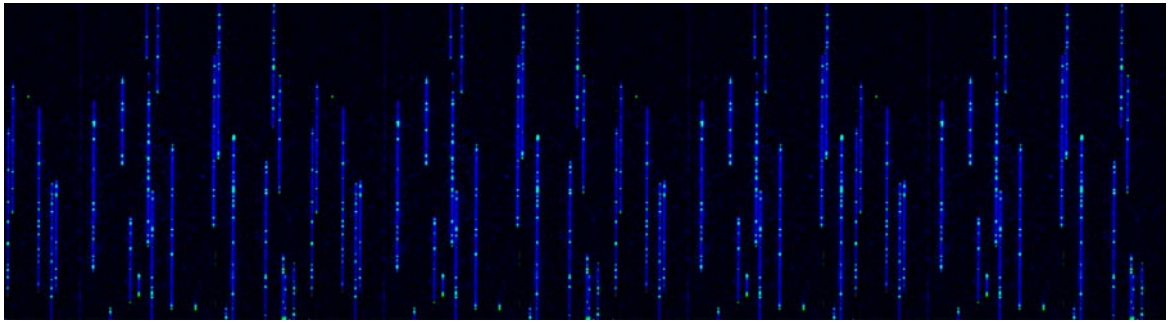
BioNano map to facilitate the assembly of the *Aegilops tauschii* genome



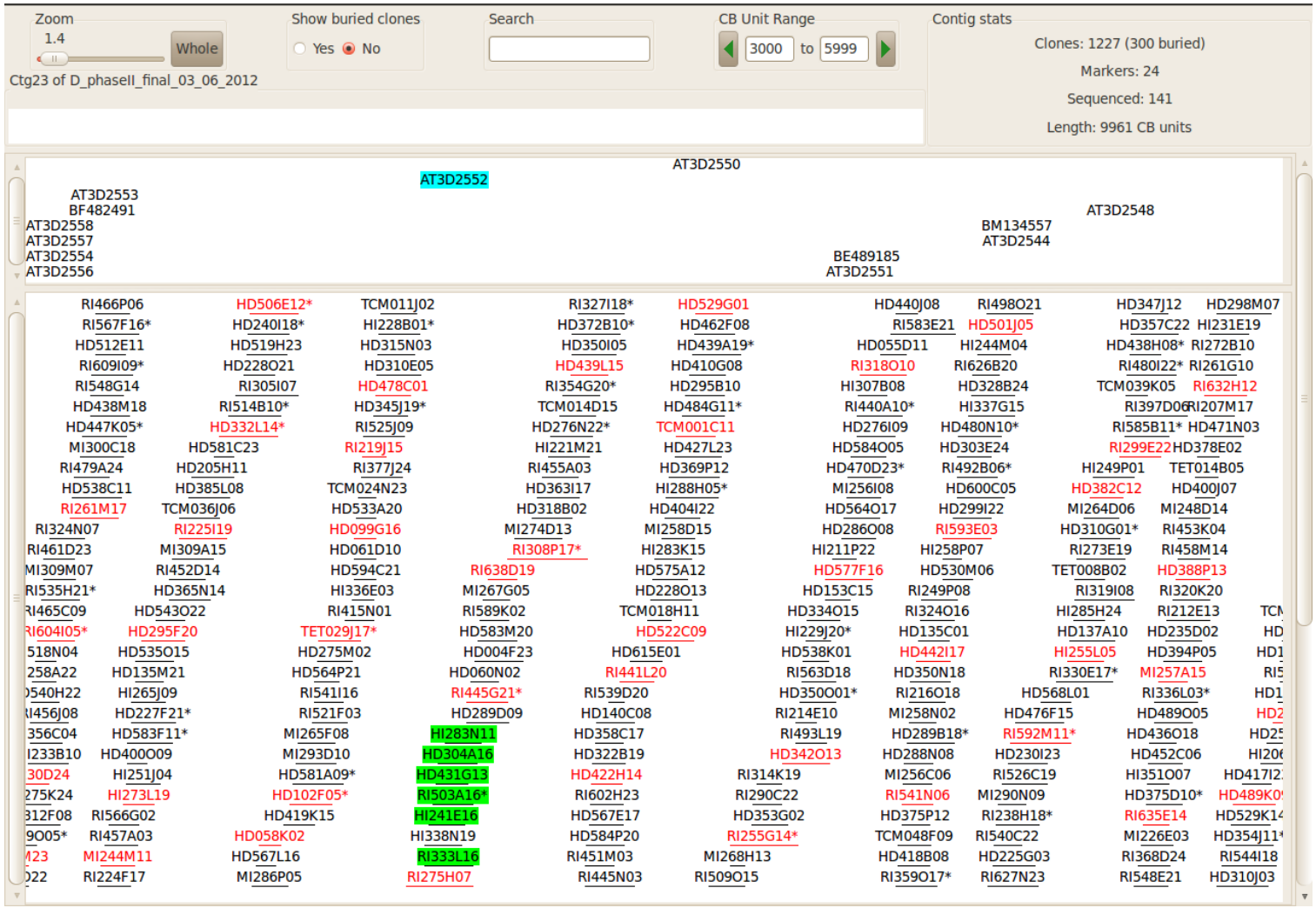
Ming-Cheng Luo ***(mcluo@ucdavis.edu)***

