



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

Diploid
2X

Medicago sativa
subsp. *caerulea*



Medicago sativa
subsp. *falcata*



Medicago sativa
subsp. *hemicycla*



Tetraploid
4X

Medicago sativa
subsp. *sativa*



Warm weather

Medicago sativa
subsp. *falcata*



Cold tolerant

Medicago sativa
subsp. *varia*



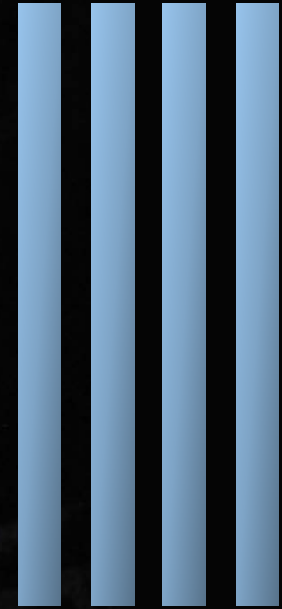
Alfalfa (*Medicago sativa*) Genetics

Basic chromosome number (x) = 8

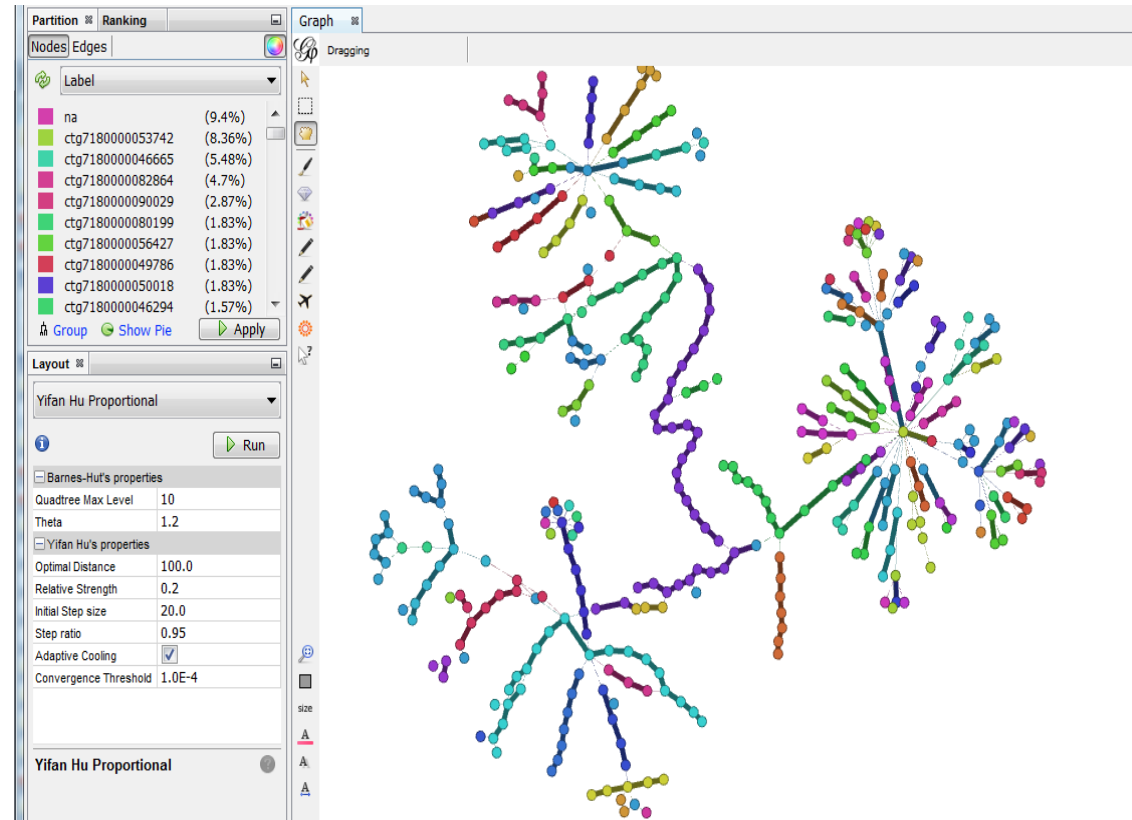
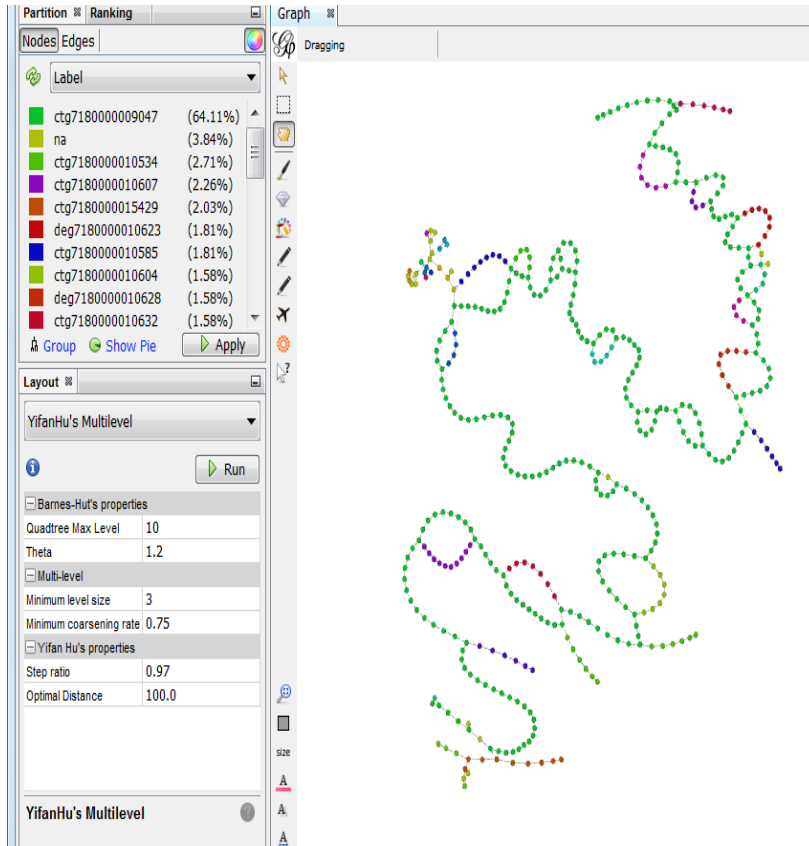
Diploid
 $2n = 2x = 16$



