between flies raised on monocultures of different bacteria. In contrast, the presence of the yeast *S. cerevisiae*, either alone or in combination with bacteria, has a large effect on adult gut gene expression, suggesting that yeast plays a more important role than bacteria in shaping physiology of the adult gut.

**Insensitivity of adult female gut to bacterial species**

Contrary to our expectations, we did not find significant differences in gene expression between the guts of *D. melanogaster* individuals mono-associated with three different bacterial species (*A. pasteurianus*, *L. brevis* or *L. plantarum*) commonly found in laboratory-reared *D. melanogaster*. This was surprising not only because of previous reports that these mono-associations had phenotypic consequences, but also because we expected, and confirmed, high variability in bacterial load between individuals within the same treatment group. We believe the most parsimonious explanation for our data is that the adult *D. melanogaster* gut is largely indifferent to the identity of bacteria species that occupy it under laboratory colonization levels that reflect those found in wild flies. But several important caveats warrant discussion.
First, that there is no difference in gut gene expression in response to these three strains does not mean that the adult gut is insensitive to bacterial identity. Indeed, while the species found in association with laboratory *D. melanogaster* are also found in wild flies, it is possible that the bacterial species that thrive amongst laboratory-reared flies are not representative of natural populations, and that the *Drosophila* gut would respond in different ways to different bacterial species. We also remind the reader that while the *A. pasteurianus* strains used in these experiments was collected from flies, the *Lactobacillus* strains came from non-fly sources (we used the sequenced strains of *L. plantarum* from the ATCC and *L. brevis* from the NBRC), and there may be important inter-strain differences that this choice obscured. For example, work with *L. plantarum* has shown that some strains but not others are competent to promote larval growth under nutrition-limiting conditions (41). Second, although the guts of five day old adult females show no differences between different bacterial mono-associations, it does not indicate that their guts did not respond differently at earlier points in development; we may simply have missed a critical period for these interactions. In fact, compelling work from several groups has demonstrated that certain fly-associated bacterial strains make significant contributions to larval development (41, 42, 110) and gut physiology (98, 100). Third, it is possible that the adult female *D. melanogaster* gut is capable of a differential response to these bacteria, but does so only under different conditions (e.g. different diet, different genetic background). Finally, it is possible that the guts in the flies we sampled were responding differently to these bacteria, but that these responses were not reflected at the level of gene expression. For example it could be that the common metabolic demands of the gut dominate its gene expression, and that the gut transduces bacteria-specific signals to other tissues.

Nonetheless, we think it is interesting that gene expression in the adult female *D. melanogaster* gut is so consistently insensitive to varying composition and levels of these bacteria under these conditions. There is a certain logic in insulating the gut from the effects of the constantly varying mix of diverse microbes that wild *D. melanogaster* are exposed to on their preferred substrate of rotting microbes, and our gene expression data may reflect precisely such an insulation. This is not to argue in any way that bacteria in the gut have no effect on the fly, just simply that the epithelial cells of the adult gut do not appear to be the locus for these effects.

**Effect of yeast on gut gene expression**

The results from this study reflect a dramatic effect of living *S. cerevisiae* in the environment on gene expression in the adult gut. The strong effect of yeast on *Drosophila* physiology is unsurprising, given the important role that fungi play in the *Drosophila* life cycle, especially as food for larvae. However, the fact that yeast have a large impact on gut gene expression despite the near complete absence of living yeast cells in the gut suggests an indirect effect, such as differences in how the host responds to living versus dead yeast (either at adulthood or earlier in development), or differences in nutritional value or other properties of the media due to yeast growth. One obvious possible factor is ethanol. We did not perform a chemical analysis of the media, but it is all but certain that the presence of *S. cerevisiae* led to significant ethanol production. *D. melanogaster* preferentially oviposit in substrates emitting volatile fermentation products, and they have evolved a tolerance for relatively high ethanol levels as both larvae and adults. The direct effect of ethanol of gene
expression in the *D. melanogaster* gut has not been studied, but dietary ethanol has been shown to have significant effects on gut morphology and physiology (111).

Interestingly, the gut expression data revealed that axenic or bacteria mono-associated guts overexpressed genes involved in fatty acid production and transport relative to yeast mono-associated or conventional guts. This observation is consistent with the finding that axenic animals contain more lipids than conventional animals (43), and suggests that sensation by the gut underlies the amount of fat produced by different gnotobiotic animals.

**Whole animal gene expression paradox**

Intriguingly, our experiments showed that, while yeast-containing treatments were sufficient to drive conventional-like transcription in the adjacent host tissue (i.e. guts), yeast alone or as part of a simplified microbial community (*A. pasteurianus, L. brevis, L. plantarum* and *S. cerevisiae*) was not sufficient to produce an animal-wide transcriptional program comparable to that of conventional animals. At first glance, these two findings seem to present a paradox. That is, one might think that if the difference between animals is what microbes are contained in their gut and if the gut is not responding to this change then one should also not observe changes outside of the gut tissue.

We can imagine several ways in which to reconcile these seemingly contradictory results. First, given the repeated observation that different microbial taxa have profound influences on host biology at the larval stages (41, 42, 98, 100, 110), we believe that differences in host gut gene expression would likely be observed at earlier time points. If this were the case, differences in response to the conventional samples (either due to species, fungal or bacterial, that were not introduced in our simplified community and/or strain-specific effects that could not be recapitulated by our chosen isolates) could lead to differences that accumulate over developmental time and result in animals with markedly different global transcriptional programs. In turn, these accumulated differences could manifest in several ways. The simplest case is that developmental differences accumulated to result in transcriptional differences in an otherwise physically indistinguishable animal. More likely, it could be that certain tissues are over- or underrepresented in conventional animals versus other treatments (which would appear in our data as either higher or lower contributions of those tissues to the overall transcriptional program). Alternatively, the metabolic state of the adult conventional animal could be distinct from those of other treatments (feeding back on transcription to cause big changes).

Lastly, we observed excessive microbial growth on the food in the conventional vials and the conventional animals were much more likely to die between eclosion and collection than any other treatment. It is possible that this sort of stressful environment selected for animals that were acutely adapted to deal with excessive microbial growth, which manifested at the level of global transcription.

**Our data in the context of the Drosophila microbiome literature**

While this study is, to the best of the authors’ knowledge, the first to look at the effects of individual microbial taxa on gene expression of *D. melanogaster* adults, two other studies have proceeded ours in investigating the difference in gene expression between axenic and conventional adult flies (97, 106). In one study, guts were dissected from either axenic or conventional adult animals and pooled in groups of sixty for RNA extraction and evaluation
of gene expression by microarray (106). The authors found 152 genes that were significantly over- or under-expressed (1.3-fold or greater) between axenic and conventional young adults (4-6 days old) in both OreR and CantonS strains (106). A direct comparison of our data with the core set of (106) shows agreement between the two sets for about two-thirds of the transcripts (Figure 2.S5). Given the differences in fly rearing practices (e.g. diet), microbial content and gene expression measurement methods between our study and theirs, this strikes us as a reasonable level of similarity between the two sets and perhaps identifies a smaller “core” set that is consistent across conditions.

In another study, 5-day old whole adult flies, reared axenically until adulthood and either left axenic or introduced to a mixture of fly-associated microbes upon eclosion, were pooled in groups of 20, and evaluated for differences in gene expression by microarray hybridization (97). The authors determined 105 transcripts to be significantly overexpressed in conventionalized flies compared to axenic flies. A direct comparison of our data with this set shows agreement between the two sets for 70% of the transcripts (Figure 2.S6). Again, given the differences in fly rearing practices (particularly gnotobiotic generation) and gene expression measurement, this strikes us as a reasonable level of similarity between the two sets.

Interaction of the fly gut microbiome with the host immune system is well documented (112). Because the fly senses gram positive bacteria, gram negative bacteria, and yeasts through different PGRPs which activate different downstream signaling pathways of Toll and IMD (29), we suspected we might see differential immune activation between different bacterial mono-associations. However, immunomodulatory mechanisms to tolerate commensals have also been described (113), which would dampen any effects we see. To detect these differences, we specifically looked at AMP production because different AMPs are produced downstream of Toll versus IMD activation (114). While immune related genes as a group showed significant differences between conventionally reared whole adult flies and other treatments, when we examined individual genes across our dataset by ANOVA, only AttacinD stood out as a significantly overexpressed AMP in whole, conventionally reared flies (Figure 2.S7). We did not observe statistically-significant differences in the levels of specific AMPs either among bacterial treatments or between bacterial treatments and conventionally reared flies in either guts or whole flies (Figures 2.S7, 2.S8). This finding is consistent with that reported in an earlier study looking at conventional versus germ-free flies (99) as well as with studies showing immunomodulatory mechanisms to maintain commensal populations in the gut by silencing downstream immune responses despite immune activation (113), suggesting that the fly detects these bacteria as commensals and dampens its response to them. We note, however, that this study and these analyses were designed to detect consistent differences in gene expression between treatment groups; our data do not speak to associations between the expression of AMPs or any other gene and overall bacterial load or the presence of specific bacteria in individual samples.

**Challenges in utilizing Drosophila as a model system to study gut microbe-host biology**

Laboratory-reared flies and their associated bacteria offer a convenient and powerful model for studying microbe-host interactions. However, the diversity and composition of the microbiomes of laboratory flies is limited compared to wild flies (28), likely because we
have selected for microbes that thrive on the substrates that we use to culture flies in the laboratory. A consequence of this selection is that the effects of microbes associated with D. melanogaster in the laboratory may not reflect those of microbes they associate with in the wild.

It is also unclear exactly what it means for microbes to be associated with a fly. Though we generally refer to the microbes found within the fly gut as having colonized the gut, it has yet to be firmly established whether stable colonization of the gut occurs and under what conditions. Analogous to the mucus layer in the mammalian gut, the Drosophila gut possesses a barrier, the peritrophic matrix (PM), that impedes the direct contact between objects greater than ~250 KDa (including food particles and microbial cells) and host cells (115, 116). Similar to the mammalian mucosal barrier, recent evidence suggests that the PM is a dynamic structure capable of responding to pathogenic microbes (117). However, while it has been established that some microbes in the vertebrate gut dine on mucus protecting the gut epithelium (118), it has not yet been firmly established whether Drosophila-associated microbes actually adhere to or subsist on the molecules comprising or associated with the PM. Additionally, flies appear to lack the crypt structures, present in vertebrate guts, that can provide hideout for microbes allowing them to persist during hard times (116). At present, we know that laboratory-reared flies can lose their microbiota by continually transferring onto sterile food (110, 119). The turnover of fly microbiota under these conditions suggests that stable colonization of the gut tissue is not occurring under these circumstances, although it is possible that there are different conditions under which stable colonization could be observed.

A recent study of wild flies illustrates the importance of microbes to fly biology and highlights the distinct manner in which flies associate with microbial partners. Yamada et al. (2015) discovered that a yeast species, Issatchenki orientalis, increases amino acid availability and extends lifespan in flies reared on a protein-deficient diet (119). They found that lifespan could be extended not just by providing flies with live I. orientalis, but also by providing heat-killed I. orientalis. Of note, the authors observed that daily transfer of flies onto sterile medium resulted in the loss of association with I. orientalis. Thus, there is evidence for flies benefiting from a non-stable association with a microbial species, and the benefit is conferred even when the microbes are dead. This might not be how we typically conceptualize important microbial interactions in animals, but it’s possible that these are the most critical microbial interactions for flies.

Materials and Methods

Fly and microbial stocks

Wolbachia-free CantonS fly stocks were reared on medium from UC Berkeley’s Koshland fly kitchen (Koshland diet; 0.68% agar, 6.68% cornmeal, 2.7% yeast, 1.6% sucrose, 0.75% sodium tartrate tetrahydrate, 5.6 mM CaCl2, 8.2% molasses, 0.09% tegosept, 0.77% ethanol, 0.46% propionic acid) supplemented with activated dry yeast pellets at 25C on a 12:12 light:dark photoperiod. Acetobacter pasteurianus (CNE7) was isolated from laboratory-reared WT Drosophila melanogaster and grown on de Man, Rogosa and Sharpe (MRS) agar (Research Products International Corp) at 30C. Lactobacillus brevis (NBRC 107147) and Lactobacillus plantarum (ATCC 8014) were acquired from NBRC and ATCC culture collections, respectively, and grown on MRS agar at 30C. Saccharomyces cerevisiae
(ASQ HI) was isolated from wild Hawaiian Drosophila by Alli Quan and grown on YPD agar at 30°C (A. Quan, personal communication).

**Shotgun DNA sequencing of WT fly guts**

Female CantonS flies age five days post-eclosion (PE) were collected via cold anesthesia, surface-sterilized with 10% bleach for ten minutes and rinsed with sterile 1x PBS before dissecting out guts. Forceps used for dissection were treated with 3.5% H₂O₂ between animals to remove DNA contaminants. Dissections included the proventriculus to the rectal ampulla, leaving the Malpighian tubules attached. Each individual dissected gut was flash frozen in sterile-filtered Buffer ATL (QIAamp Micro Kit, QIAGEN) and stored at -80°C. DNA was extracted according to QIAGEN’s QIAamp Micro Kit tissue protocol, with modifications. After the overnight digestion with proteinase K, 0.1 mm Zirconium bead and 1 volume buffer WJL (2M Guanidinium thiocyanate, 0.5 M EDTA, 1.8% Tris base, 8% NaCl, pH 8.5) were added to each sample. Samples were then bead beat twice for one minute at 4°C with a 30 second break in between, spun 5 minutes at ~14,000xg and the supernatant was transferred to a new tube. Beads were resuspended in two volumes buffer WJL and beat again an additional minute before spinning down and pooling supernatant. Beads were washed once more with two volumes buffer WJL before spinning down and pooling supernatant a final time. An additional spin was performed to pellet any carried-over beads; supernatant was transferred to a new tube. Each sample then received 1 ug of carrier RNA dissolved in buffer AE before proceeding with ethanol precipitation and elution per the manufacturer’s protocol. DNA samples were quantified (Qubit dsDNA HS assay kit, ThermoFisher Scientific) then treated with 120 ng/ul RNase A (QIAGEN) for 1 hour at 37°C. Each RNase-treated sample was used to generate an indexed next-generation sequencing library using the TruSeq Nano DNA kit (Illumina). Samples were pooled and sequenced at 100 paired-end reads at the UC Davis Genome Center on a HiSeq 2500. Reads were first filtered by aligning to the D. melanogaster genome (version 6.01) using bowtie2, then aligned to the Green Genes 16S database (release 13-5). Species that were identified by two reads or fewer were removed before rarefaction in R (http://www.jennajacobs.org/R/rarefaction.html, Jenna Jacobs). Samples that did not rarely to saturation were discarded.

**Preparation of axenic, gnotobiotic and conventional flies for RNAseq**

Freshly-laid embryos from CantonS reared on Koshland diet were used to generate axenic, gnotobiotic and conventional animals. Two days before starting embryo collection, ~500 CantonS WF adults were put into large embryo collection cages with a grape juice agar plate. Grape juice agar plates The collection cages were kept in incubators set at 25°C, 60% humidity, and 12hr light:12 hr dark cycle. On the morning of embryo collection, a fresh grape juice agar plate with a small dollop of yeast paste was placed in the embryo collection cage. Flies were allowed to lay eggs for 4-6 hours. After 4-6 hours, embryos were transferred into a sterile 100 micron cell strainer (BD Falcon #352360 and rinsed with Milli-Q water (EMD Millipore). Conventional animals were generated by transferring ~100 water-rinsed embryos to sterile YG diet (10% active dry yeast, 10% glucose, 1.2% agar, 0.42% propionic acid (120) sterilized by autoclaving).

To generate axenic and gnotobiotic flies, water-rinsed embryos were subjected to 5 minutes in 10% household bleach, replacing bleach solution once halfway through, then
briefly rinsed in 70% ethanol before being rinsed with copious sterile water. For axenic animals, ~100 cleaned embryos were transferred to sterile YG diet. For gnotobiotic animals, ~100 embryos were transferred to YG vials inoculated with 50 μL of 1 x 10^8 cell/mL of microbial suspension. Microbial suspensions consisted of one species (mono-associated) or an equal mixture of *A. pasteurianus* (CNE7), *L. brevis* (NBRC 107147) and *L. plantarum* with or without *S. cerevisiae* (HI ASQ) (poly-associated), as appropriate. To prepare cell suspensions, overnight liquid monocultures of microbes were pelleted, washed and resuspended to 1 x 10^8 cell/mL in 1x PBS following OD600 conversions (given in [43]) for *A. pasteurianus, L. brevis* and *L. plantarum* or 3 x 10^8 cell/mL per OD600 of 1.0 for *S. cerevisiae* (121).

All embryos were reared from embryo to five days post-eclosion in in 28.5 mm x 95 mm vials (FlyStuff cat # 32-121) containing 10-15 mL YG medium (spiked with microbial preps, as appropriate) without added dry active yeast pellets. Vials were capped with sterile Drosophila plugs (FlyStuff cat # 59-201) and housed at 25°C on a 12:12 photoperiod with 60% humidity. Flies were maintained in their original vials before sampling (i.e. adults were never transferred onto new medium).

Bacterial mono-associations were prepared on three different occasions for gut sampling (7/18/14, 7/29/14, 8/7/14) and one occasion for whole fly sampling (11/25/14). Yeast mono-associations were prepared on two different occasions for gut sampling (11/25/14, 2/26/15) and one occasion for whole fly sampling (11/25/14). All other treatments were prepared on one occasion for gut or whole fly sampling (11/25/14, 2/26/15).

**Fly collection and validation of gnotobiotic treatment**

Gnotobiotic and conventional CantonS females reared on YG diet were collected at age five days PE (starting at 1-3 pm for all preps except first bacterial mono-association gut prep) via cold anesthesia using aseptic technique. For gut samples, individual guts were dissected from each of three female flies (from proventriculus through rectal ampulla including Malpighian tubules) in sterile 1x PBS using Dumont 55 forceps. Cleanly dissected guts were immediately deposited in Trizol (ThermoFisher Scientific) and flash frozen in liquid nitrogen. Between samples, forceps were treated with 70% ethanol, flamed then dipped in 3.5% hydrogen peroxide and sterile water to prevent nucleic acid carryover between samples. For whole fly samples, individual cold-anesthetized animals were transferred to Trizol and frozen as above. *A. pasteurianus, L. brevis* and *L. plantarum* mono-associated animals for gut samples were collected for each of three independent gnotobiotic preparations. All other samples were collected from a single experiment, with preparations for different treatments occurring over the course of several months. Each time animals were prepared and sampled for sequencing three individuals (females when possible, males when not) were independently homogenized and plated on suitable agar medium to verify they possessed (or in the case of axenic animals, lacked) the expected microbial species. For mono-associated animals species verification was achieved by first confirming that plates showed only one morphology, performing 16S or ITS colony PCR on five representative colonies for each treatment and Sanger sequencing of the resultant products. For conventional and poly-associated animals treatment was first verified by checking morphology of resultant colonies performing 16S or ITS colony PCR on five representative colonies for each treatment and Sanger sequencing of the resultant products.
RNA preparation and sequencing

RNA was prepared from each thawed sample by homogenizing with an RNase-free pestle (Kimble Chase), washing the pestle with 750 uL Trizol, then proceeding using the manufacturer’s protocol with 10 μg glycogen carrier per sample. RNA quality was checked by running on a RNA 6000 Pico chip on a Bioanalyzer 2100 (Agilent Technologies) and quantified using a Qubit Fluorometer (Qubit RNA HS assay kit, ThermoFisher Scientific). High quality RNA was then treated with Turbo DNase (ThermoScientific) per the manufacturer’s protocol. For gut samples, RNaseq libraries were prepared using the TruSeq RNA v2 kit (Illumina) starting with 100-200 ng of DNase-treated total RNA for each sample. For whole fly samples, RNaseq libraries were prepared using 400-500 ng of RNA per sample. Samples were multiplexed and sequenced using 100-150 bp paired-end reads on a HiSeq 2500 at the QB3 Vincent J. Coates Genomic Sequencing Facility at UC Berkeley. Reads were aligned to the D. melanogaster genome (version 6.01) using Tophat using the “-no-mixed” option (S1 Table). Read alignments (BAM files) were submitted to the NCBI Sequence Read Archive under accession SRP076493. Transcript abundances were calculated from aligned reads using Cufflinks based on a reference transcriptome lacking RNA genes, as we find these to be wildly variable in any RNAseq experiment (M. B. Eisen, personal communication). Data were analyzed using hierarchical clustering by gene (Cluster 3.0), ANOVA between grouped treatments (scipy.stats) and GO term analysis (Panther(108)). Data were plotted using matplotlib (Python) and Prism 6 (GraphPad). Heatmap color scales were defined with custom_cmap.py (http://schubert.atmos.colostate.edu/~cslocum/custom_cmap.html, Chris Slocum). Samples that were in major disagreement with replicates (as determined by aberrant hierarchical clustering when compared to replicates of the same treatment and the sample set at large) were excluded from analysis (noted in S1 Table).

Determining gut microbe colonization of gnotobiotic flies

Gnotobiotic (mono-associated), axenic and conventional flies were prepared and reared on YG diet five separate times as described above. CantonS females were collected at age five days PE via cold anesthesia using aseptic technique. Animals were surface sterilized using a one minute incubation in 95-100% ethanol then rinsed in sterile water. For conventional and gnotobiotic treatments, a ten flies or as many as available were transferred to individual tubes containing 1x PBS, homogenized with sterile pestles and plated at the dilutions and on the media as shown in Table 2.1.

Table 2.1. Plating media and dilutions for determining gut microbial load of laboratory-reared flies.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Plate A</th>
<th>Plate B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional</td>
<td>1/200, MRS</td>
<td>1/200, YPD</td>
</tr>
<tr>
<td><em>A. pasteurianus</em></td>
<td>1/10, MRS</td>
<td>1/20, MRS</td>
</tr>
<tr>
<td><em>L. brevis</em></td>
<td>1/50, MRS</td>
<td>1/100, MRS</td>
</tr>
</tbody>
</table>
For axenic preparations, three females were homogenized in each of two tubes of 1x PBS without prior surface sterilization. One homogenate was plated on MRS and one on YPD. A negative control was performed for each set of flies sampled by homogenizing an equal volume of 1x PBS and plating the half on each MRS and YPD. All plates were incubated at 30°C for three days or until plates showed visible colonies. Colonies were counted by hand with the aid of a light board and a handheld counter. The count from each plate was corrected for dilution and averaged over two plates for each fly.

**Determining gut microbe colonization of wild flies**

To collect wild flies of unknown age, flies were directly caught either by baiting a closed bottle trap with banana or by directly aspirating from an uncovered plastic dishwashing pan (heretofore referred to as a “fendel”) that was baited with organic watermelon and an assortment of other organic fruits. All baiting and capture was performed in the spring of 2015 at a personal residence in Berkeley, CA. Flies were captured in the morning and transferred into sterile, empty vials (to avoid providing a diet that might manipulate their gut microbiota) before sampling shortly thereafter. Flies were recovered from vials via cold anesthesia using aseptic technique. Males were discarded.

To collect wild flies of roughly five days PE, a piece of tap-water rinsed organic fruit (purple grape, banana or Fuji apple) was placed in the open-bait fendel and left for 24 hours for provide an oviposition substrate for wild females. After 24 hours, the fruit pieces were placed in sterile vials and the deposited embryos were allowed to develop at room temperature. Vials were monitored daily for newly-eclosed flies. All flies from a vial were collected when the majority of the adults present in the vial were 5 days PE via cold anesthesia using aseptic technique. Males were discarded.

All wild flies were then surface sterilized using a one minute incubation in 95-100% ethanol and rinsed in sterile water before homogenization with a sterile pestle in 1x PBS. Half of the homogenate was diluted to plate 1/200 across MRS agar and YPD agar for each fly and plates were diluted at 30°C for three days or until colonies emerged. The remaining half was then processed for DNA extraction as described for shotgun DNA sequencing or frozen at -30°C for later processing. A negative control was performed for each set of flies collected by homogenizing an equal volume of 1x PBS with a sterile pestle and splitting in half as for fly samples. Colonies were counted by hand with the aid of a light board and a handheld counter. The count from each plate was corrected for dilution and averaged over two plates for each fly.

All sampled wild females were genotyped by PCR and Sanger sequencing cytochrome c oxidase II primers tLYS (GTGTTAAGAGACCAGTACTTG) and tLEU (ATGGCAGATTAGTGCAATGG) (122). Colonization data from flies that were not identified as *D. melanogaster* (e.g. *D. simulans, D. immigrans, D. persimilis*) were discarded.

To provide a culture-independent method for estimating gut colonization in wild animals, all DNA samples from confirmed wild *D. melanogaster* females were subjected to two sets of triplicate qPCR reactions, one using universal bacteria 16S primer pair P1 (CCTACGGGAGCGACGAG) and P2 (ATTACCGCGCTGCTG) at 100 nM per reaction and the

<table>
<thead>
<tr>
<th>L. plantarum</th>
<th>1/50, MRS</th>
<th>1/100, MRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. cerevisiae</td>
<td>1/4, YPD</td>
<td>1/4, YPD</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L. plantarum</th>
<th>1/50, MRS</th>
<th>1/100, MRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. cerevisiae</td>
<td>1/4, YPD</td>
<td>1/4, YPD</td>
</tr>
</tbody>
</table>
other using fungal ITS primer pair Y1 (GCGGTAAATTCCAGCTCCATAG) and Y2 (GCCACAAGGACTCAAGGTTAG) at 800 nM per reaction (123). Reactions were performed on a Roche480 LightCycler using Sybr Green mastermix (Roche) templated with 10-20 ng of total fly DNA for each reaction. Four-point standard curves were run for each experiment for A. pasteurianus and L. plantarum (P1, P2 primer pair) and S. cerevisiae (Y1, Y2 primer pair). The amplification program used for qPCR is as follows: 95C for 5 min then 45 cycles of 95 C for 10 seconds, 65C for 10 seconds (decreasing 1 C for the first 10 cycles until reaching 55C), and 72C for 10 seconds. Primers were validated using DNA extracted from monocultures of A. pasteurianus, L. brevis, L. plantarum and S. cerevisiae. The number of bacterial and yeast cells present in the whole animal was estimated assuming an average genome size of 3 Mb for bacteria and 25 Mb for yeast.
Supporting Information

**Figure 2.S1.** Gut bacterial community analysis of *Wolbachia*-free CantonS five-day mated females reared on Koshland diet. A) Relative species abundance of gut bacteria as determined by shotgun sequencing reads to the Green Genes 16S rRNA database release 13-5. “Other” category includes species not listed in key. “Unknown” category includes reads that aligned to 16S rRNA sequences included in the Green Genes database annotated as “unknown” (e.g. unknown compost). B) Rarefaction curve using data shown in A.