Tingting Zhu ***(tngzhu@ucdavis.edu)***

UCDAVIS

DEPARTMENT OF PLANT SCIENCES

Portion of a contig on 3D: 15 Mb



Genetic map of the *Aegilops tauschii* genome

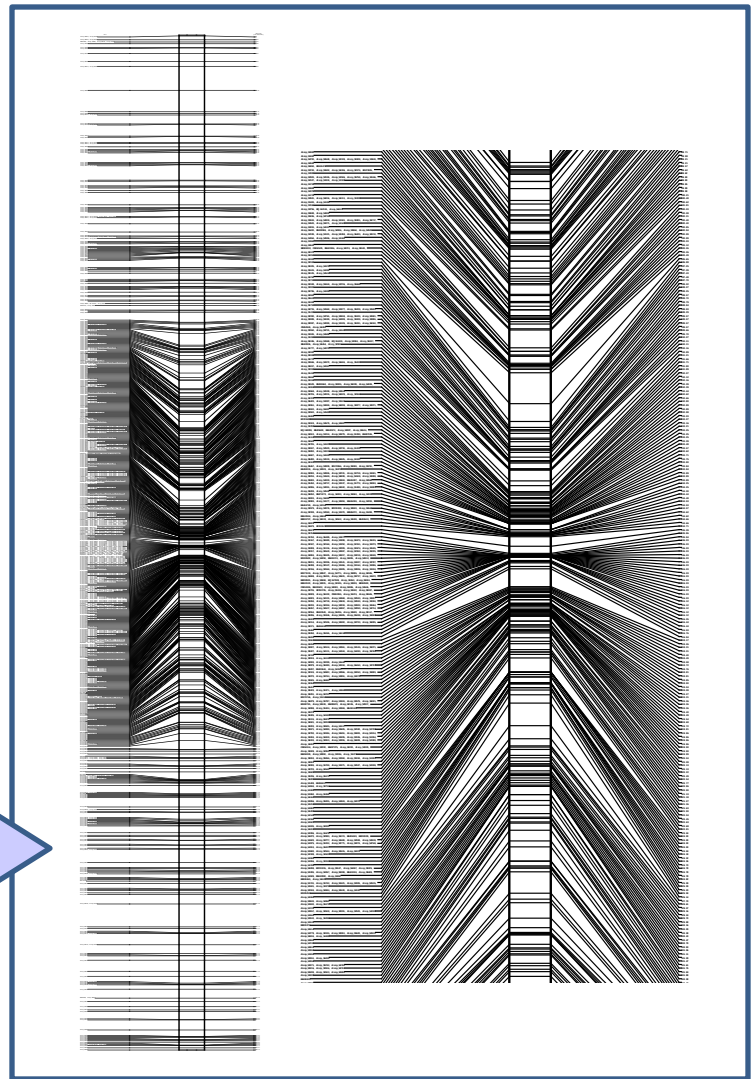
1,102 F_2 plants

7,185 SNP gene loci mapped

Chromosome 3D:

1,101 gene loci

204 cM

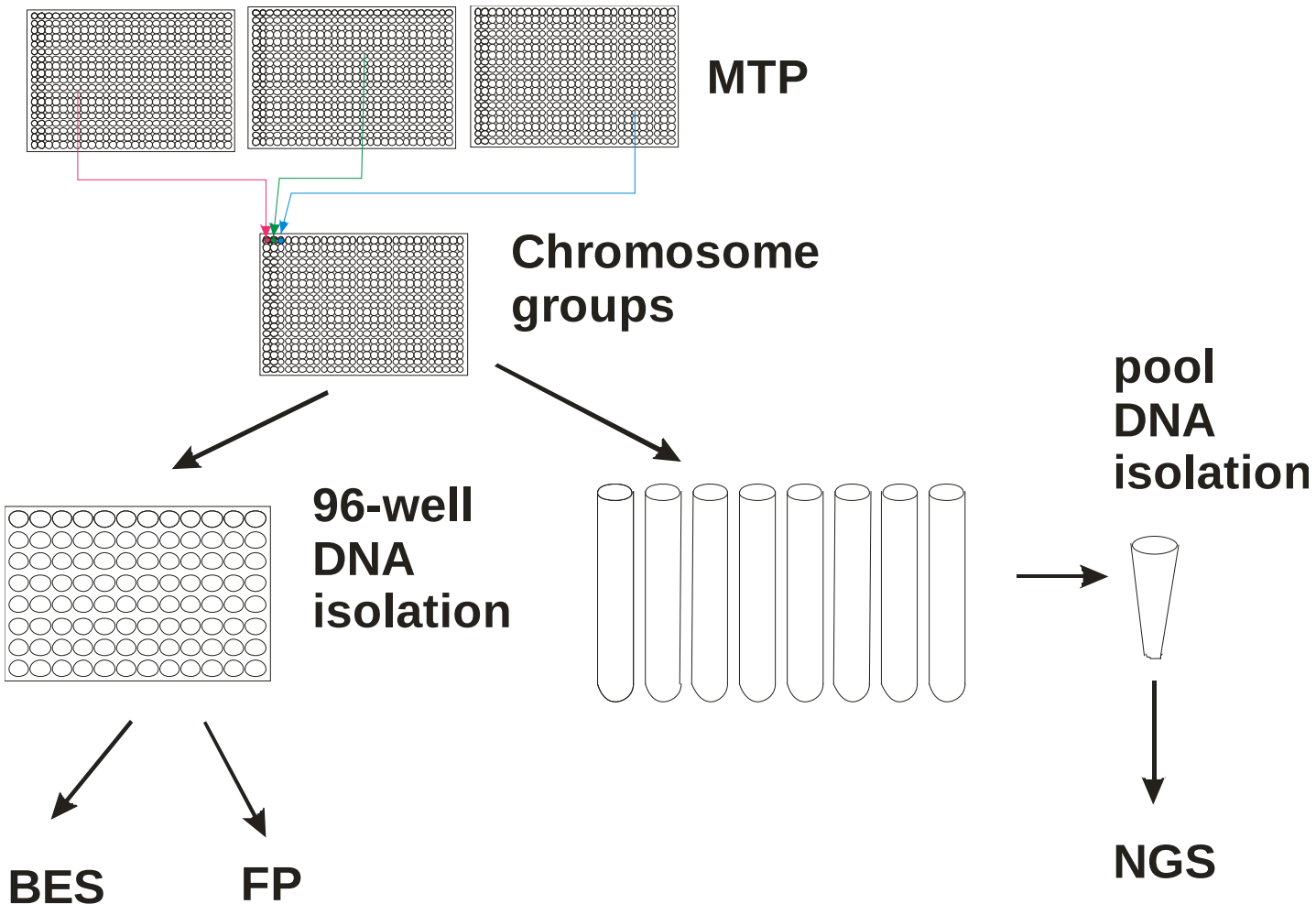


BAC-based physical maps

3,578 contigs

85% anchored on linkage maps

BAC DNA prep


MTP
Chromosome groups
96-well DNA isolation
pool DNA isolation
BES
FP
NGS

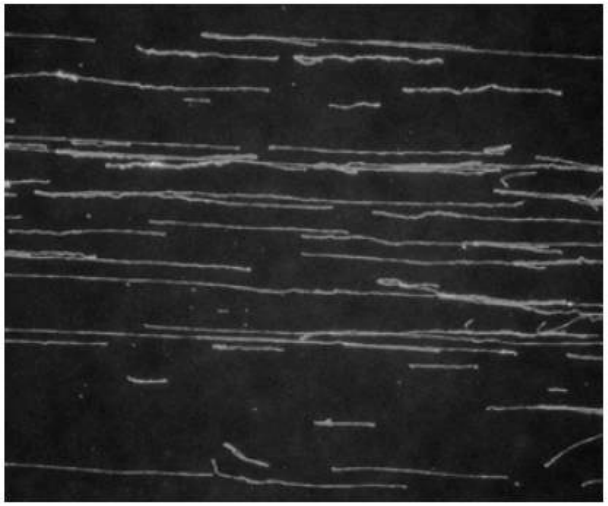
Challenges of next-generation sequencing

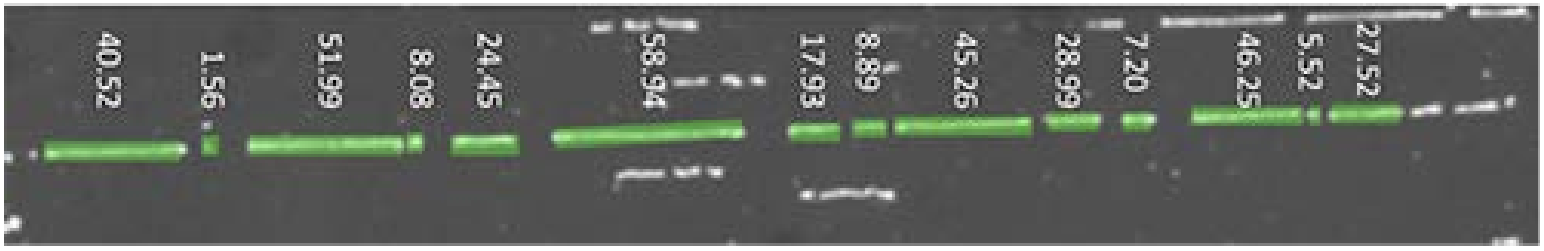
- errors in contigs and scaffolds
- ordering and orienting scaffolds (limitation of BAC contigs, contamination issue)
- estimating gap size between scaffolds
- creating superscaffolds and building pseudomolecule

