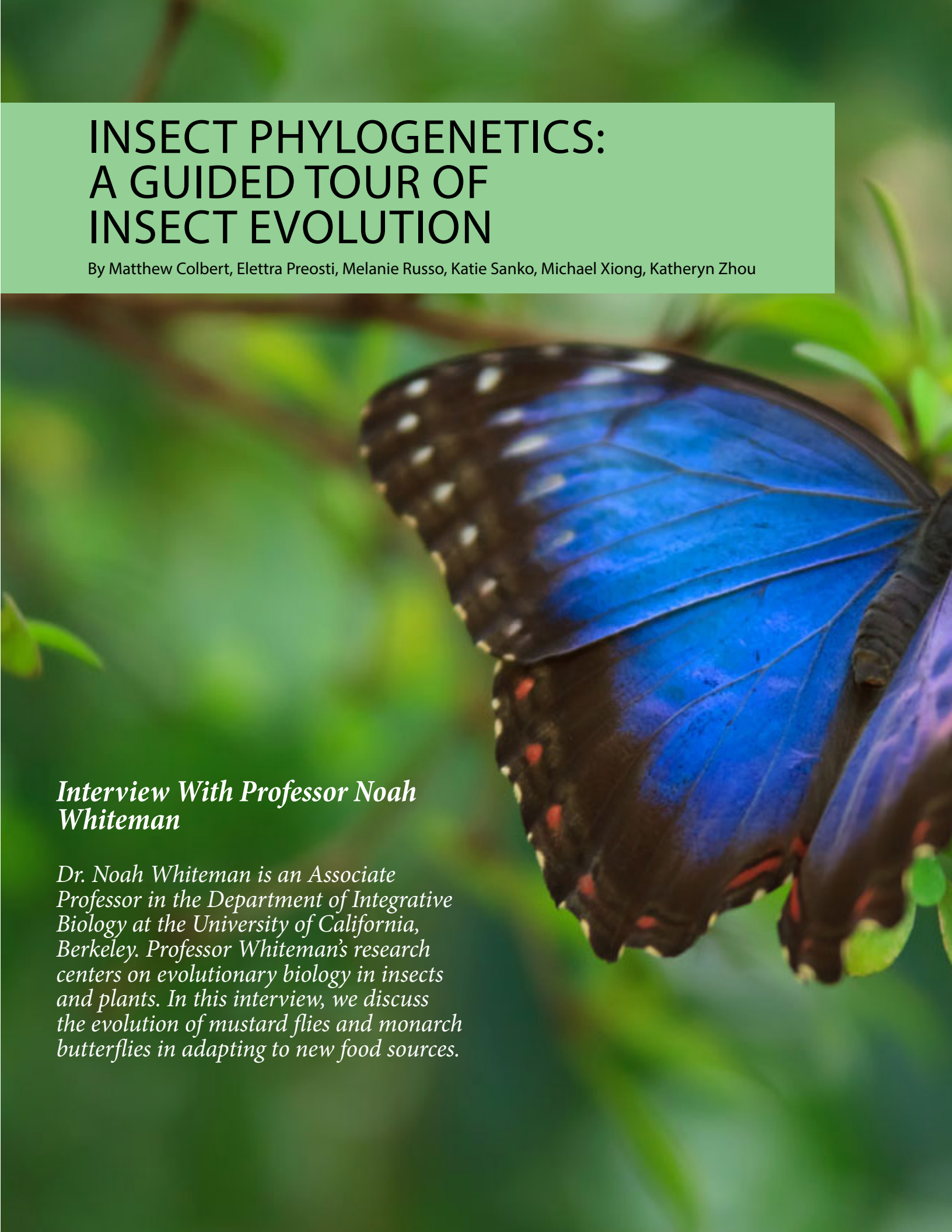
INSECT PHYLOGENETICS: A GUIDED TOUR OF INSECT EVOLUTION

By Matthew Colbert, Elettra Preosti, Melanie Russo, Katie Sanko, Michael Xiong, Katheryn Zhou



Interview With Professor Noah Whiteman

Dr. Noah Whiteman is an Associate Professor in the Department of Integrative Biology at the University of California, Berkeley. Professor Whiteman's research centers on evolutionary biology in insects and plants. In this interview, we discuss the evolution of mustard flies and monarch butterflies in adapting to new food sources.

