A Comprehensive High-Density Genetic Linkage Map for Loblolly Pine (*Pinus taeda* L.)

Pedro J. Martínez-García [1]
Kristian Stevens [2]
Jill L. Wegrzyn [1]
John Liechty [1]
Marc Crepeau [2]
Charles H. Langley [2]
David B. Neale [1]

[1] Dept. of Plant Sciences;
[2] Dept. of Evolution and Ecology
University of California, Davis

UCDAVIS PLANT SCIENCES

UCDAVIS Evolution & Ecology

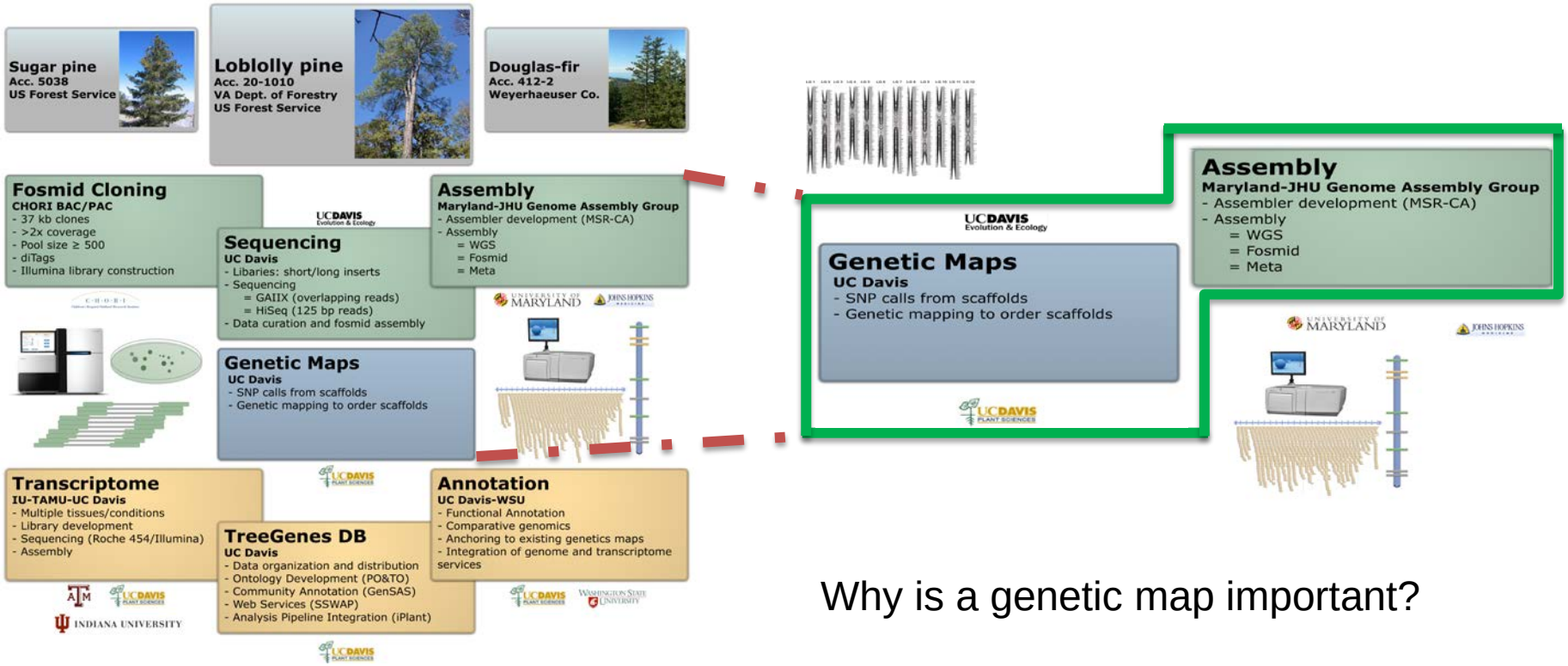
PAG XXI,
January 12th 2013

USDA United States Department of Agriculture
National Institute of Food and Agriculture

