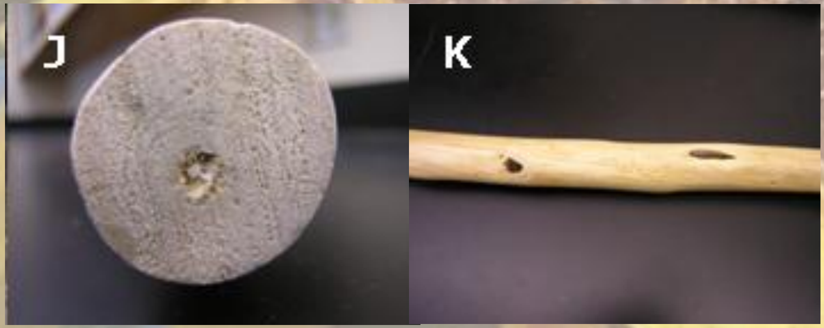
Genomics of Cellulosic Biomass Traits in Sunflower

Loren Rieseberg,
University of British Columbia

Sunflower



•Globally important oilseed crop ~ $8 billion

Ranks 11th among world food crops in terms of area harvested

•Only major crop to be domesticated in North America

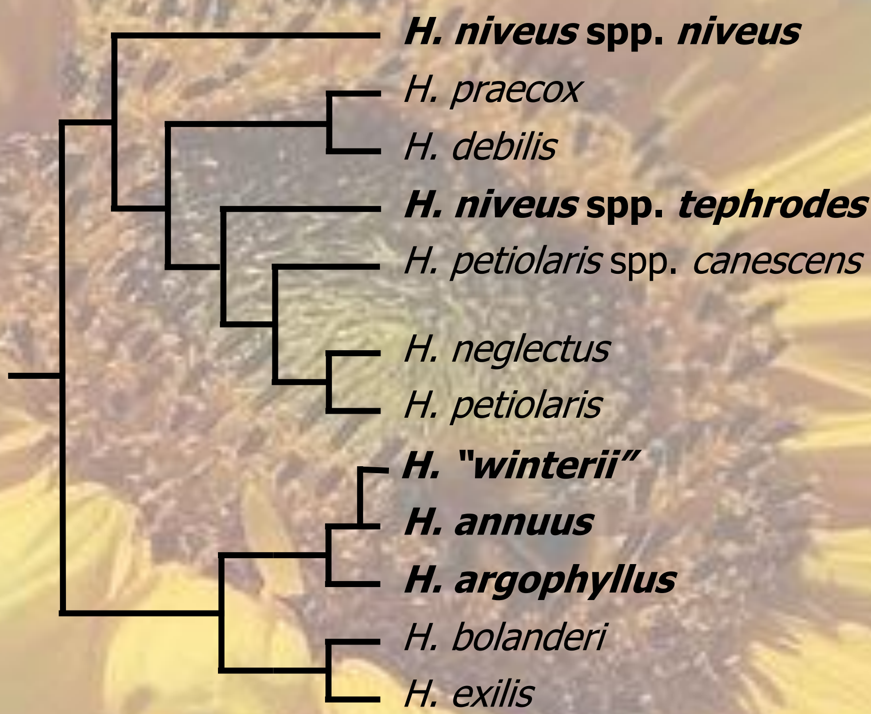
•Potential for cellulosic biomass production

•Drought tolerant wild sunflowers have woody stems

Sunflower as a Biofuel

- Biodiesel – sunflower seed has high oil content, but oil more valuable as food than fuel
- Biogas – sunflower currently employed as a biogas (methane) crop in Europe
- Cellulosic biomass
  - Tremendous potential as dual use (food and fuel) crop, especially in Africa
    - Slow deforestation, eliminate need to search for wood, improve educational opportunities for girls and women
  - Cellulosic ethanol?

Helianthus Woody Taxa



Helianthus winterii

Questions:

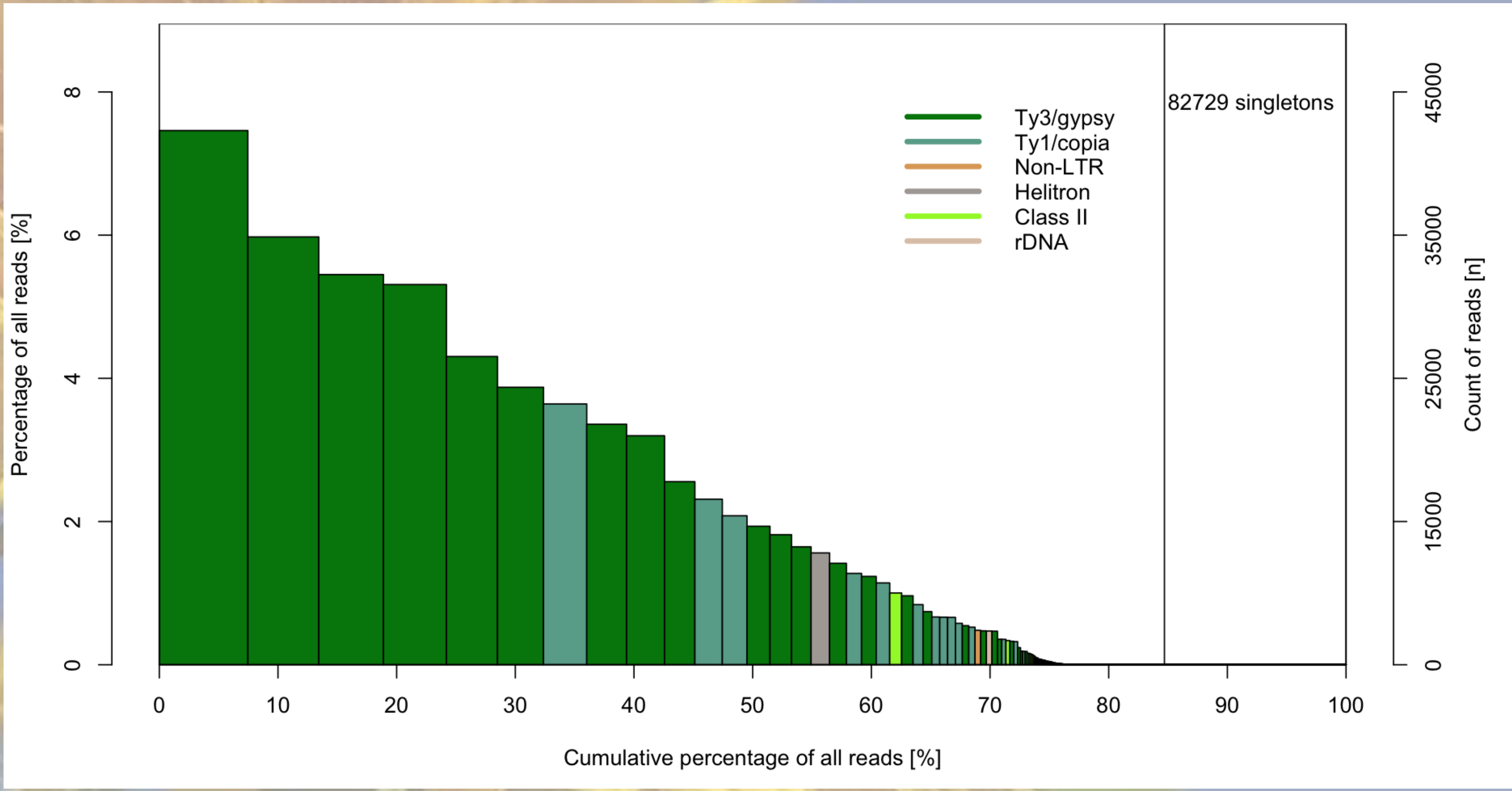
- When will we have a reference sequence for sunflower?

- Can we develop sunflower cultivars with favorable cellulosic biomass traits?

Complication #1: sunflower genome is very large

Organism	Genome Size	Completed
Arabidopsis	119 Mb	2000
Oryza (rice)	420 Mb	2002
Vitis (grapevine)	490 Mb	2007
Populus	550 Mb	2006
Solanum (potato)	844 Mb	2011
Glycine (soybean)	1,100 Mb	2010
Zea mays (maize)	2,800 Mb	2009
Homo sapiens (human)	3,200 Mb	2006
Monodelphis (opossum)	3,500 Mb	2007
Helianthus (sunflower)	**3,600 Mb**	**2013?**