BSJ: What led you to pursue a career in evolutionary biology and what challenges have you faced?

NW: I was an undergraduate at a small liberal arts college in Minnesota, and I was pre-med because that was the option that I thought existed for someone interested in biology. I took an entomology course, and I secretly loved insects but I didn't want to tell anyone. That really changed my view of what I wanted to do. It was around then that I realized that someone has to be the professor, someone who doesn't just regurgitate knowledge but generates it. I was kind of battling with myself, thinking, "Who am I going to let down by saying I don't want to be a doctor anymore?" Everyone in my family had been expecting that. It's the first normal break with family that hopefully every person makes when they realize that it's their life and not their parent's life. It's your life, you're an adult, and you get to decide what to do with your life, not your parents.

I didn't realize that there was an honors program at the college until my friend, who was in it, encouraged me to apply and do an honors thesis. In my junior year, pretty late, I asked the person who taught the entomology course if I could do an honors thesis with him, and he agreed. It was on an existing project on social wasps. They have a very interesting social system like honeybees, where there's a queen that's reproductive and workers that are sterile. To me, it was fascinating to think about the evolution of these insects, and that kind of opened the door. I thought, I might not get into medical school, although at that point I didn't really want to anymore. But then I thought I might not be able to pass the GRE. I took it and sort of walked out of the quantitative section; I had a math phobia. I got rejected from every school I applied to for PhD programs in insect evolution, but I got into one master's program at the University of Missouri, Columbia, and I thought, OK, that's what I'm going to do.

During my first semester into my PhD program, I decided that I didn't want to be in the lab I was in. I thought about dropping out. I came out of the closet then too, so there was a lot of tumult going on in my life. A professor named Patty Parker knew what was going on with me and said, "Why don't you come to Galapagos with us? We're starting a new research program there on disease ecology with birds." So, I started on a project working on the Galapagos hawk. We found that feather lice are transmitted from mother to baby like genes are, so they get their initial dose of lice from their mother. I did my dissertation on this, and we used the lice as a marker of the hawks' colonization history by studying the genetics of the lice. So that's how I slowly, slowly got more interested in evolutionary questions.

BSJ: Your current studies use phylogenetics to examine broad evolutionary questions. Could you explain what phylogenetics is and how you use it to understand all of these processes?

NW: Phylogenetics estimates the evolutionary relationships among species: who's related to whom. You use homology, which is the shared ancestry and single origin of a trait. In this case, we use DNA sequences to infer the evolutionary history of any trait or gene. You need as detailed a phylogeny as possible to reconstruct the evolutionary steps for a particular trait. You also need the biological background of the trait, because it can be complicated due to hybridization. Species sometimes interbreed and leave traces of their genome, which confuses the evolutionary trees of phylogenies, as you might imagine. You want a majority tree for the species, but each gene has its own evolutionary history, and the ability to infer that history gets confounded by things like natural selection, convergence, gene loss, and gene duplication.

Professor Noah Whiteman[1]

Roughly speaking, you can obtain the phylogeny of any group of organisms. That's a starting point for asking questions about evolution, at least at the macroevolutionary scale. Phylogenies mostly tell you fixed differences between species, not how the process of evolution works. For that, you need information on what's going on within a species now. Natural selection works and operates on dynamic genetic variants that emerge. The idea is to link the population's genetic microevolution processes, which we can study here and now, to the phylogenies that are macroevolutionary and between species.

BSJ: We read your paper on horizontal gene transfer, "Horizontal transfer of bacterial cytolethal distending toxin B genes to insects."[2] Could you describe what horizontal gene transfer is and how it occurs?

NW: I'm the senior author on it, but it's really part of Kirsten's dissertation, and was a collaboration with other professors at other places, including Jennifer Wisecaver who's at Purdue University, Donald Price who's at the University of Nevada, Las Vegas, and students or former students as well.

