Exploring the transcriptomics of polyphagy in plant-eating insects

by

Nicholas Christodoulides

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
August 5, 2017

Keywords: gene expression, plasticity, host use, generalist

Approved by

Nate B. Hardy, Chair, Assistant Professor, Department of Entomology and Plant Pathology
Tonia S. Schwartz, Assistant Professor, Department of Biological Sciences
Alana L. Jacobson, Assistant Professor, Department of Entomology and Plant Pathology

Abstract

The aim of my thesis research was to gain a better understanding of the genetics underlying insect polyphagy. My thesis consisted of three objectives, each exploring a different aspect of plant-feeding insect genetics. For my first objective, I assembled and analyzed transcriptome data to study how gene expression varies across hosts of an extremely generalist insect, the lobate lac scale (*Paratachardina pseudolobata*). For my second objective, I examined published transcriptomes of 42 species of plant-feeding insects to test the simple idea that more generalist insects have more biochemical tools for herbivory, and express more herbivory genes. As a final part of my thesis research, I lay the foundation for an experiment to examine the role of phenotypic plasticity in the evolution of host use in aphid species.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Overview

Plant-feeding insects comprise about a quarter of the described terrestrial species on earth [1]. The extreme biodiversity of herbivorous insects is commonly attributed to coevolution with their host plants, which they depend on for food, shelter, and mating sites [2]. About 90% of plant-feeding insect species are specialists and only use one type of host plant, or a few similar types. Some species however, are generalists with diets ranging up to hundreds of different plant species [3]. Although generalists play important roles in both human and natural environments, little is known about the genetics underlying generalism and its evolution in insect species. What genetic mechanisms might enable polyphagy in generalist plant-feeding insects? Why are some generalist insects able to consume significantly more hosts than others? Are there heritable gene expression patterns that are associated with generalism? With the advent of next-generation sequencing technologies and the number of available insect gene sequences steadily increasing, we are now in a position to potentially answer these questions and test different hypothesis about the genetic architecture of generalism.

My MS thesis was motivated by the need for general knowledge of the genetics of niche generalism. I was interested in 1) obtaining and analyzing plant-eating insect transcriptome data to find specific biological processes and molecular functions that are affected by the host environment, 2) testing the idea that broad diets depend on the proliferation of herbivory genes, and 3) designing an experimental evolution study to assess the role of phenotypic plasticity in host-use adaptation.

Through my work on this research, I have gained insights into the genetics of generalist insect herbivory, and have become adept at differential gene expression analysis. I have also learned how to use comparative statistical methods to test hypotheses about host-use adaptation, as well as design experimental evolution studies of generalist gene expression plasticity.

Theoretical Background

About a third of metazoan species are plant-feeding insects. Most are specialists. Previous studies on specialist insect diversity have linked their high diversification rates with the adoption of phytophagy [4]. Despite being the minority of herbivorous insect species, generalists consume a disproportionate amount of plant material relative to what their species richness would suggest. Generalist insects are also important because they are enriched for pest and invasive species that can damage and transmit plant diseases to hosts that are important to natural ecosystems and human agricultural activities [5]. In addition, they serve as a good model for understanding how broad ecological niches can develop in a group of organisms with typically very narrow ones [6]. We are just beginning to understand the molecular basis of insect-host interactions.

Host Plant Defenses against Herbivores

In order to understand the broad diets of some plant-eating insect species, we need to know what limits diet breadth. Over their history of antagonism with herbivorous insects, plants have evolved numerous defensive mechanisms to protect themselves from being eaten [7]. Some plant defenses are constitutive, that is, they are always “on” and continuously deter feeding by insects, regardless of whether or not feeding is actually happening. Constitutive defenses may be gross structural barriers, such as bark or waxy cuticles. They may also take the form of secondary plant

metabolites such as terpenes, cardenolides, and alkaloids [8]. Because such chemicals may be costly to constantly produce, plant defenses may also be inducible, that is activated only in response to stresses, such as insect feeding. Inducible defense mechanisms are expressed as a part of plant innate immunity, and are triggered in response to foreign proteins and pathogens. In response, some insects are able to secrete effector molecules into their host plants to disable systemic resistance responses such as the synthesis of antimicrobial secondary metabolites. These inducible defenses tend to be more efficient for plants that only occasionally encounter stress [9].

Detoxification Genes

In order for plants to be a part of a generalist insects' host range, the insect must be able to circumvent their defenses, both constitutive and inducible. Early research suggested that plant-feeding insects use detoxification genes as their main tools in circumventing plant defense chemicals [10]. If this is true, then the evolution of detoxification genes could play an important role in the evolution of plant-eating insect niches. Moreover, the same enzymes are thought to be critical for the evolution of resistance to man-made insecticides [11]. Thus, an understanding of the evolutionary genetics of host-use could help us to better manage populations of agricultural pests.

What genes do plant-eating insects use to detoxify plant defensive chemicals? Cytochrome P450 monooxygenases, or P450s, are the best-studied class of detoxification genes in insects, and use oxidation in association with reducing NADPH to detoxify furanocoumarins, alkaloids, and other plant chemicals and man-made insecticides [12, 13]. Other insect detoxification enzymes include carboxylesterases, glutathione peroxidases, glutathione S-

transferases, UDP-glycosyltransferases, and other proteases [14]. Glutathione S-transferases (GSTs), glutathione peroxidases, and UDP-glycosyltransferases make toxic compounds more hydrophilic and easier to process by adding bulky side groups to them. Carboxyl/cholinesterases (CCEs) have been found to cleave compounds in both plant allelochemicals and synthetic pyrethroids, known as carboxylesters, into products with reduced toxicity [14]. How do generalists use these genes to feed on multiple host plants?

Genomics of Generalism

Plasticity

Generalist insects must possess genomes that equip them to deal with the constitutive or inducible defenses of multiple host plants. That is, insect populations that routinely encounter a range of host defenses are under significant selective pressure to produce phenotypes that can overcome all of them [11]. Generalist genomes could do that by making one, all-purpose phenotype, or by tailoring the phenotype to the specific host environment. The ability of one genotype to produce multiple phenotypes in different environments is referred to as phenotypic plasticity. For the case of plant-eating insects, the organism would need to sense differences in the host environment and respond accordingly [15]. Phenotypic plasticity has been observed to evolve in organisms with broad ecological niches, and in organisms that are exposed to constant changes in environmental factors such as temperature or salinity [16]. Does plasticity in the expression of detoxification genes explain the diets of generalist plant-eating insects?

Plastic Response Mechanisms

Early comparative genomics studies suggested that the genomes of generalist plant-eating insects simply have a wider variety of genes related to herbivory [17, 18]. An insect with more

detoxification proteins can detoxify more kinds of chemicals and can eat more kinds of plants. This may seem like it is so obvious that it has to be true, however, several alternative hypothesis have been proposed about how insects develop broad diets. For example, some of the early work on the molecular biology of diet diversity focused on variation in the substrate specificity of individual enzymes [19] – maybe generalists don't have more specialist enzymes, but rather have more generic versions of the same enzymes that specialists have. In line with this idea, Li et al. (2004) found that the CYP P450 enzymes of more generalist butterflies have more substrate channels and are able to catalyze the degradation of more substrates than those of more specialized butterflies.

Transcriptome Regulation

Alternatively, a generalist insect species' ability to feed on many plants may depend on the regulation of expression of herbivory genes; and it may have little to do with the number or specificity of herbivory genes [11]. Gene expression can evolve via several processes, for example genetic and epigenetic changes in (retro) transposon mobility, sequence rearrangements and losses, gene silencing, DNA methylation changes, and chromatin remodeling. Epigenetic modifications may be especially important in the evolution of diet breadth in plant-eating insects, as they can result in phenotypic changes much more rapidly than is expected through mutation in genome sequences [20].

Recent advances in next-generation sequencing technologies has provided powerful tools for examining plasticity in generalist insect species. RNA sequencing (RNA-Seq) approaches could be particularly useful. RNA-Seq experiments are able to measure the expression of thousands of transcribed genes and can be used to identify previously unknown transcripts that

are differentially regulated by insect species across multiple conditions. Modern RNA-Seq experiments do not require a reference genome, and recently have been used to analyze the transcriptomes of numerous non-model insect species [11]. Transcriptome studies have provided insight into how gene expression affects insect traits such as development, sexual dimorphism, polyphenism, and generalist host use [21, 22, 23].

Host Plasticity Genes

Gene expression and RNA-Seq studies on generalist plant-eating insects have identified previously unknown genes and functional pathways that that are plastically regulated in response to features of the host environment. These genes represent new potential targets for pesticide development. A study by Matzkin 2012 on host range evolution of the cactophilic fruit fly, *Drosophila mojavensis* revealed that over 20% of the fly's trasnscriptome was differentially expressed across defended and relatively non-defended cactus species. Upregulated differentially expressed genes on defended cacti were related to mRNA binding and toxin response. Downregulated genes were related to carbohydrate metabolism and cell structure [24]. Note the apparent tradeoff here between primary metabolism and detoxification – a subject we will revisit below. Another example of gene expression studies identifying new possible herbivory genes was conducted by Celorio-Mancera et al. [23] on host expansion in the polyphagous butterfly *Polygonia c-album*. Proteins of no *a priori* interest such as transport and membrane-binding proteins were differentially regulated across ancestral and more recently-acquired hosts. Cuticular proteins were also found to be more heavily expressed by caterpillars on the new host, possibly to reinforce the insect gut and keep toxic plant metabolites from diffusing into the insect body [23]. By seeing how gene expression changes across hosts, we are gaining new insights into what genes are involved in tailoring plant-eating insect phenotypes to specific hosts.

Insect Performance and Tradeoffs

Gene expression studies of generalist insect feeding also serve to highlight how changes in gene expression might affect their performance across hosts, and whether there are costs or tradeoffs to feeding on different hosts. Chi et al. (2009) examined gene expression change in the cowpea bruchid, *Callosobruchus maculatus*, while feeding on well-defended soybean plants [25]. They found that cowpea aphids decrease expression of structural, defense and stress-related genes on less defended plants, and that there may be a tradeoff between the expression of plant-defense compensation and other performance-improving genes. However, studies such as the one by Alon et al. [26] suggest that there may not be tradeoffs between detoxification expression and performance. Their assay of reproductive performance in the whitefly *Bemisia tabacci* revealed that increased expression of genes involved in metabolism and defense did not result in lower reproductive performance. The costs of gene expression plasticity are unclear, and could vary across species (in a way that could be predicted by other traits which vary across species, such as trophic and dispersal modes).

Thesis Research Design

For the most part, we have focused our studies of the genetics of host-use in plant-eating insects on detoxification genes. But there is surely more to it, and many questions about the genetics of insect herbivory are unanswered. What are the costs of detoxifying plant poisons? Are there tradeoffs between detoxification and primary metabolism? How does the genetics of herbivory in generalists differ from that in specialists? Have generalists evolved to have more detoxification genes? Is gene expression in generalists more plastic than in specialists? I

conducted a multi-part research project to address these questions that can be divided into three specific modules.

1) Characterize gene expression across hosts in wild populations of a generalist

In the first part of my project, I analyzed transcriptome data from an extremely generalist insect, the lobate lac scale, *Paratachardina pseudolobata*, sampled across three different host plants in multiple locations in southern Florida. We were interested in looking for three things in particular: 1) differential expression of detoxification genes across hosts, 2) differential expression of effector genes across hosts, and 3) tradeoffs in the expression of plant-defense compensation genes and primary metabolism genes. I constructed transcriptomes assemblies *de novo* and I tested for significant differential expression and functional transcript enrichment across host plant species.

2) Do generalist species have more detoxification genes?

For my second objective, I tested a simple idea about the genetic architecture of diet generalism with a statistical analysis of published plant-eating insect transcriptomes. The idea is that plant-eating insect with broader diets should have and express more detoxification genes. To test the gene-proliferation hypothesis for diet-breadth expansion, we analyzed published transcriptome and diet data for 42 species of lepidopterans (moths and butterflies) and hemipterans (true bugs) – the two insect orders with the best characterized plant diets. Each species' transcriptome was compared against a custom database of insect detoxification proteins and the number of transcripts matching detoxification genes was recorded for each insect. Each species' diet breadth was characterized as a count of host genera and families. We estimated the effect of detoxification gene diversity on host-use diversity using a Bayesian approach to generalized

linear mixed model regression. The model includes the phylogenetic relationships among transcriptome samples to account for non-independence due to shared ancestry.

3) Experimental Evolution of Gene Expression Plasticity

For this objective, I designed (but did not perform) a selection experiment to assay changes in gene expression and performance accompanying host-use adaptation. The experiment needs a generalist species and a specialist species; I use *Myzus persicae* (the peach-potato aphid) as the generalist and *Acyrthosiphon pisum* (the pea aphid) as the specialist. The basic idea is to use quasi-natural selection to evolve aphid host-use in the greenhouse. In the past few years, many studies have explored the process of host switching in a specialist species, or in a generalist species that has been maintained on one host for many generations and then is switched to another host for many generations. None have examined the evolution of niche breadth *per se*. Our plan is to start with our generalist population of *M. persicae* and select for monophagy by subjecting lines to a constant single-host environment for several generations, and allowing the individuals that perform the worst across multiple hosts to reproduce. Alternatively, we select for polyphagy in our specialist population of *A. pisum* by exposing them to a revolving door of different host plants for several generations. We then compare gene expression in control lines (maintained in their original host environment) to those that had been selected for new diet variation. And we compare performance of the single-host- and polyphagy-selected lines on a few hosts that are novel to both. To be clear, I did not actually perform this experiment. This part of my thesis is purely prospective, and served to give me the opportunity to think about other ways of addressing the main questions of my thesis research.

