Summary N. benthamiana reference genome v2.6.1

Aureliano Bombarely
Instituto de Biologia Molecular y Celular de Plantas (IBMCP)
(CSIC-UPV)
abombarely@ibmcp.upv.es

Genome assembly in the *Nicotiana benthamiana* v2.6.1

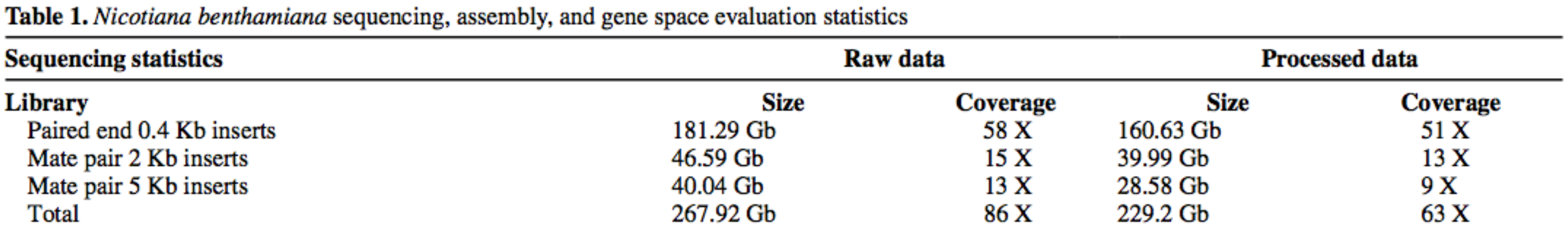


Table 1. *Nicotiana benthamiana* sequencing, assembly, and gene space evaluation statistics

