THE GENETICS OF ADAPTATION IN *DROSOPHILA SECHELLIA*

Eric Jay Earley

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biology (Molecular, Cellular, and Developmental).

Chapel Hill
2013

Approved by,

Dr. Stephen Crews

Dr. Corbin D. Jones

Dr. Joel Kingsolver

Dr. Pelin Cayirlioglu Volkan

Dr. Christopher Willett

ABSTRACT

ERIC JAY EARLEY: The Genetics of Adaptation in *Drosophila sechellia*
(Under the direction of Dr. Corbin D. Jones)

Drosophila sechellia, an ecological specialist on the ripe fruit of *Morinda citrifolia* (Morinda), displays a suite of adaptations which allow it to both prefer and tolerate Morinda and its toxic compounds. Other Drosophilids, like *D. melanogaster*, find Morinda repellent and toxic. Despite years of effort to dissect the genetic basis of this behavioral and physiological divergence, we still do not understand what genes allow *D. sechellia* to prefer and tolerate Morinda and what genes drive aversion in other species. In this dissertation I dissect the genetic basis of preference and tolerance using both traditional genetic and molecular tools along with new whole genome sequencing methods. I find that preference is genetically complex (Chapter Two), requiring up to 27 different genetic loci (Chapter Three). At the same time, however, a gene expressed in the fly's peripheral nervous system, *gustatory receptor 22c, Gr22c,* is responsible for nearly 50% of the transition between aversion and preference (Chapter Four). Surprisingly, extant *D. sechellia Gr22c* is likely a pseudogene, suggesting that preference evolution proceeded in two steps: loss of aversion then gain of preference. Finally, I, along with collaborators, introgress *D. sechellia* Morinda tolerance factors into a *D. simulans* genome and identify 17 candidate genes, among then three *Odorant binding proteins* (Chapter Five). We find that tolerance alone is not enough to confer altered behavior, conflicting with evolutionary models that predict preference and tolerance loci

would evolve to be genetically linked. In sum, this dissertation shows that the genetic basis of *D. sechellia* specialism on Morinda is complex and suggests that the evolution of preference and tolerance occurred in multiple steps: loss of aversion, gain of tolerance, and finally gain of preference.

Dedicated to my wife, Alice Drozdiak, for her unceasing love and support.

ACKNOWLEDGEMENTS

This work would not have been possible without the extraordinary help of so many people. First, thanks to my adviser Corbin Jones for his time, energy, and passion. Thank you for the journal clubs, lab meetings, and cheap beer. For the hundreds of hours of practice talks, manuscript edits, meetings, working groups, and – most importantly – the flies.

Thanks to all my lab mates (and recent pod mates) who have listened to me bitch, given edits on writing, listened to practice talks, consulted on projects, and generally enhanced my work, in particular Anais Monroy, Alicia Brandt, and Sophia Shih. Thanks especially to Josie Reinhardt, Cass Jabara, and Artur Romanchuk who have helped me in so many tangible and intangible ways.

Thanks to my amazing undergraduate assistants, Brooke Wolford and Betty Wanjiru, for their tireless help, and thanks to my collaborator Eric Hungate for letting work on his great project.

This work would not have been possible without funding from the National Science Foundation (Pre-Doctoral Fellowship and NSF#MCB0920196) and NIDDK grant (P30DK056350) to the UNC Nutrition Obesity Research Center.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

Morinda: *Morinda citrifolia* fruit

OA: Octanoic Acid

HA: Hexanoic Acid

MeHex: Methyl hexanoate

bp: base pair

GFP: Green fluorescent protein

SNP: Single nucleotride polymorphism

Indel: Insertion/deletion

Or: Olfactory receptor

Gr: Gustatory receptor

Obp: Odorant binding protein

Ir: Ionotropic receptor

Orco: *Odorant co-receptor*

poxn: pox-neuro

PSIseq: Phenotype-based Selection and Introgression with Whole genome sequencing

ORN: Olfactory receptor neuron

GRN: Gustatory receptor neuron

I. INTRODUCTION

Animals live in a world of chemicals. To find mates, food, and avoid life threatening situations, animals must process information from chemical signals into action. At the simplest level, toxic chemicals should be avoided, and beneficial chemicals should be preferred. Innate behaviors, those controlled primarily by genetic factors, ensure that organisms respond appropriately to novel chemicals with minimal risk. This is especially relevant for organisms that must chose regularly between beneficial and toxic food resources.

While we are beginning to understand how animals sense their chemical world and how chemical information is translated into electrical information within the brain (MONTELL 2009; SU *et al.* 2009; TOUHARA and VOSSHALL 2009), we do not yet understand how this information ultimately transforms into a behavior. More broadly, we do not know why some organisms prefer specific chemicals while others avoid them. What suite of genetic changes must occur for an individual to avoid or prefer? Is this transition genetically easy, or does it require many genetic modifications? An ideal system to study this question would be a group of related species with members expressing clear preference and clear avoidance of a specific chemical(s), and is amenable to genetic dissection.

The behavioral divergence between *Drosophila sechellia* and *D. simulans* is such a system (JONES 2005; RAMDYA and BENTON 2010; TSACAS and BACHLI 1981). These

two species are closely related (~0.5 MYA since most recent common ancestor (KLIMAN *et al.* 2000) and exhibit clearly divergent behavior toward the ripe fruit of *Morinda citrifolia* (Morinda, (AMLOU *et al.* 1998a; LEGAL *et al.* 1992; R'KHA *et al.* 1991). *D. sechellia* prefers Morinda and has evolved a suite of behavioral and physiological adaptations that enable it to consume and oviposit upon it (JONES 2005). *D. simulans*, on the other hand, avoids Morinda and finds it highly toxic (AMLOU *et al.* 1998a; AMLOU *et al.* 1998b; JONES 1998; R'KHA *et al.* 1991). This divergence appears to be driven in part by toxic fatty acids, Octanoic acid (OA) and Hexanoic acid (HA), abundant in Morinda (AMLOU *et al.* 1998a; FARINE *et al.* 1996; HIGA and FUYAMA 1993; PINO *et al.* 2010). Both fly species are also close relatives of *D. melanogaster,* which behaves like *D. simulans*, and provides access to an array of genetic, molecular, and cellular tools. Finally, all three species possess fully sequenced and annotated genomes (CLARK *et al.* 2007).

In the 30 years since its discovery (TSACAS and BACHLI 1981) studies of *D. sechellia* have shed some light on the genetic basis of its specialism on Morinda, yet there are gaping holes in our knowledge (RAMDYA and BENTON 2010; WHITEMAN and PIERCE 2008). We do not yet know, for example, what specific genetic modifications allowed the transition from Morinda aversion to Morinda to preference. Given Morinda's toxic potential for non-*sechellia* species, we do not know if behavior toward Morinda is genetically linked to or a pleiotropic consequence of tolerance of Morinda. Finally, we would like to identify general patterns in this genetic transition, if any exist. That is, should we expect to see certain types of mutations driving the evolution of preference

(deletions, duplications, etc.) and should we expect certain classes of genes to evolve more than others (*e.g.,* genes functioning in the peripheral or central nervous system)?

Combining all that we know about *D. sechellia* preference and tolerance for Morinda, we can begin to test hypotheses on how the transition toward specialism took place. In the current work, I propose three. First, behavioral specialism in *D. sechellia* involves multiple genes but not many genes of infinitesimally small effect size. Additionally, tolerance and preference are genetically correlated and likely require separate loci. Second, the evolution of *D. sechellia* specialism likely took place in a step-wise manner. Preference and tolerance appears to have evolved separately, and preference itself likely evolved in multiple steps—*i.e*. ancestral and aversive *D. sechellia* evolved indifference for Morinda followed by subsequent gain of preference. Third, *D. sechellia* possesses an unusually large number of pseudogenes (McBRIDE 2007; McBRIDE *et al.* 2007), and that some of these loss of function mutations contributed to specialized preference and are not simply a consequence of specialism itself.

Natural history

Drosophila sechellia is an island endemic species of the Seychelles archipelago, a granitic island chain off the coast of east Africa. *D. sechellia* is found throughout the archipelago, including Mahé, Frigate, Praslin, and Cousin islands, consuming and ovipositing upon ripe fruit of *Morinda citrifolia* (Morinda, (TSACAS and BACHLI 1981). While humans may have visited the islands much earlier, permanent settlement did not occur until roughly 400 years ago, and with them arrived many alien species of animals and plants. DNA evidence suggests, however, that *D. sechellia* colonized—and speciated

from its close relative *D. simulans*—on the islands long before humans arrived (KLIMAN *et al.* 2000).

D. sechellia's host plant, Morinda, is found throughout the archipelago as well as across the entire southeastern Pacific and Indian Ocean (BAKER 1970; ROBERTSON 1989; SAUER 1967). Its seeds can survive on salt water for more than a year, so Morinda may have floated from other locales. However, some controversy exists over its colonization on the Seychelles. Lachaise and Silvain (LACHAISE and SILVAIN 2004) argue that Morinda was most likely introduced recently by humans and that *D. sechellia* could have evolved its specialist lifestyle on chemically related plants (*e.g.,* species within the *Pandanus* genus or coconuts) and shifted to Morinda after its introduction. If Morinda was not the original host for *D. sechellia*'s evolved specialism, we can at least be sure that *D. sechellia* evolved preference and tolerance for the specific chemicals that are abundant in Morinda..

D. sechellia specialization of Morinda and/or its chemicals could have been driven by a number of genetic and ecological factors. Given the high rate of gene loss in its genome (CLARK *et al.* 2007; MCBRIDE 2007), and the paucity of genetic variation across its range (LEGRAND *et al.* 2009), ancestral *D. sechellia,* like other island endemic species, probably experienced a severe genetic bottleneck during colonization. Today *D. sechellia*'s range includes a number of sympatric non-specialist *Drosophila*, including *D. simulans* and *D. malerkotliana* (LOUIS and DAVID 1986; R'KHA *et al.* 1997). These species are strong competitors. One model of adaptive radiation on islands is that founder populations diverge allopatrically from mainland populations, evolving intrinsic reproductive isolating barriers, and subsequent re-introduction of mainland individuals

onto the island facilitates adaptive character displacement between these two groups since gene flow is reduced or absent (LOSOS and RICKLEFS 2009). Thus, competition with other species could have contributed to character displacement in nascent *D. sechellia* and pushed it to consume non-optimal food sources. If this food was only weakly toxic – or toxic for a short time – then its caloric benefit could have outweighed the risk of lethal exposure. Other ecological drivers of specialism could have also included escape from predators, such as the parasitoid wasp *Leptopilina boulardini* (LOUIS and DAVID 1986). By ovipositing on toxic Morinda, female *D. sechellia* could have ensured an "enemy-free" environment.

D. sechellia is resistant to Morinda and its toxins

Morinda fruit have a distinctive smell and are avoided by other Drosophilids (DEKKER *et al.* 2006; JONES 2005; R'KHA *et al.* 1991). Gas chromatography of Morinda headspace at various ripening stages identified 96 volatile compounds emanating from the fruit, including organic acids, alcohols, esters, ketones, and lactones (FARINE *et al.* 1996; PINO *et al.* 2010). However, the most abundant of Morinda volatiles are Octanoic and Hexanoic Acids (OA, 58-70% and HA, 8-19%) and their ester derivatives, and these components give Morinda its characteristic "rancid cheese" smell. Ripe *M. citrifolia* fruit has the highest concentration of these acids and is highly toxic to Drosophilids except *D. sechellia* (JONES 2005; R'KHA *et al.* 1991). OA and HA are the main causes of the toxicity (AMLOU *et al.* 1998a). However, this toxic effect appears limited to just a few days after dropping from its tree. Pre-ripe (green) and rotten fruits (dark brown) are not toxic (LEGAL *et al.* 1992).

D. sechellia is tolerant of Morinda and its predominant toxic compounds, OA and HA (AMLOU *et al.* 1998a; AMLOU *et al.* 1997; JONES 1998; JONES 2005; LEGAL *et al.* 1994; LEGAL *et al.* 1992; LEGAL *et al.* 1999; R'KHA *et al.* 1991). Other Drosophilids, among them *D. melanogaster, D. simulans,* and *D. mauritiana*, are sensitive to Morinda and OA in particular. Both sensitivity and tolerance are consistent throughout all life history stages.

Adult Tolerance

Sensitivity studies of flies exposed either directly to Morinda, indirectly via its vapor, or pure OA have shown non-*sechellia* flies are knocked down (a prelude to death) within 30 minutes, whereas *D. sechellia* remains unaffected for the duration of tests (JONES 1998). The physiological effects of Morinda and its compounds are currently unknown. Flies exposed to Morinda or OA exhibit stereotypical symptoms – hyper-grooming, loss of equilibrium, and finally irreversible coma – which resemble the effects of neurotoxins (LEGAL *et al.* 1992).

Interestingly, while *D. sechellia* is highly tolerant of OA, HA, and Morinda, it is less tolerant of common rotten fruit volatiles, including ethanol (MERCOT *et al.* 1994). This is surprising, given that Drosophilids in general have evolved a saprophagous diet and often prefer volatiles associated with rotten fruit. *D. sechellia* is thus unusual in that it appears to have evolved preference and tolerance for a *ripe* fruit, that is only available during a narrow time frame. This may explain why *D. sechellia* preference is so sensitive (see discussion of preference below).

Early genetic studies of *D. sechellia* tolerance to Morinda and its compounds discovered that tolerance is dominant in interspecific hybrids (AMLOU *et al.* 1997) and biometric analysis estimated at least three genetic factors were involved (R'KHA *et al.* 1991). More recent studies have relied upon marker assisted backcross experiments with visibly marked *D. simulans*. Jones (1998) first reported that *D. sechellia* OA tolerance was oligogenic and factors were present across all major chromosomes. At least two factors were present on the X, one on the 2nd, and the two strongest factors near the centromere of the 3rd, with epistatic interactions present between all chromosomes.

Larval Tolerance

D. sechellia larvae tolerate Morinda toxins, while eggs and larvae of other Drosophilids are highly sensitive (JONES 2005; R'KHA *et al.* 1991). Initial work by AMLOU, MORETEAU, and DAVID (AMLOU *et al.* 1998b) investigating the genetic basis of OA and HA tolerance discovered that *D. simulans, D. melanogaster,* and *D. mauritiana* all exhibited delayed larval development at low doses of OA and HA, whereas *D. sechellia* development was only affected at higher concentrations. The authors also found evidence that *D. sechellia* tolerance was recessive; however, previous work by others had shown that tolerance was dominant (R'KHA *et al.* 1991). One explanation for this discrepancy is that AMLOU, MORETEAU, and DAVID (AMLOU *ET AL.* 1998B) only used *D. simulans* mothers in their crosses, allowing for the possibility that maternal effects were blocking dominant zygotic tolerance. JONES (2001) accounted for this effect by repeating the analysis with reciprocal F1 hybrids, reciprocal backcrosses, and compound-*X*

chromosomes. In this case, *D. sechellia* tolerance was semi-dominant and did involve a maternal effect.

JONES (2001) also dissected the genetic basis of larval tolerance in finer detail. Using 11 genetic markers and reciprocal hybridization with *D. simulans*, the author found tolerance to be oligogenic with no effect of the fourth or *X* chromosomes. The third chromosome harbored at least one intermediately dominant factor, and the second chromosome harbored at least two mostly dominant factors. These factors are likely a subset of those important for adult resistance.

Chemotaxis and Oviposition Behavior

In addition to its tolerance to Morinda, *D. sechellia* has also evolved high sensitivity and preference for Morinda (R'KHA *et al.* 1991) and its compounds (AMLOU *et al.* 1998a; DWORKIN and JONES 2009; EARLEY and JONES 2011; MATSUO *et al.* 2007). Other Drosophilids actively avoid ripe Morinda and several of its constitutive compounds (AMLOU *et al.* 1998a; EARLEY and JONES 2011; LEGAL *et al.* 1992; MATSUO *et al.* 2007; R'KHA *et al.* 1991). The common ancestor between *D. simulans, D. mauritiana,* and *D. sechellia* was probably aversive to Morinda. How ancestral aversion evolved to preference, and what genetic mechanisms drove this transition, remains an unanswered question.

Work in the last 20 years has begun to uncover potential genetic drivers of the transition between aversion and preference. Early studies showed that *D. sechellia* preference for Morinda was recessive within interspecific hybrids with *D. simulans* (R'KHA *et al.* 1991, AMLOU *et al.* 1998a). Genetic mapping via interspecific backcross

with *D. simulans* posessing visible markers uncovered one locus on the second chromosome controlling behavior to HA (HIGA and FUYAMA 1993). A similar study of OA on egg laying behavior also using a backcross with *D. simulans*, but this time with more markers, discovered a locus on the left arm of chromsome two and the right of chromosome three (JONES 2005). No study found any effect of the X chromosome.

Role of Chemosensory System in the Behavior of D. sechellia.

The genetic basis of how *D. sechellia* finds and chooses Morinda or how *D. melanogaster* avoids this fruit remains unknown. A fly's chemosensory system, however, is a promising starting point for finding genes influencing *D. sechellia's* behavior (RAMDYA and BENTON 2010; WHITEMAN and PIERCE 2008). The past two decades have provided new insight into the molecular and genetic mechanisms of *Drosophila* olfactory and gustatory response (BENTON *et al.* 2009; MONTELL 2009; VOSSHALL and STOCKER 2007). Fly chemosensation is first processed by sensory neurons for taste (gustatory receptor neurons, GRNs) or smell (olfactory receptor neurons, ORNs) which are housed within sensory hairs ("sensilla"), scattered across the fly's antennae, proboscis, maxillary palps, legs, wings, and ovipositor. These neurons express seven-pass transmembrane receptor proteins which bind chemicals. These molecular receptors are coded for by three families of genes: *Olfactory receptors (Ors*), *Ionotropic receptors* (*Irs*, another set of olfaction genes), and *Gustatory receptors* (*Grs*). A fourth class of proteins, Odorant binding proteins (coded by *Obps*), appear to aid in the transport of organic chemicals through the aqueous lymph surrounding the sensory receptor neurons and enhance the sensitivity of receptors to specific chemicals. Once receptors bind a suitable chemical,

neural signals are transmitted to the brain via the antennal lobe (olfaction) or the suboesophageal ganglion (SOG, gustation). ORNs expressing the same receptor type transmit signal to the same glomerulus within the antennal lobe. GRNs, however, transmit signal to a variety of locations scattered across the SOG. From these regions, taste and smell signals are transmitted into higher order brain neurons and processed ultimately into behavior in ways not fully understood.

Recent work by a number of groups have mapped Morinda behavior genome wide, queried specific genes, and have identified specific olfactory and gustatory responses elicited by Morinda in *D. sechellia*. STENSMYR et al. (2003) and DEKKER et al. (2006) identified specific antennal sensilla which respond neurophysiologically to Morinda vapor. *D. sechellia* is 10 times more sensitive to the Morinda volatile methyl hexanoate relative to *D. melanogaster*. The ab3A type sensilla, located on the antennae, respond to methyl hexanoate (MeHex), and *D. sechellia* has an unusually large number of these sensilla relative to *D. melanogaster*. Supporting this model, experiments genetically perturbing the numbers and types of sensilla using alleles of lozenge (*lz*) affect how *D. melanogaster* responds to OA (JONES 2007). Similarly, the glomerulus receiving signal from these neurons is also enlarged in *D. sechellia* (IBBA *et al.* 2010). It appears that *D. sechellia* is hypersensitive to methyl hexanoate, and hence Morinda, because of this morphological transition.

However, several complications of this model prevent it from completely explaining *D. sechellia* preference for Morinda. *D. simulans*, which is strongly aversive to Morinda, also harbors ab3A sensilla which are highly sensitive to methyl hexanoate (STENSMYR *et al.* 2003). *Or22a* and *Or85b* are the only known ab3A expressed receptors

to respond to methyl hexanoate in *D. melanogaster* (HALLEM and CARLSON 2006), but their differential expression across species cannot fully explain Morinda behavior. *Or22a* is upregulated in both *D. simulans* and *D. sechellia* relative to *D. melanogaster* (which itself has highly variable expression uncorrelated with ab3A density), and the combined high expression of *Or22a* and *Or85b* in *D. sechellia* should promote heightened sensitivity to ethyl-hexanoate rather than methyl-hexanoate (KOPP *et al.* 2008), but this is the opposite of what is observed (STENSMYR *et al.* 2003). Direct genetic tests, for example by knockouts of *Or22a* and *Or85b* within *D. sechellia* itself, are needed to know for sure whether these genes actually mediate preference.

MATSUO and colleagues (MATSUO *et al.* 2007) through direct genetic tests discovered that a coexpressed pair of odorant binding proteins, *Obp57d/e,* mediated oviposition behavior to OA and HA in *D. melanogaster*. *D. sechellia*'s ortholog (*Dsech/Obp57d/e*) harbors a 4bp insertion just upstream of the shared *Obp57d* and *Obp57e* transcription start site relative to other members of the *D. melanogaster* species complex, and this insertion appeared to disrupt expression of *Dsech/Obp57d/e,* but only in the tarsi. *D. melanogaster* females with knocked-in *Dsech/Obp57d/e* preferred ovipositing on high concentrations of HA and OA, a similar preference in *D. sechellia*. Given its presumed role as a chemical chaperone (GALINDO and SMITH 2001), OBP57d/e could act as a facilitator to nearby chemosensory receptors and enhance sensitivity to Morinda compounds.

In parallel with the anotomaical and gene-specific search above, multiple studies have correlated patterns of molecular evolution with *D. sechellia* preference. MCBRIDE (2007) discovered a surprisingly high rate of pseudogene accumulation in *D. sechellia*

chemosensory genes – almost 10 times faster than *D. simulans.* Given its small population size, *D. sechellia* could have many pseudogenes because of genetic drift, but the rate of chemosensory pseudogene fixation in *D. sechellia* is significantly higher than its non-chemosensory genomic background. Chemosensory genes in *D. sechellia* also show signatures of positive natural selection relative to the genome average. *Gr*s in particular appear to be evolving faster than *Or*s in *D. sechellia*. Compared to *D. simulans*, *D. sechellia* has more *Gr* pseudogenes than *Or* pseudogenes (17.8% vs. 9.5%) and mean *Ka/Ks* ratios, a test for positive selection, increased by 94% in *Gr*s compared to only 67% increase in *Or*s. Thus, evidence points to positive selection acting on both *Or*s and *Gr*s within *D. sechellia*, although recent relaxation of selection could also explain these patterns. In a separate study MCBRIDE and ARGUELLO (2007) expanded their focus to more species, including another specialist *D. erecta,* which specializes seasonally on members of the *Pandanus* family (LACHAISE and TSACAS 1974). *D. erecta* also experienced an increased rate of *Gr* family contraction, and both *D. sechellia* and *D. erecta* show strong signatures of positive selection (LINZ *et al.* 2013). *Ka/Ks* in larval expressed *Or*s are higher than expected, and *Ka/Ks* is particularly high within specialist lineages in *Gr*s responding to sweet compounds compared to *Gr*s responding to bitter compounds.

DWORKIN and JONES (2009) measured whole genome expression differences between *D. simulans* and *D. sechellia* both before and after being exposed to Morinda compounds (OA/HA mixture). A variety of chemosensory genes were strongly up- and down-regulated in *D. sechellia* relative to *D. simulans* including *Or22a* (up in *D. sechellia*), *Or85b* (up in *D. sechellia*), *Obp83ef* (up in *D. sechellia*), *Obp83cd* (down in

D. sechellia). These results match the over representation of ab3A sensilla in *D. sechellia*, as both *Or22a* and *Or85b* are expressed within these sensilla in *D. melanogaster*. *Obp56e* was particularly interesting for a few reasons. It was down regulated in *D. sechellia* relative to *D. simulans*, but after exposure to Morinda compounds its expression increased significantly in *D. sechellia*. This pattern is surprising given that *Obp56e* in *D. sechellia* is likely a pseudogene as it harbors a stop codon roughly 50% through its open reading frame relative to *D. simulans*. When *Obp56e* was knocked down in *D. melanogaster,* flies lost some aversion to Morinda compounds. How a putative pseudogene could influence *D. sechellia* preference is not clear. One possibility is that the shared ancestor of *D. melanogaster* and *D. sechellia* used up-regulation of *Obp56e* to avoid Morinda-like plants emiting OA or HA and a deletion leading to a premature stop codon in the *D. sechellia* lineage shut this aversive process down, and remnant up-regulation of *Obp56e* in *D. sechellia* remains despite the protein's putative loss of function.

Combining all findings to date, we are beginning to piece together the evolutionary history of *D. sechellia* and its behavioral adaptation toward Morinda fruit. The trait involves at least a handful of genes, two of which are are known—*Obp57d/e* and *Obp56e*. Loss of the activity of these genes in *D. sechellia's* progenitor likely contributed to the reduction of ancestral aversion of Morinda. However, neither of these genes alone, or presumably in combination, appears sufficient to remove aversion in *D. melanogaster* or reconstitute *D. sechellia* like preference. Similarly, while the morphological evolution of ab3A sensilla in *D. sechellia* could contribute to *D. sechellia's* preference, we do not yet have any direct genetic tests of this model.

It must also be noted that specializing on Morinda must have required a suite of traits, including behavioral preference and physiological tolerance. Any behavioral adaptation toward Morinda in adults must have required eggs and larvae to survive on the fruit. One of two conditions could have allowed this: pre-existing tolerance for octanoic acid in eggs and larvae or behavioral indifference in adults manifested only toward over-ripe Morinda which was not toxic to eggs and larvae.

Evolution to Preference: One step or two?

How many mutations were initially needed to transition from aversion to preference? The simplest model predicts one mutation. While unlikely, this model is not outside the realm of possibility. At least two separate transgenic experiments in mice (MUELLER *et al.* 2005) and *C. elegans* (TROEMEL *et al.* 1997) show how changes in chemosensory regulation, and where expression occurs in space, can instantly toggle behavior between preference and aversion for the same compound. A slightly more complicated model, involving two separate mutations, is more likely.

In the two step model (first articulated by (MATSUO *et al.* 2007), aversion is first lost in *D. sechellia's* ancestor, leading to behavioral indifference. If this food is not toxic (or only mildly toxic), then indifferent flies could benefit from its consumption. Resource competion or heavy predation load could enhance this benefit. Subsequent mutations solifidy this benefit through gain of preference alleles (and perhaps tolerance alleles), which then allow the fly to actively seek the unused food.

This two step model is possible in *D. sechellia*. Multiple *Drosophila* species have been found on over ripe/rotten (non-toxic) Morinda (LOUIS and DAVID 1986; R'KHA *et*

al. 1991; R'KHA *et al.* 1997), showing that Morinda is a rich resource and accessible to flies expressing aversion. Flies that could accommodate Morinda at earlier ripening stages (more toxic) would gain . Similarly, flies that were able to oviposit on the ripe fruits could ensure a safe and abundant resource for their offspring and only larvae would have to be strongly resistant. The main difference between the one and two step models of preference evolution is that in the two step model, tolerance could evolve within an organism that is regularly exposed to Morinda toxin hence providing immediate benefit. In the one step model, flies would have to either possess a pre-adapted tolerance or evolve tolerance and preference in synchrony.

The current work

My goal is to reveal the genetic basis of host specialization in *D. sechellia* and uncover whether this process required multiple genetic steps. We know that specializing on Morinda required at least two traits - preference and tolerance – but we do not know if these traits are controlled by the same gene, nor if they arose in multiple steps. Genetic complexity in this system would contradict a number of studies showing that new host specializations can evolve "easily" from changes at relatiely few loci. I show that these two traits, preference and tolerance, are required to consume Morinda, and these traits evolved from separate genetic factors. Preference alone likely evolved gradually in at least two steps – loss of ancestral aversion and gain of preference. I support this model by identifying genetic factors affecting behavior and tolerance and ruling out alternative scenarios that predict few genetic changes.

In Chapter Two, I present a method of mapping genome-wide the genetic basis of additive traits. This method links classical genetic techniques – artificial selection and backcrossing – with modern whole-genome sequencing to track in high resolution the introgression of chromosomal elements from one background into another a known gemomic background. I test this method by introgression *D. simulans* genes contributing to Morinda aversion into a *D. sechellia* background and sequencing the genomes of the the resulting flies at low coverage to map introgression break points. Results show at least six factors contribute to aversion behavior.

In Chapter Three, I repeat this process with more biological and technical replication and with higher sequencing coverage. I find that some regions introgress repeatedly, while other regions introgress with more difficulty. I verify certain introgressions by functionally testing specific genes located within introgression breakpoints and by creating independent backcrosses of *D. simulans* and *D. sechellia* and track the influence of *D. simulans* chromosomal regions on hybrid Morinda behavior. I compliment this mapping screen with a candidate gene approach. Surprisingly, there was little overlap between these two approaches. These results suggest that Morinda behavior is genetically complex, likely requiring at least nine and up to 27 separate factors.

In Chapter Four, I show that *D. melanogaster* lacking *Gustatory receptor Gr22c, Gr22c*, lose aversion. This factor explains roughly 40-50% of the behavioral transition between aversion and preference, and a pseudogene of *Gr22c* is segrating in extant populations of *D. melanogaster*. *D. sechellia* itself harbors a loss of function allele for this gene, which suggests that loss of *Gr22c* could have been an intermediate step between ancestral Morinda aversion and derived preference. Thus, not only is the

transition to specializing on Morinda genetically complex, as seen in Chapters Two and Three, but evolving from aversion to preference alone likely required multiple steps.

In Chapter Five I, along with my collaborators, fine-map the genetic basis of OA tolerance within interspecific hybrids of *D. simulans* and *D. sechellia* and test whether tolerance alone is enough to alter behavior. We identify a small candidate region of 18 genes, three of which are *Obps*. We show that innate tolerance to OA and Morinda does not modify innate preference for the same chemical. This again confirms that multiple genetic steps were necessary for *D. sechellia* to specialize on Morinda.

Finally, in Chapter 6, I lay out my conclusions from this series of studies and propose a series of follow-up experiments.

REFERENCES

AMLOU, M., B. MORETEAU and J. R. DAVID, 1998a Genetic analysis of Drosophila sechellia specialization: oviposition behavior toward the major aliphatic acids of its host plant. Behav Genet **28:** 455-464.

AMLOU, M., B. MORETEAU and J. R. DAVID, 1998b Larval tolerance in the Drosophila melanogaster species complex toward the two toxic acids of the D. sechellia host plant. Hereditas **129:** 7-14.

AMLOU, M., E. PLA, B. MORETEAU and J. R. DAVID, 1997 Genetic analysis by interspecific crosses of the tolerance of *Drosophila sechellia* to major aliphatic acids of its host plant. Genet. Sel. Evol. **29:** 511-522.

BENTON, R., K. S. VANNICE, C. GOMEZ-DIAZ and L. B. VOSSHALL, 2009 Variant ionotropic glutamate receptors as chemosensory receptors in Drosophila. Cell **136:** 149-162.

CLARK, A. G., M. B. EISEN, D. R. SMITH, C. M. BERGMAN, B. OLIVER *et al.*, 2007 Evolution of genes and genomes on the Drosophila phylogeny. Nature **450:** 203-218.

DEKKER, T., I. IBBA, K. P. SIJU, M. C. STENSMYR and B. S. HANSSON, 2006 Olfactory shifts parallel superspecialism for toxic fruit in Drosophila melanogaster sibling, D. sechellia. Curr Biol **16:** 101-109.

DWORKIN, I., and C. D. JONES, 2009 Genetic changes accompanying the evolution of host specialization in Drosophila sechellia. Genetics **181:** 721-736.

EARLEY, E. J., and C. D. JONES, 2011 Next-generation mapping of complex traits with phenotype-based selection and introgression. Genetics **189:** 1203-1209.

FARINE, J. P., L. LEGAL, B. MORETEAU and J. L. LEQUERE, 1996 Volatile components of ripe fruits of Morinda citrifolia and their effects on Drosophila. Phytochemistry **41:** 433-438.

GALINDO, K., and D. P. SMITH, 2001 A large family of divergent Drosophila odorant-binding proteins expressed in gustatory and olfactory sensilla. Genetics **159:** 1059-1072.

HALLEM, E. A., and J. R. CARLSON, 2006 Coding of odors by a receptor repertoire. Cell **125:** 143-160.

HIGA, I., and Y. FUYAMA, 1993 Genetics of food preference in Drosophila sechellia. I. Responses to food attractants. Genetica **88:** 129-136.

IBBA, I., A. M. ANGIOY, B. S. HANSSON and T. DEKKER, 2010 Macroglomeruli for fruit odors change blend preference in Drosophila. Naturwissenschaften **97:** 1059-1066.

JONES, C. D., 1998 The genetic basis of Drosophila sechellia's resistance to a host plant toxin. Genetics **149:** 1899-1908.

JONES, C. D., 2001 The genetic basis of larval resistance to a host plant toxin in Drosophila sechellia. Genet Res **78:** 225-233.

JONES, C. D., 2005 The genetics of adaptation in Drosophila sechellia. Genetica **123:** 137-145.

JONES, C. D., 2007 Genetic ablation of antennae or basiconic sensilla reduces *Drosophila melanogaster's* ability to perceive odors from the fruit of *Morinda citrifolia*. Drosophila Information Service **90:** 45-49.

KLIMAN, R. M., P. ANDOLFATTO, J. A. COYNE, F. DEPAULIS, M. KREITMAN *et al.*, 2000 The population genetics of the origin and divergence of the Drosophila simulans complex species. Genetics **156:** 1913-1931.

KOPP, A., O. BARMINA, A. M. HAMILTON, L. HIGGINS, L. M. MCINTYRE *et al.*, 2008 Evolution of gene expression in the Drosophila olfactory system. Mol Biol Evol **25:** 1081-1092.

LACHAISE, D., and J. F. SILVAIN, 2004 How two Afrotropical endemics made two cosmopolitan human commensals: the Drosophila melanogaster-D. simulans palaeogeographic riddle. Genetica **120:** 17-39.

LACHAISE, D., and L. TSACAS, 1974 Les drosophilidae des savanes preforestieres de la region tropicale de Lamto (Cote d'Ivoire). II. Le peuplement des fruits de *Pandanus candelabrum* (Pandanacees). Ann. Univ. Abidjan Ser. E Ecol. **7:** 153-192.

LEGAL, L., B. CHAPPE and J. M. JALLON, 1994 Molecular Basis of *Morinda citrifolia* (L.): Toxicity on Drosophila. Journal of Chemical Ecology **20:** 1931-1943.

LEGAL, L., J. R. DAVID and J. M. JALLON, 1992 Toxicity and attraction effects produced by *Morinda citrifolia* fruits on the *Drosophila melanogaster* complex of species. Chemoecology **3:** 125-129.

LEGAL, L., B. MOULIN and J. M. JALLON, 1999 The Relation between Structures and Toxicity of Oxygenated Aliphatic Compounds Homologous to the Insecticide Octanoic Acid and the Chemotaxis of Two Species of *Drosophila*. Pesticide Biochemistry and Physiology **65:** 90-101.

LEGRAND, D., M. I. TENAILLON, P. MATYOT, J. GERLACH, D. LACHAISE *et al.*, 2009 Species-wide genetic variation and demographic history of Drosophila sechellia, a species lacking population structure. Genetics **182:** 1197-1206.

LINZ, J., A. BASCHWITZ, A. STRUTZ, H. K. DWECK, S. SACHSE *et al.*, 2013 Host plant-driven sensory specialization in Drosophila erecta. Proc Biol Sci **280:** 20130626.

LOSOS, J. B., and R. E. RICKLEFS, 2009 Adaptation and diversification on islands. Nature **457:** 830-836.

LOUIS, J., and J. R. DAVID, 1986 Ecological Specialization in the *Drosophila melanogaster* species subgroup: A case study of *D. sechellia*. Acta Oecologica, Oecological Generalis**:** 215-219.

MATSUO, T., S. SUGAYA, J. YASUKAWA, T. AIGAKI and Y. FUYAMA, 2007 Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in Drosophila sechellia. PLoS Biol **5:** e118.

MCBRIDE, C. S., 2007 Rapid evolution of smell and taste receptor genes during host specialization in Drosophila sechellia. Proc Natl Acad Sci U S A **104:** 4996-5001.

MCBRIDE, C. S., J. R. ARGUELLO and B. C. O'MEARA, 2007 Five Drosophila genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. Genetics **177:** 1395-1416.

MERCOT, H., D. DEFAYE, P. CAPY, E. PLA and J. R. DAVID, 1994 Alcohol Tolerance, ADH Activity, and Ecological Niche of *Drosophila* Species. Evolution **48:** 746-757.

MONTELL, C., 2009 A taste of the Drosophila gustatory receptors. Curr Opin Neurobiol **19:** 345-353.

MUELLER, K. L., M. A. HOON, I. ERLENBACH, J. CHANDRASHEKAR, C. S. ZUKER *et al.*, 2005 The receptors and coding logic for bitter taste. Nature **434:** 225-229.

PINO, J. A., E. MARQUEZ, C. E. QUIJANO and D. CASTRO, 2010 Volatile compounds in noni (Morinda citrifolia L.) at two ripening stages. Ciencia E Tecnologia De Alimentos **30:** 183-187.

R'KHA, S., P. CAPY and J. R. DAVID, 1991 Host-plant specialization in the Drosophila melanogaster species complex: a physiological, behavioral, and genetical analysis. Proc Natl Acad Sci U S A **88:** 1835-1839.

R'KHA, S. B., B. MORETEAU, J. A. COYNE and J. R. DAVID, 1997 Evolution of a lesser fitness trait: egg production in the specialist *Drosophila sechellia*. Genet. Res. Camb. **69:** 17-23.

RAMDYA, P., and R. BENTON, 2010 Evolving olfactory systems on the fly. Trends Genet **26:** 307-316.

STENSMYR, M. C., T. DEKKER and B. S. HANSSON, 2003 Evolution of the olfactory code in the Drosophila melanogaster subgroup. Proc Biol Sci **270:** 2333-2340.

SU, C. Y., K. MENUZ and J. R. CARLSON, 2009 Olfactory perception: receptors, cells, and circuits. Cell **139:** 45-59.

TOUHARA, K., and L. B. VOSSHALL, 2009 Sensing odorants and pheromones with chemosensory receptors. Annu Rev Physiol **71:** 307-332.

TROEMEL, E. R., B. E. KIMMEL and C. I. BARGMANN, 1997 Reprogramming chemotaxis responses: sensory neurons define olfactory preferences in C. elegans. Cell **91:** 161-169.

TSACAS, L., and G. BACHLI, 1981 *Drosophila sechellia*, n.sp., huitieme espece du sous-goupe melanogaster des Iles Sechelles [Diptera, Drosophilidae]. Revue Fr. Ent. NS**:** 146-150.

VOSSHALL, L. B., and R. F. STOCKER, 2007 Molecular architecture of smell and taste in Drosophila. Annu Rev Neurosci **30:** 505-533.

WHITEMAN, N. K., and N. E. PIERCE, 2008 Delicious poison: genetics of Drosophila host plant preference. Trends Ecol Evol **23:** 473-478.

II. NEXT-GENERATION MAPPING OF COMPLEX TRAITS WITH PHENOTYPE-BASED SELECTION AND INTROGRESSION

Authors: Eric Jay Earley and Corbin D. Jones

Published in Genetics, Vol. 189, 1203-1209, December 2011

ABSTRACT

Finding the genes underlying complex traits is difficult. We show that new sequencing technology combined with traditional genetic techniques can efficiently identify genetic regions underlying a complex and quantitative behavioral trait. As a proof of concept we used phenotype-based introgression to backcross loci that control innate food preference in *Drosophila simulans* into the genomic background of *D. sechellia*, which expresses the opposite preference. We successfully mapped *D. simulans* introgression regions in a small mapping population (30 flies) with whole-genome resequencing using light coverage (~1x). We found six loci contributing to *D. simulans* food preference, one of which overlaps a previously discovered allele. This approach is applicable to many systems, does not rely on laborious marker development or genotyping, does not require existing high quality reference genomes, and needs only small mapping populations. Because introgression is used, researchers can scale mapping population size, replication, and number of backcross generations to their needs. Finally,

in contrast to more widely used mapping techniques like F_2 bulk-segregant analysis, our method produces near isogenic lines that can be kept and re-used indefinitely.

INTRODUCTION

Complex traits are inherently difficult to dissect genetically. Quantitative Trait Loci (QTL) studies, Genome-Wide Association Studies (GWAS), forward- and reverse-genetics are all powerful tools; however, each technique has inherent weaknesses that limit ability to find causal loci (MACKAY *et al.* 2009; MANOLIO *et al.* 2009). New methods using next-generation sequencing (NGS) technology have successfully captured single loci underlying Mendelian traits generated from mutagenesis screens (BLUMENSTIEL *et al.* 2009; FLIBOTTE *et al.* 2010; LAITINEN *et al.* 2010; SARIN *et al.* 2010; SARIN *et al.* 2008; XIA *et al.* 2010; ZURYN *et al.* 2010); recombinant inbred lines, RILs (HUANG *et al.* 2009; SCHNEEBERGER *et al.* 2009); and backcross populations using dominant markers (ANDOLFATTO *et al.* 2011). However, we lack a time- and cost-effective method that maps multiple loci simultaneously without *a priori* knowledge of their location, number or effect size.