References

1. Daly H V., Doyen JT, Purcell AH. Introduction to insect biology and diversity. Introduction to insect biology and diversity. McGraw-Hill Book Company.; 1998. 680 p.

2. Janz N, Nylin S, Wahlberg N, Reveal J, Soltis D, Soltis P, et al. Diversity begets diversity: host expansions and the diversification of plant-feeding insects. BMC Evol Biol. BioMed Central; 2006 Jan 18;6(1):4.

3. Forister ML, Novotny V, Panorska AK, Baje L, Basset Y, Butterill PT, et al. The global distribution of diet breadth in insect herbivores. Vol. 112, Proceedings of the National Academy of Sciences of the United States of America. 2014. 1-6 p.

4. Mitter C, Farrell B, Wiegmann B. The phylogenetic study of adaptive zones: has phytophagy promoted Insect diversification? Am Nat. The University of Chicago PressThe American Society of Naturalists; 1988;132(1):107–28. 5. Keane RM, Crawley MJ. Exotic plant invasions and the enemy release hypothesis. Vol. 17, Trends in Ecology and Evolution. 2002. p. 164–70.

6. Nyman T. To speciate, or not to speciate? Resource heterogeneity, the subjectivity of similarity, and the macroevolutionary consequences of niche-width shifts in plant-feeding insects. Biol Rev. Blackwell Publishing Ltd; 2010 May 1;85(2):393–411.

7. Futuyma DJ, Agrawal AA. Macroevolution and the biological diversity of plants and herbivores. Proc Natl Acad Sci. National Academy of Sciences; 2009 Oct 27;106(43):18054–61.

8. Wittstock U, Gershenzon J. Constitutive plant toxins and their role in defense against herbivores and pathogens. Vol. 5, Current Opinion in Plant Biology. 2002. p. 300–7.

9. Rosenthal GA. The distribution of secondary compounds within plants. GA Rosenthal DH Janzen. Academic; 1979;1–718.

10. Simon JC, D'alenon E, Guy E, Jacquin-Joly E, Jaquiry J, Nouhaud P, et al. Genomics of adaptation to host-plants in herbivorous insects. Vol. 14, Briefings in Functional Genomics. 2015. p. 413–23.

11. Oppenheim SJ, Baker RH, Simon S, Desalle R. We can't all be supermodels: The value of comparative transcriptomics to the study of non-model insects. Vol. 24, Insect Molecular Biology. 2015. p. 139–54.

12. Feyereisen R. INSECT P450 ENZYMES. Annu Rev Entomol. 1999 Jan;44(1):507–33.

13. Snyder MJ, Glendinning JI. Causal connection between detoxification enzyme activity and consumption of a toxic plant compound. J Comp Physiol A. 1996 Aug;179(2):255–61.

14. Misra JR, Horner MA, Lam G, Thummel CS. Transcriptional regulation of xenobiotic detoxification in Drosophila. Genes Dev. Cold Spring Harbor Laboratory Press; 2011 Sep 1;25(17):1796–806.

15. Garland T. Phenotypic plasticity and experimental evolution. J Exp Biol. 2006;209(12):2344–61.

16. Price TD, Qvarnstrom A, Irwin DE. The role of phenotypic plasticity in driving genetic evolution. Proc R Soc B Biol Sci. The Royal Society; 2003 Jul 22;270(1523):1433–40.

17. Shi H, Pei L, Gu S, Zhu S, Wang Y, Zhang Y, et al. Glutathione S-transferase (GST) genes in the red flour beetle, Tribolium castaneum, and comparative analysis with five additional insects. Genomics. 2012 Nov;100(5):327–35.

18. Ramsey JS, Rider DS, Walsh TK, De Vos M, Gordon KHJ, Ponnala L, et al. Comparative analysis of detoxification enzymes in Acyrthosiphon pisum and Myzus persicae. Insect Mol Biol. 2010 Feb 23;19(SUPPL. 2):155–64.

19. Li X, Baudry J, Berenbaum MR, Schuler MA. Structural and functional divergence of insect CYP6B proteins: From specialist to generalist cytochrome P450. Proc Natl Acad Sci. National Academy of Sciences; 2004 Mar 2;101(9):2939–44.

20. Burggren W. Epigenetic Inheritance and Its Role in Evolutionary Biology: Re-Evaluation and New Perspectives. Biology (Basel). Multidisciplinary Digital Publishing Institute (MDPI); 2016 May 25;5(2):24.

21. Ragland GJ, Almskaar K, Vertacnik KL, Gough HM, Feder JL, Hahn DA, et al. Differences in performance and transcriptome-wide gene expression associated with Rhagoletis (Diptera: Tephritidae) larvae feeding in alternate host fruit environments. Mol Ecol. 2015 Jun;24(11):2759–76.

22. Govind G, Mittapalli O, Griebel T, Allmann S, Böcker S, Baldwin IT. Unbiased transcriptional comparisons of generalist and specialist herbivores feeding on progressively defenseless Nicotiana attenuata plants. Plaistow S, editor. PLoS One. Chapman & Hall; 2010 Jan 15;5(1):e8735.

23. De La Paz Celorio-Mancera M, Wheat CW, Vogel H, Söderlind L, Janz N, Nylin S. Mechanisms of macroevolution: Polyphagous plasticity in butterfly larvae revealed by RNA-Seq. Mol Ecol. 2013 Oct;22(19):4884–95.

24. Matzkin LM. Population transcriptomics of cactus host shifts in Drosophila mojavensis. Mol Ecol. Blackwell Publishing Ltd; 2012 May 1;21(10):2428–39.

25. Chi YH, Salzman RA, Balfe S, Ahn JE, Sun W, Moon J, et al. Cowpea bruchid midgut transcriptome response to a soybean cystatin - Costs and benefits of counter-defense. Insect Mol Biol. Blackwell Publishing Ltd; 2009 Feb 1;18(1):97–110.

26. Alon M, Elbaz M, Ben-Zvi MM, Feldmesser E, Vainstein A, Morin S. Insights into the transcriptomics of polyphagy: *Bemisia tabaci* adaptability to phenylpropanoids involves coordinated expression of defense and metabolic genes. Insect Biochem Mol Biol. 2012 Apr;42(4):251–63.

Chapter 2: Gene expression plasticity across hosts of an invasive scale insect species

Nicholas Christodoulides[1*], Alex R Van Dam[2,3], Daniel A Peterson[4], Rasmus John Normand Frandsen[2], Uffe Hasbro Mortensen[2], Bent Petersen[5], Simon Rasmussen[5], Benjamin B Normark[4], Nate B Hardy[1]

[1]Department of Entomology and Plant Pathology, Auburn University, Auburn, AL 36849

[2]Biosynthetic Pathways Engineering, Department of Bioengineering, Denmark Technical University, Søltofts plads, Building: 223, 226, 2800 Kgs. Lyngby, Denmark

[3]Present Address: Department of Biology, University of Puerto Rico, Mayagüez, Call Box 9000 Mayagüez, PR 00681-9000.

[4]Department of Biology and Graduate Program in Organismic and Evolutionary Biology, University of Massachusetts, Amherst, MA 01003

[5]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Kemitorvet, Building 208, 2800 Kgs. Lyngby, Denmark

[*]Corresponding Author

Email: nzc0035@auburn.edu (NC)

Abstract

For plant-eating insects, we still have only a nascent understanding of the genetic basis of host-use promiscuity. Here, to improve that situation, we investigated host-induced gene expression plasticity in the invasive lobate lac scale insect, *Paratachardina pseudolobata* (Hemiptera: Keriidae). We were particularly interested in the differential expression of detoxification and effector genes, which are thought to be critical for overcoming a plant's chemical defenses. We collected RNA samples from *P. pseudolobata* on three different host plant species, assembled transcriptomes *de novo*, and identified transcripts with significant host-induced gene expression changes. Gene expression plasticity was pervasive, but the expression of most detoxification and effector genes was insensitive to the host environment. Nevertheless, some types of detoxification genes were more differentially expressed than expected by chance. Moreover, we found evidence of a trade-off between expression of genes involved in primary and secondary metabolism; hosts that induced lower expression of genes for detoxification induced higher expression of genes for growth. Our findings are largely consonant with those of several recently published studies of other plant-eating insect species. Thus, across plant-eating insect species, there may be a common set of gene expression changes that enable host-use promiscuity.

Keywords

phytophagy, herbivory, phenotypic plasticity, polyphagy, generalism

Introduction

The overwhelming majority of plant-eating insect species are host-specialists that consume only one or a few closely-related plant species [1, 2]. But plant-eating insect communities predictably include a few species that are host-generalists, some of which can have extremely broad diets [2, 3]. These generalists are critical elements in terrestrial food webs; their high degree of ecological connectedness is a critical part of what makes ecosystems resilient and many specialist niches possible [4]. Generalists are also enriched in invasive and pest insect faunas, which gives them more economic weight than their species richness might suggest [5]. What genetic systems enable host-use promiscuity in plant eating insects? Some key host-use genes and proteins have been identified [6], and in the last few years there has been a number of studies examining how gene expression varies across hosts [7, 8]. Nevertheless, at this stage, few plant-eating insect species and host interactions have been studied, and we still know little with confidence about the genetic architecture of generalism in plant-eating insects.

We typically assume that the main limiters on plant-eating insect diets are plant defenses, and that a plant-eating insect's diet breadth is determined by how many of those defenses it can overcome [9]. Researchers have paid particular attention to a plant-eating insect's ability to detoxify plant defensive chemicals with a core suite of enzymes in a handful a large gene families: carboxylesterases, UDP-glycosyltransferases, glutathione-S-transferases, cytochrome oxygenase P450s, glutathione peroxidases, and ATP-binding cassette transporters (ABC transporters) [9, 10, 11]. However, this list may be too restrictive. For example, zinc-binding dehydrogenases have been implicated in plant-defense compensation in peach aphids [12]. Recently, researchers have come to recognize that the fitness of some plant-eating insects on

their host plants depends on effectors, that is, secreted macromolecules that attenuate a plant's inducible defenses [13, 14]. Disrupting a host plant's inducible defenses may be especially important for phloem-sap-sucking insects – such as aphids, whiteflies and scale insects – as they have close and persistent contact with their hosts [15]. Less attention has been paid to how factors other than plant defense, such as nutrition, could limit a plant-eating insect's diet, but non-defensive factors may be important.

The recent spate of comparative transcriptomics studies indicates that broad diets are, in fact, facilitated by differential expression of detoxification genes [8, 16, 17]. However, those studies also indicate that detoxification genes are only a small part of the picture. Most of the significant host-induced changes in gene expression do not appear to be related to detoxification. For example, researchers have found consistent host-dependent expression of many genes involved in primary metabolism [8, 18]. Furthermore, some studies have recovered evidence of a trade-off between expression of genes for detoxification and primary metabolism [7, 18]; for some plant-eating insects, increased investment in the expression of detoxification genes may come at the cost of reduced investment in growth and reproduction. But this is not the case for all insects. In the silverleaf whitefly, *Bemisia tabaci*, we find just the opposite: expression of genes for detoxification and primary metabolism rise and fall together [19]. Thus, the generality of this trade-off is not clear. Also unclear is the extent to which expression of effector genes varies across host plant species (but see [16] for evidence that it is important in host-use adaptation). Effector genes are involved in many biological processes and have many molecular functions. Consequently, there is no effector category in standard gene ontologies, and researchers would not detect enriched differential expression of effectors across hosts using standard enrichment

analyses. Furthermore, with one exception (Eyres et al. 2016), the published RNAseq-based differential expression analyses of insect diet variation have studied species with transient connections with their host plants (at least in comparison to sap-sucking insects), and might not be expected to strongly manipulate their host's induced defenses.

The lobate lac scale, *Paratachardina pseudolobata* (Hemiptera: Coccoidea: Kerriidae), is a good example of host-generalism (polyphagy) in plant-eating insects, and why it is worth studying. Thought to be native to Southeast Asia [21], the lobate lac scale is invasive and widespread in nearly all areas of Florida. It has been recorded from more than 307 plant species, including several of ecological and economic importance. Females are parthenogenic and sessile phloem feeders. They form dense aggregations on branches and stems, and excrete sugar-rich waste that feeds harmful sooty molds [22]. Some native plant species seem to be particularly vulnerable to lobate lac scale infection, including wax myrtle (*Myrica cerifera*), a favorite nesting area for many water-wading birds [23]. Controlling lobate lac scale populations has proven difficult; natural enemies are lacking in Florida, and thick resinous wax protects females from many insecticides [21].

