Title
Complex Traits and Simple Systems: An Interview with Leonid Kruglyak

Journal
Genetics, 203(3)

ISSN
0016-6731

Author
Kruglyak, Leonid

Publication Date
2016-07-01

DOI
10.1534/genetics.116.191866

Peer reviewed

Complex Traits and Simple Systems: An Interview with Leonid Kruglyak

Leonid Kruglyak
Department of Human Genetics, Department of Biological Chemistry, and Howard Hughes Medical Institute, University of California, Los Angeles, California 90095

The Genetics Society of America's Edward Novitski Prize recognizes an extraordinary level of creativity and intellectual ingenuity in the solution of significant problems in genetics research. The 2016 winner, Leonid Kruglyak, has made innovative contributions to the fields of linkage analysis, population genetics, and genomics, while drawing on a combination of mathematical, computational, and experimental approaches. Among other achievements, his work on statistical standards for genome-wide linkage studies has transformed their experimental design, and the linkage analysis program GENEHUNTER has been used to identify hundreds of human disease loci. Kruglyak's group also pioneered expression quantitative trait locus studies, which enabled variation in global gene expression to shed light on the genetics of complex human diseases. In recent years, his laboratory has focused on using genomic technology to establish *Saccharomyces cerevisiae* and *Caenorhabditis elegans* as model organisms for studies of complex genetic variation.

You Trained as a Physicist. What Got You Interested in Biology?

Initially I was interested in artificial intelligence and neural networks, so I started taking neurobiology classes in grad school. Then I realized that I had learned a lot of neurobiology, but no molecular biology, so I started reading about that. I discovered that I found the questions and approaches in genetics and genomics extremely interesting. It also seemed like having a quantitative background would be valuable in that field, so after grad school I looked for a postdoc position in genomics.

How Crucial to Your Success Has Your Physics and Quantitative Training Been?

I think it's been absolutely essential. In the early part of my career, my work depended on my ability to develop algorithms, show why they work, and implement them in code. Later, when I started a wet lab, I began doing the types of experiments in which simulation, statistical analysis, and computational analysis are integral; experiments you could not design without starting out with a pretty good quantitative understanding of what you expect. So even though my lab now primarily does experimental genetics and genomics, quantitative and computational approaches still permeate everything we do.

Which Biological Questions Intrigue You the Most?

My central interest is a very old question: How are traits transmitted from one generation to another? Of course, we understand the core principles of Mendelian inheritance, but I'm interested in complex traits, which involve multiple genes and multiple variants. The basic principles in this area are also old—they go back to [Ronald] Fisher and the modern synthesis. Although these principles give us a pretty good general description of complex inheritance, it has remained very challenging to elucidate the precise genetic basis of specific complex traits. In the last decade or two we've been able to generate data addressing these questions at a much larger scale, which has brought out new problems, like the question of missing heritability.

Yeast Is Traditionally Used to Study Molecular Genetics and Mendelian Traits. What Inspired You to Use Yeast to Study Complex Traits?

I think it comes from my physics background. There's a famous maxim that you should study a phenomenon in the simplest possible system that captures it (but not one so simple it fails

doi: 10.1534/genetics.116.191866
Available freely online.

to capture the phenomenon). It started back in the early days of microarrays, when I had the idea for what are now known as eQTL [expression quantitative trait locus] experiments. Since we could now measure in parallel all the gene expression levels of all the different genes in a system, I thought it would be really interesting to treat those expression levels as a phenotype. That would allow us to look at individual variation and inheritance of molecular traits at a large scale. For years I had seen many beautiful talks on yeast genomics, and when I thought of the practicalities of eQTL experiments and starting a wet lab, yeast was just the obvious choice. I quickly recognized the power of using an organism in which we wouldn't have to deal with a lot of the extra issues that pop up in other systems, complications that aren't central to the problem we are trying to understand.