**Figure 2.S2.** Genes expressed more similarly within yeast-containing and conventional whole flies versus all other samples. A) heatmap of 72 genes that are overexpressed in conventional, yeast- mono- and poly-associated whole flies compared to other whole fly samples (Bonferroni p-value>0.05, ANOVA). B) Heatmap of 67 genes that are overexpressed in axenic and bacteria mono- and poly-associated whole flies compare to other whole fly samples (Bonferroni p-value>0.05, ANOVA). Abbreviations: Ap = *A. pasteurianus*-mono-associated; Lbrev = *L. brevis*-mono-associated, Lp = *L. plantarum*-mono-associated, 3bac = poly-associated without yeast, Ax = axenic, Conv = conventional, Yeast = *S. cerevisiae*-mono-associated, 4mic = poly-associated with yeast.
Figure 2.S3. Genes showing greatest difference in expression values between different bacteria mono-associations in whole adults as determined by one-way ANOVA. A) Scatterplot of log10-transformed FPKM values for each bacteria mono-associated whole fly replicate. Genes are ordered from lowest ANOVA p-value (top) to highest (bottom). P-values have undergone a Bonferroni correction for multiple testing. B) Data from A presented as a heatmap. FPKM values for each gene are normalized to range from -1 to 1 before plotting. Black line above heatmap denotes bacteria mono-association samples. Abbreviations: Ap = A. pasteurianus-mono-associated; Lbrev = L. brevis-mono-associated, Lp = L. plantarum-mono-associated, 3bac = poly-associated without yeast, Ax = axenic, Conv = conventional, Yeast = S. cerevisiae-mono-associated, 4mic = poly-associated with yeast.
Figure 2.54. Loads of bacteria and yeast in wild flies estimated from qPCR. Log10 transformed average number of estimated bacteria cells or yeast cells by qPCR for individual, female, wild *D. melanogaster* raised from embryos (ranging from 3-10 days post-eclosion) or caught as adults (of unknown age). The mean for each group is plotted as a horizontal line.
Figure 2.55. Comparison of gut RNAseq data of present manuscript with expression of “core” gene set identified in Broderick et al, 2014 (106). Left) Log10 of FPKM values for each of the genes in (106) core set in present manuscript’s gut RNAseq data. Individual replicates are shown as dots. Standard deviation is shown as error bars with mean bolded. Right) Ratio of expression in conventional over axenic treatments for each of the genes in the core set in (106) Affymetrix data and in present manuscript’s gut RNAseq data. For both plots, genes are ordered by the ratio of gene expression (conventional over axenic) as reported in (106). For right plot, genes with a ratio lower than -25 have been plotted as exactly -25; genes with ratio higher than 25 have been plotted as exactly 25. NB ratio = data from Broderick et al, 2014; CE ratio = data from this study.
Figure 2.S6. Comparison of whole fly RNAseq data of present manuscript with findings in Erkosar Combe et al, 2014 (97). Left) Log10 of FPKM values for each of the genes in (97) in present manuscript’s whole fly RNAseq data. Individual replicates are shown as dots. Standard deviation is shown as error bars with mean bolded. Right) Ratio of expression in conventional over axenic treatments for each of the genes in the core set in (97) Affymetrix data and in present manuscript’s whole fly RNAseq data. For both plots, genes are ordered by the ratio of gene expression (conventional over axenic) as reported in (97). For right plot, genes with ratio higher than 25 have been plotted as exactly 25. EB ratio = data from Erkosar Combe et al, 2014; CE ratio = data from this study.
Figure 2.57. Expression of antimicrobial peptide genes across all whole fly samples. Data are plotted as boxes and whiskers using boxplot function in Matplotlib (Python). Each box is drawn from the lower to upper quartile values for all replicates for the indicated sample. The horizontal line in the box indicates the median. Whiskers extend 1.5 times the length of the box. Points beyond whiskers (outliers) are plotted as crosses. AMP genes for which FPKM values are zero across all treatments (i.e. are universally unexpressed) are omitted.
Figure 2.S8. Expression of antimicrobial peptide (AMP) genes across all dissected gut samples. Data are plotted as boxes and whiskers using boxplot function in Matplotlib (Python). Each box is drawn from the lower to upper quartile values for all replicates for the indicated sample. The horizontal line in the box indicates the median. Whiskers extend 1.5 times the length of the box. Points beyond whiskers (outliers) are plotted as crosses. AMP genes for which FPKM values are zero across all treatments (i.e. are universally unexpressed) are omitted.
<table>
<thead>
<tr>
<th>Lane</th>
<th>Read type^</th>
<th>Sample type</th>
<th>Sample ID</th>
<th>Total input read pairs*</th>
<th>Concordantly-aligned pairs*</th>
<th>Concordant alignment rate*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100 PE</td>
<td>Gut</td>
<td>Ap1a</td>
<td>4.53E+07</td>
<td>2.84E+07</td>
<td>62.6</td>
</tr>
<tr>
<td>1</td>
<td>100 PE</td>
<td>Gut</td>
<td>Ap1b</td>
<td>2.74E+07</td>
<td>1.70E+07</td>
<td>61.8</td>
</tr>
<tr>
<td>1</td>
<td>100 PE</td>
<td>Gut</td>
<td>Ap1c</td>
<td>4.96E+07</td>
<td>3.10E+07</td>
<td>62.5</td>
</tr>
<tr>
<td>1</td>
<td>100 PE</td>
<td>Gut</td>
<td>Lbrev1a</td>
<td>1.39E+07</td>
<td>8.61E+06</td>
<td>61.9</td>
</tr>
<tr>
<td>1</td>
<td>100 PE</td>
<td>Gut</td>
<td>Lbrev1b</td>
<td>2.08E+07</td>
<td>1.30E+07</td>
<td>62.2</td>
</tr>
<tr>
<td>1</td>
<td>100 PE</td>
<td>Gut</td>
<td>Lbrev1c</td>
<td>2.06E+07</td>
<td>1.22E+07</td>
<td>59.1</td>
</tr>
<tr>
<td>1</td>
<td>100 PE</td>
<td>Gut</td>
<td>Lp1a</td>
<td>1.94E+07</td>
<td>1.17E+07</td>
<td>60.6</td>
</tr>
<tr>
<td>1</td>
<td>100 PE</td>
<td>Gut</td>
<td>Lp1b</td>
<td>2.46E+07</td>
<td>1.51E+07</td>
<td>61.3</td>
</tr>
<tr>
<td>1</td>
<td>100 PE</td>
<td>Gut</td>
<td>Lp1cx</td>
<td>5.41E+07</td>
<td>3.53E+07</td>
<td>65.2</td>
</tr>
<tr>
<td>2</td>
<td>100 PE</td>
<td>Gut</td>
<td>Ap2a</td>
<td>1.14E+07</td>
<td>8.56E+06</td>
<td>74.9</td>
</tr>
<tr>
<td>2</td>
<td>100 PE</td>
<td>Gut</td>
<td>Ap2b</td>
<td>2.12E+07</td>
<td>1.63E+07</td>
<td>76.9</td>
</tr>
<tr>
<td>2</td>
<td>100 PE</td>
<td>Gut</td>
<td>Ap2c</td>
<td>2.05E+07</td>
<td>1.57E+07</td>
<td>76.5</td>
</tr>
<tr>
<td>2</td>
<td>100 PE</td>
<td>Gut</td>
<td>Lbrev2a</td>
<td>1.71E+07</td>
<td>1.28E+07</td>
<td>75.2</td>
</tr>
<tr>
<td>2</td>
<td>100 PE</td>
<td>Gut</td>
<td>Lbrev2b</td>
<td>1.84E+07</td>
<td>1.39E+07</td>
<td>75.7</td>
</tr>
<tr>
<td>2</td>
<td>100 PE</td>
<td>Gut</td>
<td>Lbrev2c</td>
<td>1.34E+07</td>
<td>1.05E+07</td>
<td>78.0</td>
</tr>
<tr>
<td>2</td>
<td>100 PE</td>
<td>Gut</td>
<td>Lp2a</td>
<td>2.37E+07</td>
<td>1.80E+07</td>
<td>76.0</td>
</tr>
<tr>
<td>2</td>
<td>100 PE</td>
<td>Gut</td>
<td>Lp2b</td>
<td>4.64E+07</td>
<td>3.55E+07</td>
<td>76.5</td>
</tr>
<tr>
<td>2</td>
<td>100 PE</td>
<td>Gut</td>
<td>Lp2c</td>
<td>3.92E+07</td>
<td>3.01E+07</td>
<td>76.8</td>
</tr>
<tr>
<td>3</td>
<td>100 PE</td>
<td>Gut</td>
<td>Ap3a</td>
<td>3.88E+07</td>
<td>3.06E+07</td>
<td>78.9</td>
</tr>
<tr>
<td>3</td>
<td>100 PE</td>
<td>Gut</td>
<td>Ap3b</td>
<td>1.57E+07</td>
<td>1.22E+07</td>
<td>77.5</td>
</tr>
<tr>
<td>3</td>
<td>100 PE</td>
<td>Gut</td>
<td>Ap3c</td>
<td>2.19E+07</td>
<td>1.70E+07</td>
<td>77.4</td>
</tr>
<tr>
<td>3</td>
<td>100 PE</td>
<td>Gut</td>
<td>Lbrev3a</td>
<td>1.54E+07</td>
<td>1.20E+07</td>
<td>77.9</td>
</tr>
<tr>
<td>3</td>
<td>100 PE</td>
<td>Gut</td>
<td>Lbrev3b</td>
<td>2.07E+07</td>
<td>1.60E+07</td>
<td>77.4</td>
</tr>
<tr>
<td>3</td>
<td>100 PE</td>
<td>Gut</td>
<td>Lbrev3c</td>
<td>2.21E+07</td>
<td>1.71E+07</td>
<td>77.4</td>
</tr>
<tr>
<td>3</td>
<td>100 PE</td>
<td>Gut</td>
<td>Lp2c</td>
<td>1.70E+07</td>
<td>1.34E+07</td>
<td>78.4</td>
</tr>
<tr>
<td>3</td>
<td>100 PE</td>
<td>Gut</td>
<td>Lp3a</td>
<td>2.61E+07</td>
<td>2.00E+07</td>
<td>76.8</td>
</tr>
<tr>
<td>3</td>
<td>100 PE</td>
<td>Gut</td>
<td>Lp3b</td>
<td>2.05E+07</td>
<td>1.57E+07</td>
<td>76.5</td>
</tr>
<tr>
<td>3</td>
<td>100 PE</td>
<td>Gut</td>
<td>Lp3c</td>
<td>2.19E+07</td>
<td>1.68E+07</td>
<td>76.7</td>
</tr>
<tr>
<td>4</td>
<td>100 PE</td>
<td>Gut</td>
<td>Ax1</td>
<td>1.58E+07</td>
<td>1.19E+07</td>
<td>75.7</td>
</tr>
<tr>
<td>4</td>
<td>100 PE</td>
<td>Gut</td>
<td>Ax2</td>
<td>3.34E+07</td>
<td>2.56E+07</td>
<td>76.6</td>
</tr>
<tr>
<td>4</td>
<td>100 PE</td>
<td>Gut</td>
<td>Ax3</td>
<td>2.96E+07</td>
<td>2.28E+07</td>
<td>77.2</td>
</tr>
<tr>
<td>4</td>
<td>100 PE</td>
<td>Gut</td>
<td>Conv1</td>
<td>3.86E+07</td>
<td>2.98E+07</td>
<td>77.3</td>
</tr>
<tr>
<td>4</td>
<td>100 PE</td>
<td>Gut</td>
<td>Conv2</td>
<td>2.50E+07</td>
<td>1.92E+07</td>
<td>77.0</td>
</tr>
<tr>
<td>4</td>
<td>100 PE</td>
<td>Gut</td>
<td>Conv3</td>
<td>3.33E+07</td>
<td>2.54E+07</td>
<td>76.4</td>
</tr>
<tr>
<td>4</td>
<td>100 PE</td>
<td>Gut</td>
<td>Yeast1x</td>
<td>3.26E+07</td>
<td>2.67E+07</td>
<td>81.8</td>
</tr>
<tr>
<td>4</td>
<td>100 PE</td>
<td>Gut</td>
<td>Yeast2</td>
<td>1.65E+07</td>
<td>1.25E+07</td>
<td>75.7</td>
</tr>
<tr>
<td>4</td>
<td>100 PE</td>
<td>Gut</td>
<td>Yeast3</td>
<td>2.10E+07</td>
<td>1.59E+07</td>
<td>75.5</td>
</tr>
<tr>
<td>5</td>
<td>100 PE</td>
<td>Whole</td>
<td>ApW1</td>
<td>6.75E+06</td>
<td>5.73E+06</td>
<td>84.9</td>
</tr>
<tr>
<td>5</td>
<td>100 PE</td>
<td>Whole</td>
<td>ApW2</td>
<td>7.51E+06</td>
<td>6.30E+06</td>
<td>83.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>-----</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>100 PE</td>
<td>Whole</td>
<td>ApW3</td>
<td>3.77E+06</td>
<td>3.23E+06</td>
<td>85.6</td>
</tr>
<tr>
<td>5</td>
<td>100 PE</td>
<td>Whole</td>
<td>AxenicW1</td>
<td>1.05E+07</td>
<td>8.91E+06</td>
<td>85.0</td>
</tr>
<tr>
<td>5</td>
<td>100 PE</td>
<td>Whole</td>
<td>AxenicW2</td>
<td>7.64E+06</td>
<td>6.45E+06</td>
<td>84.4</td>
</tr>
<tr>
<td>5</td>
<td>100 PE</td>
<td>Whole</td>
<td>AxenicW3</td>
<td>7.97E+06</td>
<td>6.71E+06</td>
<td>84.3</td>
</tr>
<tr>
<td>5</td>
<td>100 PE</td>
<td>Whole</td>
<td>ConvW1</td>
<td>2.17E+06</td>
<td>1.84E+06</td>
<td>84.6</td>
</tr>
<tr>
<td>5</td>
<td>100 PE</td>
<td>Whole</td>
<td>ConvW2x</td>
<td>1.04E+07</td>
<td>8.81E+06</td>
<td>85.1</td>
</tr>
<tr>
<td>5</td>
<td>100 PE</td>
<td>Whole</td>
<td>ConvW3</td>
<td>7.14E+06</td>
<td>6.06E+06</td>
<td>84.9</td>
</tr>
<tr>
<td>5</td>
<td>100 PE</td>
<td>Whole</td>
<td>LbrevW1</td>
<td>7.77E+06</td>
<td>6.69E+06</td>
<td>86.1</td>
</tr>
<tr>
<td>5</td>
<td>100 PE</td>
<td>Whole</td>
<td>LbrevW2</td>
<td>8.81E+06</td>
<td>7.60E+06</td>
<td>86.3</td>
</tr>
<tr>
<td>5</td>
<td>100 PE</td>
<td>Whole</td>
<td>LbrevW3</td>
<td>8.03E+06</td>
<td>6.88E+06</td>
<td>85.7</td>
</tr>
<tr>
<td>5</td>
<td>100 PE</td>
<td>Whole</td>
<td>LpW1</td>
<td>7.72E+06</td>
<td>6.55E+06</td>
<td>84.8</td>
</tr>
<tr>
<td>5</td>
<td>100 PE</td>
<td>Whole</td>
<td>LpW2</td>
<td>1.09E+07</td>
<td>9.40E+06</td>
<td>86.1</td>
</tr>
<tr>
<td>5</td>
<td>100 PE</td>
<td>Whole</td>
<td>LpW3</td>
<td>1.22E+07</td>
<td>1.05E+07</td>
<td>86.4</td>
</tr>
<tr>
<td>5</td>
<td>100 PE</td>
<td>Whole</td>
<td>YeastW1</td>
<td>1.07E+07</td>
<td>9.21E+06</td>
<td>86.0</td>
</tr>
<tr>
<td>5</td>
<td>100 PE</td>
<td>Whole</td>
<td>YeastW2</td>
<td>1.37E+07</td>
<td>1.18E+07</td>
<td>85.6</td>
</tr>
<tr>
<td>6</td>
<td>150 PE</td>
<td>Gut</td>
<td>3bac1</td>
<td>4.38E+06</td>
<td>2.39E+06</td>
<td>54.6</td>
</tr>
<tr>
<td>6</td>
<td>150 PE</td>
<td>Gut</td>
<td>3bac3</td>
<td>5.02E+06</td>
<td>2.83E+06</td>
<td>56.4</td>
</tr>
<tr>
<td>6</td>
<td>150 PE</td>
<td>Gut</td>
<td>3bacW1</td>
<td>9.11E+06</td>
<td>5.03E+06</td>
<td>55.2</td>
</tr>
<tr>
<td>6</td>
<td>150 PE</td>
<td>Gut</td>
<td>3bacW2</td>
<td>6.13E+06</td>
<td>3.41E+06</td>
<td>55.6</td>
</tr>
<tr>
<td>6</td>
<td>150 PE</td>
<td>Gut</td>
<td>3bacW3</td>
<td>2.04E+06</td>
<td>1.15E+06</td>
<td>56.3</td>
</tr>
<tr>
<td>6</td>
<td>150 PE</td>
<td>Gut</td>
<td>4mic1</td>
<td>3.53E+06</td>
<td>1.97E+06</td>
<td>55.8</td>
</tr>
<tr>
<td>6</td>
<td>150 PE</td>
<td>Gut</td>
<td>4mic2</td>
<td>4.70E+06</td>
<td>2.71E+06</td>
<td>57.7</td>
</tr>
<tr>
<td>6</td>
<td>150 PE</td>
<td>Gut</td>
<td>4mic3x</td>
<td>3.78E+06</td>
<td>2.20E+06</td>
<td>58.1</td>
</tr>
<tr>
<td>6</td>
<td>150 PE</td>
<td>Gut</td>
<td>4micW1</td>
<td>5.86E+06</td>
<td>3.21E+06</td>
<td>54.7</td>
</tr>
<tr>
<td>6</td>
<td>150 PE</td>
<td>Gut</td>
<td>4micW2</td>
<td>6.97E+06</td>
<td>3.51E+06</td>
<td>50.4</td>
</tr>
<tr>
<td>6</td>
<td>150 PE</td>
<td>Gut</td>
<td>4micW3</td>
<td>1.09E+07</td>
<td>5.90E+06</td>
<td>54.2</td>
</tr>
<tr>
<td>6</td>
<td>150 PE</td>
<td>Gut</td>
<td>Yeast4</td>
<td>2.80E+06</td>
<td>1.60E+06</td>
<td>57.0</td>
</tr>
<tr>
<td>6</td>
<td>150 PE</td>
<td>Gut</td>
<td>Yeast5</td>
<td>3.76E+06</td>
<td>2.08E+06</td>
<td>55.4</td>
</tr>
<tr>
<td>6</td>
<td>150 PE</td>
<td>Gut</td>
<td>Yeast6</td>
<td>3.68E+06</td>
<td>1.97E+06</td>
<td>53.5</td>
</tr>
</tbody>
</table>

* Value obtained from the “alignment_summary.txt” output following Tophat alignment of reads to v6.01 *D. melanogaster* genome.

^ All libraries were sequenced on the Illumina HiSeq2500 platform

x Aberrant behavior compared to rest of sample set, omitted from analysis.
Chapter 3: Or22 allelic variation alone cannot account for differences in discrimination of yeast-produced volatiles by *D. melanogaster*

*This chapter and work detailed therein was a joint effort by Carolyn Elya and Allison Quan.*

Abstract

Different lines of the fruit fly *Drosophila melanogaster* show variation in the ability to discriminate between volatiles produced by the yeast *Saccharomyces cerevisiae* under natural (nitrogen-limiting, YVN) or laboratory (sugar-limiting, YVL) conditions. Previous work in our laboratory uncovered a strong correlation between heightened sensitivity to YVN and a chimeric variant of the highly variable odorant receptor 22 (Or22) locus of *D. melanogaster*. We sought to determine if this trend held for an extended set of *D. melanogaster* lines, if observed variation within chimeric and non-chimeric lines could be explained by nucleotide polymorphisms and if replacing Or22 with a chimeric allele in a non-chimeric background could confer an enhanced ability to detect YVN. In parallel, we performed crosses of chimeric and non-chimeric fly lines and assayed the behavior of their progeny for enhanced sensitivity to YVN to assess the heritability of the Or22 locus. Ultimately, we found that, while the overall trend of chimeric lines being more sensitive to YVN holds, there are exceptions and variation that cannot be explained by sequence variation at the Or22 locus. In addition, we did not observe increased sensitivity for YVN upon replacing the Or22 allele in a non-chimeric line (OreR) with that from our most yeast-sensitive line (ME). Though our results cannot support our hypothesis that Or22 is the primary driver of sensitivity to YVN, Or22 remains an interesting locus in the context of fly-yeast ecology.

Introduction

The fruit fly *Drosophila melanogaster* and the yeast *Saccharomyces cerevisiae* are close natural partners: flies require yeast for development and nutrition (46) and yeast depend on flies to be vectored to new substrates (52). Flies can sense a variety of compounds that are produced by fermenting yeast via olfaction (59) and have demonstrated a preference to yeasted over non-yeasted fruit in the context of the laboratory (47). Evidence to date suggests that chemical communication is the basis of the co-occurrence of flies and yeast in nature, but the specific components that mediate this molecular conversation are incompletely understood.

Olfactory sensing in *Drosophila* begins in the antenna and maxillary palp, the two main odor sensing organs in adult flies (58). Both the antenna and maxillary palp are covered with sensory hairs (sensilla) which house one to four olfactory receptor neurons (56). These neurons express transmembrane odorant receptors and project onto distinct glomeruli in the antennal lobe, the central olfactory processing center (56). Olfaction is sensed when a volatile compound (odorant) diffuses into a sensillum and binds its cognate olfactory receptor (59) thereby eliciting a stimulus that is processed by the antennal lobe (56).
The *Drosophila* genome encodes 62 different olfactory receptors, each of which is expressed in a particular type of olfactory neuron either alone or in conjunction with up to two additional olfactory receptor types (59, 124). All neurons expressing a given olfactory receptor project onto the same glomerulus within the antennal lobe (56). Extensive work has profiled the repertoire of each odorant receptor by recording responses of neurons ectopically-expressing olfactory receptors to a panel of 110 odors and revealed that *D. melanogaster* odorant receptors can detect a diverse set of organic compounds with varying sensitivity and response kinetics (60).

Previous work in our laboratory showed that the wild-type fly line Ral437 (125) can differentiate between volatiles produced by yeast under natural (nitrogen-limiting, YVN) or laboratory (sugar-limiting, YVL) conditions and that six volatile compounds mediate this attraction (55). Three of these compounds, ethyl hexanoate, ethyl octanoate, and isoamyl acetate, are recognized by the same odorant receptor, Or22a (60, 126). Intriguingly, genomic comparison of *Scaptomyza flava*, an herbivorous drosophilid, and *D. melanogaster* found that Or22a is one of two olfactory receptors conserved among drosophilids but completely lost in *S. flava*, suggesting that Or22a plays a role in the fungivorous lifestyle of *D. melanogaster* (127).

In *D. melanogaster*, Or22a is one of two tandem copies of Or22 (the other copy being Or22b) present at the Or22 locus on chromosome 2L (128). Both odorant receptors are expressed in basiconic sensilla of the ab3A olfactory neuron (59, 61). A tandem duplication of Or22 occurred in the *D. melanogaster* lineage prior to the divergence from *D. simulans* but after divergence from the *D. erecta* and *D. yakuba* lineage (128). The Or22 locus is functionally variable between *Drosophila* species, indicating that it is a quickly evolving region and likely under selective pressure (129, 130). In *D. erecta*, Or22 has evolved to sense odors from the host plant *Pandanus* spp (131). In *D. sechellia*, Or22a has specialized to detect odors that emanate from the host plant *Morinda citrifolia* while Or22b has decayed into a pseudogene (132).

In addition to being highly variable between species, studies have observed significant sequence variability at the Or22 locus between different lines of *D. melanogaster* (128, 133). A set of *D. melanogaster* lines were found to segregate by two variants at the Or22 locus: one variant contained both copies Or22a and Or22b (non-chimeric) while the other contained a chimera (termed Or22ab) consisting of the first exon of Or22a fused to the last three exons of Or22b (128). In addition to the length variants observed in *D. melanogaster* lines, some lines were also observed to possess an inversion on 2L whose breakpoint is just 0.7 Mb away from the Or22 locus; however, no association between the inversion and the length variant was observed (128). Despite tolerating quite a bit of variation, the Or22 locus has been implicated as a region undergoing positive selection in comparative population genetic studies of *D. melanogaster* in African and Europe (134). In Australia, the presence of the length variants is cinal, where all southern lines were non-chimeric at the Or22 locus and almost all northern flies were chimeric (133). Altogether, comparison between *D. melanogaster* lines and between related species suggests that the Or22 locus has recently undergone positive selection, though the conferred benefit that such selection indicates is unknown (128).

Previous work in our laboratory had demonstrated a strong correlation between sensitivity to YVN over YVL and the chimeric allele of Or22 using a trap-based olfactory assay (135). We hypothesized that the chimeric variant of the Or22 locus confers a heightened sensitivity to differences in yeast volatile bouquets and consequently contributes to flies’ ability to locate yeast in nature. Here, we sought to explore this
hypothesis by expanding our behavioral set with wild, inbred lines, assaying the progeny of reciprocal crosses from this set, analyzing Or22 sequences for polymorphisms that covaried with preference for YVN over YVL and, finally, use genome editing to swap allele types (chimeric for non-chimeric) in an otherwise identical genetic background.

**Results**

**Expanding the Or22 behavioral panel shows similar trend to original set**

We first sought to assay additional fly lines from around the globe in order to ascertain if the correlation between Or22 allele and increased sensitivity to YVN held up in a larger group. We obtained ten additional fly lines from Africa and Australia (Table 3.1) in addition to the existing panel of 14 lines, determined their allele type at the Or22 locus, then tested their sensitivity to YVN using a trap-based olfactory assay (55). These additional lines behaved consistent with our hypothesis that the chimeric allele mediates increased sensitivity to YVN (Figure 3.1).

![Figure 3.1. Behavior of each fly line in our Or22 behavioral panel in olfactory trap assay. Replicate behavioral experiments are plotted as blue dots; red lines indicate mean and black lines indicate standard deviation for all replicates. Positive preference index indicates a preference for yeast grown on limiting nitrogen over limiting sugar (i.e. sensitivity to YVN); preference index of 0 indicates lack of sensitivity to YVN; negative preference index indicates preference for YVL. Lines to the left of the black vertical line have chimeric Or22 alleles (Or22ab); lines to the right have non-chimeric (Or22a and Or22b) alleles. Orange asterisks next to fly line indicate that these are new additions to the D. melanogaster line.](image)
behavioral panel. Black octothorpes next to fly line indicate that these were included in the original panel (135) but retested for behavior for this study.

**Sequencing all panel fly lines at the Or22 locus failed to reveal nucleotide variation accounting for behavioral differences**

Next, we set out to clone and sequence each Or22 locus present in our behavioral panel. Although this seemed like a straightforward task, it proved to be immensely challenging. This difficulty, in fact, had been the reason that we had no sequence information for these alleles prior to the work described in this chapter. While the chimeric alleles can be amplified, cloned and sequenced with relative ease, non-chimeric alleles require a particular set of atypical conditions during PCR and a very large amount of template (see Methods). Additionally, Sanger sequencing across the non-chimeric alleles required different primers than for sequencing chimeric alleles, probably due to mis-priming issues in the presence of the non-chimeric tandem duplication. After much struggle, we were able to clone, Sanger sequence and assemble at least three Or22 amplicons from each fly line from which we generated a line consensus and called nucleotide polymorphisms (e.g. SNPs and indels). Analysis of these polymorphisms did not reveal any obvious pattern underlying sequence variants that consistently tracked with mean preference index for YVN (Figure 3.2).

![Figure 3.2. Sequence polymorphism analysis for 24 D. melanogaster lines in Or22 behavioral panel. For each polymorphism, a t-test was performed between the set of preference indices for YVN (PI) for lines where the variant was present or absent. Polymorphisms are ranked by p-value (shown below heatmap). No significant correlation between variance across trials for a given line and polymorphisms were found (data not shown). Fly lines are ordered by preference index (highest at top). Mean preference index](image-url)
for each line is given on the right. Black lines to the right of strain names indicate strains with chimeric Or22 allele. All other strains are non-chimeric.

**Replacement of a non-chimeric Or22 allele with a chimeric one does not confer sensitivity to YVN**

In order to directly test our hypothesis that chimeric alleles drive sensitivity to YVN, we first took advantage of the empty neuron odorant receptor system established by the Carlson laboratory to test odorant receptor function (59). First, we cloned a chimeric allele (Ral437-Or22) into a vector under the control of UAS expression, then, through a series of crosses, generated flies with this UAS construct and GAL4 expression under the control of the Or22 promoter in an Or22 null background (Δhalo, (61)). Unfortunately, the Δhalo homozygotes are very sick and their health was not improved by expressing Or22 in our final animals. We were unable to generate sufficient numbers of animals for our behavioral assay and moved to adopt a different approach.

We next turned to the CRISPR-Cas9 gene-editing system and began implementing a two-step allelic replacement scheme (Figure 3.S1). We opted to perform this swap in two steps rather than one due to the technical challenges we had encountered with amplification of the Or22 locus. Namely, we were concerned that if we moved to swap, for example, a non-chimeric allele with a chimeric allele that, while we would be able to robustly detect heterozygotes, when generating the homozygote it would be unclear whether the non-chimeric allele was successfully removed. In the converse swapping experiment, we would have the opposite problem: detecting heterozygotes would be difficult due to the preferential amplification of the chimeric over the non-chimeric allele. In order to aid our detection of transformants, we designed the first step to replace the Or22 allele with a visible marker (beta-tubulin GFP cassette) so we could use visual screening to identify heterozygotes during the first round of replacement and homozygotes during the second.

In the first round of replacement of a non-chimeric allele (OreR) with our place holder cassette, we learned that our visible marker was not the reliable indicator of transformation that we hoped it would be. Though we expected global GFP expression in our transformed heterozygotes, we observed a weak symmetric GFP signal in the thorax and abdomen (Figure 3.S1). This did lead us to identify some heterozygotes which were confirmed by non-lethal genotyping. Consistent with our expectations, the resultant homozygotes from these transformants behaved as the parental line (Figure 3). We continued with our second round of replacement to swap in a chimeric allele (ME) in the place of our visible marker and obtained homozygotes (MEΔOreR) and tested the behavior of these animals (Figure 3.3). These animals did not demonstrate an increased sensitivity to YVN and so did not support our hypothesis.
Figure 3.3. Behavior of donor lines (OreR and ME), intermediate (GFP\textDelta OreR) and swapped line (ME\textDelta OreR) in olfactory trap assay. Replicate behavioral experiments are plotted as blue dots; red lines indicate mean and black lines indicate standard deviation for all replicates. Positive or preference index indicates a preference for YVN; preference index of 0 indicates lack of sensitivity to YVN; negative preference index indicates preference for YVL.

**D. melanogaster line**

Crosses between chimeric and non-chimeric lines do not show a consistent inheritance pattern.

In parallel to functional studies, we performed crosses between chimeric and non-chimeric fly lines to determine the heritability of the Or22 locus with respect to behavioral sensitivity for YVN. We first crossed two fly lines with consistent, yet strikingly different behavioral responses to YVN over YVL. The OreR fly line is homozygous for the non-chimeric allele of Or22 and has no behavioral preference for yeast grown on YVN or YVL while the ME fly line is homozygous for the chimeric allele and exhibits a strong preference for YVN (Figure 3.1).

The OreR x ME cross was performed in both directions (i.e. one cross used an OreR mother and ME father; the other an OreR father and ME mother) and progeny were assayed for sensitivity for YVN. When a ME female was crossed to an OreR male, F1 progeny phenocopied the ME chimeric parental and preferred YVN over YVL. However, when an OreR female was crossed to a ME male, the F1s showed exhibited an intermediate phenotype (Figure 3.4).
Figure 3.4. Behavioral preference for YVN over YVL for F1 crosses between ME and OreR fly lines in comparison to parental behavior. Each point represents a single behavior assay and is colored by Or22 genotype. Thin black lines represent mean preference index while thick black lines represent standard deviations. Virgin fly lines for each cross are listed first. Positive preference index indicates a preference for yeast grown on limiting nitrogen over limiting sugar (i.e. sensitivity to YVN); preference index of 0 indicates lack of sensitivity to YVN; negative preference index indicates preference for YVL.

The directional inconsistencies of the cross suggest that the genetics underlying the sensitivity for YVN may be sex linked. In Drosophila melanogaster, the Or22 locus is located on chromosome 2L. Therefore, we did not expect a sex-linked inheritance pattern. To confirm these results, we crossed the chimeric Ral437 line to three different non-chimeric lines, OreR, CantonS, and Ral324, and phenotyped the F1 progeny for sensitivity to YVN (Figure 3.5). As a control, we also crossed the three non-chimeric lines to each other and phenotyped the F1 progeny. Again, the observed behaviors of these flies were inconsistent with our hypothesis that the Or22 allele is responsible for mediating sensitivity to YVN.
Figure 3.5. Behavioral preference for YVN over YVL for F1 crosses between one chimeric fly line and three non-chimeric fly lines. Each point represents a single behavior assay and is colored by Or22 genotype. Thin black lines represent mean preference index while thick black lines represent standard deviations. Virgin females for each cross are listed first. The only chimeric line tested here is Ral437, which is denoted as chimeric by a black underline. Positive preference index indicates a preference for yeast grown on limiting nitrogen over limiting sugar (i.e. sensitivity to YVN); preference index of 0 indicates lack of sensitivity to YVN; negative preference index indicates preference for YVL.

Within this set, we were particularly puzzled by the outcome of the Ral437 x Ral437 cross. Our previous experiments had found that Ral437 flies preferred YVN over YVL (Figure 3.1), but in this experiment Ral437 x Ral437 F1s showed no preference at all. We later determined that this inconsistency in behavior was a result of our Ral437 stock having passed through a population bottleneck between the time of these assays (for various reasons, the total number of adults in our Ral437 stock plummeted after the initial assays; all progeny of the second were considerably more inbred than previously). After genotyping the Or22 locus of the stocks before and after Ral437 bottleneck, we found that the original Ral437 stock was actually heterozygous at the Or22 locus. Most flies carried the chimeric allele but the non-chimeric allele was present and maintained at low abundance. During the bottleneck, the non-chimeric allele became over-represented, thus shifting the allele frequencies of the Ral437 stock from chimeric to non-chimeric. We believe that this explains the weaker preference for YVN over YVL in the original Ral437 fly line (Figure 3.1) the complete loss of preference in our subsequent cross experiment (Figure 4.5). Ultimately, we were unable to clarify the heritability of the Or22 locus from these data. However, at face value, these crosses suggest that the Or22 locus does not or is not the only locus underlying behavioral sensitivity for YVN.
Discussion

Though we initially observed a strong correlation observed between increased preference to YVN and a chimeric variant of Or22 and this correlation held when expanding the number of fly lines examined, our hypothesis that chimeric Or22 alleles confer heightened sensitivity to YVN did not hold. Here we consider unexplored aspects of Or22 and how to move forward from this negative result.

Possible epistatic interactions between Or22 sequence variants

Though it’s possible that there are epistatic interactions between polymorphisms that could significantly correlate with preference index, we opted to hold off on these analyses until we learned the results of our functional experiment, reasoning that if replacement of a non-chimeric Or22 allele with a chimeric one didn’t result in the expected behavior, these analyses would be in vain. As this turned out to be the case, these analyses were never performed.

ME∆OreR Or22 locus exhibits aberrant amplification behavior

We are puzzled by the observation that, despite all of our non-lethal PCR screening showing us exactly what we would have expected for transformants during the second round of replacement, PCR genotyping of the final homozygotes gave unexpectedly small bands (Figure 3.S2). When sequencing these bands and those from the screening steps prior, we observed the expected sequence, with the caveat that, in the non-lethal genotyping amplicons, the sequences became heterozygous after about half way through. We are hard-pressed to explain why, by all apparent measures, these animals appear to be our desired transformants and yet show this unexpected PCR phenotype. Though we don’t believe that our transformants have the wrong genotypes, we thought this was an important caveat that needs to be provided for future work on this project. Frankly, given all of the bizarre molecular behaviors of the Or22 locus that we’ve encountered, we’re not entirely surprised to find that this is the case.