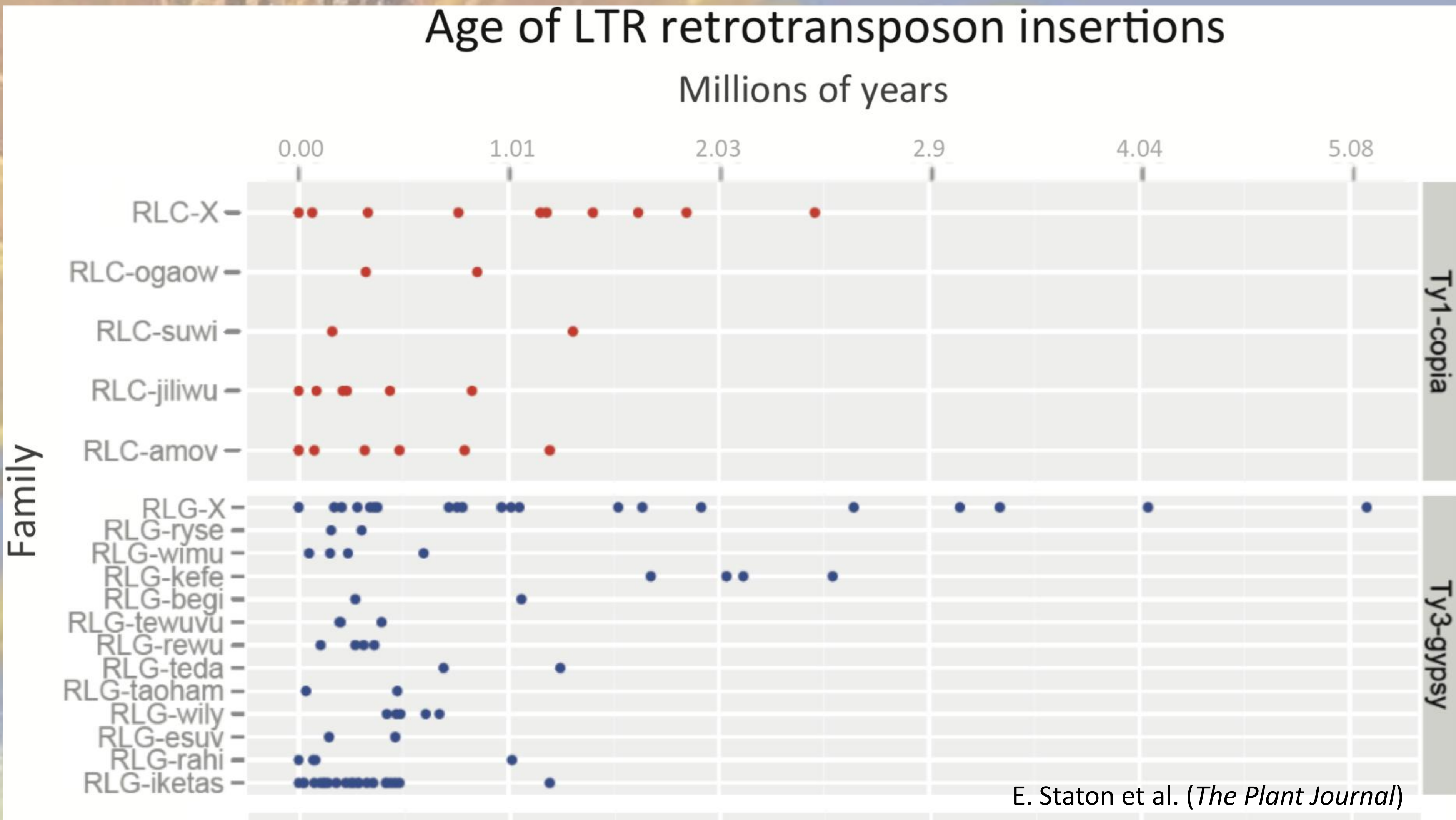
Complication #2: sunflower genome is highly repetitive



81% repetitive; 58% Ty3/gypsy

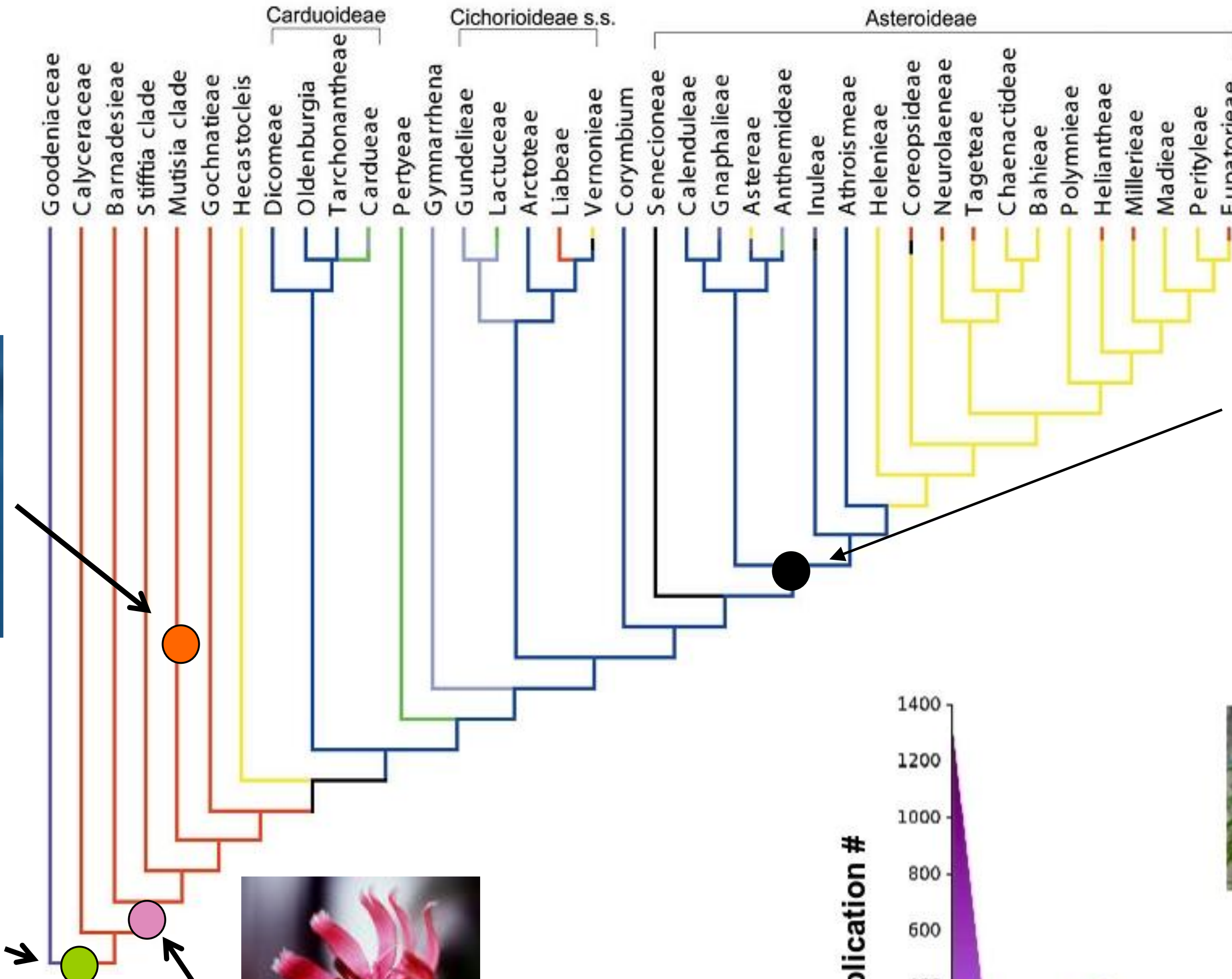
E. Staton et al. (*The Plant Journal*)