Imagine that a fly is munching on a leaf as a larva, and a wasp comes up and injects an egg into the larva. When it does that, maybe a virus gets injected as well, and it has the ability to integrate itself into the genome of the fly larva. If the virus per-

"When you build a phylogeny of the thousands of *cdtB* sequences from the thousands of bacteria in GenBank, the closest relative of the sequence in the flies is the one from *Hamiltonella defensa*. Clearly that lineage is insect-associated in some way, and is moving around."

sists and is transmitted to the fly's babies through the germline genome, then it becomes a case of horizontal gene transfer. This is one way that new genes arise in the lineage, and it also complicates phylogeny because it's not a reflection of vertical transmission, which is parent to offspring with no outside genetic information. It's a gene that moves between lineages of the phylogenetic tree. Horizontal gene transfer is rampant in bacteria; they take up DNA from the environment all the time. In animals, it's pretty rare to observe horizontal gene transfer resulting in a new function. Nancy Moran found in pea aphids—some of them are red and some of them are green—that the red ones actually have a fungal gene that encodes carotenoids and gives the aphid the ability to be red. That's a good example of how a horizontally transferred gene can result in a new function.

BSJ: Insects don't normally produce cytolethal distending toxin B (*cdtB*). Can you explain what role *cdtB* normally has in bacteria and how you traced its transfer into your flies?

NW: When you sequence a new genome of any animal species, you first try to find out what genes are not animal in origin. We studied a fly that transitioned from feeding on rotting fruit to living leaves. We thought these flies might have horizontally transferred genes that allow them to live on this plant leaf. We ran every part of the genome through an index that builds a phylogeny of every gene and identifies its closest relatives. When we did that, we got exactly one non-animal hit: a gene called *cdtB* encoded in a bacteria, *Hamiltonella defensa*, and a bacteriophage. The protein cdtB encodes is called cytolethal distending toxin subunit B. The B subunit of this three-part toxin is an enzyme that cuts DNA, which kills a cell. We've probably all had cdtB in our bodies; it's a marker for irritable bowel syndrome in humans. The cell goes through the apoptotic cycle and blows up, which is why it's called a cytolethal distending toxin.

We searched for *cdtB* using BLAST in GenBank, and we found it in two other fly lineages and in the green peach aphid. Among other *Scaptomyza* fly species, *cdtB* is located in the same position in the genome, flanked by two conserved genes, indicating this iteration had a common origin. We also found it in an unrelated *Drosophila*, *Drosophila ananassae*, and some of its relatives. If you put it all together, it seems like there were at least three, but probably four independent ancient transfer events into these insects. When you build a phylogeny of the thousands of *cdtB* sequences from the thousands of bacteria in GenBank, the closest relative of the sequence in the flies is the one from *Hamiltonella defensa* (Fig. 1). Clearly that lineage is insect-associated in some way, and is moving around, maybe through phage, and you can imagine the horizontal transfer events required for that to work.

We think that the insects are deploying this toxin to kill parasitoid wasp eggs themselves. We don't know how it's deployed. We think maybe through immune cells. When the parasitoid wasp egg gets injected into the insect, its immune cells surround the egg and melanize it. They turn the egg black, seal it off, and kill it. But some of these flies that have cdtB kill wasp eggs in a non-melanin-dependent manner. Our hypothesis is that they are somehow using this toxin to do that. To test this, we generated flies that have cdtB knocked out, and we're currently working with those.

BSJ: In your paper on the evolution of herbivory, you describe the coevolution of insects, mustard plants, and *S. flava*. Could you describe how the exposure to plant toxins drives the diversification in the *flava* species?

NW: Let's define coevolution first. The broadest definition in the context of plant-insect coevolution includes the overall interactions going on between plants and in-