PineRefSeq

Introduction

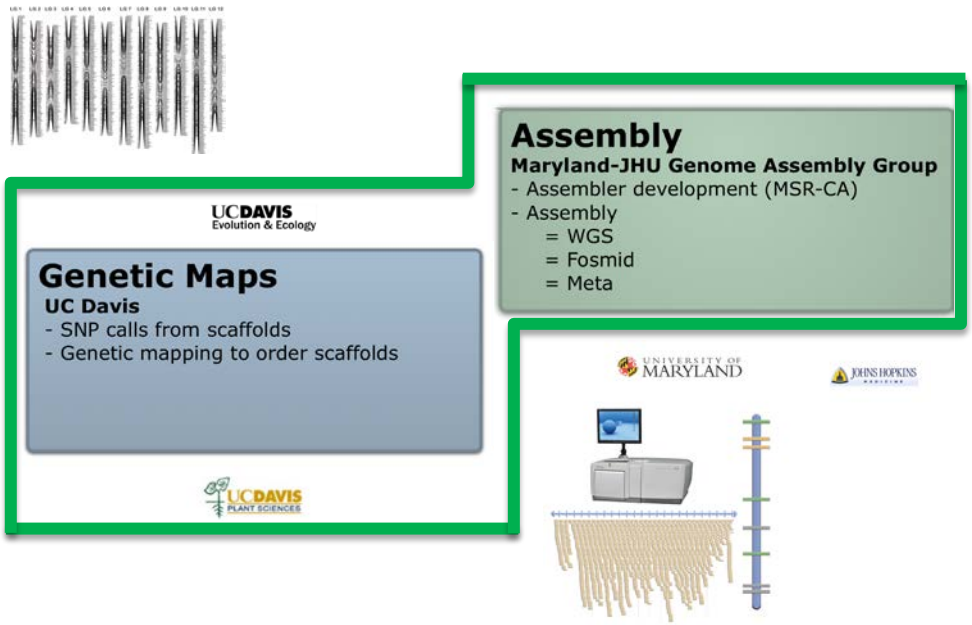
Pine Genome Reference Sequence Project

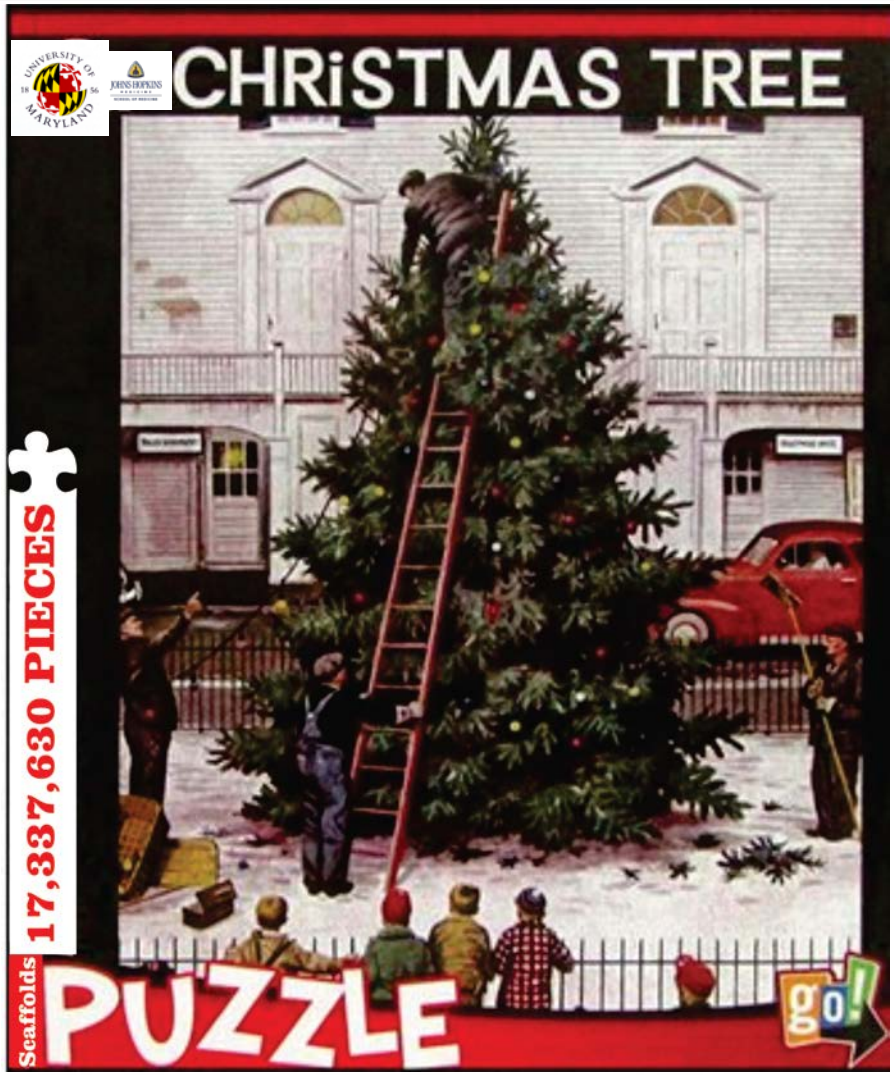


Why is a genetic map important?

- One main objective of a genetic map is to help to order scaffolds in the assembly
- Previous plant genome projects of several plants used genetic maps to validate and improve quality of their genome (Watermelon (Ren et al. 2012); Banana (D'Hont et al. 2012); Sunflower (Grassa et al.2012); *Populus trichocarpa* (Tuskan et al. 2006)).

