

Boyce Thompson Speaks #1

Jay Worley: Welcome!

Suzy Strickler and Lindsay Wyatt: Thank You!

Worley: First, I was hoping you could tell us a little about yourselves, where your alma mater is and how you came to Ithaca.

Wyatt: I am from near Cleveland, Ohio and I went to college at Duke University majoring in biology. And then I did a year of research at the National Institutes of Health, studying fruit fly heart development, before returning to plants – my first love. And now I'm here at Cornell working with Michael Mazourek in the plant breeding and genetics section.

Worley: Wonderful. And Suzy?

Strickler: I'm from near Pittsburgh, Pennsylvania and I did my undergrad at Penn State where I worked on tomatoes with Majid Foolad and screening them for disease resistance and things of that nature. Then I came to Ithaca in 2002 and I've been here ever since. I did my Ph.D. at Cornell and I studied self-incompatibility. And then in 2010, I joined Lukas Mueller's group and tried my hand at bioinformatics and kind of came back to tomato again.

Worley: Alright, so Lindsay, what first got you interested in plant science?

Wyatt: I have had a long-standing interest in plant science because my family is a third-generation owner of a retail garden center in Ohio. So that's what my father does for a living, I grew up there and so I've always really been interested in plants. And then combining that with a really great advanced placement biology course in high school kind of set me on this track. It's been really great.

Worley: That's great. Suzy?

Strickler: I guess sort of a similar story. It was my family, mainly my dad. He is a real sort of living-off-the-land type so we started gardening. I think there are pictures of me when I'm like, two, picking broccoli and eating it in the garden. Yeah, so one thing led to another and here I am.

Worley: Do either of you still garden up here in Ithaca?

Strickler: Absolutely!

Wyatt: I have what I call my very large squash garden, which is all of our research plots and at home all I manage right now is a couple of pots of herbs.

Worley: So you have quite enough gardening here at work and so you go home to be away from it?

Wyatt: Not necessarily to get away but at the end of the day sometimes you just need to eat dinner and stop thinking about plants for a while.

Worley: So I have to ask, it sort of gets into the research topic we are going to be talking about today. Do you ever get to bring any of those squash home?

Wyatt: Yes. We have tons, literally, of squash to give away at the end of the season so I take a lot home, we share a lot with collaborators, and give some to food pantries as well so I have developed a good repertoire of squash recipes.

Worley: And what sort of squash is this?

Wyatt: We do all sorts. We especially work on a lot of butternut squash so we have a butternut squash called honeynut that we are very proud of that's available now in seed catalogs. And it's a miniature butternut squash with a really rich flavor and a dark color but we have acorn squash, delicatas, jack-o'-lanterns, summer squash—all sorts.

Worley: So I see your paper here is on acorn squash—what sort of properties makes this important for us as a source of food?

Wyatt: Well, the reason we are starting with acorn squash in our investigation of fruit development is because it is one of the most popular winter squashes in the species *Cucurbita pepo*, which is the same species as jack-o'-lantern and zucchini and is the species with the most preexisting genetic and genomic resources. We wanted to build on what has already been studied and picked acorn squash for that reason as an important member of the species.

Worley: What else, outside of squash, is a Cucurbit?

Wyatt: Other Cucurbits are cucumbers, watermelon and melon. Those are the main ones in addition to squash.

Worley: And are those very related to each other on a sort of genomic level?

Wyatt: They seem to be, some of our comparisons in the paper are indicating as such.

Worley: So all of these Cucurbits have very different tastes and textures and is that sort of what your research is starting to get into?

Wyatt: Our research is more specifically focused on, within squash specifically, what causes the differences between taste and texture. So if you think about a spaghetti squash, or if you try to cook up your carving pumpkin, your jack-o'-lantern, they're stringy and they're less sweet versus something like an acorn squash or moving up in more flavor and texture would be a butternut squash and a buttercup squash. To kind of give you a series of what we see as increase in quality—that's the comparisons that we are looking at.

Worley: And do we really understand a whole lot already about development and ripening, or is this sort of a new area for us as plant researchers?

Wyatt: There's a lot of really great research that's been done on fruit development and ripening in other fruits and vegetables. We are some of the first looking at a global analysis of fruit development in squash. Some is known about the physical changes that happen, like how sugars increase or color will increase during fruit development, but this is a new analysis of all the gene expression that is causing those differences.