Alternative hypothesis to explain variation in yeast attraction in D. melanogaster lines

At this juncture, it’s clear that Or22 does not explain why some D. melanogaster lines demonstrate sensitivity to YVN. Given the notable variance at this locus, it still seems possible that Or22 is in some way involved in attraction to yeast, though at this point we don’t understand the role it plays. Given the complexity in chemical signaling between yeast and flies, it seems more likely that the molecular basis for this attraction in flies lies not in one gene but in the combined or epistatic effects of many. This hypothesis would be best addressed by taking advantage of the Raleigh line collection, a set of recently established, iso-female D. melanogaster lines (125). As all of these lines have been sequenced, it would be feasible to find a subset of flies that vary in their response to YVN and perform a genome-wide association study to determine sequence polymorphisms that correlate with this preference.
Is the yeast attraction phenotype robust enough?

However, before such a study is performed, it should be considered whether the behavioral differences in the chosen panel of fly lines are consistent enough from generation to generation to make this a feasible feat and the variance isn’t so great that differences between lines cannot be relied upon. This can be ensured by increasing the number of replicate behavioral assays run: while the number that we ran was certainly appropriate for measuring previous phenotypes (55) it may need to be increased for subsequent experiments in this line of inquiry. It’s possible that different testing conditions (e.g. yeast strains) could reveal more robust behavioral differences between these lines. Though this is the best that we’ve encountered so far, it is certainly not the only set of conditions that could be used.

Given the importance of yeast to D. melanogaster, the variation in preference towards yeasts grown under different conditions observed in fly lines collected from around the world is likely to have ecological relevance. Understanding the basis of this variation can only improve our understanding of the complex relationship between flies and yeast.

Materials and Methods

Fly stocks

All Drosophila melanogaster lines used in behavioral panel are shown in Table 1. Additional lines used were Attp64 (BestGene) and w; Δhalo/CyO; Or22a-GAL4/TM3 (J.R. Carlson, personal communication). All lines were reared on medium from UC Berkeley’s Koshland fly kitchen (0.68% agar, 6.68% cornmeal, 2.7% yeast, 1.6% sucrose, 0.75% sodium tartrate tetrahydrate, 5.6 mM CaCl₂, 8.2% molasses, 0.09% tegosept, 0.77% ethanol, 0.46% propionic acid) supplemented with activated dry yeast pellets at 25°C on a 12:12 photoperiod unless otherwise stated.

Table 3.1. D. melanogaster lines used in behavioral panel.

<table>
<thead>
<tr>
<th>D. melanogaster line</th>
<th>Collection site</th>
<th>Or22 genotype</th>
<th>Source^A</th>
</tr>
</thead>
<tbody>
<tr>
<td>CantonS</td>
<td>Canton, OH, USA</td>
<td>Non-chimeric</td>
<td>Eisen laboratory stock (136)</td>
</tr>
<tr>
<td>OreR</td>
<td>Roseburg, Oregon, USA</td>
<td>Non-chimeric</td>
<td>Eisen laboratory stock (136)</td>
</tr>
<tr>
<td>Ral437</td>
<td>Raleigh, NC, USA</td>
<td>Chimeric</td>
<td>(125)</td>
</tr>
<tr>
<td>Ral324</td>
<td>Raleigh, NC, USA</td>
<td>Non-chimeric</td>
<td>(125)</td>
</tr>
<tr>
<td>Ral705</td>
<td>Raleigh, NC, USA</td>
<td>Non-chimeric</td>
<td>(125)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>----------------</td>
<td>------------</td>
<td>-----</td>
</tr>
<tr>
<td>GRAC</td>
<td>Crete, Greece</td>
<td>Non-chimeric</td>
<td>KMS &amp; ASQ</td>
</tr>
<tr>
<td>GR2</td>
<td>Crete, Greece</td>
<td>Chimeric</td>
<td>KMS &amp; ASQ</td>
</tr>
<tr>
<td>GR21</td>
<td>Crete, Greece</td>
<td>Chimeric</td>
<td>KMS &amp; ASQ</td>
</tr>
<tr>
<td>RW1001*</td>
<td>Montebello, CA, USA</td>
<td>Chimeric</td>
<td>KMS &amp; ASQ</td>
</tr>
<tr>
<td>RW1005*</td>
<td>Montebello, CA, USA</td>
<td>Chimeric</td>
<td>KMS &amp; ASQ</td>
</tr>
<tr>
<td>RW1008</td>
<td>Montebello, CA, USA</td>
<td>Non-chimeric</td>
<td>KMS &amp; ASQ</td>
</tr>
<tr>
<td>RW1011</td>
<td>Montebello, CA, USA</td>
<td>Non-chimeric</td>
<td>KMS &amp; ASQ</td>
</tr>
<tr>
<td>Cellar8.3*</td>
<td>Lytton Springs, CA</td>
<td>Chimeric</td>
<td>ASQ</td>
</tr>
<tr>
<td>ME</td>
<td>Bowdoin, ME, USA</td>
<td>Chimeric</td>
<td>(137, 138)</td>
</tr>
<tr>
<td>PEN</td>
<td>Media, PA, USA</td>
<td>Chimeric</td>
<td>(137, 138)</td>
</tr>
<tr>
<td>FL</td>
<td>Homestead, FL, USA</td>
<td>Non-chimeric</td>
<td>(137, 138)</td>
</tr>
<tr>
<td>MAU9*</td>
<td>Rockhampton, AU</td>
<td>Non-chimeric</td>
<td>(139)*</td>
</tr>
<tr>
<td>MAU24*</td>
<td>Rockhampton, AU</td>
<td>Chimeric</td>
<td>(139)*</td>
</tr>
<tr>
<td>MAU31*</td>
<td>Rockhampton, AU</td>
<td>Chimeric</td>
<td>(139)*</td>
</tr>
<tr>
<td>FP6*</td>
<td>Sydney, AU</td>
<td>Non-chimeric</td>
<td>(139)*</td>
</tr>
<tr>
<td>FP8*</td>
<td>Sydney, AU</td>
<td>Non-chimeric</td>
<td>(139)*</td>
</tr>
<tr>
<td>FP16*</td>
<td>Sydney, AU</td>
<td>Chimeric</td>
<td>(139)*</td>
</tr>
<tr>
<td>CW105*</td>
<td>Mbengwi, Cameroon</td>
<td>Chimeric</td>
<td>(134)*</td>
</tr>
<tr>
<td>EZ2*</td>
<td>Ziway, Ethiopia</td>
<td>Chimeric</td>
<td>(134)*</td>
</tr>
<tr>
<td>SP90*</td>
<td>Phalaborwa, South Africa</td>
<td>Non-chimeric</td>
<td>(134)*</td>
</tr>
<tr>
<td>ZS56*</td>
<td>Sengwa, Zembabwe</td>
<td>Chimeric</td>
<td>(134)*</td>
</tr>
</tbody>
</table>

*Fly lines that were added to initial set examined by (135).
^KMS = Kelly Schiabor, unpublished; ASQ = Allison Quan, unpublished
#Provided by the Begun laboratory (UC Davis)

Olfactory behavior assay

The behavior assays in this chapter were performed as described in (55) Briefly, *Drosophila melanogaster* lines were raised at room temperature (21-23C) on Koshland diet. Newly eclosed flies were pushed onto new food daily and aged at room temperature for at least
four days under ambient lighting conditions (i.e. adjacent to a window) before being used in behavior assays.

The day prior to the start of the behavior assay, the *Saccharomyces cerevisiae* strain, I14 (140), was plated onto either YVN or YVL media (Table 2) and grown at 30°C for 22 hours. Two plates for each media type were streaked out per arena.

### Table 3.2. Yeast media recipes used in behavior assays.

<table>
<thead>
<tr>
<th>Media*</th>
<th>Recipe</th>
</tr>
</thead>
<tbody>
<tr>
<td>YVN</td>
<td>1.7g YNB without amino acids and ammonium sulfate (BD Difco), 2.0g SC amino acid mixture (MP Biomedicals), 50g dextrose (BD Difco), 20g agar (BD Difco), MilliQ water to 1L</td>
</tr>
<tr>
<td>YVL</td>
<td>6.7g YNB without amino acids (BD Difco), 50g dextrose (BD Difco), 20g agar (BD Difco), MilliQ water to 1L</td>
</tr>
</tbody>
</table>

*One liter batches were made every one to two weeks and poured into 60 x 10mm petri dishes (55).

The following day, grown plates were removed from the incubator, fitted with a custom, 3D printed lid, and secured with Parafilm. Lids were fitted with a 50mL conical centrifuge tube (Falcon) with the end removed. A funnel was folded from 150mm filter paper (Whatman, Grade 1) and a 5mm hole snipped off the tip. This funnel was used to top the centrifuge tube and secured with tape. Two traps for each media type were placed into behavior arenas (*Drosophila* population cages, 24” x 12” clear acrylic cylinders, TAP plastics) fitted with netting (Genesse Scientific) as shown in Figure 6. All possible orientations of YPD and SC plates within were tested to control for environmental effects (e.g. attractiveness to light).

![Image of behavior assay](image.png)

**Figure 3.6. Schematic of behavior assay per (55).** *Saccharomyces cerevisiae*, strain I14 (140), was grown on either YVN or YVL media and used to bait custom made traps. One hundred and twenty adult *Drosophila* (4-10 days old) of mixed sex were allowed to choose between traps over an 18 hour period. Preference was quantified by the number of flies in each trap at the end of the assay.
One hundred and twenty 4-10 day old mixtures of male and female *Drosophila melanogaster* were anesthetized with CO2 and allowed to recover on Koshland diet for 2 hours before being used in behavior assays. Flies were introduced into behavior arenas at 3pm and allowed survey traps. After 18 hours, traps were removed from the arena and the number of flies in each trap were counted, sexed and recorded. Flies were discarded after counting so flies were only used in behavior assays once. A preference index was calculated from the number of flies in each trap as follows:

For A = total number of flies in YVN traps (SC)

For B = total number of flies in YVL traps (YPD)

Preference Index = (A - B)/(A + B)

**Genomic DNA extraction from behavior panel fly lines**

For each line, three females were pooled in a single DNA extraction using either the QIAamp DNA Micro (QIAGEN) following the manufacturer’s instructions for the isolation of genomic DNA from less than 10 mg tissue or the PureGene Tissue kit (Gentra). Concentration of each DNA sample was quantified using the Qubit High Sensitivity dsDNA kit.

**Cloning Or22 alleles via TOPO TA**

Or22 alleles were cloned by amplification with GoTaq mastermix (Promega) using 240 ng of template gDNA, and 400 nM each o2F and o2R (Table 3.3) in a 50 uL reaction. Reactions were cycled using a specialized thermocycler protocol (Montserrat Aguadé, personal communication): an initial melting step of 94 C for 3 min followed by 35 cycles of 96 C for 10 seconds, 55 C for 10 seconds, 65 C for 4.5 min, then a final polymerase elongation step of 65 C for 7 min. Expected bands were excised from 1% agarose gels after running at 100V and gel purified using the QIAquick Gel Extraction (QIAGEN) kit eluting in 30 uL of buffer EB. Adenosine tails were added to these fragments in anticipation of TOPO TA (Invitrogen) cloning with 5U Taq polymerase (NEB), 280 uM dNTPs in 1x standard Taq buffer (NEB) for 20 minutes at 72C. A-tailed products were then immediately cloned into TOPO TA 2.1 vector using manufacturer’s instructions. Fresh TOPO TA reactions were drop-dialyzed on 0.025 um membrane (Millipore) floated in a 100x15 mm petri dish with sterile DI water (~25 mL) for 15 minutes at RT. Drop-dialyzed TOPO TA reactions were then transformed into DH5alpha *E. coli* via electroporation, rescued immediately with room temperature SOC and outgrown 15 min at 37 C with 180 rpm shaking before plating all cells pre-warmed LB + carbencillin (100 ug/mL) agar plates. Plates were incubated overnight at 37 C. Colonies were picked and dissolved into 5 uL of LB + kanamycin (50 ug/mL) in 96-well plates. One uL of cell suspensions were then used to template 20 ul colony PCR reactions with GoTaq mastermix (Promega) using primers o2F and o2R (800 nM each) with the following thermocycler settings: initial melt at 95C for 5 min followed by 35 rounds of 95 C for 30 seconds, 51 C for 30 seconds and 72 C for 2.5 min, then a final extension step at 72 C for 10 min. Positive hits were those that gave a 2.5 kb bands when run on a 1% agarose gel. Up to five colonies for each fly line were grown overnight in LB + kanamycin (50 ug/mL) and plasmids were extracted via MiniPrep (QIAGEN).
Cloning Or22 alleles via pUC19 Gibson assembly

Or22 alleles were cloned by amplification with GoTaq mastermix (Promega) using 240 ng of template gDNA, and 400 nM each o2F-pUC19 and o2R-pUC19 (Table 3.3) in a 50 uL reaction. Reactions were cycled using a specialized thermocycler protocol (Montserrat Aguadé, personal communication): an initial melting step of 94°C for 3 min followed by 35 cycles of 96°C for 10 seconds, 55°C for 10 seconds, 65°C for 4.5 min, then a final polymerase elongation step of 65°C for 7 min. PUC19 backbone was amplified from pUC19 (Invitrogen) using pUC19-PCR-F1 and pUC19-PCR-R1 (500 nM each, Table 3.3) with Q5 High-Fidelity DNA Polymerase (NEB) with the following conditions: 98°C for 30 sec followed by 30 rounds of 98°C for 10 sec, 62°C for 30 sec then 72°C for 1 min, finishing with 72°C for 2 min. Expected bands were excised from 1% agarose gels after running at 100V and gel purified using the QIAquick Gel Extraction (QIAGEN) kit eluting in 30 uL of buffer EB. Or22 bands were mixed with PUC19 backbone and assembled with NEBuilder HiFi DNA Assembly (NEB) incubating 1 hour at 50°C but otherwise following manufacturer’s instructions. Gibson reactions were then dialyzed and transformed into DH5alpha E. coli; transformants were screened and plasmid was extracted as with TOPO TA cloning with the difference that only LB + carbenicillin (100 ug/mL) agar plates were used for selection.

Table 3.3. All primers used in Chapter 3.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
<th>Source*</th>
</tr>
</thead>
<tbody>
<tr>
<td>o2F</td>
<td>TAACACCGCCAATGGTCAAC</td>
<td>(128)</td>
</tr>
<tr>
<td>o2R</td>
<td>TCTTGCTGTGTGACCCATCTC</td>
<td>(128)</td>
</tr>
<tr>
<td>o3F</td>
<td>GGGTGGAAGAGTTTTGAA</td>
<td>(128)</td>
</tr>
<tr>
<td>o2F-pUC19</td>
<td>TTGTAAAACGACGCGCAGTAACACCGCAATGGTC</td>
<td>CNE</td>
</tr>
<tr>
<td>o2R-pUC19</td>
<td>CTATGACCAGGATTACGCTGACCCATCTTC</td>
<td>CNE</td>
</tr>
<tr>
<td>pUC19-PCR-F1</td>
<td>CACTGGCGCTCGTTTTTACAA</td>
<td>CNE</td>
</tr>
<tr>
<td>pUC19-PCR-R1</td>
<td>TGGCGTAATCATGTCGTCACAG</td>
<td>CNE</td>
</tr>
<tr>
<td>o4F</td>
<td>GAGAGAATAACAGGGAAAATG</td>
<td>(128)</td>
</tr>
<tr>
<td>o4R</td>
<td>CATTGCCCTTGTATTCTCTCACAA</td>
<td>(128)</td>
</tr>
<tr>
<td>Or22_long_3</td>
<td>GATTTGATGAGCCGTAAGTCTTTTT</td>
<td>KMS</td>
</tr>
<tr>
<td>Or22b_1</td>
<td>CAACTTTGCATGACCATTGTTG</td>
<td>(130)</td>
</tr>
<tr>
<td>Or22b_P</td>
<td>TTGAAACTTTTCTGCCAGG</td>
<td>(130)</td>
</tr>
<tr>
<td>Or22b_J</td>
<td>CAGGAAGGACGGAAGATGAG</td>
<td>(130)</td>
</tr>
<tr>
<td>Or22_long_3-flip</td>
<td>AAAAGGACTTACGGTCATCAATC</td>
<td>KMS</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Or22_sl_5</td>
<td>AAAACAAAGCCACGGACAAG</td>
<td>KMS</td>
</tr>
<tr>
<td>Or22a_E</td>
<td>AAGTCTCCATGGACAGTC</td>
<td>(130)</td>
</tr>
<tr>
<td>Or22a_G-flip</td>
<td>GTGCATTCCGGATCATCGAT</td>
<td>(130)</td>
</tr>
<tr>
<td>Or22-pWALIUM-F</td>
<td>GGAATTTGGAATTCCGAAGCTGAAATGTAACCTGC</td>
<td>CNE</td>
</tr>
<tr>
<td>Or22-pWALIUM-R</td>
<td>GAACTAGTGTGCTCTAGAGTGCAGAAAGACGACAAGACACTG</td>
<td>CNE</td>
</tr>
<tr>
<td>pWALIUM-F</td>
<td>TCTAGAGCAAAACTAGTTCTG</td>
<td>CNE</td>
</tr>
<tr>
<td>pWALIUM-R</td>
<td>TGCAGATTTCCAATTCCC</td>
<td>CNE</td>
</tr>
<tr>
<td>pUAST-MCS-F1</td>
<td>AGCGCAGCTGAACAGCTA</td>
<td>CNE</td>
</tr>
<tr>
<td>pUAST-MCS-R1</td>
<td>TGTCCAAATTTATGTCACACCA</td>
<td>CNE</td>
</tr>
<tr>
<td>5'-out-pCFD4-F2</td>
<td>TATATAGGAAGAGATATCGGGTGAATTCGGAACGGCAATGCAAGCATCGAAATAGCAAG</td>
<td>CNE</td>
</tr>
<tr>
<td>3'-out-pCFD4-R2</td>
<td>ATTTAATCTGCTTTTCTAGCTCTAAACACCATTGATTTGATGATGAGCGACGTTAATGGTC</td>
<td>CNE</td>
</tr>
<tr>
<td>pCFD4-seq</td>
<td>GACACAGCAGTACGTCTTCG</td>
<td>(141)</td>
</tr>
<tr>
<td>Or22-5'flank-pUC19-F1</td>
<td>TTGTAAAACGACGGCCAGTGGTGTCGCACTGACGACGTAATCCATTG</td>
<td>CNE</td>
</tr>
<tr>
<td>GFP-HR-5'-R</td>
<td>CCGCAGCTTACATTTCTAGCTGAAACACGGCAATATCAATTCGTCC</td>
<td>CNE</td>
</tr>
<tr>
<td>Or22-Ral437-NEB-F1</td>
<td>AATTCCATTTCAGCTGAAATGTAACCTGC</td>
<td>CNE</td>
</tr>
<tr>
<td>Or22-Ral437-NEB-R1</td>
<td>GTTGACCACATCTCAATCACCATCG</td>
<td>CNE</td>
</tr>
<tr>
<td>GFP HR tubP-F</td>
<td>TTCAGCTGAAATGTAACCTCGGTGGCACACTGCGGCCATCG</td>
<td>CNE</td>
</tr>
<tr>
<td>GFP HR tubP-R</td>
<td>CCTGCACCCTTGGTCAACCATAACACAAACTGTCGCGC</td>
<td>CNE</td>
</tr>
<tr>
<td>GFP-HR-GFP-F</td>
<td>GCGGACAGTTGTGTATGTCAGAGGCCAGG</td>
<td>CNE</td>
</tr>
<tr>
<td>GFP-HR-GFP-SV40-R</td>
<td>ACCCATTCCTACATGCCCAGGTTACTGTAACAGTCCGTCC</td>
<td>CNE</td>
</tr>
<tr>
<td>GFP-HR-3'-F2</td>
<td>CCTGGGTGAGATGGTGCTCATCATCAATC</td>
<td>CNE</td>
</tr>
</tbody>
</table>
Sequencing and assembling Or22 loci for each fly line in behavioral panel

Chimeric Or22 alleles were Sanger sequenced with six primers (o2F, o2R, Or22_long_3, Or22b_1, Or22b_P, Or22b_J) and non-chimeric Or22 alleles were Sanger sequenced with 10 primers (o2F, o2R, o3F, o4F, o4R, Or22_long_3-flip, Or22_sl_5, Or22b_1, Or22a_E, Or22a_G-flip) using third party services (ELIM, Barker Hall Sequencing facility) (Table 3.3). Loci were assembled from these sequences using SeqMan Pro (DNASTAR Lasergene v.10) after lowering signal threshold to 2 and manually checking and resolving any disagreements between reads. A consensus for each line was obtained by aligning all individual clones for a given fly line in SeqMan Pro.

Polymorphism analysis for Or22 sequences

The consensus sequence for each line was assembled to the Or22 genomic reference using Geneious (version 5.1.7); chimeric sequences were split at the first intron in order to achieve alignment of the entire locus. Indels and SNPs were called manually for each consensus compared to the consensus reference of all sequenced Or22 loci to generate a
presence/absence matrix of all observed polymorphisms in our set of sequenced Or22 loci. T-tests comparing the set of preference indices or variances (data not shown for latter) for all lines possessing versus lacking a given allele were performed using the stats library in Python. Data were plotted with Prism 7 (GraphPad).

Empty neuron (Δhalo) experiment

UAS-Or22Ral437 was generated by cloning the open reading frame of Ral437 Or22 downstream of the 5x UAS in pWALIUM10 (M.R. Stadler, personal communication). To do this, Or22 was amplified from Ral437 Or22 in TOPO TA vector (Invitrogen) using primers Or22-pWALIUM-F and Or22-pWALIUM-R (500 nM each, Table 3.3) and pWALIUM backbone was amplified from pWALIUM using primers pWALIUM-PCR-F and pWALIUM-PCR-R (500 nM each, Table 3.3) with Q5 High-Fidelity DNA Polymerase (NEB) with the following conditions: 98C for 30 sec followed by 30 rounds of 98C for 10 sec, 62C for 30 sec then 72 for 45 sec (Or22) or 3:15 min (pWALIUM), finishing with 72C for 2 min. The resultant products were gel-purified, Gibson assembled (NEBuilder HiFi DNA Assembly Master Mix, NEB) transformed into chemically-competent DH5alpha E. coli (NEB #C2987) and selected for on LB + carbencillin (100 ug/mL) agar plates. Plasmid was extracted from 2-4 transformant clones (QIAGEN miniprep) and sequenced with pUAST-MCS-F1 and pUAST-MCS-R1 (Table 3.3) to confirm proper insertion had taken place. Plasmid was extracted from a verified clone (QIAGEN midiprep) and quantified using the Qubit HS dsDNA kit (Thermo Fisher Scientific). This plasmid was injected into AttP64 flies in the presence of PhiC31 recombinase and progeny were backcrossed, screened and balanced with TM3,Sb (BestGene). These w; +; UAS-Ral437or22/TM3,Sb flies were crossed according the scheme from (59). First, UAS-Ral437or22/TM3,Sb were crossed to w; Δhalo/Cyo; Or22a-GAL4/TM3 to generate w; Δhalo/+; UAS-Ral437or22/TM3 and w; Cy0/+; UAS-Ral437or22/TM3 progeny. These progeny were crossed to generate w; Δhalo/Cyo; UAS-Ral437or22/TM3 which were crossed back to w; Δhalo/Cyo; Or22a-GAL4/TM3 to generate w; Δhalo/Δhalo; UAS-Ral437or22/Or22a-GAL4 flies.

CRISPR-Cas9 Or22 allele replacement

First and second round CRISPR targets were selected using http://tools.flycrispr.molbio.wisc.edu/targetFinder/ (Table 3.4). Primers 5'-out-pCFD4-F2 and 3'-out-pCFD4-R2 or 5'-in-CFD-R and 3'-in-CFD-F were used to clone both sgRNAs into pCFD4 for the first or second round of CRISPR editing, respectively, per http://www.crisprflydesign.org/ (141). Constructs were verified by Sanger sequencing with pCFD4-seq (141). To generate a homologous recombination template plasmid for the first round of replacement, five fragments (pUC19 backbone (pUC19-PCR-F1 and pUC19-PCR-R1, one kilobase 5’ upstream of OreR Or22 locus (Or22-5’ flank-pUC19-F1 and GFP-HR-5’-R), beta-tubulin promoter from OreR (GFP HR tubP-F and GFP HR tubP –R), GFP with SV40 3’ UTR from pGREEN-Pelican (GFP-HR-GFP-F and GFP-HR-GFP-SV40-R) and one kilobase 3’ downstream of OreR Or22 locus (GFP-HR-3’-F2 and Or22-3’ flank-pUC19-R2) were amplified with Q5 High-Fidelity DNA Polymerase (NEB) using 500 nM each forward and reverse primer (Table 3.3) with the following conditions: 98C for 30 sec followed by 30 rounds of 98C for 10 sec, 62C for 30 sec then 72 for 2 min, finishing with 72C for 2 min. Fragments were assembled using NEBuilder HiFi DNA Assembly Master Mix (NEB), transformed into DH5alpha electrocompetent cells and plated on LB + carbencillin (100
ug/mL). Plasmid was isolated from 2-4 transformants and sequenced with primers M13F, M13R, o2F and CNE-Or22-R2 (ELIM) to confirm assembly (Table 3.3).

<table>
<thead>
<tr>
<th>Table 3.4. CRISPR targets for Or22 allelic replacement.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Round I (out)</strong></td>
</tr>
<tr>
<td>GAAAGGCAATGATATTGGGGGCGGG</td>
</tr>
<tr>
<td><strong>Round II (in)</strong></td>
</tr>
</tbody>
</table>

*PAM sites are underlined

Plasmid from a sequence-verified clone was prepared (QIAGEN Midiprep) and quantified using the Qubit HS dsDNA kit (Thermo Fisher Scientific). OreR flies were co-injected with pHsp70-Cas9, pCFD4 containing the two synthetic guides for round I editing (outer CRISPR targets) and the GFP homologous recombination donor (Rainbow Transgenics). Injected animals were individually backcrossed to OreR; progeny were screened using a compound fluorescence microscope and by extracting DNA from pools of 50 animals from each cross (VDRC stock center protocol “Good quality Drosophila genomic DNA extraction”) then amplifying the 5’ and 3’ PAMs using a cocktail of primers 5’ PAM F, GFP-5’PAM-R, Or22-5’PAM-R or primers 3’ PAM R, SV40-3’PAM-F, Or22-3’PAM-F, respectively (Table 3.3), at a total final concentration of 1 μM for each forward and reverse primer(s) with GoTaq 2x mastermix (Promega) with the following thermocycling conditions: 95°C for 5 min followed by 35 iterations of 95°C for 30 seconds, 61°C for 30 seconds then 72°C for 30 sec then 72°C for an additional 10 minutes. Sibling virgins from “hit” founder crosses were screened by non-lethal genotyping using each 5’ PAM (5’ PAM F, GFP-5’PAM-R, Or22-5’PAM-R) and 3’ PAM (3’ PAM R, SV40-3’PAM-F, Or22-3’PAM-F) primer cocktails per (142). Heterozygotes were crossed and progeny screened as above to identify homozygotes. Homozygotes were crossed, progeny were screened as above and genomic DNA from two batches of three females each was extracted with the QIAamp Micro kit (QIAGEN) then PCR genotyped and Sanger sequenced using three sets of primers to confirm homozygosity: 5’ PAM (5’ PAM F, GFP-5’PAM-R, Or22-5’PAM-R), 3’ PAM (3’ PAM R, SV40-3’PAM-F, Or22-3’PAM-F) and whole locus (5’ PAM F, 5’PAM-R) PAM primer sets and 5’PAM/3’PAM. Sibling flies were propagated as GFPΔOreR.

An analogous process was used for the second round of editing; this time pCFD4 contained the synthetic guides for round II (inner) CRISPR targets and the donor plasmid contained the Or22 allele from ME sandwiched between 5’ and 3’ Or22 flanking regions (assembled identically to the first round homologous donor template instead using four fragments (pUC19 backbone (pUC19-PCR-F1 and pUC19-PCR-R1, one kilobase 5’ upstream of OreR Or22 locus (Or22-5’flank-pUC19-F1 and GFP-HR-5’-R), ME Or22 locus (Or22-Ral437-NEB-F1 and Or22-Ral437-NEB-R1) and one kilobase 3’ downstream of OreR Or22 locus (GFP-HR-3’-F2 and Or22-3’flank-pUC19-R2) (Table 3.3)). These constructs and pHsp70-Cas9 were co-injected into GFPΔOreR (Rainbow Transgenics). Injected animals were individually back-crossed to GFPΔOreR then screened and homozygosed as above to establish line MEΔOreR.
Chimeric and non-chimeric crosses

Fly lines were raised at 25°C and virgins and males were collected twice a day. After five days, virgins were confirmed. Five females of a single fly line were crossed to five males of another line. Three replicates of each cross were set up and crosses were performed in both directions. As a control, virgins and males of parental fly lines were collected and crossed in parallel. F1 progeny were collected and aged for behavior assay.
Supporting information

**Figure 3.S1. Two step Or22 allelic replacement scheme using CRISPR-Cas9 genome editing.** In the first round of CRISPR editing, embryos of line X are injected with a plasmid expressing both 5’ and 3’ CRISPR #1 target synthetic guide RNAs (sgRNAs), a plasmid expressing Cas9 and a plasmid bearing a dominant visible marker (β-tubulin promoter GFP cassette) flanked by one kb sequences lying upstream and downstream, respectively, of the cut sites to template homologous recombination. The homologous template has mutated PAM sites for the #1 guides, so the resultant recombinant product will not have usable PAM sites. In the second round of CRISPR editing, embryos of LineX Or22ΔGFP are injected with a plasmid expressing both 5’ and 3’ CRISPR #2 target synthetic guide RNAs (sgRNAs), a plasmid expressing Cas9 and a plasmid bearing an alternative allele of Or22 flanked by one kb sequences lying upstream and downstream, respectively, of the cut sites to template homologous recombination. The homologous template has mutated PAM sites for the #1 guides, so the resultant recombinant product will not have usable PAM sites.
Figure 3.2. Aberrant amplification behavior of MEOreR Or22 locus. A) Hyperladder 1 kb (Bioline) used for B) and C). B) PCR behavior of GFPΔOreR with 5’ PAM cocktail (5’PAM-F, Or22-5’PAM’-R and GFP-5’PAM-R), 3’ PAM cocktail (3’PAM-R, Or22-3’PAM-F and SV40-3’PAM-F) or 5’PAM-F and 3’PAM-R (Table 3.3). First three lanes after ladder are one DNA prep from three GFPΔOreR females; second three lanes are a second DNA prep from another three GFPΔOreR females. C) PCR behavior of MEOreR as in B as well as primer pairs o2F/o2R, o1F/Or22b-P, o1F/Or22b-2, Or22b-1/Or22-CNE-long-R2 and Or22b-J/Or22-CNE-long-R2 (Table 3.3). For two leftmost gels, adjacent lanes with same primer pairs are templated by two different DNA preps, each from three MEOreR females. In rightmost gel, only one of these DNA preps is used to template all reactions. Red boxed indicate bands of unexpected sizes from these reactions. D) Relative positions and orientations of primers in ME (blue) and GFP-SV40 (green) constructs. Note that primers marked in red are outside of the 5’ and 3’ homology flanks used in the donor plasmid. Diagram is not precisely to scale.
Chapter 4: Discovery and isolation of *Entomophthora muscae* CNE1, a strain of a behavior-manipulating fungal pathogen that naturally infects *D. melanogaster*

Abstract

*Entomophthora muscae* is a pathogenic fungus that infects, alters the behavior of and kills dipterans (true flies). On the last day of life, always around sunset, infected flies climb upward, extend their proboscis, move their wings away from their dorsal abdomen, and die. The *E. muscae*-induced end of life behaviors by have been observed and studied for over 160 years but we still have no understanding as to how they occur owing to a lack of genetic and molecular tools in either host (predominantly house fly) or fungus. Recently, I discovered an *E. muscae* strain (CNE1) that naturally infects the model organism *Drosophila melanogaster*. I successfully infected laboratory-reared wild-type *D. melanogaster* with this strain and can grow the isolate *in vitro* in liquid culture. I have developed a robust protocol for propagating the infection *in vivo* in CantonS flies. Under these conditions, *E. muscae* CNE1 kills the majority of exposed fruit flies 96-168 hours after co-housing with sporulating cadavers. Importantly, *E. muscae* CNE1 is not a contamination risk for other healthy fly stocks. The *D. melanogaster-E. muscae* system presents a unique opportunity to study myriad facets of host-pathogen biology, including the basis of behavioral manipulation.