In this study we investigated the genetic architecture of host-generalism in *Paratachardina pseudolobata.* Specifically, we examined gene expression variation across field-collected samples taken from multiple locations and hosts in southern Florida. We were particularly interested in 1) determining how expression of effector genes depends on host species, 2) confirming the importance of the differential expression of detoxification genes across hosts, and 3) assessing if there are trade-offs between the expression of genes involved in primary

metabolism and those involved in plant defense compensation.

Materials and Methods

Sample Collection

P. pseudolobata samples of whole, adult females were collected from four locations in southern Florida (Table 2.1). No specific permissions were required for these collections, and the insects are not protected or endangered species. Each sample was made up of about twenty specimens. A total of nine samples were taken from three host plants: two samples from *Tetrazygia bicolor* (Melastomataceae), three samples from *Myrsine cubana* (Myrsinaceae), and four samples from *Psychotria nervosa* (Rubiaceae). Specimens were plucked from their host plants and immediately homogenized in tubes filled with Trizol reagent. The ~20 specimens in each tube were pooled.

Table 2.1. Samples of lobate lac scales.

Location	Host species
Hugh Taylor Birch State Park, Ft Lauderdale, FL	*Psychotria nervosa*
IFAS Fort Lauderdale Research & Education Center, Davie, FL	*Tetrazygia bicolor*
IFAS Fort Lauderdale Research & Education Center, Davie, FL	*Myrsine cubana*
IFAS Fort Lauderdale Research & Education Center, Davie, FL	*Psychotria nervosa*
IFAS Fort Lauderdale Research & Education Center, Davie, FL	*Psychotria nervosa*
Navy Wells Pineland Preserve, Homestead, FL	*Tetrazygia bicolor*
Navy Wells Pineland Preserve, Homestead, FL	*Myrsine cubana*
Tree Tops Park, Davie, FL	*Myrsine cubana*
Tree Tops Park, Davie, FL	*Psychotria nervosa*

RNA Extraction, Library, Preparation, and Sequencing

Each RNA sample was homogenized using the Fast Prep FP120 Homogenizer Cell Disrupter (Thermo Fisher) for three pulses of 20 seconds using small steel beads. Then, 700µL of each

sample was transferred to a fresh centrifuge tube with 200µL of chloroform and shaken vigorously for 15 seconds and then incubated on ice for 20 minutes, shaking every 10 minutes. Samples were then centrifuged for 15 minutes at 14,000rpm at 4ºC. Approximately 400µL of each sample was added to an equal amount of 100% EtOH at 4ºC. This mixture was then passed through the standard protocol of the TRIzol® Plus RNA Purification Kit (Thermo Fisher). RNA yields and qualities were assessed using a Qubit fluorometer and an Agilent 2100 Bioanalyzer at the Denmark Technical University Multi-Assay Core (DMAC). CDNA Libraries were prepared at the University of California Davis Genome Center DNA Technologies Core. Samples were poly-A tailed, normalized, and pooled using Illumina TruSeq adapters. Samples were then loaded across three lanes of a single flow cell and sequenced in 100 base paired-end reads using an Illumina HiSeq3000.

Quality Control and Transcriptome Assembly

We removed adapters and low-quality sequences (quality score cutoff = 25) from the raw sequence reads using Trimmomatic 0.35 [24] and discarded reads less than 36 bp long. We then merged and normalized the quality-trimmed reads by kmer coverage and length using Khmer [25]. We assembled normalized reads into transcriptomes *de novo* using both Velvet 1.2.08 and Trinity [26, 27]. We constructed the Trinity assembly using default parameters (kmer = 25). We constructed eight Velvet assemblies, each with scaffolding enabled, and a minimum transcript length of 200bp, but with eight different kmer lengths: 27, 29, 31, 33, 35, 43, 53, 63. We then combined all assemblies from both programs into a single Fasta file and passed it through the EvidentialGene pipeline [28] to cluster sequences based on similarity, and find the best consensus transcripts for a final assembly. We assessed the quality of the assembly with BUSCO

and Transrate [29, 30]. The Transrate assembly score was 0.30, which is considered a passing score based on the number of input reads mapping to the assembly. BUSCO found that 89% of highly conserved arthropod sequences were present as single-copy or duplicated transcripts in the assembly, suggesting that it is mostly complete. For annotation and gene ontology assignments we passed the assembly through Annocript, BLASTing against the Uniref90, SwissProt and Conserved Domain databases with an e-value cut-off value of 1e-5 [31, 32]. We used these annotations for all subsequent analysis.

Differential Expression and GO Enrichment Analysis

We quantified the number of each transcript represented in the quality-trimmed read libraries by mapping them to the final assembly using eXpress and Bowtie2 [33, 34]. To determine how gene expression in lobate lac scale populations varied across host plant species, we first created a matrix of transcript counts using Trinity scripts. Next, we used the Bioconductor package Limma to convert the transcript counts to log-counts per million and the voom function to model the mean-variance relationship with precision weights across the three host plant species with three pair-wise comparisons [35]. We excluded low-abundance transcripts from the analysis. We used a false discovery rate (FDR) cutoff value of 0.05 to classify transcripts as differentially expressed, and made an expression heat map via Trinity scripts. We also used TopGO with gene ontology (GO) annotations provided by Annocript to perform GO term enrichment analysis [36]. TopGO categorizes differentially expressed transcripts under GO terms relating to general biological processes and molecular functions. We summarized these results with REVIGO [37].

We then looked specifically at the expression of detoxification and effector genes across hosts.

For detoxification genes, we examined transcripts which had been annotated as carboxylesterases, UDP-glycosyltransferases, glutathione-S-transferases, cytochrome oxygenase P450s, glutathione peroxidases, and ABC transporters. Investigation of insect effector genes was less straightforward, as they have only recently been characterized, fall into a variety of gene ontology classes, and are unlikely to be in curated protein databases. We began with a list of 67 candidate effector genes in the aphids *Acrythosiphon pisum* and *Myzus persicae* provided by [38]. We then extended this list to include 20 predicted effectors from aphid salivary gland transcriptomes [10]. We collated sequences for the genes in this list and identified lobate lac scale homologs in our assembly using the tblastx function in BLAST+ with an e-value cutoff of 0.00001.

Results

Assembly and Annotation

Sequencing generated ~120 million paired-end reads per library. The final, consensus assembly contained 113,670 transcripts, with a contig N50 of 1055 base pairs and a longest contig length of 33,906 base pairs. Annotation yielded 82,024 transcripts with at least one database hit. In comparison to published transcriptomes, the lobate lac transcriptome is most similar to those of the two-spotted spider mite, *Tetranychus urticae,* and the pea aphid, *Acrythosiphon pisum*. The most diverse genes in the assembly included sugar transporters, protein kinases, Ras family proteins, ubiquitins, and cytochrome P450s. Transcripts annotated as detoxification genes in the assembly included 354 cytochrome P450 oxygenases, 12 carboxylesterases, 96 glutathione S-transferases, 78 UDP-glycosyltransferases, 10 glutathione peroxidases, and 646 ABC transporter transcripts. Additionally, we found 1015 transcripts closely matching candidate aphid effectors.

Differential Expression

We identified 2,028 transcripts as differentially expressed across all comparisons. Most differentially expressed transcripts were upregulated on *M. cubana* relative to the other two plant species, but 280 and 58 genes were upregulated in *T. bicolor* and *P. nervosa* respectively (Figure 2.1). Only 23 transcripts matching detoxification genes were significantly differentially expressed: 3 UDP glycosyltransferases, 4 cytochrome P450s, 2 glutathione S-transferases, 1 glutathione peroxidase, and 13 subunits of ABC transporters. Each was upregulated in samples collected from *M. cubana,* except for 2 ABC transporter subunit transcripts and 1 glutathione S-transferase, which were upregulated in samples were from *T. bicolor* compared to *P. nervosa*. Of the 1015 transcripts which closely resemble possible aphid effectors, 8 were significantly differentially expressed (Table 2.2). Seven effector homologs were upregulated on *M. cubana* and one on *T. bicolor*.

Table 2.2 Differentially expressed aphid effector homologs and what is known about their function.

Comparison & Upregulation	Contig ID	Effector Blast Hit	Protein Name	Mode Of Action	References
• ***P. nervosa* vs *T. bicolor*** · Upregulated on *T. bicolor*	c60978_g1_i1	ACYPI009755-RA	Disulfide isomerase	Increases salivary protein formation in nematodes	Geldhof et al. (2003)[39]
• ***M. cubana* vs *P. nervosa*** · Upregulated on *M. cubana*	c29924_g1_i1	ACYPI002622-RA	Calreticulin	May circumvent calcium-mediated wound responses of host plant, prevents sieve tube occlusion	Carolan et al. (2011)
· Upregulated on *M. cubana*	c38772_g1_i1	ACYPI008001-RA	ARMET/ Endopeptidase inhibitor	Found in pea aphid saliva to assist aphid feeding	Wang et al. (2015)[40]
· Upregulated on *M. cubana*	c38738_g1_i1	ACYPI003917-RA	SCP GAPR-1	Similar to plant pathenogenesis protein (PR-1), alters defense mechanisms	Carolan et al. (2009)[41]
· Upregulated on *M. cubana*	c94746_g1_i1	ACYPI008370-RA	CLIP-domain serine protease	Inhibits phenol oxidase-based innate defenses of plants	Carolan et al. (2011)
· Upregulated on *M. cubana*	c37751_g1_i1	ACYPI009427-RA	M1 zinc metalloprotease	Deactivation of plants defense signaling peptides and dietary plant protease	Carolan et al. (2011)

			inhibitors in insect gut	
· Upregulated on *M. cubana* c30871_g1_i1	ACYPI000288-RA	Glucose dehydrogenase	Suppresses plant defense mechanism	Nicholson et al (2012)[42]
· Upregulated on *M. cubana* c41723_g1_i1	Gi|109195254|gb|EC388700.1|EC388700	Retinol dehydrogenase	Binds retinols and fatty acids and has been described to bind to lipid jasmonate precursors in *M. javanica*	Iberkleid et al. (2013)[43]

These homologs are significantly upregulated on one host plant in comparison to their average expression level on the other host plant. Host plant comparisons are in bold.

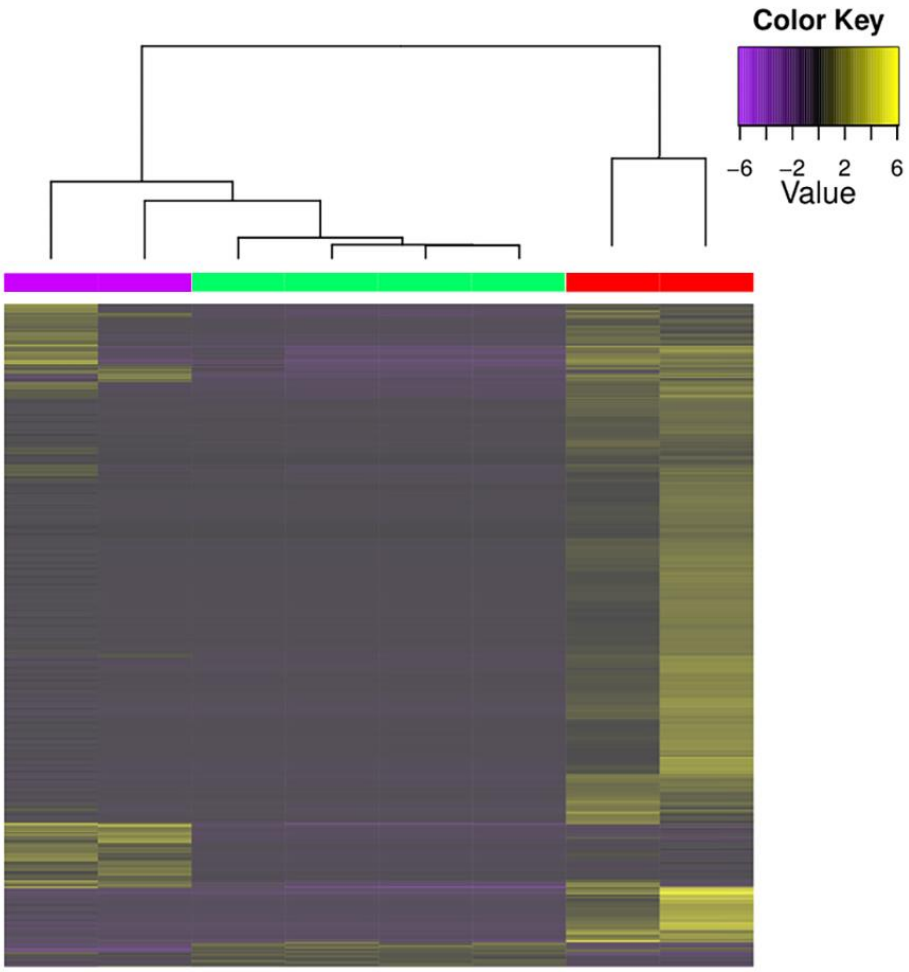


Figure 2.1 Heatmap showing differential expresssion across host plants. (Purple: *Tetrazygia bicolor*, Green: *Psychotria nervosa*, Red/: *Myrsine cubana*). Heatmap colors correspond to log2-transformed RPKM (Fragments Per Kilobase of transcript per Million mapped reads) values for each transcript. Yellow genes are upregulated in at least one pairwise comparison between the three host plants and purple genes are downregulated.