A sequence-independent technology that can aid "*de novo*" sequence assembly is needed.

Optical mapping: OpGen technology

"**Optical Mapping** is a *de novo* process that generates whole genome, ordered restriction maps with no requirement for previous sequence information."

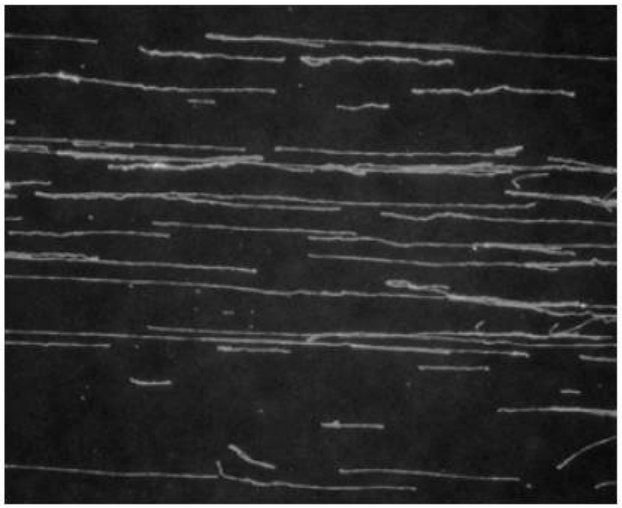




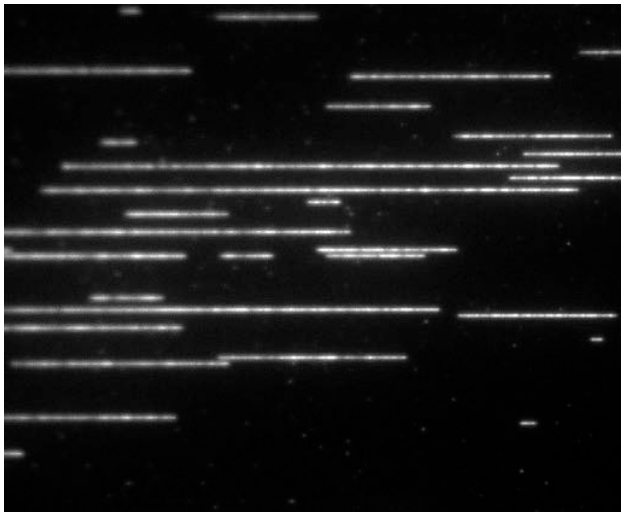
Advanced optical mapping: BioNano technology

