The American Dahlia Society

ADS Genome Project Update
June 2025

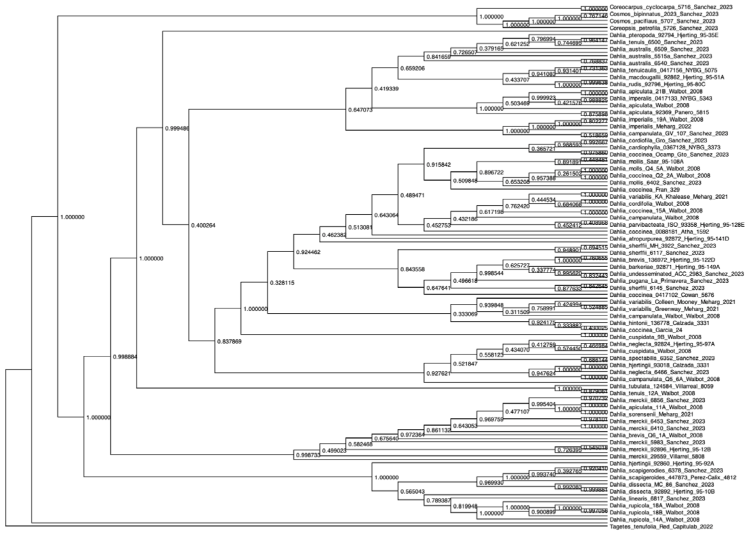
From: Alex Harkess

With summer here, Zach is close to finishing his 4th year of PhD. In particular, we've made progress on all three of Zach's major Dissertation Chapters:

1. Chapter 1: Building a Resolved Phylogenetic Tree of Species Relationships in Dahlia
2. Chapter 2: The Dahlia Genome and the Evolution of Floral Traits in Dahlia
3. Chapter 3: Identifying the Genes Underlying the Development of Diverse Floral Forms in Dahlia

In Chapter 1, Zach collaborated internationally to collect tissue samples from more than 35 Dahlia species with a global distribution over the last 3 years. He has extracted DNA from almost 200 individuals across those 35 species, generated DNA sequencing libraries, and that data was just sequenced in late 2024. He is currently analyzing these data, building the Dahlia family tree, and that will be the first major chapter of his dissertation. He has completed a draft coalescent multi-gene species tree of these 35 species, each with ~1,200 genes sequenced per individual plant. This is the largest generation of Dahlia data ever and a major push forward in the phylogeny of Dahlia.

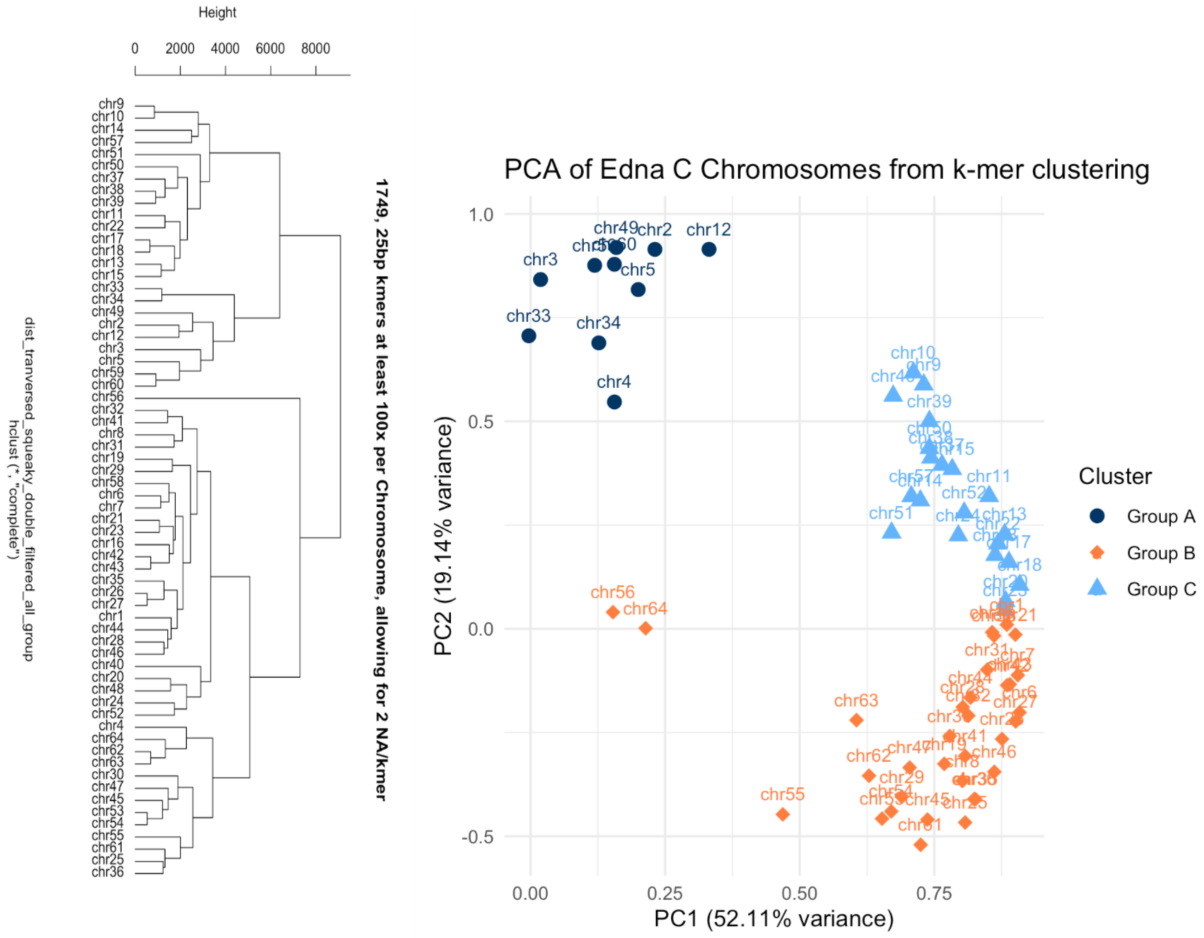
Results: We are finding a reasonable amount of concordance with the more dated phylogenetic trees (from 10-20 years ago). We are still refining this phylogenetic tree, and using more sophisticated methods to disentangle the ongoing and repeated hybridization in the genus. Below is a draft figure of this revised phylogenetic tree.