Mapping complex traits is more challenging than mapping Mendelian traits. QTL studies in the past decades have uncovered a plethora of loci underlying complex traits, but QTL methods lack the power to resolve candidate regions to individual genes (MACKAY *et al.* 2009). New NGS approaches have the potential to capture multiple causative loci; however, these methods may also lack sufficient power. HUANG *et al.* (2009) proposed "whole-genome resequencing" (WGR), and Baird *et al.* (2009) proposed "restriction site associated DNA" (RAD) genotyping, which both use NGS-based

mapping on bulk segregant populations (MICHELMORE *et al.* 1991). These methods require large mapping populations to detect multiple loci of weak effect (*e.g.,*EHRENREICH *et al.* 2010). ANDOLFATTO *et al*. (2011) developed a "multiplex shotgun genotyping" (MSG) method, yet MSG also relies on large backcross populations, and it is not clear if their approach can track quantitative or additive loci of relatively weak effect or if closely linked loci can be isolated.

We have developed a new approach that efficiently maps multiple loci contributing to a complex trait. Our method uses **p**henotype-based **s**election and **i**ntrogression followed by whole-genome re**seq**uencing (PSIseq). Our method can be easily scaled from rough mapping of a single small population to fine-scale mapping of large and replicated populations. With minimal replication, low cost, and few genomic resources, we can map any complex trait divergent between any two inter-fertile populations. This approach takes advantage of the statistical power of window-based mapping algorithms on NGS data (HUANG *et al.* 2009) and the ease of using relatively small mapping populations. Because introgression is used instead of F_2 bulk segregant analysis, mapping populations can be treated as near isogenic lines and be re-used indefinitely.

MATERIALS AND METHODS

Overview

Populations with a divergent complex trait are hybridized and then selected for a specific phenotype across multiple generations of backcrosses. Our proof of concept uses a species-level phenotype, but this method can work on any two inter-fertile populations.

The trait of interest is selected for each generation, and offspring are mated to the other parental line expressing the unselected phenotype (introgression and backcrossing). Over multiple generations of selection and backcrossing this hybrid population becomes homozygous for the majority of the unselected parent's genome while loci from the selected parent, which contribute to the selected trait, remain. Using high throughput sequencing, we map the breakpoints of these introgressions and, therefore, map the regions harboring genes influencing the trait. This scheme is analogous to introgression based mapping approaches that use marked transposable elements or molecular markers (DESJARDINS *et al.* 2010; LAURIE *et al.* 1997; TRUE *et al.* 1996). A recessive trait may also be introgressed, although this would require an extra inbreeding step following each backcross.

Phenotype-based Selection and Introgression

We tested our method on a putative complex behavioral trait divergent between two *Drosophila* sister species: *D. simulans* (c167.4) and *D. sechellia* (SynA). *D. sechellia* is an island endemic and phytophagous specialist that prefers the smell and taste of *Morinda citrifolia* (Morinda) fruit (JONES 2005; MATSUO *et al.* 2007; R'KHA *et al.* 1991). *D. simulans*, avoids the fruit and dies when in proximity to it and its constituent organic acids (JONES 1998; R'KHA *et al.* 1991). As a proof of concept, we introgressed *D. simulans* Morinda aversion behavior into the *D. sechellia* genome over 15 generations of backcrossing and selection. For systems with longer generation times where 10+ generations of backcrossing is not a viable option, one can easily reduce the number of backcross generations (*e.g.,* 1-5) while increasing introgression replication (see Results

and Figure 2.2 for a discussion of how to balance replication size versus backcross generation number).

Virgin *D. simulans* females were mated to *D. sechellia* males to create a large population of fertile F_1 females. These hybrids express *D. simulans* behavior and were backcrossed to *D. sechellia* males. Roughly 30 F_2 females were subjected to our behavioral assay, and individuals displaying *D. simulans* behavior were collected and backcrossed to *D. sechellia* males. The assay is identical to that described in DWORKIN & JONES (2009). Briefly, octanoic (45ul) and hexanoic (15ul) acids were added to instant *Drosophila* media (4.1g + 22mL diH_20; Carolina Biological Supply), or not, to create test and control food, respectively. Flies had 48 hrs to chose a medium. Flies settled on control media were considered *D. simulans* phenotype. This cycle – an assay of ~30 females then backcrossing – continued for 15 generations. The final generation was inbred for 2-3 generations to ensure that introgressed loci were mostly in a homozygous state. Thirty females were pooled for Illumina library preparation.

Introgression Mapping

To map introgression breakpoints, we used reference genomes of these species to identify single-nucleotide species differences (analogous to SNPs) that identify genomic regions as particular to a parental genome. While our proof of concept used the relative high quality *Drosophila* genomic assemblies, non-model systems without a finished assembly can still be used. In principle, a sequenced transcriptome or an rough *de novo* assembly, for example, can be used to capture introgression blocks (or transcripts), as well.

We found on average one SNP per 100bp (1% divergence, compared to the reported 2% from KLIMAN *et al.* 2000). Because these SNPs represented genetic changes between these species, markers may have also contributed to phenotypic divergence. To create our species-specific SNP map, we aligned the *D. simulans* and *D. sechellia* reference genome sequences (r1.3) reciprocally using BWA (LI and DURBIN 2009). Small insertion-deletion (indels) and microsatellites could also be used, though we did not use them in this study. We also showed that the genome of *D. melanogaster* can be used as a proxy for a reference genome and showed that SNPs distinguishing *D. simulans* and *D. sechellia* could be identified by aligning short sequencing reads from these species to the *D. melanogaster* reference using a high mismatch tolerance in the alignment (data not shown). In non-model systems this could be particularly beneficial. If neither parental line possessed an assembled genome or transcriptome, a closely related model-system assembly could be used.

Individual flies from this 15th generation introgression (BC15) lineage were pooled and sequenced *en masse* using one lane of Illumina Genome Analyzer 1.0. Sequences were deposited in the NCBI Short Read Archive (SRA) database (SRA039418.2). High quality reads were aligned via BWA to both *D. sechellia* and *D. simulans* reference genomes (84% and 72% mapped, respectively). A mismatch call required at least two confirming reads. Hybrid-parent mismatches were correlated to our parent-parent SNP database. All hybrid-parent mismatches also present in the parent-parent SNP database were considered true hybrid-parent SNPs. Despite quality filtering hybrid-parent SNPs through these two processes (multiple confirming reads and existence of parent-parent SNP), we could not be sure if a given SNP call was accurate

due to sequencing and mapping errors. To overcome this, we mapped introgression breakpoints using a window approach (as described in HUANG *et al.* 2009). Any individual mis-called SNP had minimal effect on our search for large-scale SNP enrichment. Every 1,000 parent-parent SNPs were considered a "bin," and we counted the number of hybrid-parent SNPs within this bin. Our null expectation was that hybrid-parent SNPs were binomially distributed within each bin. Any bin that harbored a significant enrichment of *D. simulans* SNPs over the chromosomal mean *D. simulans* SNP content was considered an introgression block member (Student's *t*-test). This window approach required multiple independent significance tests, which we corrected for via a false discovery rate (FDR) calculation.

Confirmation of X Chromosome Effect

D. sechellia males with recessive genetic markers evenly spaced along the X chromosome -- *zn* (*1*-25) and *f* (*1*-56) -- were mated to *D. simulans* females. The resulting F1 females were backcrossed to males from the *D. sechellia* parent. Thus, we can distinguish *D. sechellia* X chromosomes from *D. simulans* X chromosomes. We compared the influence these two classes of chromosomes have on adult food preference using a single fly two-choice oviposition assay (MATSUO *et al.* 2007). Prior work has shown that bulk population assays produce similar results to individual fly assays (AMLOU *et al.* 1998; MATSUO *et al.* 2007). The single fly assay was simpler for genotyping and it was different than the selection assay, yet assayed a similar phenotype. Each female was allowed to oviposit for two days in a chamber containing control or test

media (above). Preference was determined by which medium received the most eggs. All assays were conducted at 20°C with relative humidity 50-70%.

RESULTS

In our proof of concept study where *D. simulans* preference loci were introgressed into the *D. sechellia* genome over 15 generations within one lineage, we found that our hybrids chose non-Morinda medium 71% of the time, comparable to *D. simulans* behavior (94%), and in contrast to *D. sechellia* (18%). We found six bins showing significant *D. simulans* identity (Figure 2.1; $P < 0.0002$). Two of these bins were relatively large (1-1.5 Mb), whereas four others were smaller (20-200 Kb). Three independent introgression loci on chromosome arm *2*L were found within a 7 Mb region, illustrating the power of our method in resolving closely linked loci. We pinpointed a small region on *2*R that contains a gene for *Odorant Binding Protein 56e* previously shown to weakly affect preference (DWORKIN & JONES 2009). Combined, these six loci account for 75% of *D. simulans* aversive behavior. Our bins harbored genes underlying diverse traits: fatty-acid metabolism (*bubblegum*), cuticle tanning (*rickets*), vision-directed behavior (*black*), insecticide response (*nAcRα-30D*), temperature-directed behavior (*pickpocket*), antennal development (*elbow B*), and olfaction-directed behavior (*Smi35a*).

To confirm that our method was accurately enriching for regions affecting our complex trait, we used three approaches. First, we used an introgression model to give us expected sizes of introgression blocks given recombination rate, number of backcross generations, and the level of replication (Figure 2.2). Second, we created a backcross

simulation to estimate the size and frequency of non-backcross parental blocks remaining after 15 generations of backcrossing without selection – that is, how often would we expect to see introgression blocks by chance alone. Finally, to ensure that *D. simulans* enrichment was not simply an artifact of bin size, we re-binned chromosomes at fewer (500) and greater (10,000) SNPs per bin (Figures 2.3 & 2.4).

Introgression Modeling

To confirm that our method mapped regions of an appropriate size, we created a forward simulation to estimate the mean and variance of donor introgression block sizes (Figure 2.2). Our model uses recombination rates from *Drosophila* (TRUE *et al*. 1996) within the recombination model of FOSS *et al*. (1993), which incorporates crossover interference. We assumed a *D. simulans* genetic map of 460 cM and relatively simple interference (a non-crossover event must occur between adjacent crossovers, $m = 1$ from FOSS *et al*. 1993). Over multiple backcross generations, an introgressed block size decreases asymptotically to zero cM. With multiple independent introgressed replicates, the variance in block size also decreases asymptotically to 0. In our case, one replicate over 15 generations is expected to have a block size of ~5±5cM. Our experimentally observed blocks were roughly 0.5cM to 1cM, falling well within the expected block size. In *Drosophila*, 1cM is roughly equivalent to 0.5 Mb, depending on chromosomal position.

Backcross simulation

D. simulans enriched regions could also remain due to drift alone. To reveal the variation of introgression size by chance (that is, without selection), we performed a simulation of introgression with backcrossing. We populated a virtual pool of chromosomes made up of two parental genotypes, *A* and *B*. Each generation an *A* chromosome was recombined with a *B* chromosome, experiencing one crossover event at a random position. Each event created two daughter chromosomes that received reciprocal products. To maintain a stable population size, one daughter from each crossover was picked at random to propagate the next generation. This daughter was then "backcrossed" to a *B* chromosome. This cycle – recombination to *B* (the "backcross") followed by randomly choosing daughter recombinants ("drift") – was repeated 15 times with two levels of replication (2,000 and 20,000; see below). We tracked the size of *A* chromosome blocks maintained at each generation and calculated the population-wide *A* allele content after 15 generations.

In infinitely large populations, theory predicts that, on average, half of the *A* alleles will remain after each cycle of recombination and drift. In non-infinite populations, however, drift will likely remove more than half of *A* alleles. This is because once an *A* allele is lost, it is lost forever, and the probability of *A* alleles being lost due to drift is additive across generations. To get a sense of allele loss due to drift, consider an infinite population experiencing recombination and drift over 15 generations. This should produce a frequency of *A* alleles equal to $(½)^{15}$, or 3.05×10^{-5} % of the population allele frequency. In our simulation of 2,000 independent backcross lineages we found only one surviving *A* block within one chromosome for a population frequency of 3.8×10^{-9} % (152bp in one 20Mb chromosome).

Disregarding the low probability (1 in 2,000) of introgression by chance alone, we measured the variation in *A* frequency under a selection-free scenario. Simulations were repeated until 14 independent *A* blocks were found (20,600 replicates). Conditioning on an *A* block persisting ($N = 14$), sizes ranged from 0.7% to 27% of one chromosome (avg = 6.1±7.4%). At a population level ($N = 20{,}600$), sizes ranged from 3.4×10^{-7} % to 1.3×10^{-5} %. In our pilot experiment, we introgressed one lineage with selection and found at least 6 blocks ranging in size from 20Kb to 1.5Mb (0.1-7.5% of a 20Mb chromosome).

Binning Effect

In our effort to find regions of the 15th generation hybrid genome enriched for *D. simulans*, we binned chromosomes into overlapping 1,000 SNP (~100kb) sections. This size was chosen primarily with consideration of *Drosophila* genic density. However, choice of bin size influences one's ability to detect significant SNP enrichment. For example, picking a large bin size will mask enriched regions by including adjacent non-enriched regions. A bin size that is too small will force many comparisons and the corrected α threshold will be too low. Ultimately, bin size determines one's ability to detect enrichment.

To illustrate the influence of binning on our analysis, we re-binned generation 15 chromosome arms *2*L and *2*R at different sizes. Starting with a bin size of 500 SNPs (~50kb) we tracked clustering of *D. simulans* SNPs, calculated *P*- values then repeated this process, increasing bin size to a maximum size of 10,000 SNPs/bin. All bins overlap by 10% of their bin size. Because bin size scales inversely with the number of bins

(hence the number of significance tests), the threshold of significance according to a False Discovery Rate (FDR) changes.

Figures 2.3 & 2.4 show data from chromosome arms *2*L and *2*R, respectively, produced by four sample bin sizes. The two large regions on *2*L showing significant *D. simulans* SNP enrichment maintained this significance under all bin sizes. As expected, regions are not identified as significant for all bin sizes. As arm *2*R (Figure 2.4) shows, while regions harboring sharp peaks in *D. simulans* enrichment stand out, large regions with relatively shallow enrichment can still approach significance (*e.g*. centromeric region of *2*R). One's ability to detect significance at a given bin size is influenced by the physical size of the introgression block and the magnitude of SNP enrichment, thus it is difficult to predict what is the "right" bin size. When choosing a bin size it is more important to consider the biological and experimental conditions: recombination rate, generation number, gene density, influence of recombination hotspots, etc.

Independent Confirmation of the Effect of the X chromosomes.

The X chromosome appeared not to harbor any genes affecting aversion behavior. An alternate explanation for this pattern is that our single introgression did not capture all regions affecting aversion. (As noted above, we only captured about ¾ of the *D. simulans* phenotype with this single introgression line.) This will most likely occur within chromosomes where local recombination landscape can affect the efficacy of selection. We did not expect this problem within an independent linkage group.

Backcrossed *D. simulans/D. sechellia* flies with X-linked markers were subjected to an oviposition assay. When controlling for the influence of the X chromosome, we

confirmed the absence of its effect. Of the 169 flies with *D. sechellia* X, 77% avoided the Morinda medium. Of the 202 flies with *D. simulans* X, 73% avoided the Morinda medium. This pattern goes in the opposite direction predicted and is not statistically significant ($P = 0.4229$).

DISCUSSION

We have shown that PSIseq can efficiently map the genetic basis of a complex trait. This method can find multiple loci with small mapping populations that remain usable beyond mapping experiments. Other NGS methods rely on transient F_2 mapping populations which are irrevocably lost once nucleic acids are extracted. On the other hand, introgression stocks can be tested in future assays to further resolve candidate regions. Only low sequencing coverage is needed for rough mapping, and replicate populations can be created over time. We found at least six loci contributing to *D. simulans* aversion of Morinda compounds, and we confirmed the influence on behavior of an odorant binding protein locus discovered previously (*Obp56e*; DWORKIN and JONES 2009).

NGS mapping is typically composed of three main steps: 1) create a mapping population, 2) extract DNA and prepare libraries, 3) assemble short reads and genotype. PSIseq improves the first and third steps. First, PSIseq uses introgression lines propagated for more generations instead of an F_2 backcross line because this allows mapping of smaller candidate regions. Closely linked loci can be decoupled and mapped separately, especially if multiple replicate introgressions are made. For example, ANDOLFATTO *et al.* (2011), BAIRD *et al.* (2008), BLUMENSTIEL *et al.* (2009), HUANG *et*

al. (2009), SARIN *et al.* (2008), SCHNEEBERGER *et al.* 2009, and ZURYN *et al.* (2010) all used F_2 backcross populations and successfully mapped single loci from mutagenesis screens or dominant marker mapping. PSIseq was able to isolate at least six distinct loci in our pilot study, three of which occurred within a 7 Mb region (See Figure 2.1).

PSIseq also improves the third step in NGS mapping: genotyping. Instead of using only common SNPs (*a la* GWAS), PSIseq uses all possible SNPs and avoids potential bias in mapping to particular regions. A major issue with current short read technology is the high per base sequencing and mapping error rates. PSIseq overcomes this by using a window-based mapping approach, as described by Huang *et al.* (2009). A group of SNPs are collected into bins ("windows"), and statistical tests are performed on these bins. Thus, while a small percentage of SNPs may be inaccurate, these false SNPs are engulfed by surrounding populations of true SNPs. This method works particularly well when mapped regions are fixed for one parent and contain few polymorphisms (*eg.* RILs), and for rough mapping of hybrid genomes sequenced lightly. As an alternative to the window approach, Andolfatto *et al.* (2011) developed a customized Hidden-Markov Model (HMM) to assign the probability of a SNP's ancestry. The power of this method lies in its ability to map with high resolution the boundaries surrounding the recombination breakpoint itself. We take a more simplistic approach in identifying ancestry by measuring SNP enrichment under a binomial expectation.

PSIseq has additional advantages. First, the experimental replication needed to increase mapping resolution can be spaced out over time, and populations sizes within a replicate need only be large enough to maintain the introgression line - as small as one lineage per replicate. This approach allows us to increase introgression replicate sizes and

the quality and number of phenotypic measurements. Second, like sequenced RILS (HUANG *et al.* 2009), these selected introgression lines are potentially shelf stable and can be used in future experiments. Third, complex traits can be mapped in organisms lacking a high quality genome assembly. For example, with transcriptome sequence from parental and introgression lines one could identify gene transcripts enriched for alleles from the selected parent. Finally, in addition to being flexible and extensible, our approach only needs low sequence coverage (~1X) for rough mapping, given enough SNPs (*e.g.* ~100kb resolution). To increase resolution of candidate regions one could replicate introgression crosses (see Figure 2.2) or take advantage of new targeted enrichment sequencing (ALBERT *et al.* 2007; ANTSON *et al.* 2000) to increase coverage for only interesting regions. This strategy would be particularly useful for regions with repetitive elements, duplications, pseudogenes, and other elements difficult to align at lower coverage.

Our study measured a divergent trait between two species, but our approach can be applied to two genetically diverged individuals. The parental line would be sequenced, unique SNPs identified, and then introgression and selection. Typically, these strains would have a lower SNP density. We provide a simple model for estimating the expected mapping resolution for a trait given sequence divergence rate and bin size (Appendix A). For example, populations with relatively low sequence divergence (0.1%) and bins of 1000 SNPs each could be mapped to 2Mb regions within one replicate introgression lineage, depending on recombination rates, strength of phenotypic selection during introgression, and mapping quality.

REFERENCES

ALBERT, T. J., M. N. MOLLA, D. M. MUZNY, L. NAZARETH, D. WHEELER *et al.*, 2007 Direct selection of human genomic loci by microarray hybridization. Nat Methods **4:** 903-905.

AMLOU, M., B. MORETEAU and J. R. DAVID, 1998 Genetic analysis of Drosophila sechellia specialization: oviposition behavior toward the major aliphatic acids of its host plant. Behav Genet **28:** 455-464.

ANDOLFATTO, P., D. DAVISON, D. EREZYILMAZ, T. T. HU, J. MAST *et al.*, 2011 Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Res **21:** 610-617.

ANTSON, D. O., A. ISAKSSON, U. LANDEGREN and M. NILSSON, 2000 PCR-generated padlock probes detect single nucleotide variation in genomic DNA. Nucleic Acids Res **28:** E58.

BAIRD, N. A., P. D. ETTER, T. S. ATWOOD, M. C. CURREY, A. L. SHIVER *et al.*, 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One **3:** e3376.

BLUMENSTIEL, J. P., A. C. NOLL, J. A. GRIFFITHS, A. G. PERERA, K. N. WALTON *et al.*, 2009 Identification of EMS-induced mutations in Drosophila melanogaster by whole-genome sequencing. Genetics **182:** 25-32.

DESJARDINS, C. A., F. PERFECTTI, J. D. BARTOS, L. S. ENDERS and J. H. WERREN, 2010 The genetic basis of interspecies host preference differences in the model parasitoid Nasonia. Heredity **104:** 270-277.

DWORKIN, I., and C. D. JONES, 2009 Genetic changes accompanying the evolution of host specialization in Drosophila sechellia. Genetics **181:** 721-736.

EHRENREICH, I. M., N. TORABI, Y. JIA, J. KENT, S. MARTIS *et al.*, 2010 Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature **464:** 1039-1042.

FLIBOTTE, S., M. L. EDGLEY, I. CHAUDHRY, J. TAYLOR, S. E. NEIL *et al.*, 2010 Whole-genome profiling of mutagenesis in Caenorhabditis elegans. Genetics **185:** 431-441.

HUANG, X., Q. FENG, Q. QIAN, Q. ZHAO, L. WANG *et al.*, 2009 High-throughput genotyping by whole-genome resequencing. Genome Res **19:** 1068-1076.

JONES, C. D., 1998 The genetic basis of Drosophila sechellia's resistance to a host plant toxin. Genetics **149:** 1899-1908.

JONES, C. D., 2005 The genetics of adaptation in Drosophila sechellia. Genetica **123:** 137-145.

KLIMAN, R. M., P. ANDOLFATTO, J. A. COYNE, F. DEPAULIS, M. KREITMAN *et al.*, 2000 The population genetics of the origin and divergence of the Drosophila simulans complex species. Genetics **156:** 1913-1931.

LAITINEN, R. A., K. SCHNEEBERGER, N. S. JELLY, S. OSSOWSKI and D. WEIGEL, 2010 Identification of a spontaneous frame shift mutation in a nonreference Arabidopsis accession using whole genome sequencing. Plant Physiol **153:** 652-654.

LAURIE, C. C., J. R. TRUE, J. LIU and J. M. MERCER, 1997 An introgression analysis of quantitative trait loci that contribute to a morphological difference between Drosophila simulans and D. mauritiana. Genetics **145:** 339-348.

LI, H., and R. DURBIN, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:** 1754-1760.

MACKAY, T. F., E. A. STONE and J. F. AYROLES, 2009 The genetics of quantitative traits: challenges and prospects. Nat Rev Genet **10:** 565-577.

MANOLIO, T. A., F. S. COLLINS, N. J. COX, D. B. GOLDSTEIN, L. A. HINDORFF *et al.*, 2009 Finding the missing heritability of complex diseases. Nature **461:** 747-753.

MATSUO, T., S. SUGAYA, J. YASUKAWA, T. AIGAKI and Y. FUYAMA, 2007 Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in Drosophila sechellia. PLoS Biol **5:** e118.

MICHELMORE, R. W., I. PARAN and R. V. KESSELI, 1991 Identification of Markers Linked to Disease-Resistance Genes by Bulked Segregant Analysis - a Rapid Method to Detect Markers in Specific Genomic Regions by Using Segregating Populations. Proceedings of the National Academy of Sciences of the United States of America **88:** 9828-9832.

R'KHA, S., P. CAPY and J. R. DAVID, 1991 Host-plant specialization in the Drosophila melanogaster species complex: a physiological, behavioral, and genetical analysis. Proc Natl Acad Sci U S A **88:** 1835-1839.

SARIN, S., V. BERTRAND, H. BIGELOW, A. BOYANOV, M. DOITSIDOU *et al.*, 2010 Analysis of multiple ethyl methanesulfonate-mutagenized Caenorhabditis elegans strains by whole-genome sequencing. Genetics **185:** 417-430.

SARIN, S., S. PRABHU, M. M. O'MEARA, I. PE'ER and O. HOBERT, 2008 Caenorhabditis elegans mutant allele identification by whole-genome sequencing. Nat Methods **5:** 865-867.

SCHNEEBERGER, K., S. OSSOWSKI, C. LANZ, T. JUUL, A. H. PETERSEN *et al.*, 2009 SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat Methods **6:** 550-551.

TRUE, J. R., J. M. MERCER and C. C. LAURIE, 1996 Differences in crossover frequency and distribution among three sibling species of Drosophila. Genetics **142:** 507-523.

XIA, Y., S. WON, X. DU, P. LIN, C. ROSS *et al.*, 2010 Bulk segregation mapping of mutations in closely related strains of mice. Genetics **186:** 1139-1146.

ZURYN, S., S. LE GRAS, K. JAMET and S. JARRIAULT, 2010 A strategy for direct mapping and identification of mutations by whole-genome sequencing. Genetics **186:** 427-430.

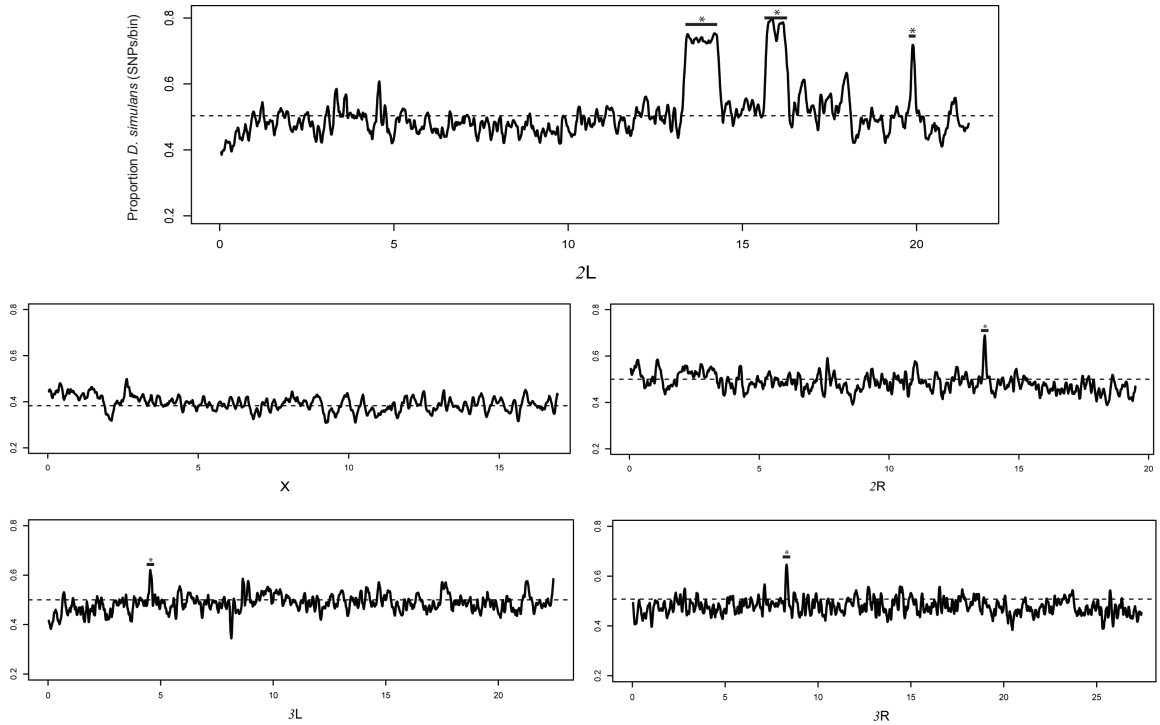

Figure 2.1 Fifteen generations of introgression show *D. simulans* enrichment for six regions across autosomes**.** Proportion *D. simulans* per bin was calculated via comparing hybrid SNPs to all possible species-level SNPs within a bin (1000 SNPs, overlapping, sliding). If all hybrid SNPs correspond to all possible *D. simulans* SNPs, the proportion is 1. If the opposite is true, all hybrid SNPs correspond to *D. sechellia* SNPs, the proportion is 0. Regions showing significant enrichment for *D. simulans* SNPs are marked (*, $P <$ 0.0002 or within-arm-FDR-correction). We estimate that these factors explain approximately 75% of *D. simulans* aversion behavior. Three loci were found clustered within a 7-Mb region on 2L (1, 0.5, and 0.3 Mb, in size, respectively, moving left to right).

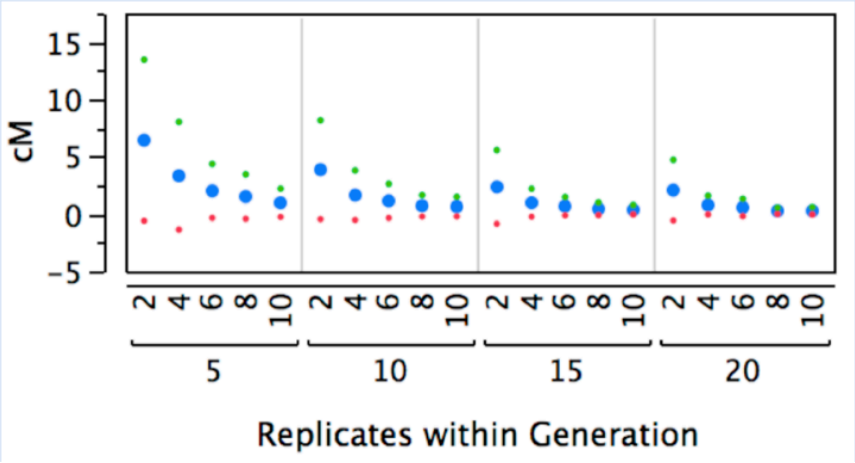

Figure 2.2 Mean size of candidate chromosomal region. Shown are results from a forward model estimating the size of introgressed blocks in a backcross with selection scenario. The block, *x,* is the chromosomal region uniquely overlapped by all introgression lines, *L*. The size of *x* will depend on the number of generations of introgression and the number of independent lines (Replicates). We simulated this process for 2–10 replicate lines and 5–20 generations. The average *x* is blue; green and red dots represent 1 SD above and below the average, respectively.

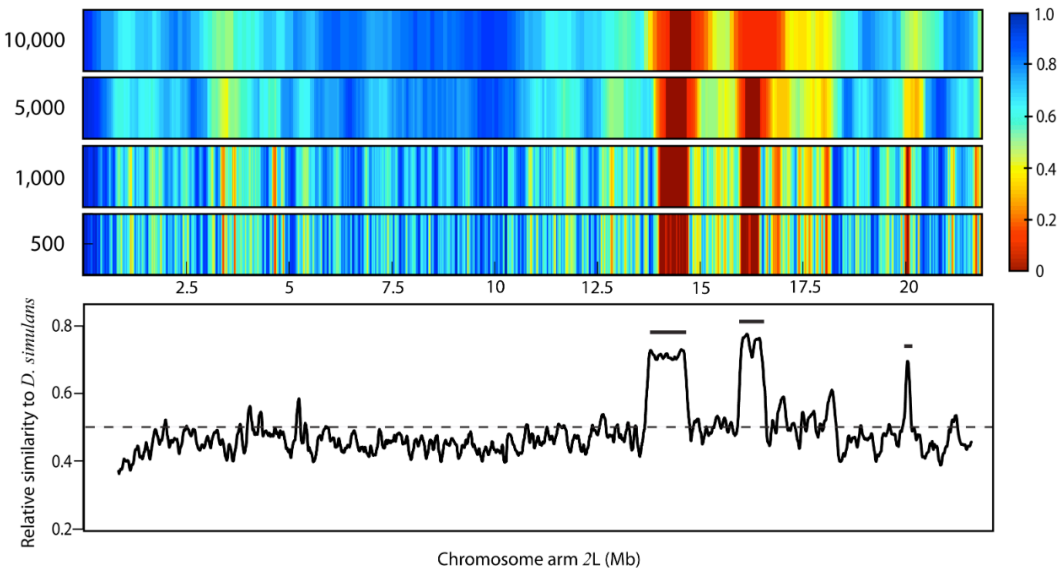

Figure 2.3 Ability to detect *D. simulans* enrichment with dynamic binning**.** Shown are sections from a heat plot of *P*-values on chromosome arm 2L from generation 15 hybrids, the same region shown in Figure 2.1. Four bin sizes were chosen to illustrate the binning effect described above: 500, 1000, 5000, and 10,000 SNPs/bin moving from middle to top. Heat plot colors correspond to corrected *P*-values (normal approximation of binomial; significance set by FDR); dark red regions have met the FDR threshold and are considered significantly enriched for *D. simulans*. The color gradient from light red to dark blue corresponds to scaled *P*-values, which are not significant. Bottom shows reproduction of Figure 2.1 for reference. Horizontal positions in the heat plots correspond to physical location on bottom.

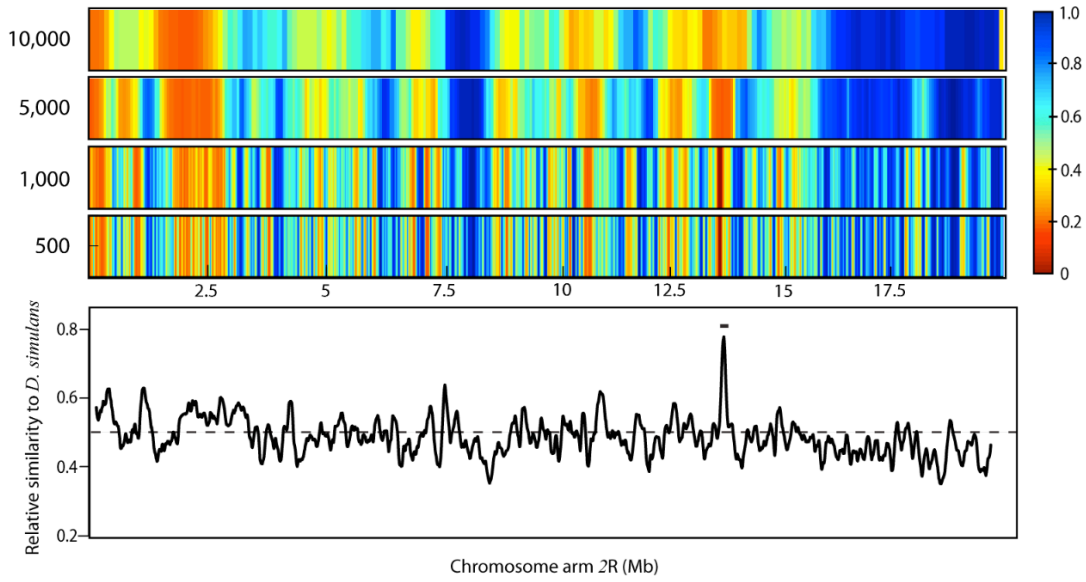

Figure 2.4 Dynamic binning on *2*R Identical to Figure 2.3, but for chromosome arm 2R. The single peak on this chromosome encompasses *Obp56e*, which is known to affect aversion behavior

III. HIGH RESOLUTION GENETIC ANALYSIS OF THE DIFFERENCE IN HOST PREFERENCE BEHAVIOR BETWEEN *DROSOPHILA SIMULANS* AND *D. SECHELLIA*

Authors: Eric J. Earley and Corbin D. Jones

ABSTRACT

Studies over the past 20 years suggest that relatively few genetic loci contribute to differences in host seeking behavior between closely related species of phytophagous insects. With so few loci involved and ample genetic variation, evolution of new behaviors could proceed rapidly. The methods and sample sizes used to find these genetic loci, however, were inherently biased towards mapping few loci of large effect size. A key question is whether different experimental approaches would yield fundamentally different numbers of loci. We took two complementary approaches to identifying genetic differences between *Drosophila simulans* and *D. sechellia* that contribute to differences in behavior towards the fruit and volatiles of *Morinda citrifolia* (Morinda), the preferred host of *D. sechellia.* Using whole genome sequencing and a recently introduced mapping method that tracks loci of variable effect size, we identified up to 27 loci contributing to Morinda host seeking behavior – far more loci than seen in previous genetic studies of host preference. In a parallel screen, we tested the function of specific candidate genes expressed in the peripheral nervous system – a likely target of host preference evolution – and found that genetic disruption of two receptors, *Ionotropic receptor 92a* and *75a*,

affected behavior towards Morinda or its volatiles. Our high resolution study suggests that the genetic basis of host preference is likely more genetically complex than previously thought and that evolution of preference may not evolve as easily as previously suggested.

INTRODUCTION

At its simplest, host seeking behavior is a dichotomy between generalists and specialists (FORISTER *et al.* 2012; FUTUYMA and MORENO 1988). Generalists can find nutrients from a diversity of sources, whereas specialists must focus on one or a few sources. Both groups often need particular morphological or physiological traits to consume nutrients, and evolution of these adaptations is critical to the evolution of a new specialism or the emergence of generalism. Ultimately, the evolutionary rate of any additive trait is determined by its standing genetic variation, which in turn is shaped by the allelic diversity and the number of contributing genes (TURELLI and BARTON 1994). If multiple genetic changes drive adaptation, then each new allele must incrementally increase fitness towards a new adaptive optimum. However, in many other scenarios, the complexity of an adaptation can be limiting (ORR 2000). For example, if evolving toward a new adaptive optimum requires traversing a fitness valley, then allelic combinations must be assembled in relative synchrony while minimizing deleterious pleiotropic effects. Thus, the number of genes necessary for adaptation has been suggested to affect the ease with which a new adaptation spreads through a population (GAVRILETS and VOSE 2005; ORR 1998).

A simple genetic path may be particularly important for ecological specialism, especially in phytophagous insects, which often need to evolve multiple traits to accommodate new plant hosts (MATSUBAYASHI *et al.* 2010). More than a dozen studies have mapped the genetic basis of host preference across a variety of these species and have discovered only 1-5 loci of moderate to strong effect in every case. This suggests that host preference can evolve rapidly and revert to an ancestral state easily. However, almost all of these studies used QTL mapping or a backcross with sparse genetic markers and relatively small sample sizes, biasing against discovery of minor effect loci. Adaptive traits in other systems also show surprisingly few contributing loci (NADEAU and JIGGINS 2010), yet in almost every case experimental procedure biased toward finding few loci of strong effect. Because of this, some have argued we may be misapprehending the true polygenic nature of adaptive genetics (ROCKMAN 2012).

To overcome the biases of QTL and linkage mapping, we investigated the genetics of a naturally occurring, adaptively divergent trait using two complimentary approaches to identify loci important for host preference: a high resolution mapping approach based on phenotype selection and introgression coupled with whole-genome resequencing (PSIseq; EARLEY and JONES 2011) and a targeted disruption of specific candidate genes. Both of these approaches can identify individual genes, including those with relatively weak effects.

We focused on the difference in host preference between *Drosophila sechellia* and its sister species*, D. simulans* and *D. melanogaster*. *D. sechellia* is a host specialist on the plant *Morinda citrifolia* (Morinda), whereas its sister species avoid this host (JONES 2005; TSACAS and BACHLI 1981). As host seeking behavior in phytophagous

insects is often mediated by chemotaxis (BERNAYS and CHAPMAN 1994; RAMDYA and BENTON 2010), genes expressed specifically in the peripheral nervous system – in particular chemoreceptors – likely contribute greatly to adaptive host preference. Changes in these genes could be critical to evolving preference toward new hosts, and mutations within a single chemosensory gene have been shown to change insect host seeking behavior significantly (MATSUO *et al.* 2007). Prior work, however, largely ignored *Ionotropic receptors* (*Ir*s; BENTON *et al.* 2009) and therefore we focused our efforts on these genes.

We found up to 27 loci contributing to host preference and resolved one interval to a single gene. Another region harbors a known contributing gene, *Odorant binding protein 57d/e* (MATSUO *et al.* 2007). We also verified three loci from an earlier mapping experiment (EARLEY and JONES 2011). To confirm the effects found in our genomic survey, we conducted an independent backcross experiment examining the effect of specific loci on chemotaxis. While some regions were confirmed, we also discovered that our genomic map has likely missed a region of strong effect at chromosome arm *2*L. Our candidate gene screen found no strong overlap between chemoreceptor genes and our PSIseq map, nor were many of the tested chemosensory genes important to avoidance behavior. However, we did find that two chemoreceptors, *Ir93a* and *Ir75a* increased Morinda aversion when knocked out.

MATERIALS AND METHODS

Introgression and backcrossing

Replicate populations of *D. simulans* (c167.4) females and *D. sechellia* (SynA) males were crossed and progeny female were backcrossed to *D. sechellia*. Backcrossed flies were subjected to a behavior assay that uses the chief volatiles of *Morinda citrifolia*, octanoic and hexanoic acid (FARINE *et al.* 1996; PINO *et al.* 2010), as previously published (DWORKIN and JONES 2009; EARLEY and JONES 2011). Our prior work shows that this assay captures on average ~68% of the parental difference in phenotype. Briefly, flies were introduced without anesthesia to a behavior arena (2L glass beaker, Fisher Scientific, Pittsburgh, PA) overnight (25°C, 50-60% humidity) containing two glass milk bottles, one with control food (22mL diH_2O, 4.1g instant fly media, Carolina Biological Supply Co.) or test food (control media with 0.2% Octanoic Acid and 0.06% Hexanoic Acid, v/v). Flies on control food – hence expressing dominant *D. simulans* aversion – were selected and backcrossed to *D. sechellia*. This cycle, phenotypic selection for *D. simulans*-like aversion and backcross to *D. sechellia*, was repeated for 15 generations (Figure 3.1). Aversion was tracked in this study for two reasons. Since aversion is dominant over preference in hybrids, we avoided the need to homozygose flies each generation to expose recessive alleles. We also assume that since these species are recently diverged, tracking aversion alleles will provide insight for preference, as well.