Introduction

*Entomophthora muscae* is a fungal parasite that infects, behaviorally manipulates, and kills its fly hosts. The most striking of the behavioral manipulations induced by *E. muscae* is the

---

3 XKCD comic by Randall Munroe, available under Creative Commons Attribution-NonCommerical 2.5 Generic license.
sequence of actions an infected fly takes on the last day of its life. Critically-infected flies start acting bizarrely around sunset (68). First, they climb a nearby substrate (a behavior known as “summit disease”) and extend their proboscis (68). Fungal material is secreted from the proboscis, serving to “glue” the fly in place (67). Then, the fly raises its wings up and away from its dorsal abdomen and dies (68). By dying in this specific position, flies serve as an ideal platform for fungal dispersal: over the next few hours, the fungus emerges through the intersegmental membranes of the dorsal abdomen and ejects infectious conidia, or spores, with the goal of landing on a new host (70, 73, 143). The conidia do not need to be ingested to establish infection, but rather are able to bore through the host cuticle to access the hemolymph (69). There they grow, using the fly’s fat body as an energy source until these reserves are gone (69). Then, around sunset, the fungus instructs the fly to summit, extend its proboscis and raise its wings so the cycle can start anew.

_E. muscae_ has been studied since its first description in 1855 (65). Most of what we know about _E. muscae_ comes from work done observing natural epizootic events in Muscoidea species and a handful of laboratory studies using house flies. Despite these efforts, little progress has been made in terms of understanding _E. muscae_ infection largely due to the inability to perform experimental manipulations in either the host or fungus. Though experimenters have tried to infect the model host _Drosophila melanogaster_ with _E. muscae_ to overcome this hurdle, these efforts haven’t been successful (Ben de Bivort, personal communication).

Here, I describe the discovery of an _E. muscae_ isolate that naturally infects _D. melanogaster_ and the protocols I developed to culture this isolate both _in vivo_ (in fruit flies) and _in vitro_ (in liquid culture). The _D. melanogaster-E. muscae_ system offers a unique opportunity to study the basis of parasitic manipulation of a host’s nervous system and, in so doing, better understand the molecular logic that governs normal nervous system function.

**Results**

**Discovery of an Entomophthora muscae strain infecting wild Drosophila**

In June 2015, I established a stable food source (organic watermelon in a clean dish pan, referred to henceforth as the “fendel”) on my back porch in Berkeley, CA so I could collect wild _Drosophila_ and determine their microbial load (see Chapter 2 Materials and Methods). While tending to the fendel in late July 2015, I noticed several dead flies that had their wings extended at an abnormal angle (approximately 90 degrees to the body axis). I inspected these flies under a dissecting microscope and observed remnants of fungal growth piercing through their intersegmental membranes as well as old conidia stuck to their wings and body (Figure 4.1 A, B). (For reference, “fresh” cadavers (those that had sporulated within 24 hours of observation) that were later discovered in the fendel are shown in Figure 4.1C and D.) I knew from previous experience that these features were hallmarks of infection by _Entomophthora muscae_.
Figure 4.1. Wild Drosophila killed by Entomophthora muscae. A, B) Representatives of first of E. muscae infected Drosophila cadavers found in fendel on 7/25/2015 and transferred to laboratory for examination. These cadavers had sporulated more than 24 hours prior to observation. C, D) E. muscae infected cadavers found in fendel on 9/25/15 and 7/27/15, respectively and left in situ. They had sporulated within 24 hours of observation.

To confirm that the fungus was E. muscae, I then genotyped representative cadavers using two sets of fungal-specific primers, one at the internally transcribed spacer locus (ITS) and the other at the ribosomal large subunit locus (LSU)(Figure 4.S2). Comparing the ITS sequence to what’s available in the NCBI database showed that the sequence belonged to E. muscae. All four cadavers yielded identical sequences at these loci which was consistent with (though not definitive evidence of) a single strain of E. muscae circulating in this population of flies. As expected, genotyping the cadavers at the cytochrome oxidase II locus confirmed that they were all drosophilids (two cadavers were D. melanogaster, two were D. hydei). To distinguish this strain from other described strains of E. muscae species complex, I will refer to this isolate as E. muscae CNE1⁴.

Isolating E. muscae CNE1 in vitro and in vivo

Given that E. muscae is known to evoke very striking behavioral changes in its host, I immediately knew that I wanted to bring this strain into the laboratory for study. In order

---

⁴ As previously mentioned, there is a dearth of molecular data for Entomophthora; sequencing is not the primary factor in species determination. My use of sequencing here is primarily to demonstrate that the fungus belongs within the E. muscae species complex, and not to pinpoint the exact species designation.
to isolate the fungus either *in vivo* (growing in fly hosts) or *in vitro* (growing axenically in liquid culture), I needed to find the cadaver of an infected fly before prior to sporulation (i.e. before it became infectious). The lifestyle of the fungus helped me in this regard, as the fungus only kills the host once a day around sunset and doesn’t begin to emit spores until some hours after (68, 75). I reasoned that the old cadavers I had found in the fendel likely would have infected other flies that were patrons of the fendel, so if I sampled from the fendel population and monitored the flies eventually I should observe an infected fly that would give rise to an infectious cadaver. I could then place that cadaver in a vial with healthy flies to spread the infection *in vivo* or use the ascending method of conidia collection to aseptically gather spores and transfer to grow *in vitro* in liquid medium, where the fungus would grow as protoplasts (144).

I monitored the fendel daily for two months following this approach. Spores were successfully collected from one of the first cadavers I collected (dubbed the “fendel mama”) and used to start a liquid *in vitro* culture of the fungus (Figure 4.2). Genotyping at the ITS and LSU loci confirmed that the cadaver used to start the culture and the culture itself matched what I had observed in the initial cadavers. After a few unsuccessful attempts, I was able to cryopreserve *in vitro* culture in 10-25% glycerol (see Methods).

**Figure 4.2. Isolating *E. muscae* CNE1 *in vitro*.** A) Conidia from a single cadaver (“fendel mama”) were deposited on the lid of a sterile petri dish. B) Fendel mama. C) Initial culture morphology after two weeks of growth.

Next, I focused on passing the infection from fendel flies to laboratory-reared flies via collecting spores from a single, sporulating fly placing any observed cadavers on Koshland fly diet with different strains of wild-type, laboratory-reared flies and monitored these “exposure” vials in the laboratory for any signs of infection or death daily for two weeks. To improve my chances of success, I rotated through a set of 14 wild type fly lines, including four isofemale lines that I had established from flies caught on my balcony, reasoning that if at least one of these strains should be susceptible to fungal infection. I checked for cadavers among my sampled wild flies every night (e.g. Figure 4.3A), finding that any flies that would die of infection on a given day would do so by two hours after sunset. I then used cold anesthesia (2-3 minutes in my home freezer) to slow the living flies co-housed with the cadavers and quickly grab the cadaver(s) to place in a new vial of Koshland diet. I then transferred available wild-type flies into the same vial housing the cadaver(s) and pushed the plug (fly vial plug) down to force the living flies to encounter the
cadaver(s) (Figure 4.3B). For the first several weeks, I was unable to infect any laboratory-reared flies using this approach.

Figure 4.3. Initial approach to passaging in vivo E. muscae CNE1 infection to laboratory-reared flies. A) Representative sampling vial from fendel population. Flies were aspirated daily around 8 am, housed on Koshland diet and monitored nightly approximately 2 hours after sunset for cadavers. B) Representative exposure vial. Available cadavers were placed on Koshland diet to expose healthy flies. Healthy flies were confined near cadaver to improve chances of infection. C) Number of fendel flies that died from E. muscae infection versus the number of days these flies were housed on Koshland diet.

Initially it was unclear of which aspect of the fendel that I was failing to recreate to propagate the infection. I first hypothesized that the infection wasn’t spreading because I was incubating my flies under temperature and humidity conditions that weren’t favorable to the fungus. Previous work had demonstrated that mild temperature and high humidity are correlated with increased observance of fungal infections (145). From using environmental data collectors, I knew that it was colder and more humid at night in the fendel than in the laboratory, so I transitioned to working with infected flies exclusively at my house to allow cadavers to experience these ideal nighttime conditions. After a few weeks, it became apparent that this wasn’t solving the problem, so I considered other ways in which I could set up exposure vials in a way that more closely mimicked the fendel.

I was aware that the Koshland fly diet has a small amount of the preservative and antifungal compound tegosept, but had initially not thought that this would be problematic since infected wild flies still died of infection after being housed on this diet for up to eight days (Figure 4.3C). However, I decided that perhaps the small amount of antifungal was, in fact, preventing the establishment of the fungal infection.

I continued to sample wild flies and house them on Koshland diet, but when a cadaver was found I instead placed it on a banana-based diet (no preservatives or antifungals) then added healthy flies, a la Figure 4.3B. The first wild-type, laboratory-reared fly strain I was able to infect multiple individuals was CantonS WF and so this became the strain of choice for passaging CNE1 in vivo infection. While I initially reared CantonS WF on banana-based diet prior to exposure, I found that this was unnecessary and returned to rearing on Koshland diet. Under ambient conditions, CantonS WF flies died four to seven
days after exposure to cadavers. After I established a stable infection in CantonS WF flies using wild cadavers (using approximately 15 wild cadavers over two weeks), I continued the infection using only CantonS WF cadavers.

Optimization of *E. muscae* CNE1 infection of *D. melanogaster* in the laboratory

In order to work with the fungus during typical laboratory hours, I reared CantonS WF flies on a shifted light cycle (12:12 L:D, “sunset” at 10 am) and exposed these flies to the fungus on this light cycle. This rearing regime resulted in flies dying within a few hours of 10 am, allowing me to collect cadavers at noon. In order to determine the time window in which I could collect cadavers and place them in new vials before they began sporulating, I imaged fresh cadavers every minute for 21 hours to generate a timelapse video of sporulation (Figure 4.4). This revealed that sporulation starts approximately five hours after artificial sunset (under ambient conditions), so I determined that we would need to collect cadavers and place them with healthy flies prior to this time to ensure considerable exposure to shooting spores.

![Timelapse video of sporulation](https://youtu.be/AvgNe0QH84c)

**Figure 4.4. Fungal growth timelapse of fresh *E. muscae* CNE1 cadavers.** Fresh cadavers with no evidence of fungal growth through the cuticle were collected 2.25 hours after sunset (elapsed time 0:00) and an image was taken every minute for 21 hours. Time since the start of the video is shown at bottom right. At 2:39 the first ejected primary conidium is observed (arrow). At 3:37 the cadavers reach their maximum volume in the course of this video; many ejected primary conidia are visible in foreground (out of focus). As timelapse continues, cadavers shrivel until resembling those in Fig 3.1A, B. Video available at [https://youtu.be/AvgNe0QH84c](https://youtu.be/AvgNe0QH84c).

When I initially began propagating the infection in our laboratory, I did not have access to CO2 anesthesia. As such, I developed an anesthesia independent method for collecting cadavers and starting exposure vials (see Methods). Through this method, I was
able to consistently infect flies and could therefore collect infectious cadavers on a daily basis to propagate the in vivo infection. I next began gently tweaking this protocol to determine ways in which I could improve cadaver yield for each vial and reduce vial to vial variability (Figure 4.5).

Infection rate was observed to drop off significantly after the flies had aged seven days (Figure 4.5A). I failed to observe significant differences in the rate of infection between young males and females (Figure 4.5B). Temperature has an impact on which day flies would die; flies kept at 18°C died later than flies kept at 21°C (Figure 4.5C). Temperatures below 18°C were not intentionally tried, but the incubator malfunctioned (as did the heating in the building) and we observed poor sporulation of cadavers and infection of flies during this time (data not shown). Cold anesthesia of exposed flies in the first 48 hours of infection reduces overall mortality dramatically. Exposure to tegosept (the antifungal present in the Koshland diet) starting 48 hours after exposure was found to significantly reduce mortality, even if that exposure only lasted 2 hours (Figure 4.5E).

Sporulation of cadavers is visibly enhanced by housing vials at ~100% humidity for the first 24 hours before incubating at 60% humidity, but beyond that is very difficult to predict or control. To distribute cadavers (and, as a result, spores) evenly throughout the vial, I began embedding cadavers in an agar matrix. This change also permitted us to more easily control the confinement space between vials as well as assess how well sporulation occurred. In an effort to reliably infect 100% of flies in a vial, we tried reducing the number of animals and decreasing the spacing (Figure 4.5D). Infection rates improved, but we never consistently reached 100% mortality. Finally, when CO2 became available, I determined that anesthetizing flies over the course of infection did not significantly hamper infection rate (Figure 4.5F).

It should also be noted that it is possible to kill flies prematurely (i.e. without producing infectious cadavers) by exposing them to too many spores. I have observed several instances where flies, especially smaller ones, have died and are heavily melanized, indicating invasion by multiple spores (data not shown). It’s possible that E. muscae CNE1 is producing some sort of resting spore in these animals, but for the purposes of in vivo propagation, premature death is not desirable.
Figure 4.5. Optimization of *in vivo* *E. muscae* infection of CantonS WF *D. melanogaster* under laboratory conditions. Percentage of infected cadavers at 96 hours (4 days), 120 hours (5 days) or 144 hours (6 days) after exposure to *E. muscae* CNE1 varying A) age; B) sex; C) temperature and humidity; D) number per vial; or subjecting to E) cold anesthesia or diet containing 0.09% tegosept for 2 or 120 hours after the first 48 hours of infection; F) CO2 anesthesia for 20 minutes at the indicated post-exposure time. 18C > 21C indicates
that vials were incubated 24 hours at 18C, 100% humidity then transferred to 21C, ~60% humidity. If no percent is indicated, then humidity was ~60%. All vials used 4-6 cadavers to establish infection and were set up using the anesthesia-independent protocol. Replicate vials for each condition are shown above the graph. For all panels except D, each vial contained 50 flies. In panel D, each vial contained the number of flies indicated (for example, all “18C – 10” vials contained 10 flies each).

Combining all of these observations, I arrived at a set of conditions for propagating the infection among CantonS WF flies with a high percentage of mortality (summarized in Figure 4.6, see Methods for details). This protocol allows for reliable maintenance of the in vivo infection of CantonS WF flies, thus allowing us to study any desired phenomenon within the D. melanogaster-E. muscae system, and is the basis of all experiments reported in subsequent chapters.

**Fig 4.6. Summary of method for in vivo propagation of E. muscae in CantonS WF D. melanogaster.** Briefly, 50 healthy, young (eclosed within the last 24 hours) CantonS flies of mixed sex are confined within 2 cm of a circle of 6 cadavers embedded headfirst in sucrose agar. Vials are incubated 24 hours at 18C and ~100% humidity. On Day 1 (24 hours since exposure) the flies are raised to the top of the vial and incubation continues for the next 24 hours at 21C with ~60% humidity. On Day 2 (48 hours since exposure), flies are moved away from cadavers and onto GB+ diet. From Days 3-7 (96 to 168 hours since exposure), vials are monitored 2 hours following “sundown” to collect fresh cadavers. These cadavers are then used to begin new vials (Day 0). See Materials and Methods for full more details or Appendix-I for laboratory protocol.
Discussion

Here, I have described the process by which I brought a wild, Drosophila-infecting isolate of Entomophthora muscae, E. muscae CNE1, into the laboratory and the methods by which it is reliably cultured both in vivo and in vitro. Through these efforts, we are now able to study the bizarre biology of E. muscae in the context of its natural host, D. melanogaster.

Other occurrences of E. muscae infection in wild Drosophila

While, to my knowledge, I am alone in studying a Drosophila-infecting strain of E. muscae in the laboratory, I am just one of many to observe E. muscae infection in wild Drosophila. In 1927, Goldstein reported finding Empusa muscae (now E. muscae)-infected cadavers of D. repleta as well as Musca domestica at Columbia University in New York state, stating that an epidemic of E. muscae had been observed for the previous four years in this location (146). In 1969, Turian and Wuest reported observing E. muscae-infected cadavers of wild D. hydei in a rotting fruit bait in Geneva, Switzerland (147). These authors made a brief attempt at exposing other wild flies to a D. hydei cadaver and were successful in infecting one D. hydei individual. Curiously, despite declaring a need to determine host range for this fungal strain, I can find no evidence that this observation was ever followed up. The most recent instance that I have seen in the literature from Keller in 2002 reported morphological parameters (e.g. number of nuclei, size of nuclei etc. in fungal structures) of E. muscae (putatively identified as E. ferdinandii, a member of the E. muscae species complex) found on Drosophila spp in Switzerland (148). Outside of scientific literature, there is photographic evidence that E. muscae infections have been observed in wild fruit flies by others, most notably an individual that lives in Oakland, CA (David Tighe, iNaturalist.com).

Importantly, my ability to recognize the signs of an E. muscae infection occurring in the fendel is owed to a serendipitous discovery in our laboratory. As it happened, wild Drosophila were collected by Quan and Schiabor during a field visit to a field site in Santa Cruz County in October 2014 to establish isofemale lines of wild D. melanogaster. A few days after collection, one of the females was found dead, stuck to the cap of her vial, wings up and mycosed (Figure S4.1). Bronski visually identified the fly as a victim of Entomophthora muscae, which was later confirmed by PCR-based genotyping of the ITS and LSU loci. This was informative as I had not previously seen E. muscae infected flies and was unfamiliar with E. muscae morphology. It was the following summer that I encountered this morphology in the fendel and knew what it meant.

Interestingly, our laboratory has continued to encounter instances of E. muscae-infect wild Drosophila. In September 2015, I visited the field site from which the first observed Entomophthora-infected Drosophila had originated and observed several spent (i.e. already mycosed and sporulated) cadavers. Then again, in October 2016, Quan collected wild Drosophila from a field site in Sonoma County and one died from infection by E. muscae the following day. It’s worth noting that all of the Northern California specimens recovered and typed by our laboratory, including the fendel isolate E. muscae CNE1, are identical at the LSU and ITS loci (Figure S4.2). These observations are consistent with the existence of a strain (or possibly species) of Drosophila-infecting E. muscae is widely distributed over the San Francisco Bay Area.

Given that all of these specimens came from locations that were densely populated with wild Drosophila, I suspect that E. muscae infection is most frequently observed when
lots of flies are present. This would make sense since *E. muscae* relies on fly hosts to propagate and requires specific conditions and timing in order to successfully infect a new host. Certainly, much work needs to be done to understand the ecology of *E. muscae* and *Drosophila*. I hope that the publication of these methods and the experiments described in subsequent chapters will spark an interest to begin this work.

**E. muscae CNE1 is not a contamination risk for *Drosophila* labs**

The first question that sharing this discovery with my committee members was “Isn’t studying a fly pathogen in a fly laboratory dangerous?”. Happily, the answer to that question appears to be no. When following the anesthesia-independent *in vivo* propagation protocol, I was initially terrified that any escaped exposed flies would cause epizootic outbreaks in the fly room. Consequently, all of the handling of *E. muscae* CNE1 and exposed flies is performed in a different room on a different floor. Any consumables that come into contact with spores or an infected fly are autoclaved, fly tape and traps are set up within the *E. muscae* workspace to catch any potential infected escapees, and work surfaces are routinely wiped with 70% ethanol.

All evidence to date suggests that the risk of inadvertently infecting other fly stocks is minimal to none; I house a variety of uninfected lines in the same incubator with exposed cadavers and have never encountered *E. muscae* CNE1 infection where there isn’t meant to be. I’ve also performed several experiments where mock treatment controls are run alongside exposed vials. These experiments have used the same fly pads, brushes, forceps etc. and no contamination has been encountered. Spraying down surfaces with 70% ethanol between working with cadavers and healthy flies appears to be sufficient to prevent cross-contamination. That I have trained five individuals ranging from undergraduate to principal investigator to work with *E. muscae* CNE1 *in vivo* culture and am comfortable with them handling the infection speaks to my confidence that contamination is not an issue. Not only is propagation of the infection dependent on a large density of susceptible hosts (healthy adults) within range of a sporulating cadaver, cadavers are only infectious for the first several hours and are then inert. Most importantly, the food on which all fly labs house their stocks contains a low level of antifungal (0.09% tegosept) which is sufficient to prevent the passage of infection and to cure flies at early stages of infection, even if the fly is only transiently housed on it.

**Future techniques and improvements in *E. muscae* CNE1 laboratory culture**

While the *in vivo* infection robustly propagates using this protocol, it is time consuming and inconvenient to maintain the infection daily 365 days a year. Additionally, in its current state, it would be difficult to share the infection with a laboratory that’s more than a few hours’ drive from UC Berkeley. We have tried to freeze cadavers (as reported in (149, 150)) in the presence or absence of silica gel, but none of these animals were not infectious after just 24 hours at -20C. Presently, we are trying to develop a protocol to passage *in vitro* culture back into fruit flies using a house fly vector ([88](#)). House flies are large and easily injected with a standard syringe; previous trials using ARSEF *E. muscae sensu lato* strains #556 and #6918 were successful in establishing patent infection in house flies via *in vitro* injection, though infection was not spread to *D. melanogaster*, likely due to housing [88](#).

---

5 Vials started with already sporulated cadavers yield no infected animals; vials exposed to old spores yield no infected animals (data not shown).
these animals on diet containing tegosept. A recent attempt succeeded in inducing patent infection eight of roughly 120 house flies with the in vitro culture that’s been passaged since August 2015, but these animals were not infectious to D. melanogaster. It’s possible that this strain of E. muscae cannot infect house flies. I believe it’s more likely is that the fungus has lost its virulence after being in culture for long and that restarting an in vitro culture with fresh D. melanogaster cadavers will restore virulence. Ultimately, this will make the fungus more convenient to culture and share but will be essential to develop methods for genetic modulation as the fungus will need to be manipulated outside of the host.

Several additional issues are worth considering with respect to the in vivo infection. The in vivo propagation protocol herein presented reliably results in infected cadavers, but there is still significant variability between vial to vial in terms of numbers of flies that die of infection, the day that most of the flies die and the position of the flies upon death. As others have reported, there is a clear inverse relationship between the size of an exposed animal and the amount of time before it becomes an infected cadaver: the smaller the fly, the sooner it dies (85). Vial to vial variability is improved when just one person manages the infection, selects flies or similar sizes and uses anesthesia to manage vials. Notably, vials that are housed for the first 24 hours at 100% humidity, 21C and the remainder of infection at 21C yield higher cadaver rates than those using the method outlined here and more consistency as to when flies die (almost exclusively days 4-5). However, it is not ideal because flexibility with maintenance is lost – the window in which the infection can be lost narrows.

However, inter-vial variability cannot be eliminated even under the best conditions. This is likely explained by differences in sporulation efficiency by cadavers, which is hard to predict. For the purposes of future work with E. muscae, it would be ideal to establish a way to control how many conidia each fly comes into contact with. We briefly attempted to apply suspensions of spores suspended in an agricultural solvent (Silwet L-77) as has been described for Beauveria bassiana (151) to healthy flies, but wound up killing these animals by drowning. It’s possible that revisiting this technique could yield success.

Finally, it’s important to keep in mind that, as an active in vivo infection, the fungus is subject to changing over time. Though the hosts are always naïve to the fungus, the fungus has repeatedly seen genetically identical hosts and may be adapting to its new lifestyle as pampered parasite. For this reason, it’s crucial to be able not only to generate viable freezer stocks but to also develop a way to passage these stocks back into D. melanogaster should things start to drift.

Materials and Methods

Fly husbandry

Healthy wild-type, CantonS Wolbachia-free (WF) D. melanogaster were reared on Koshland diet (0.68% agar, 6.68% cornmeal, 2.7% yeast, 1.6% sucrose, 0.75% sodium tartrate tetrahydrate, 5.6 mM CaCl2, 8.2% molasses, 0.09% tegosept, 0.77% ethanol, 0.46% propionic acid) supplemented with activated dry yeast pellets (Red Star) at 21C on a 12:12 light:dark photoperiod. Any time the photoperiod needed to be adjusted, flies were reared from third instar or earlier on the desired photoperiod to ensure that emerging adults were properly entrained.
PCR genotyping

DNA was extracted from individual cadavers or 1.5 mL of in vitro culture using the QIAamp Micro Kit (QIAGEN) following the tissue protocol. These DNA preparations were used to amplify the desired sequences. Entomophthora-specific ITS primers (emITS: emITS-1 5’-TGTTAGAGAATGATGGCTGTTG-3’, emITS-4 5’-GCCTCTATGCGCTAATTGCTT-3’) or fungal-specific large subunit primers (LSU: LR3-1 5’-GGTCCGTGTATTCAAGAC-3’, LR0R-4 5’-GTACCCGCTGAACTTAAGC-3’) were used to genotype Entomophthora (James et al. 2006); cytochrome oxidase II primers (tLEU: 5’ ATGGCAGATTAGTGCAATGG 3’ and tLYS: 5’ GTACCCGCTGAACTTAAGC-3’) were used to genotype infected Drosophila hosts (Liu and Beckenbach 1992). Each reaction was performed using GoTaq 2x colorless mastermix (Promega) with 800 nM of each forward and reverse primer with the following thermocycling conditions: 95°C for 5 min followed by 35 iterations of 95°C for 30 seconds, 51°C for 30 seconds then 72°C for 1 min/kb then 72°C for an additional 10 minutes. Reactions were checked by gel electrophoresis in 1% agarose. Successful reactions were prepared for sequencing using ExoSap-IT (Affymetrix) per manufacturer’s instructions and submitted with each amplification primer for sequencing (ELIM, Barker sequencing facility). Assembly of forward and reverse reads was attempted to generate a consensus sequence using Seqman Pro (DNA Lasergene v.10). Sequences were searched against the BLAST NT database using blastn.

Isolating E. muscae CNE1 in vitro

To grow E. muscae CNE1 in vitro, first spores were collected using the ascending conidia collection method (i.e. by placing a fresh cadaver in the bottom of a sterile petri dish and allowing the cadaver to sporulate overnight) (144). The following morning, the lid of the dish was rinsed with 10 mL of Grace’s insect medium (1x) supplemented with L-glutamine, 3.33g/L lactalbumin hydrolysate, 3.33g/L yeastolate (ThermoFisher Scientific # 11605-094) containing 5% fetal bovine serum (Invitrogen) and transferred to a vented, treated T25 tissue culture flask (Corning # 10-126-10) using sterile technique. The culture was then incubated at room temperature in the dark until growth was apparent (non-homogenous white spheres floating in the medium). The culture was genotyped with emITS and LSU primers to confirm that it was E. muscae and was an identical match to the cadaver that produced the spores which started the culture (“Fendel mama”) at these loci. The culture was periodically examined at 100-400x on a compound microscope to confirm proper morphology and absence of contamination.

Cryopreservation of E. muscae in vitro culture

Cultures were preserved per (152). Briefly, 100% glycerol was gently added to turbid culture (Fig S4.3) to generate a final volume of 5 mL with final percent glycerol of 10, 15, 20 or 25. Cultures were gently rocked by hand until homogeneous then were incubated 42 hours at 4C without movement to “cure”. Cured cultures were aliquoted into 3-4x cryovials and placed in a Mr. Frosty freezing container (Nalgene) at room temperature. The container was then placed at -80C for at least 24 hours before testing stocks for viability. Cryopreservation of in vitro E. muscae stock was attempted twice before success on 7/3/16. Glycerol stocks of E. muscae were tested by warming at 37C until thawed. The
entire volume (1-1.5 mL) was then immediate transfer into 9 mL room temperature medium (1x supplemented Grace’s + 5% FBS) in a T25 culture flask. After 3 weeks of growth, cultures were densely grown in a morphology resembling a mycelial mat.

**Anesthesia-independent method for propagating *in vivo* *E. muscae* CNE1 infection**

All of the exposure vials were setup as follows: chunks of organic banana (with kimwipes to sop up excess moisture) were provided as a food source, not-yet-sporulating cadavers (anywhere between 1 and 10 depending on availability) were placed on the banana chunks and approximately 50 healthy flies were transferred onto the prepared vial by flipping (no anesthesia was used). The flag of the vial was pushed down to confine the flies within a few centimeters to improve the likelihood that they would encounter flying spores. Leaving the exposed flies with the spent cadavers was initially problematic as we were working without access to anesthesia or a microscope and had to identify new cadavers by naked eye. Additionally, the raw banana began to ferment and break down, leading to excess moisture which was prematurely killing some of our exposed flies. To avoid these issues, the exposed flies were transferred to a new banana/kimwipe vial after the first 48 hours. This was done by allowing the living flies to climb into an empty “holding” vial then flipping them onto the fresh vial. The flies were monitored daily for deaths; cadavers were removed after allowing healthy flies to climb into a “holding” vial and used to start new exposure vials.

**Anesthesia-dependent method for *in vivo* propagation of *E. muscae* CNE1 infection**

Cadavers are collected daily 2-4 hours after the end of the 12 hour light cycle from exposure vials that are between 96 and 168 hours (4 and 7 days) old. All flies that will die on this day should be dead and beginning to mycose by this time, making them obvious among the living flies. First, the position of the cadavers in the vial is noted (either on flag, side or medium) and recorded on a vial template worksheet (Appendix-III) then CO2 is used to anesthetize the living flies in exposure vials. Any cadavers that have fallen onto the CO2 pad are sorted out from the healthy flies and placed in a petri dish with a piece of Whatman paper wetted with DI water to fight static. All cadavers from the vial are removed and counted. Position of the cadavers wings are counted (either stuck up, folded down or not changed) and recorded. Any dead flies that do not appear to be mycosing (i.e. shriveled, wings not perturbed) are counted as “not infectious” and discarded. Molten cadaver embedding medium is preparing by microwaving solidified AS solution (1.5% agar, 10% sucrose) and poured into a clean 100 x 15 mm petri dish just enough to cover the bottom of the dish. As soon as the agar has set, six cadavers are embedded head first in a circle of diameter <2 cm with their wings on the outside of the circle. The wings are pressed into the agar to ensure they do not intercept any launched conidia. The agar is allowed to completely set before continuing. The cadaver circle is cut out from the set agar by using an empty, wide-mouth *Drosophila* vial (FlyStuff). The agar disc containing the fly circle is then transferred, cadaver-side up, into an empty, wide-mouth *Drosophila* vial. A ruler is used to mark 2 cm above the surface of the agar. Young (<24 hour old) CantonS flies reared on the same 12 hour light cycle on Koshland medium are collected from eclosing vials using CO2
anesthesia. Fifty healthy flies (i.e. not undersized, undamaged wings) are added to the vial container the agar disc with cadavers and tapped down until all flies are under the 2 cm mark. A Droso-plug (FlyStuff) is pushed into the vial such that its bottom is flush with the 2 cm mark. The vials are incubated for the first 24 hours at 18°C in a humid chamber (2L plastic beaker lined at the bottom with wetted paper towels or kimwipes and covered with foil), to encourage sporulation. After 24 hours, the Droso-plug is lifted to relieve the confinement of the flies and the vial is moved to a 21°C incubator (~60% humidity). After 48 hours, the exposed flies are transferred onto GB+ medium (40% organic banana [w/v], 2% agar, 0.3% propionic acid, Appendix-II) without anesthesia; incubation continues at 21°C. Cadavers are collected daily 2-4 hours after the end of the 12 hour light cycle from exposure vials that are between 96 and 168 hours (4 and 7 days) old. The process is repeated daily to ensure that the infection is reliably maintained⁷.

