Finding floral scent in the genome- assembly to association

Trinity Barnes[1,2], Haley Carter[2,3], Daja Dortch[2,4], Jeremie Fant [2], Krissa Skogen [2] , Norman Wickett[2]

[1]Rhodes College, Memphis, TN [2]Chicago Botanic Garden, Glenco, IL [3]Northwestern University, Evanston, IL [4]Disney II Magnet School Chicago, IL

Sampling

DNA Extraction

Sequencing

Assembly

BLAST

Annotation

Introduction

An organism's traits, from how tall you are or the shape of a tree's leaves, result at least in part from the expression of its genes. An individual's genes are located in their genome, the collection of all their DNA. By analyzing a plant genome, you can be able to determine why some plants are certain colors or produce certain smells. For this project we are assembling, or piecing together, the genome of the plant *Oenothera harrigntonii* in order to find the gene or genes responsible for its floral scent. Some plants of this species have a floral scent that is characterized by the chemical compound linalool. Linalool is produced primarily in the petal tissues via a chemical reaction that is controlled by an enzyme called linalool synthase. To produce linalool synthase, a plant needs a linalool synthase gene in its genome. Because only some individuals of this species produce linalool, we hypothesize that these plants have different versions of linalool synthase genes - one that is active and one that is not.

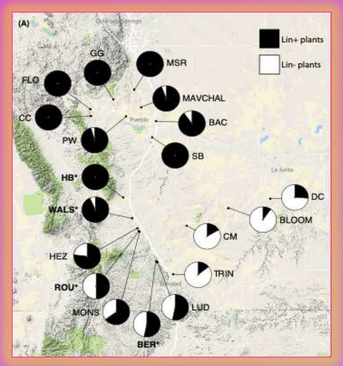

Figure 1 Map with proportion of linalool-producing plants at sites in Colorado, USA[1].

Figure 2 T. Barnes holding extracted DNA.

Methods

- Assembled long (PacBio) sequence reads into contigs with Canu[2]
- Searched the NCBI DNA sequence database[3] for linalool synthase genes & created a FASTA file containing sequences labeled as R linalool synthase genes
- Created a BLAST[4] database from our assembled genome contigs
- Used BLAST to find matches between these genes and our database
- Annotated the matching contig with the Augustus gene predictor[5]
- Converted the predicted gene DNA sequence to protein sequence
- Analyzed protein sequences for potential terpene synthase genes with Terzyme[6]
- Extracted DNA[7] for genome-wide association study (GWAS) that will identify linalool synthase alleles with linalool production

```
tig00008027

25953   TGAACGCGGAGCCCGTTCCGAAT-GCTT----CTACGATCCGGAAGTTCTCGCGAAGAGAA  26008
        |||| ||||||||||||||||||||||    || ||||||||||||||||||||||||||||
134285  TGAATGCGGAGCCCGTTCCGAATGCTTGCTTCTCCGATCCGGAAGTTCTCGCGAAGAGAA  134226

26009   TAAGATGCTGCTCCCCTTCCCTCTTTCTTTTCCCGCTTTGCTAATCTTCCCCTCTAACGC  26068
        ||||||||||||||||          |   |  | |    | ||||||||||||||| ||
134225  TAAGATGCTGCTCCC----------AC---T--C-C----C-AATCTTCCCCTCTAAGGC  134187

26069   GGGCCGGGCGGCGACGGGCGCGGGAGGAAGAAACCTAAGAGAAGA  26113
        ||||||||||||| ||||||||||||||||||||||||||||||||
134186  GGGCCGGGCGGCGGCGGGCGCGGGAGGAAGAAACCTAAGAGAAGA  134142

25953   TGAACGCGGAGCCCGTTCCGAA----TGCTTCTACGATCCGGAAGTTCTCGCGAAGAGAA  26008
        ||||| ||||||||||||||||    ||||||| ||||||||||||||||||||||||||
67359   TGAACTCGGAGCCCGTTCCGAATGCTTGCTTCTCCGATCCGGAAGTTCTCGCGAAGAGAA  67418

26009   TAAGATGCTGCTCCCCTTCCCTCTTTCTTTTCCCGCTTTGCTAATCTTCCCCTCTAACGC  26068
        ||||||||||||||    |  ||   |    |          |||||||||||||| ||
67419   TAAGATGCTGCTCCCA--C--TC---C----C----------AATCTTCCCCTCTAAGGC  67457

26069   GGGCCGGGCGGCGACGGGCGCGGGAGGAAGAAACCTAAGAGAA  26111
        |||||||||||| ||||||||| ||||||||||||||||||||
67458   GGGCCGGGCGGCGGCGGGCGCGGTAGGAAGAAACCTAAGAGAA  67500
```

Figure 3 Blast hit alignment between *Oenothera* contig and *Nicotiana* linalool synthase.

Results & Conclusion

Our BLAST search returned a match between our assembled *Oenothera* sequences and a R linalool synthase gene from *Nicotiana tomentosiformis*. Augustus annotated 34 genes within one of these contigs, including one monoterpene synthase gene from subfamily TPS-a, as identified by Terzyme. Because linalool is a monoterpene, the combination of a BLAST hit to a known *Nicotiana* R-linalool synthase and the annotation of a monoterpene synthase gene on this contig provides strong evidence for this to be the location of a linalool synthase gene in *Oenothera harringtonii*. We also extracted DNA from 118 individuals, to be included in the GWAS which will link these findings to our measured trait (linalool production).

References

1 Skogen unpub. data **2** Koren *et al.* 2017. Genome Research. **3** NCBI Resource Coordinators. 2018. Nucleic Acids Res. **4** Altschul *et al.* 1990. J. Mol. Bio **5** Stanke & Morgenstern 2005. Nucleic Acids. Res. **6** Priya *et al.* 2018. Plant Methods **7** Doyle & Doyle. 1987. Phytochemical Bulletin

Figure 4 *Oenothera haringonii.* Common name- Arkansas Valley evening primrose

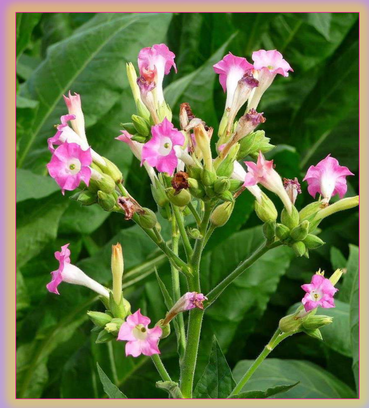

Figure 5 *Nicotiana tomentosiformis.* Source: Blühende Tabakpflanze 3268zauber (CC BY-SA 3.0)

Further Research

Our contigs will be further merged into a chromosome-level genome assembly. This genome will be re-annotated using transcriptome evidence from past projects, as well as comparisons with other species (as in the *Nicotiana* example above). The DNA extracted from over 100 individuals this summer will be prepared into libraries for Illumina sequencing. These sequences will be assembled against the finished genome and specific alleles of linalool-producing and non-producing plants will be compared in a genome-wide association study. This study should identify those nucleotide polymorphisms that correspond to the difference in floral scents, which we hypothesize will be located within a linalool synthase gene, likely the candidate gene we identified here.

Acknowledgements

We would like to thank Hillary Noble, Jeremie Fant, Norman Wickett and the REU Cohort for an amazing experience this summer. We'd like to thank NSF-REU grant DBI-1757800 and NSF grant DEB-1342873 (to KS, JF, NW) for support.

Sampling

DNA Extraction

Library Preparation

Sequencing

Assembly to Genome

GWAS