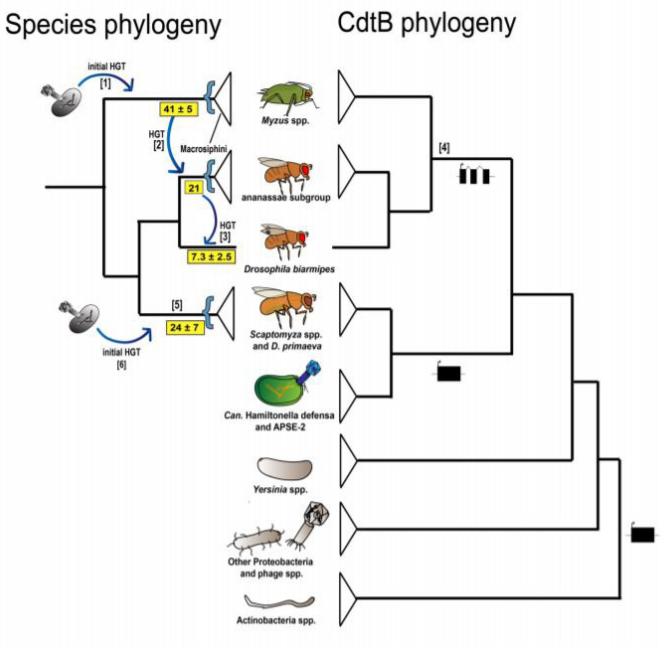


Figure 1: *Simplified paired CdtB and species phylogenies. Arrows point to potential horizontal gene transfer events.*[2]

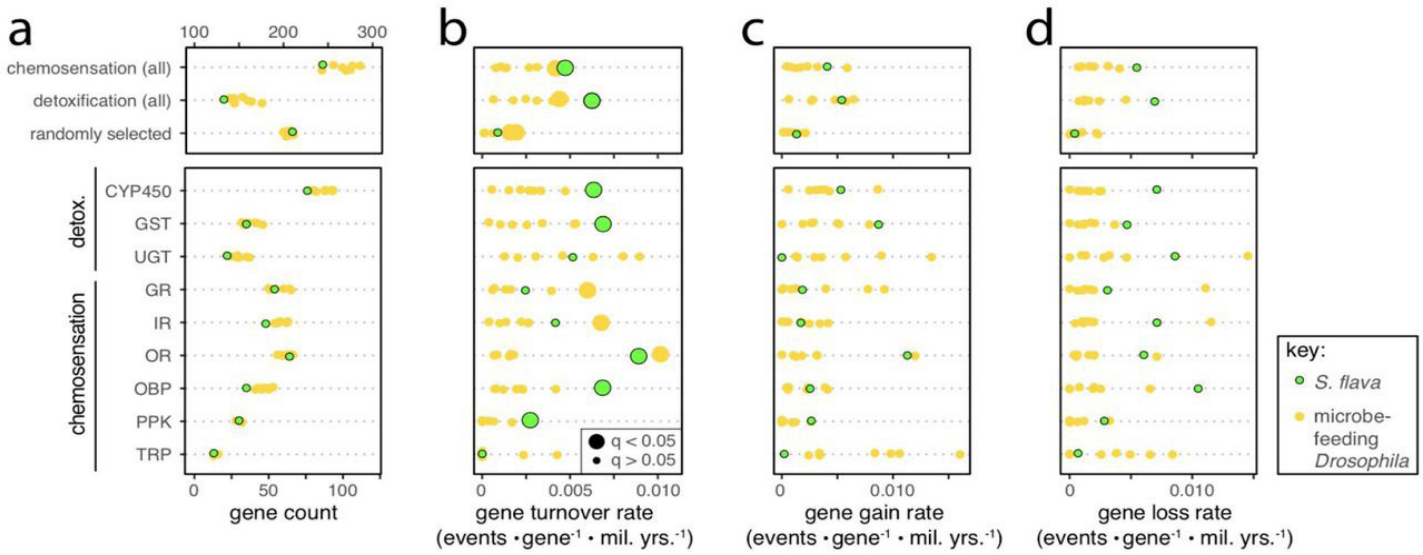


Figure 2: *Gene turnover rates of detoxification and chemosensation genes among S. flava and microbe-feeding Drosophila.*[3]

sects: insects trying to eat plants, and the plants trying to kill the insects in return. Half of all insect species that are living right now are herbivorous, meaning they feed on living plants only. Herbivorous insects make up about a quarter of all named species of eukaryotic life, which is a lot. So why is the world still so green? Well, the plants are trying to kill those insects. Herbivorous insects are very successful in part because the vast majority of them are specialized to a particular set of plants. For example, monarch butterflies are specialized on milkweeds, and they'll eat any milkweed. One hypothesis put forth by Peter Raven and Paul Ehrlich in the sixties was the idea that plants are evolving in response to the insects that are attacking them. If a plant evolves a new defense chemical, that will give it a competitive advantage compared to other plants, and it will spread around the landscape, increase its fitness, and become more diverse. The insects will eventually overcome those specific defenses. The insects become good at detoxifying the chemical, so then the plants ratchet it up. That's called the escape and radiate hypothesis, to explain the diversity of plants and herbivorous insects today. That's broadly what coevolution is in the context of plants and insects.