Tetraploid
 $2n = 4x = 32$



Diploid Genome vs. Tetraploid Alfalfa



Images provided by Haibao Tang

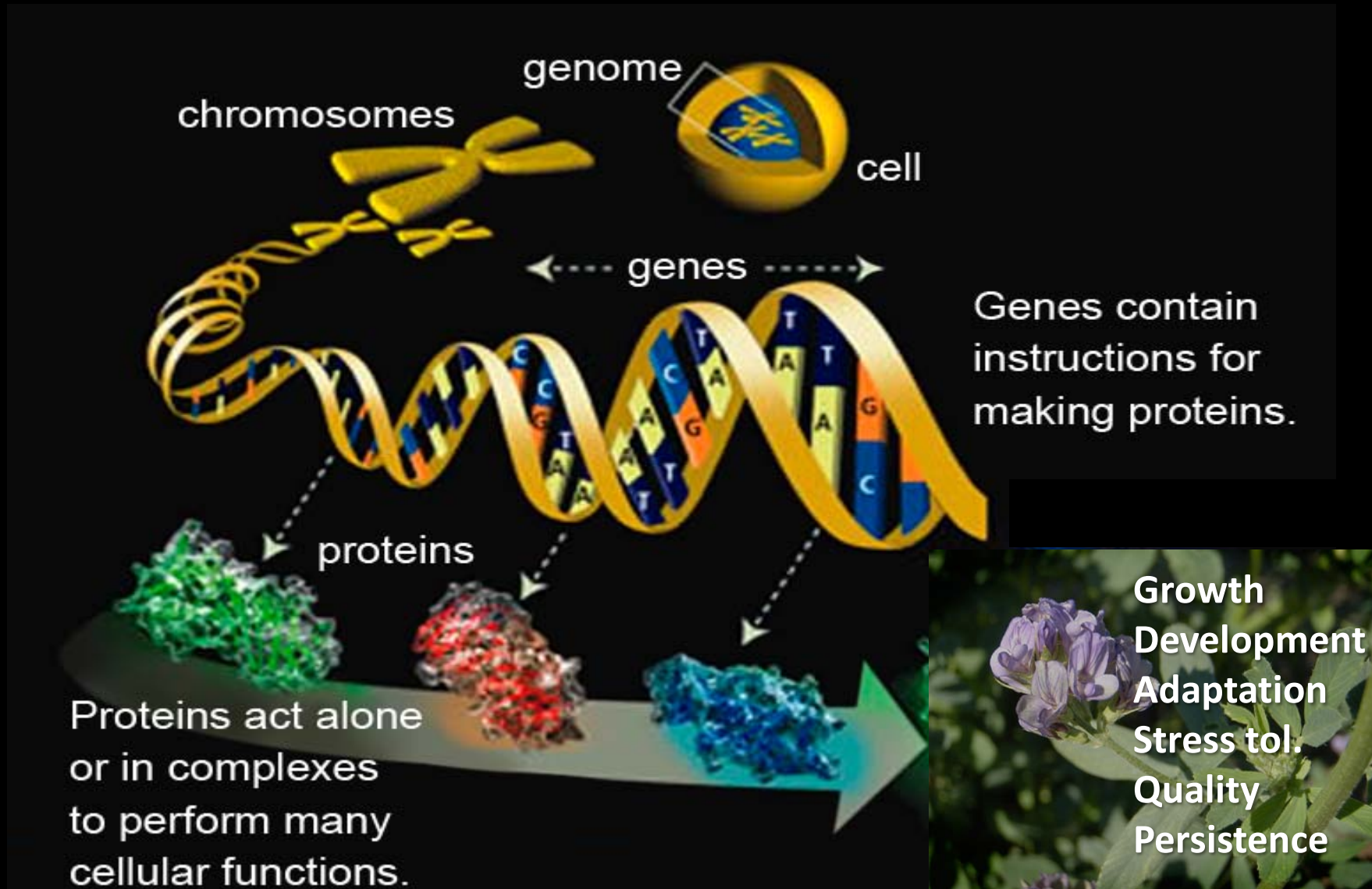
Cultivated Alfalfa at the Diploid Level (CADL)