Worley: And so what you've done here—I'll give it away, I've read the paper—you've sequenced a transcriptome and I was wondering if Suzy could just give us a little bit of insight into, well, what is a transcriptome? That word may not be familiar to everybody listening.

Strickler: So a transcriptome is sort of the collection of expressed sequences from a genome. A lot of the time, these come from genes. So genes are the things that oftentimes are expressed and we can sort of visually see these things so whenever you sequence a transcriptome you're actually only getting a very, very small percentage of the genome because the genome is much, much bigger and most of the genome isn't actually made up of genes.

Worley: Ok, so there is a lot of other stuff in a genome but a transcriptome is pure information?

Strickler: I guess you could say that, yeah.

Worley: Some of our listeners might be familiar with the fact that a lot of RNA, which is what a transcriptome really monitors, is ribosomal RNA. But I'm assuming that in getting these transcriptome profiles ready we've subtracted out those molecules, yes?

Strickler: Yes, that's part of the sequencing step in the preparations.

Worley: Ok, so I was hoping we could get into just a little bit more about what's really different about what you learn from a genome and what you learn from a transcriptome. So let's go from the other angle, if we sequence a genome, what exactly would we learn from that?

Strickler: You would get all of the information from the chromosomes. So you would get things that are sequenced that's from in between the genes—so inner genic sequence—and you would get things like promoters, so these are things that control gene expression.

Worley: Just for, like, the on and off switches of the genes?

Strickler: Exactly.

Worley: So you would basically be able to see what genes are there and you might get a little bit of information about some of when they're turned on but I'm guessing that a transcriptome could be much more specific about when these genes are actually active as opposed to their mere presence.

Strickler: Correct. So, whenever you have a transcriptome, you can actually see which genes are important in that particular tissue or at that particular time in development. Where, if you just had a genome sequence, you can't tell that from the genome sequence and from the genome sequence alone, you also can't tell where the genes are. So the transcriptome kind of allows you to see where the genes are in the genome as well.

Worley: There is sort of an analogy that I thought of for this. It was sort of as if you were at a library and you wanted to see what recipes people were making in their homes. And so you started monitoring, you had a recipe book in the library, and then you started keeping track of what recipes people were writing down and taking home. The recipe book in the library might have, let's say a recipe for making escargot, but if you were to actually monitor the recipe people were taking out of the library you'd find out that nobody was making escargot here, and that they were all making like, brownies or quiche or whatever, that is more common than you find cooked snails.

Strickler: Right, so that's something that we were really thinking about when we were sequencing this transcriptome because there is an already a published squash transcriptome but that is by a group in Spain who works in zucchini mostly. So theirs is looking at the genes that are turned on in roots and leaves and flowers. But we're working on fruit because we really want to make tasty fruit, is our ultimate goal so we instead focused on looking at the genes that are just expressed in the fruit seed in development to get this new set of information that compliments the previously published one.

Worley: So your paper says transcriptome, but what you really did was you took several different transcriptomes for different developmental stages correct?

Strickler: Also, we consider our product one transcriptome but one that encompasses, purposely, all of fruit and seed development. So what I did was grow squash in the greenhouse and then at five, 10, 15, 20 and 40 days after I made a pollination to start the fruit growing, I collected tissues. So then I took the fruit flesh, which is the part that you would eat, and then also seeds and then extracted RNA from those samples. And in this instance, we actually combined all of that RNA together and then sequenced it so we're getting what is a comprehensive look at all of the genes that are active during fruit development. And then the next step, which is not in this paper but we are working on right now, is going to be looking at each of those individual points – what genes are being turned on and off.

Worley: So then you can really begin to get a sense of what genes are being turned on when as this fruit develops. Do you expect that these different—, so we have all these different squash, like a yellow squash is obviously different than an acorn squash—so a yellow squash is very soft, you sort of cook up the entire thing, the acorn squash can be a little bit more woody on the outside. So, I guess, a big point here is that the actual genes that make up these plants are not that different from each other, correct?

Strickler: Right. I think in a lot of ways it is the expression that is different and that is exactly what we are trying to capture.

Worley: So how many sequences did you actually have to sequence in order to make a transcriptome? Like are we talking on the order of thousands, hundreds of thousands?

Strickler: Millions!

Worley: Millions and millions. I think we are actually cheating here, we have the paper in front of us, but just like it was a huge number—150 million reads.

