Kelsey Byers:
Hi there, my name is Kelsey Byers. I'm currently a group leader at the John Innes Center in Norwich, United Kingdom. And the work I'm going to talk about is actually from my second postdoc in the butterfly genetics group at the University of Cambridge. My contact information is at the bottom of the slide. It's Kelsey.buyers@jic.ac.uk, or you can reach me on twitter @plantpollinator.

I'm going to talk to you today about the genetic basis of wing and genital sets in heliconius butterflies. The slide shows a pair of two different species of Heliconius butterflies mating with some chemicals that they use during the mate choice.

So just a couple of slides about me. I did my PhD at the University of Washington in the States from 2008 to 2014. I then did a postdoc from 2014 to 2017 at the University of Zurich in Switzerland, that a second postdoc from 2017 to 2020, at the University of Cambridge in the United Kingdom, and of course now I'm a Group Leader at the John Innes Centre also in the United Kingdom. Since this is The Bio-Diverse Festival, I thought it also important to mention, I am multiply disabled, I have Ehlers Danlos Syndrome, Postural Orthostatic Tachycardia Syndrome and some other stuff. And I'm also queer. I'm asexual and non-binary, using she/her is fine. This is just a picture of me in my wheelchair in case you want to see what I look like. I'm looking happy. I'm sitting there wearing a dark shirt and jeans and sitting in a blue manual wheelchair.

So I want to start off the talk by acknowledging some folks who've contributed a lot to the science over the last couple of years when I was working in Cambridge, in particular PhD student Kathy Darragh, who's showed in the lower left corner, as well as Sylvia Fernando Garza, Diana Abondano, and of course my advisor Chris Jiggins. Also the staff at the insectary is at the Smithsonian tropical Research Institute, especially Oscar and the members of the Jiggins group, including Ian Warren, and our collaborators, Owen McMillan, Stefan Schulz, Richard Merrill, Pasi Rastas, Marek KuÅka, and Frank Chan. I want to also mention again, since this is The Bio-Diverse Festival, this is important, the work in Panama was conducted on the lands of the Embera-Wounaan people, it's hard to tell in Panama whose lands belongs to who at this point, but this is the most likely originator of these lands.

So what keeps butterfly species apart, and particularly what keeps these two species of Heliconius apart, that are shown on the top of the slide, one species is blue, with white marks, the other is black with red and yellow marks. So, one thing that keeps them apart is habitat. So the blue species prefers more forest cover and the red species less forest cover. Another, the fact that f1 females are sterile and therefore can't lay eggs. Another, the fact that f1 individuals actually look different from their parents, and therefore they don't fit the sort of picture that predators are looking for, to avoid eating because Heliconius are toxic, and so because they don't fit this normal picture, they get eaten. And another factor that influences their, what we call reproductive isolation, is the fact that they choose their own species to mate with, so mate choice. And I'm particularly interested in this aspect of mate choice.

So, these two species of Heliconius that I work on H.melpomene, and H.cydno are part of a larger group of Heliconius butterflies that are distributed across Central and South America. And their outgroup is the silvaniform butterflies of Heliconius. Although these two species are each other's closest relatives, within the area that we study them in Panama, they are the only two species in this group that are found. So, in the lower right corner of the slide, there's a map showing that in Central America down
into Panama and Northwestern South America, you see Heluconius cydno, that's the blue species, which will be in blue for the rest of the talk. And then from Southern Central America, down into South America, you see Heliconius melpomene, that's the red species which will be in red for the rest of the talk.

So, there are two types of male pheromones in Heliconius. First of all, there's winged pheromones. Those we think are used as aphrodisiacs. They're used by the males during courtship, so they flutter around, and they expose their wings to the females. This has been shown to play a role in female choice, it also likely plays a role in species isolation or reproductive isolation. And these hormones are present in a region of the dorsal hind wing called the androconia, where there are scales specific to males. Genital pheromones, we think of as anti-aphrodisiacs. They're transferred by males, to females during mating, and they benefit the male because it avoids female remating. So basically, his sperm get precedence. And they have mixed benefits to the female because on the one hand, they decrease harassment, but they also prevent her from mating with other males.