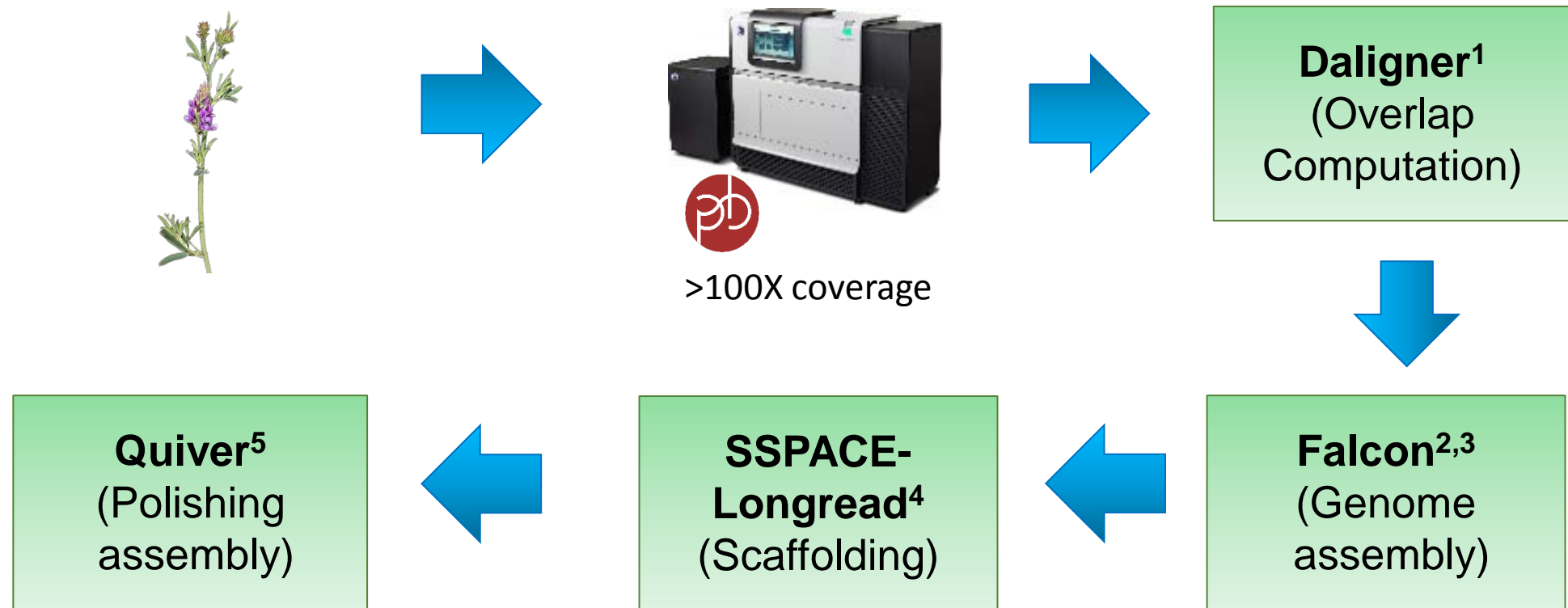
- Population developed by Ted Bingham
- Developed from cultivated tetraploids over 10 years to produce a diploid form
- Used 4x-2x cross method and backcrossing to increase cultivated germplasm background
- Simpler to analyze and assemble the genome than the tetraploid alfalfa grown commercially
- Genotype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

Bingham ET and McCoy TJ. 1979. Cultivated Alfalfa at the Diploid Level: origin, reproductive stability, and yield of seed and forage. *Crop Science* 19: 97-100.