Illumina Sequencing and Assembly

Pooled females from two replicate lines were split into two groups per line producing two technical replicates for each biological replicates. DNA libraries were prepared and barcoded by standard Illumina protocols, producing an average 60,156,803 sequences (2x100) from a HiSeq 2000.

Bioinformatic methods are after (EARLEY and JONES 2011). Briefly, *D. sechellia* reference genome (r1.3) was fragmented *in silico* to an average 20x coverage, and these fake reads were aligned against the *D. simulans* reference genome (r1.4) to generate species-specific SNPs. We also performed a reciprocal alignment for select genomic regions to test the influence of mapping reference bias and found that using *D. simulans* as reference provided significantly higher quality SNP calls (measured with SAMtools mapping scores), which was not surprising since the *D. sechellia* genome is still in scaffold phase. Experimental hybrid sequences were aligned to the same genome. SNPs shared between the hybrid read assembly and the *D. sechellia* assembly were collected. For every 1,000 possible species SNPs, hybrid SNPs were counted along a sliding window with adjacent windows overlapping by 90% . Windows exhibiting SNP content significantly different from its chromosome arm mean were identified via repeated *t*-tests, with corrected alpha threshold via FDR=0.05. Alignments were performed with BWA (LI and DURBIN 2009), processed by SAMtools (LI *et al.* 2009), SNPs were called, correlated, and binned using custom Perl scripts, and statistics were performed in R (R DEVELOPMENT CORE TEAM 2012). All reported interval sizes were calculated from narrowed overlapping replicate coordinates, technical or biological, unless none were available.

Candidate Gene Knockdowns

All flies were reared at 22-25°C on agar-yeast-cornmeal medium. Gene knockdown tests were performed with *D. melanogaster* stocks from Bloomington stock center at 25°C with controlled humidity (50-60% humidity). *Ir* knockdowns were created

from P-element insertions within coding regions. *Or* knockdowns were created by expressing Tetanus toxin (TeTxLC), an inhibitor of synaptic vesicle function, driven by *Or* promoter elements via the GAL4/UAS system. Behavior was measured using the same behavior assay described above. Flies in test or control food were counted, and a response index was calculated: *RI = logit[#flies test / (#flies test + # flies control)]*.

Verification of Candidate Loci Effects Using a Backcross and Oviposition Assay

We used a standard backcross and an alternative assay to test the completeness of our PSIseq screen and to confirm candidate regions. Genetically marked *D. simulans* females (*Cy*, *Ubx,* and *Serr*) were crossed to *D. sechellia* males to make D. *simulans/D. sechellia* F_1 hybrids. Females were backcrossed to *D. sechellia* males, and F_2 progeny were assayed for oviposition preference. Visible markers in the backcross generation denote chromosome regions and not entire chromosomes because of recombination occurring in the F_1 hybrid females. Flies exhibiting *Cy* (genetic position 2L: 6.1), *Ubx* (physical position 3R: 0.02 Mb - 8.9 Mb; COYNE and SNIEGOWSKI 1994), or *Serr* (physical position *3*R: 22.8Mb) were heterozygous for *D. simulans* loci in those regions, whereas flies without these markers were homozygous *D. sechellia* on those regions. Gravid females were given a choice of oviposition substrates: one with octanoic acid (OA, 0.07%, a non-lethal dose, Sigma) and one without. Substrates were made with diH_2O and instant *Drosophila* medium (Carolina Biological Supply Co.). Females were allowed to oviposit for two days (25°C, 50-70% humidity), after which eggs were counted on each substrate and the female was tested again on fresh substrate. One replicate was one female tested over four total days. Combined totals on both test and

control media were calculated into a response index: *RI=logit[(#eggs test) / (#eggs test + #eggs control)]*

Molecular Evolutionary Analysis

dN/dS calculation for *GD13873 (CG4597*). was performed on *D. simulans* and *D. sechellia* coding sequences from flybase (GELBART *et al.* 1999) using MEGA5 (TAMURA *et al.* 2011).

RESULTS

Assembly and SNPs

We identified 1,778,101 SNPs in an alignment between *D. sechellia* and *D. simulans* (18.4x average coverage), which is an average of 1.7% divergence across 100Mb. Assembled introgression hybrid genomes had an average coverage of 13.7x (compared to 1x coverage in our previous study; EARLEY and JONES 2011) across two biological replicate lines with two technical replications each, and were predominantly *D. sechellia* (out of 1.7 million potential SNPs 1.6 million were *D. sechellia*).

Morinda seeking behavior is not genetically simple

We identified up to 27 PSIseq intervals contributing to host preference difference between *D. simulans* and *D. sechellia* (Figure 3.2, Table 3.1, Table 3.2). Roughly half of these loci (13) were confirmed by biological replication. We were able to map one region down to a single candidate gene, *GD13873 (CG4597*). All autosomes possessed multiple

loci, except for *4*. The *X* chromosome had no effect as previously shown (EARLEY and JONES 2011).

Three regions, two on chromosome *2*L and one on *3*L (Figure 3.1, Table 3.2), overlapped our previous lower resolution study and across both biological replicates (EARLEY and JONES 2011). The sizes of these intervals ranged from 1Mb to as small as 72.5Kb. The magnitude of *D. simulans* SNP enrichment and similarity of overlap of intervals on *2*L was particularly striking, potentially representing multiple linked loci of strong effect. In contrast, the small shared region on *3*L contains a single annotated gene, *GD13873*. Within these three regions were a total of 162 annotated genes.

Focusing on the overlap of the two bio-replicates within the current study produced 12 additional candidate intervals in addition to the three previous regions. All autosomes possessed at least one locus, and the region sizes ranged from 771.7Kb to 65.4Kb. Within these regions were 813 annotated genes with an average of 68 per region. Three more loci were mapped with at least technical replication but no biological replication. Sizes for these regions were consistent – 530Kb to 555Kb – and possessed a total of 71 genes. Without regard to overlap across either techno- or bio-replicates, we found nine additional candidate regions. In total, regardless of replication, up to 27 intervals contributed to the difference in host preference.

In our earlier study, we detected *Odorant Binding protein 56e (Obp56e),* which was previously identified as affecting behavior, but did not detect another locus, *Obp57d/e*, known to affect preference (MATSUO *et al.* 2007). In contrast, our new map identified *Obp57d/e,* but not *Obp56e*. In both cases, the signal was not strong (Figure 3.2 – see region surrounding position 15Mb on *2*R. The black bar, an interval from previous

work, captured *Obo56e*, whereas the current study captured *Obp57d/e,* within the small adjacent peak).

The two most striking PSIseq intervals were on *2*L. These were consistent across all experiments and technical replicates. Given their relatively large size we might expect multiple genes within these regions to be contributing to Morinda behavior. Surprisingly, out of 159 genes there are few annotated chemosensory or behavioral genes in either of these intervals. Some genes stand out due to their functional annotation within *D. melanogaster* orthologs: *bubblegum* (fatty acid metabolism—the Morinda toxins are medium chain fatty acids), *pickpocket* (post-mating behavior), *smell impaired 35A* (chemosensory behavior). Gene Ontology cluster analysis (HUANG *et al.* 2009a; HUANG *et al.* 2009b) for this region identified a number of enriched categories, including EGF-like (12 genes, IPR006210: *arrow,* nine members of the *nimrod* gene cluster, and one gene without an annotated *D. melanogaster* ortholog, *GD25758*, all of which have EGF-like protein domains), hormone and juvenile hormone binding (four genes, IPR013053: *CG33306, CG7916, CG7953, CG7968*), DNA metabolic process (seven genes, GO:0006259), and vesicle-mediated transport (eight genes, GO:0016192). None of these stand out as contributors to host seeking behavior, although enrichment of EGF-like domains suggests that nervous system development factors could be playing a role in this behavior.

Morinda aversion increased when Ir92a *and* Ir75a *are disrupted*

In addition to our genome-wide screen for loci contributing to behavior between *D. sechellia* and *D. simulans*, we also enacted a candidate gene screen within the more

genetically tractable species *D. melanogaster*, a close relative to *D. simulans* and *D. sechellia* (Table 3.2). Behaviorally, *D. melanogaster* and *D. simulans* were not significantly different ($P = 0.3509$), whereas the difference between *D. melanogaster* and *D. sechellia* was ($P = 0.0072$).

We initially targeted the *Ir*s because at least one, *Ir64a* , had been shown to affect response to some acids (AI *et al.* 2010) and the role of this family of chemoreceptors in *D. sechellia's* preference had not been investigated. We then broadened our screen to include some *Or*s due to their hypothesized effects (Or22a; DEKKER *et al.* 2006), species specific expression patterns (Or9a, Or65c; DWORKIN and JONES 2009), or because a weak PSIseq peak was near a candidate (*Or69a*). In all cases we were limited by the available genetic reagents at the time of the screen.

We disrupted chemoreceptor function using a variety of genetic methods in *D. melanogaster* and measured change in aversion to Morinda medium. None of the chemoreceptors tested reduced aversion to Morinda medium, but two appeared to *increase* aversion when knocked-out (*Ir75a*: $P = 0.0546$; and *Ir92a*: $P = 0.00792$). Neither of these *Irs* were present in a PSIseq interval, although *Ir92a* (Figure 3.2, at position *3*R: 5.3Mb) is close to a cluster of peaks on *3*R. *Ir75a* (position *3*L: 17.1Mb) results were particularly surprising as it exists in one of the few regions completely devoid of peaks. Two other receptors, although not significant, also showed a trend for increased aversion when knocked-out (*Ir31a*: $P = 0.0997$;*Ir93a*: $P = 0.0959$). *Ir93a* was the only *Ir* present in any of our intervals (Table 3.2, position *3*R: 4.7Mb). *Ir31a* (position *2*L: 10.2Mb), like *Ir75a* was also in a region with few peaks. None of these *Irs* exhibit expression divergence between *D. simulans* and *D. sechellia* (DWORKIN and JONES

2009). *Ir64a*, an acid sensing chemoreceptor, also had no effect ($P = 0.1214$). Intriguingly, none of the Ors affected behavior in our assay. This was surprising given that both *Or9a* and *Or65c* show significant expression divergence between *D. simulans* and *D. sechellia* (DWORKIN and JONES 2009) and *Or22a* has been implicated as important to preference and detection of hexanoic acid from neuro-physiological studies (DEKKER *et al.* 2006; HALLEM and CARLSON 2006).

Independent backcross confirms some intervals, but also reveals missed loci

Our data showed that some candidate regions had consistent effects across replicates and experiments. Other loci, including known factors, were more variable (*e.g. Obp56e* and *Obp57d/e*). This pattern could be a product of either the mapping approach, or the phenotypic assay (as noted previously, our assay conditions only capture about 68% of the parental phenotype). Earlier work using an oviposition-site preference assay confirmed the absence of an effect of the X (EARLEY and JONES 2011), suggesting that large regions without preference factors were correctly identified by PSIseq. Here we targeted three more subtle classes of intervals: (i) a region that appears to have peaks, but they are not significant (left end of *2*L), (ii) a region with a single peak in one experiment (right end of 3R), and (iii) a large multi-peak region with effects shared between experiments (left end of 3R).

We conducted an independent single generation backcross between *D. simulans* and *D. sechellia* using genetic markers (*Cy* at genetic position *2*L: 6.1*, Serr* at physical position *3*R: 22.79Mb, and *Ubx* , which is associated with an inversion spanning physical position *3*R: 0.1Mb to 9Mb) to track *D. simulans* chromosomal regions within a *D.*

sechellia genomic background. Chemo-tactic behavior, assayed with the same apparatus used to generate introgression flies above, was significantly altered when *Cy* was present in hybrid flies (Figure 3.3a, $P = 0.0067$). Surprisingly, this region of chromosome *2*L was not introgressed in our genome mapping population (Figure 3.2), although a cluster of peaks was proximate. We also measured oviposition behavior toward Morinda compounds and found *D. simulans* chromosomal regions on left *2*L (*Cy,* $P = 0.001$) and left *3*R (*Ubx,* $P < 0.0001$) significantly reduced preference for ovipositing on Morinda medium compared to siblings homozygous for *D. sechellia* in this region (Figure 3.3b). *Ubx* and its associated inversion confirmed the four peaks in this region. The middle right region of *3*R, however, did not significantly alter hybrid behavior (*Serr,* $P = 0.23$), although this appears to be driven more by the relative indifference of siblings with *D. sechellia* alleles in this region (Figure 3.3b).

DISCUSSION

Several studies have found relatively few genetic loci contributing to differences in host preference behavior among closely related phytophagous insects species (MATSUBAYASHI *et al.* 2010), suggesting that evolving new host preference could be relatively easy. While these early studies have provided an invaluable initial view of the genetics of these traits, building generalized models from them may be problematic. QTL of moderate or small effect are difficult to detect (MACKAY *et al.* 2009), especially if only a few markers are used and samples sizes are small. Studies of QTLs underlying standing genetic variation suggest that strong effect QTL themselves are often not single effect loci but are instead a complex cluster of multiple interacting loci (MACKAY *et al.* 2009).

Given these limitations, some have argued that our current understanding of adaptive genetics is fundamentally shaped more by our techniques than by actual evolutionary processes. That is, many loci of relatively small effect – or infinitesimally small – create adaptations (ROCKMAN 2012).

To potentially overcome these shortfalls, we combined two complementary approaches to identify genes important for differences host preference between *D. sechellia* and *D. simulans*. We used PSIseq to map genes important for *D. sechellia*'s aversion of the fatty acids in Morinda. Along with loci of major effect, this method can also potentially track loci of relatively small to moderate effect. In parallel, we genetically ablated or altered the expression of specific candidate genes of the chemosensory system. We discovered up to 27 loci contributing to host preference behavior and found that knockdown of *Ir92a* and *Ir75a* significantly influenced response to OA and HA.

Previously, a number of genes and genetic loci have been implicated in *D. sechellia* preference for Morinda and its compounds (DEKKER *et al.* 2006; DWORKIN and JONES 2009; EARLEY and JONES 2011; HIGA and FUYAMA 1993; MATSUO *et al.* 2007), but only two genes were tested directly: *Obp57d/e* (MATSUO *et al.* 2007) and *Obp56e* (DWORKIN and JONES 2009). At first glance, these earlier results appear to be consistent with the simple genetic model for host preference. However, our recent data (EARLEY and JONES 2011) and the present study suggest that many other factors are potentially contributing to this behavior. We estimate there are at minimum nine and as many as 27 loci contributing to this behavior (Figure 3.2, Table 3.1, Table 3.2). We also note that our introgression did not track any recessive aversion loci. Additionally, although we focused

predominantly on behavior, our introgression assay allowed free movement of flies between alternate foods, and thus we could have captured loci governing performance on these foods.

Compared with our earlier work, the additional biological and technical replication and increased the sequence coverage (13x more) in the present study allowed far higher power to map regions in high resolution and make highly accurate SNP calls. We confirmed three major loci from the earlier work and uncovered many previously unidentified loci. However, these data are not complete as our independent backcross data suggest we have missed another locus near the tip of *2*L. In agreement with previous work, the *X* and dot *4* chromosomes harbored no factors (EARLEY and JONES 2011; JONES 2005).

The varying levels of technical and biological replication among these 27 loci could be caused by a number of reasons. Although our theoretical coverage is high (~50x), our realized mapping coverage is lower (~13x) so false positive SNP calls may skew our ability to identity species of origin for a given bin. This is unlikely for two reasons: first, we cross referenced SNP calls in our introgression hybrid read alignment with a separate *D. simulans/D. sechellia* alignment and only species-specific SNPs shared between these alignments were considered; second, we calculated the average percent identity of a bin across 1,000 possible SNPs so any given false positive from both alignments would be obscured by its neighborhood of accurate SNPs. More importantly, all of these technical issues were more likely to lead to missing true SNPs (false negatives) than miscalling false SNPs (false positive) and would thus conservatively bias our map and reduce the number of significant loci.

Two pieces of evidence suggested that a strong factor for Morinda behavior exists on the left part of *2*L. *D. sechellia* has an overabundance of a particular type of sensory hair (ab3) that is sensitive to Morinda, and *Or22a,* located in this region, is expressed within these hairs (DEKKER *et al.* 2006). However, our gene knockout experiments showed that *Or22a* had no effect on aversion in *D. melanogaster.* Oviposition preference for Morinda compounds also appears to be controlled by one or more factors in this region (JONES 2005). We were surprised that no *significant* peaks were present in our map for this region (Figure 3.2), but results from our independent backcross showed that at least one factor does exist here (Figure 3.3). One possibility explaining why no introgressions occurred here across three separate experiments is that hybrid fitness effects in this region could be killing flies. Early generation hybrids will likely suffer from intrinsic reproductive incompatibilities like sterility or reduced fertility (COYNE and ORR 1998), and middle to late generation flies could experience reduced fitness relative to flies which lose introgressions due to competitive exclusion (FANG *et al.* 2012). These processes are unlikely to maintain non-causal introgressions over 15 generations, but they are likely to remove introgressions that contribute weakly to behavior or that are genetically linked to hybrid fitness effects.

Chemoreceptor genes have been suggested to be important for the evolution of host preference (RAMDYA and BENTON 2010). Indeed *D. sechellia* has an elevated rate of chemosensory gene evolution (MCBRIDE 2007; MCBRIDE *et al.* 2007). Some of these genes are known to influence chemo taxis and/or ovipositional behavior towards Morinda and its compounds: *Obp56e* (DWORKIN and JONES 2009; EARLEY and JONES 2011) and *Obp57de* (MATSUO *et al.* 2007). Thus, we expected to see an enrichment of

chemoreceptors and other peripheral nervous system genes in our map. Surprisingly, this was not the case. We found 12 chemoreceptors out of 1,143 genes (~1%) across all intervals with at least technical replication, which roughly corresponds to the genome-wide ratio of chemoreceptors to all genes. We did find *Obp57d/e* within one region, which is known to affect *D. melanogaster* ovipositional behavior (MATSUO *et al.* 2007). *Ir93a*, also found within one of our regions appeared to increase *D. melanogaster* aversion to OA/HA when disrupted, but this was not significant.

Given the potential importance of chemosensory genes, we knocked down or disrupted 16 individual genes within the IR and OR chemoreceptor super families (Table 3.3). None of our tested genes reduced *D. melanogaster* aversion, however, two genes appeared to *increase* aversion: *Ir92a* and *Ir75a*. These genes were not in PSIseq intervals nor are they reported to have species-specific expression patterns. *Or22a*, a candidate olfactory receptor for *D. sechellia* preference for Morinda (DEKKER *et al.* 2006), did not alter Morinda behavior when knocked down in *D. melanogaster*. While the possibility remains that *D. sechellia* uses *Or22a* to prefer Morinda via its attraction for methyl hexanoate, it does not appear to be needed for aversion of OA and HA. Disrupting *Ir64a*, an acid sensing chemoreceptor (AI *et al.* 2010), also had no effect on aversion. In sum, our targeted gene approach was not particularly effective at identifying genes contributing to the behavioral difference among these species.

In contrast, our introgression identified a single gene, *GD13873* (*CG4597*, Figure 3.2, Table 3.2). This region was among the most consistently introgressed, occurring in three separate experiments. Unfortunately, little is known about this gene and there are minimal genetic reagents available for testing its effect. No Gene Ontology annotation is

present, and no expression difference is observed between *D. sechellia* and *D. simulans* (DWORKIN and JONES 2009). In a *D. melanogaster* screen for transcripts in the mushroom body, the first olfactory processing center within the fly's brain, KOBAYASHI et al. (2006) found that chemical ablation of the mushroom body in larvae qualitatively reduced expression of this gene in adults, but this reduction was not significant and occurred in only one out of three experiments. If this gene contributed to *D. sechellia*'s specialization on Morinda, then it may have experienced either purifying or positive natural selection. We tested for signatures of natural selection between *D. sechellia* and *D. simulans* coding regions using *dN/dS,* a ratio of non-synonymous (protein coding) nucleotide differences versus synonymous (silent) differences. This test ignores insertions/deletions and only measures coding differences between homologous codons. No evidence of positive selection ($dN/dS > 1$, $P \sim 1.0$), nor strong purifying selection ($dN/dS < 1$, $P = 0.123$) was present, suggesting that the molecular differences between these species are evolutionarily neutral.

Overall, biological and technical replication improved our power to identify loci contributing to Morinda compound behavior divergence between *D. sechellia* and *D. simulans.*. Using high resolution markers and high depth sequence coverage, we discovered many loci participating in host seeking behavior. Our data suggests that experimental methods *can* bias our view of genetic complexity, and that the evolution of host preference is more genetically complex than we thought. Candidate gene tests did not correlate well to our genomic map, although this could have reflected our *a priori* decision to test the *Ir* family. On the other hand, *Ir64a* results showed that just because a gene *can* respond to a chemical cue within a species (AI *et al.* 2010) does not mean it will

play a role between species. However, extant genetic factors may not represent those that were required for evolution in the first place.

Our study provides a deeper understanding of how novel host specialization and related complex behaviors evolve and should act as a springboard for future studies. In light of recent controversy over whether the genetics of adaptations is generally simple or complex (ROCKMAN 2012), our work confirms that studies tracking few causal loci of strong effect are indeed methodologically biased. Yet, our results also show that while the genetics of adaptation is complex, it is not caused by many loci of infinitesimal effect size, and that with a suitable design key genetic players can be identified.

ACKNOWLEDGEMENTS

The authors would like to thank K. Peck, A. Romanchuk, L. Bono, C. Burch, A. Whitlock for editorial comments. Funding was provided by NSF Pre-doctoral fellowship (EJE).

REFERENCES

AI, M., S. MIN, Y. GROSJEAN, C. LEBLANC, R. BELL *et al.*, 2010 Acid sensing by the Drosophila olfactory system. Nature **468:** 691-695.

BELLEN, H. J., R. W. LEVIS, G. LIAO, Y. HE, J. W. CARLSON *et al.*, 2004 The BDGP gene disruption project: single transposon insertions associated with 40% of Drosophila genes. Genetics **167:** 761-781.

BENTON, R., K. S. VANNICE, C. GOMEZ-DIAZ and L. B. VOSSHALL, 2009 Variant ionotropic glutamate receptors as chemosensory receptors in Drosophila. Cell **136:** 149-162.

BERNAYS, E. A., and R. F. CHAPMAN, 1994 *Host-plant selection by phytophagous insects*. Chapman and Hall, New York.

COYNE, J. A., and H. A. ORR, 1998 The evolutionary genetics of speciation. Philos Trans R Soc Lond B Biol Sci **353:** 287-305.

COYNE, J. A., and P. D. SNIEGOWSKI, 1994 A third-chromosome balancer in *Drosophila simulans*. DIS **75:** 36-37.

DEKKER, T., I. IBBA, K. P. SIJU, M. C. STENSMYR and B. S. HANSSON, 2006 Olfactory shifts parallel superspecialism for toxic fruit in Drosophila melanogaster sibling, D. sechellia. Curr Biol **16:** 101-109.

DWORKIN, I., and C. D. JONES, 2009 Genetic changes accompanying the evolution of host specialization in Drosophila sechellia. Genetics **181:** 721-736.

EARLEY, E. J., and C. D. JONES, 2011 Next-generation mapping of complex traits with phenotype-based selection and introgression. Genetics **189:** 1203-1209.

FANG, S., R. YUKILEVICH, Y. CHEN, D. A. TURISSINI, K. ZENG *et al.*, 2012 Incompatibility and Competitive Exclusion of Genomic Segments between Sibling Drosophila Species. PLoS Genet **8:** e1002795.

FARINE, J. P., L. LEGAL, B. MORETEAU and J. L. LEQUERE, 1996 Volatile components of ripe fruits of Morinda citrifolia and their effects on Drosophila. Phytochemistry **41:** 433-438.

FORISTER, M. L., L. A. DYER, M. S. SINGER, J. O. STIREMAN and J. T. LILL, 2012 Revisiting the evolution of ecological specialization, with emphasis on insect-plant interactions. Ecology **93:** 981-991.

FUTUYMA, D. J., and G. MORENO, 1988 The Evolution of Ecological Specialization. Annual Review of Ecology and Systematics **19:** 207-233.

GAVRILETS, S., and A. VOSE, 2005 Dynamic patterns of adaptive radiation. Proc Natl Acad Sci U S A **102:** 18040-18045.

GELBART, W. M., M. CROSBY, B. MATTHEWS, J. CHILLEMI, S. R. TWOMBLY *et al.*, 1999 The FlyBase database of the Drosophila genome projects and community literature. Nucleic Acids Research **27:** 85-88.

HALLEM, E. A., and J. R. CARLSON, 2006 Coding of odors by a receptor repertoire. Cell **125:** 143-160.

HIGA, I., and Y. FUYAMA, 1993 Genetics of food preference in Drosophila sechellia. I. Responses to food attractants. Genetica **88:** 129-136.

HUANG, D. W., B. T. SHERMAN and R. A. LEMPICKI, 2009a Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Research **37:** 1-13.

HUANG, D. W., B. T. SHERMAN and R. A. LEMPICKI, 2009b Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc **4:** 44-57.

JONES, C. D., 2005 The genetics of adaptation in Drosophila sechellia. Genetica **123:** 137-145.

KOBAYASHI, M., L. MICHAUT, A. INO, K. HONJO, T. NAKAJIMA *et al.*, 2006 Differential microarray analysis of Drosophila mushroom body transcripts using chemical ablation. Proc Natl Acad Sci U S A **103:** 14417-14422.

LI, H., and R. DURBIN, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:** 1754-1760.

LI, H., B. HANDSAKER, A. WYSOKER, T. FENNELL, J. RUAN *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics **25:** 2078-2079.

MACKAY, T. F., E. A. STONE and J. F. AYROLES, 2009 The genetics of quantitative traits: challenges and prospects. Nat Rev Genet **10:** 565-577.

MATSUBAYASHI, K. W., I. OHSHIMA and P. NOSIL, 2010 Ecological speciation in phytophagous insects. Entomologia Experimentalis Et Applicata **134:** 1-27.

MATSUO, T., S. SUGAYA, J. YASUKAWA, T. AIGAKI and Y. FUYAMA, 2007 Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in Drosophila sechellia. PLoS Biol **5:** e118.

MCBRIDE, C. S., 2007 Rapid evolution of smell and taste receptor genes during host specialization in Drosophila sechellia. Proc Natl Acad Sci U S A **104:** 4996-5001.

MCBRIDE, C. S., J. R. ARGUELLO and B. C. O'MEARA, 2007 Five Drosophila genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. Genetics **177:** 1395-1416.

NADEAU, N. J., and C. D. JIGGINS, 2010 A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. Trends Genet **26:** 484-492.

ORR, H. A., 1998 The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. Evolution **52:** 935-949.

ORR, H. A., 2000 Adaptation and the cost of complexity. Evolution **54:** 13-20.

PINO, J. A., E. MARQUEZ, C. E. QUIJANO and D. CASTRO, 2010 Volatile compounds in noni (Morinda citrifolia L.) at two ripening stages. Ciencia E Tecnologia De Alimentos **30:** 183-187.

R DEVELOPMENT CORE TEAM, 2012 R: A language and environment for statistical computing, pp. R Foundation for Statistical Computing, Vienna, Austrai.

RAMDYA, P., and R. BENTON, 2010 Evolving olfactory systems on the fly. Trends Genet **26:** 307-316.

ROCKMAN, M. V., 2012 The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution **66:** 1-17.

TAMURA, K., D. PETERSON, N. PETERSON, G. STECHER, M. NEI *et al.*, 2011 MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **28:** 2731-2739.

TSACAS, L., and G. BACHLI, 1981 *Drosophila sechellia*, n.sp., huitieme espece du sous-goupe melanogaster des Iles Sechelles [Diptera, Drosophilidae]. Revue Fr. Ent. NS**:** 146-150.

TURELLI, M., and N. H. BARTON, 1994 Genetic and statistical analyses of strong selection on polygenic traits: what, me normal? Genetics **138**: 913-941.

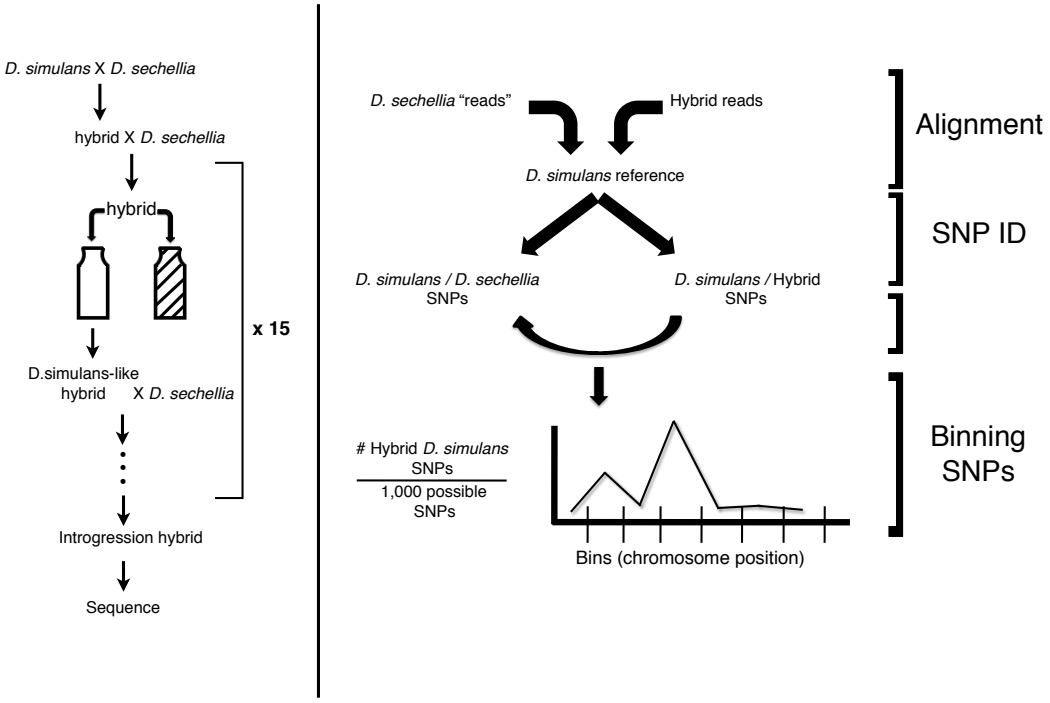

Figure 3.1 Creating hybrid introgressions and informatic pipeline. Left, species were hybridized and backcrossed to *D. sechellia,* and F_2 offspring were assayed for *D. simulans*-like aversion to Morinda compounds. Aversive flies were again backcrossed to *D. sechellia*. This process, selection for aversion and backcross, was repeated for 15 generations, after which introgression flies were selfed for ~5 generations. Right, females from two replicate lines were pooled into two technical replicate pools for each introgression line (15 each for four total pools). Introgression hybrid sequences were assembled and compared to *D. simulans/D. sechellia* SNPs in a separate assembly. These confirmed SNP counts were binned into windows of 1,000 possible species SNPs, and each window's SNP content was tested against its chromosome arm's average SNP content.

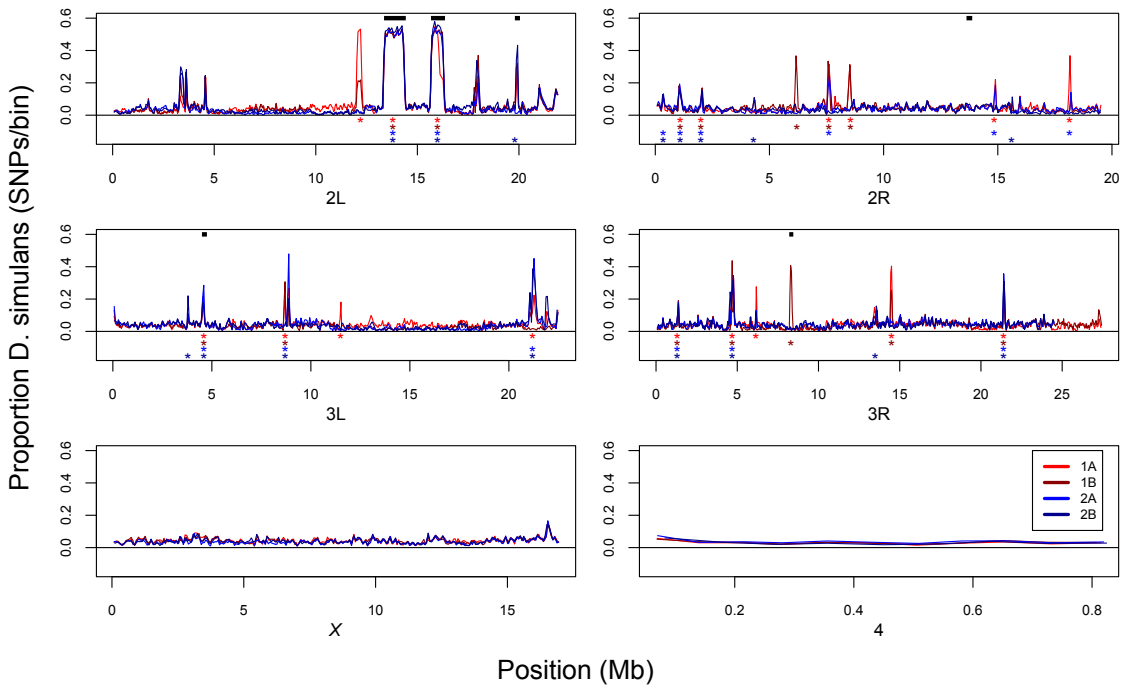

Figure 3.2 Genome map of introgression break points. Sliding windows of *D. simulans* SNP content over 1,000 possible species SNPs for each chromosome arm. Y-axis tracks the percent *D. simulans* SNPs out of 1,000 possible. X-axis tracks physical position along chromosome arm (Mb). Red (light and dark) and blue (light and dark) lines are results from each biological replicate, with darker and lighter shades denoting its technical replicate. Displayed results were averaged over six bins to remove extraneous noise. Stars beneath x-axis denote bins significantly enriched for *D. simulans* SNPs for that replicate (unpaired Student's t test, $P <$ FDR for that chromosome arm). Black bars track bins from a previous study (EARLEY and JONES 2011).

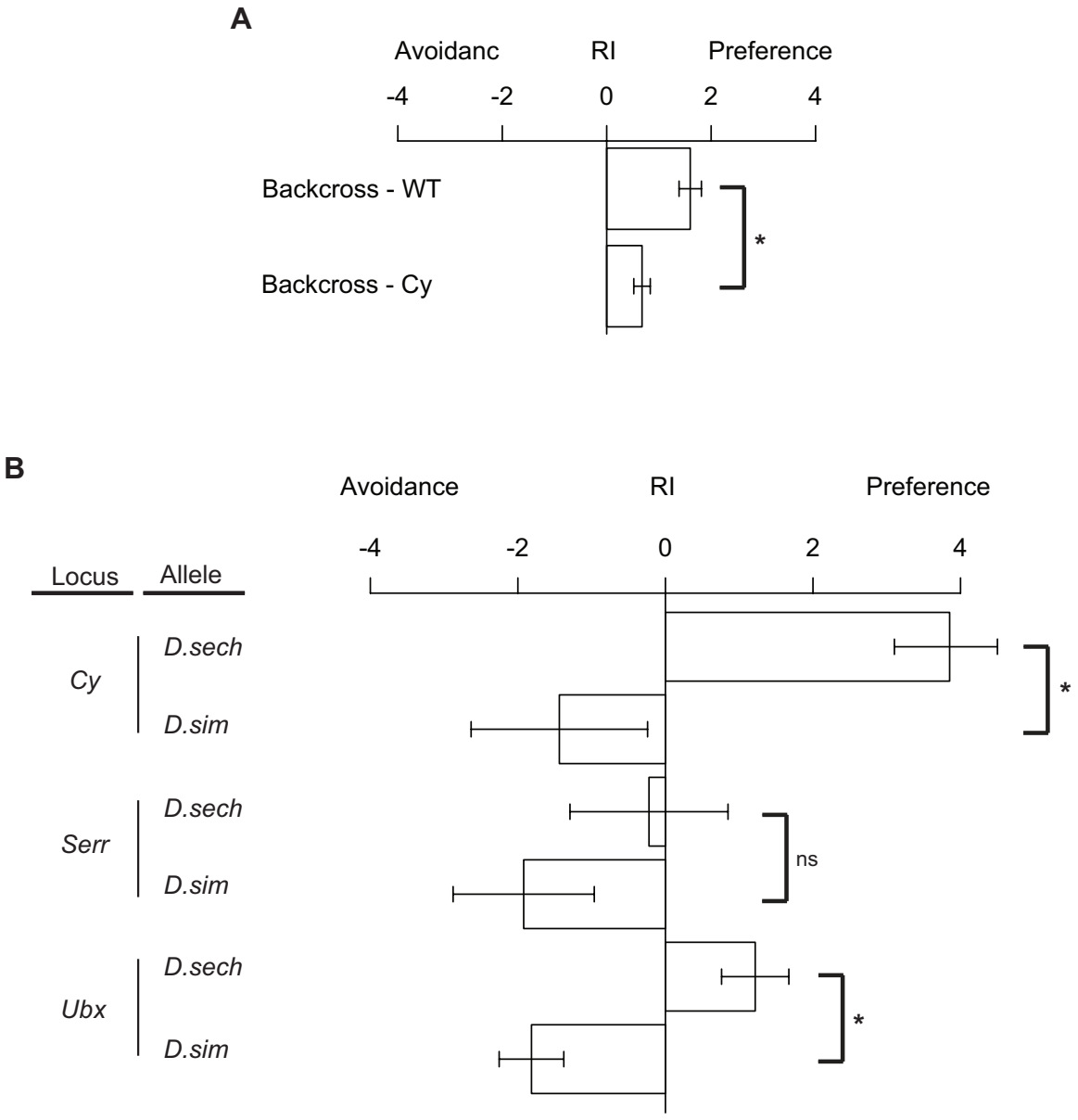

Figure 3.3 Independent backcross testing specific *D. simulans* regions. (A) Using the same phenotypic assay from the PSIseq map (Figure 3.2), one generation backcross flies showed significantly less preference for Morinda compounds when harboring *D. simulans Cy* and surrounding loci compared to siblings with homozygous *D. sechellia* alleles in this region. (B) Using an oviposition assay, backcross flies harboring *D. simulans* regions around *Cy* and *Ubx* also show significant oviposition aversion to Morinda compounds compared to siblings harboring homozygous *D. sechellia* alleles in the same regions (unpaired Student's t test comparing marked flies – *Cy*, *Ubx, Serr* – versus siblings without markers, $*P < 0.01$).

Table 3.1 Count of loci across studies, biological and technical replications

	Number of Loci
Previous Study[a]	6
Current Study	
Biological Rep. 1 Total[b]	20
Technical Rep. A	18
Technical Rep. B	14
Biological Rep. 2 Total[b]	19
Technical Rep. A	14
Technical Rep. B	16
Overall Total[b]	**27**

[a] (EARLEY and JONES 2011)
[b] Cumulative number of loci with overlap between technical replicates and/or studies; not a sum.

Table 3.2 Regions with significantly high *D. simulans* SNP content in the current study.

Chromosome arm	Begin[a]	End[a]	Number of Replicates[b]	Number of Genes	Interesting Candidates within Region
2L	13426524	14429664	2[c]	107	
2L	15762616	16360555	2[c]	51	
2L	19933433	19998886	2	7	
2R	356346	886722	1 (w/ technical replicate)	69	*Or42*
2R	1063857	1608601	2	63	
2R	2056061	2621232	2	107	*Gr43b, Or43b*
2R	7570243	8110418	2	66	
2R	8521954	9047859	1 (w/ technical replicate)	41	*Obp50*
2R	14896491	15381909	2	61	*Obp57, Gr57a*
2R	18199938	18730174	2	113	
3L	4568152	4640704	2[c]	1	*GD13873*
3L	8882610	9359951	2	76	*Or67, Ir67a*
3L	21267565	22039230	2	64	
3R	1382007	1910653	2	89	
3R	4689542	5280768	2	83	*Or92a, Ir93a*
3R	8243750	8449654	2	14	
3R	14446950	15002666	1 (w/ technical replicate)	61	
3R	21435650	21938376	2	70	

[a] Coordinates in *D. simulans* reference genome (r1.4).
[b] Biological replication, unless otherwise noted.
[c] Including both biological replication and matching previous study (EARLEY and JONES 2011)

Table 3.3 Chemosensory disruption results in *D. melanogaster*

Genotype[a]	Response different from *D. melanogaster*?
D. sechellia	**0.007187**
D. simulans	N.S.
Ir8a[b]	N.S.
Ir21a[b]	N.S.
Ir31a[b]	0.0997
Ir64a[b]	N.S.
Ir75a[b]	**0.05456**
Ir75d[b]	N.S.
Ir76a[b]	N.S.
Ir76b[b]	N.S.
Ir84a[b]	N.S.
Ir92a[b]	**0.00792**
Ir93a[b]	0.0959
Or9a[c]	N.S.
Or9a[d]	N.S.
Or22a[d]	N.S.
Or65c[c]	N.S.
Or69a[c]	N.S.