So, let's look briefly at the pheromones of the parent species. So, you can see on the left hand side there's a picture of the wings of Heliconius cydno and melpomene and the brown region is the wing pheromone region that I talked about the androconia. On the right side is a gas chromatograph mass spectrometry trace that shows the actual chemical composition of the pheromones themselves. Don't worry too much about the details of the plot unless you've spent a lot of time reading these already. The important thing is to notice that the picture of these two plots is a little different. So on the top is melpomene in red on the bottom is cydno in blue. And essentially, you can see that cydno has only a few peaks, especially this big peak labelled P4. melpomene has lots of smaller peaks, so basically, the melpomene phermone is a bit more complex to the cydno pheromone. If we plot these using NMDS, which is a technique that's somewhat similar to PCA, we see that cydno clusters on one side on the left side here, and melpomene clusters on the other side, so there's not really any overlap between their main pheromone bouquets.

What about the genitals? So, the genitals are even more divergent than the wings. Melpomene is dominated by this one compound E-\(\alpha\)-ocimene which is labeled 'oci' on the slide. I'm showing another gas chromatograph mass spectrometry trace. Cydno is dominated by a macrolides. Those are labeled M1-M3 and isobutyl oleate, which is labeled oleate. It's worth noticing that essentially melpomene is quite simple, it really only contains ocimene and a few other compounds, where cydno is much more complicated in this case containing lots of these macrolides. If you try to do the same kind of NMDS plot using Heliconius genitals, you see they're very, very divergent, they're, in fact, so divergent that the NMDS can't actually resolve on just two axes with over 20 samples. That's due to the presence of ocimene. So, if you remove the ocimene from the data table, the NMDS now resolve successfully. But you can see again, that essentially there is no overlap between the pheromone compounds of the two species, they're completely distinct.

So, what do we already know about pheromones, and melpomene and cydno? So in melpomene, we know that there's a wing pheromone. And the specific chemical that we know is important is called octadecanal. We have evidence from electroantennograms, which is measuring the electricity across the antennae of the butterfly in response to the stimulus. And I here I'm showing with my hands a sort of mock antenna. We also have done behavioral experiments that have shown that octadecanal is important. We know that the genital pheromone, a genital pheromone, is this E-\(\alpha\)-ocimene. And, you might remember that was one of the really dominant pheromones earlier. And
we again have electroantennogram evidence as well as behavioural evidence. Now, we don't know what the bioactive wing pheromone, actually is. This is going to require some more advanced technologies, specifically GC coupled electrophysiology. We do know that there is a response because again, those electroantennogram, the test of the antenna, has shown that there's a neurological response to the compounds, but we don't know which compound it responds to. And we do know a genital pheromone, that's hexyl-3-methylbutyrate, which has behavioural evidence.

So great, okay, we know something about the pheromones, but what about their genetic structure? One of the really interesting things about Heliconius is that many of the traits that people have studied in the past, especially color come down to just a few genetic loci. So just a few genes are important for controlling the massive variation we see in Heliconius color pattern. And we were curious if the same thing was true of pheromone composition as well, is it just a few genes are there lots of genes, we had no idea. So we created a mapping population to do something called QTL mapping or quantitative trait locus mapping. We cross melpomene males and females, with cydno males and females, we then got F1s. Males specifically, because remember, the females are sterile. We then cross those two either melpomene or cydno females. And that led us to our backcross mapping population, this population consisted of 25 families. And for each experiment, mapping either winged pheromones or genital pheromones, or the androconial area, which I will talk about otherwise, we ended up with about 200 butterflies per population. So one of the things that's worth noting is that your resolution for QTL mapping depends on how many individuals you have. And so we don't have as good of a resolution as some studies that might use thousands of individuals. So our peaks are relatively broad, it's hard to pinpoint the exact loci with this small of a sample size. But if you've ever reared butterflies in the tropics, this is a pretty respectable sample size.

So I'm just going to walk you through what a QTL plot looks like because you're gonna see a number of these in the next few slides. So there's a square plot in the center of the slide on the x axis is labeled chromosome from 1-21. 21 is actually the Z chromosome, the sex chromosome, and on the y axis is labeled LOD score, and basically that goes from in this case, 0-10. There's now a background showing a squiggly line that indicates the actual QTL plot as well as 2 horizontal lines. The top horizontal line is the bonferroni corrected significance threshold across the genome. And the dotted line is the uncorrected significance threshold across the genome. And you can see on chromosome one, that there's a peak that goes over the bonferroni corrected significance threshold, and I've labeled that significant peak. So that's what these plots will look like.