Introduction





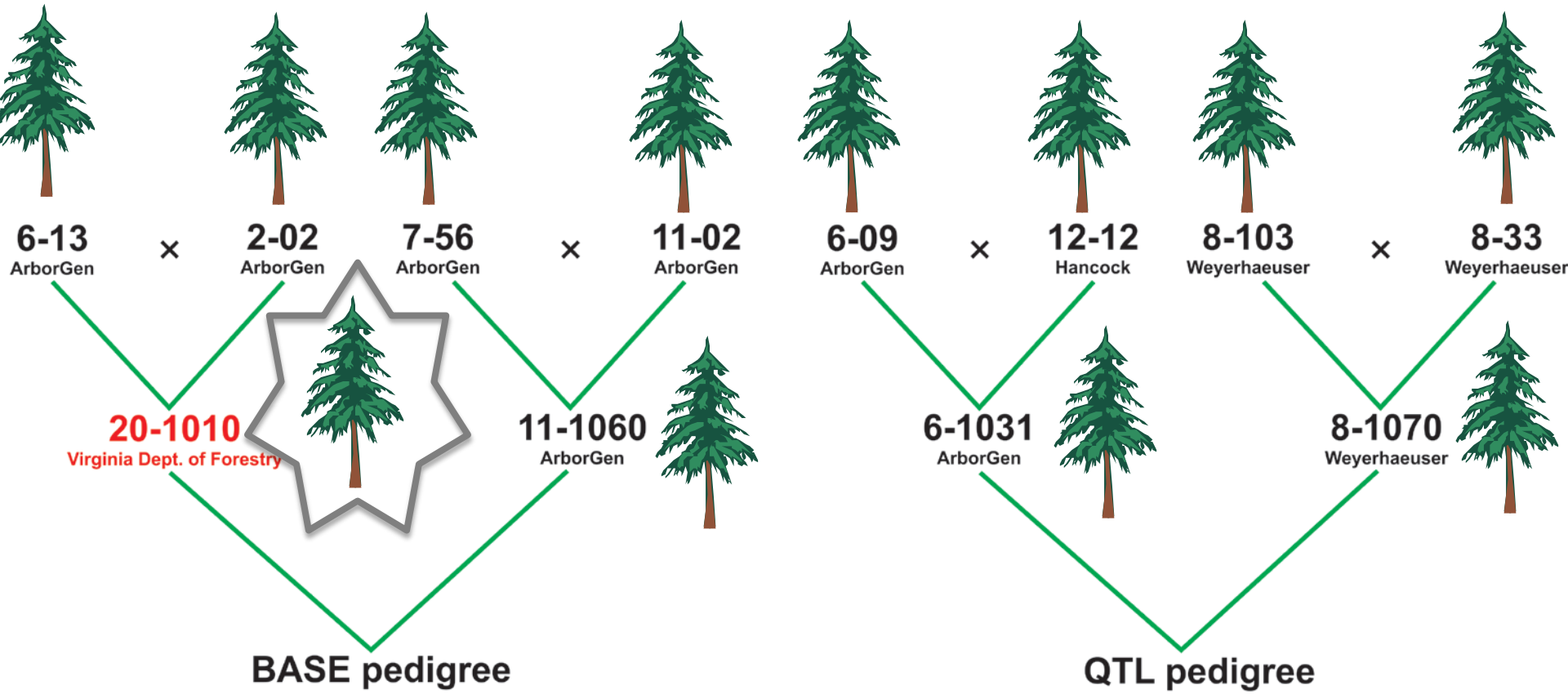
01/11/2013

Loblolly pine assembly v0.8

Christmas gifts

A high-density map can be a useful platform for the assembly of the pine genome

Reference Mapping Populations



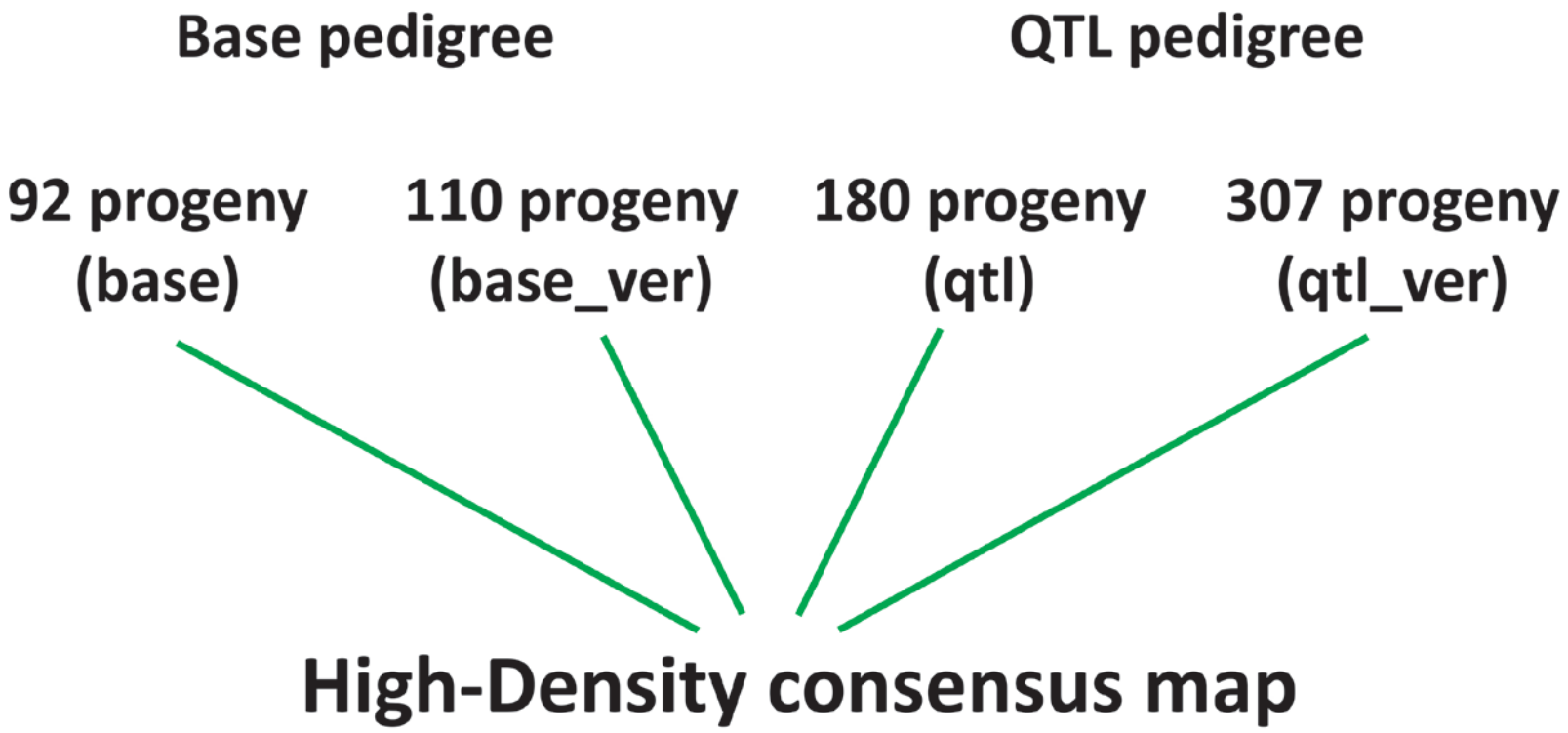
Reference Mapping Populations

Authors	Type of population	Type of map	LG	cM (average)
Devey et al. 1994	base pedigree	Sex-averaged map	20	632
Sewell et al. 1999	base and qtl pedigree	Integrated map (consensus)	12	1300
Brown et al. 2001	base and qtl pedigree	Integrated map (consensus)	12	1165
Sewell et al. 2002	qtl pedigree	Sex-averaged map	14	1237
Zhou et al. 2003	base pedigree	Sex-averaged map	15	795
Groover et al. 2004	qtl pedigree	Maternal map	18	519
Eckert et al. 2009	base and qtl pedigree	Integrated map (consensus)	12	1227
Eckert et al. 2010	qtl pedigree	Sex-averaged map	12	1898 (1.8cM)
Echt et al. 2011	base and qtl pedigree	Integrated map (consensus)	12	1429 (3.3cM)

Source: TreeGenes (CMap)

Map Construction

Selected Mapping Progeny



Selected Markers

- Isozyme
- Random Amplification of Polymorphic DNA (RAPD)
- Restriction Fragment Length Polymorphism (RFLPs)
- Expressed Sequenced Tags (ESTs)
- Single Sequence Repeat (SSRs)
- Single Nucleotide Polymorphism (SNPs)

4168 total makers

~ 90% SNPs

Devey et al. 1994; Sewell et al. 1999; Brown et al. 2001; Sewell et al. 2002; Zhou et al. 2003; Groover et al. 2004; Eckert et al. 2009; Eckert et al. 2010

National Science Foundation
WHERE DISCOVERIES BEGIN

Allele Discovery of Economic Pine Traits 2 (ADEPT2) Re-sequencing project

- ***Objective 2 – Discover alleles for 5000 candidate genes using a high-throughput sequencing pipeline at Agencourt Bioscience Corporation.***