To answer your question about our paper, we know that mustard plants have been around for 100 million years. The *Scaptomyza* flies that feed on the mustards have only been around for 10-15 million years. This paper asks how these insects deal with toxins when they colonize these mustard plants. The mustard flies feed only on mustards and they're really good at it, but they don't have an efficient way of detoxifying mustard oils. Mustard oils are also super toxic to the plant, so, smartly, the plants keep the two oil precursor components separate in the cell. Our flies have adopted the same way that we detoxify oils: detoxification enzymes. We discovered that over evolutionary time, one particular glutathione S-transferase (GST) got turned into five copies through gene duplication. Three of these new GSTs are really good at detoxifying mustard oils and one is the best of any GST that has ever been studied in animals. Previously, the paradigm was that all these mustard oils specialists prevent the oil bomb from going off, so they don't interact with mustard oils at all. But our flies have found a way around it that's good enough, and it's through a gradual adaptive process, not this big leap change of a brand-new gene coming in from somewhere (Fig. 2).

> "But our flies have found a way around it that's good enough, and it's through a gradual adaptive process, not this big leap change of a brand-new gene coming in from somewhere."

We also think we've found out how the flies are attracted to mustard plants, and it's a story of gene duplication and neofunctionalization. As gene families copy themselves or are lost, they alter the olfactory receptor produced, changing what the fly will respond to. We found the first odorant receptor in flies that is co-opted to be sensitive to volatile mustard oils. The flies had co-opted an old gene and completely changed the function of the gene to find mustard oils rather than a set of ligands that are present in rotting fruit. For that, we used something called the empty neuron mutant. A native olfactory receptor gene is normally expressed in a particular neuron in the fly, which is located on the third antennal segment in a sensilla, or a hair. Scientists can easily manipulate the receptors in these hairs to test responses to different stimuli. We stuck the candidate gene that was important in finding mustard oils into our fly. Then, we used the tools of *Drosophila* to figure out what the function of that gene

is and what it's responding to. We screened all these chemicals, and we couldn't find any that we thought it might be responsive to. In a last-ditch effort, we tried mustard oils, and unexpectedly, it worked! It seems obvious in hindsight, but we thought it would be tuned to more general plant smells and not just mustard oils.

BSJ: So, about your most recent publication on the evolution of monarch butterflies…

NW: First, I would like to give credit to two postdocs who worked on this project, Marianthi Karageorgi, who underwent a Herculean effort to complete all of our fly phenotyping in a year, and Niels Groen, who is a co-first author on the paper. My colleagues Anurag Agrawal and Susanne Dobler initially found these convergently evolved substitutions in the sodium pump of insect species that feed on milkweeds and foxgloves. This suggests that there's an adaptive value to those substitutions, which we think is relatively rare. You have independently evolved insects that feed on toxic milkweeds or foxgloves, and they don't always have the resistant mutations. Thus, these mutations are not the only way up to this adaptive peak. In fact, there are probably peaks in phenotype space. This one was the route we chose to investigate. At the time, I was a new professor at the University of Arizona, and my friend and I thought we should write a "News and Views" piece about this for *Nature*. In Agrawal and Dobler's paper in *PNAS*, they found that these sodium pump mutations were repeated at several positions, especially at positions 111 and 122 in the first extracellular loop of the sodium pump. The mutations had evolved from a conserved amino acid residue (Fig. 3). There was some evidence that monarchs and beetles, which both had the mutations, might have conferred resistance to the toxin. The potential mechanism is target site insensitivity, where the toxin binds to a particular spot on that pump, preventing the pump from working. The sodium pump is really, really important. Three quarters of the ATP in our brains is being used by the sodium pump right now! Messing with it, even a single amino acid changes, has major consequences. At the time of the paper, CRISPR had just been announced as a tool. After reading the paper, I naively said that someone needs to test these mutations to see if their gain-of-function in a species that doesn't feed on milkweed is sufficient for resistance. I thought it would be easy to do and it was obvious; take the conserved sodium pump in the fruit fly and change it into the monarch one. For Drosophila, many tools to do this already exist, so you wouldn't necessarily have to use CRISPR. Agrawal, an author on the *PNAS* paper, read our News and Views, and he asked me if I would like to do just that. I was initially unsure, but I thought it would be unwise to turn down a grant if it got funded, even for a side project. And it eventually did get funded.