[a] All genotypes are *D. melanogaster*, unless noted.
[b] Disrupted coding region with transposon insertion (BELLEN *et al.* 2004)
[c] Knock out with cell apoptotic driver, *reaper*.
[d] Knock down with tetanus toxin light chain (*TNTE*), preventing synaptic transmission between neurons.

IV. FIRST STEP TO SPECIALISM: PSEUDOGENIZATION OF A GUSTATORY RECEPTOR WAS KEY INNOVATION DURING EVOLUTION OF SUPERSPECIALISM IN *DROSOPHILA SECHELLIA*

Authors: Eric J. Earley and Corbin D. Jones

ABSTRACT

To find food and mates, and to avoid threats, animals must process information from chemical signals into action. While models like *Drosophila melanogaster* are elucidating the genetic and molecular basis of chemically induced behavior, we still do not yet understand how evolution shaped these behaviors in the first place. We present a case study that supports a multi-step evolutionary model (MATSUO *et al.* 2007) of chemically induced behavioral preference toward the plant, *Morinda citrifolia* (Morinda), in the hyperspecialist, *Drosophila sechellia.* Other Drosophilids, like *D. melanogaster*, find Morinda repellent and toxic, and this aversion is likely ancestral. We systematically ablated function of major odorant and gustatory systems in *D. melanogaster* to identify the necessary genetic components of Morinda aversion with the goal of identifying factors potentially driving preference in *D. sechellia*. We found that antennae and tarsi were necessary for aversion, but *Odorant receptors* were not. One specific *Gustatory receptor*, *Gr22c*, was necessary, and *D. sechellia* itself harbors a *Gr22c* pseudogene (MCBRIDE 2007). Naturally occurring pseudogenized *Gr22c* alleles are segregating in wild *D. melanogaster* populations, suggesting this allele is neutral or weakly deleterious

and also showing the possibility that similar alleles could have been segregating in the standing genetic variation of *D. sechellia's* ancestor (LANGLEY *et al.* 2012; MACKAY *et al.* 2012). Together, these data suggest that ancestral flies initially lost aversion to Morinda and subsequently gained preference in at least two genetic events.

RESULTS/DISCUSSION

D. sechellia prefers, *D. melanogaster* avoids, specific Morinda compounds

Drosophila sechellia, a genetic model of host specialism, prefers the ripe fruit of *Morinda citrifolia* (Morinda; JONES 2005; R'KHA *et al.* 1991) and its constituent chemicals, including octanoic acid (OA; AMLOU *et al.* 1998; DWORKIN and JONES 2009; EARLEY and JONES 2011; MATSUO *et al.* 2007), hexanoic acid (HA; AMLOU *et al.* 1998; DEKKER *et al.* 2006; DWORKIN and JONES 2009; EARLEY and JONES 2011; HIGA and FUYAMA 1993), and methyl hexanoate (DEKKER *et al.* 2006). All other Drosophilids, including *D. melanogaster*, avoid Morinda and several of its constitutive chemicals, suggesting this aversion is ancestral (Figure 4.1a) (AMLOU *et al.* 1998; EARLEY and JONES 2011; LEGAL *et al.* 1992; MATSUO *et al.* 2007; R'KHA *et al.* 1991). The most abundant volatile organic compound in Morinda, OA (50-70%; FARINE *et al.* 1996; PINO *et al.* 2010), is unique in that it is both highly toxic and aversive to most Drosophilids (AMLOU *et al.* 1997) yet tolerated and strongly preferred by *D. sechellia*. HA, the second most abundant Morinda volatile (8-19%), is also toxic and aversive at high concentrations to non-*sechellia* species (AMLOU *et al.* 1997). We hypothesized that a mixture of OA and HA are two of the main chemical cues driving preference and

avoidance of Morinda, and we sought to identify the genes responsible for the transition between aversion and preference toward OA and HA.

To more accurately mimic Morinda fruit, we tested OA and HA as a mixture with relative concentrations (3:1, OA:HA) similar to that of Morinda. Our behavioral assay allowed free movement of flies between OA:HA and control foods and likely captured a more complex mix of phenotypes, including settling behavior, oviposition preference, and seeking behavior (Figure 4.1b). Compared to ripe Morinda, pure OA:HA induced similar behaviors in wild type *D. melanogaster* and *D. sechellia* at low acid concentrations (OA = 0.2%, HA = 0.06%, v/v). *D. melanogaster* is slightly less aversive to OA:HA than Morinda (Figure 4.1c), suggesting that the repellent nature of OA:HA is stronger in the context of other chemicals.

Tarsi and antennae necessary for OA:HA aversion.

Preference toward Morinda appears to be driven in part by olfaction in *D. sechellia* (R'KHA *et al.* 1991).Whether aversion in other species is driven primarily by gustation or olfaction is not fully understood. Like many insects, *Drosophila* use olfactory hairs (or "sensilla", located along antennae and maxillary palps) to smell and gustatory sensilla (on the proboscis, tarsi, wings, and ovipositor) to taste. Previous work suggested that *D. melanogaster* aversion is mediated in part by olfaction, as well (DEKKER *et al.* 2006; HIGA and FUYAMA 1993; MATSUO *et al.* 2007). We tested the influence of 3rd segment antennae, a major olfactory organ in *Drosophila*, on both aversion and preference (Figure 4.2a). When both antennae are removed, *D. melanogaster* lost aversion and *D. sechellia* lost preference compared to intact

individuals. With only one antenna removed *D. sechellia* showed a qualitative step-wise loss of preference compared to intact *D. sechellia* (Figure 4.2a). *D. melanogaster*, however, had greatly reduced aversion with only one antenna. Our results confirm that antennae are important for *D. sechellia* to seek out Morinda and that one antennae is sufficient for this behavior, supporting reports that *D. sechellia* is hypersensitive to Morinda compounds (DEKKER *et al.* 2006; R'KHA *et al.* 1991). In contrast, *D. melanogaster* requires at both antennae to avoid OA:HA.

To test the influence of contact chemosensation via gustation on OA:HA aversion and preference, we physically removed foreleg tarsi from both species (Figure 4.2a). Gustatory sensilla on the tarsi allow the fly to sample tastants without actually ingesting them. Without tarsi, *D. melanogaster* lost aversion compared to intact *D. melanogaster,* and *D. sechellia* without tarsi trended towards a loss of preference compared to intact *D. sechellia,* but not significantly. These data suggest that gustatory receptor neurons on the tarsi could be important for OA:HA aversion and potentially preference, suggesting that *D. melanogaster* requires both antennae and tarsi to avoid OA:HA.

Gustatory Receptors, But Not Olfactory Receptors, Are Necessary

Housed within olfactory and gustatory sensilla are sensory neurons that express transmembrane receptors (chemoreceptors) that translate chemical information into electrical signals (BENTON *et al.* 2009; MONTELL 2009; VOSSHALL and STOCKER 2007). Three main families of chemoreceptors, olfactory receptors (*Or*), ionotropic receptors (*Ir*), and gustatory receptors (*Gr*). The first two detect odorants (*Or*, *Ir*); the latter detect tastants (*Gr*). Along with other organs, tarsi sensilla house neurons expressing *Grs*, while

antennae predominantly house neurons expressing *Or*s and *Ir*s. From results above, we know that both antennae and tarsi are needed for *D. melanogaster* OA:HA aversion, suggesting that some combination of *Or*s, *Ir*s, and *Gr*s are necessary. To identify which of these gene families were necessary for *D. melanogaster* aversion, we disrupted critical genetic regulators of receptor function.

Knock-outs of *Odorant-coreceptor* (*Orco*) disrupt normal localization of ORs, destroying all *Or* function but not *Ir*s or *Gr*s (BENTON *et al.* 2006; LARSSON *et al.* 2004). *Orco* mutants showed no loss of OA:HA aversion (Figure 4.2b). We verified the lack of *Orco* effect on OA:HA aversion in two ways. First, to ensure that we successfully disrupted *Orco*, we also knocked down *Orco* expression by inducing programmed cell death in neurons expressing *Orco* using *reaper*, a positive regulator of apoptosis (WHITE *et al.* 1996). Second, to make sure that *Orco* mutants behaved appropriately against other known *Or*-mediated chemicals, we showed that *Orco* disruption exhibited reduced aversion to benzaldehyde (Figure B.1)(HALLEM *et al.* 2004).

D. melanogaster lacking functional *Or*s maintain aversion to OA:HA, but flies lacking antennae, the main olfactory organ, lose aversion to OA:HA. Thus, there must be some non-OR olfactory factor, such as an *Ir*, on the antennae that detects OA:HA. One *Ir* in particular, *Ir64a*, mediates detection of acidic food (AI *et al.* 2010). Disruptions in *Ir64a*, however, do not alter *D. melanogaster* aversion to OA:HA (EARLEY AND JONES submitted)(Chapter Three). Alternatively, antennae also express *Odorant binding proteins* (*Obps)*, which can influence the detection of odors (GALINDO and SMITH 2001). OBPs are thought to shuttle non-aqueous chemicals through aqueous sensory organ lymph to the appropriate chemoreceptor (GALINDO and SMITH 2001). Earlier work

showed that knockdown of antennally expressed *Obp56e,* caused a weak loss of aversion in *D. melanogaster* (DWORKIN and JONES 2009). The difference between the *Orco* knockouts and antennalectomy results could therefore be due to *Obps,* and perhaps *Irs*, expressed in the antennae.

To look at the role of taste (*Gr*s), we disrupted *pox-neuro* (*poxn*) expression, causing all gustatory receptor neurons to revert to mechanosensory neurons (AWASAKI and KIMURA 1997; BOLL and NOLL 2002). Flies with no *Gr* function lost aversion (Figure 4.2b). We then partially rescued *poxn* with a transgene (*full-1*) that restores all gustatory bristles except for those on the labellum. These flies cannot taste using their primary gustatory organ, but other taste bristles are functional. Flies lacking labellular *Grs* avoided OA:HA. Thus, it appears *D. melanogaster* specifically needs non-labellular GRs to detect Morinda compounds. This result must be treated with caution, however, as *poxn* also affects the central nervous system and may influence olfactory signal processing.

Gr22c is necessary for aversion

Obp57d/e, which is strongly expressed in the tarsi, is a known mediator of OA:HA behavior (MATSUO *et al.* 2007). This locus has two genes that share a common regulatory region. *D. sechellia's* ortholog (*Dsech/Obp57d/e*) harbors a 4bp insertion just upstream of the *Obp57d/e* transcription start site relative to other members of the *D. melanogaster* species complex, and this insertion appears to disrupt expression of *Dsech/Obp57d/e.* Knockouts of *Obp57d/e* in *D. melanogaster* females reduces aversion to HA, and replacement of *Dsech/Obp57d/e* in these flies also decreases aversion for

oviposition on HA. Given *Obp57d/e's* role in aversion, one might expect the OBP57d/e protein to be shuttling non-aqueous HA, OA, or both to a relevant chemosensory receptor such as a *Gr*. An obvious candidate *Gr* would have a clear genetic disruption in *D. sechellia* but appear to maintain function in *D. melanogaster*. McBride (MCBRIDE 2007), in a survey of pseudogenized chemosensory genes within the melanogaster species complex, found an unusually large number of *Gr* pseudogenes in *D. sechellia*. Included in this set was *Dsech/Gr22c*, whose ortholog in *D. melanogaster* is expressed in tarsi near the cells that express OBP57d/e (DUNIPACE *et al.* 2001; SCOTT *et al.* 2001).

If OBP57d/e shuttles OA and HA to GR22c, then knocking out *Gr22c* should halt OA:HA aversion at least as strongly as *Obp57de* knockouts. To this end, we screened the population genomics panel of *D. melanogaster* ("Raleigh lines" LANGLEY *et al.* 2012; MACKAY *et al.* 2012) for lines possessing a premature termination codon in *Gr22c*. Line *RAL-437* harbors a stop codon at roughly 5% into *Gr22c* open reading frame relative to the *D. melanogaster* genomic reference (assembly r5.9; GELBART *et al.* 1999). In comparison, *D.sech/Gr22c* has a stop codon roughly 50% into the open reading frame (Figure 4.3a)(assembly r1.3; GELBART *et al.* 1999). Compared to Raleigh lines with full length *Gr22c*, *RAL-437* lost aversion to OA:HA (Figure 4.3c). To account for some of the genetic background of *RAL-437* potentially driving this loss independently of *Gr22c* we crossed the *RAL-437 Gr22c* (*437/Gr22c*) locus into our wild type *D. melanogaster* lab strain. These flies also lost aversion (Figure 4.3c) and were no different from uncrossed *RAL-437*. To ensure that *RAL-437* loss of aversion is specific to OA:HA, we tested its behavior on a panel of alternate chemicals. *RAL437* responded as other *RAL* lines for all alternate chemicals (Figure B.2). As a further test of *Gr22c*, we knocked down *Gr22c*

expression via RNAi with a *Gr22c*-specific promoter (SCOTT *et al.* 2001) and observed reduced OA:HA aversion, although in this case loss appeared to be disproportionately driven by male behavior (Figure B.3).

Gr22c is required for *D. melanogaster* aversion, but is it the receptor for *Obp57de*? To confirm that *Gr22c* was expressed in tarsi near cells expressing *Obp57de*, we used a *Gr22c* promoter (SCOTT *et al.* 2001) to drive GFP expression in *D. melanogaster* and detected signal in distal most tarsi, as expected (Figure 4.3b)(no signal detected in labellar sensilla, conflicting with reported weak signal in SCOTT *et al.* 2001; WEISS *et al.* 2011). If OPB57de is shuttling OA, HA, or both to GR22c, then disruption in *Gr22c* but not *Obp57de* should cause loss of aversion to these compounds. On the other hand, functional *Gr22c* with a disrupted *Obp57de* might have a step-wise loss of aversion. We found that *Gr22c* disruption alone is enough to cause loss of OA:HA aversion, but *Obp57d/e* disruption is not (Figure B.4).

Gr22c affects oviposition behavior in *D. melanogaster* and *D. simulans*

Two issues complicate interpretation of the effect of *Gr22c* on aversion. First, our assay (Figure 4.1b) likely captures at least two separate behavioral phenotypes: settling behavior and ovipositional preference. In each test, flies were given the opportunity to sample each food source, but most flies settled on one food or the other within 90-120 minutes and remained near that food through the duration of the test (data not shown). However, where a fly spends its time may not be where it oviposits (HARADA *et al.* 2008; JOSEPH *et al.* 2009; MATSUO *et al.* 2007). Using an oviposition assay, we measured *D. melanogaster* OA oviposition behavior after MATSUO (2012) and

found flies lacking *Gr22c* exhibited reduced oviposition aversion to OA medium at low concentrations relative to intact controls (Figure 4.4a). Flies maintained aversion at higher concentrations, and this may have been caused by other factors (*eg,* antennal expressed *Irs* or *Obps*).

The second issue is that behavioral effects observed in *D. melanogaster* mutants may not capture what occurred in the ancestor of *D. sechellia*. *D. simulans* is one of *D. sechellia*'s closest relatives and thought to be most similar to *D. sechellia's* ancestor (R'KHA *et al*. 1991). Thus, *D. simulans* is less genetically divergent from *D. sechellia* than *D. melanogaster*, and *D. simulans* is more likely to share genetic mechanisms governing Morinda behavior with *D. sechellia*. We used a standard backcross between these two species to test the effect of the *Gr22c* region on aversion (Figure 4.4b). Backcross progeny were scored for oviposition behavior toward OA media, and hybrids with the functional *D. simulans Gr22c* alleles were significantly more aversive to OA than siblings with the pseudogenized *D. sechellia Gr22c* allele (Figure 4.4c). While we cannot exclude another linked locus in this experiment, the consistency of these results with the *D. melanogaster* data strongly suggest that the *Gr22c* pseudogene in *D. sechellia* is important for its host preference behavior.

Gr22c does not affect tolerance of Morinda

In addition to affecting behavior**,** OA and HA are also the primary toxins in Morinda. One prediction of host specialization evolution on a toxic host is that genetic loci controlling preference for the host and performance on the host should be linked, either physically or epistatically to facilitate rapid evolution (JAENIKE 1990). Complete

linkage, where one locus controls both preference and performance, would ensure that adults did not oviposit upon food that was toxic to their offspring (either as larvae or as adults). To test whether *Gr22c* influenced tolerance of Morinda, we assayed time to adult knock-down (a prelude to death) in *Gr22c* knockout and control lines. *D. melanogaster* lacking *Gr22c* performed just as poorly as *D. melanogaster* with intact *Gr22c*, whereas *D. sechellia* remained unaffected throughout the test (Figure B.5). Thus, altering preference for Morinda compounds is not sufficient to alter tolerance (see Chapter Five).

Pseudogenization of *Gr22c* likely a key innovation during evolution of superspecialism in *D. sechellia*

Avoidance of the Morinda toxins, OA and HA, is mediated in part by the gustatory receptor *Gr22c*, which is located on the foreleg tarsi. Loss of a functional copy of this allele in the ancestor of *D. sechellia* was likely a key evolutionary innovation during *D. sechellia*'s shift to using Morinda as its host. However, it was not the only innovation. Neuro-physiology studies show *Or85b* and *Or22a,* expressed in ORNs within ab3 type sensilla, respond to OA and HA, respectively (HALLEM and CARLSON 2006). *D. sechellia* have an over abundance of ab3 sensilla relative to *D. melanogaster*, correlated with hypersensitivity to certain Morinda compounds (DEKKER *et al.* 2006; STENSMYR *et al.* 2003). *D. sechellia* preference, however, cannot be explained solely by ab3 sensilla. For instance, *D. simulans*, which is strongly aversive to Morinda compounds, also has an over abundance of some ab3 sensilla (KOPP *et al.* 2008; STENSMYR *et al.* 2003). Moreover, our *Orco* knockout, which disrupted both *Or85b* and *Or22a* function, did not reduce *D. melanogaster* aversion.

While ORs may yet be critical for elements of *D. sechellia* preference (such as attraction to methyl hexanoate; DEKKER *et al.* 2006), *D. melanogaster* does not need them to avoid OA:HA. Also expressed in the antennae are a subset of *Obps*, among them *Obp56e* (GALINDO and SMITH 2001), and knockdown of *Obp56e* causes loss of aversion in *D. melanogaster* (DWORKIN and JONES 2009); however, the effect of this loss is weak and explains only ~15% of *D. melanogaster* aversion. Thus, while *Obp56e* may participate in aversion, there must be some other factor that explains the necessity of antennae. *Irs* remain the best candidates.

D. melanogaster need tarsi to avoid OA:HA. Modification of *Obp57d/e*, which is expressed in the tarsi, alters female positional behavior to HA and oviposition behavior to both HA and OA (MATSUO *et al.* 2007). *Obp57d/e* expression appears to be proximal to that of *Gr22c* (DUNIPACE *et al.* 2001; SCOTT *et al.* 2001), allowing for the possibility that *Obp57d/e* acts as a chaperone for *Gr22c*. In contrast, *D. sechellia* did not need tarsi to prefer OA:HA, matching the down-regulation of *Dsech/Obp57de* in tarsi (MATSUO *et al.* 2007).

Marker assisted mapping studies on the genetic basis of HA aversion also match our results from *Gr22c*. A strong factor driving *D. simulans* behavior to HA was found on chromosome 2 (HIGA and FUYAMA 1993) and specifically on the arm *2*L (JONES 2005), where *Gr22c* resides. Recent genome-wide surveys of *D. simulans* aversion to OA:HA uncovered multiple loci spread across all autosome arms, however, none of these contained *Gr22c* (EARLEY and JONES 2011)(Chapter Two, Chapter Three). A couple factors explain the apparent conflict between this map and the current study. First, the earlier studies only recovered ~68% of the parental phenotype. Aversion is known to be

additive in *D. simulans*, thus the method of mapping – hybrid introgression with backcrossing – could have easily missed even loci of major effect. Second, if *Gr22c* is near a hybrid incompatibility locus, *Gr22c* would not have introgressed. Supporting this last possibility, FANG *et al*. (2012) found large chromosomal regions that appear to drive competitive exclusion, a mechanism of ecological speciation, between *D. simulans* and *D. sechellia*, one of which encompasses *Gr22c*. Our oviposition results with *D. simulans/D. sechellia* F_2 hybrids (Figure 4.4c) control for such interspecific incompatibilities since only one generation of backcrossing precludes competitive exclusion from driving one species' loci to fixation. A similar result was seen in Chapter Three.

First step to Specialism: support for the two-step model of specialized behavior

Preference for Morinda could have evolved in one step – simultaneous loss of aversion and gain of preference. This could occur, for example, via misexpression of chemoreceptors in alternate sensory neurons which transmit signals to glomeruli that process these signals into novel preference instead of aversion. Precedence exists for this model in lab studies, both in mice (MUELLER *et al.* 2005) and nematodes (TROEMEL *et al.* 1997).

A two-step model of the evolution of preference towards a plant host posits that before specialist preference evolves, ancestral aversion is first lost, and this intermediate stage confers higher fitness than aversion (MATSUO *et al.* 2007; R' KHA *et al.* 1997). Potential hosts that are otherwise viable food sources are often aversive due to the production of toxic secondary plant metabolites created as a defense against insect

predators (BERNAYS and CHAPMAN 1994). If the potential host is only toxic for a short time, as is the case with Morinda (LEGAL *et al.* 1994), then opportunities exist for non-aversive consumers to benefit from a rich resource without exposing themselves to a toxin. This benefit can be enhanced if non-aversion also promotes escape from resource competitors and predators. Both of these conditions apply to extant *D. sechellia* and its ecological neighbors: generalist species like *D. simulans* and *D. mauritiana* are often found on over-ripe Morinda (LOUIS and DAVID 1986; R' KHA *et al.* 1997; R'KHA *et al.* 1991) and one of Drosophila's main parasitoid wasp predators, *Leptopilina boulardi,* avoids OA (T. Schlenke, pers. comm.). It is possible that escape from competitors and predators facilitated the original transition from aversion to indifference in *D. sechellia*'s ancestor.

One prediction of the two-step model is that specialist preference and generalist aversion should work via separate genetic mechanisms. For example, if aversion is governed by gustation, as appears to be the case in *D. melanogaster* and *D. simulans*, then preference should be governed by either a separate gustation factor or by olfaction. That is, breaking an avoidance detection system makes it harder to use the same system for attraction. Early experiments in the *D. simulans* species complex do suggest that preference and aversion work through different modes of action (AMLOU *et al.* 1998), and *D. sechellia* indeed appears to primarily use olfaction to prefer Morinda (DEKKER *et al.* 2006; DWORKIN and JONES 2009; IBBA *et al.* 2010; R'KHA *et al.* 1991; STENSMYR *et al.* 2003).

Pseudogenes are a route towards specialism

We believe that a pseudogenization of *Gr22c* in the ancestor of *D. sechellia* caused loss of aversion to OA:HA. *D. sechellia* harbors an unusually large number of pseudogenized chemoreceptors (*OR*s: 10%, *GR*s: 17%) compared to *D. simulans* (*OR*s: 0%, *GR*s: 3%) and *D. melanogaster* (*OR*s: 2%, *GR*s: 8%), even when controlling for *D. sechellia*'s genome-wide pseudogene frequency (MCBRIDE 2007). It may not be surprising that specialists often have many pseudogenes (BURKE and MORAN 2011; MCBRIDE 2007; VAN DE GUCHTE *et al.* 2006). Ecological specialism equates to genetic specialism. This could occur through an adaptive process, or superspecialism can promote "genetic decay" – that is, disused genes accumulate neutral disruptive mutations. Specialists often have small population sizes and thus genetic drift will drive these mutations to fixation faster (GARDINER *et al.* 2008).

While genetic drift promotes gene loss after specialism, it may also be possible that gene loss itself promotes specialization in the first place. Olson (1999) proposed the "less is more" hypothesis of adaptive genetic evolution, which predicts that adaptive gene loss should occur more frequently than adaptive gain. Mutations that break a gene are much more likely to occur than mutations that shift a gene to some new function—as evinced by the segregating *Gr22c* pseudogene in *D. melanogaster*. Adaptation relies on mutations as a resource. Given their overwhelming presence in the pool of mutational events, if even a small percent of pseudogenes are adaptive, they would outnumber genetic gain of function events, which although highly beneficial, occur rarely. Some evidence for this hypothesis exists in humans (STEDMAN *et al.* 2004; VARKI 2001; WANG *et al.* 2006), although see (PERRY *et al.* 2005), and bacteria (MORRIS *et al.* 2012). In the case of *Gr22c* in Drosophila, a nonsense mutation leading to a premature stop codon

could have been beneficial to ancestral *D. sechellia,* removing unnecessary aversion behavior and promoting use of a novel resource. In theory, then, the preponderance of pseudogenes in *D. sechellia* could have been created via a combination of genetic drift and adaptive evolution, and not simply as a consequence of specialization itself. That is, some pseudogenes were a cause and not a consequence of specialization. If true, this would provide even stronger barriers against evolutionary reversions to generalism, promoting the "dead end" model of specialism, since reversions in pseudogenes that rescue function would themselves be maladaptive.

MATERIALS AND METHODS

Fly stocks

All flies were reared in laboratory conditions at 25C, 50% humidity, ~12hr:12hr day/night cycle. Control lines were w^{1118} (*D. mel*), *SynA* (*D. sech*), and *c167.4* (*D. sim*). *poxn* stocks, *Gr22c-GAL4,* and *Obp57de* stocks were generous gifts from R. Joseph, K. Scott, and T. Matsuo, respectively. All other lines were from Bloomington stock center, *Drosophila* species stock center (San Diego), or custom made.

Behavior test

Milk bottle assay. Tests conducted as in (EARLEY and JONES 2011). Briefly, between 30-60 flies, aged 2-10d post-eclosion, were introduced into a behavior arena without anesthesia. Within the arena (2L glass beaker, Fisher, Pittsburgh, Pa.) were two milk bottles containing media - 22mL diH_2O, 4.1g instant fly media (Carolina Biological) – with or without acid inoculate (45ul OA, 15ul HA, Sigma-Aldrich). Before mixing with

solid food, organic acids were ejected into the water with force to create an even dispersion of acid droplets atop the surface of the media. Flies were allowed to chose between bottles for 20-24hrs, after which bottles were stopped, and flies within each were counted. A response index was calculated: RI= logit[(#flies test)/(#flies test + #flies control)]. One replicate test was a population of 30-60 flies. At least three replicates were obtained for each line.

A similar set up was used for comparing response between fruits. ~3g of wet, ripe fruit (test = Morinda, control = banana) was placed at the bottom of separate milk bottles. However, in this case, it is impossible to distinguish between strong preference for the control fruit or strong aversion for Morinda.

Oviposition behavior and Morinda tolerance

D. simulans/D. sechellia hybrid oviposition was assayed as in Jones (2005). *D. melanogaster* oviposition behavior was assayed as found in Matsuo (MATSUO 2012). Tolerance was assayed on ripe Morinda after Jones (1998). Roughly a pea sized mass of ripe, wet fruit (no seeds) was placed in the center of a 60mm pyrex petri dish (Corning), and individual flies were then placed inside without anesthesia. Tests were conducted at 25C, 60% humidity, and knockdown was scored starting at 30 min. then every 10 min. until 60 min.. Unaffected flies were scored as 70 min.

Cloning and transgenics

Transgenic *D. melanogaster* were made via phiC31-mediated integration (GROTH *et al.* 2004). Coding regions of *Gr22c* from *D. melanogaster, D. simulans,* and *D.*

sechellia were synthesized with flanking *AgeI* and *NheI* sites and cloned separately into the pUASg.attB vector, and insertion was confirmed with restriction digest banding pattern (performed by Genewiz, South Plainfield, NJ). Midi-prepped purified plasmid was injected into an attP-containing fly line and balanced with TM3 (performed by BestGene, Chino Hills, CA). Transgenic flies were the crossed into a *Gr22c*$^{-}$ background and subsequently crossed to a *GAL4* driver line. *GAL4/UAS* flies were confirmed to be in *Gr22c*$^{-}$ homozygous background via AFLP and RFLP. DNA was collected using squish buffer on individual flies (GLOOR *et al.* 1993).

ACKNOWLEDGEMENTS

We thank Betty Wanjiru and Sophia Shih for assistance. We also thank Steve Crews, Joel Kingsolver, Pelin Volkan and Chris Willett for comments. We appreciate the generous gift of Drosophila stocks and technical/editorial advice from Pelin Volkan, Jeff Sekelsky, John Carlson, Kristin Scott, Dan McKay, Joe Pearson, Steve Crews, Joel Kingsolver, Ryan Joseph, Takashi Matsuo, Artur Romanchuk, Kayla Peck, Tony Perdue, and Chris Willett. This work was supported by the National Science Foundation DEB-0212686 (CDJ), NSF-Graduate Research Fellowship Program (EJE), and by NIDDK grant P30DK056350 to the UNC Nutrition Obesity Research Center.

REFERENCES

AI, M., S. MIN, Y. GROSJEAN, C. LEBLANC, R. BELL *et al.*, 2010 Acid sensing by the Drosophila olfactory system. Nature **468:** 691-695.

AMLOU, M., B. MORETEAU and J. R. DAVID, 1998 Genetic analysis of Drosophila sechellia specialization: oviposition behavior toward the major aliphatic acids of its host plant. Behav Genet **28:** 455-464.

AMLOU, M., E. PLA, B. MORETEAU and J. R. DAVID, 1997 Genetic analysis by interspecific crosses of the tolerance of *Drosophila sechellia* to major aliphatic acids of its host plant. Genet. Sel. Evol. **29:** 511-522.

AWASAKI, T., and K. KIMURA, 1997 pox-neuro is required for development of chemosensory bristles in Drosophila. J Neurobiol **32:** 707-721.

BATEMAN, J. R., A. M. LEE and C. T. WU, 2006 Site-specific transformation of Drosophila via phiC31 integrase-mediated cassette exchange. Genetics **173:** 769-777.

BENTON, R., S. SACHSE, S. W. MICHNICK and L. B. VOSSHALL, 2006 Atypical membrane topology and heteromeric function of Drosophila odorant receptors in vivo. PLoS Biol **4:** e20.

BENTON, R., K. S. VANNICE, C. GOMEZ-DIAZ and L. B. VOSSHALL, 2009 Variant ionotropic glutamate receptors as chemosensory receptors in Drosophila. Cell **136:** 149-162.

BERNAYS, E. A., and R. F. CHAPMAN, 1994 *Host-plant selection by phytophagous insects*. Chapman and Hall, New York.

BOLL, W., and M. NOLL, 2002 The Drosophila Pox neuro gene: control of male courtship behavior and fertility as revealed by a complete dissection of all enhancers. Development **129:** 5667-5681.

BURKE, G. R., and N. A. MORAN, 2011 Massive genomic decay in Serratia symbiotica, a recently evolved symbiont of aphids. Genome Biol Evol **3:** 195-208.

DEKKER, T., I. IBBA, K. P. SIJU, M. C. STENSMYR and B. S. HANSSON, 2006 Olfactory shifts parallel superspecialism for toxic fruit in Drosophila melanogaster sibling, D. sechellia. Curr Biol **16:** 101-109.

DUFFY, J. B., 2002 GAL4 system in Drosophila: a fly geneticist's Swiss army knife. Genesis **34:** 1-15.

DUNIPACE, L., S. MEISTER, C. MCNEALY and H. AMREIN, 2001 Spatially restricted expression of candidate taste receptors in the Drosophila gustatory system. Curr Biol **11:** 822-835.

DWORKIN, I., and C. D. JONES, 2009 Genetic changes accompanying the evolution of host specialization in Drosophila sechellia. Genetics **181:** 721-736.

EARLEY, E. J., and C. D. JONES, 2011 Next-generation mapping of complex traits with phenotype-based selection and introgression. Genetics **189:** 1203-1209.

FANG, S., R. YUKILEVICH, Y. CHEN, D. A. TURISSINI, K. ZENG *et al.*, 2012 Incompatibility and competitive exclusion of genomic segments between sibling Drosophila species. PLoS Genet **8:** e1002795.

FARINE, J. P., L. LEGAL, B. MORETEAU and J. L. LEQUERE, 1996 Volatile components of ripe fruits of Morinda citrifolia and their effects on Drosophila. Phytochemistry **41:** 433-438.

GALINDO, K., and D. P. SMITH, 2001 A large family of divergent Drosophila odorant-binding proteins expressed in gustatory and olfactory sensilla. Genetics **159:** 1059-1072.

GARDINER, A., D. BARKER, R. K. BUTLIN, W. C. JORDAN and M. G. RITCHIE, 2008 Drosophila chemoreceptor gene evolution: selection, specialization and genome size. Mol Ecol **17:** 1648-1657.

GELBART, W. M., M. CROSBY, B. MATTHEWS, J. CHILLEMI, S. R. TWOMBLY *et al.*, 1999 The FlyBase database of the Drosophila genome projects and community literature. Nucleic Acids Research **27:** 85-88.

GLOOR, G. B., C. R. PRESTON, D. M. JOHNSON-SCHLITZ, N. A. NASSIF, R. W. PHILLIS *et al.*, 1993 Type I repressors of P element mobility. Genetics **135:** 81-95.

GROTH, A. C., M. FISH, R. NUSSE and M. P. CALOS, 2004 Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. Genetics **166:** 1775-1782.

HALLEM, E. A., and J. R. CARLSON, 2006 Coding of odors by a receptor repertoire. Cell **125:** 143-160.

HALLEM, E. A., M. G. HO and J. R. CARLSON, 2004 The molecular basis of odor coding in the Drosophila antenna. Cell **117:** 965-979.

HARADA, E., D. HABA, T. AIGAKI and T. MATSUO, 2008 Behavioral analyses of mutants for two odorant-binding protein genes, Obp57d and Obp57e, in Drosophila melanogaster. Genes Genet Syst **83:** 257-264.

HIGA, I., and Y. FUYAMA, 1993 Genetics of food preference in Drosophila sechellia. I. Responses to food attractants. Genetica **88:** 129-136.

IBBA, I., A. M. ANGIOY, B. S. HANSSON and T. DEKKER, 2010 Macroglomeruli for fruit odors change blend preference in Drosophila. Naturwissenschaften **97:** 1059-1066.

JAENIKE, J., 1990 Host Specialization in Phytophagous Insects. Annual Review of Ecology and Systematics **21:** 243-273.

JONES, C. D., 1998 The genetic basis of Drosophila sechellia's resistance to a host plant toxin. Genetics **149:** 1899-1908.

JONES, C. D., 2005 The genetics of adaptation in Drosophila sechellia. Genetica **123:** 137-145.

JOSEPH, R. M., A. V. DEVINENI, I. F. KING and U. HEBERLEIN, 2009 Oviposition preference for and positional avoidance of acetic acid provide a model for competing behavioral drives in Drosophila. Proc Natl Acad Sci U S A **106:** 11352-11357.

KOPP, A., O. BARMINA, A. M. HAMILTON, L. HIGGINS, L. M. MCINTYRE *et al.*, 2008 Evolution of gene expression in the Drosophila olfactory system. Mol Biol Evol **25:** 1081-1092.

LANGLEY, C. H., K. STEVENS, C. CARDENO, Y. C. LEE, D. R. SCHRIDER *et al.*, 2012 Genomic variation in natural populations of Drosophila melanogaster. Genetics **192:** 533-598.

LARSSON, M. C., A. I. DOMINGOS, W. D. JONES, M. E. CHIAPPE, H. AMREIN *et al.*, 2004 Or83b encodes a broadly expressed odorant receptor essential for Drosophila olfaction. Neuron **43:** 703-714.

LEGAL, L., B. CHAPPE and J. M. JALLON, 1994 Molecular Basis of *Morinda citrifolia* (L.): Toxicity on Drosophila. Journal of Chemical Ecology **20:** 1931-1943.

LEGAL, L., J. R. DAVID and J. M. JALLON, 1992 Toxicity and attraction effects produced by *Morinda citrifolia* fruits on the *Drosophila melanogaster* complex of species. Chemoecology **3:** 125-129.

LOUIS, J., and J. R. DAVID, 1986 Ecological Specialization in the *Drosophila melanogaster* species subgroup: A case study of *D. sechellia*. Acta Oecologica, Oecological Generalis**:** 215-219.

MACKAY, T. F., S. RICHARDS, E. A. STONE, A. BARBADILLA, J. F. AYROLES *et al.*, 2012 The Drosophila melanogaster Genetic Reference Panel. Nature **482:** 173-178.

MATSUO, T., 2012 Contribution of olfactory and gustatory sensations of octanoic acid in the oviposition behavior of *Drosophila melanogaster* (Diptera: Drosophilidae). Appl Entomol Zool **47:** 137-142.

MATSUO, T., S. SUGAYA, J. YASUKAWA, T. AIGAKI and Y. FUYAMA, 2007 Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in Drosophila sechellia. PLoS Biol **5:** e118.

MCBRIDE, C. S., 2007 Rapid evolution of smell and taste receptor genes during host specialization in Drosophila sechellia. Proc Natl Acad Sci U S A **104:** 4996-5001.

MONTELL, C., 2009 A taste of the Drosophila gustatory receptors. Curr Opin Neurobiol **19:** 345-353.

MORRIS, J. J., R. E. LENSKI and E. R. ZINSER, 2012 The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. MBio **3**.

OLSON, M. V., 1999 When less is more: gene loss as an engine of evolutionary change. Am J Hum Genet **64:** 18-23.

PERRY, G. H., B. C. VERRELLI and A. C. STONE, 2005 Comparative analyses reveal a complex history of molecular evolution for human MYH16. Mol Biol Evol **22:** 379-382.

PINO, J. A., E. MARQUEZ, C. E. QUIJANO and D. CASTRO, 2010 Volatile compounds in noni (Morinda citrifolia L.) at two ripening stages. Ciencia E Tecnologia De Alimentos **30:** 183-187.

R' KHA, S., B. MORETEAU, J. A. COYNE and J. R. DAVID, 1997 Evolution of a lesser fitness trait: egg production in the specialist Drosophila sechellia. Genet Res **69:** 17-23.

R'KHA, S., P. CAPY and J. R. DAVID, 1991 Host-plant specialization in the Drosophila melanogaster species complex: a physiological, behavioral, and genetical analysis. Proc Natl Acad Sci U S A **88:** 1835-1839.

SCOTT, K., R. BRADY, JR., A. CRAVCHIK, P. MOROZOV, A. RZHETSKY *et al.*, 2001 A chemosensory gene family encoding candidate gustatory and olfactory receptors in Drosophila. Cell **104:** 661-673.

STEDMAN, H. H., B. W. KOZYAK, A. NELSON, D. M. THESIER, L. T. SU *et al.*, 2004 Myosin gene mutation correlates with anatomical changes in the human lineage. Nature **428:** 415-418.

STENSMYR, M. C., T. DEKKER and B. S. HANSSON, 2003 Evolution of the olfactory code in the Drosophila melanogaster subgroup. Proc Biol Sci **270:** 2333-2340.

VAN DE GUCHTE, M., S. PENAUD, C. GRIMALDI, V. BARBE, K. BRYSON *et al.*, 2006 The complete genome sequence of Lactobacillus bulgaricus reveals extensive and ongoing reductive evolution. Proc Natl Acad Sci U S A **103:** 9274-9279.

VARKI, A., 2001 Loss of N-glycolylneuraminic acid in humans: Mechanisms, consequences, and implications for hominid evolution. Am J Phys Anthropol **Suppl 33:** 54-69.

VOSSHALL, L. B., and R. F. STOCKER, 2007 Molecular architecture of smell and taste in Drosophila. Annu Rev Neurosci **30:** 505-533.

WANG, X., W. E. GRUS and J. ZHANG, 2006 Gene losses during human origins. PLoS Biol **4:** e52.

WEISS, L. A., A. DAHANUKAR, J. Y. KWON, D. BANERJEE and J. R. CARLSON, 2011 The molecular and cellular basis of bitter taste in Drosophila. Neuron **69:** 258-272.

WHITE, K., E. TAHAOGLU and H. STELLER, 1996 Cell killing by the Drosophila gene reaper. Science **271:** 805-807.