National Science Foundation
WHERE DISCOVERIES BEGIN

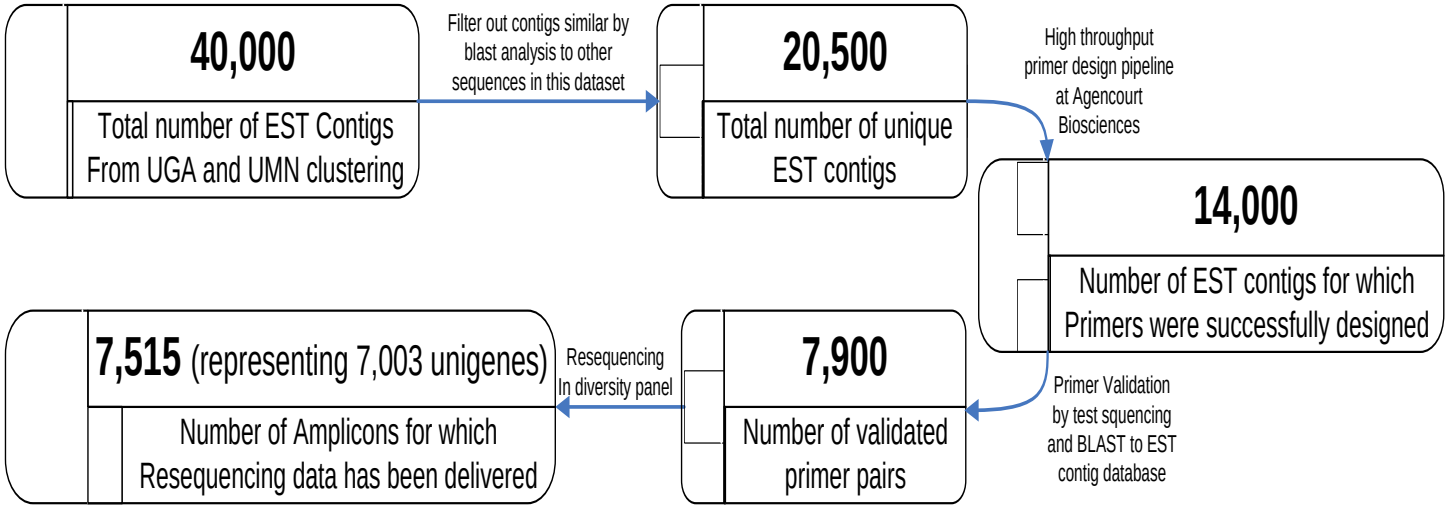
Allele Discovery of Economic Pine Traits 2 (ADEPT2) Re-sequencing project

- ***Objective 2 – Discover alleles for 5000 candidate genes using a high-throughput sequencing pipeline at Agencourt Bioscience Corporation.***

Diversity panel

	Provenance	State	County
1	COASTAL	AL	SUMTER
2	PIEDMONT	AL	TALLAPOOSA
3	SW AR	AR	Pike
4	FLORIDA	FL	LEVY
5	MARION	FL	MARION
6	COASTAL	GA	LIBERTY
7	PIEDMONT	GA	MERIWETHER
8	Livingston Parish	LA	Livingston
9	COASTAL	MS	JASPER

	Provenance	State	County
10	PIEDMONT	MS	TISHOMINGO
11	PIEDMONT	NC	ANSON
12	COASTAL	NC	ONSLOW
13	PIEDMONT	SC	EDGEFIELD
14	COASTAL	SC	JASPER
15	COASTAL	SC	WILLIAMSBURG
16	East Central	TX	Grimes
17	PIEDMONT	VA	GREENSVILLE
18	COASTAL	VA	SUSSEX



- ***Objective 2 – Discover alleles for 5000 candidate genes using a high-throughput sequencing pipeline at Agencourt Bioscience Corporation.***

High Quality Validated Subset of ADEPT2 Loci

- 5599 of the 7500 loci have one or more assembled read pairs that meet our quality criteria
- Total base count: 22,622,243
- Average successful sequence count per locus: 11.5
- Average phred/phrap quality score: 80.6 (and we currently reject any assembled read pair if any of its bases has a quality score below 30, among other criteria).

Genotyping

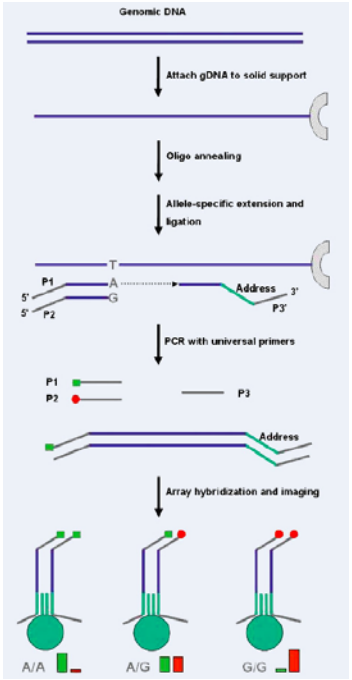
3750 SNPs

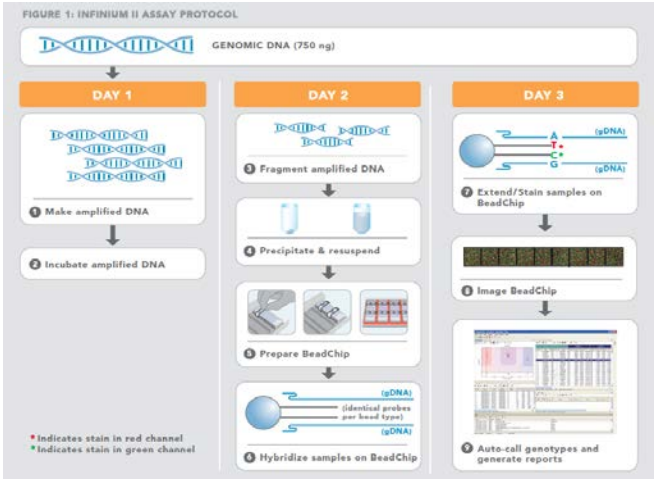
- 1 GoldenGate™ assay 384 SNPs

- 3 Infinium assay (between 7K to 8K SNPS on each)