- High throughput
- Uniform DNA stretching

  Precise DNA length measurement
- Low error rates in assembling

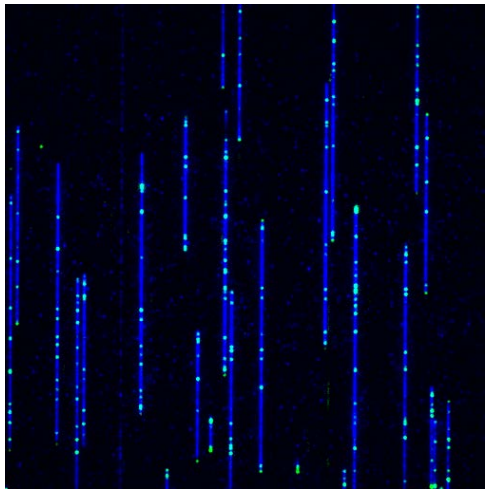


OpGen

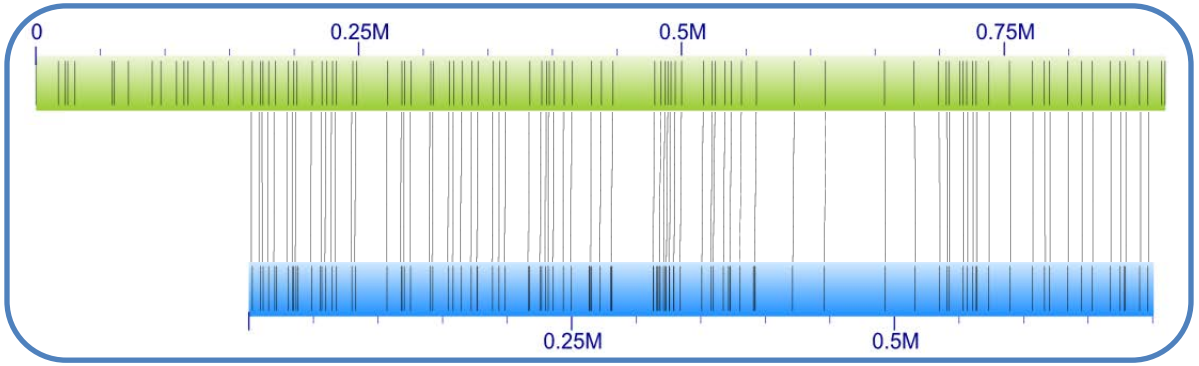


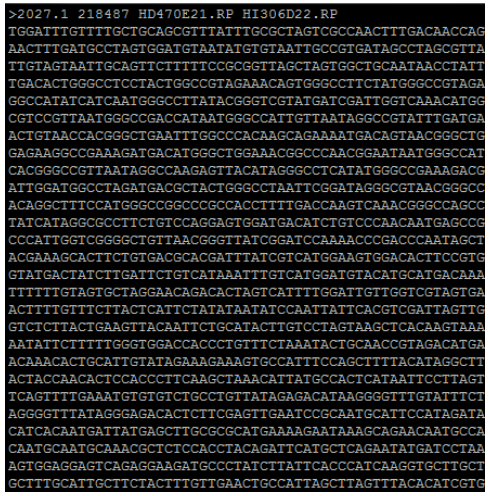
BioNaNo

Comparisons with NGS assembly



Nanomap assembling





in silico digestion

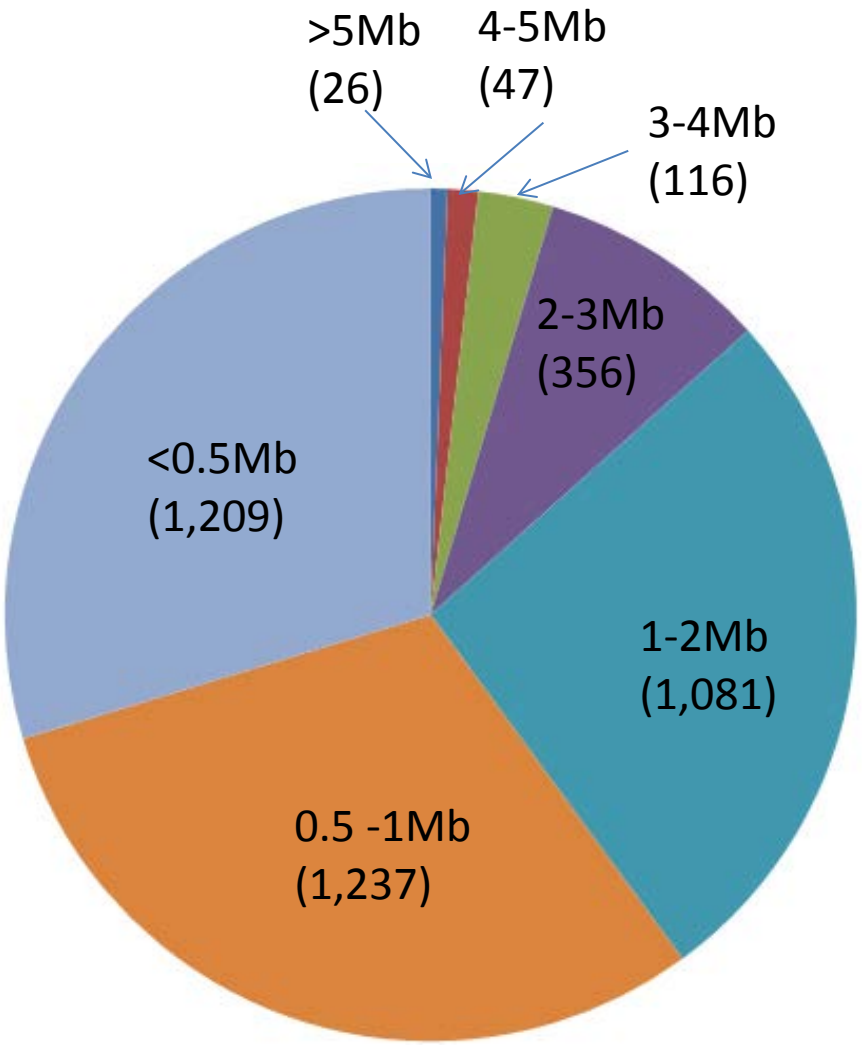
BioNano maps for the *Ae. tauschii* genome