Complication #3: many of the LTR repeats are very young



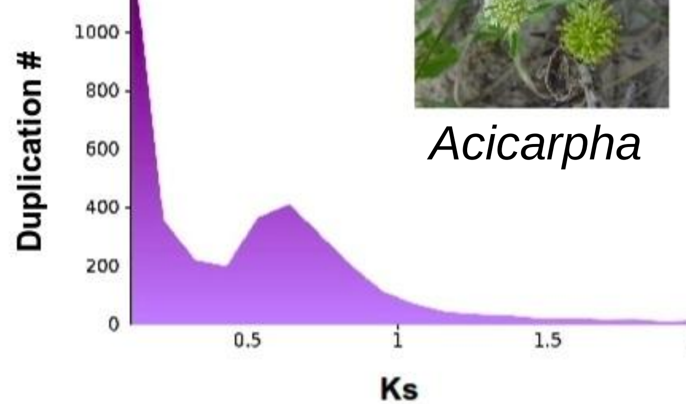
E. Staton et al. (*The Plant Journal*)

Complication #4: sunflower is an ancient octaploid



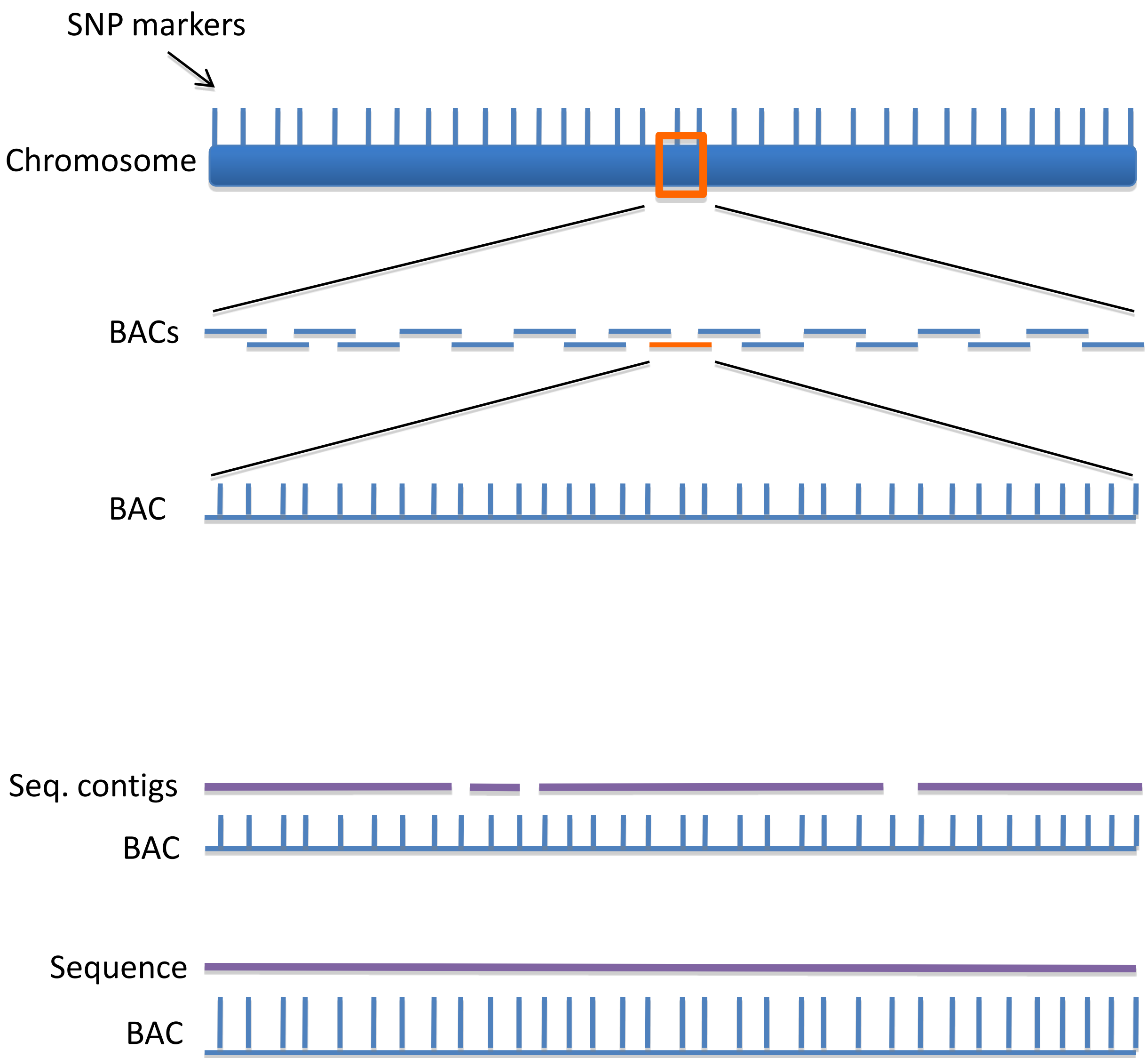
Mike Barker

Sunflower



Acicarpha