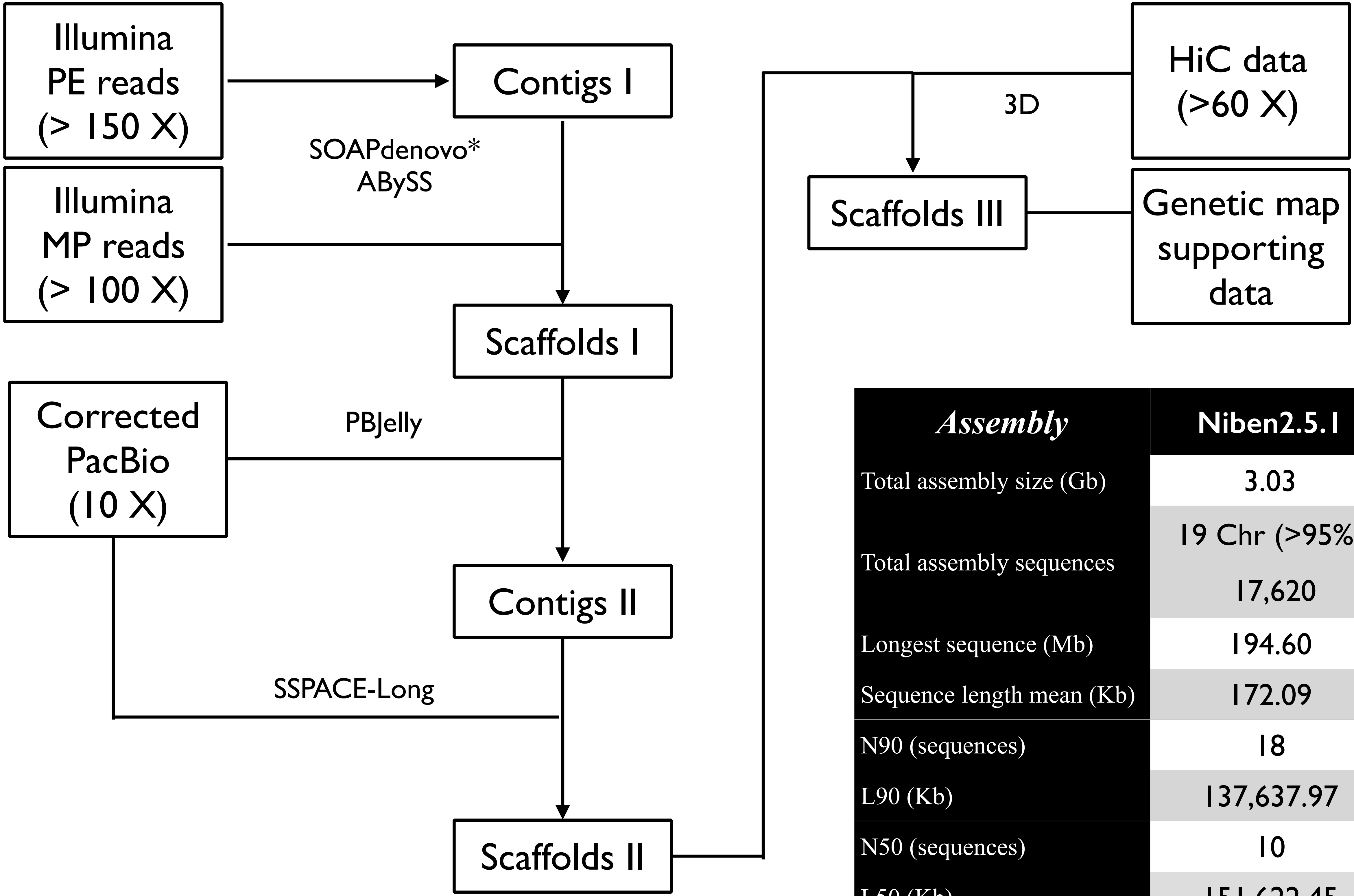
Sequencing statistics	**Raw data**		**Processed data**	
Library	**Size**	**Coverage**	**Size**	**Coverage**
Paired end 0.4 Kb inserts	181.29 Gb	58 X	160.63 Gb	51 X
Mate pair 2 Kb inserts	46.59 Gb	15 X	39.99 Gb	13 X
Mate pair 5 Kb inserts	40.04 Gb	13 X	28.58 Gb	9 X
Total	267.92 Gb	86 X	229.2 Gb	63 X