- **BAC-based BioNano maps (Hastie *et al*. 2013, PLoS One)**
- **Whole-genome approach**

Data collection

- Nick restriction: Nt.BspQ1
- BioNano Irys System, chip v. 2
- Total molecules collected (>20 Kb): 901 Gb
- Average label density: 12 sites per 100 Kb
- Molecules used for assembly (>150 Kb): 411 Gb (100X)

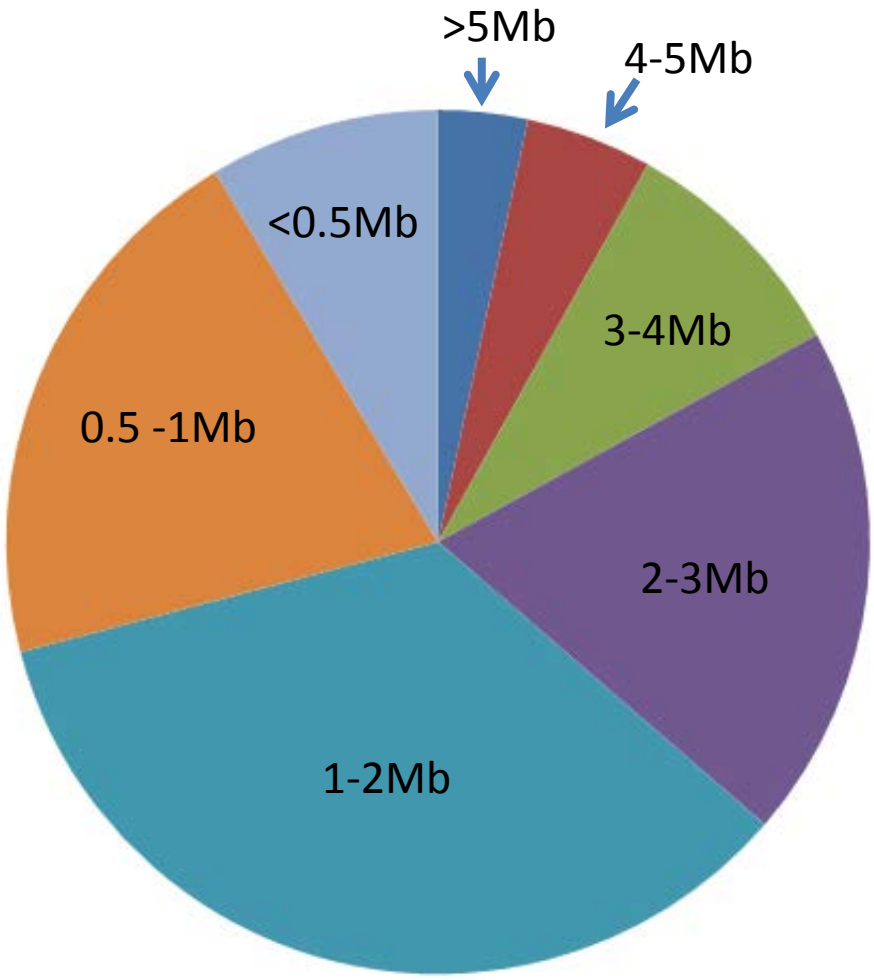
Results of initial assembly



Average length: 1.09 Mb