We originally decided to try a one-step CRISPR approach, where we would try to edit the gene directly with no additional marker. That did not work—any kind of perturbation to the pump turned out to be really difficult for the flies to handle. We failed for about two and a half years, and I was ready to give up on the project. However, my postdoc, Niels Groen, asked to try one more time. We had a new strategy: a two-step CRISPR, in which we mutated the sodium pump and additionally knocked in a green fluorescent protein (GFP) fused to a gene that's expressed in the flies' eyes. Thus, we could see when we got a deletion line in the region we wanted to edit, as those flies would now have green eyes.

Eventually, we got homozygous viable mutants for all of our genotypes of interest. The single mutants (leucine (L) at 111, serine (S) at 119 and histidine (H) at 122) revealed that S is neutral, but provides some resistance to animals. L and V also provided some resistance, but caused neurological damage in the flies unless paired with the S mutation. The H mutation causes a lot of damage, but grants a lot of resistance to mustards. That's why in our other insects that evolved to feed on mustards the H always appears with the S, to mitigate its neurological impacts. That suggests there's a constraint in the adaptive walk. Think about this as base camps on the way up to Mt. Everest. You have to go through each base camp to get to the last one. There are other peaks or different solutions at the end, so it's not the only way, but it's the one that was taken multiple times. We built a fly with those three mutations that is as resistant to the cardenolides as the monarch is at the physiological level. We took our fly and butterfly brains, ground them up, and ran an assay that allowed us to isolate the activity of the sodium pump itself. The monarch flies and the monarch butterflies basically have an identical kinetic response, meaning those three mutations, VSH, are important and provide the most resistance in the assay.

Then we confirmed this with cell line experiments, where our collaborator Susanne Dobler created the same mutations in moth cells and found the same thing. That's why gain-of-function studies, in my mind, are easier than loss-of-function ones. There's

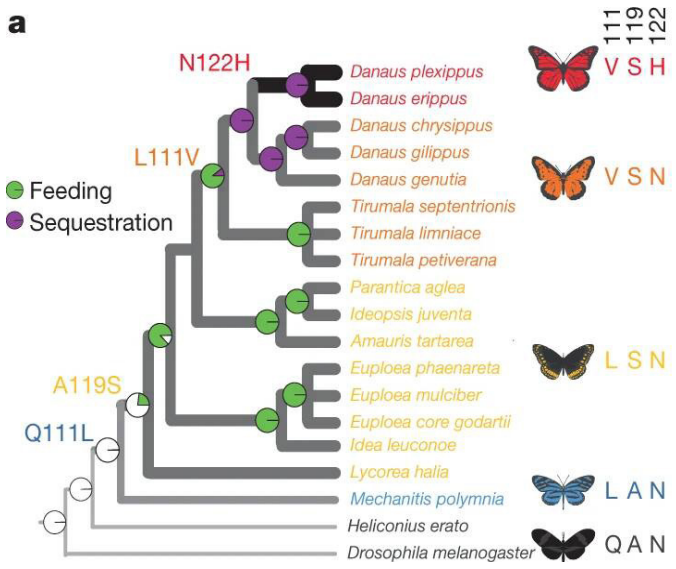


Figure 3: *Phylogeny of monarch resistance to milkweed toxins. Amino acid positions 111, 119, and 122, and the mutations in different butterflies, are shown. Feeding and sequestration nodes indicate whether that mutation genotype fed on milkweed and didn't interact with the toxin or sequestered the toxin away.*[4]

"We built a fly with those three mutations that is as resistant to the cardenolides as the monarch is at the physiological level."