Success rate 63%

GoldenGate™ assay





Infinium assay

- Isozyme
- RAPD
- RFLPs
- EST
- SSR

Several published sources of genetic markers

Devey et al. 1994; Sewell et al. 1999; Brown et al. 2001,2003; Sewell et al. 2002; Krutovsky et al. 2004; Eckert et al. 2009; Echt et al. 2011

Map Construction

- 689 progeny
- Number total markers 4168

- JoinMap® 4.1
The ML algorithm for outbred species does map integration over the two separate parental maps by averaging and interpolation (Van Ooijen et al. 2011)

- Four different sex-averaged maps

- Consensus map
Map integration over multiple segregating populations is done by regression mapping, as in all previous JoinMap versions.
(Van Ooijen et al. 2011)

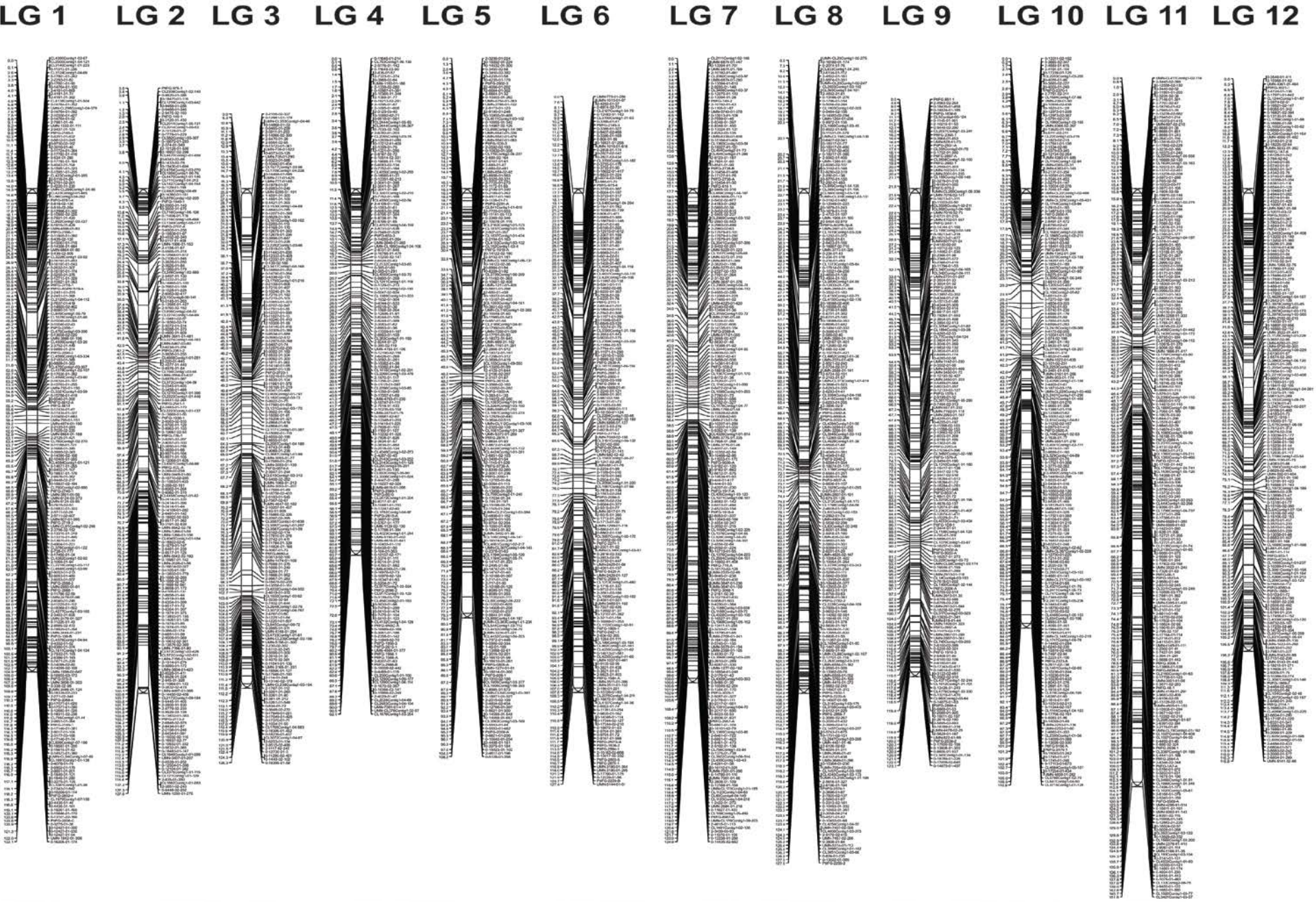
Summary High-Density consensus map

Linkage group	Markers	Genetic Distance (cM)	Unique position	Largest GAP
1	222	122.114	214	5.819
2	200	127.63	198	10.799
3	184	126.259	179	8.602
4	186	92.068	181	2.904
5	198	108.171	186	6.409
6	195	127.449	191	7.634
7	222	124.945	213	2.681
8	228	127.618	218	7.291
9	189	123.315	185	4.628
10	206	110.846	201	4.394
11	232	151.57	225	5.855
12	194	116.83	191	5.302
Total	**2456**	**1458.815**	**2382**	

1458cM/2382markers= **0.612433 cM**

It is the most saturated map in *Pinus taeda* to date

All pines have 12 pairs of chromosomes (Saylor 1961) with an estimated total map length of 1500 cM (Remington et al. 1999)