Functional Enrichment Analysis

The genes upregulated in lobate lac scale feeding on *M. cubana* were enriched for GO terms related to functions such as ion transport, ATP hydrolysis-coupled proton transport, regulation of gene expression, GTP binding, oxidoreductase activity on NADPH, phosphatase activity, ion binding, and metalloendopeptidase activity (Table 2.3). In contrast, on *P. nervosa*, upregulated expression was significantly enriched for macromolecule localization and protein catabolism. Enriched GO terms on *T. bicolor* included actomyosin structure, cellular component organization, and regulation of gene expression.

Table 2.3. Functional Enrichment Analysis

Host Plant Comparison	**Upregulated Genes**	**Top Enriched Biological Processes**	**Top Enriched Molecular Functions**
Myrsine cubana vs *Psychotria nervosa*	❖ *M. cubana* · 1738 ❖ *P. nervosa* · 57	❖ *M. cubana* · **Membrane fusion** · **Ion transport** · Cellular protein modification ❖ *P. nervosa* · Ubiquitin-dependent protein catabolism · Macromolecule localization	❖ *M. Cubana* ·GTP binding ·*Oxidoreductase activity on NADPH* ·Phosphoprotein phosphatase activity ·**Phospholipid binding** ·Endopeptidase inhibitor activity
Myrsine cubana vs *Tetrazygia bicolor*	❖ *M. cubana* · 1521 ❖ *T. bicolor* · 195	❖ *M. cubana* · Post-transcriptional regulation of gene expression · *ATP hydrolysis coupled proton transport* · Response to external stimulus ❖ *T. bicolor* · Actomyosin structure organization · Gene silencing by RNA	❖ *M.cubana* · Protein binding · *Metalloendopeptidase activity* · **Ion binding** · Zinc binding · **Calcium binding** ❖ *T. bicolor* · Structural molecule activity
Psychotria nervosa vs *Tetrazygia bicolor*	❖ *P. nervosa* · 36 ❖ *T. bicolor* · 246	❖ *P. nervosa* · Macromolecule localization ❖ *T. bicolor* · Negative regulation of gene expression · Cellular component	❖ *P. nervosa* · Translation factor activity, RNA binding ❖ *T. bicolor* · Motor activity · ATP binding

Host Plant Comparison	Upregulated Genes	Top Enriched Biological Processes	Top Enriched Molecular Functions
		organization	

Enriched GO terms among genes significantly upregulated on a host plant species relative to another in a pairwise comparison. GO terms potentially related to detoxification are italicized. GO terms that are potentially related to effector activity are in bold. Terms related to primary metabolism are underlined.

Many of the GO terms enriched for differential expression on *M. cubana* are thought to be involved in detoxification (Table 2.3). Specific mechanisms are not clear, but increased expression of these genes has been associated with exposure to plant chemical defenses [44]. By contrast, these GO terms are not enriched on *T. bicolor* and *P. nervosa.* Instead, on those host plant species, we saw increased expression of genes involved in primary metabolism.

Discussion

The recent comparative genomics work is beginning to show us general features of the genetic architecture of host-use variation in plant-eating insects. In this study, we find additional support for the importance of host-dependent expression of detoxification genes. We also find further evidence that detoxification genes account for only a small part of all host-induced gene expression changes, and that there is a trade-off between expression of genes for detoxification and those for primary metabolism. We recover evidence of a striking diversity of effector proteins in lobate lac scale; more than a thousand transcripts, ~ 1% of all transcripts in the sample, are near matches to one of 87 putative aphid effectors, and several of these transcripts are differentially expressed in the lobate lac scale across host plant species. Nevertheless, effectors as a group were not enriched for differential expression; most of them are expressed at the same level across hosts.

Detoxification Genes

Of the 1196 putative detoxification genes that were expressed, only 23 were significantly upregulated on any one host. Pervasive host-insensitivity in the expression of detoxification genes appears to be the trend across plant-eating insects [18, 45]. However, we did find that ABC transporters were more differentially expressed across hosts than expected by chance. The pool of differentially expressed genes was significantly enriched for genes involved in ATP hydrolysis-coupled proton transport (p-value = 1.6e-5) (Table 2.3), a process carried out by ATPase proteins, which are subunits of ABC transporter proteins. ABC transporter expression is associated with xenobiotic elimination and insecticide resistance in several insect species [44].

Only a few of the genes that we identified *a priori* as being involved in detoxification were found to be differentially expressed, but perhaps our *a priori* assignments were too exclusive. In fact, some of the enriched GO terms in lobate lacs scale from *M. cubana* could be indicative of host-induced differences in the expression of genes involved in detoxification pathways (Table 2.3). Most notably, oxidoreductase activity on NADPH is often linked to detoxification, as several detoxification proteins use NADPH as an electron acceptor [46]. These proteins include cytochrome P450 oxygenases, and glutathione peroxidases [47]. In other words, we found significantly enriched differential expression of some genes that are associated with detoxification genes in metabolic networks. Furthermore, oxidoreductase activity and zinc binding, other enriched GO terms, are functions of zinc-binding dehydrogenases, which appear to be important for plant defense compensation in peach aphids [12]. Thus, if we consider a more inclusive set of genes involved in plant-defense compensation, our expression data provides more support for the importance of host-induced differential expression.

Aphid Effectors on *M. cubana*

Each of the differentially expressed effector transcripts is likely to have a similar function to its closest matching aphid transcript (Table 2.2). These effectors have several functions. M1 zinc metalloprotease is one of the most well-understood effector proteins. Sap-sucking insects use it to destroy plant signaling defense peptides and improve phloem sap quality by increasing free amino acid content [10]. A second effector that we know something about is calreticulin. Calcium ions transmit information in plant signaling pathways. Calreticulin binds calcium ions and disrupts these signals. This prevents plants from closing off compromised sieve tube elements, which would cut off the supply of phloem sap [10]. The over-expression of a calreticulin homolog may account for the enrichment of calcium binding processes we found in lobate lac scale on *M. cubana.* A third effector with characterized function is retinol dehydrogenase. It attracts retinols and fatty acids and can bind lipid jasmonate precursors to prevent jasmonate pathway defenses [43]. This may relate to the enriched lipid and phospholipid binding we observe in lobate lac scale on *M. cubana*.

Induction of Effectors and Detoxification Genes in *M. cubana*

The great majority of differentially expressed genes were upregulated on *M. cubana.* This could indicate that *M. cubana* is simply better defended than the other plant species. Plants in the Myrsinaceae family are known to produce a variety of benzoquinone compounds, characterized by an aromatic ring and two carbonyl groups [48]. Benzoquinones and their derivatives have antibacterial, insecticidal, and anti-feeding properties [49]. Benzoquinones are toxic to several species of aphid as well as the extremely polyphagous red spider mite, which shows that they could be difficult to process by generalist insect species [50]. Many species of Myrsinaceae also

accumulate saponins in their tissues. Saponins are cytotoxic chemicals composed of a hydrophobic aglycone (non-sugar) attached to glycoside (sugar residue) [48].They play an important role in plant defense against insects and parasites, and they also elicit immune responses in people [51]. Myrsinaceous plants can also be induced to produce proteinase inhibitors known as phytostatins in response to wounding [52].These proteinase inhibitors make their way to the insect gut and inhibit digestion. In comparison to *M. cubana*, *P. nervosa* and *T. bicolor* are thought to be less protected by defensive chemistry [53, 54]

DE of Gene Expression Regulation

We found that some of the most conspicuous changes in gene expression across hosts occurs at genes involved in regulating gene expression. In lobate lac scale on *T. bicolor,* we found enrichment of genes for gene silencing by RNA. This was previously shown on some hosts of *Myzus persicae* (in which case miRNA activity is negatively correlated with the abundance of a P450 enzyme) [55]. In lobate lac scales on *P. nervosa,* we found enriched expression of genes for RNA binding and translation factor activity, both of which are thought to play a role in eukaryotic post-transcriptional regulation [56]. Previous RNA-seq studies of how gene expression differs across hosts of plant-eating insect species have also found differential expression of genes for ribosomal proteins and nucleic acid binding. This kind of differential expression may integrate across GO categories and be indicative of broad-scale expression changes that may nonetheless not amount to significant enrichment of differential expression within GO categories.

Caveats

In this study, some of the gene expression variation that we observed across hosts could be due to biotic and abiotic factors that covary with hosts across sites. For example, gene expression variation could correspond to variation in local light and soil conditions, or assemblages of other plant-eating insects and natural enemies. In future studies, we could minimize the noise caused by such factors by using common garden experimental designs, or sampling much more intensively across wild populations. The strictly bioinformatics approach that we have taken to functional analysis also has limitations. The function of candidate genes for host-generalism could be more accurately characterized with genome editing approaches such as RNA interference. And that may yield insights that could yield new tools to control populations of pests.

Conclusion

For decades, we assumed that host use in a plant-eating insect is limited by their capacity to cope with the chemical defenses of plant species, but we knew little about the details. In the last few years, comparative transcriptomics analyses of plant-eating insects have identified hundreds of genes that are differentially expressed across hosts, including many genes suspected to be involved in the detoxification of plant chemicals. They have also shown us that host-use adaptation is much more complex than activating or silencing the expression of a few detoxification genes, and may entail gene expression trade-offs between plant-defense compensation and primary metabolism. Here, by analyzing gene expression plasticity across three host species of the lobate lac scale insect, we find further support for each of these insights. Moreover, we find evidence that the genes directly involved in plant-defense-compensation may

vary across multiple plant-eating insect species, and may be more diverse than commonly thought. We also find evidence that in some cases, host-defense compensation may depend on differential expression of genes that are adjacent to detoxification enzymes in metabolic pathways. These insights have the potential to pave the way for a more useful theory of host-use adaptation, and for new, more economical and sustainable tools for managing pest insect populations.

Acknowledgments

Thanks to Robin Giblin-Davis, William Kern, Ian Stocks, and Catherine Mannion for helping us find scale insects and learn the southern Florida flora. Thanks also to Mayrolin García-Morales for helping with the field work. NC thanks the Auburn University Department of Entomology and Plant Pathology for research assistantship support. We thank Marlene Dalgaard (DMAC) for use of their Qubit and Bioanalyzer. We thank Andrea Schreier for sample processing at UC Davis. We also thank Lutz Froenicke at UC Davis for advice on the sequencing scheme for this project. Funding: This work was supported by the National Science Foundation [DBI 1306489] awarded to AVD; and the Alabama Agricultural Experiment Station.

References

1. Bernays E, Graham M. On the Evolution of Host Specificity in Phytophagous Arthropods. Ecology. 1988;69(4):886–92.

2. Forister ML, Novotny V, Panorska AK, Baje L, Basset Y, Butterill PT, et al. The global distribution of diet breadth in insect herbivores. Vol. 112, Proceedings of the National Academy of Sciences of the United States of America. 2014. 1-6 p.

3. Hardy NB, Peterson DA, Normark BB. Scale insect host ranges are broader in the tropics. Biol Lett. 2015;11:20150924.

4. Bascompte J. Plant-animal mutualistic networks: the architecture of biodiversity. Appl Math. 2007;101(2):187–203.

5. Ross L, Hardy NB, Okusu A, Normark BB. Large Population Size Predicts the Distribution of Asexuality in Scale Insects. Evolution (N Y). 2012;

6. Czosnek H, Ghanim M. Management of insect pests to agriculture: Lessons learned from deciphering their genome, transcriptome and proteome. Management of Insect Pests to Agriculture: Lessons Learned from Deciphering their Genome, Transcriptome and Proteome. 2016. 1-290.

7. Govind G, Mittapalli O, Griebel T, Allmann S, Böcker S, Baldwin IT. Unbiased transcriptional comparisons of generalist and specialist herbivores feeding on progressively defenseless Nicotiana attenuata plants. PLoS One. 2010;5(1).

8. Ragland GJ, Almskaar K, Vertacnik KL, Gough HM, Feder JL, Hahn DA, et al. Differences in performance and transcriptome-wide gene expression associated with Rhagoletis (Diptera: Tephritidae) larvae feeding in alternate host fruit environments. Mol Ecol. 2015;24(11):2759–76.

9. Li X, Schuler MA, Berenbaum MR. Molecular Mechanisms of Metabolic Resistance to

Synthetic and Natural Xenobiotics. Annu Rev Entomol. 2007;52(1):231–53.

10. Carolan JC, Caragea D, Reardon KT, Mutti NS, Dittmer N, Pappan K, et al. Predicted effector molecules in the salivary secretome of the pea aphid (Acyrthosiphon pisum): A dual transcriptomic/proteomic approach. J Proteome Res. 2011;10(4):1505–18.

11. You M, Yue Z, He W, Yang X, Yang G, Xie M, et al. A heterozygous moth genome provides insights into herbivory and detoxification. Nat Genet. 2013;45(2):220–5.

12. Cooper WR, Dillwith JW, Puterka GJ. Salivary proteins of Russian wheat aphid (Hemiptera: Aphididae). Environ Entomol. 2010;39(1):223–31.