So octadecanal, that component that was really important in the wing pheromone in melpomene, maps to chromosome 20. The cool thing about this is we didn't have any other previous lows on chromosome 20. So the color low siren on chromosome 20, the main choice loci, we looked at run on chromosome 20. So this is kind of new. When we backcrossed instead to Heliconius cydno, we see that in fact, it also maps to chromosome 20. So that's independent confirmation with a different set of families that this locus on chromosome 20 is in fact legitimate. Ocimene production maps to chromosome six and backcrosses to Heliconius cydno. So there's a peak on chromosome six, and then hexyl-3-methylbutyrate, that's that cydno genital pheromone, maps to chromosome three and backcrosses to melpomene.

So what do we know now? We've got our octadecanal, that wing pheromone in melpomene, we see that there are 16 candidate genes and one quantitative trait locus, I'll explain those genes in a minute. For ocimene, we actually have a functionally characterized terpene synthase that biosynthesizes
ocimene in melpomene. So we basically have riddled this enzyme down pretty thoroughly, and we know what's going on. In cydno, we still don't know what the wing pheromone is, and so therefore, we can't map it. And we did find a QTL on chromosome three for a genital pheromones in cydno, which is hexyl-3-methylbutyrate.

So what about the other pheromone components? So I showed you a plot in the beginning that showed you lots and lots of little peaks. That's because the pheromone is relatively complex. It's not just one or two compounds, there's other peaks as well. So, what about them from a genetic mapping point of view? So there are a couple of compounds that are shared between the wings and the genitals, and they mostly map in parallel. So for benzyl cyanide, for example, there's a peak on chromosome 17, and the wings and a peak on chromosome 17 - the genitals, sounds good.

And henicosane. Again, there's a peak on chromosome 20, for the wings at peak on chromosome 20, for the genitals, so the genital and the wing samples agree, which is good. The wind compounds really seem to cluster on chromosome 20. So on the left hand of the slide, I'm showing a plot where we've plotted all of the QTL peaks on top of each other. And you might be able to see that on chromosome 20. There's a massive overlap of lots and lots of different peaks. So we have 10 compounds that have significant QTL on the wings. And seven of these compounds actually cluster on chromosome 20. So now I'm showing you just chromosome 20, on the x axis. And you can see that the QTL plots all line up at the end of chromosome 20. And they're actually quite consistent in exactly where they're falling. But because we don't have massive mapping resolution, I can't tell you that these definitely are due to, for example, the same gene. The genital compounds really cluster on chromosome eight. So again, showing you all the QTLs plotted on top of each other, and the fact that there's a massive overlap on chromosome eight. There are nine compounds with significant QTL, not including E-\(\Delta^2\)-ocimene. And six of these compounds cluster significant on chromosome eight, again, their peaks all overlap each other. But again, I can't tell you because of our resolution, that this is due to a single gene.

So what compounds are actually clustering now that we do see clustering? For the wings, we see fatty acid derived compounds. So for example, octadecanal and henicosane, that I talked about earlier. And all but one of the significant fatty acid derived compounds are FADs, cluster on chromosome 20. For the genitals, what we see is macrolides. So 12-octadecanolido, hexadecenolide, etc, all the significant macrolides cluster on chromosome eight. So in other words, what I'm trying to tell you, what this slide is essentially that the clustering that we're seeing on chromosomes 20, and eight seems to be driven by chemicals that are from the same chemical families, that being fatty acid drive compounds and macrolides.

So we know a little bit about the biosynthesis of octadecanal, partially because we focused on that compound in some earlier studies. So octadecanal is synthesized first by primary metabolism, as you might expect, that produces stearic acid. From stearic acid, you then produce either one octadecanol, the alcohol and then turn that into an octadecanal, or you might produce octadecanal directly from stearic acid, we're not entirely sure which is happening.

So the likely key players in this enzyme pathway are fatty acyl-CoA-reductase or FARs, of which there are 13 to 14 in this QTL region. And alcohol dehydrogenase is of which there are one to two in this region. And those two catalyzed the two important steps in the production of octadecanal. One of the cool things is that we did some selection tests and
one of these alcohol dehydrogenases is actually under positive selection, it has an Omega value of 1.25, which is pretty cool. The other genes are all under purifying selection. So essentially, they're, they're under selection to maintain their function.