Strategy



High-density genetic map

↓

Sequence-based physical map

↓

Sequence-tags every 5-6 kb

↓

Assembly of WGS sequence

↓

Sequence contigs aligned against genetic & physical map

↓

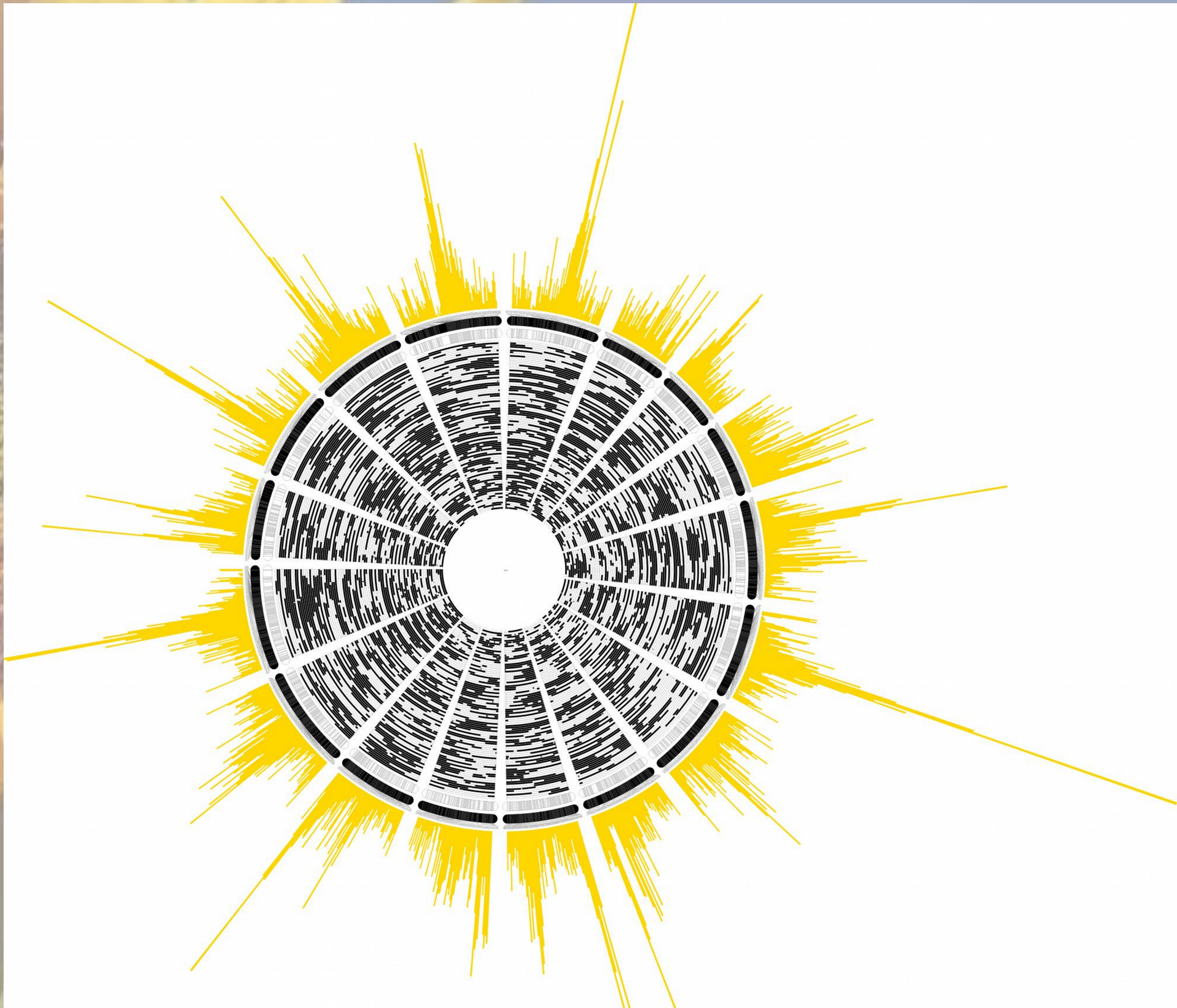
Local scaffolding with mate pairs

↓

Annotation