Role of Plant Genomes



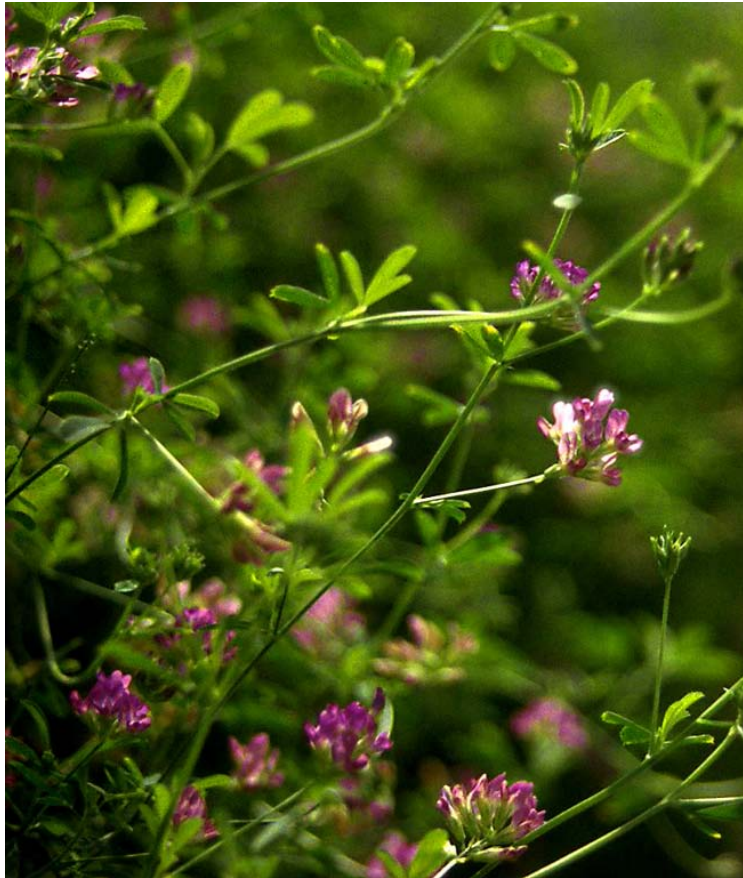
CADL Genome Assembly Pipeline (V.0.95)



1. Myers G. 2014. The Daligner Overlap Library. <https://github.com/thegenemyers/DALIGNER>.
2. Chin C, et al. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods*. 10:563–569.
3. Chin J. 2015. FALCON: experimental PacBio diploid assembler. <https://github.com/PacificBiosciences/FALCON>.
4. Boetzer M, Pirovano W. 2014. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics*. 15:211.
5. PacBio® variant consensus caller (Quiver algorithm). <https://github.com/PacificBiosciences/GenomicConsensus>

Source: Joann Mudge, NCGR

CADL - PacBio *de novo* Assembly (V.0.95)

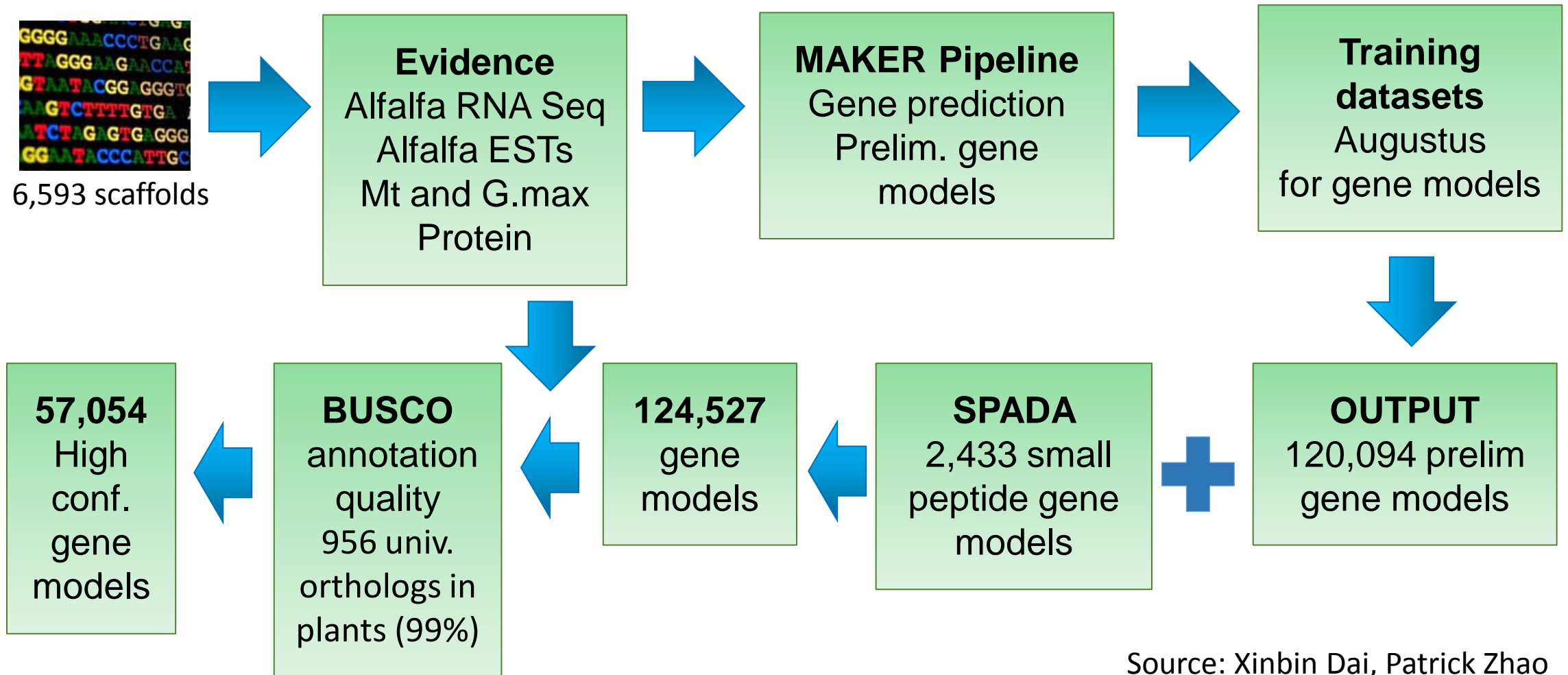


Falcon 0.2 + SSPACE + Quiver

Contigs	6,593
Contig Length	1,254,734,629
Contig N50	547,092
Max Contig Size	4,047,589