Bombarely A. et al. 2012. MPMI 25, 1523-1530

+

Scaffolding Libraries	Size (Gb)	Coverage (X)
Illumina Mate Pair 5 Kb (BTI)	44.38	14.32
Illumina Mate Pair 15 Kb (BTI)	61.34	19.79
Illumina Mate Pair 20 Kb (BTI)	0.76	0.24
PacBio RS II (Brian Kivtko)*	20.99	6.77
HiC (Silin Zhong)	134.06	43.25

Genome assembly in the *Nicotiana benthamiana* v2.6.1



Illumina PE reads (> 150 X)

Contigs I

SOAPdenovo*
ABySS

Illumina MP reads (> 100 X)

Scaffolds I

Corrected PacBio (10 X)

PBJelly

Contigs II

SSPACE-Long

Scaffolds II

HiC data (>60 X)

3D

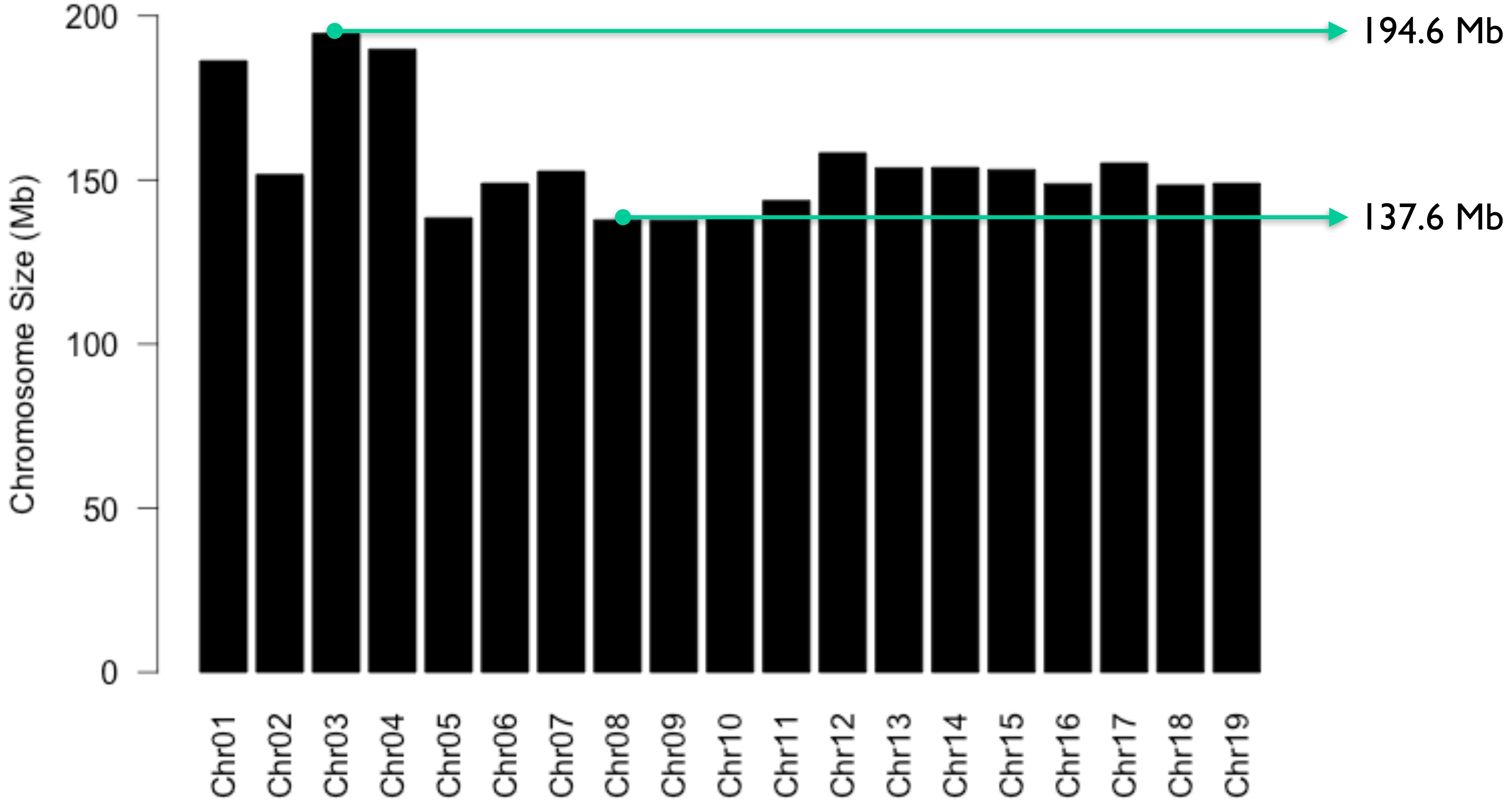
Scaffolds III

Genetic map supporting data

Assembly	**Niben2.5.1**
Total assembly size (Gb)	3.03
Total assembly sequences	19 Chr (>95%) 17,620
Longest sequence (Mb)	194.60
Sequence length mean (Kb)	172.09
N90 (sequences)	18
L90 (Kb)	137,637.97
N50 (sequences)	10
L50 (Kb)	151,622.45

Genome assembly in the *Nicotiana benthamiana* v2.6.1

Anchored assembly: 2.94 Gb (97%)



Unanchored assembly: 17,620 sequenced counting for 95.74 Mb (3%)
(N90=13,337 sequences / L90=1,738 bp)