13. Bellafiore S, Shen Z, Rosso M-N, Abad P, Shih P, Briggs SP. Direct Identification of the Meloidogyne incognitaSecretome Reveals Proteins with Host Cell Reprogramming Potential. PLoS. 2008;

14. Harris MO, Friesen TL, Xu SS, Chen MS, Giron D, Stuart JJ. Pivoting from Arabidopsis to wheat to understand how agricultural plants integrate responses to biotic stress. Vol. 66, Journal of Experimental Botany. 2015. p. 513–31.

15. Will T, Furch ACU, Zimmermann MR. How phloem-feeding insects face the challenge of phloem-located defenses. Front Plant Sci. 2013;4(August):336.

16. Wybouw N, Zhurov V, Martel C, Bruinsma KA, Hendrickx F, Grbic V, et al. Adaptation of a polyphagous herbivore to a novel host plant extensively shapes the transcriptome of herbivore and host. Mol Ecol. 2015;24(18):4647–63.

17. De La Paz Celorio-Mancera M, Wheat CW, Vogel H, Söderlind L, Janz N, Nylin S. Mechanisms of macroevolution: Polyphagous plasticity in butterfly larvae revealed by RNA-Seq. Mol Ecol. 2013;22(19):4884–95.

18. Hoang K, Matzkin LM, Bono JM. Transcriptional variation associated with cactus host plant

adaptation in Drosophila mettleri populations. Mol Ecol. 2015;24(20):5186–99.

19. Alon M, Elbaz M, Ben-Zvi MM, Feldmesser E, Vainstein A, Morin S. Insights into the transcriptomics of polyphagy: Bemisia tabaci adaptability to phenylpropanoids involves coordinated expression of defense and metabolic genes. Insect Biochem Mol Biol. 2012;42(4):251–63.

20. Eyres I, Jaquiry J, Sugio A, Duvaux L, Gharbi K, Zhou JJ, et al. Differential gene expression according to race and host plant in the pea aphid. Mol Ecol. 2016;25(17):4197–215.

21. Pemberton RW. Potential for biological control of the lobate lac scale, Paratachardina lobata lobata (Hemiptera : Kerriidae). Florida Entomol. 2003;86:353–60.

22. Howard FW, Pemberton R, Hamon A, Hodges GS, Steinberg B, Mannion M, et al. (Chamberlin) (Hemiptera : Sternorrhyncha : Coccoidea : Kerriidae) 1. Univ Florida. 2014;1–14.

23. Pemberton RW. Invasion of Paratachardina lobata lobata (Hemiptera: Kerriidae) in South Florida: A snapshot sample of an infestation in a residental yard. In: Florida Entomologist. 2003. p. 373–7.

24. Bolger A, Lohse M UB. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114.

25. Brown CT, Howe A, Zhang Q, Pyrkosz AB, Brom TH. A Reference-Free Algorithm for Computational Normalization of Shotgun Sequencing Data. arXiv. 2012;1203.4802(v2):1–18.

26. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8(8):1494–512.

27. Zerbino DR, Birney E. Gibbs, RA, and Buetow, KH 2005. Velvet: algorithms for de novo

short read assembly using de Brujin graphs. Genome Res. 2008 May;18(5):821–9.

28. Gilbert D. EvidentialGene: mRNA Transcript Assembly Software [Internet]. Genome Informatics Lab of Indiana University Biology Department. 2013. Available from: http://arthropods.eugenes.org/genes2/about/EvidentialGene_trassembly_pipe.html

29. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–2.

30. Smith-Unna RD, Boursnell C, Patro R, Hibberd JM, Kelly S. TransRate: reference free quality assessment of de-novo transcriptome assemblies. bioRxiv. 2015;21626.

31. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST plus: architecture and applications. BMC Bioinformatics. BioMed Central; 2009 Dec 15;10(421):1.

32. Musacchia F, Basu S, Petrosino G, Salvemini M, Sanges R. Annocript: A flexible pipeline for the annotation of transcriptomes able to identify putative long noncoding RNAs. Bioinformatics. 2015;31(13):2199–201.

33. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol. 2011;12(3):R22.

34. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.

35. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15:R29.

36. Alexa A, Rahnenfuhrer J. topGO: enrichment analysis for gene ontology. Vol. 2, R package version. 2016.

37. Supek F, Bosnjak M, Skunca N, Smuc T, S ,. REVIGO summarizes and visualizes long lists

of Gene Ontology terms. PLo e2. 1800;6.

38. Bos JIB, Prince D, Pitino M, Maffei ME, Win J, Hogenhout SA. A functional genomics approach identifies candidate effectors from the aphid species Myzus persicae (green peach aphid). PLoS Genet. 2010;6(11).

39. Geldhof P, Vercauteren I, Knox D, Demaere V, Van Zeveren A, Berx G, et al. Protein disulphide isomerase of Ostertagia ostertagi: An excretory-secretory product of L4 and adult worms? Int J Parasitol. 2003;33(2):129–36.

40. Wang W, Dai H, Zhang Y, Chandrasekar R, Luo L, Hiromasa Y, et al. Armet is an effector protein mediating aphid-plant interactions. FASEB J. 2015;29(5):2032–45.

41. Carolan JC, Fitzroy CIJ, Ashton PD, Douglas AE, Wilkinsonl TL. The secreted salivary proteome of the pea aphid Acyrthosiphon pisum characterised by mass spectrometry. Proteomics. 2009;9(9):2457–67.

42. Nicholson SJ, Hartson SD, Puterka GJ. Proteomic analysis of secreted saliva from Russian Wheat Aphid (Diuraphis noxia Kurd.) biotypes that differ in virulence to wheat. J Proteomics. 2012;75(7):2252–68.

43. Iberkleid I, Vieira P, de Almeida Engler J, Firester K, Spiegel Y, Horowitz SB. Fatty Acid- and Retinol-Binding Protein, Mj-FAR-1 Induces Tomato Host Susceptibility to Root-Knot Nematodes. PLoS One. 2013;8(5).

44. Vogel H, Musser RO, de la Paz Celorio-Mancera M. Transcriptome Responses in Herbivorous Insects Towards Host Plant and Toxin Feeding. Annu Plant Rev Insect-Plant Interact. 2014;47:197–233.

45. Strode C, Wondji CS, David JP, Hawkes NJ, Lumjuan N, Nelson DR, et al. Genomic analysis of detoxification genes in the mosquito Aedes aegypti. Insect Biochem Mol Biol.

2008;38(1):113–23.

46. Miles PW. O. The significance of antioxidants in the aphid plant interaction - the redox hypothesis.pdf. Entomol Exp Appl. 1993;275–83.

47. Mao W, Berhow M, Zangerl A, Mcgovern J, Berenbaum M. Cytochrome P450-Mediated Metabolism of Xanthotoxin by Papilio multicaudatus. Vol. 32, Journal of Chemical Ecology. 2006. p. 523–36.

48. Lukhele T, Krause, M C-SRW, Olivier MDK. Isolation, Characterisation and Biological Activity of Some Compounds From Rapanea Melanophloeos (L.) Mez. Mez. 2009.

49. Midiwo J, Mwangi R. Insect antifeedant, growth-inhibiting and larvicidal compounds from Rapanea melanphloes (Myrsinaceae). Int J. 1995;

50. Akhtar Y, Isman MB, Lee CH, Lee SG, Lee HS. Toxicity of quinones against two-spotted spider mite and three species of aphids in laboratory and greenhouse conditions. Ind Crops Prod. 2012;37(1):536–41.

51. Augustin JM, Kuzina V, Andersen SB, Bak S. Molecular activities, biosynthesis and evolution of triterpenoid saponins. Vol. 72, Phytochemistry. 2011. p. 435–57.

52. Volpicella M, Leoni C, Costanza A, De Leo F, Gallerani R, Ceci LR. Cystatins, serpins and other families of protease inhibitors in plants. Curr Protein Pept Sci. 2011;12(5):386–98.

53. Adonizio A, Downum K, Bennett B, Mathee K. Anti-quorum sensing activity of medicinal plants in southern Florida. J Ethnopharmacol. 2006;105(3):427–35.

54. Melendez PA, Capriles VA. Antibacterial properties of tropical plants from Puerto Rico. Phytomedicine. 2006;13(4):272–6.

55. Peng T, Pan Y, Gao X, Xi J, Zhang L, Ma K, et al. Reduced abundance of the CYP6CY3-targeting let-7 and miR-100 miRNAs accounts for host adaptation of Myzus persicae nicotianae.

Insect Biochem Mol Biol. 2016;75:89–97.

56. Glisovic T, Bachorik JL, Yong J, Dreyfuss G. RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett. 2008;582(14):1977–86.

Chapter 3: Genetic architecture of evolvability: Generalist insect herbivores have more detoxification genes.

[1*]Nicholas Christodoulides and [1]Nate B. Hardy

[1]Department of Entomology and Plant Pathology, Auburn University, Auburn, AL 36849.

*corresponding author. Email: nzc0035@auburn.edu

Abstract

Some herbivorous insect species eat more plant species than others, and species with broader diets appear to more readily evolve new host associations. How do we explain this enhanced evolvability of host use? Here we test the role of one genome-level property thought to affect evolvability: genetic redundancy. Specifically, we look to see if the diet breadth of herbivorous insect species positively co-varies with the diversity of protein transcripts involved in the detoxification of plant defensive chemicals. With comparative phylogenetic analyses of transcriptomes from 43 herbivorous insect species, we find a strong and significant effect. As the theory predicts, more polyphagous insect herbivores have more diverse detoxification proteins.

Background

Looking across herbivorous insect species, we see that the variation in diet breadth has a long tail of specialist species. In fact, it fits a power-law distribution [1]. Wherever power-law distributions occur, they point to an underlying positive-feedback process [2]. Simply, the bigger a herbivorous insect's diet, the better the odds that it will evolve an even bigger diet. Why? Here, we focus on potential evolutionary feedbacks involving the genetic architecture of host use. The evolution of diet breadth in herbivorous insects is a striking example of how the evolvability of a trait can vary across species. Evolutionary geneticists have nominated several genetic architectural properties which could govern evolvability [3]. One of the most salient is redundancy; having multiple versions of the same gene relaxes stabilizing selection on at least one of them, and facilitates adaptive evolution [4]. In theory, herbivorous insects with more herbivory genes should have a greater potential to evolve host use.

Plants have evolved many defenses against herbivorous insects. Chemical defenses (i.e. poisons) are thought to be especially important [5]. For a herbivorous insect to eat a plant, it needs to be able to survive exposure to that plant's poisons. Metabolic detoxification is one solution. Hence, one might predict that species with broader diets should have broader arrays of detoxification proteins that are transcribed or expressed. This prediction is corroborated by some anecdotal evidence. Generalist species of *Papillio* swallowtail butterflies have more P450 detoxification genes than specialist congeners [6]. The polyphagous cotton bollworm, *Helicoverpa armigera,* expresses more detoxification genes than its more specialist relative, the tobacco budworm, *H. assulta* [7], and the super-polyphagous peach-potato aphid (*Myzus persicae*) has more P450s than the less polyphagous pea aphid (*Acyrthosiphon pisum*) [8].

Despite that evidence, the authors of a recent comparative genomic study claim that across herbivorous insects, there is no relationship between diet breadth and detoxification gene redundancy [9]. However, that study was skewed towards saprophytic flies and pollinators, that is, species that do not feed on living and defended plant-tissues. Moreover, diet diversity was not explicitly parameterized and no formal statistical analysis was performed. As it stands, the general effect of detoxification gene redundancy on diet breadth in herbivorous insects is unclear.

Note that we have no obvious reason to expect that the genetic redundancy hypothesis is wrong, and that having fewer herbivory genes would increase diet evolvability. What is not clear is if it actually explains observed differences in evolvability and diet breadth across species. Alternative factors could also affect the evolvability of a host use, and effectively make genetic redundancy unnecessary. Examples are changes in generation time, population size, or to gene expression networks that make them more robust to variations in the host environment. It could also be that genetic redundancy is a critical governor of diet evolution, but that the detoxification of plant poisons is less important that we assume. Here, we test the effect of detoxification transcript diversity on the diets of herbivorous insects by performing a statistical comparison of published transcriptomes.

Methods

Transcriptome data. We obtained transcriptomes of 23 species of lepidopterans (moths and butterflies) and 19 hemipterans (true bugs) from the NCBI Transcriptome Shotgun Assembly database. These two orders have the most well-documented feeding patterns and host associations. To this sample, we added one more hemipteran transcriptome that we assembled *de*

novo (the scale insect species *Paratachardina pseudolobata* (Keriidae). This resulted in a total of 43 transcriptomes. We used BLAST+ to compare each transcript to a custom database composed of 44696 insect detoxification proteins downloaded from UniProtKB and NCBI protein databases [10]. This database included reference sequences from six protein classes: 1) Cytochrome P450 monooxygenases, 2) carboxylesterases, 3) glutathione peroxidases, 4) glutathione S-transferase, 5) UDP-glycosyltransferases, and 6) other proteases. For each insect species, we used the top hit for each sequence to record the number of unique proteins matching detoxification protein sequences with an e-value threshold of 1e-10. As a measure of the sequencing effort behind each transcriptome, we also noted the average lengths of detoxification transcripts, along with the total number of transcripts in each assembly.