LG 1
LG 2
LG 3
LG 4
LG 5
LG 6
LG 7
LG 8
LG 9
LG 10
LG 11
LG 12

Summary High-Density consensus map

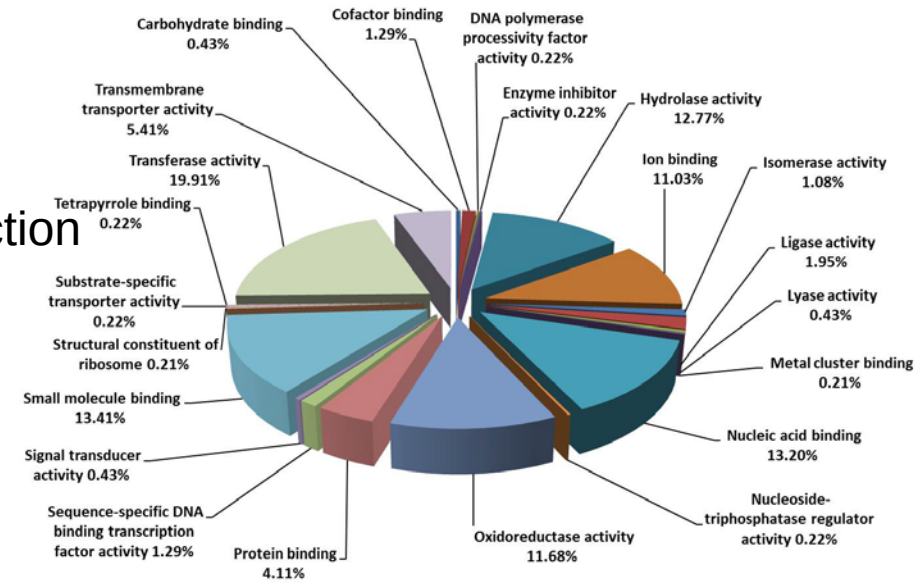
- 2298 SNPs markers mapped
- 156 others markers (SSR, RFPLs, EST)

- **70% (1633 SNPs)** are ADEPT2 loci

- 1428 SNPs are in annotated gene

- ~33% (463) present a Gene Ontology Function

GO Function

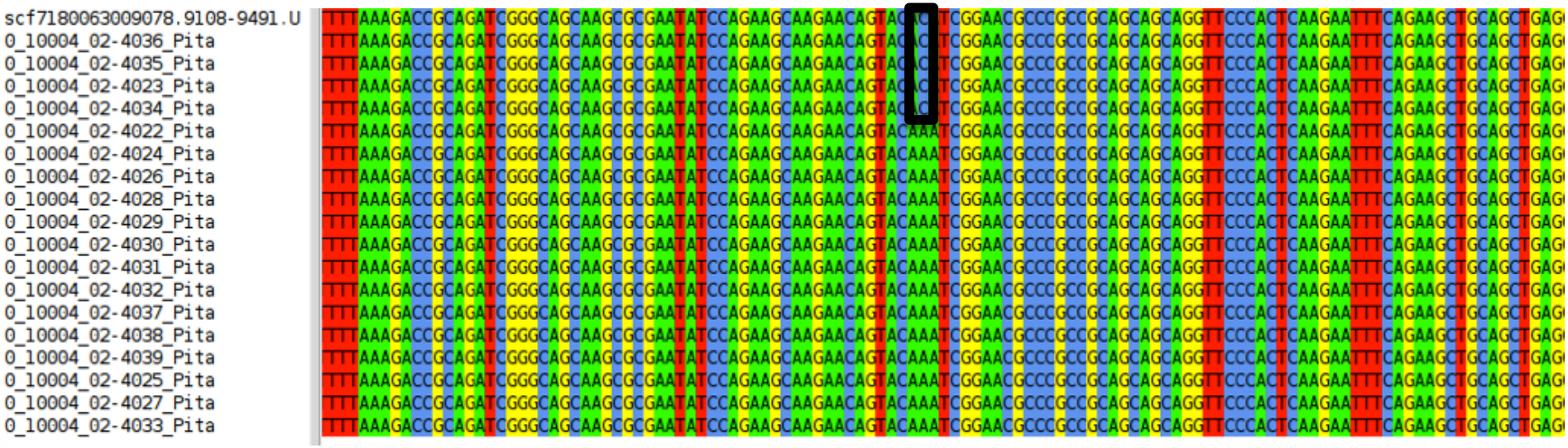


Scaffold anchoring and orientation

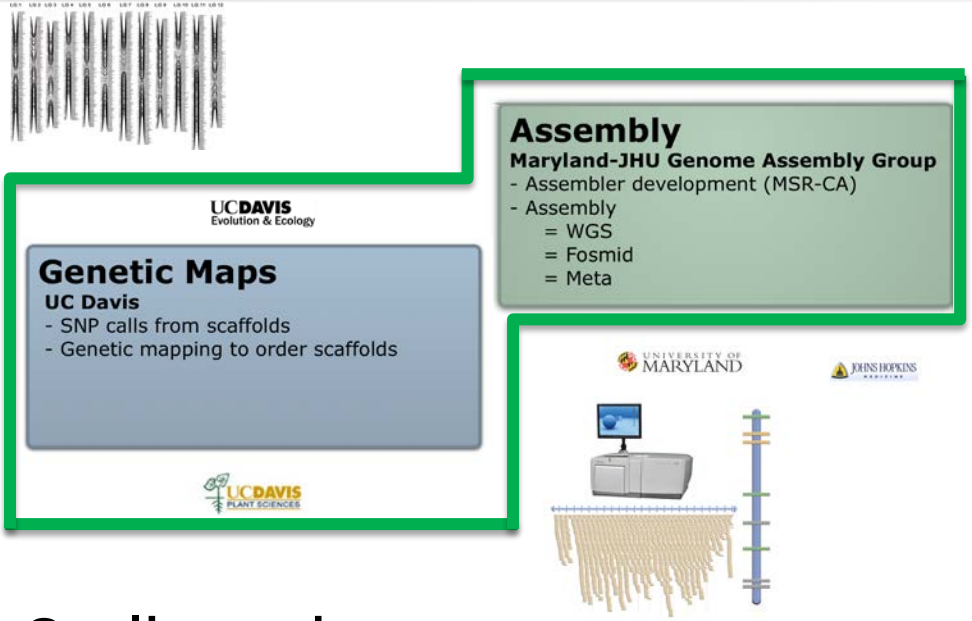
Anchoring ADEPT2 Loci to the Genome

- We aligned 5599 ADEPT2 loci to the Loblolly pine assembly v0.8
  - BLAST + MUSCLE
  - Filtered for 99% identity and 99% coverage
  - 5009 loci met filtering criteria (90%)

- For 3733 of the 5009 aligned loci the genome exactly matches one or more haplotypes of the diversity panel.