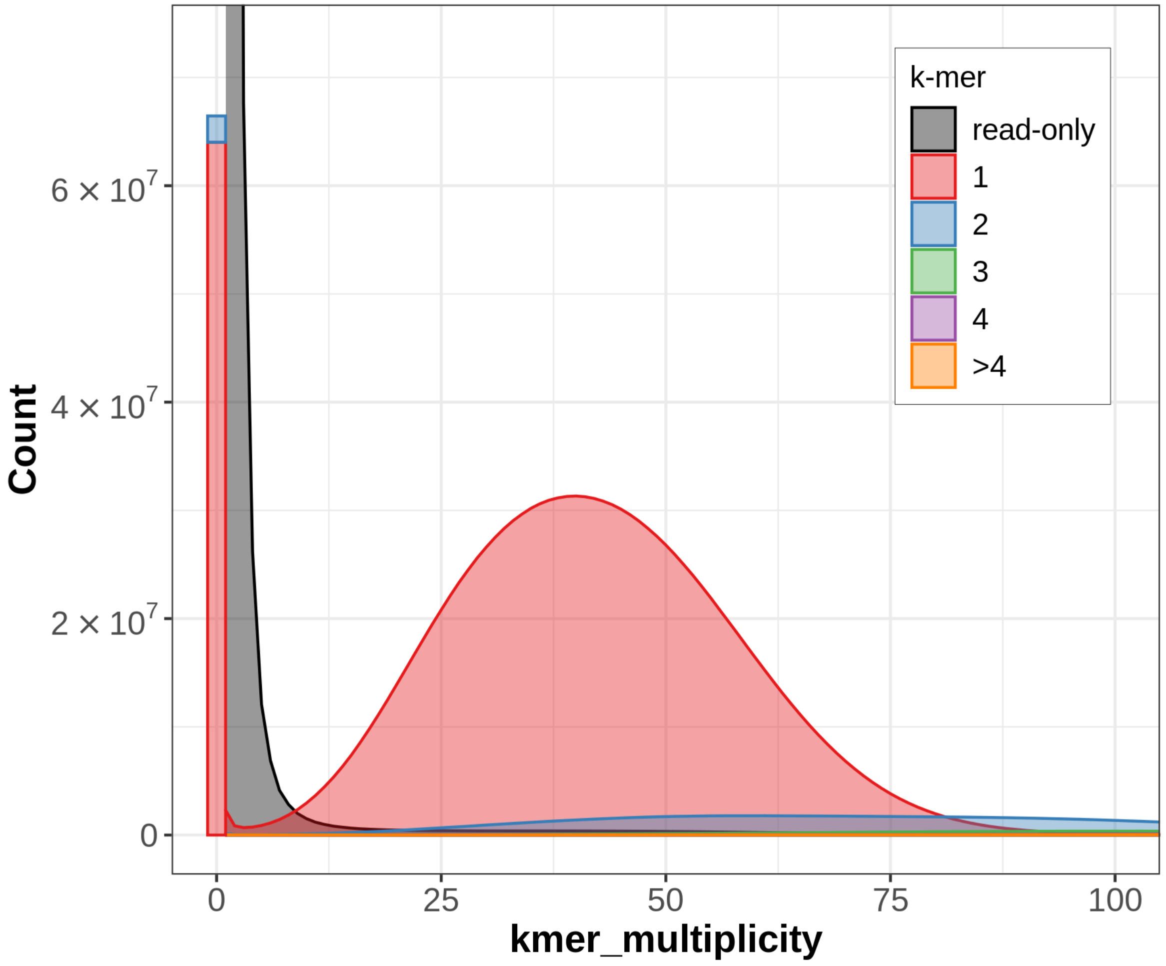
Genome assembly in the *Nicotiana benthamiana* v2.6.1

Quality evaluation

Estimated genome size:

```
•  Flow Cytometry (Narayan,1987)    = 3.13 Gb
•  Kmer Count*                      = 3.07 Gb
•  Assembly size (scaffolds v2.6.1) = 3.04 Gb
•  Assembly size (contigs   v2.6.1) = 2.75 Gb
```

Merqury
Completeness = 98.12%
QV = 29.4
Consensus error rate = 0.001



Gene annotation in the *Nicotiana benthamiana* v2.6.1

Paleoploid species with 2n = 4x = 38

Assembled into 19 chromosomes and 17,620 scaffolds.