Total assembled length: 4.4 Gb

Results of initial assembly

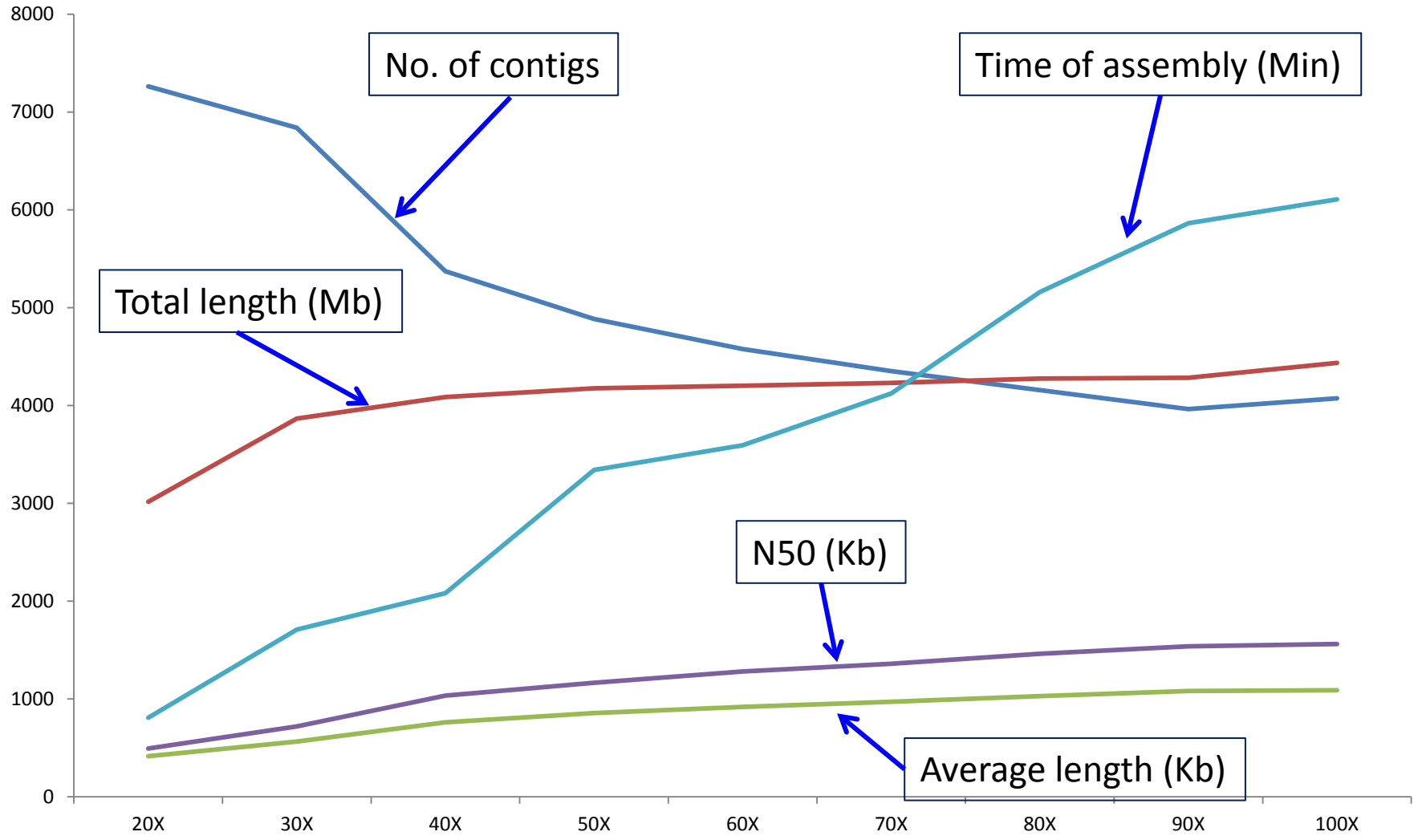


N50 = 1.56 Mb

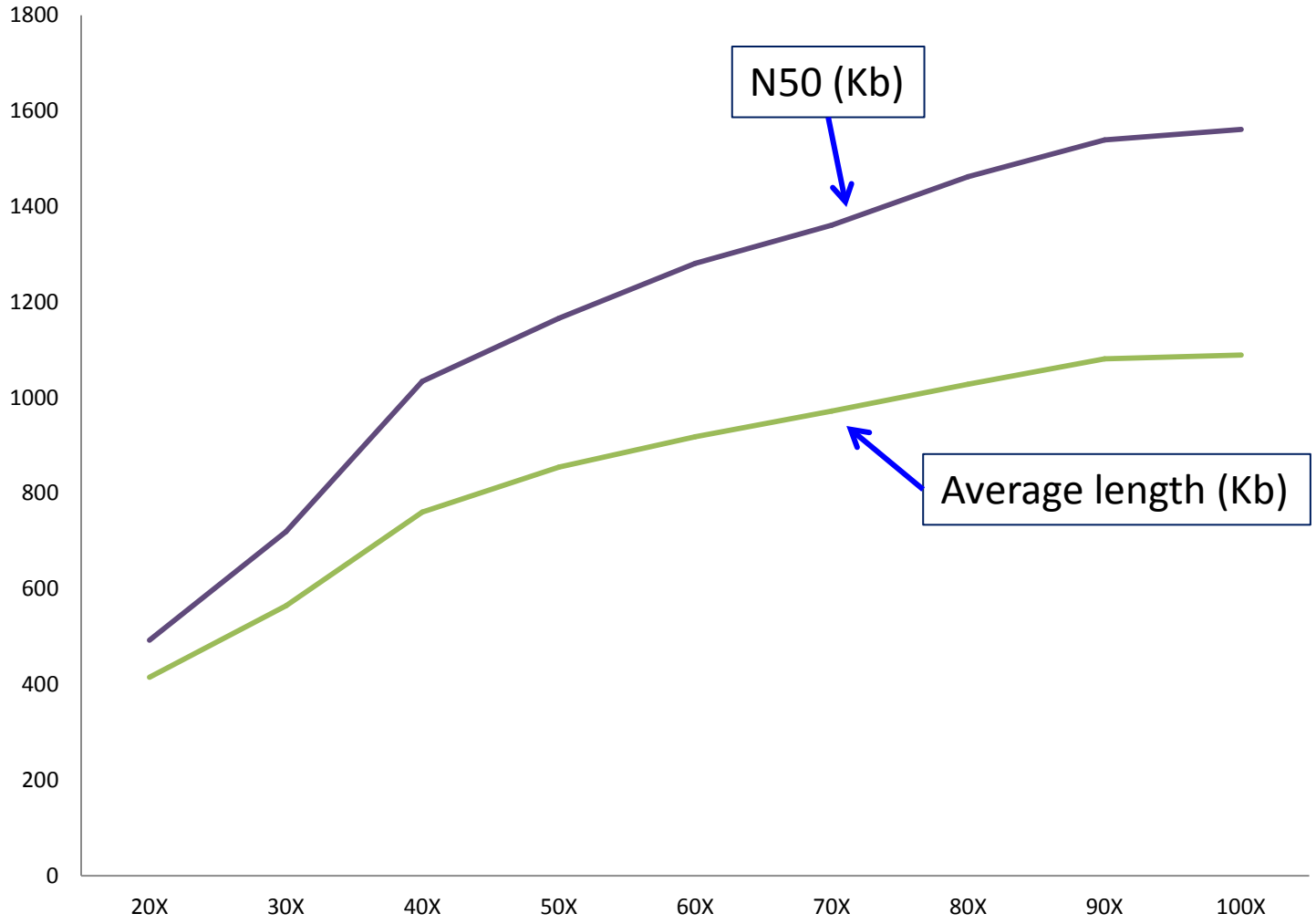
Coverage and assembly

Coverage	No. contigs	Length (Mb)	Average (Kb)	N50 (Kb)	Time (hr)
20X	7,261	3,015	415	493	14
30X	6,838	3,866	565	720	29
40X	5,372	4,087	761	1,034	35
50X	4,884	4,175	855	1,166	56
60X	4,576	4,202	918	1,281	60