Genetic Map - completed

- •96 RILs sequenced to 1x depth
- •2.6 million SNPs
- •C. Grass, J. Bowers et al.



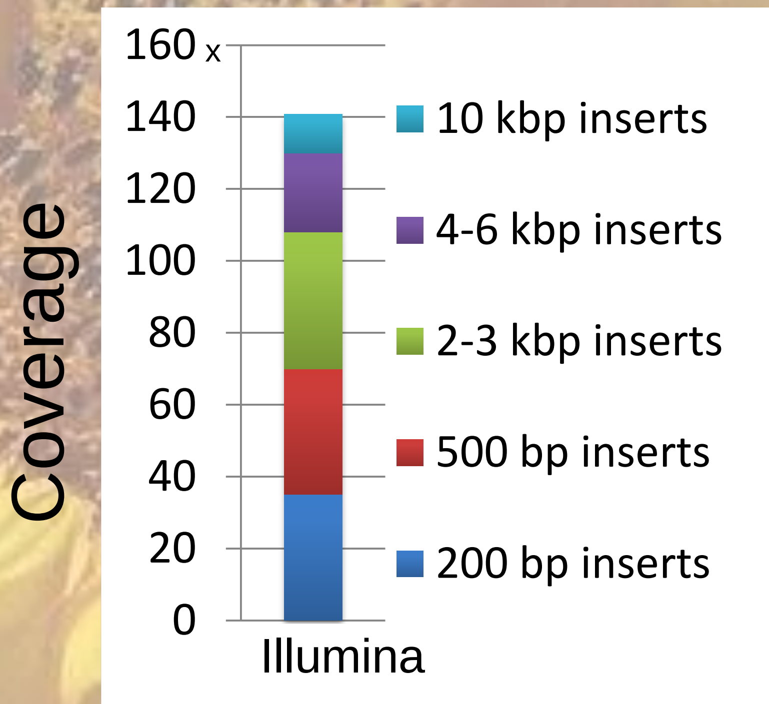
Physical Map – Completed

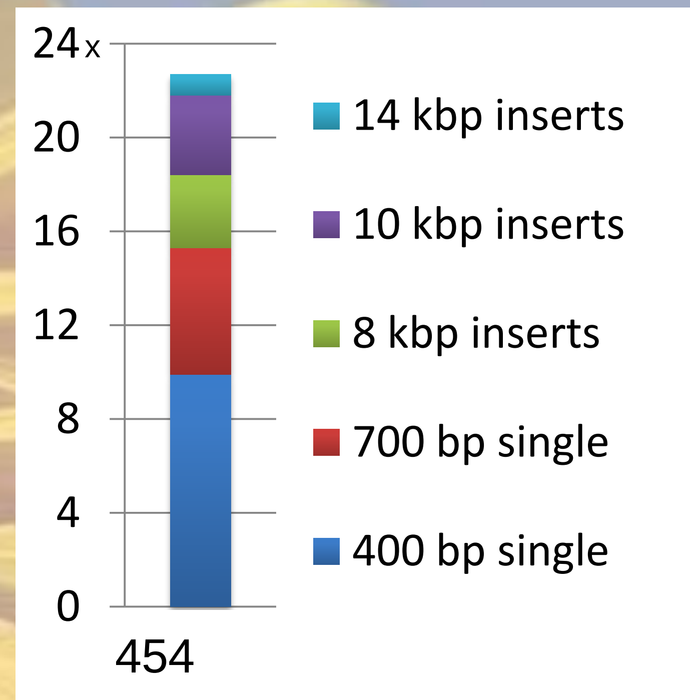
(iterative FPC assembly; N. Gill et al. unpublished)