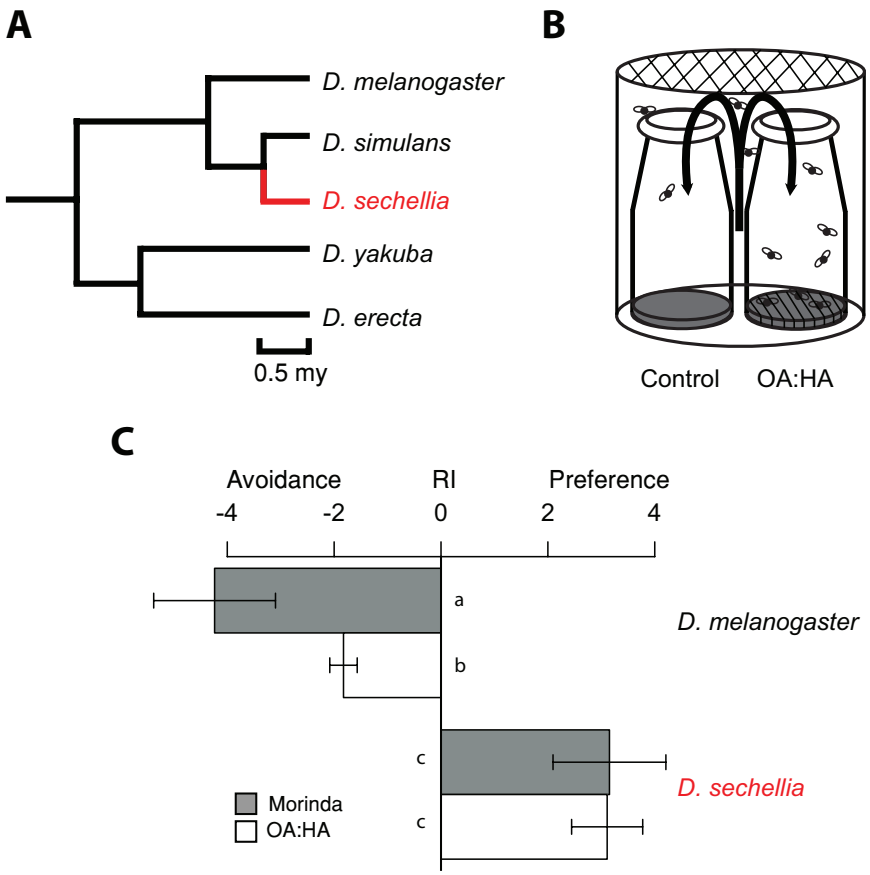

Figure 4.1 Wild-type behavior of *Drosophila* towards Morinda and its compounds. (A) Preference for ripe Morinda and its compounds occurred along the *D. sechellia* lineage (red), and all other *Drosophilids* avoid it (black). (B) Wild-type response index (avoidance or preference) measured on ripe fruit (Banana vs. Morinda) or pure chemicals (chemical free media vs. OA:HA media). Bars represent mean ± SEM. Bars with different lower case letters were significantly different from each other ($P < 0.05$, unpaired Student's t-test).

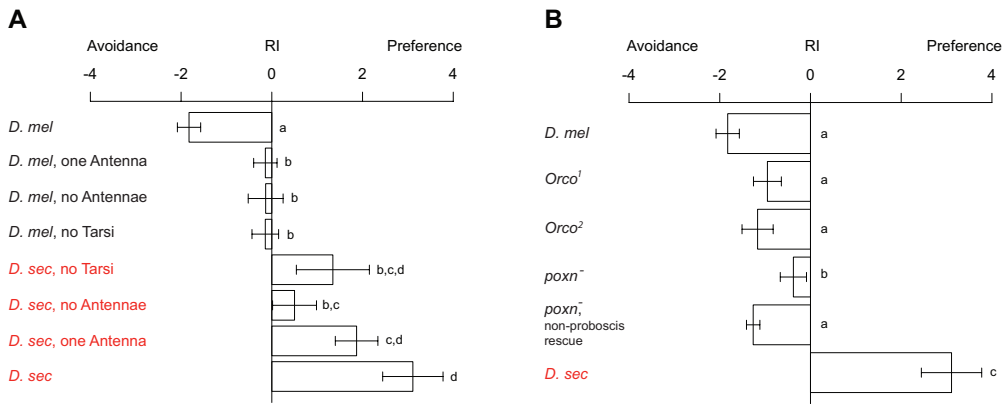

Figure 4.2 Tarsal taste necessary for OA:HA aversion. (A) Response Index (RI) of antennae- and tarsal- ectomies on wild-type *D. melanogaster* (*D.mel*, black) and *D. sechellia* (*D.sec*, red). (B) Tests of genetic disruptions on smell and taste receptors in *D. melanogaster*. *Orco = Odorant coreceptor*, *poxn = pox-neuro*. Bars with different letters were significantly different from each other ($P < 0.05$, unpaired Student's t test).

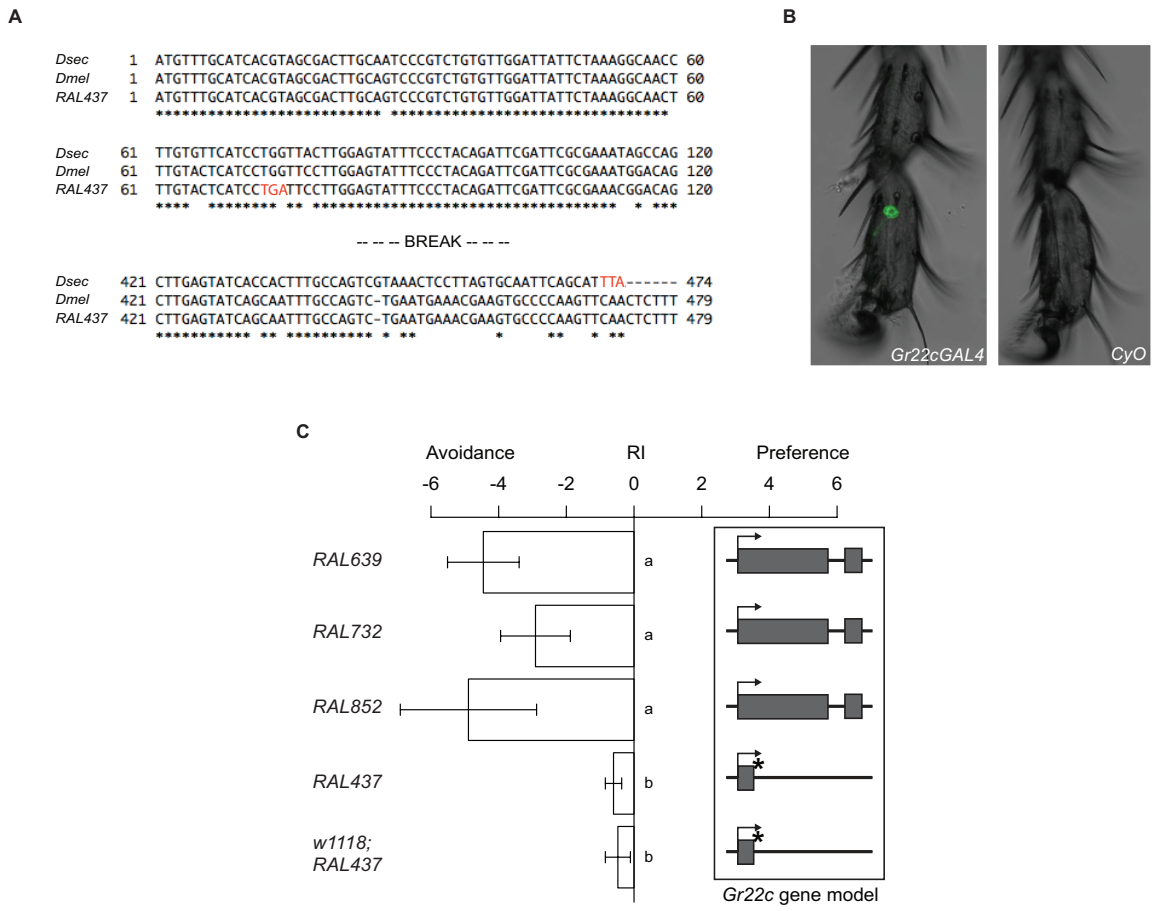

Figure 4.3 *Gr22c* necessary for OA:HA aversion. (A) *RAL-437* and *D. sechellia/Gr22c* harbor premature stop codons ~5% and ~50%, respectively, into *Gr22c* open reading frame relative to control *D. melanogaster.* Red codons denote stop codons (B) Left, GFP expression pattern of *Gr22c* (*Gr22c-GAL4/40x-UAS-GFP*) is near that of *Obp57d/e* expression (not shown, MATSUO *et al.* 2007). Right, sibling controls without *Gr22c-GAL4* do not express GFP (*40x-UAS-GFP/CyO*). (C) *Raleigh* lines strongly avoid OA:HA, except *RAL437*, which is significantly less aversive, and this correlates with *Gr22c* pseudogene state (gene models on right). After crossing to *D. melanogaster* control line from previous experiments, this loss is maintained. Note different horizontal scale from previous figures. ($P < 0.05$, unpaired Student's t test).

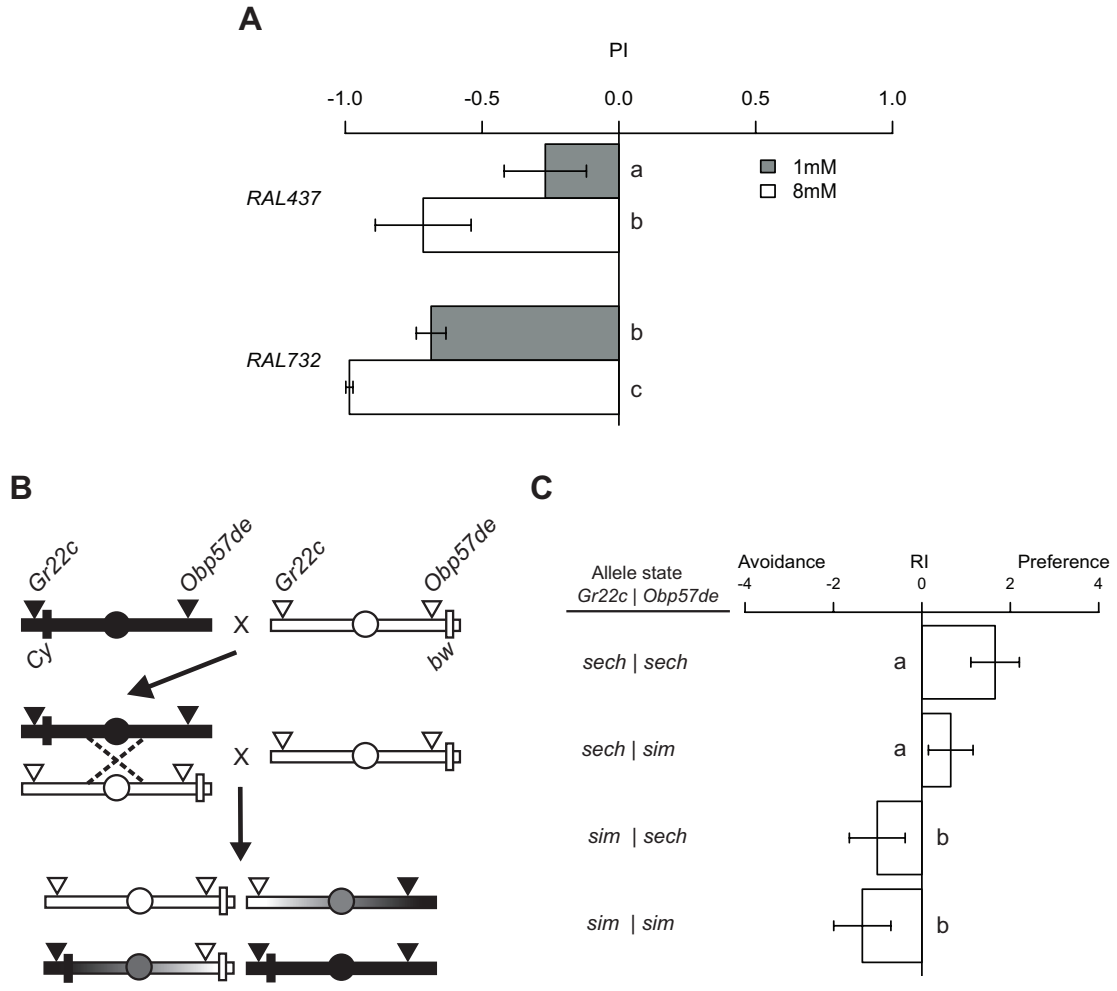

Figure 4.4 *Gr22c* locus confers oviposition avoidance in *D. melanogaster* and hybrids. (A) *RAL437* avoids ovipositing on low concentration OA media (1mM) relative to *RAL732*, which has intact *Gr22c*. Because this assay is significantly different from other assays, we calculated a preference index, *PI* = (#eggs OA)/(#eggs OA + #eggs Ctrl) ($P < 0.05$, Kruskal-Wallis rank sum test). (B) We crossed visibly marked *D. simulans* (*Cy,* dominant, black vertical bars) and *D. sechellia* (*bw,* recessive, white vertical bars), then backcrossed the progeny to *D. sechellia*. *Cy* in *D. simulans* is close to *Gr22c* (~3.5cM), and *bw* in *D. sechellia* is close to *Obp57d/e* (~8cM). Inverse triangles denote gene positions for *Gr22c* (left of centromere) and *Obp57d/e* (right of centromere). All four combinations of alleles were observed in the F_2 generation. (C) F_2 progeny were scored for oviposition behavior toward OA media, and hybrids with *D. simulans Gr22c* were significantly more aversive to OA than siblings with *D. sechellia Gr22c* ($P < 0.05$, unpaired Student's t test).

V. LOCUS IN *DROSOPHILA SECHELLIA* CONTRIBUTING TO TOLRANCE OF A HOST PLANT TOXIN DOES NOT AFFECT HOST PREFERENCE

Authors: Eric A. Hungate, Eric J. Earley, Ian A. Boussy, David A. Turissini, Chau-Ti Ting, Jennifer R. Moran, Mao-Lien Wu, Chung-I Wu, Corbin D. Jones

Author's contribution: I performed all behavioral and Morinda tolerance assays on all flies. This included OA:HA chemotaxis assays on F_1 and F_2 populations of recombinant hybrids, tolerance to ripe Morinda fruit for all flies, and all analyses measuring correlation between tolerance and preference (linear modeling, correlation coefficients). I performed all statistical tests, generated figures, and interpreted results for all behavior and Morinda tolerance experiments. I assembled *D. mauritiana* reads against *D. simulans* candidate tolerance region, generated variant calls, and constructed *Obp83* paralog phylogenies. Along with C. Jones, I recommended changing the motivation of the entire project to testing correlation of tolerance and performance. I provided major editorial revisions to all drafts of the manuscript and documented historical examples both supporting and rejecting the preference/tolerance linkage hypothesis. I contributed to revisions in response to reviewer comments.

ABSTRACT

Many insects feed on only one or a few types of host. These host specialists often evolve a preference for chemical cues emanating from their host and develop

mechanisms for circumventing their host's defenses. Adaptations like these are central to evolutionary biology, yet our understanding of their genetics remains incomplete. *Drosophila sechellia,* an emerging model for the genetics of host specialization, is an island endemic that has adapted to chemical toxins present in the fruit of its host plant, *Morinda citrifolia*. Its sibling species, *D. simulans*, and many other *Drosophila* species do not tolerate these toxins and avoid the fruit. Earlier work showed that a region with a strong effect on tolerance to the major toxin, octanoic acid, was on chromosome arm *3*R. Using a novel assay we narrowed this region to a small span near the centromere containing 18 genes, including three odorant binding proteins. It has been hypothesized that the evolution of host specialization is facilitated by genetic linkage between alleles contributing to host preference and alleles contributing to host usage, such as tolerance to secondary compounds. We tested this hypothesis by measuring the effect of this tolerance locus on host preference behavior. Our data were inconsistent with the linkage hypothesis as flies bearing this tolerance region showed no increase in preference for media containing *M. citrifolia* toxins, which *D. sechellia* prefers. Thus, in contrast to some models for host preference, preference and tolerance are not tightly linked at this locus and increased tolerance per se is not sufficient to change preference. Our data are consistent with the previously proposed model that the evolution of *D. sechellia* as a *M. citrifolia* specialist occurred through a step-wise loss of aversion and gain of tolerance to *M. citrifolia's* toxins.

INTRODUCTION

Half of all insects interact with plants (GRIMALDI and ENGEL 2005). However, most phytophagous insects use only a few plant genera for food, mating, and oviposition

(BERNAYS and CHAPMAN 1994). Changes in host use are thought to result in both new species and new adaptations (EHRLICH and RAVEN 1964; JANZ 2011). For example, the evolution of a new host specialization may have contributed to the formation of new species in pea aphids (*Acyrthosiphon*) among others (VIA 2001; MATSUBAYASHI *et al.* 2010). Adapting to the new host can drive genetic and phenotypic change that is critical for isolating the nascent species. In some cases specialization has a price: increased performance on the new host correlates with reduced performance on other hosts (FUTUYMA and MORENO 1988; JAENIKE 1990; FRY *et al.* 1996; SCHEIRS *et al.* 2005; VIA and HAWTHORNE 2005). This scenario poses a new challenge for the nascent specialist, as it must keep alleles for finding and selecting the appropriate host ("preference") along with those for utilizing that host ("performance"; e.g. physiologically adapting to that host's secondary compounds or nutritional content; JAENIKE 1990; JANZ 2011). Theory suggests that a genetic correlation between the preference and performance alleles, such as caused by pleiotropy or genetic linkage, can overcome this problem and facilitate the switch to a new host (LANDE 1979; JAENIKE 1990; FRY *et al.* 1996; JANZ 2011).

Until recently, evidence for this "genetic linkage" hypothesis in phytophagous insects has been mixed. Early genetic data in *Drosophila* by JAENIKE suggested that oviposition preference and "settling" behavior are unlinked in *D. tripunctata* (1986; JAENIKE 1987; JAENIKE 1989), while TAYLOR AND CHONDRA (1983) found linkage between preference and performance in *D. pseudoobscura*. No linkage was found in other herbivorous species, such as *Callosobruchus maculatus* (southern cowpea weevil; WASSERMAN AND FUTUYMA 1981), *Colias philodice* (butterfly; TABASHNIK 1986), *Papilionidae* (swallowtail butterflies; THOMPSON 1988; THOMPSON *et al.* 1990),

Chrysomelidae (leaf-feeding beetles; KEESE 1996), *Nilaparvata lugens* (brown planthopper; SEZER and BUTLIN 1998a; SEZER and BUTLIN 1998b), and *Oreina elongata* (leaf beetle; BALLABENI AND RAHIER 2000). However, more recent QTL mapping data for aphids (HAWTHORNE and VIA 2001; CAILLAUD and VIA 2012; SAUGE *et al.* 2012), and other genetic association studies in *Euphydryas editha* (Edith's checkerspot butterfly; NG 1988; SINGER *et al.* 1988a), *Liriomyza sativae* (leafminer fly; VIA 1986), *Phyllotreta nemorum* (flea beetle; NIELSEN 1996), and *Papilio glaucus* (eastern tiger swallowtail butterfly; BOSSART 2003), suggest that some preference and performance alleles can be genetically linked.

A major concern with most genetic studies to date is their low resolution. QTL and marker-association studies produce candidate regions with large confidence intervals, increasing the chance that preference and performance alleles will overlap. Other studies simply infer genetic linkage due to the apparent heritability of host preference to well-performing offspring (*e.g.* SINGER *et al.* 1988b). Most finer resolution genetic studies of adaptive host specialization have focused on host preference or avoidance. Few studies have focused on the genetics of tolerance of a specific compound in the host plant because few species with obvious host adaptations have the requisite genetic tool kit needed to study these traits (except studies relating to domesticated plants where selection pressures are often much different than in natural populations).

Drosophila sechellia's specialization on *Morinda citrifolia* is a model system wherein the adaptation is obvious and a genetic tool kit is available. *D. sechellia* is endemic to the islands of the Seychelles (TSACAS and BACHLI 1981), specializes on the fruit produced by *M. citrifolia* (LOUIS and DAVID 1986), and is closely related to *D.*

simulans, a well-studied human commensal and habitat generalist. *M. citrifolia* contains octanoic acid (OA; LEGAL *et al.* 1994), which is a fatty acid that is toxic to *D. simulans* and other insects but tolerated by *D. sechellia* (RKHA *et al.* 1991). OA typically comprises 58% of the volatile chemicals in a ripe *M. citrifolia* (FARINE *et al.* 1996; PINO *et al.* 2010), which makes it the main toxic component of the fruit. *D. sechellia* also prefers *M. citrifolia* over other fruit for consumption and oviposition (RKHA *et al.* 1991; LEGAL *et al.* 1992; MATSUO *et al.* 2007). As a result of this adaptation, *D. sechellia* has limited competition for access to *M. citrifolia* and may be protected from predation (JONES 2005).

Prior work coarsely mapped several tolerance factors and identified some loci underlying the preference behavior (JONES 1998; JONES 2001; COLSON 2004; JONES 2004; MATSUO *et al.* 2007; EARLEY and JONES 2011). As with other genetic studies of host preference, the earlier work lacked the resolution needed to confidently test the genetic linkage hypothesis or did not assay tolerance (e.g. MATSUO *et al.* 2007; EARLEY and JONES 2011). Moreover, because a specific gene involved in OA tolerance was not identified, the specific mechanism of OA tolerance in *D. sechellia* remains undetermined.

In this study, we ulta-fine map OA tolerance by genotyping independently derived recombinants using visible markers and a panel of molecular markers, along with a new phenotypic assay that provides reliable doses of OA vapor to flies without allowing them to directly contact the toxic chemical. We then measured the preference behavior of these recombinants in a test of the genetic linkage hypothesis. As preference and tolerance are not tightly linked, we reject the linkage hypothesis for this region. We hypothesize that

the evolution of *D. sechellia* into *M. citrifolia* specialist occurred through a step-wise, gradual gain of tolerance and loss of behavioral aversion to *M. citrifolia's* toxins.

MATERIALS AND METHODS

D. simulans/D. sechellia introgression lines

Jones (1998) identified a region harboring resistance alleles on chromosome arm *3*R between two visible markers. As this interval had the greatest effect on resistance, we dissected it further by generating a set of *D. sechellia*/*D. simulans* introgression lines. We used these fifteen original introgression lines (OILs; Figure 5.1) to recombine elements of *D. sechellia* into the *D. simulans* background, using the *Dsim\jv st e osp p* mutant line (14021-0251.173, *Drosophila* species stock center) and *D. sechellia* S9 (M. Ashburner stock collection, Cambridge, UK). The presence or absence of a *D. sechellia* introgression was monitored with these recessive visible markers. Introgressed regions were present if the dominant wild-type phenotype was seen (from *D. sechellia*) rather than the recessive visible mutation (from *D. simulans*).

The OILs were made by crossing *D. sechellia* females with the males of the mutated *D. simulans* line. F_1 hybrid females were then backcrossed to the same mutant *D. simulans* males. The recessive mutations in the *D. simulans* background were then visible in some of the F_2 recombinants. Individual females with the desired visible marker combinations were backcrossed to mutant *D. simulans* males for over 20 generations to reduce the size of the *D. sechellia* introgression on chromosome 3 and eliminate *D. sechellia* contamination from the rest of the genome. The fifteen OILs represent every combination of the four visible markers, with the presence of the *D. sechellia*

introgression in each OIL represented by the presence of a black line in Figure 5.1a. Each OIL was derived from a pool of many F_2 females sharing the same marker phenotype, so each OIL "line" is actually a population. While male F_1 *D. simulans/D. sechellia* hybrids are sterile, 20th generation introgression males are fertile. To stably maintain the *D. sechellia* introgressions of each OIL, we then backcrossed 20th generation OIL males to *D. simulans* females. The *D. sechellia* introgressions were maintained as heterozygotes.

Generation of two "gold standard" lines and large panel of recominant lines

The OILs were tested for OA tolerance (data not shown) and it was determined that OIL 10 flies could be used to generate high and low gold standard lines (called "High 10" and "Low 10", respectively). High 10 and Low 10 contain a *D. sechellia* introgression spanning the *e* locus, but exhibited very different responses to OA exposure. We assayed a large number of OIL 10 males and used three high and three low tolerance flies, respectively, to create High 10 and Low 10 at the beginning of the study. They were used to calibrate further tolerance assays.

Preliminary data indicated the OA tolerance locus was near *e*. We used OIL 8 to generate a huge population of individual recombinants in this region with unknown tolerance (Figure 5.1b). OIL 8 contains a large introgression spanning *st*, *e*, and *osp* loci. Three highly tolerant OIL 8 males were backcrossed to *D. simulans* females to create the line "High 8." To generate individual recombinant flies (Figure 5.2), we crossed High 8 females to *D. simulans* males and collected male offspring with introgressions at *st* and *osp* but not *e* (similar to OIL 6), only at *e* (similar to OIL 10), at *e* and *osp* (similar to OIL 9), and only at *st* (similar to OIL 5). All of these recombinants had a breakpoint between

e and a neighboring marker (either *st* or *osp*). We gave each line an arbitrary number followed by a dash and the number of the OIL to which it was phenotypically identical (e.g. 197-6 is an OIL 6 line). Initially, we generated 36 new recombinant lines to validate the OA tolerance assay (called the 36 "unknown" lines; see Appendix C). Once it was clear that the assay was viable, we created another 700 recombinant lines and genotyped/phenotyped them as described below to finely map the OA tolerance locus.

Genotyping - CAPS

We genotyped recombinant lines using Cleaved Amplified Polymorphic Sequence (KONIECZNY and AUSUBEL 1993). Primers were designed to amplify both *D. simulans* and *D. sechellia* sequences that contained polymorphic restriction cut sites. PCR amplicons from *D. sechellia* sequence possessed an intact restriction site, whereas *D. simulans* amplicons did not. Recombinant flies possessing an introgression between flanking CAPS sites were propagated, and recombinant offspring were subjected to further CAPS genotyping with increasingly fine-mapped CAPS marker sites (Figure 5.2). Overall, 700 unique recominant males were genotyped, using Acc65I, EcoRI, HindIII, HpyCH4IV, and SpeI.

Genomic DNA was extracted from males at each generation using a single fly purification method. Briefly, a single male fly was frozen at -80°C then homogenized with a pipette tip in a "squishing buffer" (10 mM Tris-HCL pH 8.2, 1 mM EDTA, 25 mM NaCl). To this mixture, 1ul 0.2 mg/ml of Proteinase K was added and incubated at 37°C for 20 min., then inactivated at 95°C for 2 min. The resulting DNA was PCR-ready.

Genomic positions listed herein are the *D. melanogaster* positions from *D. melanogaster* Gene Models/Evidence Release 5 (FLYBASE 1999) identified using a syntenic alignment with *D. simulans* and *D. sechellia*.

Octanoic Acid Tolerance Assay

Tolerance to OA in *D. sechellia, D. simulans*, and hybrid recombinants was assayed using a vapor delivery system (Figure 5.3). A fish tank pump, regulated by a flow meter, pushed air through plastic tubing submerged in a tube of liquid OA at ~2.2 liters/minute, followed by a second tube of OA, and finally into a third tube with flies (additional details are provided in the Supplemental methods). To ensure full OA saturation, air was pumped for at least an hour before fly testing. The entire apparatus was in a fume hood with full light and ambient temperature (20-25°C).

Flies were collected 4-7 days post-eclosion with light CO_2 anesthesia no fewer than 4 days pre-test. Between 10-60 flies were dumped in the test chamber and every two minutes the number of "knocked-down" (KD) flies were counted. Every two minutes KD flies were counted, up to a total of 30 minutes. Typically, OA exposure induces neuro-toxin-like symptoms in flies: frantic whole-body movement, leg and wing twitching, and finally KD, where flies either invert their body or collapse while upright. When needed, we tapped the test chamber to distinguish tolerant flies at rest versus KD flies.

Data Analysis

Cumulative KD counts within a line were calculated as a proportion (#flies KD/#flies total) then logit transformed, and a linear model compared these values against

log-transformed time of OA exposure (0-30 min. at 2 min. intervals). "Knock Down 50%" (KD50) was calculated as the time at which 50% of the flies in a given assay were knocked down (R library MASS; R DEVELOPMENT CORE TEAM 2012). To determine the influence of genotype, mutant phenotype, fly test chamber density, air flow rate, and sex on KD50 values, we constructed a linear model and performed an ANOVA. Significant differences in KD50 between lines and sexes were calculated using Welch's *t*-test.

Morinda *tolerance assay*

Five recombinant lines were tested for tolerance to ripe *Morinda* (High 8, High 10, Low 10, 335-6, and 197-6) and pure *D. sechellia* and *D. simulans*. Naturally fallen fruit from *M. citrifolia* trees in a climate-controlled greenhouse were stored in plastic bags and used within three days. A pea-sized fruit pulp (no seeds) was spread across the top of a 60mm pyrex petri dish. Individual flies were aspirated into this petri dish then placed in a growth chamber (25°C, 60% humidity). After 30 minutes, flies were observed for KD every 10 min. up to a maximum of 60 minutes (most flies survive to 30 min.). "Survivorship" was measured as the first block of time during which knockdown was noticed. If no KD was observed at 60 minutes, that fly was scored as 70 and the test ended. Wilcoxon rank sum test compared differences between *D. simulans* and recombinant lines (Holm corrected *P*-value).

Behavior Experiments

F_1 back-crossed flies aged 2-10 days post-eclosion were subjected to a behavioral assay as in DWORKIN & JONES (2009; EARLEY and JONES 2011). Briefly, flies of mixed

sex were introduced without anesthesia into the assay chamber (2L glass beaker, Fisher). Within the chamber were two glass milk bottles, open on the top, containing either control or test food (22mL diH20, 4.1g instant fly media 4-24, Carolina Biological Supply), where test food was identical to control food save for acid inoculate (0.2% OA and 0.06% hexanoic acid, HA, v/v). The two food types were otherwise similar in color, texture, and water content. Cheese cloth was rubber-banded over the entire chamber to prevent escape, and the chamber was placed in a growth chamber (25°C, 60% humidity) overnight. The next day, flies located within each milk bottle were counted and sexed. A response index was calculated: RI = (# flies in test bottle - # control bottle) / (# test bottle + # control bottle). This index was logit transformed to perform parametric tests. Data were analyzed using all-by-all *t*-tests with an FDR of 0.05 (R DEVELOPMENT CORE TEAM 2012), and a GLM (Normal , model: $RI = Line + Sex + Line\ X\ Sex + err$; SAS Institute, Cary, NC). This assay captures fly "settling" behavior, which likely includes both positional preference and ovipositional preference–prior work suggests that they give qualitatively similar results (EARLEY and JONES 2011). Flies that chose a medium generally do not switch to the alternate medium (data not shown).

The recessive effect of introgressed loci was tested by selfing F_1 flies to create F_2 that segregated introgressed regions in both heterozygous (*D. simulans/D. sechellia* and homozygous (*D. sechellia/D. sechellia*) state. Any F_2 flies expressing the recessive *D. simulans* markers (hence no introgression) were removed. F_2 flies were then pooled and tested in the same way as F_1.

RESLUTS

New high-throughput assay for volatile fatty acid tolerance.

We developed and validated a new apparatus for measuring tolerance to OA and other volatile fatty acids. In contrast to past studies, we focused on adult, rather than larval host performance. Adult foraging is important in host preference for some phytophagous insect species, such as the grass miner *Chromatomyia nigra* (SCHEIRS *et al.* 2005), the chrysomelid *Altica carduorum* (SCHEIRS *et al.* 2005), and *Liriomyza trifolii* (SCHEIRS *et al.* 2005). These species prefer to oviposit and feed on host plants best suited for adult performance. Also, SCHEIRS et al. (2005) points out that several studies only considering larval performance also suggest that adult performance may have been affected by host quality (e.g. KAROWE 1990; HERR AND JOHNSON 1992; LU AND LOGAN 1994). As *D. sechellia* exclusively feed on *M. citrifolia* (TSACAS AND BACHLI 1981) and fresh *M. citrifolia* can be toxic even to *D. sechellia* larvae (RKHA *et al.* 1991), we believe that adult performance, rather than larval, may be a key component of this adaptation.

The High 10 and Low 10 gold standard lines were used to initially validate our new "vapor" assay. We found a significant, and repeatable, difference in tolerance when High 10 males and females were compared to their respective sexes for Low 10 and *D. simulans*, but not between the latter two lines for each sex (*P*-values in Table 5.1; KD50s plotted in Figure 5.4a; Figure C.1). *D. simulans* was used as the low tolerance control, while *D. sechellia* was not knocked down after three consecutive hours of exposure. We used Welch's *t*-test instead of ANOVA because the variances for the High and Low 10 male and the High and Low 10 female comparisons were not equal (Bartlett Test *P*-values = 0.0079 and 0.0034, respectively).

Our data show that knockdown is highly reproducible at a given flow rate, with higher flow rates resulting in more rapid knockdown (Figure C.2). Fly density had no

effect (see Appendix C). The assay could reliably and repeatedly classify 36 "unknown" lines for resistance (Figure 5.4b/c; Appendix C). These lines were also used to test if the visible genetic markers affected tolerance. While the visible marker *ebony* affected tolerance, this effect was generally weak and background specific (e.g. the same marker did not always have the same effect across lines; Appendix C) and did not correlate with the number of markers in the genetic background.

High 10/Low 10 "gold standard" lines show differential tolerance by sex

In all lines tested for male/female differences (High 10, Low 10, and *D. simulans*; Table 5.1), the females were more tolerant to OA exposure than males. The percent of cumulative knock down over time of all individual replicates for High 10, Low 10, and *D. simulans* are plotted by sex in Figure C.3 (mean values in Table C.1).

The difference in tolerance by sex could be due to a different mechanism for tolerance in females than in males. JONES (1998) found the effects to be of different magnitudes between the sexes for every region he studied, along with an epistatic interaction between all three major chromosomes and one between markers *y* and *f* on the X chromosome in females, but not males. However, he did find effects for females in every region in which he found effects for males and explained the epistasis involving the X chromosome as possibly due to the X being hemizygous. Our results showed a significantly higher tolerance in females than males, but both sexes in highly tolerant lines exhibited significantly higher tolerances than their respective low tolerance counterparts. As the difference between the low and high tolerance lines was much larger than that between the sexes, it seems for this particular locus that the mechanism

involved in OA tolerance is the same for both sexes. It is likely that females were more tolerant due to their larger size, although we cannot specifically rule out that females have a different overall tolerance mechanism than males.

Fine mapping of tolerance using recombinant line screening with CAPS markers

Once we validated the tolerance assay with the High/Low 10 and 36 "unknown" lines, we focused on narrowing the region of interest. Using a marker panel consisting of 47 CAPS markers, we screened 700 independent *D. simulans*/*D. sechellia* recombinants. 24 of these new recombinants helped narrow the region of interest by several hundred kilobases. Ultimately, seven of the recombinants had a breakpoint that helped define the final 18-gene region (*i.e.* had a boundary marker adjacent to it). We preserved the haplotypes of these seven "boundary lines" for further testing. The boundary lines clustered into clear low and high tolerance groups, similar to the 36 unknown lines, with a significant difference between the KD50s of the lines from each group (Welch's *P*-value = 4.385e-04; Figure C.4). The final boundary markers of the region containing the tolerance locus are on *3*R at positions 1,913,252, defined by 335-6 (low tolerance), 697-6 (low), 505-10 (high), 525-10 (high), and 553-10 (high), and 2,082,441, defined by 197-6 (high) and 725-6 (high) (Figure 5.2).

Formally, these effects could be a by-product of the particular lines used in this introgression, the species chosen, or the hybrid background produced by this introgression. However, earlier work used a variety of different *D. simulans* and *D. sechellia* backgrounds and all were qualitatively similar (AMLOU *et al.* 1998b; JONES 1998; JONES 2001). Alternatively, *D. simulans* could be the outlier instead of *D.*

sechellia. Analysis of *D. mauritiana*, a susceptible sister species of *D. sechellia* and *D. simulans*, shows that in this region *D. mauritiana* alleles tend to be more *D. simulans*-like than *D. sechellia*-like or none of the three species appears to be a strong outlier (Table C.6 and Figure C.4), suggesting that *D. simulans* alleles in this region were likely not influencing tolerance any more than *D. mauritiana* alleles would have been.

Tolerance locus confirmed using fresh M. citrifolia *fruit*

Consistent with our vapor-based tolerance assay, the *M. citrifolia* tolerance assay showed clear differences between the *D. simulans* background line (*Dsim\jv st e osp p*) and several of the recombinant lines. *D. simulans* had a mean (± S.E.) knockdown of 31.67 ± 3.89 minutes in males and 32.50 ± 5 minutes in females, which is close to the minimum knockdown time allowed in the experiment (*i.e.* knockdown was first noted at 30 minutes of exposure to the fruit). A *D. sechellia* line (synA) remained upright for the entire 70 ± 0 minutes (maximum time allowed) in both males and females.

Lines that were tolerant to pure OA were generally also tolerant to *Morinda* fruit. In a mixed sex analysis comparing individual lines to their *D. simulans* background (Wilcoxon rank sum test), lines High 8, High 10, and 197-6 were all significantly more tolerant than *D. simulans*, whereas Low 10 and 335-6 were not (Table 5.2). However, this difference appears to be driven by higher tolerance in females compared to males across all lines ($P < 0.001$, Wilcoxon, Figure 5.5, Table 5.2). This was particularly pronounced in the high tolerance lines. High 8 males were knocked down at 30 ± 0 minutes, while the females had a mean of 50 ± 4.71 minutes. 197-6 males were down by 34 ± 1.89 minutes, while the females went down at 45.71 ± 3.42 minutes. High 10 males were down by 37.5

± 2.7 minutes and females by 54.5 ± 5.1. This pattern was also seen in one of the low tolerance lines, Low 10, where males averaged 30.53 ± 0.53 and females averaged 36 ± 4.00 minutes. 335-6 was the exception, with the males going down at 30.83 ± 0.58 and the females similarly at 30.69 ± 0.48 minutes. Of the highly tolerant recombinant lines, only 197-6 and High 10 had males stay upright beyond the initial 30-minute check. Thus, the mixed sex results were mostly driven by the female tolerance.

Final tolerance region includes 18 genes, including* Obp *and* Osiris *families

A list of the 18 genes found in the ~170Kb tolerance region on *3*R is in Table C.3. Of the 18 genes, only three remain unnamed. Two families of genes are represented. A cluster of three odorant-binding proteins (*Obp*) is present, along with nine *Osiris* genes. The three other named genes are *Gasp*, *Vha14-2*, and *NPFR1*.

The Ka/Ks ratios (LI 1993) from both the *D. simulans* and *D. sechellia* lineages, as well as non-synonymous sites and other sequence information, for the 18 genes are summarized in Tables S4 and S5 (also see Appendix C). There is not a strong signature of positive selection at any of these loci. Only 7 of the 17 genes in *D. simulans* had any non-synonymous changes (41%), while 13 had such changes in *D. sechellia* (76%).

Earlier work contrasted levels of transcription of these genes across species, tissues, and treatments (KOPP *et al.* 2008; DWORKIN and JONES 2009). None of the *Osiris* family genes showed differential expression. Between species' bodies, only CG31562 (a CHK kinase-like protein of unknown function) was significantly different between *D. simulans* and *D. sechellia* (*D. sechellia* expresses ~1/3 as much as *D. simulans*). In heads, *Obp83cd*, was also significantly down in *D. sechellia* compared to *D. simulans* (~1/2).

Obp83cd was not differentially expressed in the antennae. *Obp83ef* was differentially expressed in the antennae, but the difference between *D. simulans* and *D. sechellia* was substantially less than the difference between *D. simulans* and *D. melanogaster* suggesting that expression of this gene is evolutionarily labile and that this expression change is not associated with the shift to *M. citrifolia*. In an experiment looking at differential changes in gene expression associated with exposure to OA and HA, the gene CG1077 was weakly induced (~1.5X).

Tolerance alleles do not affect host preference behavior

D. sechellia exhibits strong preference for *M. citrifolia's* fruit and its constituent fatty acids, OA and HA. These compounds, in contrast, are highly aversive to *D. simulans*. To test if the tolerance conferred by the 18-gene region or the genes within this region affected behavior, we measured the preference of seven recombinant lines (High/Low 10, High 8, and *D. simulans* background stock) using our established preference assay (EARLEY and JONES 2011). Most lines did not differ from the *D. simulans* background control (Figure 5.6a; pairwise t-test and GLM, $P > 0.05$). The exceptions, High 8 (both males and females; $P = 0.017$) and Low 10 (males only, $P = 0.011$), behaved in the opposite of expectation—lower tolerance resulted in less aversion. Similarly, the trend was for flies with higher tolerance to avoid the OA medium (Figure 5.6b), although this trend was not significant ($P = 0.3506$). To improve power, we pooled high tolerance lines and low tolerance lines and compared preference behavior between these two groups (replicates: 55 high, 26 low; 4993 flies). There was no effect for either sex and the trend was in the opposite of expectation (mean $RI_{High} = -1.49$; mean $RI_{low} = -$

1.21; main effect, $P = 0.148$, sex $P = 0.157$; interaction $P = 0.207$). Similarly, there was no difference between the pooled high or low lines and the *D. simulans* background control (High: $P = 0.408$; Low $P = 0.121$). We confirmed for a subset of lines that this pattern was consistent for tolerance to *M. citrifolia* fruit (Figure 5.6c). These data suggest that the increased tolerance conferred by the introgressed *D. sechellia* region is not sufficient to change behavior and excludes the possibility of an additive or dominant acting preference locus in this interval.