Wyatt: Yeah, so you start with 150 million that are only 100 bases or nucleotides long. That's kind of the basic building block of this genomic sequence information. And then computationally that gets assembled together to make unigenes, which are presumably one gene that's being expressed in the transcriptome. So that's a difficult process that was a real learning experience for me: learning how to do a little bit of computer programming to do that and then after that point, Suzy has been a big part of all of our bioinformatics work, analyzing all that data.

Worley: So you have all of these tiny little fragments, so what do you actually do to sort of get these guys to one single gene or one single transcript?

Strickler: So, I like to compare it to putting a puzzle together where you have a lot of pieces and so you're trying to take all of these little pieces and put them back together into something that's much bigger so all these little sequences actually can go together in such a way that makes up a much larger sequence and your hope is that that larger sequence is actually capturing a full length gene.

Worley: So I'm just going to let you guys talk for a little bit, what are some cool things that you found when you actually processed all of this data you got?

Wyatt: Well one thing I thought was exciting is Suzy did a comparison of our transcriptome with the published one, the one that came from root, leaf and flower tissue. And while there was a lot of really good overlap which supports both of ours being good data, we found about 8000 new unigenes that were able to be identified as a certain kind of gene based on how they compare with genes that are known in other species. And these were known to be sequenced in squash and unique so far to fruit and seed. So that was exciting, realizing we were really adding to the body of knowledge.

Worley: That's cool. So even if you had a genome you wouldn't have been able to predict those 8000 would be expressed in like fruit.

Wyatt: Right, I don't think there would be an easy way to predict that just from looking at genomic data.

Worley: So Suzy, what sort of comparisons did you actually have to make here with the transcriptomic data? What other species could you compare this to?

Strickler: So we did compare it to the watermelon genome and we also compared to the cucumber genome.

Worley: So then what sort of different genes did you guys find that turned on the different developmental stages? So as we are going through this life cycle what are some of the major changes that occur as we go from something just like a collapsed flower all the way to a full squash?

Wyatt: So that is upcoming data. So like I said, we pulled all of the RNA together before we sequenced it for this project that's written up in the paper. So it's more of a reference set is how we think of it. And then right now we are working on our experiments to look at, actually between the different time points and between squash, with, for example, more sugar or less sugar that accumulates, so sweeter or less tasty squash. So that will be coming out hopefully reasonably soon.

Worley: Was there anything just really surprising that you guys found when you were just doing the regular transcriptome—so we don't have the genome for—I don't even think we have a genome for cucumber or watermelon—

Strickler: There is.

Worley: There is a genome?

Wyatt: There is. Yes, and Fei's lab at BTI is actually helping in the efforts to sequence the squash genome so we're really excited about that work and eagerly awaiting the publication.

Worley: How is this going to help researchers in the future who work with other fruits and vegetables to have the squash transcriptome around? And how is this going to lead to a tastier squash?

Wyatt: I think more knowledge in general about the process of fruit and seed development is always going to be helpful for other researchers. I know I have looked a lot at melon and watermelon fruit development work in guiding my own research—things like sugar accumulation can be really similar—also carotenoids, which are the color molecules in squash, can serve pathways. So things that we find in squash could translate to other fruits and vegetables as well.

Worley: And so you guys have this giant data set, I was curious how much of it is actually publicly available now and how much will be released in the future after, of course, you publish the subsequent papers that also come from such a monumental effort. This is not a small task to make. I have made a transcriptome myself but to do each of these requires a lot of work and they're sort of easy to screw up a little bit too. So it takes a lot of technique and a lot to actually get it prepared and into a form that can be sequenced so this is a really large effort is what I'm going to leave it at. There are multiple reports that can come out of this just through careful analysis, painstakingly careful analysis.

Wyatt: The information in this paper is all publicly available already. The raw sequence reads right off of the sequencing machine are available on the NTBI website, the short sequence read archive. And it's mentioned with their reference numbers in the paper. The assembled data set is available on the Horticulture Research site as a supplementary file and has been submitted to the GenBank website is pending final publication.

Worley: I just wanted to reiterate that the paper is open access so no matter who you are you can go read this right now.

Wyatt: Right, no need to have a university subscription.

Worley: Before I let you guys go, I was wondering if you had any shoutouts to people who had helped you a lot in your research?

Wyatt: I think the Cornell University Biotechnology Research Center was very helpful in the early stages of this process helping me to learn Linux and start to tackle this data set so I'd like to acknowledge them.