I think perhaps the reason complex traits weren't originally so commonly studied in yeast is that the system is extremely powerful. So if you are interested in a specific biological question, you can generally make much faster progress in yeast by studying it as a simple trait. At the same time, it's likely that most of the phenomena of complex inheritance that exist in higher organisms can also be found in yeast. The difference is that in yeast we can understand the details a lot faster and with higher experimental power.

Why Did eQTL Experiments Become So Popular?

Our initial goal in developing eQTLs was to understand complex traits; that is, what rules can we learn by looking at thousands of model quantitative traits that are all comparable to each other? But as we started working on eQTLs we realized that because this model trait provides information at the level of a transcript, it can tell you things that you can't learn from studying a trait that's further removed from the DNA level. For example, because transcripts correspond to a physical location in the genome, we could ask questions about *cis*-acting *vs. trans*-acting regulatory variation. And we identified groups of transcripts, representing genes of related function, that map to the same locus (eQTL "hot spots"), which can provide insight into transcriptional regulatory networks.

We could also combine eQTL studies with measurements of other traits, such as fitness in a particular growth environment. If these traits map to the same loci that explain expression-level variation at specific genes, you now have important intermediate level information that could connect DNA variants to organismal traits, which can otherwise be very difficult. Basically, the expression-level information gives you a link from DNA sequence to the level of gene expression, to specific biological processes, and then to a cellular or organismal phenotype.

This idea of using eQTLs as a bridge has now been widely applied in human genetics in particular. Once people started doing genome-wide association studies on a large scale, they often ended up with loci for diseases or disease-related traits mapping to noncoding regions. It was hard to figure out which genes the variants were affecting, so having that level of eQTL variation sitting between the genotype and the phenotype has been a powerful addition to the toolkit.

Leonid Kruglyak continues to change the way we think about the genome, how to navigate it, and what those changes mean for transcriptional regulation. Having said all that, I think one of Leonid's greatest assets is that he brings out the best in other scientists. I, like others, am a far better and more creative scientist when he is around me because he asks simple and far-reaching questions that make me rethink my data and its interpretation.

—Elaine Ostrander, National Human Genome Research Institute

Why Are You Now Working with *Caenorhabditis elegans* for Complex Traits Genetics?

At the moment we are doing a lot of tool development to make worms more useful for complex traits genetics—again, like yeast, the worm was not used much for these questions until recently. Our initial work includes genome sequencing to define genetic variation in various strains, finding isolates that are genetically divergent from each other, and understanding the levels of variation and the forces that shaped this variation. One of my alumni, Erik Andersen at Northwestern University, is doing a lot of work to leverage all of that information, the collections of isolates, and the crosses that we've made between different isolates.

A lot of the things that we take for granted in yeast become much more challenging in worms because of all of the complications that come along with being multicellular and having tissues and moving around and having a sensory system. But we're motivated by the fact that this also opens up a lot of phenotypic space—particularly for phenotypes that are more closely connected to what you see in other animals compared to yeast. Plus there's the challenge of taking what we have learned in yeast and scaling the technologies up to a more complex setting.

You Have Also Done Some Work on Yeast Population Genomics. Why Is Understanding Wild Yeast Important?

A lot of people in the field are doing interesting field work with yeast where they ask natural history questions: When was yeast domesticated? How did it adapt to the different types of human usage and fermentations? How do, say, the European and North American vineyard strains differ from Sake strains domesticated in the Far East? It has been really interesting to see what genomics is telling us about these questions. In terms of using wild yeast as a model for complex traits, the benefit is that you get a lot more genome variation and phenotypic variation that you just wouldn't get from lab strains.