---

⁶ If an 18°C incubator is not available, the vials can be housed the first 24 hours in a humid chamber at 21°C. This will actually result in more flies dying from infection, but they will tend to all die 96-120 hours after exposure rather than being spread out between 96-168 hours.

⁷ The occasional day may be skipped (e.g. holidays) without problem, but skipping several consecutive days will result in unreliable cadaver quantities for experiments and could result in loss of the infection.
Supporting Information

Figure S4.1. *E. muscae*-infected female *D. melanogaster* collected from winery in Santa Cruz mountains. The female was found dead, stuck to the cap of her vial, wings up and mycosed in the mid-afternoon a few days after collection. Species identity of both host and fungus were confirmed by PCR genotyping (COII and ITS, LSU loci, respectively).

Figure S4.2. Ribosomal RNA loci for wild and ARSEF isolates. Alignment of A) eight ITS sequences; B) six LSU sequences. Neighbor-joining trees for C) alignment shown in A); D) alignment shown in B). Sequences were aligned using BLAST (somewhat similar sequences). The reference for alignments was the consensus sequence of the emlITS or LSU amplicons from first four cadavers sequenced from the fendel after *E. muscae* CNE1 discovery (labeled red in trees). Sample names: iv Fendel culture = *in vitro* fendel culture started with Fendel mama. ARSEF *E. schizophorae* and ARSEF *E. ferdinandii* are from *in vitro* stocks of the indicated species from the United States Department of Agriculture Agriculture Research Service at Cornell (stock numbers #556 and #6918, respectively).
Figure S4.3. *E. muscae* in vitro culture used for successful cryopreservation starting 7/1/16. This culture had been passaged once monthly since August 2015. Here, small “beads” of growth are observed, similar to what I have observed for *in vitro* culture of *E. schizophorae* and *E. ferdinandii* (ARSEF #556 and 6918, respectively) (data not shown).
Chapter 5: Characteristics of *E. muscae* CNE1 and initial experiments probing its effect on fly behavior

Abstract

*Entomophthora muscae* is a fungal pathogen that infect, manipulates the behavior of, and kills its dipteran hosts. Recently, I isolated a strain of *E. muscae* (CNE1) that naturally infects *Drosophila* and have developed methods to propagate it in laboratory-reared flies. Primary conidia of *E. muscae* CNE1 conidia are most similar to an isolate previously described from house flies. Through high-speed videography of conidia ejection and landing, I found that conidia are ejected extremely quickly (~21 mph) and land with their sticky halo already formed, supporting the spore canon mechanism of conidia ejection. I found that secondary conidia form from primary conidia under ambient conditions in part by transferring, rather than duplicating, DNA from primary to secondary conidia. Fruit flies that are infected with *E. muscae* CNE1 reliably exhibit the characteristic fungal-manipulated end of life behaviors of summiting, proboscis extension and wing raising. Preliminary evidence suggests that E49 neurons are not required for proboscis extension, and nanchung is not required for summiting. Similar to observations in house flies, fruit flies die in a gated fashion when provided light cues; when infected and incubated in darkness, flies die according to a random schedule. Consistent with this observation, infected flies lose their circadian rhythm approximately 48 hours prior to death when housed in complete darkness. Altogether, the work in this chapter highlights the utility of having a stable, *in vivo* *E. muscae* infection in a genetically-tractable host to study the basis of behavioral manipulation or other aspects of host-pathogen biology.

Introduction

The previous chapter described how *E. muscae* CNE1, a strain of *E. muscae* that naturally infects *Drosophila*, was discovered and brought into the laboratory for culturing both *in vivo* and *in vitro*. Here, I sought to take advantage of the *in vivo* infection to both more closely observe general features of the fungal life cycle as it propagates in its natural host *D. melanogaster* and begin addressing initial hypotheses regarding how the fungus might influence host behavior. On the fungal side, I wanted to use morphological characteristics of primary conidia to determine if a similar strain has been previously reported and I wanted to resolve controversy of primary spore ejection, which should be settled by knowing if spores change size after landing. On the fly side, I wanted to observe and describe the end of life behaviors and begin addressing a few simple hypotheses to assess how to study the behaviors induced by the fungus. In particular, I wanted to test a) if E49 neurons known to induce proboscis extension in response to a tastant stimulus are required for proboscis extension during the final suite of behaviors (153), b) if Nanchung, a non-specific ion channel involved in locomotion is required for summiting and c) if timing of death has a molecular basis and, if so, determine whether fly or fungus dictates this timing. The efforts taken to address this initial series of questions is described herein.
Results

Fungal phenotypes and lifecycle

Comparing *E. muscae* CNE1 to previous *E. muscae* isolates

At present, fly-infecting species of *Entomophthora* are grouped in a taxonomically ill-defined species complex. There is a dearth of molecular data for this group; species within the complex are still distinguished primarily by fungal morphology and insect host (80, 86, 154). To compare my strain with those previously described, I therefore performed a morphological characterization of primary conidia from *E. muscae* CNE1 with the help of rotation student Hayley MacCausland (Figure 5.1). These data are most consistent with what has been reported for *E. muscae* sensu strictu (primary conidia 27-31 µm x 20-24 µm with 15-20 nuclei with a diameter of 3.9-4.4 µm; (86)).

Figure 5.1 Morphology of *E. muscae* CNE1 primary conidia. *(Figure on next page)* A) Number of nuclei per conidium. B) Length and width of conidia. C) Diameter of nuclei. For each morphological feature (i.e. number of nuclei per conidium, width and length per conidium, diameter of nuclei), a series of primary conidia was collected from a single sporulating cadaver. At least 50 conidia (A, B) or 75 nuclei (C) were counted for each set. All data are plotted; averages for each feature for each set are given in boxes above data. Example images of conidia measured are shown to the right at 200x-400x magnification imaged on a Nikon 80i microscope (A, B – Stained with Hoechst to visualize nuclei, C – Stained with lactophenol cotton blue to define conidial boundaries).
Mechanism of primary conidia discharge by sporulating cadavers

While collecting conidia for morphological study, I became curious as to how the primary conidia are actually discharged from cadavers. My experience watching sporulating cadavers indicated they were ejected too quickly to observe standard video camera or by eye (24 and 60 frames per second (fps), respectively). In order to observe the spore ejection process and measure approximate discharge speed of primary conidia, I therefore employed a high-speed camera, capable of capturing video up to 100,000 frames per second. Initial attempts at 10,000 fps were too slow to capture the spore before it left the frame (Figure 5.2A). Subsequent recordings at 54,000 fps and above were sufficient to calculate a rough estimate of initial ejection velocity of 21 miles per hour (9.4 m/s) (Figure 5.2B). These speeds are comparable to those observed in coprophilous fungi, and are among the fastest observed velocities of organisms relative to their size known in the natural world (155).

Figure 5.2 High-speed videography of primary conidia ejection. Image series taken at A) 10,000 fps and B) 73,000 fps. Both image series were taken at 50x magnification. Images are identically scaled for comparison.

There is a long-standing disagreement over the mechanism by which spores of E. muscae are so forcibly expelled. At its core, this disagreement centers on the issue of what material comprises the sticky “halo” that is observed with landed spores. Reports indicate that spores form on the tips of conidiophores as cytoplasm the flows into the spore, inflating it like a balloon (73). I have also observed this in taking a time lapse image series
of forming spores (Figure 5.3). From there, some favor a model in which the septum between conidiophore and spore pinches off and the continuing flow of cytoplasm leads to a buildup in pressure that launches the spore like a bottle rocket (72). According to this model, the halo arises from co-ejection of cytoplasm with the spore. Alternatively, a structural study of spores pre and post launch has been reported that no cytoplasm sits behind the closed septum and that the material comprising the “halo” is mucilaginous material that is released when the outermost cell wall ruptures upon landing (73).

![Figure 5.3. Time lapse of conidophore growth and primary conidia formation in E. muscae CNE1-infected fruit fly cadaver.](image)

To determine if the halo arrives with the launched spore or is formed after landing, I used the same high speed camera setup to capture video of spores landing on a petri dish. In so doing, I observed that the conidium does not change diameter immediately after landing (Figure 5.4). This observation invalidates the model where the cell wall ruptures after landing but is consistent with co-ejection of cytoplasm (i.e. the fungal canon mechanism).
Formation of secondary conidia

It has been reported that when *E. muscae* primary conidia (those launched from the infected cadaver) launch and fail to hit a target, they can sporulate again and give rise to secondary conidia who then launch via a method distinct from primary spore discharge (156). If the secondary conidia also fail to hit a target, the spore can continue to sporulate and give rise to smaller and smaller conidia until the spore runs out of resources or desiccates. For strains of *E. muscae* it has been reported that secondary conidia are actually more virulent than primary conidia (76). Indeed, I have also observed that primary conidia who fail to hit their target (i.e. hit a part of the fly which does not allow access to the hemolymph, wing) will produce secondary conidia (Figure 5.5A). Through vital nuclear staining of primary conidia, I have observed that DNA appears to be transferred to secondaries rather than duplicated, leaving the primaries behind as empty shells (Figure 5.5B). This is consistent with previous reports demonstrating “ghost” conidia left after secondary sporulation (157). I have not tested if these conidia are more infectious than their primary forebears and have not examined if primary conidia that do land on fly cuticle still produce secondary spores as reported in (78).
Figure 5.5. Formation of secondary conidia. A) Time lapse of primaries generating secondaries after landing on wing of near fly. One image was taken each minute for four hours starting after primary sporulation was well underway (5 hours after sunset). Secondary launches between 224 and 225 minutes leaving behind a primary ghost. B) Snapshots of primary/secondary conidia complexes in various stages of formation. Nuclear material is shuttled from the primary to the secondary before the primary is launched. Spent primaries have no observable nuclear material (bottom panel, left conidium; conidium to right is in process of transferring nuclear material to secondary, for comparison).

Fly phenotypes and behavior

Behavior of dying flies

In optimizing the in vivo infection, I repeatedly observed the final moments of life of moribund flies. Flies appear phenotypically normal right up until a few hours before sunset, although activity experiments show that infected flies become less active than uninfected controls starting 24 hours before death (Fig 5.6A, B). On the last day of life, the first visible signs of trouble (in the context of a fly vial) are when a fly begins to move slowly and shakily. Typically, such flies are in the process of climbing the side of the vial, though not all flies are observed to successfully summit in K-resin vials (Figure 5.6C.). When a wooden dowel is supplied as a more natural substrate for climbing, more flies are observed to climb (Figure 5.6D, E).

When a slowly moving, tremulous fly is removed from its vial and placed on the benchtop, it is apparent that the fly can no longer fly. The fly may attempt to summit if a vertical surface is within reach (e.g. petri dish). Soon thereafter, the fly will lose its ability to walk/climb and will remain immobile. At this point, already the fly will begin extending its
proboscis and become adhered to the substrate on which it stands (Figure 4.5F). The legs will continue to spasm and the abdomen may thrust back and forth as the fly struggles with its proboscis extended. This can continue for several minutes (up to an hour) before the wings begin to lift up and away from the dorsal abdomen. The process of wing lifting observed in real time resembles a balloon being blown up in fits and starts (Fig 5.6G). From start to finish I have observed the wing lifting process to take around 15 minutes. The fly's legs and antennae may continue to twitch after the wings have lifted but will soon after (within a few minutes) stop. The fly is now dead.

An interesting exception to the standard end of life phenotype: while most flies die with their wings extended (Fig 5.6H), some flies are found with their wings down on their dorsal abdomen, resembling the floppy ears of a rabbit (Fig 5.6I). I have found that by applying pressure to the thoraxes of a living fly one can “toggle” their wings from the downward to the more commonly observed raised position, suggesting a common mechanism underlying how both wing positions are achieved.

**Figure 5.6.** End of life behaviors of *E. muscae* CNE1-infected flies. (Figure on next page). A) Representative fly in late stage of infection (Day 4, 96 hours after exposure) by *E. muscae* CNE1 (infection status confirmed by RNAseq) B) Cumulative average activity for uninfected flies (black) or *E. muscae* CNE1 infected flies that died on Day 4 (blue) or Day 5 (red). Activity of 96 flies was measured using the *Drosophila* Activity Monitor (Trikinetics), 24 of which died of *E. muscae* infection on Day 4 (96 hours), 14 on Day 5 (120 hours) and 48 of which were did not die (i.e. uninfected controls). Standard deviation for each data point is indicated by gray error bar. Yellow shading indicates daily 12 hour intervals where lights were on. C) Typical exposure vial, 2 hours after sunset on Day 5 (120 hours after exposure). Infected cadavers are observed adhered to flug, sides of vial and media (black chevrons). D) Location of infected cadavers in exposure vials with and without dowels (Dowel+ N=24 vials of 50 flies each, 834 cadavers, Dowel- N=18 vials of 50 flies each, 632 flies) E) Representative exposure vial with cadavers, 4 hours after sunset on Day 5 (~120 hours after exposure). F) Immobile *E. muscae* CNE1-infected fly with proboscis extended and adhered to substrate, shortly before death. Video is available at https://youtu.be/YgH080NnEhA. G) Frames of video showing process of wing extension. Wings did not move after 7 min 17 sec. Video is available at https://youtu.be/lkHWt3XY59I. H) Cadaver with most common wing position, wings up almost 90 degrees to body axis. I) Cadaver with alternative wing position (wings down rather than up).
Evidence that E49 neurons are not involved in proboscis extension of moribund flies

During the last moments of a life, an *E. muscae* CNE1-infected fly will extend its proboscis and adhere to the surface on which it stands. The mechanism by which fungal infection induces proboscis extension has been unclear. Proposed hypotheses include mechanical force (the fly is so swollen with fungal material that it can no longer keep the proboscis retracted, (69)) and manipulation of neurons involved in extending the proboscis. My observations of proboscis extension are not consistent with a model in which the fly is simply forced to extend its proboscis by virtue of being so swollen with fungus. While the proboscis is generally extended and in contact with the substrate while the fly is in the process of dying (i.e. before and during wing raising), once the fly is dead, the proboscis is seen to retract to varying degrees. This would suggest that muscles are actively used in keeping the proboscis extended as the fly is dying as the fungal titer inside the fly does not drop the instant a fly dies (i.e. internal pressure is not relieved upon death) and there would likely be neurons underlying this active muscle utilization.

The only neurons known to be involved in proboscis extension at the time were those labeled by E49-GAL4 (153). These neurons were found to be necessary for proboscis extension in response to tastants. I decided to ask if these neurons were also necessary for proboscis extension behavior in dying *E. muscae* CNE1-infected flies. As many flies exhibit this behavior concurrently, making it hard to score in real time, I decided that I would count position where flies had died in a vial to inform whether or not they had extended their proboscis. I reasoned that if a fly could not extend its proboscis, it should not be able to adhere itself to a raised surface (i.e. side of vial or flug) upon death. Therefore, if knocking out these neurons did prevent proboscis extension, the flies would all be found on the medium. If they were not necessary, flies would be found in elevated positions. To this end, I performed reciprocal crosses for each of the E49-Gal4 line (in Gal80(ts) background) and E-49 single neuron line to UAS-TNT, exposed the appropriate progeny to *E. muscae* and counted where cadavers were found (flug, side or medium) four to seven days later (Figure 5.7A).

There was significant variation in where cadavers were found from vial to vial, but all treatments (except non-heats-shocked A) yielded elevated cadavers. Upon closer inspection, I found that elevated cadavers were not always adhered via their proboscis (e.g. Figure 5.7B). Some flies were observed to be attached to the side of vials via legs stuck in fecal spots or droplets of food; some flies stuck to the flug were observed to have their proboscis extended but not in contact with the flug surface and instead may be attached via their tarsal claws like Velcro. I had previously observed this when providing flies with a dowel for summiting (flies appeared to be held in place by “hugging” the dowel rather than proboscis-mediated adhesion to the dowel) and avoided using dowels in these experimental vials for that purpose. I went through my experimental and control vials and looked for evidence of any adhesion via the proboscis. Indeed, all of my experimental vials contained at least some animals that were undoubtedly stuck via the proboscis, which would suggest that the E49 neurons are not required for the *E. muscae* CNE1-induced proboscis extension phenotype. I observed that in wild-type flies and in my control vials, elevated cadavers are not always stuck via their proboscis, highlighting that these alternative adhesion methods were not unique to my experimental animals. Still, I am reluctant to completely rule out the involvement of E49 neurons and would advocate...
repeating this experiment using a different metric, which will be addressed in the discussion.

**Figure 5.7. Preliminary data: position of cadavers with or without functional E49 or E49SCMN neurons.** A) Location of cadavers in experimental vials (found on flug (F), side (S) or medium (M)). The first four groups contain animals with all E49 neurons (A*,B*) or just E49-SCMN neurons (C,D) knocked out via expression of tetanus toxin. Remaining groups are controls. Asterisk (*) indicates that flies were heat shocked for 35 min at 37°C at the pupal stage. Genotypes and treatments are as follows: A = E49-GAL4; tub-GAL80(ts) [male] x UAS-TNT; UAS-mcd8GFP; TM2/TM6B [female]. B = reciprocal cross of A, but UAS parent was heterozygous for UAS chromosome, balanced by CyO balancer; only Cy+ animals were assayed for B* group. C = E49-SCMN-GAL4 [male] x UAS-TNT; UAS-mcd8GFP; TM2/TM6B [female]. D = Reciprocal cross of C. B (Cy-)* are B progeny selected with the CyO balancer instead of the UAS chromosome. E49 = E49-GAL4; tub-GAL80(ts); E49-SCMN = E49-SCMN-GAL4; UAS = UAS-TNT/CyO; UAS-mcd8GFP; TM2/TM6B. Number of cadavers counted for each treatment: A* = 85, B* = 109, C = 88, D = 68, A = 10, B = 38, B (Cy-)* = 34, E49 = 69, E49-SCMN = 57, UAS = 85. At least two vials of each treatment were counted except for A and B (only 1 vial each). B) Example of flies adhered to flug without using proboscis.

**Nanchung is not necessary for moribund summiting behavior**

As a first experiment thinking about the molecular basis of summiting behavior, I considered if Nanchung is needed for summiting behavior. Nanchung is a member of the vanniloid transient receptor channel (TRPV) family involved in a variety of *Drosophila* sensory processes including walking behavior and negative geotaxis (158, 159). Flies with non-functional nanchung (nan[36a]) are observed to move less and be less coordinated than WT flies (Nick Jourjine, personal communication). I wondered if nanchung knockout flies would not summit in response to fungal infection.

To address this question, I infected nanchung knockout flies and w- flies (background line in which the nanchung knockout was generated) to *E. muscae* CNE1 then provided the infected flies with dowels for summiting. I opted to use dowels for this experiment because summiting is a more robust phenotype when dowels are present. Adhesion of cadavers to the vial, flug, medium or dowel is not very robust, and living flies
within these vials obviated the ability to simply remove the dowel and count; I didn’t think I would get an accurate measurement of cadaver locations after removing the living flies via CO2 anesthesia. Instead, I opted to pre-mark the dowel with 1 cm marks and record the location of cadavers relative to these marks as a way of measuring summitting height before I proceeded to manipulate the vial and sort living flies from cadavers (see Figure 5.8B as example). It quickly became clear that nanchung did not inhibit summitting of moribund flies (Figure 5.8A). Curious to see if summitting height between vials had any consistency, I then measured the height of cadavers in vials of WT, *E. muscae* CNE1-infected flies also provided with marked dowels. Though there is large variation from day to day and between vial to vial in where the flies wind up on the dowel, using my simplified metric I found a surprising consistency in summitting height averaged over all of these vials (Figure 5.8A). The position of cadavers from both the nanchung knockouts (nan-) and nanchung knockout background vials (w-) were also consistent with this general trend.

![Figure 5.8](image)

**Figure 5.8. Weighted average summitting height of nanchung knockout, nanchung background control and wild-type flies.** A) Flies of indicated genotypes were exposed to *E. muscae* CNE1 and placed in vials with wooden dowels that had been pre-marked into 6, 1 cm sections. The sections containing the lowest point of each cadaver on the dowel were recorded. Cadavers found in each of the scored sections (lowest = 1, highest = 6) were summed for each vial and the weighted average position of cadavers were calculated. Each data point represents all cadavers from one vial of 50 *E. muscae* CNE1-exposed flies. Total cadavers for each treatment were 27 (nan[36a]), 40 (w-) and 632 (WT). B) Example nanchung knockout (nan-) vial with scored dowel.

**Investigating the molecular basis of timing of death from *E. muscae* CNE1 infection**

One of the most striking features of *E. muscae* infection is that it kills hosts in a gated fashion: infected animals only die around sunset (68). I hypothesized that the timing of
death was under control of a molecular clock (also proposed by (68)) and wanted to determine, first, if this were true, and second, to whom the clock belonged, fly or fungus.

To this end, I designed an experimental setup using *Drosophila* activity monitors (Trikinetics) to measure the time infected flies stopped moving, reasoning that this would be a reliable proxy of death. Briefly, each DAM consists of a plate with 32 openings in which small tubes can be inserted (Figure 5.S1). Each tube contains a single fly with a food source (agar sucrose) on one end and a cotton ball on the other (to provide air flow). When the DAM is connected to a computer, infrared beams pass vertically through the plate, one for each channel. Every time a fly walks through the beam and disrupts it, this event is counted and output to the computer in bins of a specified size (e.g. 30 seconds). In this way, the movement of 32 individual flies can be monitored over the course of several days. Living flies frequently break the beam; the last time the beam was broken for a given channel can serve as a proxy for the last time that fly was seen alive.

To test this setup and confirm that I could detect fly death accurately based on DAM signals, I exposed wild-type CantonS WF flies with *E. muscae* CNE1, loaded these animals into the DAMs and recorded from these DAMs for a week on a 12:12 light:dark photoperiod. Concurrently, I visually checked each fly daily after “sundown” to see which flies had died of patent *E. muscae* infection. At the end of the experiment, I determined the time of last movement for each fly and compared the time of death as determined by DAM signal to the time of death determined by visual inspection. After minor adjustments\(^8\), all of the times of last movement determined by the DAM data were concordant with my manual observations of when flies had died, indicating that the DAM data were reliable in calling time of last movement (Figure 5.9A).

Next, I tested if CantonS WF flies reared on a 12:12 light cycle would still die in a gated fashion after exposure to *E. muscae* CNE1 (Figure 5.9B, 5.9G). Infected flies continued to die according to their entrained light cycle (i.e. non-randomly), however timing of death became more diffuse than when light cues were provided. Additionally, fewer exposed flies died in the absence of light cues and those that did die later. Following this initial promising result, I continued to perform a series of experiments where I varied the entrainment cycle of the exposed flies or the cadavers (i.e. flies were entrained on a light cycle 6 hours ahead of the “default” light cycle), or the genetic background of the exposed flies or cadavers (i.e. used flies with a non-functional clock gene, Clk[Ar]. These flies are arrhythmic under constant darkness). Generally, I observed timing of death that could not be conclusively attributed to exposed fly or cadaver entrainment and genetic background (Figure 5.9C, D, E, F, H). I also noticed that, under free-running conditions, flies lose their sense of time 48 hours before death (Figure 5.10). This observation does not narrow down whether timing of death is based on a molecular clock or who might control that clock as it is consistent with timing of death being dictated by a fly clock (which appears to shut down nearing the end of infection in the absence of light cues), a fungal clock that isn’t functional until sometime after infection but before death or no clock but rather dependence on proximal environmental cues.

---

\(^8\) Initially there were a few disagreements between manual calls and DAM calls. These were due to aberrant beam breaks (i.e. beam breaks occurring several hours after the last movement of the fly) which were easily distinguishable from actual beam breaks by simply plotting breaks versus time or simply indicating that in order for a beam break to be called as the final time of movement, it had to occur within 24 hours of a previous break. Fluorescent lighting is capable of inducing erroneous beam breaks in the DAM monitors, so these breaks were easily explained.
Fig 5.9. Last movements of *E. muscae* CNE1-infected flies of varying photoperiod entrainments and genetic backgrounds versus movement of uninfected flies. Conditions under which flies were exposed and monitored is given above each plot (i.e. 12:12 light:dark or constant darkness). Genetic backgrounds and photoperiods under which exposed flies (“Exposed”) and cadavers used to expose flies (“Cadavers”) were reared are given below lighting conditions. Flies reared on a 12:12 photoperiod that ran six hours ahead of default 12:12 photoperiod are denoted by “(-6 hr)”. For each plot, the
average activity of uninfected flies of the same entrainment and genetic background of exposed flies is shown in teal binned to either 15 minute or 30 minute intervals. Lighting conditions (A, solid line) or entrainments (B-H, dotted or dashed lines) are shown with activity data. When two different photoperiod entrainments were used, orange dashed lines indicate the photo entrainment of cadavers (and therefore fungus); red dotted lines indicate photo entrainment of exposed flies. Each green circle indicates the time of last movement for one *E. muscae* CNE1-infected fly under the indicated experimental conditions. The number of exposed flies and infected cadavers for each experiment are as follows (written as #infected (#exposed)) A) 39 (48); B) 35 (48); C) 15 (32); D) 38 (80); E) 26 (64); F) 18 (64); G) 11 (64); H) 13 (80).

Figure 5.10. Infected flies lose circadian rhythm 48 hours before death when housed in complete darkness. A) Average movement of flies under 12:12 light:dark conditions that were observed to die on day 4 (Day 4), day 5 (Day 5) or were never infected with *E. muscae* CNE1 (Control). B) Average movement of flies entrained on same light cycle as A, but housed in constant darkness. Flies observed to die on day 5 (Day 5), day 6 (Day 6) or day 7 (Day 7) are shown in color; average movement of flies not exposed to *E. muscae* CNE1 is shown in black (Control).

Discussion

The descriptions of fungal morphology and differentiation shown in this chapter confirms and expands on observations made of *E. muscae* isolates infecting other Muscoidea. Here, I’ll briefly recap the implications of these fungal observations.

*E. muscae* CNE1 morphology

By measuring morphological traits of primary conidia, I’ve determined that the most closely related *E. muscae* strain reported to *E. muscae* CNE1 is *E. muscae* sensu strictu. Curiously, this strain is predominantly reported in *Musca domestica*, not *Drosophila* (86). Much more work is needed to both move *E. muscae* into the molecular era and to understand this organism’s ecology before we can draw conclusions about the relatedness of *E. muscae* CNE1 to previously described strains and species within the *E. muscae* complex.
E. muscae CNE1 imaging and videography

Much of the E. muscae literature originates from a time when photographic technology was not easily available as a means of capturing the dynamics of E. muscae biology. Here, I've presented videos of many of the processes previously described in the literature. Specifically, I have confirmed observations made by Cohn (1855) with regards to conidiophore formation. Additionally, because I have an in vivo infection actively running, I am able to capture infected flies at any point between exposure and death, allowing me to resolve some issues that could not be settled by field observations. Using high-speed videography, I found that conidia possess a “halo” upon landing which supports the spore canon model of conidial ejection (72). Additionally, I have used time lapse imaging to observe the formation of mature conidiophores and connect the events that lead to conidial formation, ejection and landing in a way that has not been previously reported in the E. muscae literature.

Initial behavior experiments: considerations and next steps

The work in this chapter has only skimmed the surface with regards to examining the molecular bases underlying various E. muscae-induced behaviors. In performing these pilot experiments in fruit flies, I hope to have illustrated that the D. melanogaster-E. muscae system presents a host of opportunities for studying the mechanisms of behavior.

Proboscis extension

The E49 neuron knockout experiments do not support the hypothesis that these neurons are necessary for the end-of-life proboscis extension in moribund flies. However, there are caveats to this experiment that should be made clear. First, though the expected phenotypes were observed in the crosses that provided the animals for these experiments, no orthogonal method was used to confirm that these neurons were, in fact, knocked out. However, I did notice that experimental flies spent more time on food attempting to extend their proboscis (they exhibited incomplete extension). This phenotype was apparent in control sibling Gal80-ts crosses that were not heat-shocked (i.e. should have had intact E49 neurons); this could have been due to rearing the flies at 21C instead of the recommended 18C and therefore permitting some leaky expression of TNT.

More work is needed to conclusively determine if end-of-life proboscis extension is due to neuronal manipulation or mechanical pressure. As my experiment showed, the metric for proboscis extension should not be vertical adherence to a substrate (as attachment via the proboscis is not the exclusive means by which this is achieved), but rather assessment of individual dying flies for this behavior. In order to observe this, flies will need to be separated prior to the initiation of end-of-life behaviors, confined and observed using a recording device (since the timing of proboscis extension is fairly synchronous in dying flies). Scoring the proboscis extension of these animals should inform whether or not a given set of neurons is required for this response in terminally-ill E. muscae CNE1 infected flies. Additionally, these experiments should consider not only E49 neurons but also those recently identified in (160).

Of additional interest is the matter of proboscis adhesion. Within the E. muscae house fly literature, it has been acknowledged that flies are adhered via their proboscis to
the substrate on which they summit, but the means of adhesion have not been agreed upon. One study declared that fungal structures termed “holdfasts” were responsible for adhesion (67). A nearly contemporaneous study observed hyphae emerging from the proboscis of some, but not all, cadavers, and proposed that dried up food rather than fungal growth could be responsible for adhesion (69). Consistent with (67), I have observed fungal material emanating from the proboscis while the fly is dying (Figure 5.S3), but I have also observed that the flies don’t necessarily remain elevated via their proboscis. This might be an artifact of providing flies an unnatural substrate on which to adhere (cotton flug or plastic vial). As mentioned above, providing flies with a wooden dowel demonstrated that they preferentially summit and adhere to the dowel, but I can observe cases where the proboscis isn’t making contact with the dowel after death (the fly is seemingly attached via “hugging” the dowel.) Perhaps it’s worth considering that while the proboscis adhesion occurs in the majority of cases, it is not the only way that a fly can remain suspended under natural conditions.

Summiting before death

The experiment with nanchung null animals demonstrated that nanchung is not required for summiting behavior and that summiting behavior can be assayed in a relatively simple way. As nanchung is just one of many possible genes that is implicated in the processes involved in summiting (e.g. locomotion, negative geotaxis), future efforts could expand on this initial experiment to look at genes that encode TRP channels (e.g. those in (158)), receptors, neuropeptides or, using the GAL4-UAS system, systematically knock out neurons within Johnston’s organ to begin to understand the components required for this behavior.

Circadian timing and death – arguments for a fungal clock

Though the experiments herein reported showed that, at some point during infection, light cues are critical for coordinating end of life behavior. Additionally, they showed that infected flies become arrhythmic during their last 48 hours. Thus, these data are not inconsistent with the molecular clock belonging to the fly. However, these data are not also inconsistent with a fungal clock that either acts independently of the host clock and either desynchronizes or fails to initially synchronize in the absence of light cues or relies on the host clock which becomes unsynchronized in the absence of light cues.

Interestingly, Krasnoff et al reported that infected house flies housed in complete darkness die randomly but if infected flies are housed on a photoperiod 72 hours after exposure before being transferred to complete darkness, they die with a periodicity shorter than would be expected if the fly clock dictated timing of death, suggesting timing based on a fungal clock (68). Indeed, molecular circadian clocks have been found within fungi across the fungal tree of life (161). If controlling the timing of death of infected flies benefits E. muscae, it seems very likely that the fungus would have a clock of its own and not depend on the host’s.

If E. muscae has a clock and dictates the timing of death, why don’t we observe flies continuing to die in a gated manner under free-running conditions? The answer could be as simple as that the fungal clock isn’t present until later in the infection. If the clock did turn on, for example, at 72 hours after fly exposure, there would be no environmental cues to entrain to and so each fungal infection would run arrhythmically in its host, killing at random times. Consistent with this hypothesis, the fungus doesn’t start to proliferate in the
infected fly until sometime between 48 and 72 hours and expression of most transcripts is not detectable until 72 hours (Chapter 6). This timing is consistent with the fungal clock not functioning until later in infection.

