A 10 year plan for Dahlia genomics to address virus resistance

August 23, 2021
Alex Harkess, PhD
Assistant Professor, Auburn University
Faculty Investigator, HudsonAlpha Institute for Biotechnology
Profile: https://hudsonalpha.org/faculty/alex-harkess/
Publications: https://scholar.google.com/citations?hl=en&user=EkrN6nAAAAAJ

Background and rationale

Alex Harkess, Ph.D. is an Assistant Professor at Auburn University in the Department of Crop, Soil, and Environmental Sciences. He is jointly appointed as a Faculty Investigator at HudsonAlpha Institute for Biotechnology, a globally recognized human and plant genomics institution with diverse expertise in genome sequencing, molecular biology, and functional genomics. Nearly half of the high quality plant genomes that exist in the public repository Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html) were sequenced and assembled at HudsonAlpha. The Harkess Laboratory at HudsonAlpha has expertise in plant reproductive genomics, evolutionary biology, phylogenetics, and the application of these diverse tools to plant breeding.

Recently my laboratory recruited a Masters student, Zachary Meharg, to train in my laboratory for 2 years and conduct research on Dahlia. From my conversations with Dahlia breeders at the ADS meetings, here I lay out what I consider to be a feasible and realistic 10 year plan for addressing several key issues in Dahlia breeding, largely revolving around issues with virus infiltration into breeding germplasm. These are admittedly complex and multi-disciplinary, collaborative issues, and will require significant external funding efforts via federal agencies such as USDA and NSF. We propose that the ADS should consider a three pronged research approach: 1) building genomic infrastructure to support the study of virus integration and molecular breeding, 2) classical and molecular breeding approaches to introgress virus-resistance from diverse germplasm into cultivated Dahlia, and 3) a biotechnology approach using synthetic small RNAs to directly target viruses.

Aim 1: Developing genomic resources for *Dahlia* to enable molecular-assisted breeding

A key barrier to addressing any virus-related research is the lack of any genomic resources for any Dahlia species, either cultivated and wild species. The cultivated Dahlia genome is complex; it is an octoploid, meaning that there are four homologous pairs of each chromosome. It is possible to assemble the genomes of octoploid plants with the latest sequencing technologies (e.g. Pacific Biosciences Sequel-II HiFi reads), and here at HudsonAlpha we have experience and expertise in assembling these types of complex plant genomes. However, there are a handful of other Dahlia species which have less complex genomes, and are a more straightforward path to building a reference genome that we can use for all downstream plans.

Kristine Albrecht has kindly organized seeds from Dr. Walbot's collection trips, and arranged for three ADS growers nearby HudsonAlpha and my laboratory to grow plants to maturity. These species are: *Dahlia rudis, Dahlia coccinea, Dahlia tenuicaulis, Dahlia brevis, Dahlia sorensenii.* Zach Meharg, a Masters student in my laboratory, will isolate DNA from these different species, and perform preliminary sequencing analysis to identify which species have the smallest genomes. We will use "the best" species to generate more DNA, and build full-length

chromosomes using advanced DNA and RNA sequencing techniques with PacBio sequencers here at HudsonAlpha. This will be funded largely by ADS and the existing fundraising for the genome project. This genome will be used in several major ways, including collaboration with Dr. Hanu Pappu's lab to investigate virus integration and diversity.

Expected result 1: An ultra-high quality *Dahlia* reference genome, comprised of high quality chromosome assemblies and gene annotations.

Next, the Harkess Lab is supremely interested in mutations that drive floral variation. The ADS National Show is the ideal scenario to sample diverse dahlia accessions that vary in major floral characteristics (e.g. between categories). We will sample the winners from each major show category, which we know display unique floral variation. We will isolate DNA and sequence inexpensive short reads on Illumina sequencers, and using the reference genome from Expected result 1, identify and characterize genes displaying mutations in known floral and developmental pathways. These data are also immensely useful for identifying the diversity of integrated viruses into dahlia genomes. In collaboration with Dr. Pappu's lab, we will *de novo* assemble virus genomes that are present in each of the show plants in order to develop targets for Aims 2 and 3.

Expected result 2: A catalog of genes controlling major flower variation in Dahlia as a model for improving related crops (e.g. Asteraceae), as well as a catalog of virus genomes.

Aim 2: A classical and molecular breeding approach to developing virus resistance in cultivated Dahlia

Horticultural societies uniformly suffer a key issue: repeated, selective breeding for a small number of traits on a limited diversity of germplasm leads to immense genetic bottlenecks. Most cultivated show dahlias are likely highly genetically similar, which is a disastrous situation when considering disease resistance. Time is a key factor to think about. Did virus issues exist as heavily 50+ years ago? Less so, it seems. High levels of genetic diversity are crucial in all aspects of plant breeding. By leveraging this existing diversity that ADS members have collected and maintained, we propose that the ADS can assist us with thoughtful crosses with older varieties that are sturdier when it comes to virus, for instance developing introgressed lines that are crosses involving Thomas Edison and Little Beeswings. The genetics of these "older" varieties are critical. A real-life example of this working can be found in Chestnut, which has almost entirely been wiped out from North America due to Chestnut blight. Chinese Chestnut, however, is resistant to blight. By repeatedly introgressing Chinese Chestnut into American Chestnut, we maintain the "best" characteristics of native american chestnut, while also maintaining the resistance genes found in Chinese Chestnut. We propose the same type of breeding approach here: collect as much diverse germplasm as possible, find the "old" cultivars that are not used much anymore, and begin repeated crossing schemes to introgress their alleles into modern show varieties. This requires the ability to directly challenge offspring from all crosses with a virus to test the infection rate, on which we will collaborate with Dr. Pappu's laboratory.

Expected result 3: Modern cultivar breeding that contains resistance genes found in older cultivars, leading to increased virus resistance in the dahlia breeding community.

Aim 3: A biotechnology approach to virus resistance in dahlia

Classical breeding and the preservation of plant diversity can be immensely effective, and indeed is how most major crops are bred in order to be resistant to various pathogens. Small RNAs are one mechanism that nearly all plants express that can be used to target viruses for destruction. They serve diverse functions, including plant defense against fungi and viruses. Effectively like small missiles in a plant, these small RNAs have memory, and can target foreign invaders for destruction. In short, they can be used to target and destroy viruses.

Building on Dr. Pappu's laboratory protocols of generating callus (tissue culture) methods, we are proposing a collaboration to insert a construct containing a special type of small RNA that specifically targets a conserved region of the Dahlia viruses that infect breeding lines. This is a similar approach to how COVID-19 mRNA vaccines are designed, with specificity against the highly conserved spike protein of the virus. This special type of small RNA, called a synthetic trans-acting siRNA, or syn-tasiRNA, creates a feedback loop of more small RNAs that continually and perpetually target viruses. We believe that these syn-tasiRNAs may be heritable if engineered correctly, so that this "memory" is transgenerational. Consequently, while we will be genetically modifying Dahlias at first, in order to express these synthetic molecules, we believe that offspring will carry these molecules in a non-GMO way. This requires a key collaboration with Dr. Pappu's lab, who has developed a catalog of some Dahlia viruses that will serve as the blueprints for what we target using these syn-tasiRNAs.

Expected Result 4: Genetically modified Dahlias that express syn-tasiRNAs that target known viruses.