We assayed a subset of lines for a recessive preference factor linked to the tolerance factors. From the lines showing highest (high10, 197-6) and lowest OA tolerance (low10, 335-6), we sib-mated F_1 flies to create F_2 progeny segregating introgressions in homozygous (*D. sechellia/D. sechellia*) and heterozygous (*D. simulans/D. sechellia*) states, removing F_2s that had no introgression (*D. simulans/D. simulans*). If a recessive *D. sechellia* preference factor existed within an introgression line, then homozygous flies for that factor should have manifested higher OA preference and shifted F_2 behavior relative to a population of fully heterozygous introgressions. We pooled homozygote and heterozygote F_2s heterozygotes and found no significant difference in their behavior compared to fully heterozygous F_1 flies (Figure 5.7).

DISCUSSION

Many insects feed on only one or a few types of host. Genetic linkage between alleles contributing to host preference and alleles contributing to host usage, such as tolerance of secondary compounds, has been suggested to facilitate the evolution of new host specializations. We used a forward genetic approach and a novel assay to see if this

type of genetic correlation contributed to the evolution of the host specialization in *D. sechellia*. We isolated a ~170Kb region on *3*R harboring 18 genes that contains at least one locus affecting OA tolerance, a critical element of *D. sechellia's* adaptation to the toxic fruit of its host plant, *M. citrifolia*. While the *D. sechellia* introgression conferred OA tolerance in a *D. simulans* background, it had little to no effect on host-seeking behavior.

M. citrifolia *assay validates OA apparatus and methodology*

To screen the thousands of flies needed for introgression mapping of tolerance loci, we developed a new assay for volatile fatty acid resistance. This assay uses OA to mimic the toxic effects of the fruit. Our analysis revealed that exposure to *M. citrifolia* fruit is quantitatively similar to the OA assay results, in terms of consistency in the lines that exhibited highly and lowly tolerant behavior. However, the mixed sex results for lines exhibiting high tolerance to *M. citrifolia* were driven by tolerance of the females (i.e. the males showed lower tolerance levels). Unfortunately, we do not know the concentration of OA in either the actual vapor of the OA assay or in the fruit itself (there is considerable variation among fruits and across ripening stages; (LEGAL *et al.* 1992; PINO *et al.* 2010). Also, it is likely the fruit has a higher concentration of OA than the maximum our pump can produce. The consistently high tolerance of the females in both experimental setups suggests that the same tolerance mechanism is being assayed in both. Some of the variability among the sexes likely reflected the larger size of the females. Experimentally, the high variability among lines in the *M. citrifolia* assay suggests that

using the OA vapor methodology may provide sensitivity to detect moderate to weak effect loci unattainable using fruit.

The test apparatus proved to be an effective way to expose the flies to OA without killing them. Contact with OA in high doses is often fatal even to *D. sechellia* (COLSON 2004). Thus, distributing the acid as a vapor eliminated the error caused by flies coming into contact with the acid and going down before they otherwise would have. This likely explains the strong repeatability of replicates within and between lines exhibiting the same type of tolerance. The sensitivity of the assay was evident in the fact that male versus female tolerance differences could be consistently detected, as well as differences between the OIL phenotypes.

Tolerance region harbors a handful of candidate loci

Of the 18 genes in this region, two gene families represented two-thirds of the total and only three remain unnamed. A cluster of nine *Osiris* genes was present, along with three *Obps*. None of these 18 genes, however, showed a strong signature of positive selection that may be expected for a gene contributing to *D. sechellia* adaptation to its host. Likewise, gene expression data did not strongly implicate any one locus (but see *Obp83cd* below).

Osiris *genes*

According to DORER et al. (2003), the Osiris gene family is clustered at the *Triplo-lethal* locus in *D. melanogaster*. All have endoplasmic reticulum signal peptides, may be integral to the plasma membrane, may have important housekeeping functions

and are highly dosage-sensitive. In addition, their linkage and sequences are unusually highly conserved, as seen in *Anopheles gambiae* (DORER *et al.* 2003). While none of these genes can be ruled out, it seems unlikely that their functionality can be appreciably altered without dire consequences to the individual.

Odorant-binding proteins

Three genes stand out in the 18-gene region based on their gene family: *Obp83cd*, *Obp83ef*, and *Obp83g*. *Obps* are a family of proteins involved in olfactory perception, but their function is not fully understood (VIEIRA *et al.* 2007). They are water-soluble and exist in the aqueous lymph surrounding odorant receptors in the chemosensory sensilla of insects. They enhance the solubility of hydrophobic odorants by binding to them and transporting them through the extracellular lymph to the dendritic membrane of neurons (WHITEMAN and PIERCE 2008). *Obps* are typically found in olfactory tissues, but expression analysis shows that they are not limited to them (PELOSI *et al.* 2006). Besides transporting odorants, it has been suggested that *Obps* may also act as scavengers, removing toxic odorant molecules to prevent damage to cells (STEINBRECHT 1998; BLOMQUIST and VOGT 2003). The possible involvement of *Obps* in OA tolerance discovered in this study may provide evidence for this previously hypothesized link between *Obp* chemical detection and detoxification.

Obp83cd (GALINDO and SMITH 2001) and *Obp83ef* (GALINDO and SMITH 2001; KOPP *et al.* 2008) are significantly down-regulated in the head and up-regulated in the antennae in *D. melanogaster*. *Obp83cd* is expressed in the labellum (GALINDO and SMITH 2001), while *Obp83ef* is expressed in the antennae and other nonspecific tissues

(GALINDO and SMITH 2001; KOPP *et al.* 2008). *Obp83ef* is up-regulated in *D. sechellia*, and *D. simulans* has intermediate expression between *D. melanogaster* and *D. sechellia* (KOPP *et al.* 2008). Intriguingly, *Obp83cd* shows a species-specific pattern of expression: *Obp83cd* has reduced expression in *D. sechellia* relative to *D. simulans* (DWORKIN and JONES 2009).

As with most genes, the regulatory regions of these *Obps* are not well characterized. The Regulatory Element Database for *Drosophila* v3.0 (GALLO *et al.* 2011) indicated that the regulatory regions for both *Obp83cd* and *Obp83ef* are 3 kb upstream of the start codons. We aligned these upstream regions, using *D. melanogaster* as the outgroup. For *Obp83cd*, there were 34 changes in *D. sechellia*, along with a 43 bp deletion. *Obp83ef* had 28 changes in *D. sechellia*. REDfly did not specify transcription factor binding sites, but any one of these upstream changes could alter the regulation of these genes in *D. sechellia*. No regulatory information was available for *Obp83g*.

Obp83cd and *Obp83g* have a *D. sechellia* Ka/Ks higher than the mean for the region, while *Obp83ef* has a Ka/Ks that is much lower than the mean. *D. simulans* alleles for *Obp83cd* and *Obp83g* have no non-synonymous changes, while *D. sechellia* alleles have five and two, respectively. Of the genes that do not end pre-maturely or contain frameshifts, *Obp83cd* has the largest Ka (0.0076) of the *D. sechellia* alleles. *Obp83ef* has one non-synonymous change in both species, but a much higher Ka/Ks in *D. simulans* (0.6249 versus 0.1073). Coding changes may not be involved in OA tolerance, but if they are, these genes qualify as candidates with at least one amino acid change.

Obps have previously been associated with OA avoidance behavior (DWORKIN and JONES 2009), *M. citrifolia* preference (MATSUO *et al.* 2007), and host plant

preference in *Drosophila* (KOPP *et al.* 2008), but not to tolerance of the normally toxic effects of OA experienced by most *Drosophila* species. While we have been unable to eliminate the other 15 genes, these three *Obps* are our strongest candidates for future analysis.

No evidence for genetic linkage between preference and tolerance loci in this region

A positive genetic correlation between the preference and performance alleles due to pleiotropy or genetic linkage can facilitate the evolution of a new host specialization. This 18-gene region spans only ~170,000 bp and has an estimated recombination rate of 0.28 cM. These genes are therefore tightly linked. However, we find no evidence that harboring the *D. sechellia* tolerance alleles in this region causes the flies to behave in a more *D. sechellia*-like manner. Indeed the "High 8" line, which spans a much larger interval (up to 23.3 cM on chromosome *3*, although the actual boundaries have not been mapped), does not appear to be significantly different from the controls and is marginally more aversive than some low tolerance lines. Previous work noted that *D. sechellia* preference for OA was recessive to *D. simulans* aversion (HIGA and FUYAMA 1993; AMLOU *et al.* 1998a). We tested for recessive effect loci on high and low tolerance lines by selfing F_1 hybrids to segregate F_2 progeny with pooled homozygous and heterozygous *D. sechellia* introgressions (removing any F_2 with recessive homozygous *D. simulans* markers – hence, no introgression). If recessive OA preference loci exist within these introgressions, then F_2 flies should exhibit increased preference relative to F_1 flies. We did not see this (Figure 5.7).

To our knowledge, our data represent the highest resolution analysis of the genetic relationship between preference and performance. The genetic resolution of many earlier studies suggesting linkage between these traits biased towards overlap between QTLs for preference and performance, or even more disparate visible markers. Our data suggest that if these studies had resolution comparable to ours the apparent linkage between preference and performance would be reduced. However, in the current study we have a different concern. We are only looking at a single locus and the critical early association between behavior and tolerance could have occurred elsewhere in the genome. Recent work, however, has identified strong preference factors on chromosome *2*, which has the weakest effect on tolerance (JONES 1998; MATSUO *et al.* 2007; DWORKIN AND JONES 2009). Similarly, the X chromosome, which has at least two tolerance factors, has no effect on preference behavior (JONES 1998; JONES 2001; JONES 2004; EARLEY AND JONES 2011). Furthermore, the tolerance region on *3*R was previously shown to be one of the two largest contributors to resistance (the other locus is flanking; (JONES 1998). Preferring the toxic host without tolerance alleles in this region would be deleterious. Together these data suggest that linkage between preference and tolerance factors did not play a major role in the evolution of *D. sechellia's* specialization.

There are several examples of genetically unlinked development of host preference and performance. Similar to our data, earlier work by JAENIKE (1989) also ruled out tight linkage between preference and performance in *D. tripunctata*. The same was found in other herbivorous species' (see Introduction). Theoretically, the quickest way to achieve speciation through host specialization is if host preference and performance each have a simple genetic architecture and are tightly linked to each other

(FRY 2003). JAENIKE (1987) posited that it is unlikely for linkage disequilibrium to establish a genetic correlation between preference and performance in a system with more than a few interacting alleles. More likely, pleiotropy would explain such a linkage. (In the case of *D. sechellia*, we also eliminate this possibility.) Assortative mating on the new host plant may allow unlinked evolution of preference and performance during sympatric divergence (DIEHL and BUSH 1989). This scenario would likely require a few new alleles of large effect, which are seen in *D. sechellia*'s adaptation to *Morinda* fruit, but also may have resulted in a fitness gap between the old and new hosts. In this case, the deleterious effects of the new host on the performance of the unadapted fly may have been overcome by reduced competition on the new host (WALLACE 1968; BERLOCHER and FEDER 2002), since OA is toxic to *D. sechellia*'s sibling species. Alternatively, since five or more loci may be involved in OA tolerance (*i.e.* a polygenic architecture), it is possible that part or most of the speciation process occurred gradually in allopatry (TEMPLETON 1981).

Evolution of* D. sechellia's *host specialization likely occurred in a step-wise manner

D. sechellia's loss of OA avoidance and its development of preference for *M. citrifolia* seemed to derive in part from the elimination of related *Obp*s. DWORKIN AND JONES (2009) found that *Obp56e* had a premature stop codon in *D. sechellia* and that *D. melanogaster* showed reduced avoidance of *M. citrifolia* when *Obp56e* was knocked down. MATSUO et al. (2007) found a 4-bp insertion upstream of *Obp57e* in the *D. sechellia* allele affecting preference, which prevented expression when heterozygous within a *D. melanogaster Obp57e* deficiency line. If loss of functional *Obps* occurred

first in the evolution of *D. sechellia's* specialization on *M. citrifolia*, it would certainly have produced large selection pressure toward the development of tolerance on the flies that were coming into contact with the fruit.

Loss-of-function mutations are usually recessive, as they often result in the elimination or reduction of protein expression or of non-functional protein structures. However, JONES (1998) concluded that at least five loci involved in *D. sechellia* tolerance were dominant. Similarly, pesticide resistance factors are usually dominant or codominant (OTTEA and PLAPP 1984; ROUSH and MCKENZIE 1987; HOUPT *et al.* 1988; FFRENCHCONSTANT *et al.* 1993; ROUSH 1993). This asymmetry in dominance may reflect the different genetic mechanisms through which these two traits evolved. Loss of avoidance can be achieved by the loss or reduction of a sensory response, such as through pseudogenization of *Obps*. In fact, MCBRIDE (2007) discovered that *D. sechellia* harbors an unusually large number of pseudogenized chemosensory genes relative to *D. simulans* or *D. melanogaster*, even when controlling for *D. sechellia*'s small population size. Toxin resistance, in contrast, often requires *increased* expression of existing detoxification genes or *gain* of a new physiological mechanism. Both cases may result in additive to completely dominant phenotypes. This pattern may imply that during the early genetic steps of the evolution of a new specialization—or any adaptation—that "loss" of an ancestral trait may readily evolve from common recessive nulls segregating in the standing genetic variation. In contrast, dominant phenotypes associated with new traits and gain-of-function alleles may involve more new mutations and other relatively rare alleles as these dominant alleles are expected to be at a lower frequency in the ancestral population (ORR and BETANCOURT 2001).

Acknowledgements

We thank the anonymous reviewers for their comments. Also, we thank Karna Mital for his contributions and Daniel Matute for his helpful advice. This work was supported by the National Science Foundation DEB-0212686 (CDJ), NSF-Graduate Research Fellowship Program (EJE), by NIDDK grant P30DK056350 to the UNC Nutrition Obesity Research Center, by NIH R01 GM058686 and by NIH ARRA R01 supplement GM058686-08S1.

REFERENCES

AMLOU, M., B. MORETEAU and J. R. DAVID, 1998a Genetic analysis of Drosophila sechellia specialization: oviposition behavior toward the major aliphatic acids of its host plant. Behav Genet **28:** 455-464.

AMLOU, M., B. MORETEAU and J. R. DAVID, 1998b Larval tolerance in the Drosophila melanogaster species complex toward the two toxic acids of the D. sechellia host plant. Hereditas **129:** 7-14.

BALLABENI, P., and M. RAHIER, 2000 Performance of leaf beetle larvae on sympatric host and non-host plants. Entomologia experimentalis et applicata **97:** 175-181.

BERLOCHER, S. H., and J. L. FEDER, 2002 Sympatric speciation in phytophagous insects: moving beyond controversy? Annual Review of Entomology **47:** 773-815.

BERNAYS, E. A., and R. F. CHAPMAN, 1994 *Host-plant selection by phytophagous insects*. Chapman and Hall, New York.

BLOMQUIST, G. J., and R. G. VOGT, 2003 *Insect pheromone biochemistry and molecular biology: The biosynthesis and detection of pheromones and plant volatiles*. London: Elsevier Academic Press.

BOSSART, J., 2003 Covariance of preference and performance on normal and novel hosts in a locally monophagous and locally polyphagous butterfly population. Oecologia **135:** 477-486.

CAILLAUD, M. C., and S. VIA, 2012 Quantitative genetics of feeding behavior in two ecological races of the pea aphid, Acyrthosiphon pisum. Heredity **108:** 211-218.

COLSON, I., 2004 Drosophila simulans' response to laboratory selection for tolerance to a toxic food source used by its sister species D-sechellia. Evolutionary Ecology **18:** 15-28.

DIEHL, S. R., and G. L. BUSH, 1989 *The role of habitat preference in adaptation and speciation*. Sinauer Associates, Sunderland, MA.

DORER, D. R., J. A. RUDNICK, E. N. MORIYAMA and A. C. CHRISTENSEN, 2003 A family of genes clustered at the Triplo-lethal locus of Drosophila melanogaster has an unusual evolutionary history and significant synteny with Anopheles gambiae. Genetics **165:** 613-621.

DWORKIN, I., and C. D. JONES, 2009 Genetic Changes Accompanying the Evolution of Host Specialization in Drosophila sechellia. Genetics **181:** 721-736.

EARLEY, E. J., and C. D. JONES, 2011 Next-Generation Mapping of Complex Traits with Phenotype-Based Selection and Introgression. Genetics **189:** 1203-+.

EHRLICH, P. R., and P. H. RAVEN, 1964 Butterflies and Plants - a Study in Coevolution. Evolution **18:** 586-608.

FARINE, J. P., L. LEGAL, B. MORETEAU and J. L. LEQUERE, 1996 Volatile components of ripe fruits of Morinda citrifolia and their effects on Drosophila. Phytochemistry **41:** 433-438.

FFRENCHCONSTANT, R. H., J. C. STEICHEN, T. A. ROCHELEAU, K. ARONSTEIN and R. T. ROUSH, 1993 A Single-Amino Acid Substitution in a Gamma-Aminobutyric-Acid Subtype-a Receptor Locus Is Associated with Cyclodiene Insecticide Resistance in Drosophila Populations. Proceedings of the National Academy of Sciences of the United States of America **90:** 1957-1961.

FLYBASE, 1999 The FlyBase database of the Drosophila Genome Projects and community literature. The FlyBase Consortium. Nucleic Acids Research **27:** 85-88.

FRY, J. D., 2003 Multilocus models of sympatric speciation: Bush versus Rice versus Felsenstein. Evolution **57:** 1735-1746.

FRY, J. D., S. L. HEINSOHN and T. F. C. MACKAY, 1996 The contribution of new mutations to genotype-environment interaction for fitness in Drosophila melanogaster. Evolution **50:** 2316-2327.

FUTUYMA, D. J., and G. MORENO, 1988 The Evolution of Ecological Specialization. Annual Review of Ecology and Systematics **19:** 207-233.

GALINDO, K., and D. P. SMITH, 2001 A large family of divergent drosophila odorant-binding proteins expressed in gustatory and olfactory sensilla. Genetics **159:** 1059-1072.

GALLO, S. M., D. T. GERRARD, D. MINER, M. SIMICH, B. DES SOYE *et al*., 2011 REDfly v3.0: toward a comprehensive database of transcriptional regulatory elements in Drosophila. Nucleic Acids Research **39:** D118-123.

GRIMALDI, D., and M. S. ENGEL, 2005 *Evolution of Insects*. New York/Cambridge: Cambridge University Press.

HAWTHORNE, D. J., and S. VIA, 2001 Genetic linkage of ecological specialization and reproductive isolation in pea aphids. Nature **412:** 904-907.

HERR, J. C., and M. W. JOHNSON, 1992 Host plant preference of Liriomyza sativae (Diptera: Agromyzidae) populations infesting green onion in Hawaii. Environmental entomology **21:** 1097-1102.

HIGA, I., and Y. FUYAMA, 1993 Genetics of food preference in Drosophila sechellia. I. Responses to food attractants. Genetica **88:** 129-136.

HOUPT, D. R., J. C. PURSEY and R. A. MORTON, 1988 Genes-Controlling Malathion Resistance in a Laboratory-Selected Population of Drosophila-Melanogaster. Genome **30:** 844-853.

JAENIKE, J., 1986 Genetic Complexity of Host-Selection Behavior in Drosophila. Proceedings of the National Academy of Sciences of the United States of America **83:** 2148-2151.

JAENIKE, J., 1987 Genetics of Oviposition-Site Preference in Drosophila-Tripunctata. Heredity **59:** 363-369.

JAENIKE, J., 1989 Genetic population structure of Drosophila tripunctata: patterns of variation and covariation of traits affecting resource use. Evolution**:** 1467-1482.

JAENIKE, J., 1990 Host Specialization in Phytophagous Insects. Annual Review of Ecology and Systematics **21:** 243-273.

JANZ, N., 2011 Ehrlich and Raven Revisited: Mechanisms Underlying Codiversification of Plants and Enemies. Annual Review of Ecology, Evolution, and Systematics, Vol 42 **42:** 71-89.

JONES, C. D., 1998 The genetic basis of Drosophila sechellia's resistance to a host plant toxin. Genetics **149:** 1899-1908.

JONES, C. D., 2001 The genetic basis of larval resistance to a host plant toxin in Drosophila sechellia. Genetical Research **78:** 225-233.

JONES, C. D., 2004 Genetics of egg production in Drosophila sechellia. Heredity **92:** 235-241.

JONES, C. D., 2005 The genetics of adaptation in Drosophila sechellia. Genetica **123:** 137-145.

KAROWE, D. N., 1990 Predicting host range evolution: colonization of Coronilla varia by Colias philodice (Lepidoptera: Pieridae). Evolution**:** 1637-1647.

KEESE, M. C., 1996 Feeding responses of hybrids and the inheritance of host-use traits in leaf feeding beetles (Coleoptera: Chrysomelidae). Heredity **76:** 36-42.

KONIECZNY, A., and F. M. AUSUBEL, 1993 A Procedure for Mapping Arabidopsis Mutations Using Codominant Ecotype-Specific Pcr-Based Markers. Plant Journal **4:** 403-410.

KOPP, A., O. BARMINA, A. M. HAMILTON, L. HIGGINS, L. M. MCINTYRE *et al*., 2008 Evolution of gene expression in the Drosophila olfactory system. Molecular Biology and Evolution **25:** 1081-1092.

LANDE, R., 1979 Quantitative Genetic-Analysis of Multivariate Evolution, Applied to Brain - Body Size Allometry. Evolution **33:** 402-416.

LEGAL, L., B. CHAPPE and J. M. JALLON, 1994 Molecular-Basis of Morinda-Citrifolia (L) - Toxicity on Drosophila. Journal of Chemical Ecology **20:** 1931-1943.

LEGAL, L., J. R. DAVID and J. M. JALLON, 1992 Toxicity and attraction effects produced by *Morinda citrifolia* fruits on the *Drosophila melanogaster* complex of species. Chemoecology **3:** 125-129.

LI, W. H., 1993 Unbiased Estimation of the Rates of Synonymous and Nonsynonymous Substitution. Journal of Molecular Evolution **36:** 96-99.

LOUIS, J., and J. R. DAVID, 1986 Ecological Specialization in the Drosophila-Melanogaster Species Subgroup - a Case-Study of Drosophila-Sechellia. Acta Oecologica-Oecologia Generalis **7:** 215-229.

LU, W. H., and P. LOGAN, 1994 Genetic-variation in oviposition between and within populations of Leptinotarsa decemlineata (Coleoptera, Chrysomelidae). . Ann. Entomol. Soc. Am. **87:** 634–640.

MATSUBAYASHI, K. W., I. OHSHIMA and P. NOSIL, 2010 Ecological speciation in phytophagous insects. Entomologia Experimentalis Et Applicata **134:** 1-27.

MATSUO, T., S. SUGAYA, J. YASUKAWA, T. AIGAKI and Y. FUYAMA, 2007 Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in Drosophila sechellia. Plos Biology **5:** 985-996.

MCBRIDE, C. S., 2007 Rapid evolution of smell and taste receptor genes during host specialization in Drosophila sechellia. Proceedings of the National Academy of Sciences **104:** 4996-5001.

NG, D., 1988 A novel level of interactions in plant–insect systems.

NIELSEN, J. K., 1996 Intraspecific variability in adult flea beetle behaviour and larval performance on an atypical host plant, pp. 160-162 in *Proceedings of the 9th International Symposium on Insect-Plant Relationships*. Springer.

ORR, H. A., and A. J. BETANCOURT, 2001 Haldane's sieve and adaptation from the standing genetic variation. Genetics **157:** 875-884.

OTTEA, J. A., and F. W. PLAPP, 1984 Glutathione S-Transferase in the Housefly - Biochemical and Genetic Changes Associated with Induction and Insecticide Resistance. Pesticide Biochemistry and Physiology **22:** 203-208.

PELOSI, P., J. J. ZHOU, L. P. BAN and M. CALVELLO, 2006 Soluble proteins in insect chemical communication. Cellular and Molecular Life Sciences **63:** 1658-1676.

PINO, J. A., E. MARQUEZ, C. E. QUIJANO and D. CASTRO, 2010 Volatile compounds in noni (Morinda citrifolia L.) at two ripening stages. Ciencia E Tecnologia De Alimentos **30:** 183-187.

R DEVELOPMENT CORE TEAM, 2012 R: A language and environment for statistical computing, pp. R Foundation for Statistical Computing, Vienna, Austria.

RKHA, S., P. CAPY and J. R. DAVID, 1991 Host Plant Specialization in the Drosophila-Melanogaster Species Complex - a Physiological, Behavioral, and Genetic-Analysis. Proceedings of the National Academy of Sciences of the United States of America **88:** 1835-1839.

ROUSH, R. T., 1993 Occurrence, Genetics and Management of Insecticide Resistance. Parasitology Today **9:** 174-179.

ROUSH, R. T., and J. A. MCKENZIE, 1987 Ecological Genetics of Insecticide and Acaricide Resistance. Annual Review of Entomology **32:** 361-380.

SAUGE, M. H., P. LAMBERT and T. PASCAL, 2012 Co-localisation of host plant resistance QTLs affecting the performance and feeding behaviour of the aphid Myzus persicae in the peach tree. Heredity **108:** 292-301.

SCHEIRS, J., K. JORDAENS and L. DE BRUYN, 2005 Have genetic trade-offs in host use been overlooked in arthropods? Evolutionary Ecology **19:** 551-561.

SEZER, M., and R. BUTLIN, 1998a The genetic basis of oviposition preference differences between sympatric host races of the brown planthopper (Nilaparvata lugens). Proceedings of the Royal Society of London. Series B: Biological Sciences **265:** 2399-2405.

SEZER, M., and R. K. BUTLIN, 1998b The genetic basis of host plant adaptation in the brown planthopper (*Nilaparvata lugens*). Heredity **80:** 499-508.

SINGER, M. C., D. NG and C. THOMAS, 1988a Heritability of oviposition preference and its relationship to offspring performance within a single insect population. Evolution**:** 977-985.

SINGER, M. C., D. NG and C. D. THOMAS, 1988b Heritability of Oviposition Preference and Its Relationship to Offspring Performance within a Single Insect Population. Evolution **42:** 977-985.

STEINBRECHT, R. A., 1998 Odorant-binding proteins: Expression and function. Olfaction and Taste Xii **855:** 323-332.

TABASHNIK, B. E., 1986 Evolution of host plant utilization in Colias butterflies. Evolutionary genetics of invertebrate behavior. Plenum, New York**:** 173-184.

TAYLOR, C. E., and C. CONDRA, 1983 Resource Partitioning among Genotypes of Drosophila-Pseudoobscura. Evolution **37:** 135-149.

TEMPLETON, A. R., 1981 Mechanisms of speciation - A population genetic approach. Annual Review of Ecology and Systematics **12:** 23-48.

THOMPSON, J. N., 1988 Evolutionary ecology of the relationship between oviposition preference and performance of offspring in phytophagous insects. Entomologia experimentalis et applicata **47:** 3-14.

THOMPSON, J. N., W. WEHLING and R. PODOLSKY, 1990 Evolutionary genetics of host use in swallowtail butterflies.

TSACAS, L., and G. BACHLI, 1981 *Drosophila sechellia*. n. sp., huitieme espece du sous-groupe melanogaster des iles Seychelles (Diptera, Drosophilidae). Revue Francaise d'Entomologie **3:** 146–150.

VIA, S., 1986 Genetic covariance between oviposition preference and larval performance in an insect herbivore. Evolution**:** 778-785.

VIA, S., 2001 Sympatric speciation in animals: the ugly duckling grows up. Trends in Ecology & Evolution **16:** 381-390.

VIA, S., and D. J. HAWTHORNE, 2005 Back to the future: genetic correlations, adaptation and speciation. Genetica **123:** 147-156.

VIEIRA, F. G., A. SANCHEZ-GRACIA and J. ROZAS, 2007 Comparative genomic analysis of the odorant-binding protein family in 12 Drosophila genomes: purifying selection and birth-and-death evolution. Genome Biology **8**.

WALLACE, B., 1968 Polymorphism, population size, and genetic load, pp. 87-108 in *Population biology and evolution*, edited by R. C. LEWONTIN. Syracuse University Press, Syracuse, NY.

WASSERMAN, S. S., and D. J. FUTUYMA, 1981 Evolution of host plant utilization in laboratory populations of the southern cowpea weevil, Callosobruchus maculatus Fabricius (Coleoptera: Bruchidae). Evolution**:** 605-617.

WHITEMAN, N. K., and N. E. PIERCE, 2008 Delicious poison: genetics of Drosophila host plant preference. Trends in Ecology & Evolution **23:** 473-478.

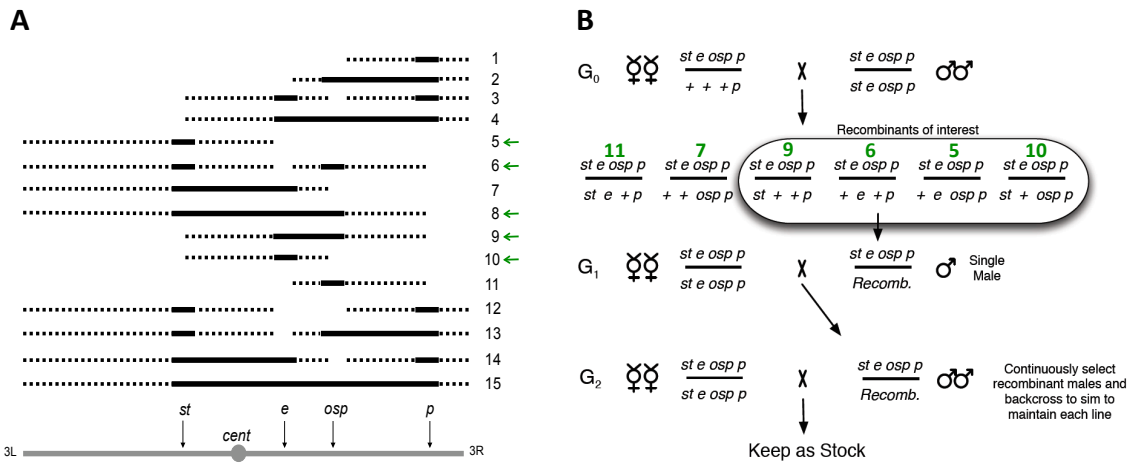

Figure 5.1 Description of OIL genotypes and the cross to construct additional recombinants. (A) Diagram of original introgression lines (OILS). Solid black lines represent *D. sechellia* chromosome in *D. simulans* background. Dashed lines indicate the possible extent of the *D. sechellia* introgressions between visible markers. Four visible markers on chromosome 3 were used to categorize the OILs (*scarlet*, *st*; *ebony*, *e*; *outspread*, *osp*; *pink*, *p*). Introgression lines used for OA tolerance testing are denoted with green arrows (i.e. have a breakpoint between either *st*/*e*, *e*/*osp*, or both). (B) Diagram of fly cross to generate new introgression lines for OA tolerance assay. G_0 females are from OIL 8 and males are pure *D. simulans*. The recombinants of interest (G_1 males) were backcrossed to *D. simulans* females. The six recombinant genotypes (second line of cross) correspond to OILs numbered in green. Visible mutant phenotype was continually selected in males to preserve the haplotype without recombination. *D. simulans* background indicated by recessive mutations (*st*, *e*, *osp*, *p*). Presence of *D. sechellia* introgression denoted by symbol for wild type (+). *Recomb.*, the recombinant chromosome inherited from the G_0 female.

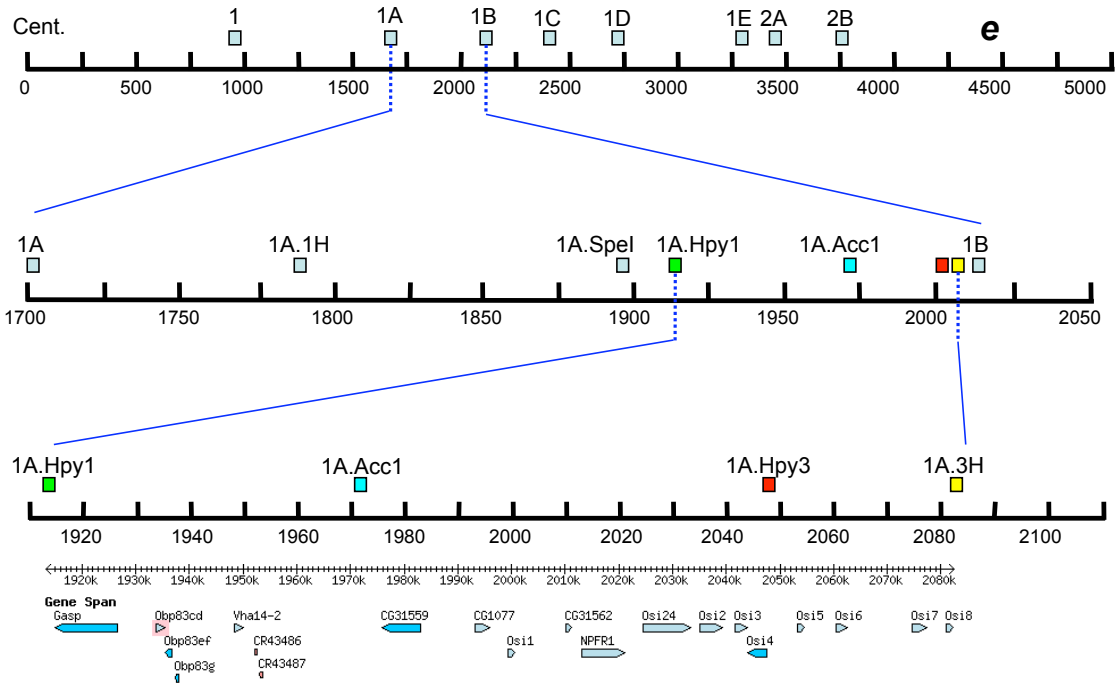

Figure 5.2 Map of CAPS markers used to narrow OA tolerance region on chromosome arm 3R. The dashed and solid blue lines indicate zooming in (note the new scale on each level). The light blue boxes on the top level are markers using EcoRI spanning the region on 3R between the centromere (Cent.) and *ebony* (*e*). Between the first and second levels, the focus becomes the region between markers 1A and 1B on the second level. The marker names ending in 'H' are for HindIII, 'SpeI' for SpeI, 'Hyp#' for HypCH4V, and 'Acc1' for Acc65I. The markers shaded with light blue are outside of the region of interest. The third level represents the ~170 kb region between markers 1A.Hpy1 and 1A.3H. These markers are color-coded for easy identification between second and third levels. The bottom panel shows the names and positions of the 18 candidate genes (Flybase). Scale for chromosome position: 1000x.

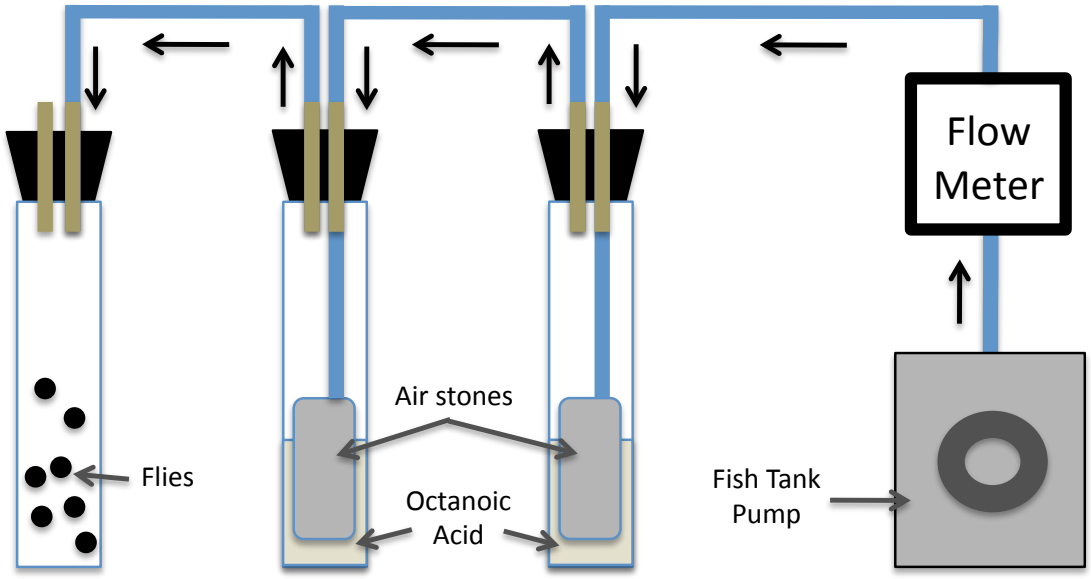

Figure 5.3 Diagram of OA tolerance test setup. OA vapor flows in the direction of the black arrows through plastic tubing (thick blue lines) and is bubbled through air stones twice to ensure saturation before reaching the flies. The OA/air stones and flies are housed in glass vials sealed with black stoppers. The stoppers have holes for glass tubing to pass through (brown lines), which are attached to the plastic tubing. The chamber containing flies is vented.

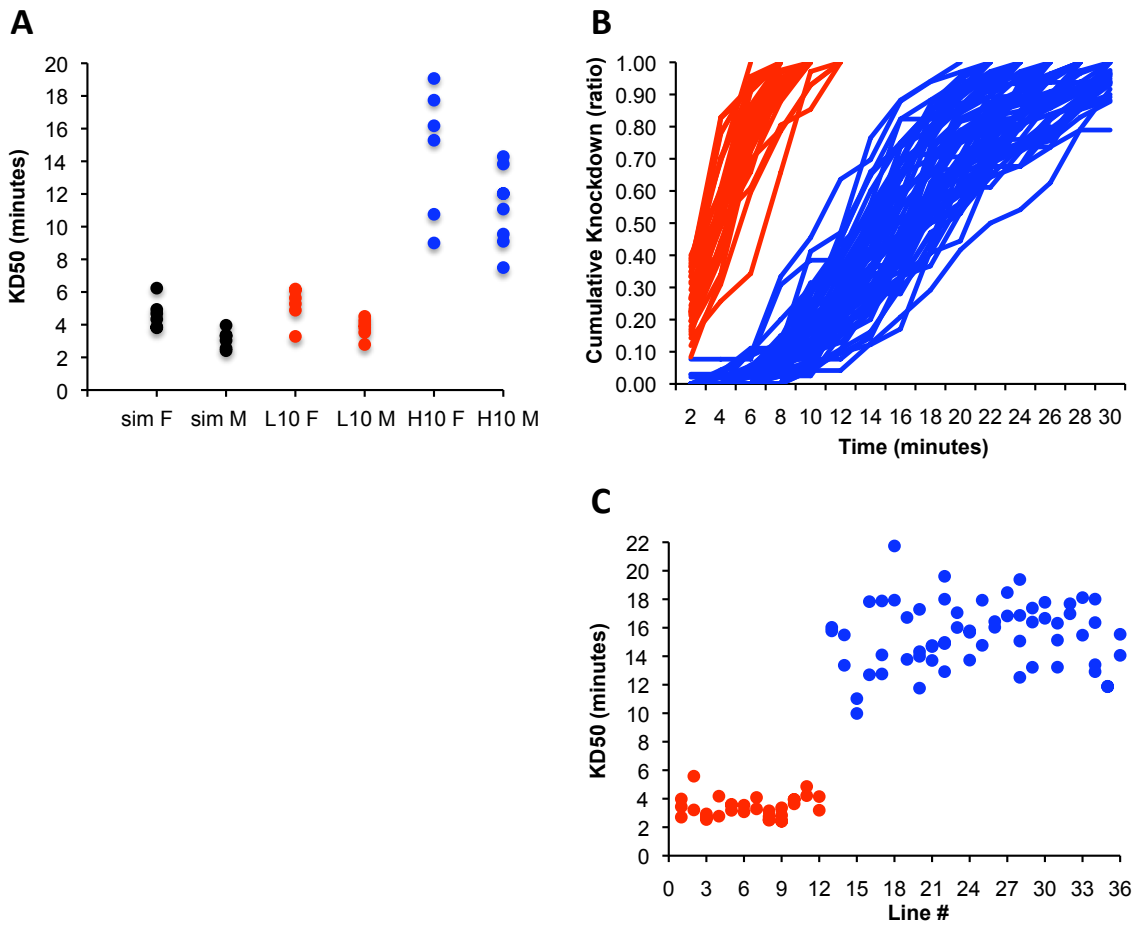

Figure 5.4 OA tolerance for gold standard and 36 unknown lines. (A) KD50 of all replicates for females (labeled 'F') and males ('M'), separately, of *D. simulans* ('*sim*'; black), Low 10 ('L10'; red), and High 10 ('H10'; blue). Plot of the (B) cumulative distribution of knockdown and (C) KD50 for all replicates of the 36 lines of unknown tolerance. The lines/dots clustered into two distinct groups. The blue lines/dots have been classified as high tolerance and the red as low tolerance. None of the lines/dots classified as high tolerance had replicates that clustered with the low tolerance lines and vice versa. Line numbers (1-36) in (C) correspond to the 'Line #' column in Table C.2.

