Genome Sequence of *Medicago sativa*: Cultivated Alfalfa at the Diploid Level (CADL)

Maria J. Monteros, Joann Mudge, Andrew D. Farmer, Nicholas P. Devitt, Diego A. Fajardo, Thiru Ramaraj, Xinbin Dai, Zhaohong Zhuang, Peng Zhou, Joseph Guhlin, Christopher D. Town, Patrick X. Zhao, Jason R. Miller, Kevin A. Silverstein, E. Charles Brummer, Nevin D. Young

NAAIC
Madison, WI
July 13, 2016
# Alfalfa - *Medicago sativa* Complex

<table>
<thead>
<tr>
<th></th>
<th>Diploid 2X</th>
<th>Tetraploid 4X</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Medicago sativa</em> subsp. <em>caerulea</em></td>
<td><img src="image1" alt="Purple Flowers" /></td>
<td><img src="image2" alt="Purple Flowers" /></td>
</tr>
<tr>
<td><em>Medicago sativa</em> subsp. <em>falcata</em></td>
<td><img src="image3" alt="Yellow Flowers" /></td>
<td><img src="image4" alt="Yellow Flowers" /></td>
</tr>
<tr>
<td><em>Medicago sativa</em> subsp. <em>hemicycla</em></td>
<td><img src="image5" alt="Purple Flowers" /></td>
<td><img src="image6" alt="Purple Flowers" /></td>
</tr>
</tbody>
</table>

- **Warm weather**
- **Cold tolerant**
Alfalfa (Medicago sativa) Genetics

Basic chromosome number (x) = 8

Diploid
2n = 2x = 16

Tetraploid
2n = 4x = 32
Diploid Genome vs. Tetraploid Alfalfa

Images provided by Haibao Tang
Cultivated Alfalfa at the Diploid Level (CADL)

- Population developed by Ted Bingham
- Developed from cultivated tetraploids over 10 years to produce a diploid form
- Used 4x-2x cross method and backcrossing to increase cultivated germplasm background
- Simpler to analyze and assemble the genome than the tetraploid alfalfa grown commercially
- Genotype HM342 sequenced as part of the *M. truncatula* HapMap project

Role of Plant Genomes

Genes contain instructions for making proteins.

Proteins act alone or in complexes to perform many cellular functions.

Growth
Development
Adaptation
Stress tol.
Quality
Persistence

Source: Joann Mudge, NCGR
CADL - PacBio *de novo* Assembly (V.0.95)

Falcon 0.2 + SSPACE + Quiver

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Contigs</td>
<td>6,593</td>
</tr>
<tr>
<td>Contig Length</td>
<td>1,254,734,629</td>
</tr>
<tr>
<td>Contig N50</td>
<td>547,092</td>
</tr>
<tr>
<td>Max Contig Size</td>
<td>4,047,589</td>
</tr>
</tbody>
</table>

Source: Joann Mudge, NCGR
CADL Gene Annotation Pipeline (V.0.95)

Evidence
Alfalfa RNA Seq
Alfalfa ESTs
Mt and G.max
Protein

MAKER Pipeline
Gene prediction
Prelim. gene models

Training datasets
Augustus for gene models

Evidence
6,593 scaffolds

BUSCO
annotation quality
956 univ. orthologs in plants (99%)

SPADA
2,433 small peptide gene models

OUTPUT
120,094 prelim gene models

57,054
High conf. gene models

124,527
gene models

Source: Xinbin Dai, Patrick Zhao

CADL Gene Functional Annotation (V.0.95)

MAKER Pipeline
Gene prediction
Prelim. gene models

SPADA
2,433 small peptide gene models

Alignment of CADL and M. truncatula

Orthologs in M. truncatula

InterPro Scan
Protein domain analysis

BLAST
against Unipro database

6,593 scaffolds

57,054 gene models

<table>
<thead>
<tr>
<th>Genome Size</th>
<th>Sequenced size</th>
<th>H_Conf. Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. truncatula 4.0</td>
<td>454-526 MB</td>
<td>370 MB</td>
</tr>
<tr>
<td>Alfalfa CADL</td>
<td>415-430 MB</td>
<td>1,200 MB</td>
</tr>
</tbody>
</table>

Source: Xinbin Dai, Patrick Zhao
Whole Genome BLASTN CADL vs Mt4.0

Good coverage of the gene space

Source: Andrew Farmer, NCGR
Divergent Haplotypes Differ in Gene Content (Chr. 7)

Source: Andrew Farmer, NCGR
Welcome to the Alfalfa Breeder’s Toolbox

The Alfalfa Breeder’s toolbox aims to provide convenient access to alfalfa genomic, genetic and phenotypic datasets deriving from U.S. and international research. We have embarked on a journey to sequence the alfalfa genome and use scientific knowledge to inform plant improvement strategies. Included are genetic markers for use in molecular breeding approaches to develop enhanced alfalfa cultivars that improve forage based production systems regionally and globally. The alfalfa genomics sequencing efforts and the toolbox are a collaboration between multiple organizations and made possible through support from the Noble Foundation and industry partners.

<table>
<thead>
<tr>
<th>Search options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genome Browser Position</td>
</tr>
<tr>
<td>Gene</td>
</tr>
<tr>
<td>Phenotype</td>
</tr>
<tr>
<td>Clone</td>
</tr>
<tr>
<td>Marker</td>
</tr>
</tbody>
</table>
Genome BLAST Search - Alfalfa Breeder’s Toolbox

http://alfalfatoolbox.org/doblast/

CADL (0.95)

*M. truncatula* (4.0)
Available at the *M. truncatula* HapMap Project

http://www.medicagohapmap.org/downloads/cadl
**CADL Dovetail assembly (Anticipated V.1.0)**

(Falcon 0.4 + Quiver + Dovetail + Quiver)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Scaffolds</td>
<td>5,751</td>
</tr>
<tr>
<td>Scaffold Length</td>
<td>1,251,060,667</td>
</tr>
<tr>
<td>Scaffold N50</td>
<td>1,271,357</td>
</tr>
<tr>
<td>Max Scaffold Size</td>
<td>6,073,685</td>
</tr>
</tbody>
</table>

Source: Joann Mudge, NCGR

![Image of plants: plants that represent the scaffolds in the assembly process.](image-url)
Acknowledgements

Univ. of Minnesota
Nevin Young
Roxanne Denny

NCGR
Joann Mudge
Andrew Farmer
Diego Fajardo
Nicolas Devitt
Thiru Ramaraj

Noble Foundation
Patrick Zhao
Xinbin Dai
Jaeyoung Choi
Chunlin He
Perdeep Mehta
Michael Udvardi

UC Davis
Charlie Brummer

Haibao Tang
Christopher Town
Ted Bingham
CADL Scaffolds “Painted” by Mt Chromosomes

Source: Andrew Farmer, NCGR