The very next experiment regarding circadian timing of death needs to ask what happens to infected wild-type and arrhythmic flies that are housed for the first 24, 48 or 72 hours of infection on a 12:12 photoperiod then transferred to complete darkness for the duration of infection. If there is a fungal clock, that, for whatever reason, isn’t effective until one of these time points after infection, the flies should die in a periodic manner.

No evidence so far for sexual attraction of male flies to female cadavers

One of the most fascinating reported behavioral modifications by *E. muscae* is observation that healthy male house flies are sexually attracted to infected female cadavers (91). This attraction would confer an obvious benefit to the fungus in increasing the number of males infected by contact with sporulating cadavers. While I have not yet attempted to test attraction of healthy fruit flies to *E. muscae* CNE1 infected cadavers, I have observed many vials with both cadavers and living *D. melanogaster*, I have never seen a male fruit fly attempt to mate with a female cadaver. It’s possible that this phenomenon is specifically absent in fruit flies either because this strain isn’t capable of inducing this behavior at all or because it evolved in another host species before jumping to fruit flies. However, it is also possible that this behavior is not observed in *D. melanogaster* due to a difference in innate behaviors between fruit flies and house flies rather than behavioral manipulation by the fungus. It has been observed that house flies will attempt to mate with any properly positioned corpse, infected or no (82), but this is not a behavior apparent in fruit flies.

The preference of male house flies for infected over non-infected female cadavers reported by Moller was based on the first cadaver which male flies contacted when placed alongside them (91). It’s possible that first contact was not a measure of sexual attraction but rather novelty of the fungus-filled corpse to the unsuspecting male fly. Follow up experiments found that infected females have decreased levels of the sex pheromone (Z)-9-tricosene, which is known to mediate male attraction to females, compared to uninfected females (92). Decrease in sex pheromone production in infected females has also been observed in the diamond back moth (*Plutella xylostella*) infected with another Entomophthoromycete, *Zoophthora radicans* (162). Reduced sex pheromone production makes sense in light of infection: the animal’s resources are depleted and so it doesn’t expend as much effort in production of sex pheromones which are not absolutely critical for survival. Though it’s possible that other factors can mediate sexual attractiveness in the absence of pheromone (e.g. visual or tactile stimuli), I find it unlikely that another stimulus could so potently counteract decreased sex pheromone as to cause a net increase sexual attraction of cadavers. In order to resolve this issue, Moller’s experiments in house flies need to be replicated and other metrics (e.g. attempted copulation duration) should be used to quantify attractiveness.

Materials and Methods

Fly lines

All fly lines used in this chapter are listed in Table 5.1
Table 5.1. *D. melanogaster* lines used in Chapter 5.

<table>
<thead>
<tr>
<th>Fly line name or description</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>CantonS WF</td>
<td>+; +; +</td>
<td>Liming Wang</td>
</tr>
<tr>
<td>nan -</td>
<td>+; +; nan[36a]</td>
<td>Nick Jourjine</td>
</tr>
<tr>
<td>w -</td>
<td>w- ; +; +</td>
<td>Nick Jourjine</td>
</tr>
<tr>
<td>E49 neurons</td>
<td>+; E49-GAL4; tub-Gal80(ts)</td>
<td>Kristin Scott</td>
</tr>
<tr>
<td>UAS TNT, GFP</td>
<td>+; UAS-TNT, UAS-cd8GFP/Cyo; TM2/TM6b</td>
<td>Kristin Scott</td>
</tr>
<tr>
<td>E49 SCMN neurons</td>
<td>+; +; E49SCMN-GAL4</td>
<td>Kristin Scott</td>
</tr>
<tr>
<td>#24513</td>
<td>y[1] w[*]; Clk[ar]</td>
<td>Bloomington Stock Center</td>
</tr>
</tbody>
</table>

Collection and staining of primary conidia

Three to six fresh cadavers (i.e. those who had not yet mycosed) were collected from exposure vials using the anesthesia dependent methods detailed above. Sporulation chambers were prepared as follows: a small piece of Whatman paper was placed in the base of a small petri dish (60 mm x 15 mm) and wetted with DI water. A bloated cadaver was chosen for each chamber and its wings were removed. The cadaver was placed in the middle of the Whatman paper and the chamber was topped with a custom, 3D-printed top (Figure 5.11) that included a square opening slightly smaller than a standard #1.5 22x22 mm coverslip. The top and bottom were sealed using parafilm and a new coverslip was placed over the opening. Cadavers were left in the chambers at room temperature to sporulate. Coverslips were changed every 30 minutes to 1 hour, as needed, and promptly stained for microscopy either by applying Hoechst (1 ug/mL, for counting nuclei) or lactophenol cotton blue (Remel, for measuring dimensions). Lactophenol blue staining was inconsistent from spore to spore and unnecessary for determining length and width of conidia. Spores were imaged on a compound microscope at 40x for measuring conidia attributes; exact distances were determined by calibration with a 0.01 mm micrometer (Omax). For each attribute (number of conidia, length and width of conidia, diameter of nuclei), at least 50 different primary conidia were counted from three different cadavers.
**Figure 5.11. 3D-printable lid to capture conidia via the ascending method.** The lid is placed atop a 35x10 mm petri dish holding a single cadaver. A 22x22 mm coverslip is placed over the hole to capture spores. Dimensions of the lid are 40 x 40 x 7mm³.

**Infection with *E. muscae* CNE1**

See “Anesthesia-dependent method for *in vivo* propagation of *E. muscae* CNE1 infection” in Chapter 4, Materials and Methods.

**Imaging**

High-speed videos (10,000 fps and faster) were filmed at 50x-100x magnification on a Axiovert 200 microscope (Zeiss) equipped with an Photron Fastcam 1024PCI. Cadavers were mounted in 1.5% agar on a glass slide and arranged such that forming conidiophores and conidia were visible (for spore ejection) or such that cadavers sporulated onto a #1 coverslip in the plane of the camera (for spore landing). Video was captured via Photron Fastcam Viewer software, running at the indicated frames per second in end trigger mode (new frames were continually captured and old frames dumped until the user triggered the filming to stop). Spores or landing sites were manually watched until a spore disappeared or appeared, then video was stopped and last ~10 seconds of footage were manually searched for the spore launching or landing event.

Videos of dying flies (extending proboscies and raising wings, Figure 5.6F, G) were taken with a USB microscope (DinoLite Digital Microscope Pro) using DinoLite software (v1.12) or using a macro lens (Luxsure) attached to a Nexus 5x (Google) or iPod Touch (Apple). Images were taken under ambient light, humidity and temperature.

Time lapse microscopy (conidiophore formation, secondary conidia formation) was taken on a Nikon 80i compound microscope equipped with a Hamamatsu black and white camera (C11440) using MetaMorph software (v. 7.8.00, Molecular Devices). Each time lapse consists of images collected once a minute for the indicated duration. Images were taken under ambient temperature and humidity.

Additional photos and videos (Figures 5.6A,C,E,H,I, 5.7B and 5.8B) were shot on a Nexus 5x mobile phone (Google) or iPod Touch MP3 player (Apple).
Proboscis extension

All E49 neurons or just the E49SCMN neurons were knocked out by crossing the appropriate GAL4 line to UAS-TNT, GFP (Table 5.1). Indicated progeny or parental lines were randomly chosen to subject to heat shock in a 37°C water bath for 35 minutes as early pupae (per (153)). Up to 50 flies of indicated genotypes younger than five days post-eclosion were exposed to E. muscae CNE1 as in Chapter 4 (anesthesia-dependent method) and monitored from 96-168 hours after exposure for deaths, noting position (medium, vial side or flug) where each cadaver was found. When it became apparent that cadavers could adhere in elevated positions by means other than their proboscies, vials were scrutinized under a dissecting scope for any cadavers that demonstrated clear attachment via proboscis to the vial side or flug.

Summiting

CantonS WF, nan- or w- flies were exposed to E. muscae CNE1 as described in Chapter 4 (anesthesia-dependent method) with one modification: at 48 hours, flies were transferred onto GB+ medium with a dowel pre-marked every 1 cm with permanent marker. Flies were monitored every 24 hours from 96-168 hours for cadavers. Position of cadavers in vial (flug, dowel, vial side or medium) and location on dowel (measured as the 1 cm increment where the lowest part of their abdomen was found) were recorded prior to anesthetizing and removing flies that were still alive. Weighted mean position was determined by summing the (# of cadavers*position1) + (# of cadavers*position2)... and dividing by the total number of cadavers.

Circadian timing of death

Flies of the indicated genotype were reared on either light cycle A (lights on 12 am – 12 pm, 21C, ~60% humidity) or light cycle B (lights on 6 pm – 6am, room temperature and humidity). For experiments run in complete darkness, 25 flies, 13 females and 12 males, were exposed to E. muscae CNE1 as described in Chapter 4. All exposure vials were set up before the utilized cadavers sporulated (~11 am for light cycle B cadavers, ~3 pm for light cycle A cadavers). Exposure vials were housed in a humid chamber in a dark 21C incubator wrapped in blackout cloth for approximately 24 hours before loading into Drosophila activity monitors (DAMs, Trikinetics, Figure 5.S1). Before loading flies, DAM tubes (5 mm x 65 mm, polycarbonate) were prepared containing such that one end of the tube held approximately one cm of 1.5% agar, 10% sucrose and was wrapped in parafilm to prevent drying out.

To load the flies, all accessible light sources (overhead fluorescents, computer monitors, LEDs on chargers etc.) were disabled before removing the humid chamber from 21C and placing on the bench at RT. Vials of flies were kept under a foil-lined box as they waited to be processed. One vial at a time was retrieved from the box, knocked out with CO2 under a dissecting scope whose LED light ring was covered with a red gel filter (Neewer), sorted by sex and loaded into individual DAM tubes with the aid of a red LED flashlight (KMD Aero) before capping each tube with an autoclaved cotton plug. For each vial, eight females and eight males were chosen for loading into DAM vials. Unexposed flies (i.e. controls) were always processed before proceeding to exposed flies. CO2 pad was
wiped down with 70% ethanol between vial types to prevent cross-contamination. DAMs were loaded from bottom to top row, filling a row and securing each tube with rubber bands before proceeding to the next. Loaded tubes were kept under a foil-lined box to prevent light exposure. When all loading was finished, DAMs were covered with blackout cloth and transported back to the 21C incubator. There they were attached to the DAM interface unit and recording began, binning counts in 30 second intervals. Recording continued for ~170 hours until the experiment was stopped.

Like with loading, experiments were stopped by first disabling all light sources, then carefully disconnecting and removing DAMs from the 21C incubator to not disturb adjacent experiments, and sealing incubator before turning on the overhead lights for manual inspection. Each DAM tube was inspected to see if the fly was dead or alive. If dead, the tube was inspected for evidence of sporulation to determine if the cause of death was patent E. muscae infection.

For experiment run on a 12:12 light cycle, flies were exposed as above but without concern for light contamination; resultant DAMs were housed on a 12:12 light cycle for the duration of the experiment. Each channel was checked daily following sundown to see which flies had died within the previous 24 hours.

To determine time of death, DAM data were processed using Python to determine time of last movement (accurate to 30 seconds) and to plot movements over time. For each channel, the reported time of last movement was manually crosschecked using the plot of activity data. In cases where there was an erroneous movement (i.e. a signal occurring more than 24 hours after the fly’s last movement), the time of last death was manually recalled. For data visualization, data were binned into 15-30 minute intervals and the average movements of unexposed animals (controls), exposed or entrained light cycle of exposed flies and cadavers and the time of last movement for each observed cadaver were plotted in Excel. There were generally no obvious differences between male and female activity for the unexposed animals so sexes were combined for the control data.
Supporting Information

Figure 5.S1. Drosophila activity monitor (DAM). Each of 32 tubes contains a source of food and water (1% agar, 10% sucrose) on one end, a breathable, cotton plug on the other and houses one fly (diagrammatic example shown above photo). Tubes are placed into DAM channels and secured with rubber bands to prevent sliding of the tube relative to the plane of the infrared beam. Each DAM is housed in an incubator and connected via phone cable to a powered interface unit that is, in turn, connected to a Mac desktop via USB. Beam breaks are collected via Trikinetics software for each channel binned in the indicated interval.
Figure 5.S2. Proboscies of E. muscae CNE1-infected cadavers stained with lactophenol cotton blue. A) Proboscis at 10x; B) Proboscis at 20x. Arrows indicate fungal morphologies visible emanating from the labella (also stained blue). Spots in the background (out of the focal plane) are hyphal bodies released by tearing the proboscis from the cadavers’ heads. E. muscae CNE1-infected D. melanogaster cadavers were collected immediately after death. Their proboscises were torn off using Dumont 55 forceps, briefly covered in lactophenol cotton blue then washed with 1x PBS and imaged on a Nikon 80i compound microscope using a Nexus 5x mobile phone (Google).
Chapter 6: Progression of E. muscae CNE1 infection in D. melanogaster: gene expression and histology

Abstract

Entomophthora muscae CNE1 is a fungal pathogen that infects, behaviorally manipulates and ultimately kills Drosophila spp, including D. melanogaster. In order to get a broad sense of the molecular changes that occur in the host and the fungus over the course of infection, I performed a transcriptomic time course which monitored gene expression from exposure until death. These data did not reveal any obvious means by which the fungus could affect behavioral changes in the host, but did reveal that the host mounts a robust immune response to the initial infection and that some features of this initial response linger until later infection stages. Using in vitro RNAseq data, I assembled an E. muscae CNE1 transcriptome to map fungal reads from this time course. Though the identity and functional output of E. muscae CNE1 transcripts are still unknown, these transcriptomic data revealed patterned gene expression in the fungus as it progressed from early infection, to late infection, to post infection. In order to understand where the fungus travels within the host during infection, I performed a histological time course analogous to the transcriptomic time course. Using contrast staining, I found that the fungus is abundant throughout the fly body cavity in infected animals at 72 hours of infection and beyond, but is only consistently located in the brain and central nervous system at 48 hours, early in the infection. I corroborated this observation using an E. muscae CNE1-specific fluorescent probe. More work is needed to understand where the fungus is at early points in infection and what morphology it adopts. Understanding fungal tropism throughout infection will allow us to assay fungal-targeted tissues for transcriptomic and other -omic responses towards revealing how the fungus ultimately changes host behavior.

Introduction

E. muscae CNE1 is a strain of the fungal pathogen E. muscae that infects, modifies the behavior of and kills D. melanogaster. In previous chapters I showed that this strain is amenable to culture both in vivo and in vitro in the laboratory and described observations of the fungal lifecycle and fly response that bolster or expand on observations made since 1855. These previous efforts were generally focused on external morphologies and phenotypes within this system, still leaving us just guessing as to what is happening inside the fungus and fly. In order to acquire a more complete understanding of how infection progresses, I wanted to look at both the molecular and physical changes that occur in both fly and fungus from host exposure until death. While I wanted to understand general features of the infection over time, I was particularly curious to see what, if anything, I could glean about how the fungus elicits changes in host behavior. Here, I describe what I found.
Results

To gain a first comprehensive look into how *E. muscae* CNE1 infection progresses in *D. melanogaster* at the molecular level, I decided to sequence mRNA from individual, exposed flies over the first 96 hours after exposure to the fungus. I knew that in any given exposure vial there are a mix of infected and uninfected animals and faced the complication that infected flies are phenotypically indistinguishable from uninfected animals. In order to ensure that I picked and sequenced infected animals, I wanted to determine if I could distinguish infected from uninfected animals by PCR. I reasoned that amplifying ribosomal RNAs would be my best bet as they would be distinguishable from fly rRNAs and present in several copies per cell, as opposed to a handful of times in the genome.

In order to determine when I could detect *E. muscae* rRNA in exposed animals, I collected exposed or control animals every 24 hours for four days, extracted total RNA individually from each animal performed reverse-transcription PCR on these total RNA samples using *Entomophthora*-specific ITS primers (emITS1, emITS4). It should be noted that I did not collect these flies randomly but rather chose animals that had visible spores on their abdomen (24 hours), melanization scars on their abdomen to indicate successful spore entry (48 and 72 hours) or that had clearly died of infection and would soon sporulate (96 hours). One out of six exposed flies gave a faint *E. muscae* ITS by 24 hours or 48 hours after exposure; all six flies collected at 72 hours gave robust ITS bands (Figure 1). Curiously, all three of the cadavers that I sampled, flies that were unquestionably infected with *E. muscae* CNE1, gave weak ITS bands.

![Figure 6.1. Reverse-transcription PCR of *E. muscae* CNE1 ITS sequence in exposed and control flies. All samples were run on the same gel with equal loading volumes; samples to the right of the dashed line were run on the lower half of the same gel containing samples to the left of dotted line (see methods for sample generation). DNA ladder (5 µL Hyperladder 1 kb, Bioline) was run in the first lane of each gel half. Samples are shown in chronological order, with the time point indicated above the left-most sample. Black lines indicated exposed flies; white lines indicate unexposed flies. Exposed flies collected at 96 hours were cadavers that had died of *E. muscae* CNE1 infection. Plus (+) indicates positive control (*E. muscae* CNE1 in vitro RNA template); minus (-) indicates additional negative control (*D. melanogaster* RNA from earlier experiment, before the discovery of *E. muscae* CNE1).](image)

Following this result, I concluded that I couldn’t guarantee that all of my 24 hour and 48 hour samples were infected by *E. muscae* CNE1 using RT-PCR against ribosomal RNAs. However, I felt confident that I would be able to distinguish infected from uninfected
exposed animals after sequencing mRNA since a) I would only need a few *E. muscae* reads to confirm that *E. muscae* was present and b) it would be unlikely for infected and uninfected exposed animals to demonstrate identical transcriptional profiles. Thus, I moved forward and collected individual female flies that had either been exposed to *E. muscae* or subjected to a “mock” exposure for every 24 hours over the first 120 hours (Fig 6.S1A), extracted total RNA and generated mRNA libraries for sequencing.

The first question I wanted to address in analyzing these data was how many reads in each sample belonged to the host versus the fungus. In order to answer this question, I would align the reads both to host and fungal references and count the resultant alignments for each. Reference genomes and transcriptomes for *D. melanogaster* are readily available, but I did not have access to a reference to use for the fungus. I couldn’t just assume that all the reads that didn’t align to *D. melanogaster* were *Entomophthora* as sequencing error and mRNA from other sources (e.g. *S. cerevisiae*) would artificially inflate this number. Therefore, I needed to generate an *E. muscae* reference transcriptome.

**Establishing a reference transcriptome for *E. muscae* CNE1**

I had previously sequenced mRNA from *in vitro* *E. muscae* CNE1 culture to aid with gene annotation of our genome assembly and had assembled these using the *de novo* transcriptome assembler TRINITY (this assembly will be referred to as TRINITY1; Table 6.S1). I aligned all reads from all samples in parallel to each ivTRINITY1 and the *D. melanogaster* reference. As a first check of these alignments, I plotted the number of reads that mapped to each reference. Surprisingly, I found that there were a significant number of reads from my control samples (i.e. unexposed flies) that mapped to the *E. muscae* reference. There were two possible explanations for this: either my *in vitro* reference was contaminated with *D. melanogaster* reads or there was significant horizontal gene transfer from *D. melanogaster* to *E. muscae*. The first possibility seemed more likely. Though the library I had sequenced was derived purely from *in vitro* E muscae RNA (completely devoid of *D. melanogaster* RNA), I had multiplexed this sample on a lane with *D. melanogaster* mRNA samples. As de-multiplexing results on sequencing barcodes and any sequencing includes some degree of error, I should have anticipated that my reads would not be *Drosophila*-free, but I didn’t anticipate to what extent this contamination would skew a TRINITY assembly.

Indeed, upon aligning the original sequencing reads from my *in vitro* *E. muscae* sample to the *D. melanogaster* transcriptome, I found that approximately 0.1% of these reads mapped. I removed these reads from my set and re-assembled an *E. muscae* reference transcriptome using TRINITY (referred to as TRINITY2). To my surprise, removing the scant 0.1% of *Drosophila* reads before assembly resulted in a transcriptome with 8,040 fewer transcripts than TRINITY1. As an additional measure, I removed 792 transcripts where expression was greater than 1 in control (unexposed) samples and that gave blastn e-values of less than 1e-50 to possible contaminant references (i.e. organisms that are not fungi or Twyford virus9). Aligning reads to this reference (TRINITY2.1) gave

---

9I did not remove sequences with viral homology because, while sequencing the genome, BLAST searches from kilobase genomic reads showed that sequences annotated as Twyford virus are actually *Entomophthora*. The sequences were presumed to belong to a virus by virtue of being found in *Drosophila* and having some, though scant, similarity to viral sequences.
minimal alignment of control samples (~2%) while aligning at least 70% of total reads per sample.

In parallel, I prepared an assembly derived from all in vivo RNAseq time course reads that failed to align to the D. melanogaster transcriptome (TRINITY3), removed transcripts that showed significant homology to non-fungal organisms (TRINITY3.1), then removed any transcripts that showed aberrant expression patterns in uninfected flies (TRINITY3.1.1). Alignment of the time course data to this reference gave virtually identical results at the level of read distribution, though the BUSCO statistics of this assembly were better than those of TRINITY2.1 (Table 6.S1). I continued my analyses with the in vivo (TRINITY3.1.1) reference.

E. muscae cadavers contain more fungal mRNA than fly mRNA

Now equipped with a reasonable E. muscae reference, I examined the percentage of reads that aligned to host or fungus within my time course samples (Figure 6.2). E. muscae reads are low abundance until 72 hours after exposure, which probably reflects that fungus doesn’t begin to actively divide until between 48 and 72 hours. The majority of mRNA in cadavers comes from E. muscae, with some flies barely having any remaining D. melanogaster RNA detectable. This is consistent with my previous observations that cadavers collected from field sites yield little host DNA, as assessed by the extremely faint bands that result from PCR genotyping to identify the host species.

![Figure 6.2. Read alignment statistics for all time course RNAseq samples.](image)

*Figure 6.2. Read alignment statistics for all time course RNAseq samples.* Samples are ordered first by treatment (unexposed samples = A-C; exposed samples = D-F, Cad. = cadavers) then by time point (numbers correspond to days since exposure or mock exposure, e.g. “1” samples are 24 hours, “2” are 48 hours etc.)
D. melanogaster gene expression throughout the course of E. muscae CNE1 infection

To examine how host gene expression changes over the course of infection, I next quantified expression for each gene in D. melanogaster and performed hierarchical clustering on this entire set to assess which samples and genes were most similar to each other (Figure 6.3A). The resultant organization shows four major groupings: (i) early controls (24-48 hours) and early exposed (24-48 hours), (ii) late exposed (72-96 hours) and cadavers (96-120 hours), (iii) early to late controls (48-120 hours) and one 48 hour exposed sample, (iv) remaining late controls (96-120 hours), early exposed (24-48 hours) and two late exposed (96 hours). As a fly’s transcriptional profile will change as it ages, it's not surprising that 24 hour controls are distinguishable from controls from later time points. The reason for the separation of exposed samples based on global gene expression is less clear, and could be due to a number of factors (e.g. nutritional status of the fly, speed of infection progression, etc.).

I next performed clustering of samples by expression of only immune genes (i.e. genes we would expect would depend on sample treatment) (Fig Figure6.3B). The resulting pattern is a clear separation of almost all exposed samples (left) and control samples (right). Interestingly, two 96 hour exposed samples still group with controls. This may indicate that either these animals were never infected or came in contact with E. muscae or that they did but were since able to recover.
Figure 6.3. Hierarchical clustering of host gene expression data in *E. muscae* CNE1-infected and age-matched uninfected *D. melanogaster* over the first 120 hours (five days) after exposure. Hierarchical clustering grouping both genes and samples for A) entire transcriptome or B) only genes with immune GO term annotations. Complete linkage hierarchical clustering was performed in Gene Cluster 3.0 after filtering genes that didn't exhibit at least two TPM (transcripts per million) expression in three out of 42, log transforming TPM values and centering genes around the median TPM value. Flies that were exposed to *E. muscae* CNE1 are denoted by a black bar above the sample name.

Next, I asked which genes were differentially expressed between age-matched samples. After 24 hours, all exposed samples show a robust anti-fungal immune response and an overall decrease in metabolism compared to controls (Figure 6.4A). This trend is
unaffected by omitting samples that clustered in group iv (Figure 6.3). After 48 hours, significant differences below p-value 0.001 become scarce. These genes exhibit no obvious relationship to one another and yield no GO-term enrichment. At 72 hours, exposed samples show decreased expression of a slew of genes related to general cell metabolism (DNA replication, RNA processing, protein translation and folding) compared to age-matched controls (Figure 6.4B).

**Figure 6.4. Genes exhibited differential expression between flies 24 hours after mock exposure or exposure to E. muscae.** A) Left: Volcano plot for all genes at 24 hours. P-value is determined by ANOVA grouping control vs. exposed samples. Genes with p-value under 0.001 are shown in color. Right: Panther GO-term analysis (complete biological process) of genes overexpressed in exposed animals (red) or control animals (blue). B) Left: Volcano plot for all genes at 72 hours. P-value is determined by ANOVA grouping control vs. exposed samples. Genes with p-value under 0.001 are shown in color. Right: Panther GO-term analysis (complete biological process) of genes overexpressed in control animals (blue). There are no significant GO term enrichments for set of genes overexpressed in exposed samples.

Next, I looked at genes that were consistently different between control and exposed samples from 24-72 hours (Figure 6.5A). I excluded my cadaver samples (both 96 and 120 hours) from this pooled analysis because the animals are dead, and variations in gene expression at that point are confounded by mRNA degradation. I also opted to exclude exposed animals at 96 hours because two of these three samples more closely resemble
control animals than exposed animals (Figure 6.3B). Genes that are under-expressed in exposed animals compare to control animals from 24-72 hours are enriched for a handful of metabolic processes, including arginine and glutamine synthesis. Interestingly, both arginine and glutamine are amino acids synthesized from the Kreb’s cycle intermediate alpha-ketoglutarate. In times of starvation, the cell would be expected to prioritize generating ATP via the Kreb’s cycle over synthesizing these amino acids. The idea that the fly is starving is not just consistent with these enrichments but also with the observation that basic cell metabolism (macromolecule synthesis) is substantially decreased at 72 hours (6.4B).

Genes overexpressed at 24-72 hours samples are enriched for immune function. As immune genes were recognized at 24 hours as being overexpressed compared to controls but not at 48 or 72 hours, I wanted to make sure that immune genes found in this analysis weren’t discovered just by virtue of being over-expressed during an initial immune response. Therefore, I plotted the fold overexpression of these genes compared to controls (Figure 6.5B.) Some genes (especially IMs) appear to wane over time, but some are consistently expressed throughout the first 72 hours, indicating that overexpression at 24 hours was not sufficient to identify immune genes in this analysis.

**Figure 6.5. Genes that are consistently over or under-expressed compared to controls over the first 72 hours after exposure to E. muscae.** A) Left: Volcano plot for all genes over the first 72 hours after exposure. P-value is determined by ANOVA grouping 24-72 hour control vs. exposed samples. Genes with p-value under 0.001 are shown in color. Right: Panther GO-term analysis (complete biological process) of genes overexpressed in exposed animals (red) or control animals (blue). B) Ratio of exposed to control expression for immune genes that are significantly different between control and exposed animals from 24-72 hours (ANOVA, p value < 0.001). Genes are ranked according to percent difference from expression at 72 hours versus 24 hours. Genes to the right of the dotted line are those that are significantly down-regulated.
vertical line exhibit a decrease of at least 15% in average expression at 72 hours compared to 24 hours.

Overall, these time course data show that a) flies exhibit a strong immune response to fungal exposure at 24 hours coupled with an overall decrease in nutrient metabolism, b) infected flies sustain increased expression of a subset of immune genes through the first 72 hours of infection and c) by 72 hours fly cellular metabolism is significantly diminished in exposed flies compared to controls.

**E. muscae CNE1 gene expression throughout the course of infection**

To understand fungal gene expression changes over the course of infection, I quantified expression for each gene in *E. muscae* using TRINITY3.1.1 and performed hierarchical clustering on this entire set to assess which samples and genes were most similar to each other (Figure 6.6). Three groups emerge from this analysis: group I is comprised of mostly early samples (exposed for one or two days), group II is comprised entirely of late samples (exposed for three to four days), group III consists of all six cadaver samples sequenced. The bulk of transcripts are not expressed until three days after exposure, which could simply be a consequence of the fungus being low abundance until this time point. Interestingly, there are three obvious groupings of genes, a-c, that demonstrate patterns that cannot be explained merely by fungal abundance in the samples. Group a demonstrates genes that are expressed early on but whose expression is depressed later on in infection, group b consists of genes that don’t turn on until after the fungus has killed the host, and group c contains genes that turn on during the later phases of growth in the living host but are turned off after the fly is killed.
Figure 6.6. Hierarchical clustering of fungal gene expression data in *E. muscae* CNE1-infected *D. melanogaster* over the first 120 hours (five days) after exposure. Hierarchical clustering grouping both genes and samples for entire 10809 out of 86510 transcripts (TRINITY3.1.1) for exposed samples only (uninfected controls omitted). Complete linkage hierarchical clustering was performed in Gene Cluster 3.0 after filtering genes that didn’t exhibit at least ten TPM (transcripts per million) expression in three out of 27, log transforming TPM values and centering genes around the median TPM value.

At present, we know don’t know what gene products these transcripts encode, as (there is little homology between entries in protein databases and the translated open reading frames in these transcripts). We look forward to understanding more about these data once we complete assembly and annotation of the *E. muscae* CNE1 genome.

*E. muscae* CNE1 is first observable in the fly brain and central nervous system at 48 hours after exposure.

The RNAseq time course gave me a broad sense of host response at 24 hour time point but didn’t suggest obvious route by which behavioral manipulation of the host could be achieved. In order to learn where the fungus localizes within the host over the course of infection, I performed a histological time course (Figure 6.S1B). Briefly, I collected animals at 24 hour intervals for the first 168 hours after exposure to the fungus, embedded them in
paraffin and applied a contrast stain to 8 micron sections of these samples to observe fungal morphology. I reasoned that by comparing an animal full of fungal material to one lacking fungal material, I could with high confidence deduce *E. muscae* CNE1 morphology and use this information to locate fungal material in earlier infection time points. The initial choice of stain was not obvious; many were tried (toluidine blue, hematoxylin and eosin, Gram stain, Grocott’s methenamine Silver stain with the help of Ciera Martinez) before settling on Safranin O/Fast Green FCF (SFW, suggested by Richard Humber, personal communication). Safranin O is a contrast stain that turns nuclei, muscle and ovaries red and remaining tissues green. Using Under the SFG regime I found that *E. muscae* CNE1 hyphal bodies have consistently-sized nuclei that stain red or purple and cytoplasm that stains purple to green (Figure 6.7 A’).

**Figure 6.7.** Histological comparison of uninfected *D. melanogaster* and *E. muscae* CNE1-infected *D. melanogaster* cadaver. A) Uninfected female adult stained with SFG imaged at 100x magnification. Contents of the boxed regions, brain, thoracic muscle and abdomen (including ovaries), are shown to the left at 400x magnification. B) Adult recently killed by *E. muscae* CNE1 infection (not yet sporulating) stained with SFG imaged at 100x magnification. Contents of the boxed regions, brain, are shown to the right at 400x magnification. Large *E. muscae* CNE1 nuclei are clearly visible in the brain (smaller than that in A because it is being consumed by fungus), thoracic muscle and abdomen. No organs are visible in the abdomen as the fungus has eaten these tissues. Abdominal inset shows fungal hyphal bodies differentiating into conidiophores, fungal structures that will form and then forcibly eject infectious conidia to spread the fungus to a new host.

