Genome Project Update
9/9/2021
Dr. Harkess

Zach Meharg has started his first semester of his MS program in Crop, Soil, and Environmental Sciences at Auburn University and HudsonAlpha Institute for Biotechnology. In Zach's first week, we travelled to Marcie Holt's farm in Ringgold, GA, who kindly agreed to collaborate with us to grow D. brevis, D. rudis, D. coccinea, and D. sorensenii. Kristine Albrecht has driven and organized this collaboration, as well as created duplicate sites at two other ADS members and growers near us in Alabama. We travelled to Marcie Holt's farm and collected leaf tissue for all four species dahlias, plus a smattering of accessions that we know to be interesting from a virus resistance standpoint.

Zach Meharg and lab manager Haley Hale isolated DNA from these flash-frozen leaf samples and assessed the DNA for purity and concentration (Table 1). Dahlias produce high quality DNA, which is helpful to know. These DNAs are now in the sequencing queue at HudsonAlpha, a global leader in plant and animal genome sequencing.

Table 1: DNA isolation quality and quantity

Extraction Date	Speicmen	Nanodrop ng/µl	Nanodrop 260/230	Nanodrop 260/280	Qubit ng/µl	Qubit ng/µl
8/13/21	D. brevis #1	1154.252	0.86	1.79	68.6	67.1
8/13/21	D. rudis #3 cutting 2	1844.125	1.26	1.93	189	185.5
8/13/21	D. coccinea	1479.043	1.65	2.05	136	135.6
8/13/21	D. sorensenii #2	846.321	1.09	1.92	60	58.8
8/13/21	Thomas Edison	1931.688	1.63	2.06	298	295
8/13/21	Edna C.	2978.652	1.29	1.99	336	331.2
8/13/21	J.K. Lauren	1576.147	1.36	1.95	188	181
8/13/21	Spartacus	1466.803	1.11	1.81	164	170.3

We are currently generating at least 60X Illumina PE150 short-read coverage of each individual. This is one full lane of Illumina NovaSeq6000 sequencing, priced at nearly $6,000. The genome size of Dahlia is quite large, at least 8 Gb for the octoploid cultivars, so it is quite expensive to generate sequencing data for these individuals. As soon as the run is finished, these data will be used to estimate the genome size, ploidy, and heterozygosity from each of these individuals. This can be achieved by running GenomeScope (https://github.com/schatzlab/genomescope) and SmudgePlots (https://github.com/KamilSJaron/smudgeplot) in tandem. We estimate that sequencing will be completed in the next three weeks, and we will have data in hand.

We have collect more than enough tissue that after that, we will move forward with generating an ultra-high quality PacBio HiFi genome on one of the species dahlia accessions that displays the lowest genome size, lowest heterozygosity, and lowest ploidy. Hopefully they are tetraploids.