

2021-2023 Dahlia plans

The Harkess Lab at Auburn University in Alabama has recruited a Masters student, Zach Meharg, to begin studies in the Fall of 2021. A two year plan incorporating two manuscripts has been developed and this will be two chapters of Meharg's thesis:

- 1) Manuscript 1: A chloroplast phylogeny of the *Dahlia* genus
 - a) Justification: The RNA-based phylogeny did not yield sufficient phylogenetic resolution, as there appears to be very little genetic diversity within the expressed genes sampled. Here we will instead take an alternative approach and use chloroplast genome sequencing and assembly to resolve the phylogeny of the Dahlia genus
 - b) 14 species Dahlias and 15 modern cultivars will be shotgun Illumina sequenced (5 gigabases of data each) in the Harkess Lab at Auburn and HudsonAlpha.
 - c) Chloroplasts will be assembled using Novoplasty (<https://github.com/ndierckx/NOVOPlasty>), and a species tree phylogeny constructed with RAxML and Astral-Pro.
 - d) We will also assemble mitochondrial genomes from the data using Novoplasty.

- 2) Manuscript 2: The Dahlia genome
 - a) Generate Illumina genome shotgun data (40X coverage) for 5 species -- *D. rudis*, *coccinia*, *tenuicaulis*, *brevis*, *sorensenii*.
 - b) Build kmer heterozygosity plots using smudgeplots (<https://github.com/KamilSJaron/smudgeplot>)
 - c) Assess genome size, ploidy, heterozygosity, and repeat content for the 5 species. Pick 2 to move forward with for PacBio genome sequencing, assembly, and annotation to build the first Dahlia genome. Hopefully one of these genomes is a tetraploid.
 - d) Kristine Albrecht identified growers nearby our lab to start seed for those species.
 - e) Resequence individuals from across the Dahlia Show winner groups to map floral mutations that drive the major floral variation between groups.