<u>Annotation</u>:
1- RepeatModeler2
2- Maker
  * Augustus + SNAP
  * RNA-Seq
  * Solanaceae proteins

Gene annotation in the *Nicotiana benthamiana* v2.6.1

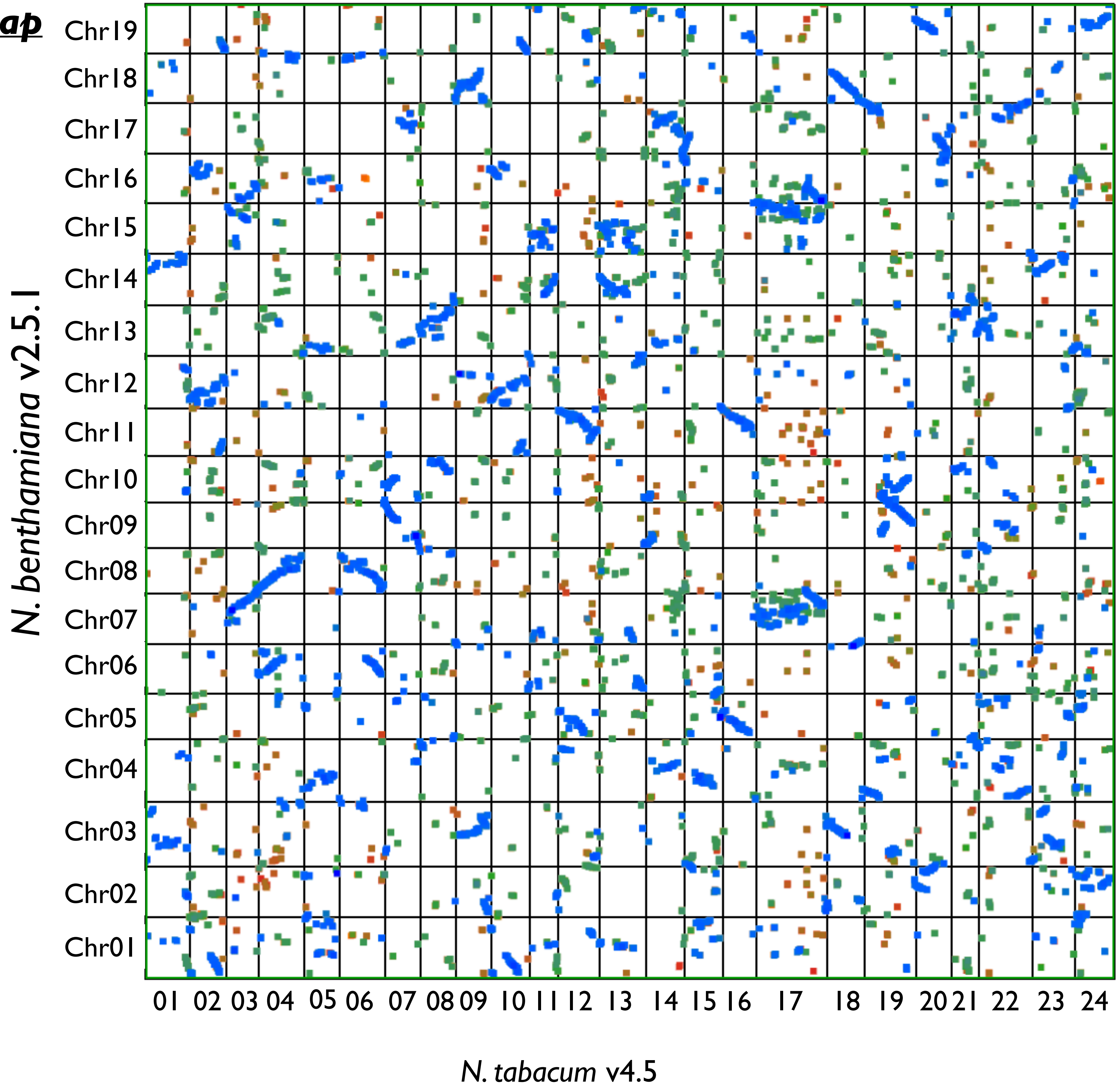
	***N. attenuata* NIATTr2**	***N. tabacum* Nitab4.5**	***N. benthamiana* v2.6.1**	***S. lycopersicum* ITAG 2.40**
Genes Models	33,320	69,500	61,328	34,725
mRNA	33,320	69,500	61,328	34,725
tRNA	785	0	0	0
exons/gene model	4.58	4.79	4.98	4.61
five_prime_UTRs	26,223	15,661	1,901	13,548
three_prime_UTRs	26,354	11,381	244	15,343
average mRNA length (bp)	1,331		1,005	1,209
average protein length (Aa)	370	351	333	344
Gene space size (Mb)	122		281	110
BUSCO	C:96.4%[S:93.7%,D:2.7%], F:1.6%,M:2.0%	C: 95.1%[S:23.8%,D:71.3%], F:2.9%,M:2.0%	C:95.9%[S:47.6%,D:48.3%], F:2.1%,M:2.0%	C:95.7%[S:95.2%,D:0.5%], F:2.7%,M:1.6%