Preliminary estimation of scaffold anchoring and orientation



1635 SNPs from Adept2 aligned to genome

1096 SNPs aligned alone to a scaffold

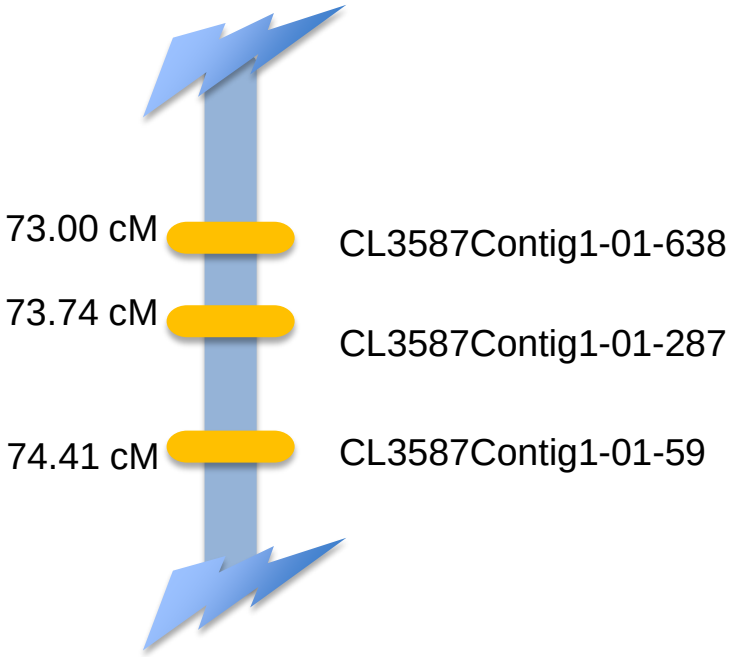
539 SNPs aligned with another SNP to a scaffold

WGS Scaffolds

- Based on the number of markers assigned, scaffolds can be hierarchically classified into not-ordered, ordered, oriented, and validated
  - 1333 WGS scaffolds were ordered by this map
  - 242 WGS scaffolds were also oriented
  - 43 WGS scaffolds were also validated