Using the morphology of *E. muscae* CNE1 hyphal bodies revealed SFG, I worked through each 24 time point in my time course to locate fungal cells. In infected animals that survived 96-144 hours, fungal grow was rampant throughout the entire body cavity (head, thorax and abdomen) (Figure 6.8D). In these animals, the gut, gonads were uninvaded, brains and muscle were invaded but intact and fat body cells were depleted compared to uninfected controls. Infected animals at 72 hours showed fungal growth throughout the
body cavity but at lower titer than observed in 96 hours and beyond (Figure 6.8C). At 48 hours, four out of the five animals examined had fungal morphology in the brain and/or central nervous system (thoracic ganglion) but growth was not apparent elsewhere in the fly (Figure 6.8B). The sample that did not have fungus in the brain or CNS at 48 hours had fungus in the head adjacent to the blood brain barrier, which would be consistent with the fungus about to enter this organ. At 24 hours, no hyphal bodies were apparent (Figure 6.8A). Knowing that these animals were exposed and that most (if not all) were successfully infected, we know that fungus is present in these samples. We must therefore conclude that the fungus adopts a morphology distinct from that seen 48 hours and later.

Figure 6.8. Morphology of D. melanogaster at 24-96 hours post exposure to E. muscae. All images shown are from sagittal paraffin sections of infected flies at the indicated time points stained with SFG. Center-most images were taken at 100x magnification; peripheral images (enlarged views of boxed regions) were imaged at 400x magnification. Each picture represents 3-5 animals examined (roughly half male, half female for each sample type). Insets to the side of each picture show the brain and abdomen at the indicated locations. All images are identically scaled for comparison. 24, 72 and 96 hour samples shown are males; 48 hour sample is female. No sex differences were observed over infection.

Like any contrast stain, SFG can aid with, but not definitively determine, cellular identify by morphology. To confirm that the morphologies I attributed to E. muscae CNE1 in my intermediate infection samples were, in fact, fungus, I used fluorescence in situ
hybridization to specifically label *E. muscae* cells within the context of an infected fly. Indeed, I confirmed that the fungal morphology observed 48 hours and beyond was *E. muscae* CNE1 (Figure 6.9).

\[\text{Figure 6.9. Fluorescence in situ hybridization with an *E. muscae*-specific probe confirms that large nuclei observed in animals 48 hours after exposure muscae is *E. muscae*. The merged images clearly demonstrate which tissues belong to *D. melanogaster* (those only stained with DAPI, blue) and which are fungal (those that stain with both DAPI and Alexafluor-633 *E. muscae* probe).}\]

Taken together, the RNAseq and histology time course data presented in this chapter indicate that *E. muscae* CNE1 infection progresses in a stereotyped manner. At 24 hours after exposure, *D. melanogaster* show a robust antifungal immune response, though the fungus is nearly undetectable within the fly, indicating that it is at very low titer. At 48 hours, the fungus has adopted the morphology which it will assume until killing the host. Overall fungal titer is still quite low, but fungus is observable within the host’s CNS. As the abdomen is the most likely point of initial entry for the fungus (it is the biggest target for the fungus to hit and provides a direct route to the hemolymph), we can infer that the
fungal titer increases and fat body decreases.

Discussion

The time course experiments described above shed light on the molecular and morphological features of *E. muscae* CNE1 infection of *D. melanogaster*. Here, we will consider more detailed repercussions of these observations as well as considerations for this and future studies.

**E. muscae** infection and host immune response

It is generally thought that *E. muscae* evades the host response by growing protoplastically (i.e. without a cell wall, components of which would be recognized and targeted by the host immune system). In both the gypsy moth and the greater wax moth, it has been shown that the host immune cells recognize walled *Entomophthora* fungal cells, but there is little cellular response to protoplasts (163, 164). Based on these findings, it has been posited that the host doesn't detect the ever-increasing fungal mass within until the end of infection when the fungus puts on a cell wall that contains epitopes that the host can recognize (165, 166). As a result of ostensibly evading the immune system, it has also been hypothesized that *E. muscae* does not generate toxins, as it would have no incentive to do so in the absence of attack by the host (165).

My data show that there is a robust initial response to *E. muscae* exposure. Many of the immune genes that are induced with *E. muscae* have also been observed to be induced by exposure to other, more generalist fungal pathogens (e.g. *Beauveria bassiana*, *Metarhizium anisopliae* (167, 168)). These data clearly indicate that the host detects an invader early on in infection. Furthermore, there is a detectable, although not as strong, response as the infection progresses (i.e. 48-72 hours, Figure 6.5B), though at this point we can’t say if this response is a slow disengagement of the initial response or stimulated *de novo* by the growing fungus. Though there is not a GO enrichment of genes that are overexpressed in exposed flies at 72 hours compared to controls, it is notable that Thor (recently implicated in immune response) shows a large increase in expression at this time point. Could this be a late wave of defense? Validation via RT-qPCR is needed to determine if Thor is genuinely significantly upregulated in exposed compared to controls or if Thor is always overexpressed, but just happened to exhibit more variable expression in my earlier samples and thus was not picked up by ANOVA.

Though five of my six cadavers have low levels of immune gene transcripts, the sixth cadaver exhibits a spike in expression of anti-fungal peptides Drosomycin and Metchkinowin, the beta-glucan receptor GNBP3-like and several I1M family genes. It’s possible that this fly is demonstrating an immune-system overload as proposed by (165). It might be the case that an immune spike occurs in all animals; I could have sampled too late
to observe it in my other five samples but was able to see it in one sample that was late to respond. However, I would imagine that in the presence of copious fungal epitopes the fly immune system would continue to go berserk until death, not drop back down to levels comparable to 72 hour samples. At present, I’m inclined to interpret this odd-sample-out as a fluke rather than an indication of a moribund immune spike. This hypothesis could be definitively addressed in the future via performing RT-qPCR on representative immune genes on several flies sampled near the end of life.

**Comparison of transcriptional response of *E. muscae* to other pathogenic fungi**

In order to understand if the trends I was seeing in my data were generally true of flies infected with non-behaviorally modifying pathogens, I compared my time course data to a published transcriptional time course of fruit flies infected with generalist fungal pathogen *B. bassiana* (169). Given that *E. muscae* and *B. bassiana* require different amounts of time to kill their host, it was unclear what meaningful comparisons could be performed between the two datasets. As a result, I was extremely conservative in my comparison. First, I looked for genes that were always up or down in my dataset but not the *B. bassiana* dataset. I found only a handful of genes, including IM4, which a later study (170) found to be induced by *B. bassiana* infection. The other comparison I felt was robust enough to be worth making was that between my 72 hour samples and the last time point of the *B. bassiana* set, 96 hours after exposure. Here, I observed a handful of genes that trended in the opposite direction between datasets, including genes involved in immunity (Thor) and metabolism (sugb) (Figure 6S2).

In the end, I concluded that any comparison of an *E. muscae* time course to a time course of infection with a different fungal pathogen is not necessarily what will be most useful in terms of understanding genes involved in behavioral manipulation. A significant hurdle with these data is that there’s signal coming from so many different tissue types and all of this information effectively ends up getting averaged over the entire animal. Thus, while I can pick up on broad trends (metabolism and immune system changes), subtler or tissue-specific changes (those which I anticipate will be informative about how the fungus works, especially to manipulate behavior) are simply not detected by such an experiment, even if I compare my dataset to that of a “control” pathogen. Subsequent transcriptomic experiments should target specific tissues where the fungus is known to localize (e.g. the brain and CNS) rather than the whole fly such that we might be better able to interpret the data.

**Variability in infection progress: different animals progress through infection at different rates**

Though grouping time points was generally useful in demonstrating trends over the course of infection, it is important to recognize that just because two animals have been exposed for the same duration of time that these two animals will not progress through infection identically. This was especially apparent for the RNAseq data. I noticed that the host gene expression in exposed animals at 24 and 48 hours tended to be more variable than those for 72 hours. This is likely due to chance: I imagine that I simply got lucky in sampling at 72 hours and managed to pick animals that were at similar points in infection, whereas I picked animals on more divergent infection trajectories at other time points. This may at
least partly explain why I observed so much differential expression at the 72 hour time point in exposed versus controls whereas less was observed at earlier time points, especially at 48 hours. It is likely that several factors play into whether or not an infection succeeds and how quickly it progresses (e.g. initial exposure titer, size of host, nutritional status of host etc.). Thus, future work should consider how to expose flies in a more reproducible manner (e.g. control fungal titer) or, if this is not feasible, determine a metric to gauge progress of infection so that similarly-progressed samples can be compared.

At present, we have no way of knowing which exposed flies are infected and which will die when. More work is needed to understand how common it is for an infection in progress to fail and the basis of this failure. As an initial step in this assessment, we should fluorescently tag a host gene that is induced upon infection and tracks with infection progress. From the RNAseq data presented in this chapter, there are several candidate immune genes that could fit this bill. Ideally, we should label *E. muscae* with a fluorescent marker to be able to determine fungal titer without sacrificing the animal. In a successful infection, fungal titer will tend to increase until death. Observing a decrease in fungal titer would definitively indicate that the host is clearly an infection.

**An unanswered question: do all exposed flies get sick?**

Interestingly, the living animals selected for RNAseq at 96 hours are inconsistent in their host transcriptional immune response: two of the three animals (4D2, 4F2) more closely resemble control animals than infected animals in host transcription (Figure 6.3B). Fungal expression in these samples is distinguishable, with 4D2 samples behaving as other late infection time point samples while 4F2 clusters with earlier time points. It’s possible that both of these animals are in the process of recovering from infection or are behind schedule compared to contemporaneous samples. The proportion of fungal reads present in these samples is lower than what would be expected for late time points, which is consistent with either scenario. At this point we simply don’t know if every instance of a spore hitting a fly leads to a productive fungal infection. There is some evidence to the contrary: I have consistently observed that exposed flies die prematurely. These animals are generally smaller than others in the vial and are often covered in spores. This could indicate that getting hit by too many spores (an unlikely outcome in the natural world) leads to an overwhelmed fly (e.g. overactive immune system or accelerated fungal growth) that dies before being manipulated. These flies do not sporulate, though it’s possible that they do produce resting spores. On the other hand, I’ve observed that survival of exposed flies is substantially increased when flies are exposed to cold anesthesia or small quantities of anti-fungal. This indicates that there are ways of either halting or slowing an infection, though whether the fly’s immune system is generally capable of doing this is unknown.

**Where is the fungus after 24 hours?**

I can clearly observe *E. muscae* cells in histological samples at 48 hours after exposure or beyond, but, despite scrutinizing my 24 hour samples, I was unable to locate cells that were unambiguously *E. muscae* (Figure 6.8). It’s possible that, for any given 24 hour sample, loss of sections could lead to not being able to observe the fungus. It seems unlikely that this would be true for five samples that each mounted and slice individually. As it turns out, I am not the first to be baffled by not being able to detect fungus soon after exposure (R.
Humber, personal communication). I hypothesize that fungal morphology at 24 hours is different from that observable at 48 hours and perhaps that it adopts a morphology similar to a fly cell type, which has so far rendered me unable to identify it. Future work should take advantage of my *E. muscae*-specific FISH probe to locate the fungus at time points earlier than 48 hours. Observing earlier forms of the fungus will also allow us to see how the host immune cells interact with it and can also inform us about how the fungus is able to travel from the point of entry to the CNS (e.g. if this early morphology has any obvious means of locomotion).

**Materials and Methods**

*E. muscae* CNE1 infection RNA time courses

For RNAseq time course and reverse-transcription PCR experiments, three mock vials and three exposure vials were started with 25 CantonS WF flies 0-1 days old with either 0 (mock) or 6 (exposure) cadavers embedded in AS. Flies were incubated for the first 24 hours at 18C confined to 2 cm with cadavers, then moved to 21C where the confinement was relieved. Flies were transferred to GB+ (40% organic banana [w/v], 2% agar, 0.3% propionic acid) at 48 hours where they continued to be housed at 21C. Vials were sampled every 24 hours for 96 hours (reverse transcription) or 120 hours (RNAseq, Figure 6.S1A). Before sampling, all materials to manipulate flies (CO2 pad, gun, forceps, brush) were treated with 10% bleach, wiped with DI water then sprayed with 70% ethanol. Control vials were always sampled first. For each control vial, all flies were anesthetized and one female was selected for sampling. She was placed in a non-stick 1.7 mL Eppendorf tube containing 250 uL Trizol (ThermoFisher Scientific), submerged in the Trizol and then flash frozen in liquid nitrogen (LN2). For each exposure vial, all flies were anesthetized and two females were selected or sampling. Animals were preferentially selected based on evidence of exposure to fungus. Each female was submerged in 250 uL Trizol in a non-stick 1.7 mL Eppendorf tube then flash frozen in LN2. All materials that handled flies (CO2 gun, pad, forceps) were treated with 10% bleach and rinsed with DI water between sampling exposure vials.

Reverse-transcription PCR

RNA was prepared from each thawed sample by homogenizing with an RNase-free pestle (Kimble Chase), washing the pestle with 750 uL Trizol, then proceeding using the manufacturer’s protocol. RNA was then treated with Turbo DNase (ThermoScientific) per the manufacturer’s protocol and quantified using a Qubit Fluorometer (Qubit RNA HS assay kit, ThermoFisher Scientific). For each sample, 1 µL or 160 ng of DNase-treated RNA, whichever was more, was added to a new, nonstick tube and mixed with two pmol primer emITS1 (5’- TGGTAGAGAATGATGGCTGTTG-3’), 770 nM dNTPs in a final volume of 13 µL. The reaction was incubated at 65C for 5 minutes then incubated on ice for at least 1 minute before proceeding. To the mixture was added 5x First Strand Buffer (1x final, ThermoFisher Scientific), 100 mM DTT (5 mM final, ThermoFisher Scientific), 1 uL RNaseOUT (ThermoFisher Scientific) then 200 units of SuperScript III RT (ThermoFisher Scientific). After thorough mixing, each tube was incubated at 55C for 60 minutes to reverse transcribe then 70C for 15 minutes to heat kill the transcriptase. To amplify *E.
**muscae**-specific cDNA, 2 uL of the reverse transcription reaction was mixed with GoTaq 2x colorless mastermix (1x final, Promega) and 500 nM each primers emITS1 and emITS4 (5’-GCCTCTATGCCATTGCTT-3’) then run on a thermocycler with the following settings: 95C for 5 min followed by 35 iterations of 95C for 30 seconds, 61C for 30 seconds then 72C for 30 seconds then 72C for an additional 10 minutes. Four µL of each reaction was analyzed by gel electrophoresis in 1% agarose.

**RNAseq: in vivo timecourse and in vitro culture**

RNA was prepared from each thawed sample by homogenizing with an RNase-free pestle (Kimble Chase), washing the pestle with 750 uL Trizol, then proceeding using the manufacturer’s protocol. RNA was quantified using a Qubit Fluorometer (Qubit RNA HS assay kit, ThermoFisher Scientific) and quality was checked by running on a RNA 6000 Pico chip on a Bioanalyzer 2100 (Agilent Technologies). High quality RNA was then treated with Turbo DNase (ThermoScientific) per the manufacturer’s protocol. RNAseq libraries were prepared with the TruSeq RNA v2 kit (Illumina) using 500 ng of input RNA per sample. Samples were multiplexed 21 samples to a lane and sequenced using 100 bp paired-end reads on a HiSeq 4000 (in vivo samples) or multiplexed 18 samples to a lane and sequenced using 150 bp paired-end reads on a HiSeq2500 at the QB3 Vincent J. Coates Genomic Sequencing Facility at UC Berkeley.

**E. muscae** reference transcriptome (TRINITY3.1.1) assembly

TRINITY3 was assembled from reads from exposed in vivo time course samples that had first failed to align as pairs to the *D. melanogaster* transcriptome (r6.11, HiSat2) then failed to align as singletons to the *D. melanogaster* genome (r.611, bowtie2) using TRINITY with the developer’s recommended settings. After assembly, all in vivo time course reads were aligned to TRINITY3 to assess contamination of non-*E. muscae* sequences. All TRINITY3 transcripts were searched using blastn for homology (evalue 1e-50 or smaller) to organisms not annotated as fungi or Twyford virus. These transcripts were removed to generate TRINITY3.1. All in vivo time course reads were aligned to TRINITY3.1 to assess contamination of non- *E. muscae* sequences. Transcripts that were not expressed by any sample (TPM = 0) or where TPM of uninfected samples accounted for more than 10% of TPM summed across all samples were removed to generate TRINITY3.1.1. Transcriptome completeness was estimated by BUSCO v1.1 analysis using the fungal reference set (1438 BUSCOs).

**RNAseq data analysis**

To calculate gene expression, reads were pseudo-aligned to the appropriate reference (dmelDBGP6.rel85 for *D. melanogaster* or TRINITY3.1.1 for *E. muscae*) and transcript abundance was estimated using Kallisto (171). Data were analyzed using hierarchical clustering by gene (Cluster 3.0), ANOVA between grouped treatments (scipy.stats) and GO term analysis (Panther(108)). Hierarchical clustering heatmaps were generated in Java TreeView; other data were plotted in matplotlib (Python) or Excel 2013 (Microsoft).
Paraffin embedding and microtomy of whole flies

Two mock and two exposure vials were started daily for seven days each with 50 CantonS WF flies 0-1 days old with either 0 (mock) or 6 (exposure) cadavers embedded in AS. Flies were incubated for the first 24 hours at 18°C confined to 2 cm with cadavers, then moved to 21°C where the confinement was relieved. Flies were transferred to GB+ at 48 hours where they continued to be housed at 21°C. Vials were sampled every 24 hours for (Figure 6.S1B). Flies were fixed for 24-36 hours in ice-cold Carnoy's (6:3:1 ethanol:chloroform:glacial acetic acid) at 4°C then dehydrated by stepping through a series of increasing ethanol concentrations. Samples were then transitioned into Histoclear (National Diagnostic) before slowly introducing Paraplast (Sigma). Samples were infiltrated with Paraplast for at least 84 hours at 60°C with gentle shaking before embedding in base molds with embedding rings (Thermo Scientific) and drying overnight. Samples were stored at room temperature until they were sectioned at 8 µm with an RM2255 microtome (Leica), applied to Polysine slides (ThermoFisher Scientific) and dried overnight at 42°C. Sections were stored at room temperature for up to three weeks before Safranin O/Fast Green FCF staining or up to one week before fluorescence in situ hybridization (FISH).

Safranin O/Fast Green FCF (SFG) staining of paraffin sections

Slide-mounted sections were dewaxed with two, 10 minute changes of Histoclear then rehydrated to 70% ethanol with a decreasing ethanol series. Sections were then stained one-at-a-time following Johansen’s Safranin and Fast Green protocol (172) then checked under a dissecting scope before mounting in DEPEX mounting medium (Electron Microscope Sciences) and drying overnight. Slides were imaged at 10x-40x on a Nikon 80i compound microscope using a Nexus 5x mobile phone (Google).

Fluorescent in situ hybridization (FISH) of paraffin sections

Slide-mounted sections were dewaxed with two, 10 minute changes of Histoclear then rehydrated to 70% ethanol with a decreasing ethanol series. Slides were incubated in 0.2 M HCl at 37°C for 45-60 minutes and rinsed in DI water before applying 80 µL of hybridization solution (20 mM Tris-HCl pH 8.0, 0.9 M NaCl, 0.01% sodium dodecyl sulphate, 30% formamide) containing 200 pmol/µL of an E. muscae-specific DNA probe (AlexaFluor633-5'-TGCCTAAAACAGCACAGTT-3', ThermoFisher Scientific). Slides were incubated overnight in a humid chamber at room temperature. The following day, slides were briefly washed in 1x PBS with 0.3% Triton-X100, rinsed in 1x PBS and mounted in ProLong Gold with DAPI (ThermoFisher Scientific). Slides were cured for 24 hours before imaging on a LSM 800 confocal microscope (Zeiss).
Supporting Information

Figure 6.S1. Sampling scheme for time course experiments. A) Sampling scheme for RNAseq time course and reverse transcription. Three vials of each control (“mock exposure”) and exposure conditions were set up using 30 newly-eclosed (less than 24 hours old) flies. Individual female flies were sampled every 24 hours and immediately deposited in Trizol (ThermoFisher Scientific), flash frozen in liquid nitrogen and stored at -80°C until RNA extraction. Asterisk indicates that 1 cadaver was sampled per vial. For 96 hour samples, the other sampled individual was exposed (living). Only one sample was collected from each experimental vial at 120 hours; these were all cadavers. B) Sampling scheme for histology time course. A total of 7 sets of duplicate vials were set up for each control (“mock exposure”) and exposure conditions. Duplicates were set up on the same day using the same starting pool of healthy flies. A minimum of 50 flies were sampled at each indicated time point (i.e. one vial was sacrificed). Paired, unsampled replicate vial was
kept to monitor through 168 hours to ensure high percentage (exposure) or no ("mock" exposure) mortality among flies in those vials. For exposure vials 96 hours and beyond, all flies over both vials were sampled with exposed, living flies separated from cadavers for fixation. All flies were fixed overnight in Carnoy’s solution at 4C then dehydrated in a series of increasingly concentrated ethanol over the course of 10 hours. A subset of 6 flies for each condition were subsequently taken through the paraffin embedding process; the remaining flies were stored at 4C in 100% ethanol for later use.

Figure 6.S2. Genes that show opposite trends during late *E. muscae* CNE1 infection (72 hours) and *B. bassiana* infection (96 hours) of *D. melanogaster* (169). A) Genes that are overexpressed in *E. muscae* CNE1-infected animals versus *B. bassiana*-infected animals. B) Genes that are under-expressed in *E. muscae* CNE1-infected animals versus *B. bassiana*-infected animals. For both plots, Em3 = log2 average ratio of *E. muscae* CNE1-exposed animals to unexposed controls at 72 hours; Bb96 = log2 ratio of *B. bassiana*-infected animals to unexposed controls at 96 hours. All genes shown are differentially expressed between *E. muscae* CNE1-infected animals and controls at 72 hours after exposure with an ANOVA p-value of less than 0.001. Genes that with green asterisks (*) are differentially expressed between all exposed animals versus unexposed animals from 24-72 hours.
Table 6.S1. *E. muscae* transcriptome assemblies.

<table>
<thead>
<tr>
<th>E. muscae transcriptome assembly</th>
<th>Reads used</th>
<th>Parent assembly, if derived*</th>
<th>Subtracted transcripts, if derived*</th>
<th>Number of transcripts</th>
<th>Average % alignment control samples#</th>
<th>BUSCO analysis^</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRINITY1</td>
<td>E. muscae <em>in vitro</em> mRNA library</td>
<td></td>
<td></td>
<td>71,418</td>
<td>41%</td>
<td></td>
</tr>
<tr>
<td>TRINITY2</td>
<td>E. muscae <em>in vitro</em> mRNA library reads that failed to align to <em>D. melanogaster</em> transcriptome or genome</td>
<td></td>
<td></td>
<td>63,378</td>
<td>16%</td>
<td></td>
</tr>
<tr>
<td>TRINITY2.1</td>
<td></td>
<td>TRINITY2</td>
<td>TPM &gt; 1 in control samples only, tBLASTx value &lt;1e-50 to off target reference</td>
<td>62,586</td>
<td>2%</td>
<td>30.5%</td>
</tr>
<tr>
<td>TRINITY3</td>
<td>E. muscae timecourse (<em>in vivo</em>) reads that failed to align to <em>D. melanogaster</em> transcriptome, exposed samples only</td>
<td></td>
<td></td>
<td>195,072</td>
<td>44%</td>
<td></td>
</tr>
<tr>
<td>TRINITY3.1</td>
<td></td>
<td>TRINITY3</td>
<td>tBLASTx value &lt;1e-50 to off target reference</td>
<td>114,258</td>
<td>14%</td>
<td></td>
</tr>
<tr>
<td>TRINITY3.1.1</td>
<td></td>
<td>TRINITY3.1</td>
<td>Removed transcripts that have a) TPM sum of 0 across all samples, b) avg TPM of control &gt;10% of avg TPM non-control</td>
<td>86,509</td>
<td>3%</td>
<td>42.6%</td>
</tr>
</tbody>
</table>

*Derived assembly ≠ not a de novo assembly; achieved via removing transcripts from a previous assembly
# Control samples are unexposed flies
^ Results are listed as percentages of complete single-copy genes | complete duplicated genes | fragmented genes | missing genes as determined by BUSCO analysis for fungal genes
Chapter 7: Progress towards sequencing the genome of *E. muscae* CNE1

This work detailed in this chapter was a joint effort between Carolyn Elya, Michael Bronski and Michael Eisen. The chapter itself was written by Carolyn Elya.

Abstract

*Entomophthora muscae* CNE1 is a fungal pathogen that infects, behaviorally manipulates and ultimately kills *Drosophila* species, including *D. melanogaster*. *E. muscae* belongs to a basal fungal phylum, Entomophthoromycota, for which genomic information is sorely lacking. As part of establishing the *D. melanogaster-E. muscae* CNE1 system, we sequenced genomic libraries from *E. muscae* CNE1 DNA using a variety of methods. Ultimately, we were able to assemble a genome via 10x Genomics sequencing, yielding an assembly with a contig N50 of ~38 kb with scaffold N50 varying according to sequencing platform. The genome is extremely repetitive (~83% repeats) and is likely diploid. We are currently working to improve our 10x assembly using long-read PacBio data.

Introduction

We know next to nothing about the content and organization of genomes from early-branching fungal lineages, especially in Entomophthoromycota. Within the entire phylum Entomophthoromycota (recently established from Zygomycota in (173)), only one organism’s genome sequence has been reported, *Conidiobolus coronatus* (174). The *C. coronatus* genome is 40 megabases (Mb), a typical size for a fungal genome (175). *E. muscae* CNE1 is a fungal pathogen within the Entomophthoromycota that infects, changes the behavior of then kills *D. melanogaster*. I have previously demonstrated how *E. muscae* CNE1 was initially isolated and can be maintained in the laboratory both *in vivo* and *in vitro*. Following this work, I demonstrated that *E. muscae* CNE1 behaves consistently with what has been described in the *E. muscae* literature and, through molecular and histological time course experiments, demonstrated how infection progresses within *D. melanogaster*. Here, we describe the steps taken to achieve a *de novo* *E. muscae* genome assembly for the purposes of beginning to establish *E. muscae* CNE1 as a molecular model.

Results

The path to an assembly

Given the small size of virtually all other available fungal genomes, we set out with the expectation that the genome of *E. muscae* would be similarly sized (~50 Mb) and developed a sequencing strategy for *de novo* assembly of a small genome (175). We elected to use a hybrid approach for assembly using short reads from small insert libraries (Illumina HiSeq) and long reads from high-molecular weight DNA libraries (PacBio).

The first issue we encountered was in extracting high molecular weight DNA. We tried three different commercial kits developed for this purpose (QIAGEN Genomic Tip,
QIAGEN Mag Attract HMW and Gentra Puregene Tissue) following protocols for extraction from fungal or plant material and consistently achieved polysaccharide contaminated DNA. This DNA was generally of high molecular weight, but we received low yields which each of these kits. Cleanup efforts tended to be lossy and result in shearing of the DNA. We were able to overcome these issues by turning to a protocol for DNA extraction from *Trichoderma* and *Gliocladium* (176). This protocol tended to be variable in terms of yield, likely a result of starting with cultures of varying growth stages. Unfortunately, as *E. muscae* grows heterogeneously in liquid culture, we could not rely on spectrophotometric analyses to analyze culture growth stage and instead relied mostly on gross culture morphology to determine which cultures would be used for gDNA extractions. We found that younger cultures, which consisted primarily of spherical growths, were more suited to giving high-yield, low carbohydrate contamination DNA samples. As cultures age, they tend to form elongated hyphal structures. Cells at this point become very hard to lyse, lowering DNA yield and increasing the problem of polysaccharide contamination.

Having overcome this initial obstacle, we proceeded with our sequencing strategy. We began preparing short-insert Illumina libraries in house. Surprisingly, short-insert library preparation was another technical challenge. Using Illumina’s PCR-free DNA sequencing library kit, we generated what appeared to be high-quality libraries of 350 bp and 550 bp inserts. However, these libraries clustered poorly on the HiSeq2500 at our local sequencing facility (QB3), resulting in a dataset highly contaminated with adapter sequences (16.76% and 40.45% bases were identified as part of adapters from the two libraries). Clustering failed completely during a subsequent attempt to rerun these libraries. While sparse, our PCR-free Illumina data allowed us to make an initial estimate of genome size as 1.3 gigabases (Gb), approximately 30-fold larger than the *C. coronatus* assembly (Figure 7.S1).

Simultaneously, we had submitted DNA to the HHMI PacBio facility at University of Washington for PacBio library generation and sequencing at what we expected to be approximately 30x coverage (2 SMRT, with ~1 Gb per SMRT cell). Our PacBio sequencing yielded high-quality data, with 203,871 reads with a mean length of 13,454 bp resulting. Surprisingly, when we BLASTED a randomly chosen subset of the PacBio reads we found high homology to Twyford virus (Michael Eisen, personal communication). The sequence deposited as Twyford virus was discovered via genomic sequencing of wild *Drosophila* (177). Given that Twyford virus itself has sparse homology to other known viruses but is an excellent match for *E. muscae* CNE1, we believe that Twyford virus (and likely other, as yet unidentified viruses) is actually *E. muscae*.

Having significantly underestimated the size of the *E. muscae* genome, we would no longer be able to achieve an assembly with just our low-yield Illumina data with our low coverage PacBio data (now ~1.5x coverage), but instead would have to rely more heavily on short read sequence data for assembly. For unknown reasons, it appeared that a PCR-free Illumina sequencing approach was incompatible with our DNA preparations, so we moved forward by preparing a short insert library (350 bp) using a PCR-dependent approach. This library was successfully sequenced (1 lane, HiSeq4000, 100 PE). Strangely, though these reads were by all measures of high quality, we discovered that they contained a large number of error kmers (kmers only occurring once in our entire set) that could not be corrected by standard methods. Still, as we now had enough data to attempt a de novo assembly, we did so. This attempt failed.

Given the size and the large amount of repeat content in the genome, we next sought to use a short-read sequencing approach that would give us long distance information that
could be combined with our already existing data to yield an assembly. The obvious solution was to generate and sequence mate pair libraries. Such libraries are not cost-effective to produce in-house as the require very specialized reagents and are very technically challenging to be done correctly, so we solicited a genomics to produce them. In response, the core suggested we instead have them prepare a 10x genomics library, which offered the advantages of requiring less input material and yielding long distance information in a strand-specific manner. This is achieved through microfluidics; high molecular weight DNA are separated into individual molecules using oil and combined with reagents necessary for barcoding. The final library consists of fragments that contain a universal barcode (for the purposes of multiplexing samples on a flow cell lane) and barcodes that correspond to the parent DNA molecule from which each fragment is derived. A 10x-specific assembler, Supernova, is used to take advantage of this strand-specific information and generate a phased assembly. Given that we had no information about the ploidy of our organism, haplotype information was extremely appealing, and we opted to adopt this approach instead.

One lane of sequencing the resultant 10x genomics library on the HiSeq4000 (150 PE) yielded a successful assembly of the expected size and reasonable quality (assembly 10X.1), despite the reverse reads being poorer quality than expected (Table 7.1). As with our previous HiSeq4000 run of our PCR-dependent, short-insert library, we observed that the majority of the unique kmers in our data were uncorrectable error kmers that nevertheless had high quality scores. Given the issues with both read two quality and recalcitrant error kmers, we performed a subsequent rerun of our 10x library on the HiSeq2500 (150 PE). This run yielded higher quality reads and gave an assembly of similar quality to that from the HiSeq4000 data (assembly 10X.2, Table 7.1). As expected, the input read metrics are noticeably better for our second assembly. Output assembly metrics are similar for 10X.1 and 10X.2: though 10X.2 gives more long contigs, 10X.1 has a higher scaffold N50. Moving forward, we opted to work with 10X.1 on the premise that higher N50 was indicative of a slightly better assembly.