One thing that we and others have been able to show is that although laboratory strains may have different names and

people are used to thinking of them as quite distinct, they almost all have a very recent common ancestor and share large stretches of the genome with each other. This means that when you find a variant in one lab strain, it's quite likely to be present in most of the other commonly used strains. Often these are mutations and adaptations that arose in the lab. For example, one of the first yeast traits we analyzed was the clumping phenotype, which is caused by incomplete daughter cell separation. We thought of it as a mutant phenotype because the vast majority of lab strains don't clump. But it turns out that most wild yeast *do* clump. We found that the lab strains have a very specific mutation in a cell cycle regulator that prevents this. It seems that it was extremely useful to early yeast geneticists to have a strain that you could easily streak for single cells, so at some point a mutation that ensured daughter cells separate from the mother was artificially selected. We were able to show that the mutation arose at some point after the domestication of yeast in the lab because we have the progenitor from which most lab strains derive, and it doesn't carry that mutation. This kind of phenomenon has been seen over and over again. We saw a similar phenomenon in *C. elegans*, and there are other examples in various model organisms; just because you move representatives from a wild species into the lab setting, evolution doesn't stop.

So I think it's important to figure out which traits naturally vary in the species as a whole and which are specific adaptations to the very different environment of the lab. Also, understanding the population genetics of the species is crucial to our analysis. Selective and demographic forces have a lot of impact on the spectrum of variation underlying a particular trait, and this can help you understand which results may be specific to yeast and which may translate to other species, or how you might need to modify particular parameters.

A few years ago we all had to wrap our heads around what types of experiments very large-scale sequencing technologies would enable. Now it has become absolutely essential to pretty much everything we do. I think we're going through the same sort of transition with genome editing tools.

—L.K.

You Recently Published a Genetic Mapping Method That Uses CRISPR to Systematically Engineer Recombination Events. What Impact Is CRISPR Genome Engineering Having in Your Work?

We're exploiting the CRISPR/Cas9 toolkit for quite a few different applications. We're a bit spoiled with yeast because we could already do many of the things that in other systems you can only do now thanks to CRISPR/Cas9. So we've been looking at applications that are either qualitatively new, like this CRISPR mapping approach, or applications that CRISPR makes possible in terms of scale. A few years ago we all had to wrap our heads around what types of experiments very large-scale sequencing technologies would enable. Now it has become absolutely essential to pretty much everything we do. I think we're going through the same sort of transition with genome editing tools.

What's Next for Your Lab?

We want to use a combination of large-scale sequencing, large-scale gene synthesis, and genome editing tools to get a much more granular view of genetic variants and their influence on traits in yeast. We also want to figure out how we can make these tools more general. I think there are a lot of approaches where the findings translate but the exact details of the design don't. So, for instance, the experiments you can do in yeast aren't necessarily feasible in mammals. But for approaches based on synthetic biology and gene editing, you can often directly take what you develop in yeast and apply similar methods to mammalian cells, so we're thinking about how best to do that.

You're a Big Supporter of Publishing Preprints. What Has Your Own Experience with Preprints Been Like?

In the physics community, online preprint servers have existed for over 25 years, and they grew out of an even older paper-based preprint culture. Typically, when you submitted a manuscript for publication, you would also send copies to all your colleagues in the field for comments. So maybe coming from a physics background made me more comfortable with preprints initially. But it's still one of those things that feels like a bit of a leap into the unknown when you do it for the first time. Then once you've done it you wonder, "OK, why haven't we been doing this all along?" because it enables people to read your work as soon as you consider it complete. I believe the traditional peer-reviewed publication process plays an important role alongside preprints, but once a manuscript is ready to be submitted to a journal, we can post it right away and, in addition to the peer reviews that we get from the journal, we also get feedback from everybody else who wants to comment. We've gotten specific criticisms in this way, and occasionally readers of our preprints have even picked up on things that peer reviewers didn't. So now we can fix those issues before publication. We've found that if people can look at your work before it's officially published, it accelerates the science. In genomics and quantitative genetics, fields which have been early adopters of preprints in biology, it's gotten to the point where if you read about something interesting that's cited as "in press," and there isn't a preprint, it's a bit of a surprise and a frustration.

Communicating editor: C. Gelling