lots of ways to break something, but to "make" something is a lot more stringent. It took about seven years to make these mutations. I think CRISPR has opened up an avenue to test multistep adaptive walks so we can reconstruct evolutionary history *in vivo*, not just in a test tube. I think that's the exciting thing for me when we got involved in this project—we're able to reconstruct evolutionary history and ask why something evolved the way it did. Furthermore, we had to have the whole animals to test that, because in a cell line, you're not going to have the neurological phenotype. The pumps seemed fine, but clearly in the whole animals it was not working as well. The best part is we put the flies on a milkweed diet. The VSH flies retained the toxin in their bodies through metamorphosis, like a monarch butterfly does when it becomes orange. The color warns predators to leave them alone. And why are the monarchs toxic? They retain toxins from their larval diet. How do they do that? They need VSH to be able to concentrate the toxin at high levels. So, our study helped understand if VSH opens the door to passive accumulation of the toxin through metamorphosis, as flies have complete metamorphosis just like butterflies.

BSJ: We wanted to expand a little on what you were talking about, because a lot of it sounds like bioengineering and genetic engineering. When we think about GMOs, we mostly think about genetically modified crops and some of the controversy surrounding them. What do you think about the potential applications of gene editing animals and humans possibly?

NW: Well, I completely agree with the moratorium on genome editing in living humans, period. I do think that it should be used for crop improvement. Humans have been selecting natural mutants for millennia, and recently, a lot of the crops that we use are the result of mutagenesis experiments. And people are happy to eat those! If you think about how mutations work, every mutation that can be tolerated by an individual is already out there, segregating at a low level. Think of a corn field. There are mutations at every single base pair that could be tolerated if the population is over a certain size, and even in a single field you'll have a large population. I think for humans and biomedicine, like everything in that realm, it will take a lot more study and careful regulation before we use genome editing for treatment. There should be a moratorium on editing human germline cells and embryos using CRISPR. CRISPR could be used to treat conditions that are genetic disorders, like muscular dystrophy or sickle cell anemia, in a way that doesn't involve germline transformation. But our study is a cautionary tale—we had a lack of viability in a lot of our transgenic flies, and we don't know why.

BSJ: Any closing remarks about science or research from your perspective?

NW: Follow your passion and ignore everyone's advice. March to the beat of your own drummer. You have got to believe in yourself and you have to have a network of people who will believe in you. No one tells you what to do in terms of research. That's the best part of the whole thing: nobody tells us what to study. If we can get funding for it, we can do it, provided it's ethical. I think the discoveries that you can make as an individual now are just incredible, even compared to when I was a PhD student. A lot of students around here want to go to medical school, but I'm really glad I made the decision I made, even though it's less financially lucrative. It turns out you only need a certain amount of money to be happy, and it's less than what doctors make.

REFERENCES

1. Noah Whiteman [Photograph]. Retrieved from: http://www.noahwhiteman.org/principal-investigator-dr-noah-whiteman.html
2. Verster, K.I., J.H. Wisecaver, R.P. Duncan, M. Karageorgi, A.D. Gloss, E. Armstrong, D.K. Price, A.R. Melon, Z.M. Ali, N.K. Whiteman (2019). Horizontal transfer of bacterial cytolethal distending toxin B genes to insects. *Molecular Biology and Evolution 36*: 2105-2110. doi: 10.1101/544197
3. Gloss, A.D., A.C. Nelson Dittrich, R.T. Lapoint, B. Goldman-Huertas, K. I. Verster, J.L. Pelaez, A.D. L. Nelson, J. Aguilar, E. Armstrong, J.L.M. Charboneau, S.C. Groen, D.H. Hembry, C.J. Ochoa, T.K. O'Connor, S. Prost, H. Suzuki, S. Zaaijer, P.D. Nabity & N.K. Whiteman. Evolution of herbivory remodels a *Drosophila* genome. *bioRxiv* [Preprint]. doi: 10.1101/767160
4. Karageorgi, M., S.C. Groen, K.I. Verster, J.M. Aguilar, F. Sumbul, A.P. Hastings, J.N. Pelaez, S.L. Bernstein, T. Matsunaga, M. Astourian, G. Guerra, F. Rico, S. Dobler, A.A. Agrawal & N.K. Whiteman (2019). Genome editing retraces the evolution of toxin resistance in the monarch butterfly. *Nature 574*: 409-412. doi: 10.1038/s41586-019-1610-8