Host-use data. The diet breadth of each species in our transcriptome dataset was parameterized as a count of host plant families. Host use data for lepidopterans came from the HOSTS database [11]. Host use data for hemipterans came from several sources: ScaleNet [12]; Aphids on the Worlds Plants (http://www.aphidsonworldsplants.info/); and the Tri-Trophic Database project (http://tcn.amnh.org/). We were unable to find host use data for four species. These were excluded from the analysis.

Phylogeny estimation. To account for non-independence of our observations due to shared ancestry, we estimated phylogenetic relationships. First, we used BLAST+ searches of our transcripts to identify orthologous sequences of cytochrome oxidase subunit 1 (COI) and elongation factor 1-a (EF1A). Next, we aligned sets of orthologs with MAFFT [13], and trimmed poorly-aligned segments with trimAL [14]. From a concatenation of the COI and EF1A alignments, we used RaxML [15] to estimate the phylogeny along with the parameters of a GTR + G nucleotide substitution model, unlinked between loci. In the tree search we constrained

relationships to conform to the NCBI classification, as they are not controversial. We then time-scaled branch lengths of the optimal tree, by using the relaxed clock method implemented in TreePL [16, 17], and four divergence time calibrations from the TimeTree aggregator [18]. Specifically, we applied uniform constraints on the ages of the crown nodes of Aleyrodidae (83.27-93.27), Lepidoptera (224 – 234 Ma), Noctuidae (60.39 – 70.39 Ma), and a fixed constrain of 357.6 Ma on the root node.

Statistical analysis. We used a Bayesian approach to perform generalized linear mixed model regressions of diet breadth (the response variable) on detoxification transcript diversity, and average detoxification gene transcript length (to account for sampling intensity). We modeled the variance in the response variable, counts of host-plant families, with an over-dispersed Poisson distribution, and related it the model covariates through a log link function. We incorporated the evolutionary relationships among species as random effects, expressed as a covariance matrix derived from our time-scaled phylogeny estimate. MCMC chains were run for 1 million iterations, sampling parameters once every 100 iterations, and discarding the first 20% as burn-in. Full prior and model specifications are given in S1.

Results

Descriptive summaries of the sampled transcriptomes, along with host-plant range estimates are provided in Table 3.2. An overview of the data is also shown as a phylogeny with tip annotations in Figure 3.1.

Total number of detoxification transcripts is a good predictor of diet breadth (Figure 3.2). While accounting for sampling effort, we found that adding one novel detoxification transcript

corresponds to an increase of ~ 1 new host plant family to the diet (pMCMC = .01) Model summaries are presented in Table 3.1.

Discussion

This study supports the notion that genetic redundancy is a key governor of evolvability. It also provides the first statistical support for the more specific idea that, in herbivorous insects, having more detoxification genes in the transcriptome opens to door to a broader diet. We expect that insects encountering diverse plant poisons will be under selection to expand their detoxification capacity [19]. Furthermore, we have seen plant-eating insects evolve metabolic resistance to synthetic insecticides – a challenge analogous to plant poisons – through detoxification gene over-expression via gene diversification [20]. This link between gene number counts and expression levels could favor the fixation of a gene duplication [21]. The subsequent evolution of alternative expression modulators could then free up one gene copy for functional divergence [4].

This study has some limitations. Most importantly, the analyzed transcriptomes were generated from a variety of experiments. For most of the species in the study, the number of transcripts of a particular class, for example P450s, closely matched published gene counts [22, 23]. But for some species we counted up to three times as many transcripts as there are putative genes. This inflation could be due to alternative splicing of transcripts, duplicated or diversified genes being counted twice during transcriptome assembly, or alternative transcriptions of detoxification genes. It could also result from biases in transcriptome assembly methods, for example multiple non-overlapping contigs for one transcript [24]. Unless these possible

assembly artifacts were more common in polyphagous species, they should not have biased our analysis.

Conclusions

More polyphagous – and more evolvable – species of plant-eating insect tend to produce more detoxification transcripts. Genetic redundancy is one aspect of genetic architecture predicted to affect evolvability. There are others, for example, the robustness of gene expression networks. We look forward to future research exploring how the evolution of diet diversity feeds back on genetic architecture.

Ethics

No ethical approval or requirements were needed or obtained for this study.

Data accessibility

Information regarding transcriptomes used in the study can be found in Table 3.2

References

1. Forister ML, Novotny V, Panorska AK, Baje L, Basset Y, Butterill PT, Cizek L, Coley PD, Dem F, Diniz IR, Drozd P. The global distribution of diet breadth in insect herbivores. Proceedings of the National Academy of Sciences. 2015 Jan 13;112(2):442-7.

2. Scanlon TM, Caylor KK, Levin SA, Rodriguez-Iturbe I. Positive feedbacks promote power-law clustering of Kalahari vegetation. Nature. 2007 Sep 13;449(7159):209-12.

3. Hansen TF. The Evolution of Genetic Architecture. Annu Rev Ecol Evol Syst. Annual Reviews; 2006 Dec;37(1):123–57.

4. Zhang J. Evolution by gene duplication: An update. Vol. 18, Trends in Ecology and Evolution. 2003. p. 292–8.

5. Futuyma DJ, Agrawal AA. Macroevolution and the biological diversity of plants and herbivores. Proc Natl Acad Sci. National Academy of Sciences; 2009 Oct 27;106(43):18054–61.

6. Li X, Berenbaum MR, Schuler MA. Plant allelochemicals differentially regulate Helicoverpa zea cytochrome P450 genes. Insect molecular biology. 2002 Aug 1;11(4):343-51.

7. Li H, Zhang H, Guan R, Miao X. Identification of differential expression genes associated with host selection and adaptation between two sibling insect species by transcriptional profile analysis. BMC genomics. 2013 Aug 28;14(1):582.

8. Ramsey JS, Rider DS, Walsh TK, De Vos M, Gordon KH, Ponnala L, Macmil SL, Roe BA, Jander G. Comparative analysis of detoxification enzymes in Acyrthosiphon pisum and Myzus persicae. Insect Molecular Biology. 2010 Mar 1;19(s2):155-64.

9. Rane RV, Walsh TK, Pearce SL, Jermiin LS, Gordon KH, Richards S, Oakeshott JG. Are feeding preferences and insecticide resistance associated with the size of detoxifying enzyme families in insect herbivores?. Current Opinion in Insect Science. 2016 Feb 29;13:70-6.

10. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. BioMed Central; 2009 Aug 15;10(16):421.

11. Robinson GS, Ackery PR, Kitching IJ, Beccaloni GW, Hernández LM. HOSTS - A database of the world's Lepidopteran Hostplants. Natural History Museum, London. 2010.

12. Morales MG, Denno BD, Miller DR, Miller GL, Ben-Dov Y, Hardy NB. ScaleNet: A literature-based model of scale insect biology and systematics. Database. 2016;2016.

13. Katoh K, Misawa K, Kuma KI, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic acids research. 2002 Jul 15;30(14):3059-66.

14. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. Oxford University Press; 2009 Aug 1;25(15):1972–3.

15. Stamatakis, Alexandros. "RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models." Bioinformatics 22.21 (2006): 2688-2690.

16. Sanderson MJ. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol. 2002;19(1):101–9.

17. Smith SA, O'Meara BC. TreePL: Divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics. 2012 Oct 15;28(20):2689–90.

18. Hedges SB, Dudley J, Kumar S. TimeTree: A public knowledge-base of divergence times among organisms. Bioinformatics. 2006;22(23):2971–2.

19. Tao XY, Xue XY, Huang YP, Chen XY, Mao YB. Gossypol-enhanced P450 gene pool contributes to cotton bollworm tolerance to a pyrethroid insecticide. Molecular ecology. 2012 Sep 1;21(17):4371-85.

20. Puinean A, Foster S, Oliphant L, Denholm I, Field LM, Millar NS, et al. Amplification of a Cytochrome P450 Gene Is Associated with Resistance to Neonicotinoid Insecticides in the Aphid Myzus persicae. PLoS Genet. 2010;6(6):e1000999.

21. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2000;290(5494):1151–5.

22. Francis F, Vanhaelen N, Haubruge E. Glutathione S-transferases in the adaptation to plant secondary metabolites in the Myzus persicae aphid. Arch Insect

23. Schuler MA. P450s in plant-insect interactions. Vol. 1814, Biochimica et Biophysica Acta - Proteins and Proteomics. 2011. p. 36–45.

24. Markova-Raina P, Petrov D. High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 Drosophila genomes. Genome research. 2011 Jun 1;21(6):863-74.

Table 3.1. Model summaries. The response variable in both models was the diet breadth of a species, measured as a count of host plant families. (Significance Codes: 0 *** .001 ** .01 *)

General Linear Mixed Model	**Mean Effect**	**Effective Samples**	**pMCMC**
Detoxification Genes	.5970	1000.0	.008**
Sequencing Depth	-.2370	814.5	.232
Phylogenetically Independent Contrasts (PIC)	**Mean Effect**	**Adjusted R^2**	**P-value**
Detoxification Genes	4.111	.1335	.027*
Sequencing Depth	.0033	.0004	.351

Table 3.2. Model data spreadsheet. The insect species transcriptomes are listed along with their TSA database accession numbers, along with their detoxification gene counts and transcriptome summary statistics.

Species	TSA Prefix	Average Seq Length	Total Detox Genes	Host Families	Mbp Length	Number Contigs	Order
Acanthosoma haemorrhoidale	GAUV02	1452.38	429	5	20.8	28218	Hemiptera
Adelges tsugae	GBJX01	2050.09	854	1	34.1	21848	Hemiptera
Bemisia tabaci	GARQ01	1986.15	500	67	19.8	33165	Hemiptera
Trialeurodes vaporariorum	GAWX02	1807.53	456	45	39.3	49895	Hemiptera
Aphis glycines	GBSU01	1771.31	451	4	37.6	55368	Hemiptera
Essigella californica	GAZF02	1749.86	472	1	40.3	56656	Hemiptera
Sitobion avenae	GAPL01	1251.44	574	27	25.8	30678	Hemiptera
Graminella nigrifons	GAQX01	2289.79	816	7	52.1	37537	Hemiptera
Graphocephala atropunctata	GEBQ01	1977.7	1324	5	51.8	32270	Hemiptera
Clavigralla tomentosicollis	GAJX01	831.298	348	5	7	10663	Hemiptera
Nilaparvata lugens	GANM01	1580.28	629	1	33.2	39866	Hemiptera
Kerria lacca	GBDP01	3022.7	1257	22	75.6	56324	Hemiptera
Paratachardina pseudolobata	N/A	1430.74	2035	62	98.91	113670	Hemiptera
Diaphorina citri	GACJ01	3280.88	643	2	52.3	27821	Hemiptera
Lygus lineolaris	GDAW01	962.856	628	18	34.5	62559	Hemiptera
Lygus hesperus	GBHO01	2419.73	1887	23	97.5	45706	Hemiptera
Halyomorpha halys	GDCO01	2986.64	2044	47	239.5	247113	Hemiptera
Planococcus citri	GAXF02	1604.63	806	82	32.1	37149	Hemiptera
Pachypsyilla venusta	GAOP01	2249.06	1211	1	112.3	142265	Hemiptera
Acanthocasuarina muellerianae	GAYY02	1707.15	413	1	30.8	42879	Hemiptera
Nemophora degeerella	GATC02	1700.98	691	1	33.7	37393	Lepidoptera
Telchin licus	GBAS01	817.134	371	4	14.3	23824	Lepidoptera
Ostrinia furnacalis	GAQJ01	1252.14	1172	10	43.4	62382	Lepidoptera
Ostrinia nubilalis	GAVD01	1385.82	750	12	32.1	40907	Lepidoptera
Ostrinia scapulalis	GAHQ01	1197.09	120	5	9.9	11212	Lepidoptera

Dyseriocrania subpurpurella	GASY02	1444.57	536	1	24.1	33856	Lepidoptera
Biston suppressaria	GCJP01	1543.63	1279	15	52.9	69761	Lepidoptera
Triodia sylvina	GAVB02	1631.66	748	5	34.1	41912	Lepidoptera
Polyommatus icarus	GAST02	1354.75	671	2	25.2	31374	Lepidoptera
Lymantria dispar	GDCN01	686.461	565	22	9	20472	Lepidoptera
Agrotis segetum	GBCW01	1287.14	1084	10	43.4	62115	Lepidoptera
Athetis lepigone	GARB01	998.952	745	1	27	41264	Lepidoptera
Helicoverpa armigera	GBXD01	2073.81	1520	26	61.5	37694	Lepidoptera
Helicoverpa assulta	GBTA01	1022.7	1967	6	53.2	64112	Lepidoptera
Spodoptera exigua	GAOQ01	769.378	1588	8	31.8	55375	Lepidoptera
Spodoptera litura	GBZS01	742.989	1322	36	31.6	60192	Lepidoptera
Trichoplusia ni	GBKU01	1828.72	967	10	57.3	70322	Lepidoptera
Parides eurimedes	GAXH02	2040.52	672	1	32.9	30635	Lepidoptera
Amyelois transitella	GDGN01	3442.63	1182	7	111	93073	Lepidoptera
Manduca sexta	GETI01	2655.69	1698	5	243.8	334069	Lepidoptera
Grapholita molesta	GADK01	1259.33	179	1	3	2872	Lepidoptera
Yponomeuta evonymellus	GASG02	1451.44	589	1	30.4	38838	Lepidoptera
Zygaena fausta	GAYB02	1443.53	649	1	27.2	33791	Lepidoptera

Figure 3.1. Phylogenetic relationships among sampled species of Hemiptera (gray branches) and Lepidoptera (green branches). Length of bars extending from tips are proportion to the host-plant range size (pink) and detoxification gene diversity (red).