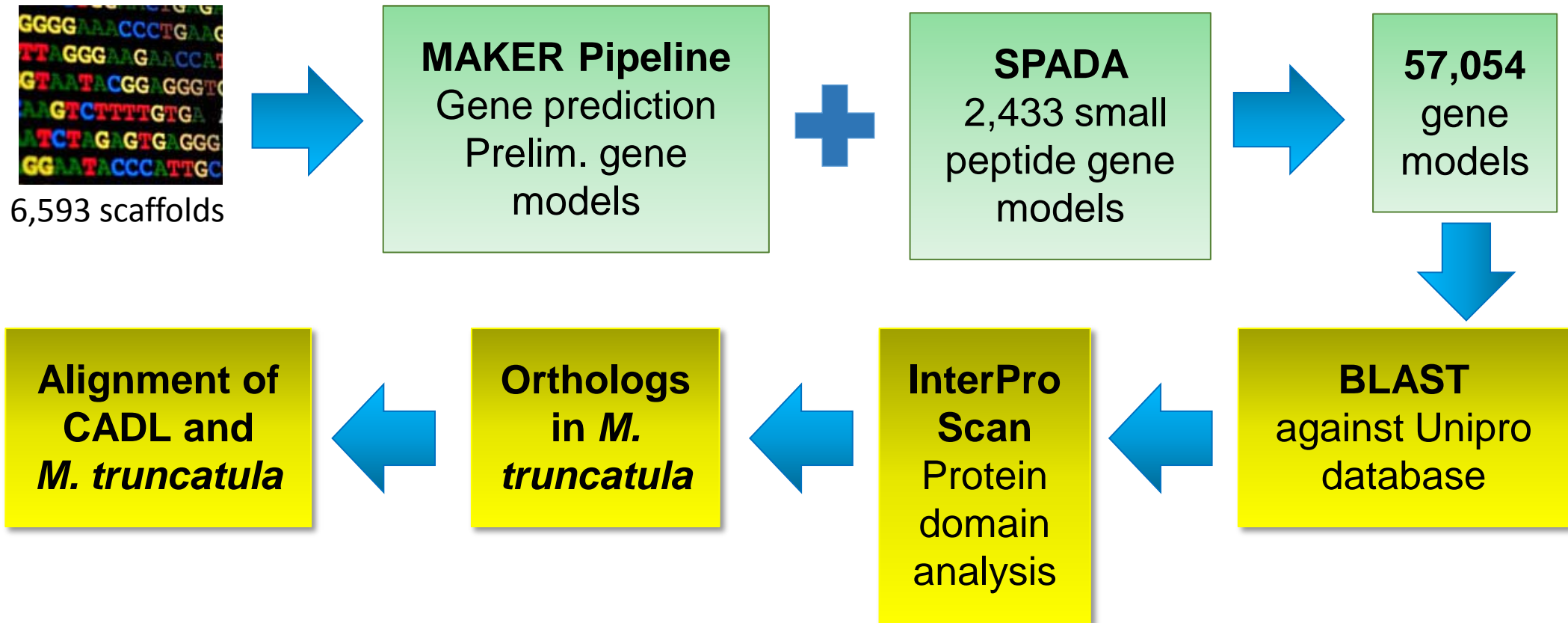
Source: Joann Mudge, NCGR

CADL Gene Annotation Pipeline (V.0.95)



Source: Xinbin Dai, Patrick Zhao

CADL Gene Functional Annotation (V.0.95)