(Figure 1 larger size)

In Chapter 2, Zach has locked the Dahlia genome assembly as of a few months ago in late 2024. The assembly is contained in 64 chromosomes as expected, and there are some incredibly complex regions of these chromosomes that were difficult to assemble over the last several years. This represents a major milestone not just in the Dahlia community, but in the comparative genomics community across the entire Asteraceae family. Zach successfully sequenced RNA (expressed genes) from 15 different tissue types from the Edna C plant, generating hundreds of millions of expressed sequence reads. He has generated a gene annotation that is high quality and matches closely to the D. pinnata genome assembly and annotation. In other words, Zach's genome assembly and annotation is now locked and very high quality. We are ready to publish this, and Zach is quickly writing the genome release manuscript that we are intending to submit to a journal "G3". In June 2025 he presented this draft manuscript to our laboratory for comments and revision.

Major results: The manuscript is quite dense with material, but one major result that Zach disentangled is that there is clear subgenome structure within the *D. variabilis* genome, that is detectable using our genome sequencing and assembly techniques. It looks like either two or three clear groups are falling out - suggesting that there is a complex structure in the Dahlia genome that we are still untangling. The phylogeny data from Chapter 3 below will be very helpful in that regard. Overall, our results recapitulate that the cultivated dahlia is indeed a tetraploid and not an octoploid as once thought.



[(Figure 2 larger size)]()

In Chapter 3, Zach is identifying genes that control major floral form diversity in Dahlia using two complementary approaches.

1. He's taking a plant breeder's perspective by sequencing a large diversity collection of cultivated Dahlia collected from the 2021 national show. Zach has recently generated whole genome resequencing data for 384 individuals to identify mutations that control traits such as floral form, shape, size, color, and architecture. This project is a very fruitful collaboration between my laboratory and Dr. Josh Clevenger's laboratory at HudsonAlpha, who is a molecular plant breeder who supports breeding programs in more than 20 different countries using this next generation of DNA sequencing technology. Zach is currently analyzing this data in an association mapping approach to identify mutations in the genome that correlate with these floral traits. This was a major milestone in Zach's PhD.
2. The second approach is that Zach is taking a developmental biologist's perspective by growing exemplar cultivars that represent the major floral forms in Dahlia (7 or 8 of them) and producing a gene expression time course of floral bloom development. By sequencing a gene

> expression time course of floral development across these different major floral forms, Zach will be able to identify different genes that are turned on and off that perhaps control the development of these diverse forms. This data will be overlapped with the diversity collection sequencing and genome-wide association study from the first part of this chapter. This work is ongoing and required the genome before we could do anything with it.

In summary, this project in the last 4 years has built the largest collection of Dahlia genetic and genomic data on Earth. We have built the family tree for the entire genus, we have generated a very challenging genome assembly for a tetraploid dahlia, and we have sequenced hundreds of dahlia cultivars to identify genes controlling major traits.

Zach's current timeline for publishing these results is:

Summer 2025 - submission of genome manuscript to *G3* journal.

Winter 2025 - submission of phylogeny manuscript

April 2026 - submission of reproductive development microscopy/transcriptomics chapter