Conserving the Molecular Profiles of California's Endangered Plants

Flint Mitchell; Weilun Tan; Hollie Mickelson; Chris Baysdorfer; Ana M.R. Almeida; Brian Perry; Joanna Garaventa; Bharti Panhar; Randall Davis

Department of Biological Sciences, California State University, East Bay, Hayward, CA

Abstract: California is home to over 6,000 plant species, a large portion of which are endemic to the state. As of November 2023, 386 plant species are at risk of extinction (ranked 181.) according to the California Native Plant Society due to factors such as climate change and urbanization driven habitat loss. Meaningful efforts in preserving these remaining populations requires foundational knowledge of the species’ ecology, evolution, and increasingly: the plants genomic characteristics, in order to guide conservation and management strategies. The Green Biome Institute (GBI) at California State University East Bay (CSUEB) aims to contribute to the conservation of California’s rare and endangered plant species by creating molecular profiles of over 300 rare plants by 2026 and making them freely available to the public and scientific community. These profiles will eventually include each plant's genome size, transcriptome, universally recognized barcode regions, epigenome, metabolome, and leaf microbiome. To date, the GBI has produced approximately 30X coverage of short-read genome sequence data for over 100 species, along with genome assemblies for around 80 of these rare plants. Additional data generation and conservation efforts include long-read sequencing, genome size estimations by k-mer analysis, metabarcoding of endophytic fungal communities, as well as germplasm and propagation efforts. In addition, the GBI genome profiling program has created unique opportunities for engagement in meaningful research of a diverse student body at the High School and College undergraduate and graduate levels. Altogether, we believe these collective efforts will contribute towards data-driven conservation of California’s plant diversity while growing tomorrow’s science-based, conservation leaders.

Over 200 plant leaf samples of various Fritillaria were obtained and DNA was isolated using the Qiagen Plant DNeasy protocol. DNA sequencing was performed using the Illumina HiSeq X Ten platform at the Stanford Genome Technology Center. DNA was submitted to the The International Nucleotide Sequence Database Collaboration.

Table 1 - Draft assemblies for short-read data using Abyss 2.0 (Jackman et al. 2017)

<table>
<thead>
<tr>
<th>Plant Name</th>
<th>Genome Size (Gb)</th>
<th>K Value BUSCO (viridiplantae_odb10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arctostaphylos morroensis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arctostaphylos silvicola</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arctostaphylos stanfordiana ssp. decumbens</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ceanothus divergens</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ceanothus roderickii</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. amcali</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lonicera subspicata var. subspicata</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ceanothus maritimus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carex xerophila</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dudleya densiflora</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

80 rare and endangered California plants were sequenced and analyzed using local and Amazon Web Services (AWS) hosted compute with the following results:

- Genome size estimates using jellyfish (Marcia & Kingsford, 2011) and Genomescope (Yurchenko et al., 2017):
  - Average genome size: 742 Mb
  - Largest: 3.35 Gb (k. halli 113)
  - Smallest: 2.16 Mb (Cercospora trivialis)
- Draft assemblies for 80 species (Table 1):
  - Complete (single and double-copy): 62.3%
  - Complete: 25.1%
  - Missing: 12.6%
- QUAST (Mihaljevic et al., 2015) metrics:
  - Average N50: ~77 Mb
  - Average L50: ~75.2 Kb
- BUSCO (viridiplantae_odb10) BUSCO completeness:
  - Smallest: 218 Mb
  - Average: 1,281 Mb
  - Largest: 22,987 Mb

80 rare and endangered California plants were sequenced and analyzed using local and Amazon Web Services (AWS) hosted compute with the following results:

- Genome size estimates using jellyfish (Marcia & Kingsford, 2011) and Genomescope (Yurchenko et al., 2017):
  - Average genome size: 742 Mb
  - Largest: 3.35 Gb (k. halli 113)
  - Smallest: 2.16 Mb (Cercospora trivialis)
- Draft assemblies for 80 species (Table 1):
  - Complete (single and double-copy): 62.3%
  - Complete: 25.1%
  - Missing: 12.6%
- QUAST (Mihaljevic et al., 2015) metrics:
  - Average N50: ~77 Mb
  - Average L50: ~75.2 Kb
- BUSCO (viridiplantae_odb10) BUSCO completeness:
  - Smallest: 218 Mb
  - Average: 1,281 Mb
  - Largest: 22,987 Mb