Parameter	Value
Genome size	3,600 Mbp
Genome equivalents of BACs tested	12.5x
Total No. unique tags	483,758
Average tags per BAC	20
Total # of BACs in FPC	335,201
Number and % BACs in contigs	287,344 (86%)
Coverage of the genome	**3,500 Mbp (97%)**
Total number of contigs	**3,332**
N50 contig size (BACs)	57
N50 contig size (Mbp)	1,783 Kb

Sequencing and Assembly

Nolan Kane





- Sequencing is finished and latest assembly covers 94% of genome

Integration and Annotation

- 63% of genome placed on chromosomes and annotated

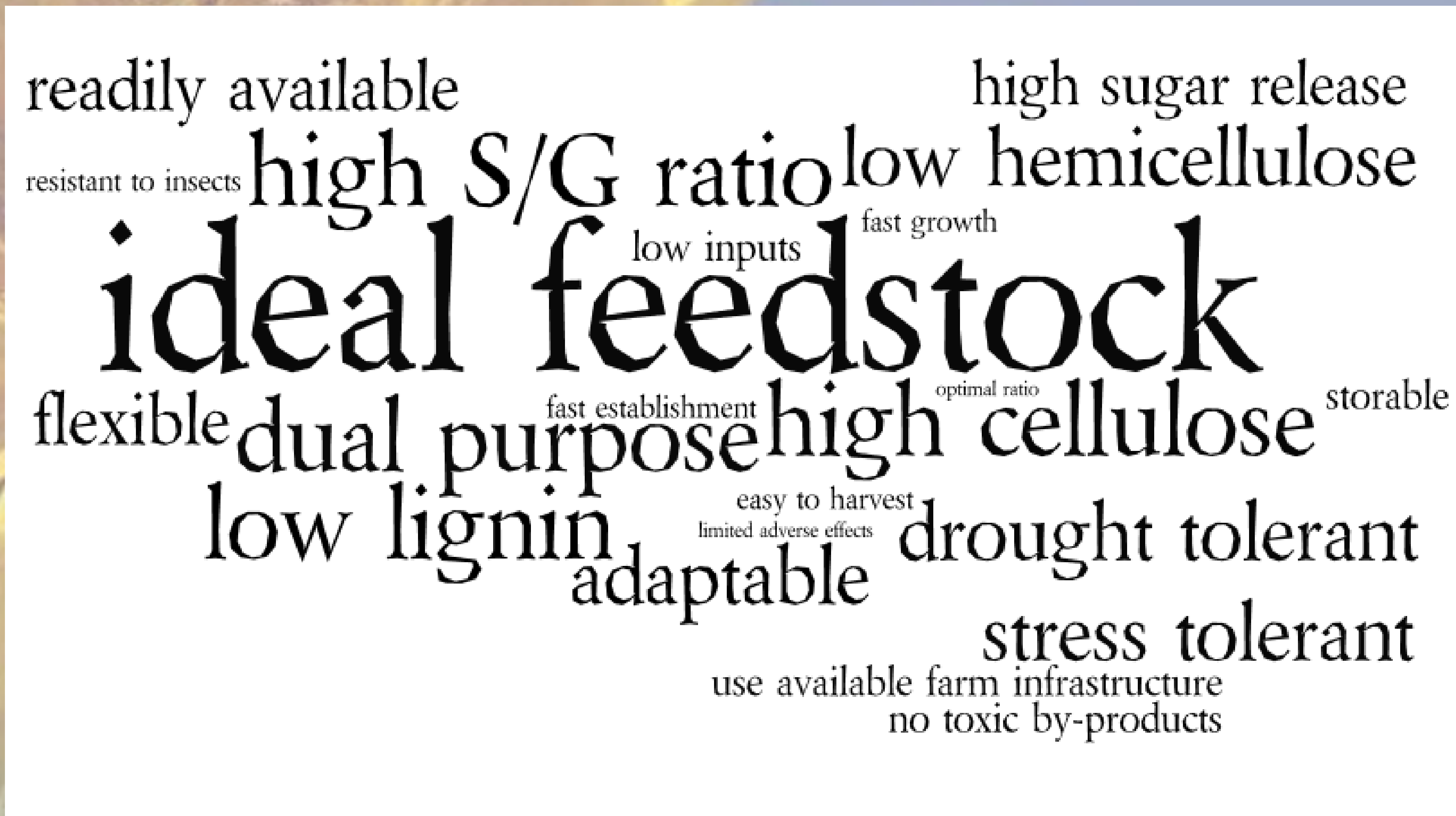
Questions:

- When will we have a reference sequence for sunflower?

- <u>Can we develop sunflower cultivars with favorable cellulosic biomass traits?</u>

What makes an ideal biofuels feedstock?