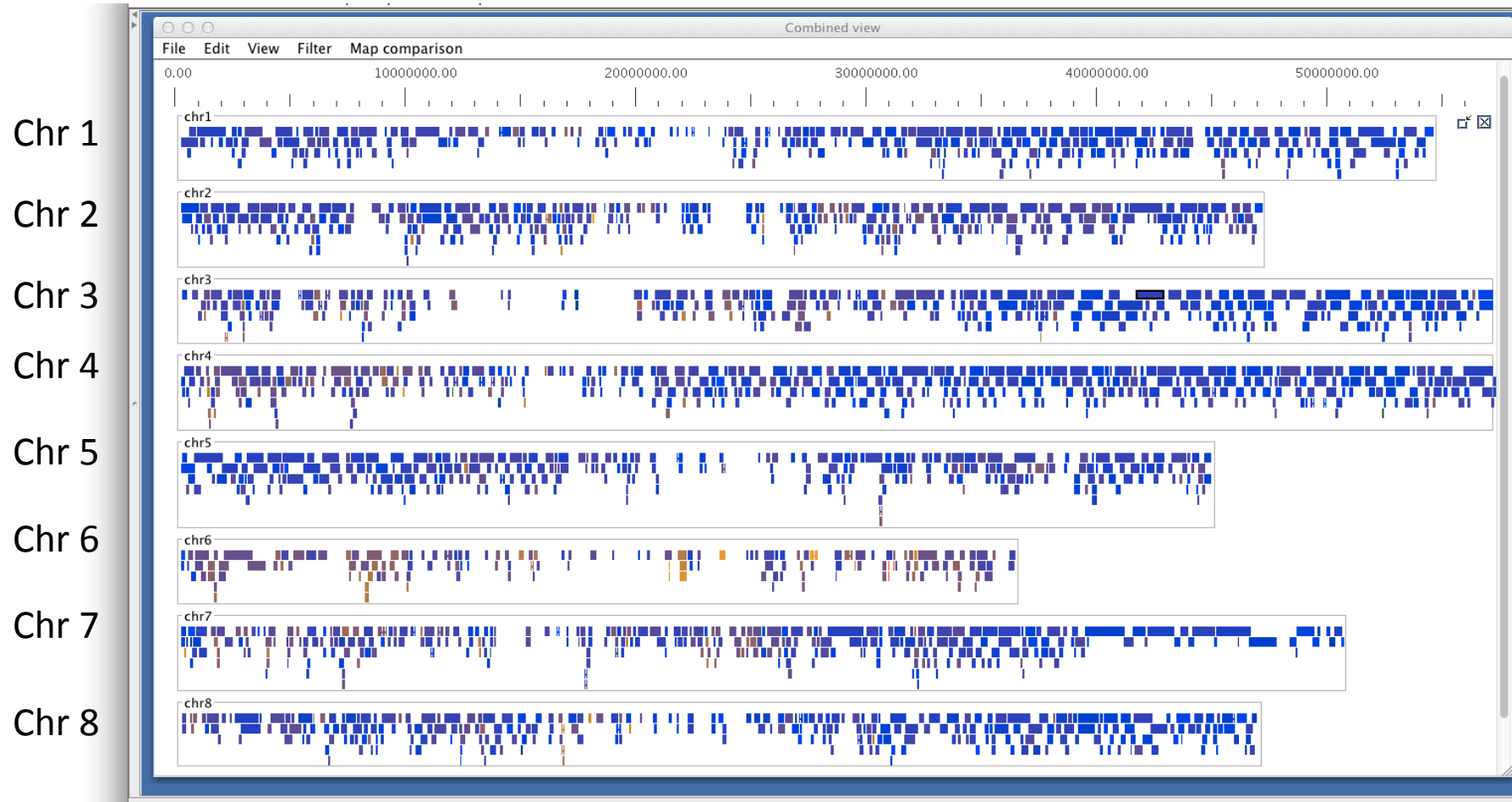


	Genome Size	Sequenced size	H_Conf. Genes
M. truncatula 4.0	454-526 MB	370 MB	31,661
Alfalfa CADL	415-430 MB	1,200 MB	57,054