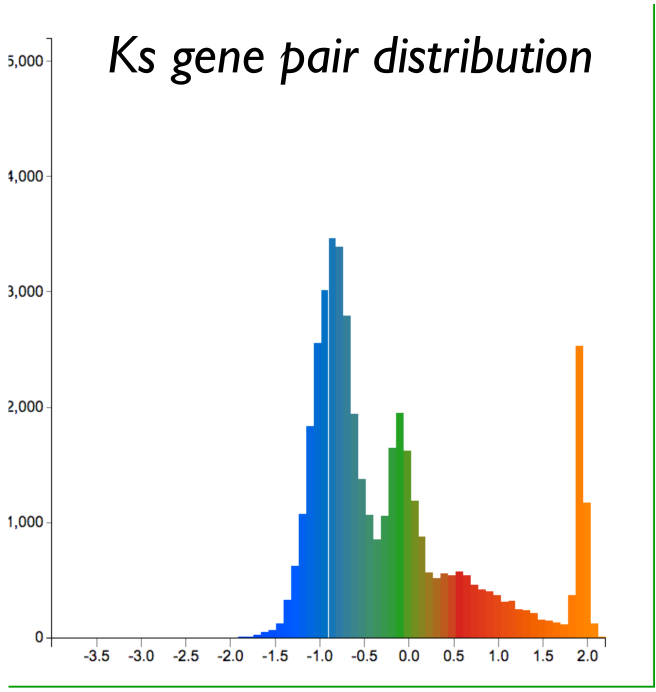
Repeat annotation in the *Nicotiana benthamiana* v2.6.1

		N.syl	***N.tom***	***N.tab***	***N.ben***
Type	***Repeat***	***% Genome***	***% Genome***	***% Genome***	***% Genome***
Low Complexity	Simple Repeat	1.34	1.40	0.25	1.27
	Nucleotide Rich/Satellites	0.39	1.39	1.85	0.24
DNA transposons	CMC	0.21	0.25	0.22	0.22
	Harbinger	0.08	0.07	0.07	0.07
	hAT	0.38	0.44	0.39	0.66
	Maverick	<0.01	<0.01	<0.01	0.01
	MULE	0.11	0.11	0.10	0.08
	TcMar	0.11	0.14	0.12	0.10
	Other	0.11	0.11	0.10	0.11
Non LTR transposons	LINE	0.87	0.98	0.91	1.57
	SINE	0.18	0.21	0.19	0.36
LTR transposons	LTR/Copia	2.98	3.12	3.06	4.60
	LTR/Gypsy	16.26	14.60	15.87	10.73
	Other LTR	0.20	0.26	0.23	0.16
RNA	rRNA	<0.01	<0.01	<0.01	0.03
	tRNA	<0.01	<0.01	<0.01	< 0.01
	snRNA	<0.01	<0.01	<0.01	< 0.01
Unknown/Unspecified		47.63	51.85	51.21	41.77
Total		**70.96**	**75.05**	**74.67**	**62.15**

- The *Nicotiana benthamiana* genome v2.6.1.

SyMap





Acknowledgements

Development of a *N. benthamiana* reference genome

VirginiaTech
Invent the Future®

UK
UNIVERSITY OF
KENTUCKY

David Zaitlin

Gao Muqiang

BTI
Boyce Thompson Institute
for Plant Research

Gregory Martin

Lukas A. Mueller

Noe Fernandez-Pozo

Hernan G. Rosli*

Marina Pombo*

UNIVERSITY OF
GEORGIA

Brian Kvitko (PacBio)

香港中文大學
The Chinese University of Hong Kong

Silin Zhong (HiC)