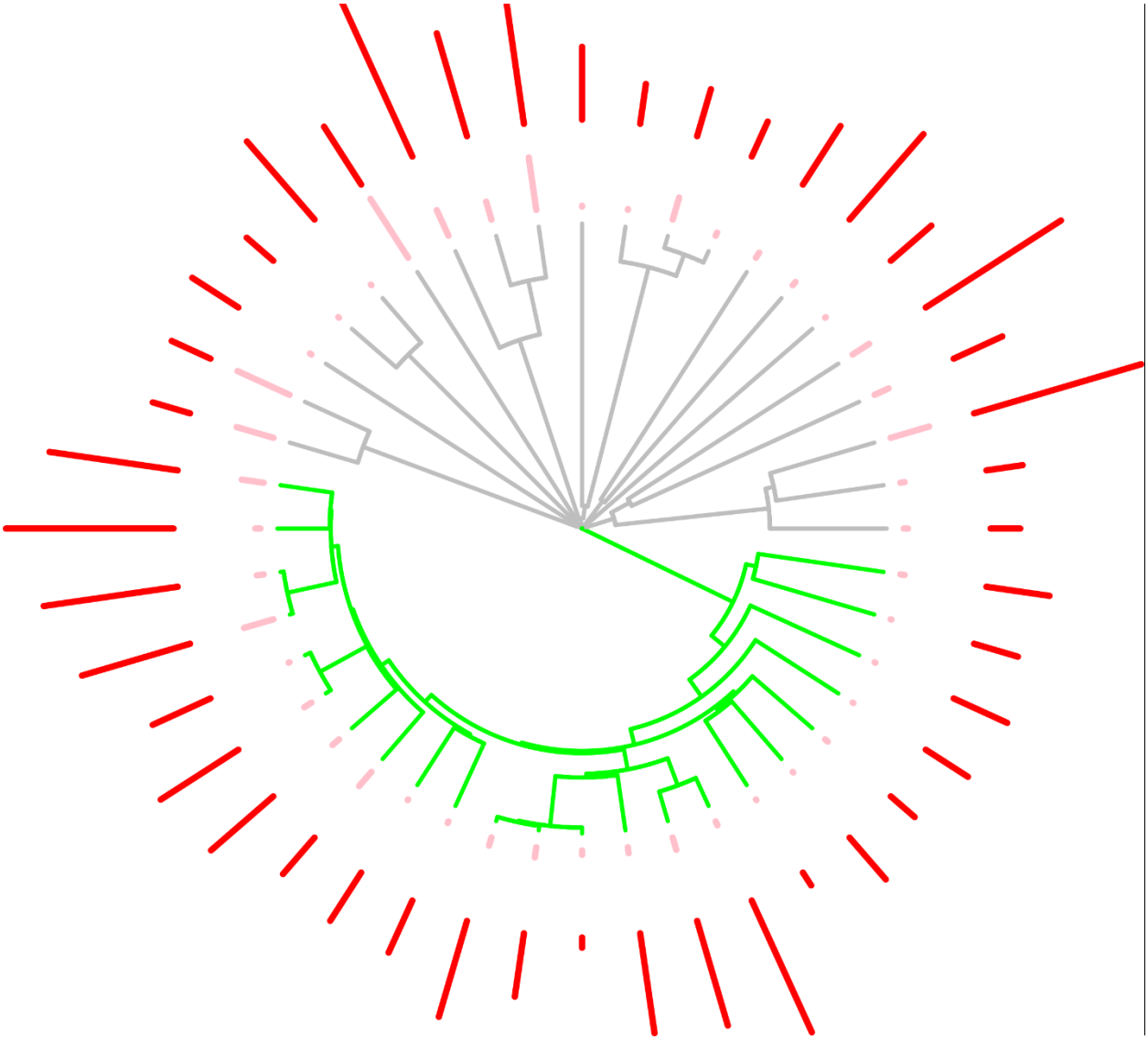
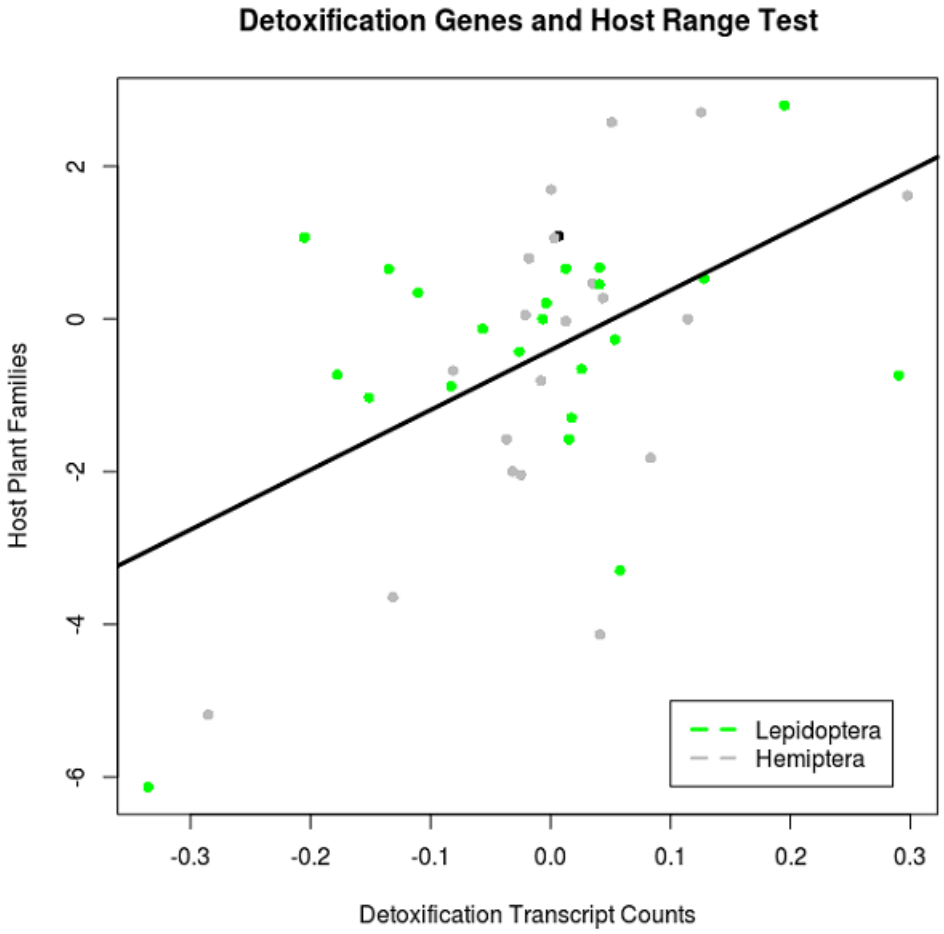


Figure 3.2. Regression of phylogenetically independent contrasts of host-plant range against detoxification gene diversity. Each point represents a pair of contrasts, each between the daughter nodes of an internal node on the phylogeny. Green points are contrasts between lineages of Lepidoptera. Gray points a contrasts between lineages of Hemiptera. The black point is the contrast from the root node, that is, between 1) the most recent common ancestor of all of the lepidopterans in our sample, and 2) the most recent common ancestor of all of the hemipterans in out sample.



Chapter 4: Experimental Evolution of Host-Use Plasticity in Two Aphid Species

Nicholas Christodoulides[1]

[1]Department of Entomology and Plant Pathology, Auburn University, Auburn, AL 36849

Abstract

Gene expression plasticity may be a major mechanism for niche generalism. What role might it play in the evolution of host use? In this section, I outline an experimental evolution study designed to test how phenotypic plasticity evolves in response to host environments that vary over time. Specifically, I plan to select for broader diets in the specialist pea aphid, *Acythsiphum pisum*, by placing them in an environment in which host species change across generations. Conversely, I will use the generalist potato aphid, *Myzus persicae* to select for decreased generalism by maintaining them on only one host for many generations. Selection will be quasi-natural; individuals will compete for host resources, and the best adapted genotypes will increase in frequency over the course of several generations. At the end of the evolutionary period, I plan to compare aphid performance and gene expression from all lines on their ancestral and new hosts.

Background

How will niches evolve in environments that vary over time and space [1]? Most organisms live in environments that vary both over the short and long term and over fine and coarse geographic scales. It seems likely that in these environments, the ability to plastically regulate certain traits could confer an advantage. We know that gene expression in generalist plant-eating insects depends on the host environment, but still understand little about how much of this plasticity is adaptive, and about how important plasticity is to host-use adaptation and the evolution of generalism. What is the role of gene expression plasticity in the evolution of generalism? How heritable is it from one generation to the next? What components of generalist insect performance are affected by evolved plasticity? What genes are important to regulate to increase insect performance and plasticity across hosts?

Gene Expression Plasticity

Phenotypic plasticity is the ability of a genotype to produce multiple phenotypes, depending on the environment [2]. This plasticity can be adaptive (and improve fitness across environmental heterogeneity) or maladaptive (exacerbate a poor match between a genotype and an environment). In the present case, the environmental variation we are interested in is that between the multiple host plant species of a generalist insect. The phenotypes we are interested in are gene expression and over-all performance. And we are especially interested in the gene expression and performance changes induced by exposure to a novel host plant environment. Do generalist species tend to have more adaptive gene expression plasticity than specialists?

Just as the value of a trait can be heritable, so can the average plasticity of that trait, which can change in response to selection at one or more important gene sequences [3, 4]. Moreover, several studies suggest that a shift in the mean value of a trait (up to a certain point) is usually accompanied by an increase in its plasticity [5]. Other researchers suggest that plasticity of a given trait often evolves without a change in the mean value of that trait, and phenotypic plasticity might develop as an alternative to the evolution and possible divergence of fixed phenotypes. Quantitative and genetic models suggest that plasticity should be most favored when habitat variability is high, all habitats are encountered at the same rate, selection is equally strong in each habitat, and different phenotypes are optimal in different environments [6]. But a much broader array of conditions are thought to select for any kind of generalism. Indeed, at the most basic level, density-dependent selection should promote the evolution of broad niches wherever broad niches can occur, that is, wherever environments are heterogeneous. This certainly applies to plant-eating insects. Plant communities are invariably heterogeneous. The question at hand is if, and when host-use generalism evolves in these species, does it depend on phenotypic plasticity, specifically, gene-expression plasticity.

While theories about the evolution of niche breadth and adaptive plasticity are plentiful, empirical data on the subject is lacking, especially on plasticity associated with host use evolution in plant-feeding insects [4]. Selection experiments can provide these data. Indeed, previous selection studies in certain organisms have played a crucial role in furthering our understanding of plasticity and generalism.

Selection Experiment Basics

The objective of a selection experiment is to contrast changes in allele frequencies and phenotype in resulting from alternative selective regimes. This is more tractable for organisms with short generation times, and many plant-feeding insects satisfy that constraint [7]. Some selection experiments involve artificial selection, where the researcher intentionally selects for a particular trait in the evolving population. Alternatively, quasi-natural selection experiments are meant to mimic natural selection in the wild. Members of an experimental population compete for several generations, and the fittest genotypes increase in frequency. Although I will mostly focus on quasi-natural selection, in the past, both types of selection experiments have provided valuable insights into the evolution of generalism and plasticity [4].

Directly Selecting for Plasticity

Past studies have focused on the evolution of insect traits such as size, fecundity, and life cycle duration, but few selection experiments have attempted to select for plasticity *per se*. A notable experiment was done by Scheiner and Lyman [1] on the plasticity of thorax size in response to temperature in *D. melanogaster*, which tend to be smaller at higher temperatures, a common trait in insects. They used flies captured from the wild to establish a genetically heterogeneous base population, and reared experimental lines with selection regimes for increased and decreased plasticity on the trait of thorax size, as well as increased and decreased thorax size, the trait itself, in both high and low temperature environments. All fly lines had their thoraxes measured, and plasticity was calculated as the difference in average thorax length among lines raised. The results showed that plasticity in thorax size in both temperatures does respond to selection with a realized heritability of 0.088 ± 0.027 SE. Additionally, they found that heritability of plasticity

was much lower than the heritability on the mean of the trait of thorax length. They concluded that this discrepancy might be the result of selection mainly acting on only one of the two loci of plasticity, response or magnitude, while selection on thorax size would inherently involve both.