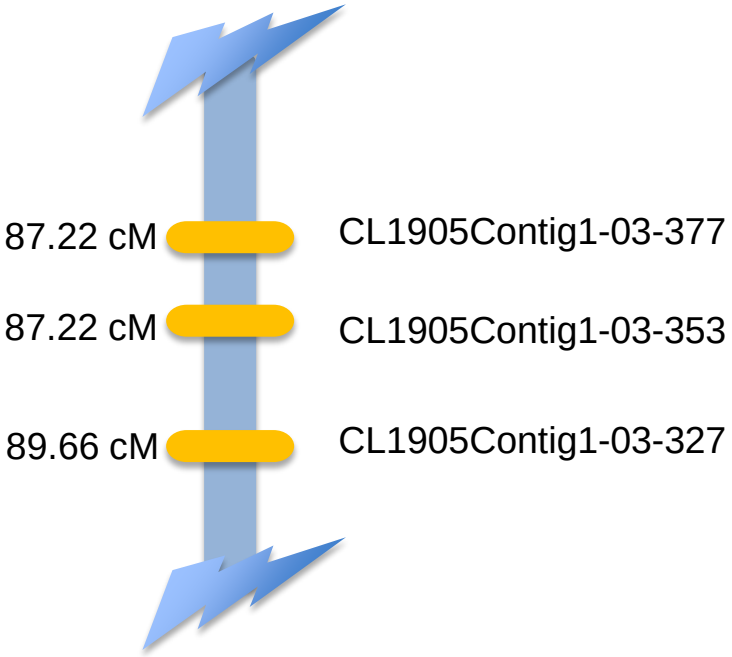
Example of validated scaffold

Scf7180063292642

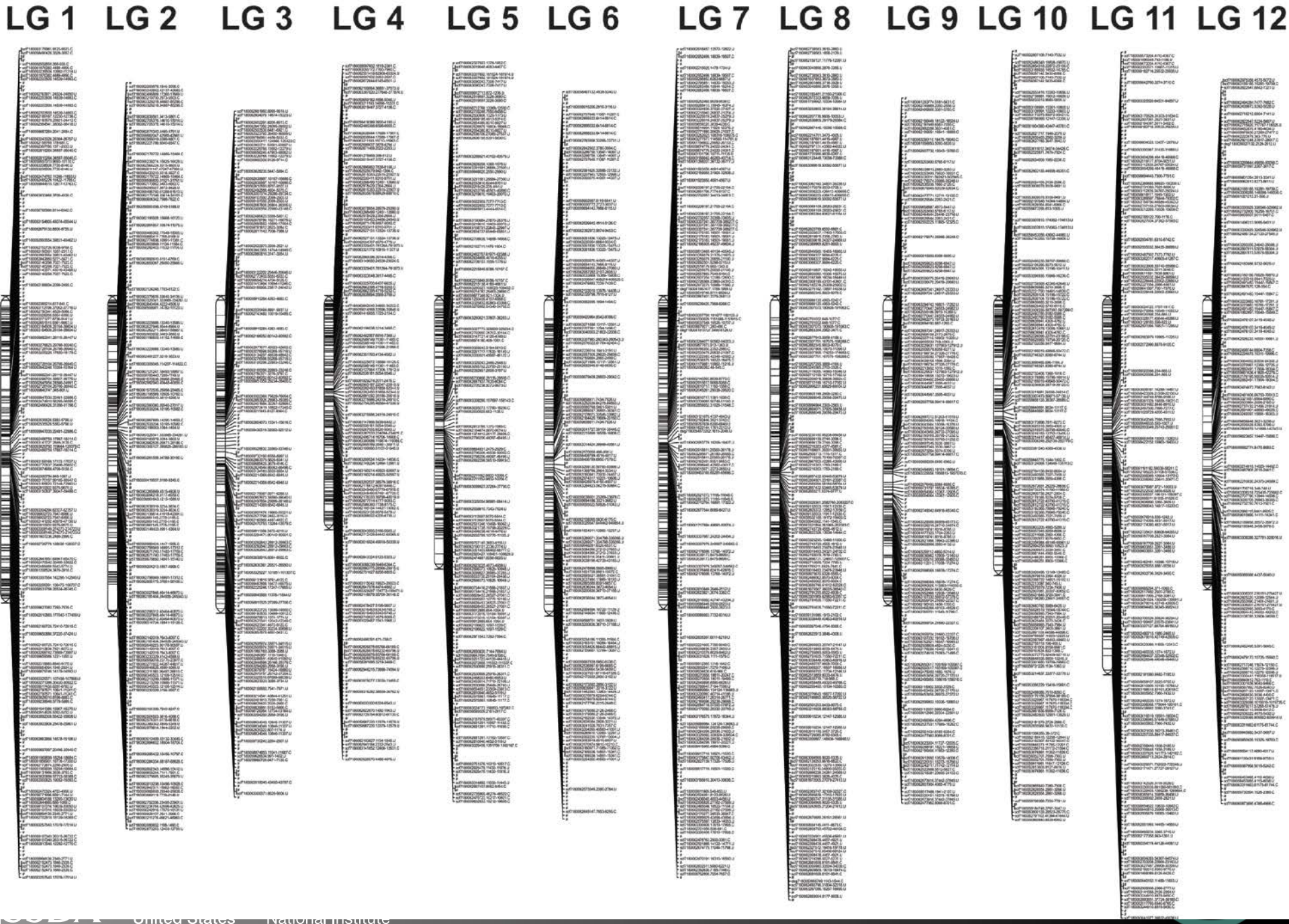


LG 3

scf7180062681291

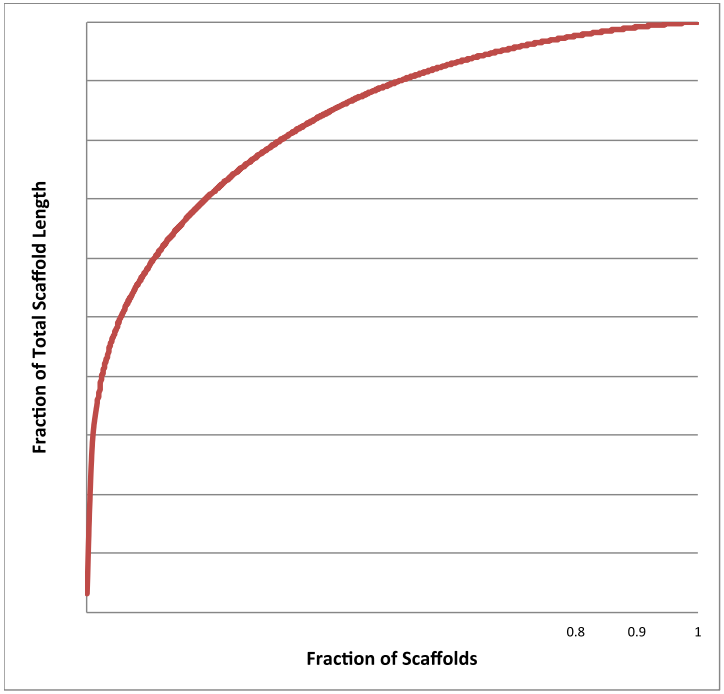


LG 5


LG 1
LG 2
LG 3
LG 4
LG 5
LG 6
LG 7
LG 8
LG 9
LG 10
LG 11
LG 12

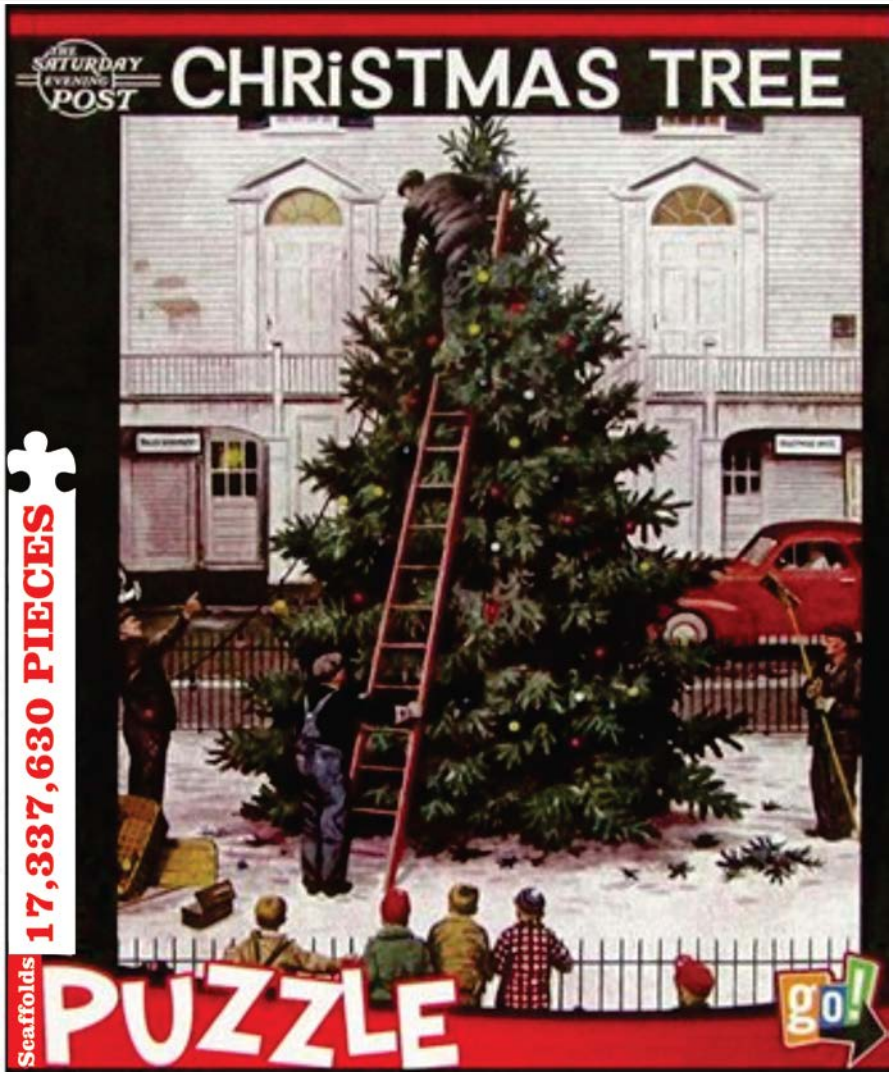
WGS Scaffolds

- 96,628,871 bp are ordered
- Most of the ordered scaffolds are short
  - 50% of the total length is in 5.6% of the scaffolds.



- Total length is less than 1% of the 22Gb assembly.

Conclusion



…the puzzle is incomplete.

BUT….

…a much higher density map is under development…

Acknowledgments

Neale lab

Trust me…a Jedi I am…

Thank you for your attention