Source: Xinbin Dai, Patrick Zhao

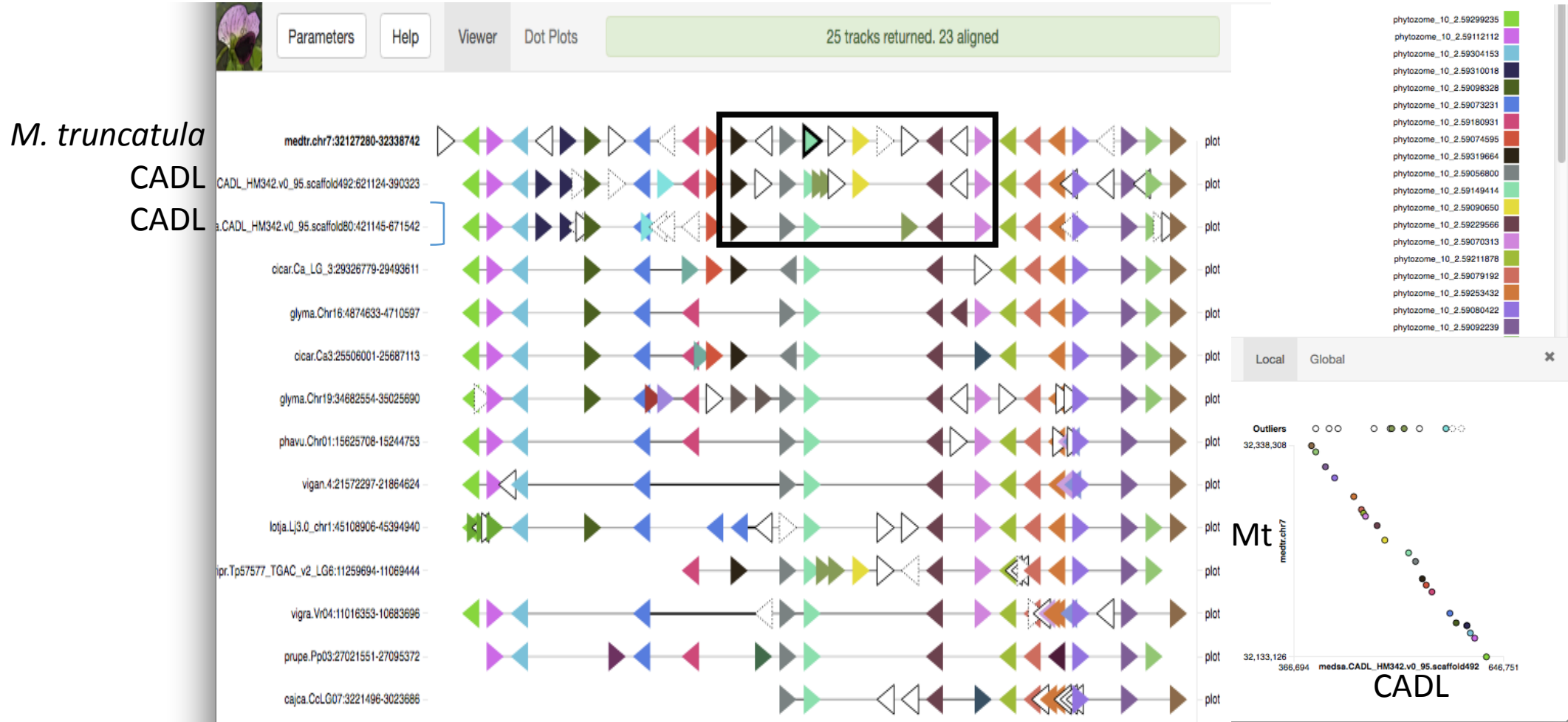
Whole Genome BLASTN CADL vs Mt4.0

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



[See Poster #25](#)

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.

Search options					
Genome Browser Position	M. sativa (CADL) ▾	chr/scaffold	scaffold1	Range 10000 - 100000	Go More options
Gene Annotation	MSAD_035820			Search	More options
Phenotype Gene Expression Condition				Search	More options
Sequence	<pre>>EF654111.2 Medicago sativa subsp. falcata DREB1 AAAACAAACTCTCTCCAATTCCGACTTTACCTCTCAACACCATTTTCCAC TCTATCCAACACATACATATATGATTACTACTAACAACCTCTTCCTATPCA CACTCCATTTCTCAAAGATTTTCTCCCTTCGACGCATCATCACCGAA</pre>			BLAST	More options
Clone Cultivar Population				Search	More options
Marker	MSCWSNP0002			Search	More options

Genome BLAST Search - Alfalfa Breeder's Toolbox

THE SAMUEL ROBERTS
NOBLE
FOUNDATION

The Samuel Roberts Noble Foundation

DOBLAST: The Cultivated Alfalfa at the Diploid Level (CADL) Genome Blast Server

Home BLAST Download

Location: BLAST

<http://alfalfatoolbox.org/doblast/>

Upload query sequence(s) in FASTA format: [\[Load Demo\]](#)
Browse...

or paste sequences below:

- file / input sequence size limit: 100K

Select a database for search:

```
CADL_cds-----NT-----119M--Alfalfa CADL v0.95P cds-----
CADL_genome-----NT-----1G---Alfalfa CADL genome scaffold v0.95P-----
CADL_pep-----PROTEIN-42M---Alfalfa CADL v0.95P protein-----
CADL_splicedTrans-----NT-----125M--Alfalfa CADL v0.95P spliced transcript-----
Mt4.0_cds-----NT-----69M---Medicago truncatula v4.0 JCVI 20130313 cds-----
Mt4.0_genome-----NT-----399M--Medicago truncatula v4.0 JCVI 20130313 genome-----
Mt4.0_pep-----PROTEIN-26M---Medicago truncatula v4.0 JCVI 20130313 protein-----
Mt4.0_splicedTrans-----NT-----88M---Medicago truncatula v4.0 JCVI 20130313 spliced transcript-----
```

BLAST Program:

Output format:

E-value: (default = 10.0) e.g. 1e-3, 10

Applying Filter:

of top hits:

Submit Reset

CADL (0.95)

M. truncatula (4.0)

Available at the *M. truncatula* HapMap Project

<http://www.medicagohapmap.org/downloads/cadl>

Medicago truncatula
HAPMAP PROJECT

[Home](#) [Hapmap](#) [Tools](#) [Downloads](#) [Resources](#) [Contact](#)

[Home](#) > [Downloads](#) > [CADL](#)

CADL Downloads

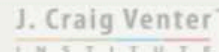
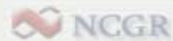
Detailed description of the data file is available from the [readme](#) file (updated 7 June 2016).