Indirectly Selecting for Plasticity

Phenotypic plasticity has also been indirectly selected as a means to generalism. Harshman et al. [8] conducted a study on the expression of detoxification enzymes in *D. melanogaster* across multiple hosts. They reared a colony from wild-caught flies, and formed six lines of flies for the study; three lines that stayed on the standard original medium they started on, and three lines selected to subsist on lemon. They allowed these lines to propagate for around 20 generations. They hypothesized that plastic individuals more adapted to lemon would express more detoxification gene activity due to either bacteria or chemicals present in the fruit. To select for the lemon host lines, they put flies in bottles with 10g of lemon for 7-10 days, and after around half of the flies in the current generation died, surviving flies were randomly placed on more freshly cut lemon in a new bottle to reproduce the next generation. At the end of 20 generations, all six lines were reared on the original common substrate for one generation to control for their original environment before testing. Afterwards, young flies from this generation were transferred to either lemon or additional normal medium for one day before their enzyme activities were measured. The researchers measured the activity levels of epoxide hydrolases and glutathione S-tranferases (GST), and found that their expression differed little between the non-selected and lemon-selected lines on the original substrate for 24 hours, but lemon selected lines had significantly greater detoxification enzyme activity than the control lines when placed on lemon prior to sacrifice. The authors noted that there was indeed ‘a pronounced change in environment-dependent expression’ that ‘appeared independently in all three lines on lemon.’

This study suggests that environmentally affected enzyme expression plasticity can evolve in a relatively constant selection regime designed to evolve generalism.

On the other hand, we also have experimental evidence that generalism can evolve without increased plasticity. Bennett and Lenski [9] examined thermal tolerance, and plastic responses to thermal stress, in the bacteria *E. coli*. From a specialized bacterium that evolved in a constant 37°C, they selected for thermal-generalist lines by placing them in variable thermal environments between 32 and 42°C and allowing them to reproduce for 2000 generations. They found that overall fitness and resource use efficiency of bacteria in the variable environment improved in temperature-fluctuating environments, but not due to increases plasticity. Additionally, they concluded that the specialized bacteria did not display any notable increase in resource use or performance in the ancestral environment compared to the more generalized lineages, so there is no apparent tradeoff associated with thermal niche generalism in these bacteria.

The studies reviewed here give us some insight into the role of plasticity in niche breadth evolution. But they leave us with only a vague idea of the specific genetic changes that occur during niche evolution. And we are not sure what to expect in plant-feeding insects. [10, 11, 12]. Today we have a variety of next-generation sequencing technologies and techniques for studying the expression of large numbers of genes in response to selection. Here, I propose a selection experiment which draws from previous selection studies, to test the idea that plasticity is a crucial part of the evolution of generalism. I also outline my methods for incorporating this experiment with modern techniques of RNA-Seq and transcriptome assembly, to provide rigorous genetic analysis on the plasticity of gene expression associated with evolved host use in insect populations.

Proposed Methods

Selecting for Increased and Decreased Plasticity

The selection experiment will take place in a greenhouse environment using two aphid species: the pea aphid *Acythsiphum pisum,* and the peach-potato aphid *Myzus persicae*. A stock population of pea aphids will be reared on *Pisum sativum* for 75 to 100 generations exclusively, while a stock of *M. persicae* will be assembled from several wild populations and maintained on a mix of host plants (species composition varying over time and space). A total of six lines, three control and three plasticity selected lines consisting of around five hundred aphids, will be founded for each aphid species, for a total of 12 lines (Fig. 1). In the generalist, *Myzus persicae*, we will select for specialism by raising them on peas (*P. sativum*) for 50 generations. While the selected lines are adapting to peas, control lines will be rotated between four hosts: Foxglove, Squash, Salvia and Cabbage – changing the host plant species once every three generations, long enough to give the aphids time to manipulate their transient gene expression patterns to match their current host, but not become fixated. These likely hosts will be part of a preliminary screening for aphid use along with ten other plants that may have interesting defensive chemistry, in order to make sure that the aphids can use them before the experiment is carried out and change them if needed. In specialist *A. pisum*, we will select for diet broadening by exposing them to a different legume species every three generations. We will also include a second, non-legume host in each insect cage, on the chance that tolerant genotypes will get a fitness boost by feeding on the alternate host when competition on the legume is high. Control lines will be kept on peas. Selection will be applied for 50 generations. We will cycle through hosts by using insect cages large enough to contain two host plants. At the start of an experiment, the cage will contain

one plant. Aphid populations will build up on that plant for two generations, then a second, new plant will be placed in the cage. After another three generations and migration onto this new host, the original plant will be removed and another new plant will be pushed into the cage.

Measuring Aphid Performance

For our assays of performance, we borrow from the design of Harshman et al. [8]. After a preliminary assessment of performance and subsequent selection for 50 generations, the aphids from both species and all twelve lines will be reared for one generation on peas, the original specialist host, so as to iron out any effects on performance that stem from the maternal environment (Fig.1). Then the progeny of the mothers raised on peas, will be divided equally across two assay environments. One environment will be a completely new host (i.e., a host that was not encountered during the selection stage of the experiment), milkweed for *M. persicae*, and tobacco for *A. pisum*. What we would like to know is whether or not selection for increased or decreased host-use generalism affects performance and gene expression on a host outside of the set of historical hosts, and possibly its ability to become invasive.

A second environment will be one of the original hosts that the stock populations of each aphid species was kept on. This will let us see if adaptation to the experimental host environments came with a cost on performance on the initial hosts. For the pea aphids this will be peas, and for the peach aphids this will be one of the rotating stock hosts (e.g., squash). Performance will be measured as a function of generation time (i.e., degree days from nymph to nymph) and fecundity (number of proliferating aphids assessed periodically). Aphids will be considered to be specialized if they perform better on only one plant, but decrease in performance across other plants,

Analyzing Plastic Gene Expression

A subset of the aphids transferred to the testing environments described above will be assayed for gene expression differences, rather than performance differences. Around twenty adult

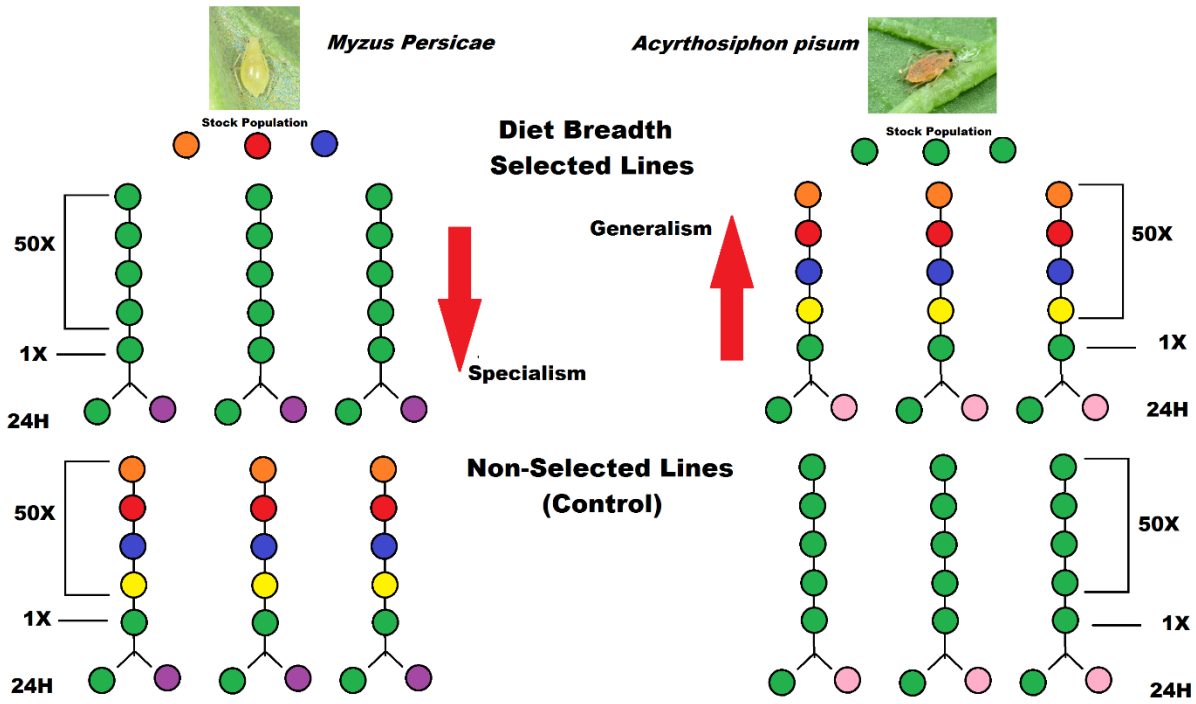


Fig. 4.1. Selection regime and sampling plan for experimental evolution of niche breadth in *A. pisum* and *M. persicae* aphids. In *A. pisum* we are selecting for a broad diet while in *M. persicae* we are selecting for a narrower one (in comparison to ancestral diets). Circles represent host plants that the aphids will be reared on as follows: Green circles are peas, orange circles are squash, red circles are foxglove, blue circles are salvia, and yellow circles are cabbage. Purple circles represent milkweed and pink circles represent tobacco, the new hosts to *M. persicae* and *A. pisum* respectively that will be used to test plasticity in the face of a completely new stimulus.

females from each testing environment will be collected and pooled into a single sample in Trizol for RNA extraction for an N size of 24 samples, using the Purelink RNA Mini Kit, with DNA removed using the DNAase kit [13]. Samples will be sequenced on a Illumina HiSeq3000 generating paired-end 100 BP reads at a depth of 20 million reads per sample, so around 480 million raw sequence reads. Since a reference genome for *A. pisum* has already been constructed, our reads from both species will be aligned to it using Bowtie and transcript counts performed using eXpress [14, 15, 16]. After constructing the counts matrix for transcripts, I will use the qualitative comparison functions in edgeR and the gene enrichment functions of to find what kinds of genes are significantly up or downregulated across hosts in lines selected for increased and decreased plasticity [17, 18]. This gene expression will be compared to preliminary gene expression from samples taken before selection on both species of aphids. I will also be able to use the quantitative models included in the package to see if any of these enriched genes are correlated to increase or decrease in individual performance, both in egg sac mass and generation time.

Conclusion

The variation in niche breadth across species is striking. Evolutionary biologists have many ideas for why that might be, some of which are incompatible with others, and most of which have not been tested empirically. Selection experiments are onerous to design and can take a long time to carry out. This may be why so few have looked at the evolution of niche breadth. But when paired with RNAseq, they have the potential to yield high-resolution insights into the process of niche breadth evolution. .

References

1. Scheiner SM, Lyman RF. The genetics of phenotypic plasticity. II. Response to selection. J Evol Biol. Blackwell Science Ltd; 1991 Jan 1;4(1):23–50.

2. Garland T. Phenotypic plasticity and experimental evolution. J Exp Biol. 2006;209(12):2344–61.

3. Via S, Gomulkiewicz R, De Jong G, Scheiner SM, Schlichting CD, Van Tienderen PH. Adaptive phenotypic plasticity: consensus and controversy. Trends Ecol Evol. 1995 May;10(5):212–7.

4. Scheiner SM. Selection experiments and the study of phenotypic plasticity. Vol. 15, Journal of Evolutionary Biology. Blackwell Science Ltd; 2002. p. 889–98.

5. Gillespie JH, Turelli M. Genotype-environment interactions and the maintenance of polygenic variation. Vol. 121, Genetics. 1989. p. 129–38.

6. Scheiner SM. Genetics and Evolution of Phenotypic Plasticity. Annu Rev Ecol Syst. Annual Reviews. 1993 Nov 28;24(1):35–68.

7. Gabriel W. How stress selects for reversible phenotypic plasticity. In: Journal of Evolutionary Biology. Blackwell Science Ltd; 2005. p. 873–83.

8. Harshman LG, Ottea JA, Hammock BD. Evolved environment-dependent expression of detoxification enzyme activity in Drosophila melanogaster. Evolution (N Y). 1991 May 1;45(3):791–5.

9. Bennett AF, Lenski RE. Experimental evolution and its role in evolutionary physiology. Am Zool. Oxford University Press; 1999;39(2):346–62.

10. Schnee FB, Thompson Jr. JN. Conditional neutrality of polygene effects. Evolution. 1984 Jan;38(1):42–6.

11. Cavicchi S, Guerra D, Natali V, Pezzoli C, Giorgi G. Temperature related divergence in experimental populations of Drosophila melanogaster. II. Correlation between fitness and body dimensions. J Evol Biol. Blackwell Science Ltd; 1989 Jul 1;2(4):235–51.

12. Weber SL, Scheiner SM. The genetics of phenotypic plasticity. IV. Chromosomal localization. J Evol Biol. Blackwell Science Ltd; 1992 Jan 1;5(1):109–20.

13. Rna P, Kit M. PureLink TM RNA Mini Kit. Cell. 2010;(12183020):1–4.

14. Ramsey JS, Rider DS, Walsh TK, De Vos M, Gordon KHJ, Ponnala L, et al. Comparative analysis of detoxification enzymes in Acyrthosiphon pisum and Myzus persicae. Insect Mol Biol. Blackwell Publishing Ltd; 2010 Feb 23;19(SUPPL. 2):155–64.

15. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol. 2011;12(3):R22.

16. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.

17. MD R, DJ M, GK S. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26:139–40.

18. Alexa A, Rahnenfuhrer J. topGO: enrichment analysis for gene ontology. Vol. 2, R package version. 2016.