Analysis showed the presence of 11 groups. Four of these correspond to known Fritillaria species, these include F. affinis, F. pudica, F. liliacea from Jepson prairie. One group corresponds to plants labeled as F. rogersi/greggii and F. leucos of Nicasio reservoir. One group corresponds to plants labeled as F. agrestis from Contra Loma and Diablo foothills and F. liliacea from Jepson prairie. A fifth group corresponds to plants labeled as F. agrestis from Contra Loma and Diablo foothills and F. liliacea from Jepson prairie. A sixth group corresponds to plants labeled as F. nigri from Lake Chabot and F. agrestis from Brushy point. Theoloto, Tilden, and Peaks. Seven group corresponds to plants labeled as F. nigri from Lake Chabot and F. agrestis from Brushy point. Theoloto, Tilden, and Peaks. Eight group corresponds to plants labeled as F. nigri from Lake Chabot and F. agrestis from Brushy point. Theoloto, Tilden, and Peaks. Nine group corresponds to plants labeled as F. nigri from Lake Chabot and F. agrestis from Brushy point. Theoloto, Tilden, and Peaks. Ten group corresponds to plants labeled as F. nigri from Lake Chabot and F. agrestis from Brushy point. Theoloto, Tilden, and Peaks. Eleven group corresponds to plants labeled as F. nigri from Lake Chabot and F. agrestis from Brushy point. Theoloto, Tilden, and Peaks.

References:

- Chris Baysdorfer, PhD (Director)
- Randy Davis (Advisor)
- Chris Baysdorfer, PhD (Director)
- Randy Davis (Advisor)

Acknowledgments:

- We would like to thank all the students who have contributed to this project, including:
  - M.S. Student: Robbi Mickelson
  - Watan8@horizon.csueastbay.edu
  - Psarn46@horizon.csueastbay.edu
  -少见dna@horizon.csueastbay.edu
  - roux17@horizon.csueastbay.edu
  - giavita@horizon.csueastbay.edu

Green Biome Institute Contact Info:

- Chris Baysdorfer, PhD (Director)
- We would like to thank all the students who have contributed to this project, including:
  - M.S. Student: Robbi Mickelson
  - Watan8@horizon.csueastbay.edu
  - Psarn46@horizon.csueastbay.edu
  -少见dna@horizon.csueastbay.edu
  - roux17@horizon.csueastbay.edu
  - giavita@horizon.csueastbay.edu

Over 200 plant leaf samples of various Fritillaria were obtained and DNA was isolated using the Gagen Plant DNeasy protocol. DNA samples were quickly chased by nanodrop and Qubit. 100 samples passed quality control and were normalized and sent to Finngene (Beaverton Oregon) for deRAD sequencing. Analysis showed the presence of 11 groups. Four of these correspond to known Fritillaria species, these include F. affinis, F. pustica, F. shinata, and one population of F. purpura. Three groups correspond to F. leucos with one including plants from LA, Riverside and San Diego counties, a second from SLO and Monterey counties and a third (close to the second) corresponding to plants from Santa Barbara County. One group corresponds to all plants labeled as F. rogersi/greggii and F. leucos of Nicasio reservoir. One group corresponds to plants labeled as F. agrestis from Contra Loma and Diablo foothills and F. liliacea from Jepson prairie. One group corresponds to plants labeled as F. agrestis from Contra Loma and Diablo foothills and F. liliacea from Jepson prairie. One group corresponds to plants labeled as F. nigri from Lake Chabot and F. agrestis from Brushy point. Theoloto, Tilden, and Peaks. Eight group corresponds to plants labeled as F. nigri from Lake Chabot and F. agrestis from Brushy point. Theoloto, Tilden, and Peaks. Nine group corresponds to plants labeled as F. nigri from Lake Chabot and F. agrestis from Brushy point. Theoloto, Tilden, and Peaks. Ten group corresponds to plants labeled as F. nigri from Lake Chabot and F. agrestis from Brushy point. Theoloto, Tilden, and Peaks. Eleven group corresponds to plants labeled as F. nigri from Lake Chabot and F. agrestis from Brushy point. Theoloto, Tilden, and Peaks. This work was funded by the San Francisco Public Utilities Commission as a project entitled Restoring Plants on Serpentine Sites: Fritillaria Genomics.