We are providing this data pre-publication to facilitate other community research. However, in accordance with the Toronto Agreement, we reserve the right to be the first to publish a genome-wide analysis of the data we have generated. Further, the pre-publication data released here is embargoed for publication except for analyses of individual gene loci or small genomic regions (< 10 kb) for single or multiple accessions. Researchers are encouraged to contact us with queries about referencing or publishing analysis based on pre-publication data obtained via this website.

CADL: Files

Released June 2016

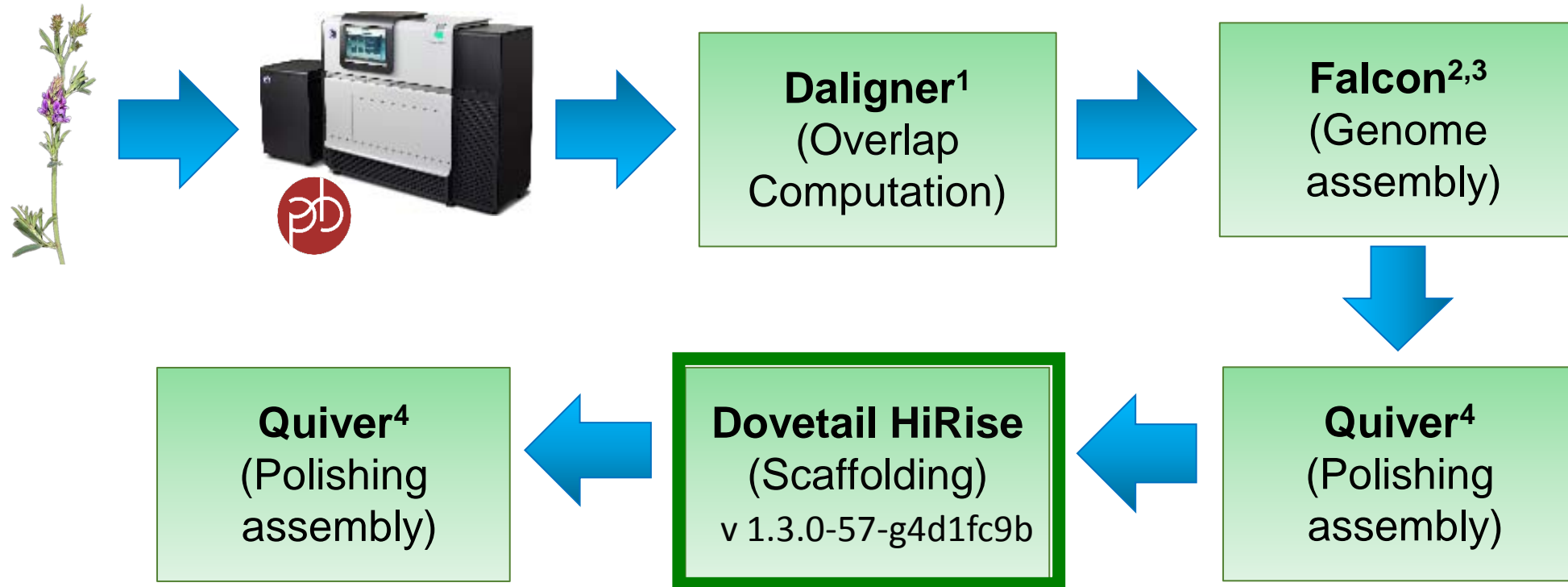
CADL_HM342.v0.95P.fasta.gz	367 MB	2016-06-24
CADL_v0.95p_README.pdf	522 KB	2016-06-24



Site hosted by: [Nevin Young's Lab](#) at University of Minnesota. Managed By: [Joseph Guhlin](#)
Primary Collaborators: [INRA - Montpellier](#) | [ENSAT - Toulouse](#) | [Noble Foundation](#)
Co-Principal Investigators: [University of Minnesota](#) N. Young, P. Tiffin, M. Sadowsky, B. Stupar, K. Silverstein | [NCGR](#) J. Mudge, A. Farmer
[Boyce Thompson Institute](#) M. Harrison | [Hamline University](#) B. Martinez-Vaz

THE SAMUEL ROBERTS
NOBLE
FOUNDATION

CADL Genome Assembly Pipeline (~V.1.0)



1. Myers G. 2014. The Daligner Overlap Library. <https://github.com/thegenemyers/DALIGNER>.
2. Chin C, et al. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods*. 10:563–569.
3. Chin J. 2015. FALCON: experimental PacBio diploid assembler. <https://github.com/PacificBiosciences/FALCON>.
4. PacBio® variant consensus caller (Quiver algorithm). <https://github.com/PacificBiosciences/GenomicConsensus>.

Source: Joann Mudge, NCGR

CADL Dovetail assembly (Anticipated V.1.0)

(Falcon 0.4 + Quiver + Dovetail + Quiver)



Scaffolds	5,751
Scaffold Length	1,251,060,667
Scaffold N50	1,271,357
Max Scaffold Size	6,073,685

Source: Joann Mudge, NCGR

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





THE SAMUEL ROBERTS
NOBLE
FOUNDATION

CADL Scaffolds "Painted" by Mt Chromosomes



Source: Andrew Farmer, NCGR