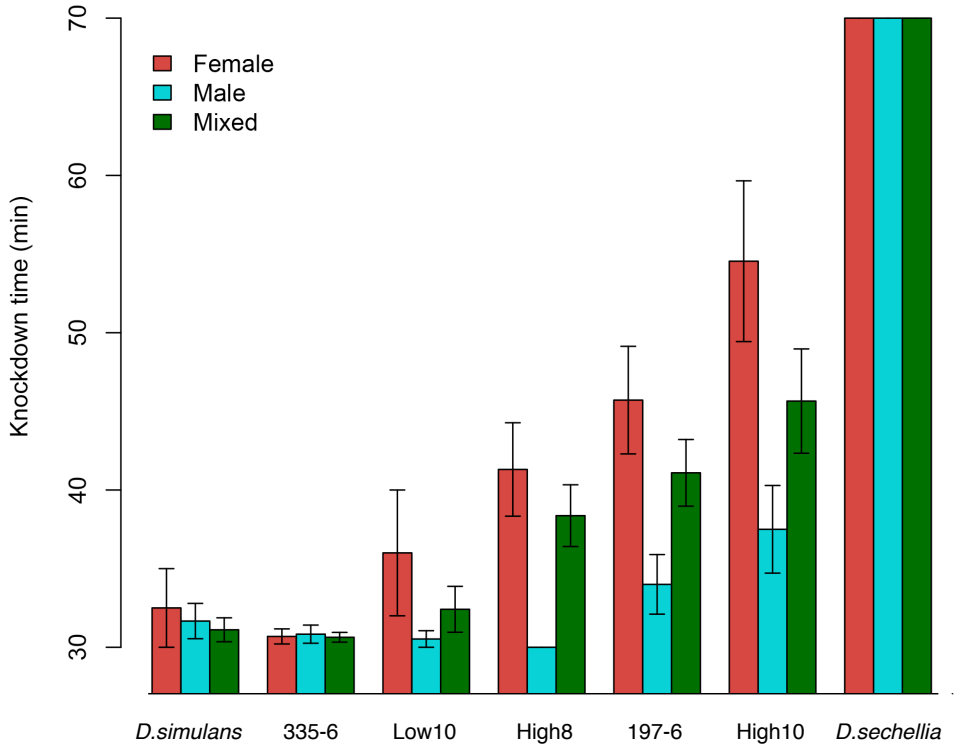

Figure 5.5 *M. citrifolia* tolerance assay. Introgression flies from high and low tolerant lines were assayed for tolerance of ripe *M. citrifolia* fruit. Individual flies were exposed, and knock-down was observed between 30-60 minutes every 10 minutes. Flies used: *D. simulans* (susceptible background of introgression lines); *D. sechellia* (SynA tolerant line); High 8, Low 10, 335-6, 197-6, High 10 (recombinant lines and high and low tolerant lines).

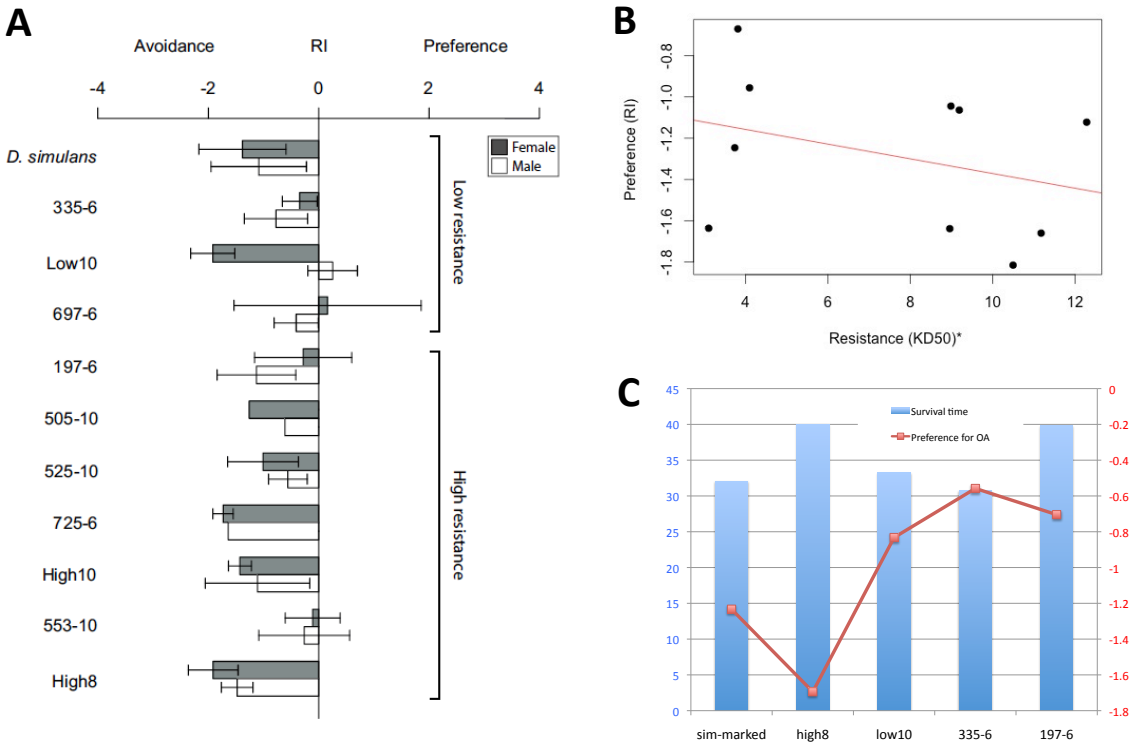

Figure 5.6 Tolerance alleles in the 18-gene region do not affect behavior in heterozygotes. (A) We the measured the settling behavior of several introgression lines, as well as High/Low 10 and *D. simulans* (also called "sim-marked"). The Response Index (RI) describes the degree of aversion (negative values) or preference (positive values). Most lines were not significantly different from controls, except for High 8 males and females, and Low 10 males. Neither (B) OA tolerance nor (C) *M. citrifolia* fruit tolerance were positively correlated with OA preference.

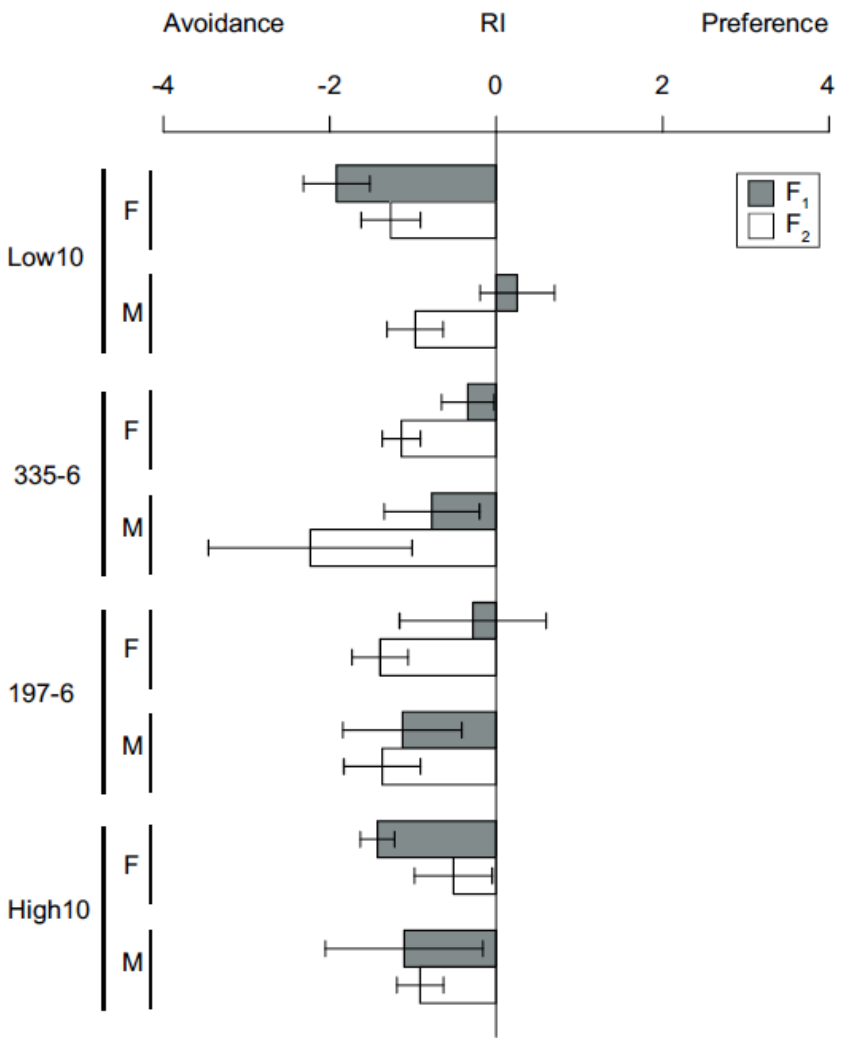

Figure 5.7 *D. sechellia* recessive alleles have no effect on behavior, as shown by the plot of the Response Index (RI). To test the effect of recessive alleles from *D. sechellia* on preference behavior, we selfed F_1 low (Low10 and 335-6) and high (High10 and 197-6) tolerant introgression lines to make F_2 progeny. Flies possessing both homozygous and heterozygous *D. sechellia* introgressions were pooled. Their behavior was not different from purely heterozygous F_1. Females ('F') and males ('M') were assayed for all four lines.

Table 5.1 Welch's t-test *P*-values for comparisons of KD50 between lines and sexes

	Female			*Male*		
Line	*D.sim*	High 10	Low 10	*D.sim*	High 10	Low 10
Female						
D.sim	----	**<0.0001**	0.3324	**0.0068**	----	----
High 10	----	----	**<0.0001**	----	**0.0072**	----
Low 10	----	----	----	----	----	**0.0229**
Male						
D.sim	----	----	----	----	**<0.0001**	0.0635
High 10	----	----	----	----	----	**<0.0001**
Low 10	----	----	----	----	----	----

Significant p-value < 0.05 (in bold)

D.sim, *D. simulans*

KD50, time that 50% of flies are knocked down

Table 5.2 Wilcox rank-sum tests (unpaired) *P*-values* comparing differences in Morinda tolerance between lines and *D. simulans* background

Line	Females	Males	Mixed
High10	**0.035**	0.076	**<<0.001**
High8	0.25	0.063	**0.03**
197-6	0.10	0.76	**0.0019**
Low10	0.86	0.32	0.96
335-6	0.27	0.48	0.16

* p-values are estimates because of rank ties between lines and *D. simulans* background.

VI. CONCLUSIONS

Studies have shown that evolving new adaptive traits, like novel host preference in phytophagous insects, is relatively easy when their genetic basis is relatively simple and genetic variation is abundant (1-5 loci; GAVRILETS and VOSE 2005; MATSUBAYASHI *et al.* 2010). This may be particularly crucial for adaptations that arise in sympatry with their ancestral morph or an incompletely isolated sister species experiencing gene flow. If only a few genes are needed, then adaptive traits could sweep a population quickly before gene flow breaks up adaptive allele combinations. Insects often must evolve multiple traits to accommodate novel plant hosts, including tolerance to plant defensive compounds and preference for these, or other, compounds (FUTUYMA and MORENO 1998). While it is formally possible that many of these traits arose from a single locus, it is much more likely that multiple loci govern each trait. Current work, however, suggests that new host preferences and similar adaptations often have a relatively simple genetic bases (GAVRILETS and VOSE 2005; MATSUBAYASHI et al. 2010). It is unclear whether this pattern reflects biology or methodological bias as the tools used to map these genetic factors bias towards finding few loci of large effect (MACKAY *et al.* 2009; MATSUBAYASHI *et al.* 2010; ROCKMAN 2012)

In my work, I have shown the genetic basis of *Morinda citrifolia* (Morinda) host use is genetically complex in *Drosophila sechellia*. At least nine, and possibly as many as 27, factors underlie the behavioral divergence between *D. sechellia* and *D. simulans*

toward Morinda compounds (Chapters Two and Three). Surprisingly, and in contrast to our expectations, most of these factors do not include genes known to play a role in the chemosensory system (RAMDYA and BENTON 2010). However, I did find that at least one chemosensory gene, *Gustatory receptor 22c*, *Gr22c*, is responsible for *D. melanogaster* aversion to Morinda compounds (Chapter Four). These data suggests a multistep model of host preference evolution: initial loss of aversion followed by gain of preference. Finally, a separate locus on *3*R is important for tolerance of Morinda toxins but is not enough to alter behavior of *Drosophila* (Chapter Five), conflicting with an evolutionary model that host preference and tolerance factors are genetically linked to facilitate more rapid host use evolution (HAWTHORNE and VIA 2001; JAENIKE 1990; JANZ 2011).

While these results confirm that evolving use of a new host is genetically complex, we still haven't yet dissected the full genetic complement of *D. sechellia*'s adaptation to Morinda. I see future work in this system taking two routs: direct genetic tests within *D. sechellia* and further dissection within *D. melanogaster*.

In the short term, we should continue to dissect the genetics of aversion in *D. melanogaster* with the proximal goal of creating a fly that prefers Morinda. First, we need to confirm genetically that *Gr22c* is indeed driving aversion in *D. melanogaster* by creating transgenic rescues of *Gr22c* in a $Gr22c^{-}$ background (see Chapter four). Towards this end, I have already created three transgenic *D. melanogaster* lines with inducible wild type *Gr22c* (*UAS-Gr22c*) using coding regions from *D. simulans* (intact coding region), *D. sechellia* (premature stop codon at ~50% through the ORF), and *D. melanogaster* itself. I have crossed these three lines separately into a *Gr22c* mutant background (that is, $Gr22c^{-}$ from *RAL437*) to produce the genotype $Gr22c^{-}$*; UAS-Gr22c.*

Importantly, the *Gr22c* transgene is silent in these flies, so they will serve as behavior controls; they should exhibit indifference to Morinda compounds. Finally, I crossed these lines to two different *GAL4* driver lines (*hairy-GAL4*, expressing *GAL4* in all neurons; *tubulin-GAL4*, expressing ubiquitously) and maintained the *Gr22c*$^{-}$ background. These flies now have the genotype: *Gr22c*$^{-}$*; UAS-Gr22c/driver-GAL4*, where "driver" refers to *hairy* or *tubulin* promoters. These flies have been created and their genotype confirmed. I anticipate phenotyping all flies, controls and tests. I predict that flies with activated *D. melanogaster/Gr22c* and *D. simulans/Gr22c* will avoid Morinda compounds, whereas activated *D. sechellia/Gr22c* and lines with inactive forms for all insertions will not avoid Morinda compounds.

Ultimately a more precise experiment is to drive the *UAS-Gr22c* with a *Gr22c-GAL4*, which we obtained from Kristin Scott. I have begun the initial crosses to move the *Gr22c*-GAL4 into the *Gr22c*$^{-}$ background. This cross is more complicated because the *Gr22c*-GAL4 is inserted at an unknown location on chromosome 2, which is where *Gr22c* itself resides. Once *Gr22c*-GAL4 is in the right background, I will cross it to the *UAS-Gr22c* and assay behavior as before. Again, I predict that the *D. sechellia* allele will fail to rescue the loss of aversion, but the *D. melanogaster* and *D. simulans* alleles will rescue aversion.

Beyond confirming *Gr22c,* we also need to identify the antennal factors that contribute to Morinda aversion. In Chapter four, I show that antennal ablations remove aversion, and JONES (2007) showed that genetic disruption of basiconic sensilla reduced aversion, as well. While *Obp56e* contributes to aversion (DWORKIN and JONES 2009), this effect is weak, and I suspect that *Ionotropic receptors* will play a role, and *Ir64a* is still a

viable candidate, despite its lack of effect in my hands (see Chapter Three). *Ir64a* appears to form a receptor complex with *Ir8a* (AI *et al.* 2013), which calls for a double knockout experiment for both *Ir64a* and *Ir8a*. DEKKER et al. (DEKKER *et al.* 2006) predict that over-representation of ab3A type sensilla makes *D. sechellia* hypersensitive to the Morinda volatile methyl hexanoate. To see if the ab3A expressed chemoreceptors *Or22a, Or22b,* or *Or85b* are used in *D. melanogaster* to respond behaviorally to methyl hexanoate, we should knockout each gene in *D. melanogaster* and measure behavior at varying concentrations of this compound. Ideally, in parallel we would also overexpress these chemoreceptors, but the genetic reagents to perform such an experiment are not available at this time. If over-representation of one or more of these receptors is responsible for *D. sechellia* hypersensitivity to methyl hexanoate, then overexpression in *D. melanogaster* should promote the same hypersensitivity, whether it be for preference or aversion.

In the long term, we need to perform genetic tests within *D. sechellia* itself, instead of leveraging *D. melanogaster* genetics and building models of inference. This will be difficult given the challenge of transforming *D. sechellia*. Gravid females tend to hold eggs beyond the syncytial blastoderm stage, a hallmark of *Drosophila* development that allows higher incorporation frequency of transgenic material. While we have attempted to instigate egg laying with octanoic acid and Morinda fruit itself prior to vector injections, we have not successfully captured insertions in *D. sechellia*. The first step should be to incorporate a higher efficiency transformation mechanism, like a phiC31 landing site. In the mean time, we can also hybridize *D. sechellia* with transgenic *D. melanogaster* to, for example, conduct gene-reporter assays with GFP to map *D.*

sechellia chemoreceptors. Aversion is dominant in hybrids, and we could also cross *D. sechellia* to *D. melanogaster* chromosomal deficiency lines which have deletions spanning particular candidate genes to see if we can uncover recessive behavioral effects in hybrids.

In my dissertation, I have shown that *D. sechellia's* host specialism is genetically complex. Using both traditional genetic tests and new methods, I have shown repeatedly that the suite of adaptive traits that extant *D. sechellia* need to prefer and accommodate *M. citrifolia* required multiple genetic loci. However, it remains unclear if there is a subset of these loci that are sufficient to shift a fly from not using Morinda to using it. Consistent with this idea, a loss of function mutation in *Gr22c* in *D. sechellia* has a large effect on preference and a single locus on *3*R can quadruple tolerance of OA. It is quite possible that a handful of other loci would be able to move flies to even greater preference and tolerance.

More broadly, my work shows that evolutionary geneticists need to reflect on what the last decade of genetic research has taught us. Some now take it for granted that we have gathered enough direct molecular evidence to build general models of genetic evolution (CARROLL 2008; HOEKSTRA and COYNE 2007; STERN and ORGOGOZO 2008), yet all of the studies supporting one model (coding-region centric) or another (regulatory-region centric) take it for granted that simple genetic changes are the rule and not the exception. Clearly, in the case of *D. sechellia*'s adaptation to Morinda, this is not the case, and it is likely not the case for many other systems (ROCKMAN 2012). Instead, we need to fit new techniques into the largely successful evolutionary theoretical models

built over a century of hard work (ORR 2005). Only then will we begin to construct a new and better theory of genetic adaptation.

REFERENCES

AI, M., S. BLAIS, J. Y. PARK, S. MIN, T. A. NEUBERT *et al.*, 2013 Ionotropic Glutamate Receptors IR64a and IR8a Form a Functional Odorant Receptor Complex In Vivo in Drosophila. J Neurosci **33:** 10741-10749.

CARROLL, S. B., 2008 Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell **134:** 25-36.

DEKKER, T., I. IBBA, K. P. SIJU, M. C. STENSMYR and B. S. HANSSON, 2006 Olfactory shifts parallel superspecialism for toxic fruit in Drosophila melanogaster sibling, D. sechellia. Curr Biol **16:** 101-109.

DWORKIN, I., and C. D. JONES, 2009 Genetic Changes Accompanying the Evolution of Host Specialization in Drosophila sechellia. Genetics **181:** 721-736.

FUTUYMA, D. J., and G. MORENO, 1998 THE EVOLUTION OF ECOLOGICAL SPECIALIZATION. Ann. Rev. Ecol. Syst. **19:** 207-233.

GAVRILETS, S., and A. VOSE, 2005 Dynamic patterns of adaptive radiation. Proc Natl Acad Sci U S A **102:** 18040-18045.

HAWTHORNE, D. J., and S. VIA, 2001 Genetic linkage of ecological specialization and reproductive isolation in pea aphids. Nature **412:** 904-907.

HOEKSTRA, H. E., and J. A. COYNE, 2007 The locus of evolution: evo devo and the genetics of adaptation. Evolution **61:** 995-1016.

JAENIKE, J., 1990 Host Specialization in Phytophagous Insects. Annual Review of Ecology and Systematics **21:** 243-273.

JANZ, N., 2011 Ehrlich and Raven Revisited: Mechanisms Underlying Codiversification of Plants and Enemies. Annual Review of Ecology, Evolution, and Systematics, Vol 42 **42:** 71-89.

JONES, C. D., 2007 Genetic ablation of antennae or basiconic sensilla reduces *Drosophila melanogaster's* ability to perceive odors from the fruit of *Morinda citrifolia*. Drosophila Information Service **90:** 45-49.

MACKAY, T. F., E. A. STONE and J. F. AYROLES, 2009 The genetics of quantitative traits: challenges and prospects. Nat Rev Genet **10:** 565-577.

MATSUBAYASHI, K. W., I. OHSHIMA and P. NOSIL, 2010 Ecological speciation in phytophagous insects. Entomologia Experimentalis Et Applicata **134:** 1-27.

ORR, H. A., 2005 The genetic theory of adaptation: a brief history. Nat Rev Genet **6:** 119-127.

RAMDYA, P., and R. BENTON, 2010 Evolving olfactory systems on the fly. Trends Genet **26:** 307-316.

ROCKMAN, M. V., 2012 The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution **66:** 1-17.

STERN, D. L., and V. ORGOGOZO, 2008 The loci of evolution: how predictable is genetic evolution? Evolution **62:** 2155-2177.

APPENDIX A: SUPPLEMENTARY MATERIAL FOR CHAPTER TWO

Predicting Physical Resolution of Mapping

How much resolution can one expect from PSIseq? The following parameters influence choice of bin size: recombination rate, gene frequency, genome-wide sequence divergence, quality of reference assemblies (or transcriptomes), and level of replication. For example, *D. simulans* and *D. sechellia* are roughly 2% divergent, and given the quality of each reference genome assembly we captured half this divergence in our SNP map (1%). At 1,000 SNPs/bin this provides 100kb resolution. We chose 1,000 SNPs/bin given *Drosophila* gene frequency (~1 gene per 12kb or ~8 genes per 100kb; calculation based on annotation from FlyBase *D. melanogaster* r5.37).

$$S = \frac{B}{(x/2)}$$

Where, B is bin size, or #SNPs/bin; x is average sequence divergence, and S is expected physical size resolution. Thus, for populations 1% divergent with SNP markers capturing half this divergence, bin sizes of 1,000 SNPs/bin will provide 500kb resolution. With an average of one crossover per chromosome per generation, fifteen generations of introgression will isolate 0-2% of the mappable genome. A resolution of 500kb in *Drosophila* easily captures 2% of the genome (~3.6mb).

APPENDIX B: SUPPLEMENTARY MATERIAL FOR CHAPTER FOUR

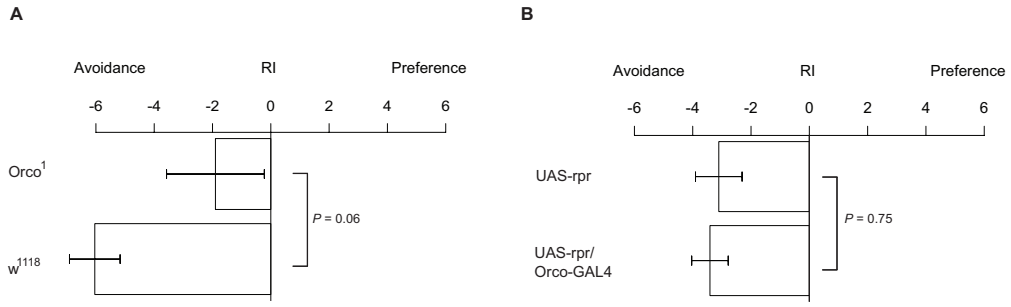

Figure B.1 Confirmations of *Orco* disruptions. (A) *Orco* mutant loses aversion to Benzaldehyde, a chemical avoided through the action of certain *Ors* (*eg, Or43a*). (B) *Orco* was also disrupted using a positive regulator of apoptosis, driven by an *Orco-GAL4* promoter element. *UAS-rpr* is a parental control; *UAS-rpr/Orco-GAL4* experience apoptosis in all *Orco*-expressing cells (unpaired Student's t test).

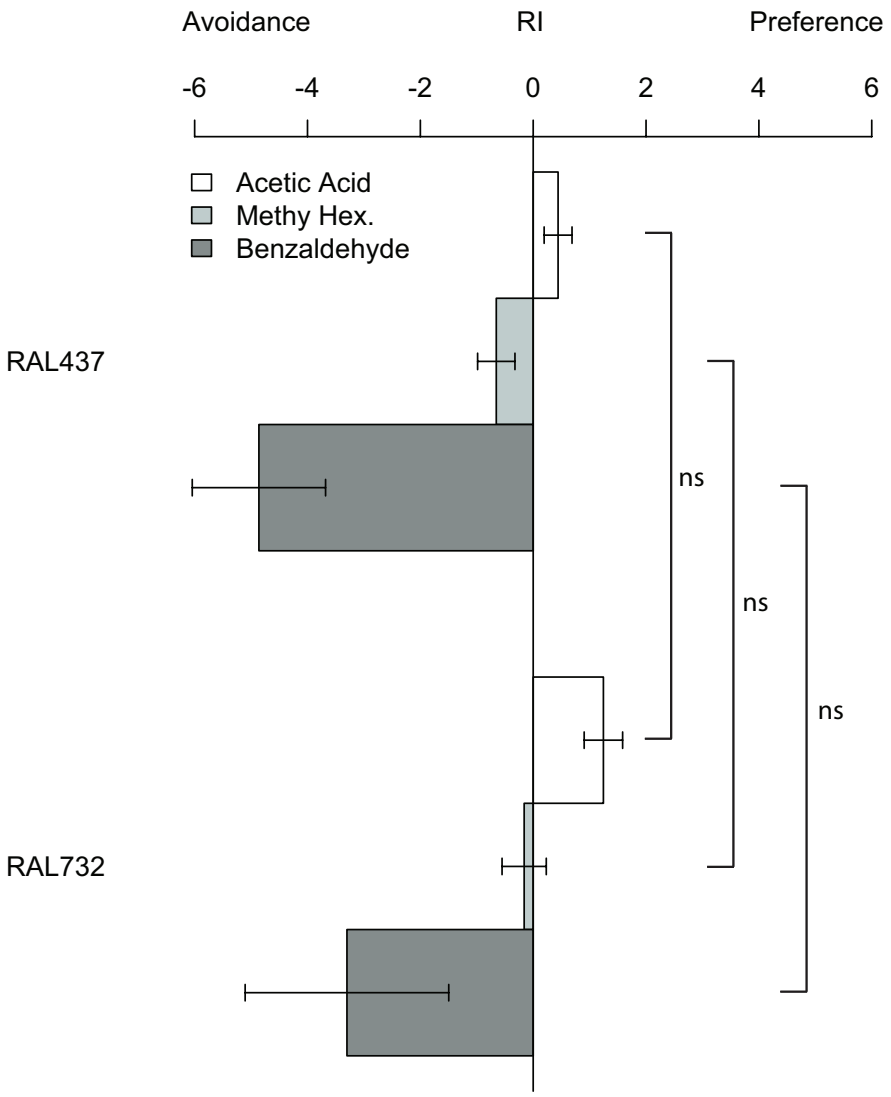

Figure B.2 *RAL437* behaves similarly to other *RAL* lines on alternate chemicals. Unlike *RAL437*, *RAL732* harbors an intact *Gr22c* (see Figure 4.3c). Both lines behave similarly on other chemicals: Acetic Acid (1% v/v), Methy hexanoate (0.1%), and Benzaldehyde (1%) (unpaired Student's t test between lines on the same chemical).

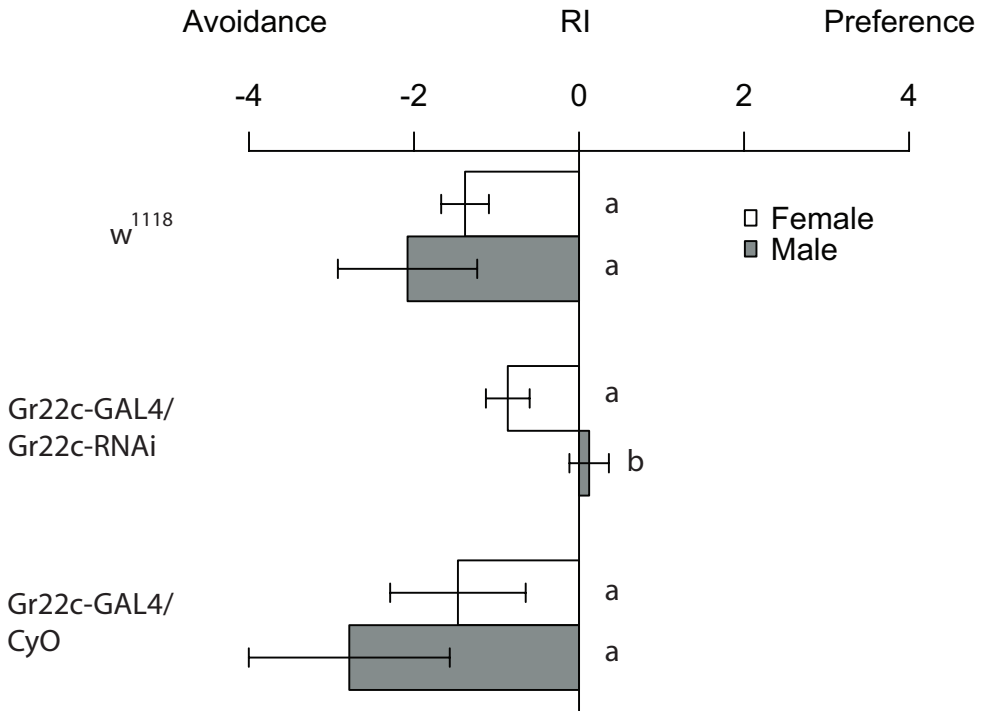

Figure B.3 RNAi on *Gr22c* using a *Gr22c-GAL4* driver. Putative *Gr22c* promoter elements (spanning 8.5kb upstream of *Gr22c* start codon) driving *GAL4* expression, leading to inducible hpRNA expression that initiates RNAi of *Gr22c* mRNA, in only cells expressing endogenous *Gr22c*. Surprisingly, only males exhibited significant loss of aversion to OA:HA, although pooled sexes do show significant loss, as well. *Gr22c-GAL4/CyO* is a sibling control lacking RNAi activity ("a" and "b" denote groups of comparisons not significantly different from each other, unpaired Student's t test).

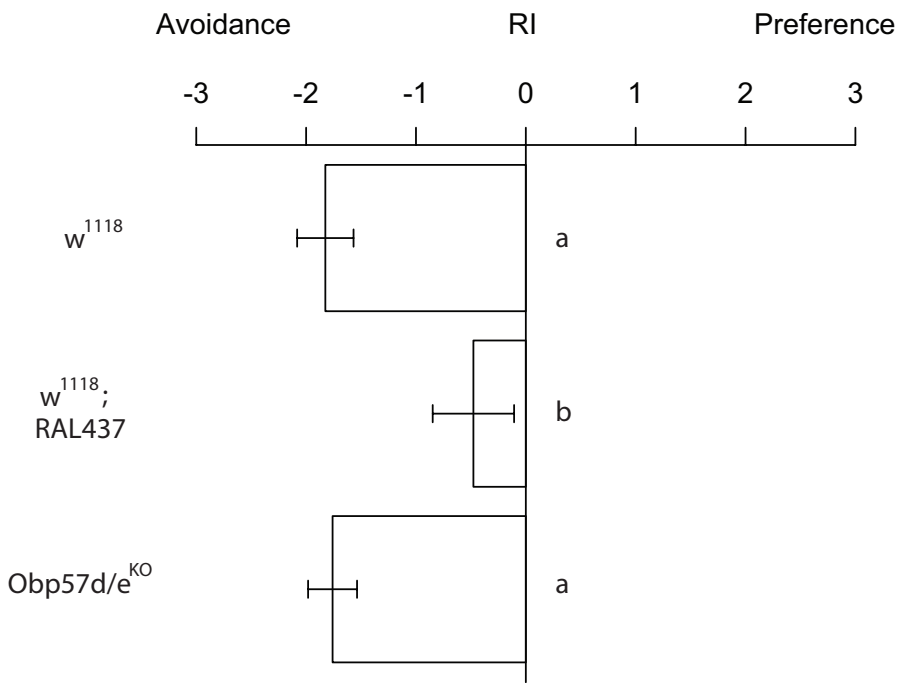

Figure B.4 *Obp57d/e* knockouts maintain aversion to OA:HA in our assay. Flies lacking *Obp57d/e* coding region express no OBP57d/e, but these flies maintain aversion to OA:HA. w^{1118} and w^{1118}*; RAL437* are data from previous figures (4.2a, 4.3c, respectively) (unpaired Student's t test).

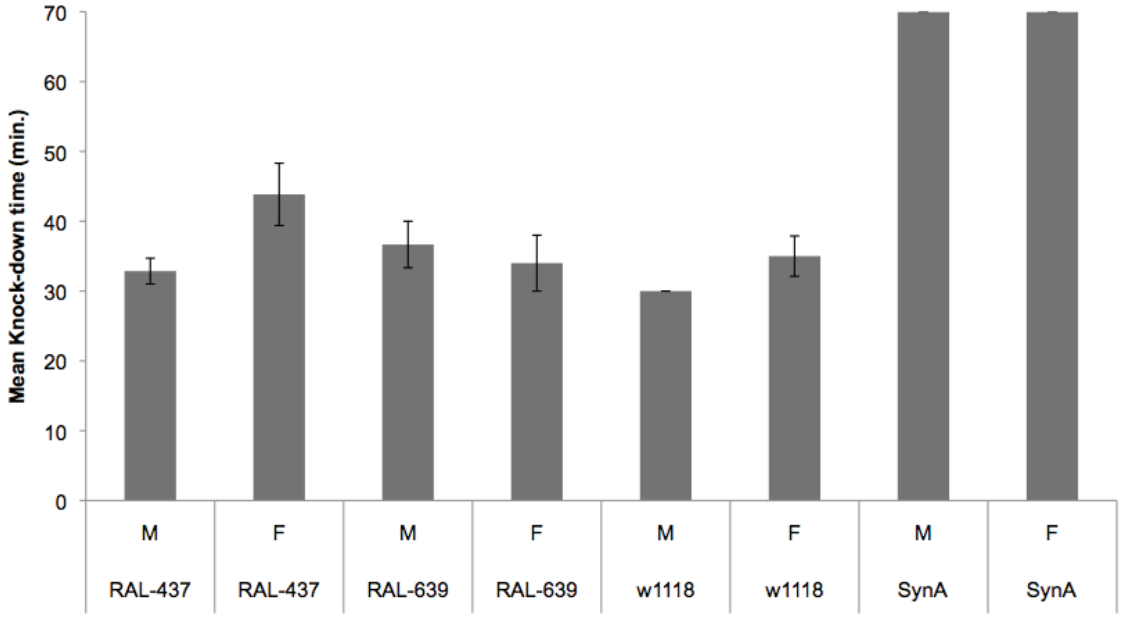

Figure B.5 *Gr22c* loss does not affect Morinda tolerance. Flies lacking *Gr22c* (*RAL437*) were still susceptible to Morinda toxic effect. All *D. melanogaster* lines exhibited similar knockdown times (prelude to death) upon exposure to ripe Morinda. Only *D. sechellia* (*SynA*) was unaffected ("70 min." was the end of observation, and no *D. sechellia* were knocked down).

APPENDIX C: SUPPLEMENTARY MATERIAL FOR CHAPTER FIVE

Supplemental Materials and Methods

Reproducibility of knock-down: To demonstrate the reproducibility of the results from the OA vapor apparatus, we extensively tested the control lines called Low 10 and High 10, along with *D. simulans* and *D. sechellia*. We set the flow rate so most of the recombinant flies were knocked down within 30 minutes, but still allowed adequate resolution to see significant differences between tolerant and non-tolerant recombinant lines. Under these conditions, *D. simulans* flies could only tolerate the acid vapor for a few minutes, while *D. sechellia* flies from three replicate tests were not knocked-down after six consecutive hours of exposure. The "tolerant" recombinant flies had an intermediate tolerance, which was typically less than 30 minutes, but significantly more than that of *D. simulans*.

Effect of density on tolerance: We used linear regression analysis to determine if the density of flies in the test vial had any impact on OA tolerance. There was no significant effect detected for either gender in any of the three lines based on the slope of the linear regression not differing from zero when the number of flies in the vial was between 10 and 60 (High 10 male p-value = 0.2175, female p-value = 0.6740; Low 10 male p-value = 0.6756, female p-value = 0.2000; and *D. simulans* male p-value = 0.5881, female p-value = 0.8801).

Flow rate: Tolerance at two extreme flow rates (0.8 and 2.8 liters/minute) was assayed for High 10 and Low 10 males and females, with *D. simulans* males as the control. The cumulative distribution curves are plotted in Figure C.2. Since only one

replicate was used for each genotype at each flow rate, the KD50s are plotted without error bars. However, the KD50 for each line decreased when the flow rate increased. This effect was especially large in High 10, where the KD50 in females dropped from 23.18 to 14.00 minutes and from 10.96 to 8.60 minutes in males. The effect was fairly large in *D. simulans* males, as well, where the KD50 dropped from 5.46 to 1.82 minutes. The smallest effect was seen in Low 10, where the KD50 in females dropped from 6.19 to 4.21 minutes and from 3.49 to 2.96 minutes in males. Despite the small sample size, the consistency of the results across lines indicated that flow rate, which is a proxy for concentration, had a substantial impact on tolerance. The significance of this conclusion will be explained during the discussion of the *M. citrofolia* assay results.

36 unknown lines and the effect of markers on tolerance: The cumulative knockdown curves of each replicate for all 36 unknown recombinant lines are plotted in Figure 5.4b. The graph revealed obvious clustering of two distinct groups. In fact, none of the lines had replicates split between the high tolerance and low tolerance groups. In other words, a given line was distinctly highly tolerant or lowly tolerant across its replicates. 12 of the 36 lines exhibited low tolerance, while the other 24 were highly tolerant. The mean KD50s for each line are listed in Table C.2 and are plotted in Figure 5.4c. Since most of the lines were assayed with only two replicates, we plotted every replicate for each line instead of using error bars. The highest low tolerance replicate had a KD50 of 5.58, while the lowest high tolerance replicate was 9.98. We did not run an ANOVA to statistically classify each line as low or high tolerance because of the small number of replicates per line, as well as the obvious clustering pattern in the data.

Since the unknown lines had the same phenotypes as OILs 5, 6, 8, 9, and 10, we was able to group them by phenotype to determine if their visible mutations had any effect on tolerance. We separately analyzed the high and low tolerance clusters. It was necessary to confirm that the different phenotypic groups had equal variances, since the samples sizes varied largely. Each pair of phenotype groups was analyzed with the Bartlett Test, which showed no significant differences between the variances within any of the pairs (all p-values > 0.05; data not shown). The ANOVA for the low tolerance cluster indicated that phenotype (i.e. visible mutations used as markers) did have a significant impact on tolerance (p-value = 0.0016). A Tukey HSD Test showed that pairs exhibiting the significant differences were OIL phenotypes 6 and 10 (p-value = 0.0014) and 5 and 10 (p-value = 0.0386). The ANOVA for the high tolerance cluster indicated that phenotype did not play a significant role on tolerance (p-value = 0.0775). However, since the p-value was very close to 0.05, we completed the Tukey analysis anyway and found that OIL phenotype pairs 6/8 (p-value = 0.0754) and 6/10 (p-value = 0.0801) were nearly significant.

The data from the 36 unknown lines showed that lowly tolerant recombinant lines expressing *ebony* (i.e. lines not containing the *D. sechellia* introgression at *ebony*) were significantly less tolerant (phenotype from OILs 5 and 6) than those without *ebony* expression (OILs 9 and 10). The pattern held, but with non-significant p-values, for the highly tolerant lines. COYNE (1984) showed that the *Dsim\jv st e osp p* markers do not affect backcross hybrid size, so there should not have been any inherent difference in size between the OILs 5, 6, 9, and 10. However, *ebony* is known to reduce viability to about 80% of the wild type (LINDSLEY and ZIMM, 1992). The reduced viability probably had an

impact on OA tolerance since the flies were being exposed to a toxic chemical and *ebony* has been shown to be pleiotropically involved in neural function (HOVEMANN et al., 1998).

Ka/Ks: We calculated *Ka/Ks* for genes within the final candidate region (Bergen Center for Computational Science's *Ka/Ks* Calculation tool), where *Ka* is the number of non-synonymous changes in a codon divided by the number of non-synonymous sites, and *Ks* is the same as *Ka*, but for synonymous sites. We compared *D. simulans* and *D. sechellia* coding sequences with *D. melanogaster* as the outgroup.