Jessica Barb



Population Development

H. annuus x (*H. annuus* x *H. argophyllus)* BC_1

NMS/RHA377 (PI560145)

- Elite apically branched oilseed
  inbred B-line

ARG1820 (PI49480)

- Tall, late flowering, drought resistant

Phenotypic Data - Pearson's Correlations

	Lignin				S-lignin				G-lignin				S/G-lignin			
	HOPI	GA	BC	IA	HOPI	GA	BC	IA	HOPI	GA	BC	IA	HOPI	GA	BC	IA
S-lignin	0.86	0.88	0.82	0.77												
G-lignin	0.47	0.62	0.59	0.69		0.19										
S/G-lignin	0.32	0.44	0.28		0.75	0.80	0.78	0.67	-0.68	-0.43	-0.60	-0.68				
Hemicelluloses	-0.69	-0.68	-0.44	-0.86	-0.37	-0.43	-0.17	-0.48	-0.65	-0.65	-0.46	-0.79			0.16	0.24
Celluloses	-0.76	-0.69	-0.42	-0.86	-0.46	-0.54	-0.26	-0.55	-0.64	-0.50	-0.32	-0.72		-0.20		
Cellulose/Hemicellulose		0.51	0.24	0.48	-0.16	0.25		0.21	0.32	0.59	0.35	0.50	-0.32		-0.23	-0.21
C6overLig	-0.88	-0.93	-0.79	-0.93	-0.61	-0.78	-0.58	-0.63	-0.63	-0.62	-0.53	-0.74		-0.35		
Height	0.32	0.47	0.32	0.45	0.36	0.53	0.25	0.52			0.23		0.25	0.43		0.30
Biomass		0.34	0.20			0.39			-0.20		0.19		0.19	0.31		
% Main Stem	N/A		0.16		N/A		0.21	0.19	N/A	-0.16			N/A		0.17	0.20
Flower			0.20				0.21							0.16		
Stem Diameter		0.20	-0.17	0.32		0.22		0.28	-0.23				0.20	0.18		
Stem Density	0.27				0.26		0.20								0.24	

Lignin and carbohydrate content are **negatively** correlated

Phenotypic Data - Pearson's Correlations

	Lignin				S-lignin				G-lignin				S/G-lignin			
	HOPI	GA	BC	IA	HOPI	GA	BC	IA	HOPI	GA	BC	IA	HOPI	GA	BC	IA
S-lignin	0.86	0.88	0.82	0.77												
G-lignin	0.47	0.62	0.59	0.69		0.19										
S/G-lignin	0.32	0.44	0.28		0.75	0.80	0.78	0.67	-0.68	-0.43	-0.60	-0.68				
Hemicelluloses	-0.69	-0.68	-0.44	-0.86	-0.37	-0.43	-0.17	-0.48	-0.65	-0.65	-0.46	-0.79			0.16	0.24
Celluloses	-0.76	-0.69	-0.42	-0.86	-0.46	-0.54	-0.26	-0.55	-0.64	-0.50	-0.32	-0.72		-0.20		
Cellulose/Hemicellulose		0.51	0.24	0.48	-0.16	0.25		0.21	0.32	0.59	0.35	0.50	-0.32		-0.23	-0.21
C6overLig	-0.88	-0.93	-0.79	-0.93	-0.61	-0.78	-0.58	-0.63	-0.63	-0.62	-0.53	-0.74		-0.35		
Height	0.32	0.47	0.32	0.45	0.36	0.53	0.25	0.52			0.23		0.25	0.43		0.30
Biomass		0.34	0.20			0.39			-0.20		0.19		0.19	0.31		
% Main Stem	N/A		0.16		N/A		0.21	0.19	N/A	-0.16			N/A		0.17	0.20
Flower			0.20				0.21							0.16		
Stem Diameter		0.20	-0.17	0.32		0.22		0.28	-0.23				0.20	0.18		
Stem Density	0.27				0.26		0.20								0.24	

S-lignin and G-lignin contents are **NOT** correlated

It should be possible to develop sunflower cultivars with favourable cellulosic biomass traits

Conclusions:

•Assembly of sunflower genome challenging because of its large size, past polyploidization events, and recent amplification of repetitive elements.

•Nonetheless, 63% of genome placed on chromosomes and annotated.

•Transgressive segregation of key cellulosic biomass traits and favorable genetic correlations should permit breeding of a sunflower biofuel feedstock.

•Dual use sunflower has promise for subsistence agriculture.

Acknowledgements

Genomics of Sunflower Project (PIs): Loren Rieseberg**,** Steve Knapp, John Burke, Patrick Vincourt, Nolan Kane

Genetic mapping: John Bowers, Chris Grassa, Nicolas Langlade, Eleni Bachlava

Physical mapping: Helene Berges, Navdeep Gill, Thuy Nguyen, KeyGene N.V.

Sequencing: Genome Quebec, Michael Smith Genome Sciences Centre, Zhao Lai, Jie Huang, Megan Stewart

Bioinformatics: Jerome Gouzy, Chris Grassa, Thuy Nguyen, Matteo Buti, Evan Staton, Navdeep Gill

Cellulosic biomass traits: Jessica Barb, Mark Davis, Sukhpreet Sandhu, Brook Moyer, Laura Marek