**Table 7.1. Metrics for 10x genomics assemblies.**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Assembly 10X.1</th>
<th>Assembly 10X.2</th>
<th>Ideal assembly^†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reads</td>
<td>671,290,000</td>
<td>311,050,000</td>
<td>800,000,000 - 1,000,000,000</td>
</tr>
<tr>
<td>R2 with &gt;=Q30 (%)</td>
<td>64.11</td>
<td>78.58</td>
<td>75-85</td>
</tr>
<tr>
<td>Median insert size (bp)</td>
<td>430</td>
<td>400</td>
<td>350-400</td>
</tr>
<tr>
<td>Proper pairs (%)#</td>
<td>83.25</td>
<td>87.68</td>
<td>&gt;=75</td>
</tr>
<tr>
<td>Weighted mean molecule size (kb)</td>
<td>24.75</td>
<td>28.37</td>
<td>50-100</td>
</tr>
<tr>
<td>Duplicate reads (%)</td>
<td>17.40</td>
<td>7.35</td>
<td>0</td>
</tr>
</tbody>
</table>
Features of our working assembly (10X.1)

Our current assembly is quite highly fragmented, with a contig N50 of just 38.3 kb for an overall assembly of 1.01 Gb. The assembly is extremely repeat rich, with at least 83% of bases identified as comprising repeat elements (Table 7.2). Additional repeat elements have been found using manual annotation (Michael Eisen, personal communication), leading us to believe that this percentage is an underestimate of repeat content. Within these repeats, we observe some families of elements that vary in their internal similarity, suggesting a diversity in times of divergence for these elements. In RNAseq data from in vitro E. muscae culture, we observed the transcription of some transposable elements, indicating that these elements are still active and consistent with observation of recently-diverged repeat families.

Table 7.2. Repeat element analysis of 10X.1 assembly.

<table>
<thead>
<tr>
<th>Element type</th>
<th>Number of elements</th>
<th>Length occupied (bp)</th>
<th>Percentage of sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SINEs</td>
<td>0</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>ALUs</td>
<td>0</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>MIRs</td>
<td>0</td>
<td>0</td>
<td>0.00%</td>
</tr>
</tbody>
</table>
BUSCO analysis of the assembly shows relatively poor completeness (Table 7.3), though whether this analysis reflects the actual state of the assembly or merely the phylogenetic distance from the fungi used to generate the BUSCO reference set is unknown.

Table 7.3. BUSCO statistics for 10X.1 assembly (Fungal reference)

<table>
<thead>
<tr>
<th></th>
<th>Complete BUSCOs</th>
<th>159 (54.8%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete and single-copy BUSCOs</td>
<td>85 (29.3%)</td>
<td></td>
</tr>
<tr>
<td>Complete and duplicated BUSCOs</td>
<td>74 (25.5%)</td>
<td></td>
</tr>
<tr>
<td>Fragmented BUSCOs</td>
<td>41 (14.1%)</td>
<td></td>
</tr>
<tr>
<td>Missing BUSCOs</td>
<td>90 (31.1%)</td>
<td></td>
</tr>
<tr>
<td>Total BUSCOs</td>
<td>290</td>
<td></td>
</tr>
</tbody>
</table>

In order to understand the ploidy of our organism, we aligned our genomic reads to single-isoform BUSCOs (SIBs), either called as single-copy or duplicated from BUSCO analysis, from our TRINITY2.1 assembly. What we observed is consistent with a diploid genome. First, there are about twice as many reads that map to the “single copy” SIBs as to the “duplicated” SIBs (Figure 7.1A). This is what we would expect to see if the “single copy” SIBs are actually collapsed homologs. Secondly, there are far more single nucleotide polymorphisms (SNPs) present in the “single copy” as opposed to “duplicated” SIBs (Figure 7.1B). Again, this is consistent with a scenario in which homologous SIBs were collapsed in the pseudo-haplotype output of the Supernova assembler. Importantly, while there are SNPs observed in the “duplicated” SIBs, these SNPs are only present on one of the two
scaffolds to which those SIBs align, consistent with a model in which there are two copies genuinely present of these “duplicated” SIBs in our assembly.

**Figure 7.1. Analysis of single-isoform BUSCOs (SIBs) from TRINITY2.1 transcriptome within 10X.1 genome assembly.** A) Coverage of SIBs according to designation as “single copy” or “duplicated” within genome assembly. B) Frequency of single nucleotide polymorphisms (SNPs) in SIBs present as either one or two copies in the genome assembly.

Aligning long reads to our assembly (sequencing an additional 5 SMRT cells has brought us to ~4.3x PacBio coverage for an estimated genome of 1.2 Gb) reveals substantial disagreements. While a large percentage of our reads map to the assembly and yield a high-quality mapping score (91.3% align with MAPQ=254), these alignments require extensive soft-clipping (i.e. exclusion of of bp 5’ or 3’ end of a PacBio read) in order to achieve this (Figure 7.2A). If we increase our stringency in what we consider a good alignment, looking not only at mapping quality but also excluding any reads where soft clipping over 1,000 bp occurred, only 58.9% aligned reads meet these criteria. Furthermore, reads that failed to align to our assembly at all (MAPQ=0) are not the “dregs” within our dataset; reads that fail to align to the assembly show virtually the same size distribution as reads that do align (Figure 7.2B). That so much soft-clipping is necessary in order for our long PacBio reads (which, while having a higher chance of error at each individual base should be correct in terms of order of elements) and that reads of substantial length (up to 30 kb) do not align at all to the assembly could indicate that contigs were erroneously misassembled.
Figure 7.2. Features of PacBio reads the align to assembly 10X.1. A) BLASR mapping quality of all PacBio reads to assembly 10X.1. B) Length distribution of all PacBio reads (including unaligned reads, top) and all PacBio reads that align to the 10X.1 assembly with a MAPQ score of 0 (bottom).

Discussion

Despite our expectation that the genome of *E. muscae* CNE1 would be relatively small and straightforward to assemble, we find ourselves grappling with improving an unexpectedly large working assembly for this fungus. Our present and ongoing efforts to complete the genome will be described below.

Reconsidering the 10X.2 assembly

Though at first appearance it seemed that choosing to move forward with either 10X.1 or 10X.2 was a toss-up, it seems like assembly 10X.1 might actually have been a poor choice. There appears to be an incompatibility between Supernova (the 10x assembler) and HiSeq4000 data. The HiSeq4000 is not the recommended platform for 10x sequencing, but, unfortunately, earlier platforms with more reliable chemistry have been discontinued and the facilities closest at hand are only equipped with this sequencer. It was on their recommendation that we proceeded with HiSeq4000 sequencing. We are still without
explanation as to why we observe a significant number of uncorrectable error kmers from identical libraries sequenced on the HiSeq4000 versus the HiSeq2500. That the scaffold N50 is so much larger for the 10X.1 assembly over the 10X.2 assembly might indicate that more mis-assemblies are present in 10X.1, potentially as a result of the HiSeq4000’s problematic chemistry. Future work should reconsider assembly 10X.2 as a better starting point than 10X.1 for scaffolding and annotation.

**Improving our assembly**

We are currently employing two approaches to improve our 10x assembly. The first approach is to use our low-coverage long-read PacBio data to scaffold a 10X-only assembly. Our efforts had been directed at 10X.1 but are being reconsidered in light of recent considerations regarding the HiSeq4000. Still, regardless of which 10x-only assembly we start with, we still have the problem that it’s substantially fragmented. Though the 10x approach is able to use long-range information in assembly, it is still a platform based on short read data. As such, 10x is still likely stymied by repetitive genomes. The long reads generated by the PacBio platform are probably our best avenue in terms of assembling such a repeat-rich genome, but sequencing at a high enough coverage to complete a PacBio-only assembly is not financially feasible. At present, we have ~4.3x PacBio coverage, which we are using after Supernova assembly to join contigs. Ideally, we would use these long-read data during, rather than after, the initial assembly in generate a correctly-ordered assembly. We are presently looking to use the hybrid assembler MaSuRCA which uses both overlap consensus and de Brujin graph approaches to assemble reads of varying lengths and error rates (178). We hope that by incorporating the long-range information we can prevent hard-to-correct mis-assemblies that may occur by relying on short-read data alone.

Our main goal for an *E. muscae* CNE1 assembly is to have a genome in which the majority of coding content has been properly sequenced and assembled. The most pressing question is what the *E. muscae* CNE1 genome encodes for and what can what functional information can we glean using homology of studied organisms. Genes that encode products that are exported or with similarity to neural effectors would be of particular interest, as they would be likely candidates in mediating behavioral manipulation of the host.

**Methods**

**DNA extraction**

Large molecular weight DNA was extracted from *in vitro* culture using the method of Bulat et al, 1998 (176). Twenty mL of spherical-phase *E. muscae* CNE1 cultures were loosely (2800g, 15 min, 4C) then tightly (21,000g, 2 min, room temperature) pelleted. After decanting supernatant, cells were flash frozen in liquid nitrogen then homogenized with 1.5 mL pestles (Kimball Chase) and resuspended in 400 µL buffer A (50 mM Tris pH 7.8, 50 mM EDTA, 150 mM NaCl, 2.5% N-lauryl sarcosine, 500 mM 2-mercaptoethanol, 600 µg/mL proteinase K). Samples were digested with 600 µg/mL proteinase K (NEB) for 4 hours at 65C then digested with 100 µg/mL RNase A (ThermoFisher Scientific) at 37C for 30 minutes before storing overnight at 4C. Concentrated NaCl was added to bring the salt concentration to 1 M before subjecting to a two to four rounds of chloroform
(CHCl₃):octanol 24:1 extraction where sample where incubated for 15 minutes at room temperature in an equal volume of CHCl₃:octanol before pelleting 2 min at 12,000g at room temperature. DNA was precipitated from pooled aqueous supernatant with 0.6 volumes of isopropanol, incubating 30 minutes at room temperature. The DNA pellet was washed with 70% EtOH before air-drying and resuspending in TE buffer (10 mM Tris pH 8.0m 1 mM EDTA) with a wide bore tip. DNA concentration was quantified using a fluorometric method (Qubit dsDNA HS assay kit, ThermoFisher Scientific), polysaccharide contamination was assessed via spectrophotometry (Nanodrop, ThermoScientific) and length was determined by gel electrophoresis in 0.4% agarose.

Library preparation and sequencing

Genomic DNA libraries were prepared as in Table 7.4. All libraries made in house were quantified using a fluorometric method (Qubit dsDNA HS assay kit, ThermoFisher Scientific) and run on an HS DNA Bioanalyzer chip (Agilent) to assess quality before sequencing submission.

Table 7.4. *E. muscae* CNE1 sequencing libraries and platforms.

<table>
<thead>
<tr>
<th>Input DNA</th>
<th>Library preparation method</th>
<th>Library preparer*</th>
<th>Sequencing platform and type</th>
<th>Sequencing facility*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 µg</td>
<td>TruSeq DNA PCR-Free</td>
<td>CNE</td>
<td>HiSeq2500, 1x100 PE</td>
<td>VCGSL, QB3, UC Berkeley</td>
</tr>
<tr>
<td>100 ng</td>
<td>Wafergen PrepX DNA</td>
<td>FGL, QB3, UC Berkeley</td>
<td>HiSeq4000, 1x100PE</td>
<td>VCGSL, QB3, UC Berkeley</td>
</tr>
<tr>
<td>11.8 µg</td>
<td>SMRTBell (PacBio)</td>
<td>UW PacBio</td>
<td>7 SMRT cells</td>
<td>UW PacBio Sequencing Services</td>
</tr>
<tr>
<td>10 ng</td>
<td>10x Chromium</td>
<td>UC Davis Genomics Core</td>
<td>HiSeq4000, 1x150 PE</td>
<td>UC Davis Genomics Core</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HiSeq2500, 1x150 PE</td>
<td>VCGSL, QB3, UC Berkeley</td>
</tr>
</tbody>
</table>

*FGL = Functional Genomics Laboratory, VCGSL = Vincent Coates Genome Sequencing Laboratory

Assembly and analysis

One lane of 150 PE HiSeq4000 or HiSeq2500 data were assembled using Supernova; output pseudohaplotype ("pseudohap") assemblies 10X.1 and 10X.2, respectively, were subjected to further analysis (10x Genomics). Repeats were called with RepeatModeler and RepeatMasker (Smit, AFA, Hubley, R & Green, P., [http://www.repeatmasker.org](http://www.repeatmasker.org)). BUSCO v2 analysis was run on assembly 10X.1 using the fungal reference set. Reads form the HiSeq2500 run of the 10x genomics library were aligned to the 10X.1 assembly using Bowtie2 to determine genomic coverage of SIBs. Bowtie2 alignments were then processed to vcf files and used to call unfiltered SNPs. PacBio circular consensus sequencing reads were aligned to the *E. muscae* CNE1 10X.1 assembly with BLASR, using setting optimized for the human genome (179).
Figure 7.S1. Kmer frequency spectrum for HiSeq2500 PCR-free library run #1. Kmer (k=31) analysis from small insert, PCR-free Illumina libraries shows mean coverage of 7.17x for 9.4x10⁹ bp, giving a size estimate of ~1.3 Gb.
Chapter 8: Concluding thoughts and future directions

Perhaps gut-associated microbes are more important as substrates than colonizers

Since my initial attempts to identify behaviors that depend on gut microbial taxa, a handful of additional behavior studies have been published. In one study, both D. melanogaster larvae and adults preferred food that had been previously colonized by conventional but not axenic larvae (104). Further experiments suggested that the most important factor for larval food preference is previous food churning irrespective of microbial presence (104). Another study examined mating preferences of three Drosophila species based on relatedness (i.e. tested pair were siblings), rearing vial and rearing food each in the presence of absence of antibiotic treatment (180). Differences in mating between antibiotic treated and untreated flies were only observed for relatedness in D. melanogaster. However, these findings are confounded in two ways. First, the experimenters claim that they prevented transmission of bacteria between mothers and progeny by separating adults and larvae (180). However, it’s been well-established that microbes are transmitted vertically via deposition on the eggshell (96). Since the embryos in this experiment were not sterilized, they received their mothers’ bacteria, contrary to the experimenter’s expectations. Secondly, it was unclear exactly how the animals were treated with antibiotic (at the start or end of development) and whether they were colonized with Wolbachia: both of these factors could influence the outcome. A more recent study tried to replicate the experiments of Sharon et al, 2010 and found that the assortative mating phenotype was more subtle and strain-dependent than initially suggested (105).

The most compelling of the behavioral studies to date is work showing that flies are more attracted to co-cultures of S. cerevisiae and Acetobacter than either microbe alone (181). Using a synthetic mixture of the attractants produced by these co-cultures, the experimenters were able to recapitulate fly attraction and through mutant studies demonstrated that Or42b was needed for this attraction. The authors also demonstrated that fly attraction to the co-culture depends on the ability of Acetobacter to metabolize ethanol. Furthermore, they provided evidence as to why this preference might have evolved in flies, by showing that a) yeast/Aacetobacter co-culture can exclude contaminant microbes (e.g. molds) and promote larval survival and b) ethanol levels in yeast/Aacetobacter co-culture are dampened compared to yeast monoculture, which likely is advantageous to larvae. These results suggest a key role for microbes in shaping the environment in which flies live and providing a readout (i.e. volatile production) that flies can use to select ideal substrates.

Consistent with this, recent work by Bill Ja’s group also suggests that microbes may play their most important role as substrate, in this case by providing nutrition to developing larvae (119). Taking this finding and and looking back to Shin et al 2011 and Storelli et al 2011, the papers that found A. pomorum and L. plantarum, respectively, to rescue larval growth under low nutrient conditions, it should be noted that microbial loads of the media were not tested (41, 42). Therefore, it’s possible that the observed influence on host development was due to these strains’ ability to grow substantially better on fly media and thereby providing more nutrition on a yeast-limited diet. Both studies focused on occupancy of the host rather than the media as the key factor in rescuing growth and, in doing so, may have missed a key variable in the experiment.
Interestingly, both the Shin et al and Storelli et al studies found that the rescue of delayed growth on yeast-poor diets were dependent on insulin or TOR signaling, respectively, which are well-known growth pathways in the fly (41, 42). Shin et al found that though PQ-ADH A. pomorum mutants (who could colonize flies equally well as wild-type) could not restore development, but adding acetic acid along with these bacteria could rescue growth (42). Crucially, providing just acetic acid to axenic flies reared on nutrient-poor food did not rescue larval growth, suggesting that the bacterial product was indirectly related to improved development. An explanation of this observation could be that acetic acid is required for A. pomorum to grow on the fly’s food; perhaps in the absence of this essential nutrient, bacteria cannot grow to high-enough titer to rescue development. Ultimately, the question of whether any gut bacterial taxa play a more nuanced role in fly biology could be assessed by consistently assessing bacterial levels on the medium and performing experiments in parallel that use heat-killed (or otherwise unviable) bacteria.

Overall, the D. melanogaster-gut microbe literature is consistent with a seldom-considered hypothesis, that gut-associated bacteria may be most important for flies as substrates rather than colonizers. Given the evidence, it seems likely that the relationship between fruit flies and microbes may not precisely parallel that between vertebrates and their microbiota. Even so, the study of fruit flies has been incredibly useful in revealing many general principles in biology; understanding their relationship with commensal bacteria and yeasts can move us to further understand the natural world. That said, we are best served by this or any model when it is taken as what it is, not what we want it to be. It is therefore crucial that, as we continue to explore the relationship between flies and microbes, we consider all explanations that are consistent with our data and reject simple ones only when there is evidence to do so.

The promise of E. muscae

I’m still shocked that E. muscae was infecting fruit flies in my own backyard. Having been observed several times now in Northern California and reported in the eastern coast of the United States as well as Europe, I imagine that E. muscae can be found most everywhere one can find humans and compost piles (146, 147). That it’s never been studied in fruit flies is baffling, but there could hardly have been a more exciting time to find it. With our current knowledge and repertoire of genetic tools in the fruit fly, the accessibility of cheap sequencing, and, of course, CRISPR-Cas9 and its derivative technologies, the D. melanogaster-E. muscae system offers the scientific community a way to understand behavioral manipulation like never before. There are too many possible experiments to list; I hope that others take advantage of this system (the E. muscae literature could definitely do with an update!)
Bibliography

65. Cohn F (1855) *Empusa muscae und die krankheit der Stubenfliegen* Available at: All Papers/C/Cohn 1855 - Empusa muscae und die krankheit der Stubenfliegen.pdf.


93. Staubach F, Baines JF, Künzel S, Bik EM, Petrov DA (2013) Host species and


118. Johansson ME V, Larsson JM, Hansson GC (2011) The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. Proc Natl Acad Sci 108(Supplement 1):4659–4665.


130. Tunstall NE, Sirey T, Newcomb RD, Warr CG (2007) Selective pressures on


159. Akitake, Bradley; Ren, Qiuting; Boiko, Nina; Ni, Jinfei; Sokabe, Takaaki; Stockand, James D; Eaton, Benjamin A; Montell C (2015) Coordination and fine motor control depend on Drosophila TRPy. *Nat Commun* 6:7288.


179. Chaisson, Mark J P; Huddleston, John; Dennis, Megan Y; Sudmant, Peter H; Malig, Maika; Hormozdiari, Fereydoun; Antonacci, Francesca; Surti, Urvashi; Sandstrom, Richard; Boitano, Matthew; Landolin, Jane M; Stamatoyannopoulos, John A; Hunkapiller, Michael W EE (2014) Resolving the complexity of the human genome using single-molecule sequencing. *Nature* 517(7536):608–611.


Appendix

I) Protocol: Propagating *in vivo Entomophthora muscae* CNE1 infection in *D. melanogaster*

**Materials needed:**
- Clean petri dishes (100 x 15 mm)
- Whatman paper
- Deionized water
- Paintbrush
- Small forceps (e.g. Dumont 55)
- Big forceps (6” long, blunt)
- New empty vials, wide (FlyStuff/Genessee)
- New autoclavable Droso-plugs (FlyStuff/Genessee)
- New cotton flugs (FlyStuff/Genessee)
- 1.5% agar, 10% sucrose solution (AS)
- Vials of GB+ diet
- Microwave
- Metric ruler
- Marker
- 18C incubator on inverted light cycle (12:12 light:dark, transition to light:dark 2-3 hours before you are available to maintain infection)
- 21C incubator on inverted light cycle (see above)
- Kimwipes
- Young CantonS WF flies (under 5 days old)

**Day 0**

**Collect new cadavers from vials that are 4-5 days old, 1:30 pm – 5 pm**

1. Retrieve flat of 4 day + vials from incubator and place under foil-lined box on cart. Remember that these flies are entrained to light from 12 am (midnight) – 12 pm (noon). We want to keep them on that light cycle with minimal interruptions.

2. Prepare a small petri dish (35 mm x 10 mm) with a piece of Whatman paper wetted with DI water.
   a. This is where cadavers will be collected from each vial, one by one, to get an accurate count.

3. Prepare a clean petri dish (100 mm x 15 mm) with a piece of Whatman paper wetted with DI water (helps mitigate static).
   a. This is where all cadavers will be pooled

4. Retrieve one vial from the foil-lined box.
a. While gently handling vial, record number of cadavers stuck to flug, side and medium on vial data sheet.

5. Use CO2 to knock out living flies and tap onto pad. (Remember that once you apply CO2, all flies stop moving; it can become difficult to distinguish live flies from dead flies. It is easier to distinguish cadaver from living flies if you wait 2-3 hours after lights out in the incubator.)

6. Using forceps and/or paintbrush (lightly wetted in DI water), remove all cadavers from vial and place in small petri dish.

7. Record final count of cadavers on data sheet, also noting any flies that are dead but not infected (i.e. no fungal banding on dorsal abdomen).

8. Discard non-infected flies in morgue. Transfer cadavers to large petri dish.

9. Repeat steps 4-8 with all available vials.

10. Return flat of vials to incubator.
   a. Leave leftover dish of cadavers on bench overnight to make sure that they sporulate.
   b. Wipe down the pad, gun and desk near the CO2 tank with 70% EtOH.

**Embed new cadavers**

11. Melt AS until molten using microwave (usually 30 seconds – 1 min).
    a. Be careful that the AS doesn’t boil over (this is very easy to do!)

12. Transfer ~20 mL of AS into a clean petri dish (100 mm x 15mm) and allow it to cool without a lid

13. Jiggle the petri dish every minute or so to assess hardness of agar. When agar has just set (no longer moves when jiggled), use small forceps to embed 6 good cadavers, head-down, into AS medium in a circle. Do this for each vial you plan to start.

   a. Push any wings that are sticking up into agar to move out of the line of fire of conidiophores.
   b. A “good” cadaver is one that is: big and puffy with clear fungal banding (preferably symmetric) and has raised wings.
14. Wait 2-3 minutes after placing last cadaver for AS to completely harden then use a vial as a cookie cutter to cut a circle slightly smaller than the vial diameter around the cadavers.
15. Transfer this disc containing the cadavers to a new vial using large forceps. Make sure cadavers are facing up, or else the flies won’t be exposed.

a. The leftover AS in dish can be discarded in the non-medical waste to be autoclaved.

16. Use a ruler to measure 2 cm above the top of the agar and mark with a pen.

17. Place a new Drosophila plug near each vial.
18. Fill out a log sheet for each new vial you start (blank sheets are behind the last tab in the binder).
   a. If you anticipate more than an hour difference between adding embedded cadavers and live flies to vial, record the time cadavers were added to vial.

Expose more WT flies
19. Using CO2, anesthetize young CantonS WF flies (<5 days post-eclosion) and count out 50 total.
a. Don’t worry about a perfectly even sex balance. However, if you notice that you’re getting a lot more of one sex than the other, try to include more of the neglected sex to compensate.

b. Avoid flies that are abnormally small or have wing defects.

20. Transfer the flies to one vial using a paintbrush making sure that all flies are below the marked 2 cm line.

21. Using the big forceps, push the Droso-plug into the vial until the bottom is a bit below the 2 cm mark then pull up to make the bottom of the Droso-plug even with the mark.

22. Record when flies were added to vial.

**Incubate exposure vials at 18C**

23. Retrieve the “humid chamber” at 18C and rewet kimwipes with DI water.

24. Place the vials in the “humid chamber” at 18C.

25. Refill the 1L beaker of DI water in 18C incubator if it’s getting low (under 400 mL)

**Day 1**

Relieve confinement of exposed flies

1. Using large forceps, pull up Droso-flug so that it’s situated at the top of the vial.
   a. As you do this, briefly check that bottom 2 cm of vials are visibly covered in spores.

2. Move vials to flat at 21C (“Need to MAC”).

3. Retrieve vials of GB+ medium from cold room (next door to GL11), one for each vial whose flug you raised today, and place at 21C to warm for tomorrow.

**Check yesterday’s cadavers**
4. Remove lid from dish of leftover cadavers and check that it is covered in spores. Discard dish with cadavers in biohazard trash under bench.

**Day 2**

*Move exposed flies away from old cadavers onto new food*

1. Transfer flies to GB+ diet by inverting AS vial and tapping onto GB+.
   a. Watch the AS disc – it’s not adhered to the bottom of the vial, so it might fall if you tap too hard. If this happens, you can knock the flies out with CO2 and remove the disc using forceps.
2. Cap GB+ vial with a cotton plug and return to incubator at 21C (“Younger than 4 days”).

**Days 4-7**

*Collect cadavers and propagate in vivo infection*

1. Go back to Day 0 to harvest cadavers and start new vials.

**Other things to note**

- Push Cs WF stocks daily to ensure that you have sufficient 0-1 day adults to start the number of desired vials

- Keep an eye on the humidity in the large incubator (tends to fluctuate between 50-70% humidity, which seems to be fine for the flies). There is a bin of DI water in the bottom to help maintain reasonable humidity.

- All used cadavers and exposed flies or consumables (kimwipes, vials etc.) should be discarded in the biohazard waste under my bench.

- Cadavers and/or spores do not come upstairs to EL151 unless they are a) fixed or otherwise killed or contained (e.g. in an eppy); b) mounted on a slide.

- Labware (especially gloves!) should not be transited back and forth between GL11 and EL151 with the exception of the AS solution, autoclave glove and the cadaver log binder. Please wipe down the binder (front, back and spine) with 75% EtOH before bringing upstairs.
II) Recipe: GB+ medium
for maintenance of Entomophthora in vivo infection
Modified from: https://groups.google.com/forum/#!topic/cct_cognitiveneuroscience/6C6iNjdX0D8

Materials
- 5-6 organic bananas. Can be just ripe (all yellow) or very ripe (yellow with spots)
- MilliQ water
- Agar
- Goggles
- Large stir plate
- 2L beaker
- 2L Erlenmeyer flasks (to hold max 1L each)
- Graduated cylinder
- Blender, cleaned beforehand with soap and water then rinse with MilliQ water
- Metal spatula
- Large stir bar
- IR thermometer gun
- Autoclave
- Propionic acid
- Pipetteman and serological pipettes (1 mL, 5 mL, 50 mL)
- New flat of K-resin, wide-mouthed Drosophila vials (FlyStuff/Genessee)

Protocol
1. Wash the blender with soap and DI water.
2. Peel bananas and weigh peeled fruits on balance. Record total weight.
3. Calculate the volume of medium that can be prepared from this mass of banana using the following equation: $2.5 \times \text{mass} = \text{mL of medium}$.
   a. The final concentration of banana in solution is 40% w/v
4. Measure MilliQ water calculated in step 2 and add ~60% of this volume to 2L beaker.
5. Homogenize the bananas using the blender. **Do not overload the blender with banana;** split the banana into batches and process each batch and transfer to beaker before continuing with the next.
6. When you are done mashing all of the banana, rinse the blender and any other utensils that have banana paste on them with MilliQ water into the 2L beaker until clean.
7. Bring volume in beaker to desired volume with MilliQ water.
8. Turn on heat, add large stir bar and begin stirring banana.
9. While mixture stirs and heats, weigh out BUT DON’T YET ADD enough agar to bring final concentration to 2% (i.e. 2 g per 100 mL liquid)
10. Monitor temperature of banana mixture with IR thermometer gun. When mixture is ~45°C, add in the agar.
11. Cover banana/agar/water mixture and heat until 70-80°C.
   a. You may need to cease stirring when the solution gets very hot to clearly see boiling and to avoid having the solution boil over onto hot plate. You only need to heat until the agar is well-dispersed throughout the solution (and doesn’t fall to bottom when you stop stirring) and is thin enough to pour.
12. Once the solution has reached a boil, transfer to Erlenmeyer flasks for autoclaving. Put no more than 1L in each 2L flask.
13. Autoclave on liquid for the minimum recommended time given the solution’s volume:

```
<table>
<thead>
<tr>
<th>Volume of Liquid in One Container (mL)</th>
<th>Minimum Recommended Sterilize Time at 121°C (250°F) (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>75</td>
<td>25</td>
</tr>
<tr>
<td>250</td>
<td>30</td>
</tr>
<tr>
<td>500</td>
<td>40</td>
</tr>
<tr>
<td>1000</td>
<td>45</td>
</tr>
<tr>
<td>1500</td>
<td>50</td>
</tr>
<tr>
<td>2000</td>
<td>55</td>
</tr>
<tr>
<td>&gt;2000</td>
<td>&gt;55+10mL</td>
</tr>
</tbody>
</table>

*Minimum sterilization times are based on obtaining a 10^−6 Sterility Assurance Level (SAL) with standard test loads. Specific labs may require different sterilization times to achieve this level of sterility, or may require a different SAL.*
```

Source: [https://medschool.vanderbilt.edu/cdb-core-laboratory/how-long-do-i-autoclave-liquids](https://medschool.vanderbilt.edu/cdb-core-laboratory/how-long-do-i-autoclave-liquids)

14. Remove container from autoclave and record the current volume of solution (some volume tends to be lost during autoclaving).
15. Stir solution uncovered with heat turned off, monitoring temperature with IR thermometer gun.
16. When solution reaches ~60°C or below, bring the medium into the hood and add enough propionic acid for a final concentration of 0.3%.
   a. Add 300 uL of propionic acid for every 100 mL medium.
17. Swirl the solution to make sure the propionic acid is homogenously distributed.
18. Near flame, use 50 mL serological pipette to distribute 10-15 mL of medium into new K-resin vials until you run out of medium.
19. Cover the vials with paper towel and tape the towel down to prevent stray flies from coming into contact with medium.
20. Put beaker in sink and fill with water/soap to soak. You will need to come back later to wash this out.
21. Leave vials on bench overnight to dry.
22. The next morning, flush and label vials then transfer to 4°C for storage.
III) Template: Exposure vial record

Vial name: Date:

Flies to be exposed:
#

Age (days): 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 Other:

Strain:

Sex:

How these flies were reared (from embryo)

Medium: Koshland Other:

Light cycle: 12a – 12 p (A) Other:

Temp: 21°C Other:

Humidity: ~60% Other:

Exposure conditions

# of cadavers:

Confinement distance (cm):

Time exposed:

Medium:

Light cycle: 12a – 12 p (A) Other:

Temp: 18°C for 24 hrs > 21 C Other:

Humidity: ~100% 24 hours > ~60% Other:

Other notes:
<table>
<thead>
<tr>
<th>Time</th>
<th>Day 0 (+24 hr)</th>
<th>Day 1 (+48 hr)</th>
<th>Day 2 (+72 hr)</th>
<th>Day 3 (+96 hr)</th>
<th>Day 4 (+120 hr)</th>
<th>Day 5 (+144 hr)</th>
<th>Day 6 (+168 hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Notes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

★ = vial started
FR = Flug raised
MAC = exposed flies moved way from exposure cadavers
NM = exposed flies moved to new medium
NI = not infected (not demonstrating wings up and mycosing phenotype of typical cadaver)
WU = wings up (typical)
WD = wings down (like floppy-eared bunny, common)
WH = wings horizontal (rare for cadaver)