Other fatty acid derived compounds are produced in similar fashions, although desaturase genes are also important, and there are no desaturase genes in these regions. I take that back. No, there are no D saturated genes, my bad. So the fatty acid, fatty acyl-CoA reductase's are highly clustered on chromosomes 19, and 20, in the heliconius genome. So just to zoom in here, showing you chromosome 19. There's a cluster of six fatty acyl-CoA-reductase's on the top of this region on chromosome 19, that I've zoomed in. And there's also a group of desaturases, as well. So there are both desaturases, and FARs on chromosome 19.

On chromosome 20, which you might remember is where most of those fatty acid derived compounds map, there's another cluster of FARs, and also some alcohol dehydrogenases, and these clusters are really densely packed. In fact, in many cases, they look like tandem duplications. The genes themselves are not super closely homologous. So it's possible these are ancient tandem duplications or that they're not actually true tandem the applications, they might have arisen through some other process. Just to show you something kind of interesting, Leonard et al, in 2014, published a paper in Nature Communications, where they looked at fatty acyl-CoA reductases in cyclist butterflies. And they found a sort of pheromone gland fatty acyl-CoA-reductase clade which is labeled in blue at the top of the tree.

So this is a phylogenetic tree that's showing all of the different fatty acyl carrier reductases they managed to identify. Now, what's interesting is that our chromosome 19, fatty acyl-CoA-reductases all fall within the pgFAR clade, this pheromone gland clade. While our chromosome 20 FARs, all 12 to 14 of them are 13 to 14 of them fall outside of this pheromone gland FAR clade. But the cool thing is, you might remember that the pheromones that we're interested in including aka octadecanal appear to be produced by FARs that are on chromosome 20, because the QTL that they map to are on chromosome 20 rather than on chromosome 19, where we didn't actually see anything mapping. So it's a little bit surprising because unlike in the Leonard paper where there is a pheromone gland FAR clade, we actually see that our pheromone FARs fall outside of that clade. So there's sort of a different genetic history to the production of pheromones and bi-cyclists than there is in our Heliconius butterflies.

So I talked in the beginning about the fact that essentially, you do often see overlap and Heliconius, where many of the different color patterns are produced by a relatively small number of genes. And these genes are responsible for massive amounts of variation. And we know that these things are important for reproductive isolation. What we found is that the pheromone component QTL don't cluster with these known QTL that we already have information on. So what I'm showing you here is across the chromosomes 1-Z, the number of QTL that had been previously identified, and then the number we've identified between melpomene and cydno. So in purple, we see low sign chromosomes 1, 10, 15, and 18, for color pattern. Male mate choice has been mapped to chromosomes 1, 17 and 18, in orange. In gray, androconia area we mapped in this study to chromosome 18. And the wing compounds then map one to chromosome one, one to chromosome 17, one to the Z chromosome, and eight to chromosome 20. So again, chromosome 20 is kind of important for the wing compounds. And then finally, the genital compounds mapped chromosomes, 3, 5, 14, and six of them mapped to chromosome eight. So again, we see cluster on chromosomes, eight and 20. And these are not
chromosomes where we've previously identified QTL for differences between melpomene and cydno.

So in conclusion, what did we see? We saw significant clustering of pheromone loci on chromosomes eight for the genital compounds, specifically the macrolides. And chromosome 20 for the wing compounds, specifically the fatty acid derived compounds. We see no overlap with a known color pattern and make choice QTL which is in contrast with other studies where often traits repeatedly map back to these QTL over and over and over. If multiple pheromone loci are important for species isolation, this clustering that we're seeing could promote speciation to fight despite the fact that there is gene flow between these two species. Gene flow is mostly historic and at very low rates in the modern day. But these pheromone loci could be really important for species isolation, and this clustering could promote that speciation process.

If you're interested in learning more about this, we do have a preprint on bio archive and our paper is under review at ecology and evolution. The preprint is titled 'Clustering of loci controlling species differences in chemical bouquets of sympatric Heliconius butterflies. You can also reach me again at Kelsey.Byers@jic.ac.uk or on Twitter @plant_pollinator no space, no underscore. And of course they'll be available for questions as well. Thanks very much.