The Ka/Ks ratios (LI 1993) from both the *D. simulans* and *D. sechellia* lineages, as well as non-synonymous sites and other sequence information, for the 18 genes are summarized in Tables C.4 and C.5. *Osi4* was not included in the following results, due to having an undefined Ka/Ks for *D. simulans*. Using MCBRIDE's (2007) control group means for Ka, Ks, and Ka/Ks in both lineages, there were 4 genes in *D. simulans* with Ka values higher than the control value of 0.002, while there were 7 genes in *D. sechellia* higher than the control of 0.004. There was only one gene with a Ks above the control value of 0.023 in *D. simulans*, whereas there were two genes above the control of 0.030 in *D. sechellia*. 6 genes in *D. simulans* had a Ka/Ks higher than the control value of 0.117, while 10 genes (not including *Osi4*) in *D. sechellia* had a Ka/Ks > 0.145. Thus, 59% of the genes in the region in *D. sechellia* had an enrichment of non-synonymous changes relative to synonymous changes when compared to the control group, while only 35% of the genes in *D. simulans* showed the same. Moreover, only 7 of the 17 genes in *D. simulans* had any non-synonymous changes (41%), while 13 had such changes in *D.*

sechellia (76%). In sum, there is not a strong signature of positive selection at any of these loci.

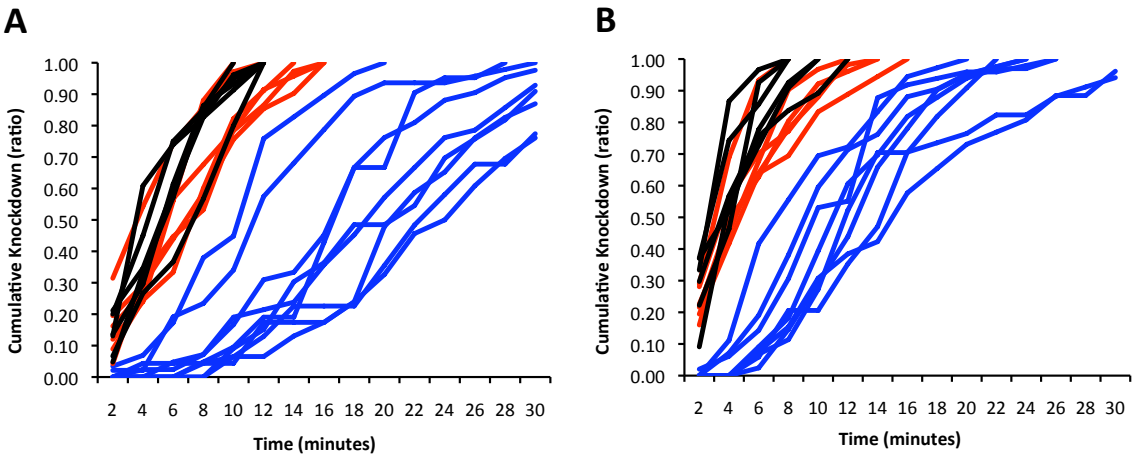

Figure C.1 Plot of cumulative distribution of OA tolerance of all replicates for the comparison between *D. simulans* (black), Low 10 (red), and High 10 (blue), (A) females and (B) males.

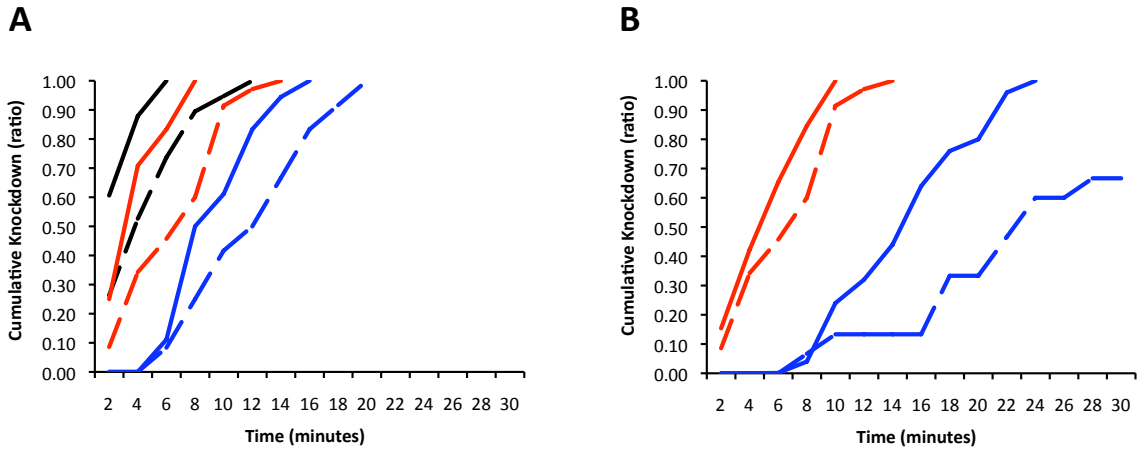

Figure C.2 Comparison of impact on tolerance using a low flow rate (0.8 L/min) and a high flow rate (2.8 L/min) with three different genotypes (High 10, blue; Low 10, red; and *D. simulans*, black) and by sex. The dashed lines represent low flow and the solid lines are high flow. (A) Comparison of males between the three genotypes at the two flow rates and (B) comparison of High 10 and Low 10 females at the two flow rates.

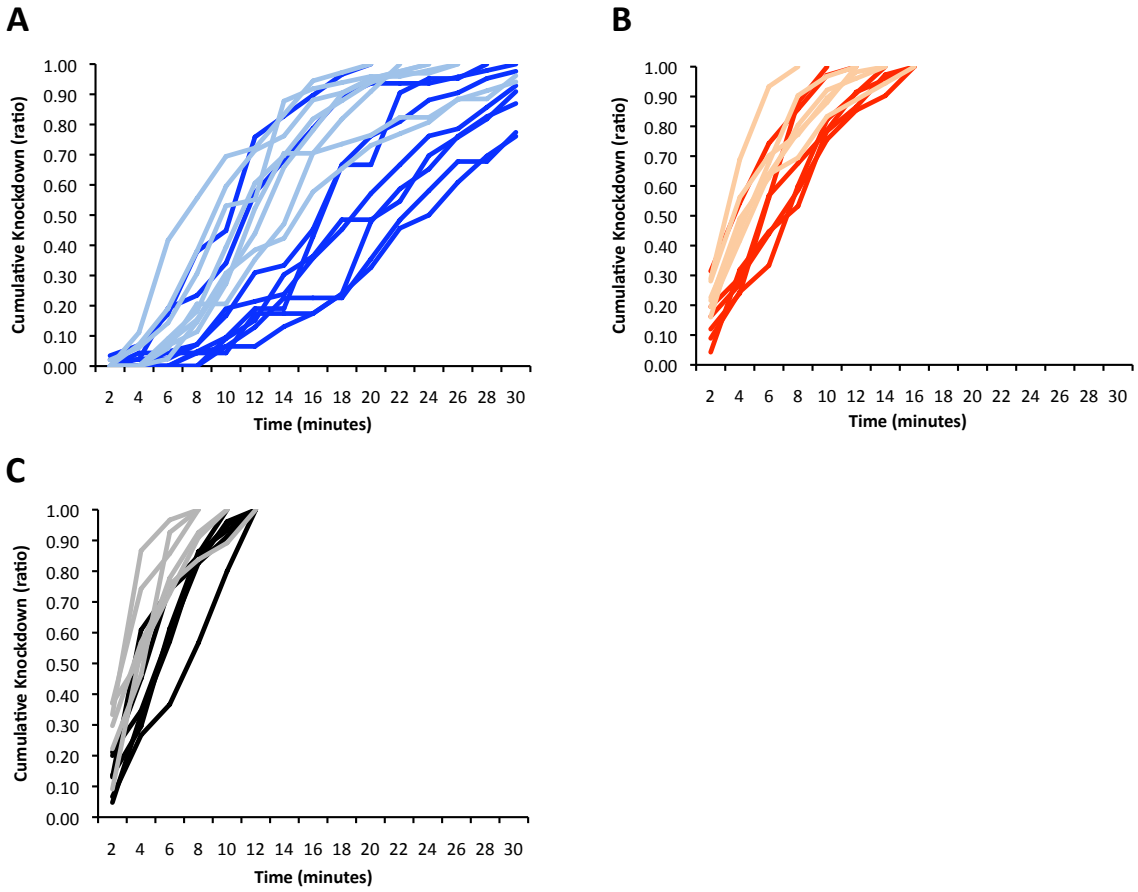

Figure C.3 Plot of cumulative distribution of OA tolerance of all replicates for the comparison between (A) High 10 males (light blue) and females (dark blue), (B) Low 10 males (orange) and females (red), and (C) *D. simulans* males (gray) and females (black).

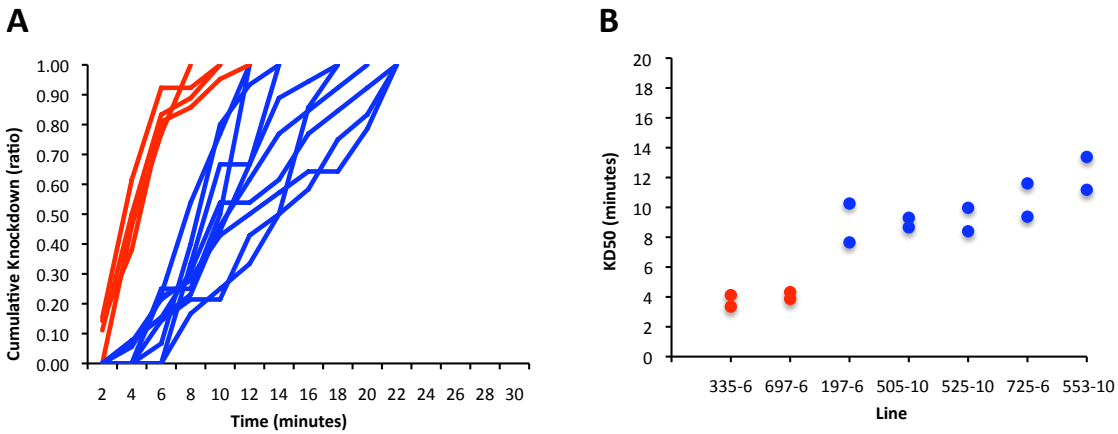

Figure C.4 Cumulative distributions of percent knockdown over time for males of the seven boundary lines. Each of these lines has a boundary marker immediately flanking the 18-gene region (i.e. the OA tolerance locus). The introgressions for 335-6 and 697-6 stop short of including the OA tolerance locus, while the other five lines contain it. (A) Plot of the cumulative distributions of all replicates of the seven boundary lines (two replicates per line). (B) Plot of the KD50 (in minutes) for all replicates of the boundary lines. Blue represents line called as high tolerance, while red is for low tolerance. These lines were used for the *M. citrifolia* tests described in the Materials and Methods.

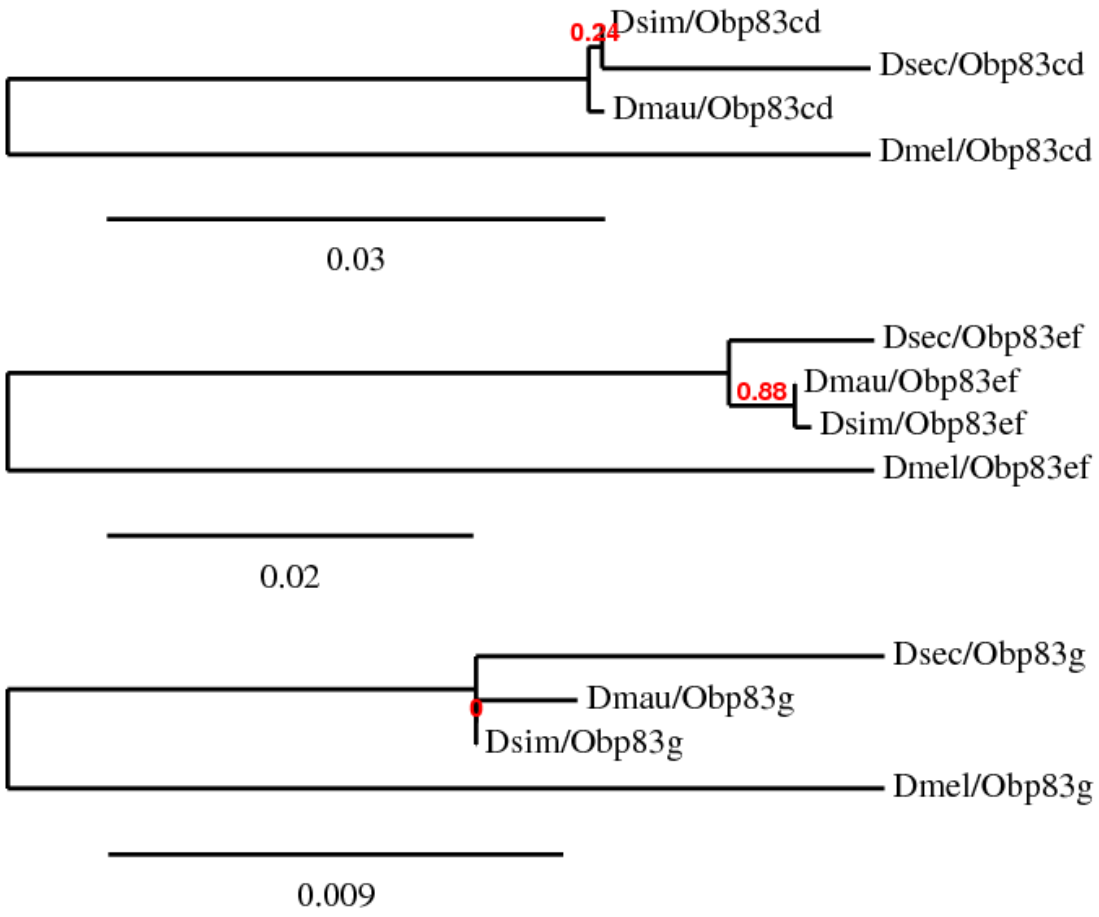

Figure C.5 Maximum likelihood phylogenies of Odorant Binding Proteins within the tolerance region. Although two of three trees are unresolved, the *Obp83ef* trees shows that *D. sechellia* tends to have the unusual allele. Phylogenies were made using http://www.phylogeny.fr/. Bootstrap values are in red.

Table C.1 KD50 values for *D. sim*, High 10, and Low 10 for both sexes

	Females				*Males*			
Line	*replicates*	*# flies*	*mean KD50**	*S.E.*	*replicates*	*# flies*	*mean KD50**	*S.E.*
D.sim	6	190	4.64	0.37	6	216	3.10	0.23
High 10	9	337	17.14	1.64	8	305	11.17	0.83
Low 10	6	230	5.23	0.44	6	270	3.81	0.25

D.sim, *D. simulans*
KD50, time that 50% of flies are knocked down
*KD50 units: minutes

Table C.2 Mean KD50 for males for 36 recombinant lines of unknown tolerance

Line name	Line #	Mean KD50	S.D.	S.E.	# replicates
1-H6	1	3.37	0.64	0.37	3
10-L10	2	4.40	1.67	1.18	2
14-L5	3	2.71	0.20	0.12	3
3-L10	4	3.48	0.99	0.57	2
3-L5	5	3.45	0.23	0.13	3
4-L6	6	3.31	0.33	0.23	2
5-L5	7	3.69	0.57	0.40	2
5-L6	8	2.73	0.30	0.15	4
6-L6	9	2.76	0.44	0.22	4
7-L10	10	3.84	0.18	0.10	3
8-L10	11	4.54	0.45	0.32	2
9-L10	12	3.67	0.68	0.48	2
1-H10	13	15.90	0.17	0.12	2
1-H9	14	14.43	1.51	1.07	2
1-L6	15	15.27	3.99	2.82	2
1-L9	16	15.32	3.63	2.57	2
2-H10	17	14.91	2.66	1.53	3
2-H9	18	19.84	2.69	1.90	2
2-L6	19	15.25	2.08	1.47	2
2-L9	20	14.34	2.27	1.14	4
3-H9	21	14.38	0.59	0.34	3
4-H10	22	16.08	2.68	1.20	5
4-H9	23	16.54	0.74	0.52	2
4-L10	24	15.06	1.16	0.67	3
5-H9	25	16.35	2.24	1.58	2
5-L9	26	16.23	0.28	0.20	2
6-H10	27	17.65	1.16	0.82	2
6-L8	28	15.96	2.90	1.45	4
6-L9	29	15.67	2.17	1.25	3
7-H10	30	17.22	0.79	0.56	2
7-H9	31	14.18	2.18	1.26	3
7-L8	32	17.33	0.50	0.35	2
7-L9	33	16.80	1.86	1.32	2
8-H10	34	15.17	2.43	1.21	4

8-L9	35	11.88	0.02	0.01	2
9-L9	36	14.80	1.04	0.74	2

S.D., standard deviation
S.E., standard error
KD50, time that 50% of flies are knocked down
*KD50 units: minutes

Table C.3 Gene ontology of genes in tolerance region

CG	Symbol	Gene Ontology Terms
CG10287	*Gasp*	Chitin binding and metabolic process; structural component of peritrophiv membrane
CG15582	*Obp83cd*	Odorant binding; sensory perception of chemical stimulus
CG31557	*Obp83ef*	Odorant binding; sensory perception of chemical stimulus
CG31558	*Obp83g*	Odorant binding; sensory perception of chemical stimulus
CG1076	*Vha14-2*	Hydrogen-exporting ATPase activity, phosphorylative mechanism
CG31559	CG31559	Electron carrier activity; protein disulfide oxidoreductase activity
CG1077	CG1077	Protease inhibitor
CG15585	*Osi1*	Unknown function; protein features similar to DUF1676; integral to plasma membrane
CG31562	CG31562	Unknown function; protein features similar to DUF227, CHK kinase-like
CG1147	*NPFR1*	Neuropeptide F, Y, and tachykinin receptor activity; GPCR signaling pathway
CG15589	*Osi24*	Unknown function
CG1148	*Osi2*	Unknown function; protein features similar to DUF1676; integral to plasma membrane
CG1150	*Osi3*	Unknown function; protein features similar to DUF1676; integral to plasma membrane
CG10303	*Osi4*	Unknown function; integral to plasma membrane
CG15590	*Osi5*	Unknown function; integral to plasma membrane
CG1151	*Osi6*	Unknown function; protein features similar to DUF1676
CG1153	*Osi7*	Unknown function; protein features similar to DUF1676; integral to plasma membrane
CG15591	*Osi8*	Unknown function; integral to plasma membrane

Table C.4 Number of non-synonymous and syntenic sites for genes in tolerance region

		Non-synonymous sites		Syntenic sites included in:		
CG	**Symbol**	***D. sim***	***D. sech***	**Ka/Ks**	**Alignment**	**Missing seq**
CG10287	*Gasp*	1	1	708	708	0
CG15582	*Obp83cd*	0	5	729	729	0
CG31557	*Obp83ef*	1	1	738	738	0
CG31558	*Obp83g*	0	2	441	441	0
CG1076	CG1076	0	1	111[a]	390	0
CG31559	CG31559	2	2[b]	1365	1365	0
CG1077	CG1077	11[c]	6	2193	2193	0
CG15585	*Osi1*	1	2[d]	513	513	414
CG31562	CG31562	2	5	765[e]	792	27
CG1147	*NPFR1*	0	1	1458	1458	0
CG15589	CG15589	2	5	1447[f]	1602	0
CG1148	*Osi2*	0	0	1173	1173	0
CG1150	*Osi3*	0	0	867	867	0
CG10303	*Osi4*	3	1	717[g]	1182	0
CG15590	*Osi5*	0	1	609	609	0
CG1151	*Osi6*	0	0	939	939	0
CG1153	*Osi7*	0	0	115	115	752
CG15591	*Osi8*	6	1[i]	309[h]	825	0

[a] Frameshift deletion at 112 in both *D. simulans* (10 bp) and *D. sechellia* (8 bp)
[b] In-frame deletion at 646 in *D. sechellia* (9 bp)
[c] In-frame deletion at 436 in *D. simulans* (6 bp)
[d] In-frame deletion at 154 in *D. sechellia* (3 bp)
[e] Pre-mature stop codon ending at 765 in *D. sechellia*
[f] Pre-mature stop codon ending at 1447 in *D. sechellia*
[g] Frameshift deletion at 718 in *D. simulans* (1 bp)
[h] Pre-mature stop codon ending at 309 in *D. simulans*
[i] One additional non-syn site in *D. sechellia* from *D. melanogaster* after pre-mature stop in *D. simulans*

Table C.5 Ka/Ks calculation for genes in tolerance region for *D. simulans* and *D. sechellia*

		D. simulans			***D. sechellia***		
CG	**Symbol**	**Ka/Ks**	**Ka**	**Ks**	**Ka/Ks**	**Ka**	**Ks**
CG10287	*Gasp*	0.2652	0.0022	0.0083	0.0696	0.0022	0.0316
CG15582	*Obp83cd*	0	0	0.0039	0.2321	0.0076	0.0325
CG31557	*Obp83ef*	0.6249	0.0016	0.0026	0.1073	0.0022	0.0207
CG31558	*Obp83g*	0	0	0	0.4153	0.0054	0.0131
CG1076	CG1076	0	0	0	0.6269	0.0164	0.0262
CG31559	CG31559	0.1304	0.0020	0.0156	0.1880	0.0023	0.0124
CG1077	CG1077	0.6445	0.0067	0.0104	0.3237	0.0042	0.0130
CG15585	*Osi1*	0	0	0	0.1791	0.0049	0.0271
CG31562	CG31562	0.7708	0.0045	0.0058	0.6480	0.0075	0.0116
CG1147	*NPFR1*	0	0	0.0114	0.0495	0.0008	0.0171
CG15589	CG15589	0.5857	0.0019	0.0033	0.3146	0.0036	0.0115
CG1148	*Osi2*	0	0	0.0024	0	0	0.0099
CG1150	*Osi3*	0	0	0.0100	0	0	0.0131
CG10303	*Osi4*	2.7557	0.0028	0.0000	1.1981	0.0046	0.0038
CG15590	*Osi5*	0	0	0.0045	0.1989	0.0026	0.0132
CG1151	*Osi6*	0	0	0.0080	0	0	0.0142
CG1153	*Osi7*	0	0	0	0	0	0
CG15591	*Osi8*	0.3306	0.0274	0.0830	0.2063	0.0050	0.0241
	Mean	0.1972	0.0027	0.0100	0.2094	0.0038	0.0171

Table C.6 Catalog of *D. simulans* and *D. mauritiana* allelic differences at the tolerance locus [a]

D.simulans **locus**	***D. melanogaster*** **homolog**	**Location in Gene (bp)**	**Location on Chromosome (bp)**	***D.simulans*** **allele**	***D. mauritiana*** **allele**
Dsim\GD19843	mel:NPFR1	447	3R:2030917	sim:AGGG	mau:AGG
Dsim\GD19843	mel:NPFR1	550	3R:2031020	sim:AAGCTCAGC	mau:AAGC
Dsim\GD19843	mel:NPFR1	1175	3R:2031645	sim:GGTG	mau:GG
Dsim\GD19843	mel:NPFR1	1405	3R:2031875	sim:GCCCGATACCCGATACCCGAT	mau:GCCCGATACCCGAT
Dsim\GD19843	mel:NPFR1	1759	3R:2032229	sim:CT	mau:C
Dsim\GD19843	mel:NPFR1	2184	3R:2032654	sim:T	mau:C
Dsim\GD19843	mel:NPFR1	2309	3R:2032779	sim:CA	mau:CAA
Dsim\GD19843	mel:NPFR1	2329	3R:2032799	sim:C	mau:A
Dsim\GD19843	mel:NPFR1	2519	3R:2032989	sim:A	mau:G
Dsim\GD19843	mel:NPFR1	2601	3R:2033071	sim:T	mau:C
Dsim\GD19843	mel:NPFR1	2850	3R:2033320	sim:T	mau:C
Dsim\GD19843	mel:NPFR1	3018	3R:2033488	sim:T	mau:C
Dsim\GD19843	mel:NPFR1	3522	3R:2033992	sim:CTTTTTT	mau:CTTTTT
Dsim\GD19843	mel:NPFR1	3998	3R:2034468	sim:GAAAA	mau:GAAAAA
Dsim\GD19843	mel:NPFR1	4077	3R:2034547	sim:TA	mau:T
Dsim\GD19843	mel:NPFR1	4112	3R:2034582	sim:A	mau:AATAGGATTCCAAG
Dsim\GD19843	mel:NPFR1	4292	3R:2034762	sim:A	mau:G
Dsim\GD19843	mel:NPFR1	4334	3R:2034804	sim:G	mau:A
Dsim\GD19843	mel:NPFR1	4463	3R:2034933	sim:C	mau:T,G
Dsim\GD19843	mel:NPFR1	4475	3R:2034945	sim:T	mau:A
Dsim\GD19843	mel:NPFR1	4562	3R:2035032	sim:T	mau:C
Dsim\GD19843	mel:NPFR1	4883	3R:2035353	sim:T	mau:C
Dsim\GD19849	mel:Osi8	89	3R:2094767	sim:ACCCC	mau:ACCC
Dsim\GD19849	mel:Osi8	275	3R:2094953	sim:GT	mau:G
Dsim\GD19849	mel:Osi8	513	3R:2095191	sim:A	mau:AGG
Dsim\GD19849	mel:Osi8	743	3R:2095421	sim:CTTTTTTTTT	mau:CTTTTTTTT

Dsim\GD19849	mel:Osi8	868	3R:2095546	sim:T	mau:C
Dsim\Obp83ef	mel:Obp83ef	141	3R:1963207	sim:T	mau:A
Dsim\Obp83ef	mel:Obp83ef	760	3R:1963826	sim:TTATATTTTTTATATA	mau:TTATA
Dsim\Obp83ef	mel:Obp83ef	821	3R:1963887	sim:TGG	mau:TGGGG
Dsim\Obp83ef	mel:Obp83ef	910	3R:1963976	sim:G	mau:C
Dsim\GD19841	mel:Osi1	401	3R:2022030	sim:C	mau:T
Dsim\GD19841	mel:Osi1	521	3R:2022150	sim:G	mau:A
Dsim\GD19841	mel:Osi1	581	3R:2022210	sim:A	mau:C
Dsim\GD19841	mel:Osi1	691	3R:2022320	sim:T	mau:C
Dsim\GD19846	mel:Osi3	106	3R:2057905	sim:A	mau:G
Dsim\GD19846	mel:Osi3	190	3R:2057989	sim:C	mau:T
Dsim\GD19846	mel:Osi3	205	3R:2058004	sim:CTTTTTTTTT	mau:CTTTTTTTT,CTTTTTTT
Dsim\GD19846	mel:Osi3	322	3R:2058121	sim:A	mau:T
Dsim\GD19846	mel:Osi3	405	3R:2058204	sim:GTATTTAT	mau:GTATTTATTTAT
Dsim\GD19846	mel:Osi3	1121	3R:2058920	sim:CAAAA	mau:CAAAAA
Dsim\GD19846	mel:Osi3	1244	3R:2059043	sim:G	mau:A
Dsim\GD19846	mel:Osi3	1380	3R:2059179	sim:A	mau:G
Dsim\GD19846	mel:Osi3	1439	3R:2059238	sim:C	mau:T
Dsim\GD19846	mel:Osi3	1490	3R:2059289	sim:A	mau:G
Dsim\GD19846	mel:Osi3	1589	3R:2059388	sim:C	mau:T
Dsim\GD19848	mel:Osi6	177	3R:2077595	sim:T	mau:C
Dsim\GD19848	mel:Osi6	441	3R:2077859	sim:C	mau:T
Dsim\GD19848	mel:Osi6	561	3R:2077979	sim:G	mau:A
Dsim\GD19848	mel:Osi6	812	3R:2078230	sim:T	mau:C
Dsim\GD19848	mel:Osi6	903	3R:2078321	sim:C	mau:T,G
Dsim\GD19848	mel:Osi6	1353	3R:2078771	sim:C	mau:T
Dsim\GD19557	mel:CG31559	172	3R:2006506	sim:C	mau:G
Dsim\GD19557	mel:CG31559	628	3R:2006962	sim:G	mau:A
Dsim\GD19	mel:CG31559	1007	3R:2007341	sim:G	mau:A

557					
Dsim\GD19557	mel:CG31559	2077	3R:2008411	sim:TC	mau:TCATTGC
Dsim\GD19557	mel:CG31559	2183	3R:2008517	sim:A	mau:T
Dsim\GD19557	mel:CG31559	2299	3R:2008633	sim:CAAAAAA	mau:CAAAAA
Dsim\GD19557	mel:CG31559	2759	3R:2009093	sim:GAGCAACAGCAGCAGCAACAGCAGCAGCA	mau:GAGCAACAGCAGCAGCA
Dsim\GD19557	mel:CG31559	2952	3R:2009286	sim:T	mau:C
Dsim\GD19557	mel:CG31559	4240	3R:2010574	sim:GC	mau:G
Dsim\GD19557	mel:CG31559	4440	3R:2010774	sim:AT	mau:A
Dsim\GD19557	mel:CG31559	4995	3R:2011329	sim:A	mau:C
Dsim\GD19557	mel:CG31559	5053	3R:2011387	sim:ATAT	mau:ATATTAT
Dsim\GD19557	mel:CG31559	5344	3R:2011678	sim:T	mau:A
Dsim\GD19557	mel:CG31559	5486	3R:2011820	sim:AATTTATAT	mau:AAT
Dsim\GD19557	mel:CG31559	5754	3R:2012088	sim:A	mau:G,C
Dsim\GD19557	mel:CG31559	5812	3R:2012146	sim:T	mau:C
Dsim\GD19557	mel:CG31559	6063	3R:2012397	sim:C	mau:G
Dsim\GD19557	mel:CG31559	6112	3R:2012446	sim:T	mau:C
Dsim\GD19557	mel:CG31559	6214	3R:2012548	sim:CGGG	mau:CGG
Dsim\GD19557	mel:CG31559	6276	3R:2012610	sim:C	mau:T
Dsim\GD19557	mel:CG31559	6474	3R:2012808	sim:T	mau:G
Dsim\GD19561	mel:Gasp	252	3R:1946656	sim:T	mau:A
Dsim\GD19561	mel:Gasp	440	3R:1946844	sim:C	mau:T
Dsim\GD19561	mel:Gasp	706	3R:1947110	sim:G	mau:A
Dsim\GD19561	mel:Gasp	1086	3R:1947490	sim:C	mau:G
Dsim\GD19561	mel:Gasp	1149	3R:1947553	sim:TA	mau:TAA,TGGTGCATCCCTAA
Dsim\GD19561	mel:Gasp	1316	3R:1947720	sim:A	mau:G
Dsim\GD19561	mel:Gasp	1493	3R:1947897	sim:T	mau:C
Dsim\GD19561	mel:Gasp	1519	3R:1947923	sim:AGGTGTAGGGAATGGG	mau:AGG
Dsim\GD19561	mel:Gasp	1943	3R:1948347	sim:T	mau:A
Dsim\GD19	mel:Gasp	2223	3R:1948627	sim:ATGTGTGTGTG	mau:ATGTGTGTGTG

561				TGTGTG	TGTGTGTG
Dsim\GD19561	mel:Gasp	2766	3R:1949170	sim:A	mau:G
Dsim\GD19561	mel:Gasp	3083	3R:1949487	sim:A	mau:T
Dsim\GD19561	mel:Gasp	3307	3R:1949711	sim:A	mau:G
Dsim\GD19561	mel:Gasp	3494	3R:1949898	sim:CAA	mau:CAAA
Dsim\GD19561	mel:Gasp	3595	3R:1949999	sim:GCC	mau:GC
Dsim\GD19561	mel:Gasp	3998	3R:1950402	sim:A	mau:G
Dsim\GD19561	mel:Gasp	4514	3R:1950918	sim:GC	mau:GCC
Dsim\GD19561	mel:Gasp	4667	3R:1951071	sim:AGCTCCTGCTCCTGCTCCTGCT	mau:AGCTCCTGCTCCTGCT
Dsim\GD19561	mel:Gasp	4986	3R:1951390	sim:T	mau:A
Dsim\GD19561	mel:Gasp	5312	3R:1951716	sim:C	mau:A
Dsim\GD19561	mel:Gasp	5885	3R:1952289	sim:T	mau:C,G
Dsim\GD19561	mel:Gasp	5905	3R:1952309	sim:TAAAAAAAA	mau:TAAAAAAAAA
Dsim\GD19561	mel:Gasp	6071	3R:1952475	sim:GCCCCCCCC	mau:GCCCCCCC
Dsim\GD19561	mel:Gasp	6652	3R:1953056	sim:C	mau:T
Dsim\GD19561	mel:Gasp	6715	3R:1953119	sim:T	mau:C
Dsim\GD19561	mel:Gasp	6793	3R:1953197	sim:G	mau:T
Dsim\GD19561	mel:Gasp	7329	3R:1953733	sim:TTTTCGTTTCGTTTC	mau:TTTTCGTTTCGTTTCGTTTC
Dsim\Obp83cd	mel:Obp83cd	65	3R:1961592	sim:C	mau:A
Dsim\Obp83cd	mel:Obp83cd	389	3R:1961916	sim:T	mau:C
Dsim\Obp83cd	mel:Obp83cd	792	3R:1962319	sim:A	mau:T
Dsim\Obp83cd	mel:Obp83cd	867	3R:1962394	sim:AA	mau:AAAATGA
Dsim\GD19844	mel:Osi24	210	3R:2046272	sim:G	mau:A
Dsim\GD19844	mel:Osi24	264	3R:2046326	sim:T	mau:C
Dsim\GD19844	mel:Osi24	849	3R:2046911	sim:AATGTTTAT	mau:AAT
Dsim\GD19844	mel:Osi24	2001	3R:2048063	sim:G	mau:C
Dsim\GD19844	mel:Osi24	2317	3R:2048379	sim:G	mau:A
Dsim\GD19844	mel:Osi24	2399	3R:2048461	sim:C	mau:T
Dsim\GD19838	mel:-	432	3R:1989928	sim:CTTTTTTTTTT	mau:CTTTTTTTT,CTTTTTTTTTTT

Dsim\GD19838	mel:-	641	3R:1990137	sim:T	mau:TAAC
Dsim\GD19838	mel:-	2101	3R:1991597	sim:A	mau:G
Dsim\GD19838	mel:-	2127	3R:1991623	sim:TCCCC	mau:TCCC
Dsim\GD19838	mel:-	2340	3R:1991836	sim:G	mau:T
Dsim\GD19838	mel:-	2393	3R:1991889	sim:C	mau:A
Dsim\GD19838	mel:-	2767	3R:1992263	sim:A	mau:T
Dsim\GD19838	mel:-	2774	3R:1992270	sim:A	mau:G
Dsim\GD19838	mel:-	2839	3R:1992335	sim:C	mau:A
Dsim\GD19554	mel:Osi4	77	3R:2060376	sim:C	mau:T
Dsim\GD19554	mel:Osi4	316	3R:2060615	sim:TTCTTCACTTATAAAATATC	mau:T
Dsim\GD19554	mel:Osi4	645	3R:2060944	sim:A	mau:G
Dsim\GD19554	mel:Osi4	765	3R:2061064	sim:CTT	mau:CT
Dsim\GD19554	mel:Osi4	954	3R:2061253	sim:TCCCCCC	mau:TCCCCC
Dsim\GD19554	mel:Osi4	1072	3R:2061371	sim:A	mau:C
Dsim\GD19554	mel:Osi4	1173	3R:2061472	sim:A	mau:G
Dsim\GD19554	mel:Osi4	1297	3R:2061596	sim:C	mau:T
Dsim\GD19554	mel:Osi4	1399	3R:2061698	sim:C	mau:G
Dsim\GD19554	mel:Osi4	1585	3R:2061884	sim:A	mau:AAAC
Dsim\GD19554	mel:Osi4	1687	3R:2061986	sim:G	mau:GT
Dsim\GD19554	mel:Osi4	2069	3R:2062368	sim:GC	mau:GCGCTGCTGCC
Dsim\GD19554	mel:Osi4	2368	3R:2062667	sim:C	mau:T
Dsim\GD19554	mel:Osi4	2840	3R:2063139	sim:A	mau:T
Dsim\GD19554	mel:Osi4	2855	3R:2063154	sim:A	mau:G
Dsim\GD19554	mel:Osi4	2895	3R:2063194	sim:C	mau:T
Dsim\GD19842	mel:CG33301,CG16898	333	3R:2025889	sim:C	mau:T
Dsim\GD19842	mel:CG33301,CG16898	861	3R:2026417	sim:G	mau:A
Dsim\GD19842	mel:CG33301,CG16898	876	3R:2026432	sim:C	mau:T
Dsim\GD19842	mel:CG33301,CG16898	906	3R:2026462	sim:GGAATTGGAG	mau:GG
Dsim\GD19	mel:CG33301,CG	1180	3R:2026736	sim:G	mau:A

842	16898				
Dsim\GD19842	mel:CG33301,CG16898	1232	3R:2026788	sim:C	mau:G
Dsim\GD19558	mel:-	56	3R:1969285	sim:TGTCCT	mau:TGTCCTTGCCGTCCT
Dsim\GD19558	mel:-	96	3R:1969325	sim:T	mau:A
Dsim\GD19558	mel:-	107	3R:1969336	sim:A	mau:G
Dsim\GD19558	mel:-	420	3R:1969649	sim:TAAAA	mau:TAAA
Dsim\GD19558	mel:-	436	3R:1969665	sim:ACTGCTAGT	mau:ACTGCTAGTTCTGCTAGT
Dsim\GD19555	mel:-	74	3R:2042841	sim:GTGT	mau:GTGTTTGT
Dsim\GD19556	mel:-	64	3R:2020986	sim:C	mau:T
Dsim\GD19847	mel:Osi5	40	3R:2069929	sim:A	mau:G
Dsim\GD19847	mel:Osi5	387	3R:2070276	sim:G	mau:A
Dsim\GD19847	mel:Osi5	609	3R:2070498	sim:C	mau:T
Dsim\GD19847	mel:Osi5	661	3R:2070550	sim:TGACATCGACATC	mau:TGACATC
Dsim\GD19847	mel:Osi5	979	3R:2070868	sim:C	mau:T
Dsim\Obp83g	mel:Obp83g	103	3R:1965138	sim:G	mau:T
Dsim\Obp83g	mel:Obp83g	410	3R:1965445	sim:A	mau:G
Dsim\GD19837	mel:Vha14-2	402	3R:1976141	sim:G	mau:A
Dsim\GD19840	mel:CG1077	207	3R:2015194	sim:C	mau:T
Dsim\GD19840	mel:CG1077	246	3R:2015233	sim:T	mau:G
Dsim\GD19840	mel:CG1077	508	3R:2015495	sim:C	mau:G
Dsim\GD19840	mel:CG1077	565	3R:2015552	sim:T	mau:C
Dsim\GD19840	mel:CG1077	602	3R:2015589	sim:C	mau:A
Dsim\GD19840	mel:CG1077	610	3R:2015597	sim:G	mau:A
Dsim\GD19840	mel:CG1077	1098	3R:2016085	sim:T	mau:G
Dsim\GD19840	mel:CG1077	1239	3R:2016226	sim:C	mau:T
Dsim\GD19840	mel:CG1077	1347	3R:2016334	sim:T	mau:C
Dsim\GD19840	mel:CG1077	1706	3R:2016693	sim:A	mau:T
Dsim\GD19840	mel:CG1077	2368	3R:2017355	sim:A	mau:G
Dsim\GD19840	mel:CG1077	2495	3R:2017482	sim:T	mau:C

Dsim\GD19840	mel:CG1077	2506	3R:2017493	sim:G	mau:A
Dsim\GD19840	mel:CG1077	2716	3R:2017703	sim:G	mau:C
Dsim\GD19840	mel:CG1077	2832	3R:2017819	sim:A	mau:G
Dsim\GD19839	mel:-	74	3R:2009286	sim:T	mau:C
Dsim\GD19845	mel:Osi2	267	3R:2053474	sim:C	mau:T
Dsim\GD19845	mel:Osi2	483	3R:2053690	sim:C	mau:G
Dsim\GD19845	mel:Osi2	1091	3R:2054298	sim:G	mau:C
Dsim\GD19845	mel:Osi2	1581	3R:2054788	sim:C	mau:A

[a] Because a published reference of *D. mauritiana* is not available, we produced these variant calls using public and other (unpublished) short read sequencing.

REFERENCES

COYNE, J. A., 1984 Genetic basis of male sterility in hybrids between two closely related species of *Drosophila*. Proc. Natl. Acad. Sci. USA **51:** 4444–4447.

HOVEMANN B. T., R. P. RYSECK, U. WALLDORF, K. F. STÖRTKUHL, I. D. DIETZEL, et al., 1998 The *Drosophila ebony* gene is closely related to microbial peptide synthetases and shows specific cuticle and nervous system expression. Gene **221:** 1-9.

LINDSLEY, D. L., AND G. G. ZIMM, 1992 *The Genome of Drosophila melanogaster*. Academic Press, Inc., San Diego.

MCBRIDE, C. S., 2007 Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. Proc. Natl. Acad. Sci. USA **104:** 4996–5001.

THIBAULT, S. T., M. A. SINGER, W. Y. MIYAZAKI, B. MILASH, N.A. DOMPE, et al., 2004 A complementary transposon tool kit for *Drosophila melanogaster* using P and *piggyBac*. Nat. Genet. **36:** 283-287.