70X	4,351	4,230	972	1,361	69
80X	4,158	4,275	1,028	1,462	86
90X	3,964	4,283	1,081	1,539	98
100X	4,072	4,435	1,089	1,561	102

Coverage and assembly



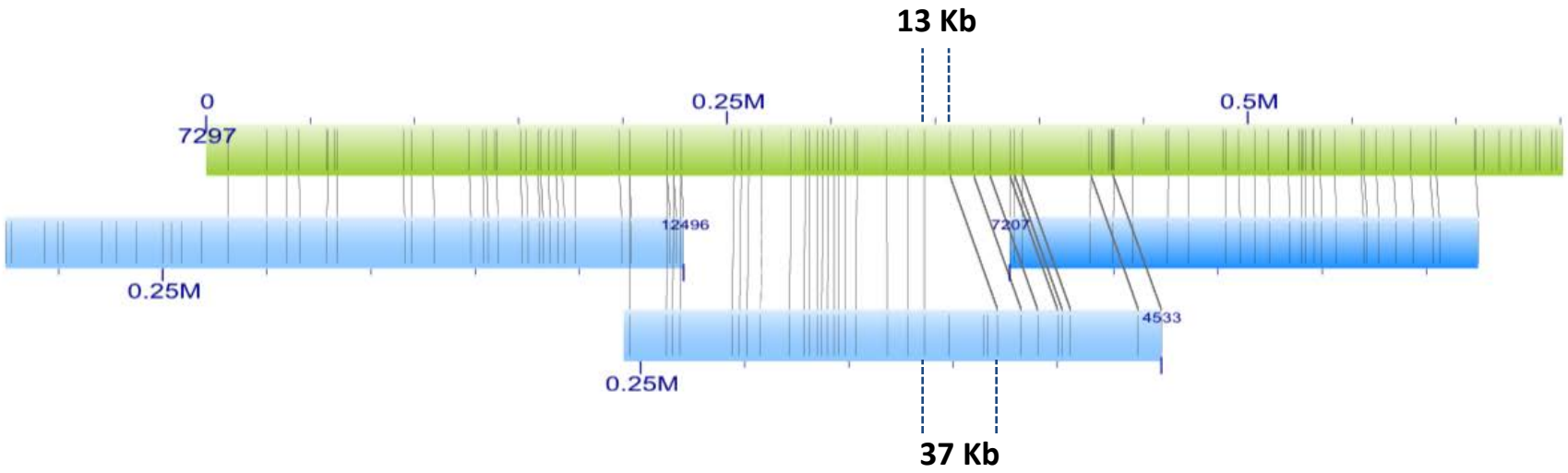
Coverage and assembly



How we are using the BioNano maps?

- **Detect contentious regions**
- **Rectify order & orientations**
- **Sizing gaps**
- **Locate unanchored BAC scaffolds**
- **Build pseudomolecules**

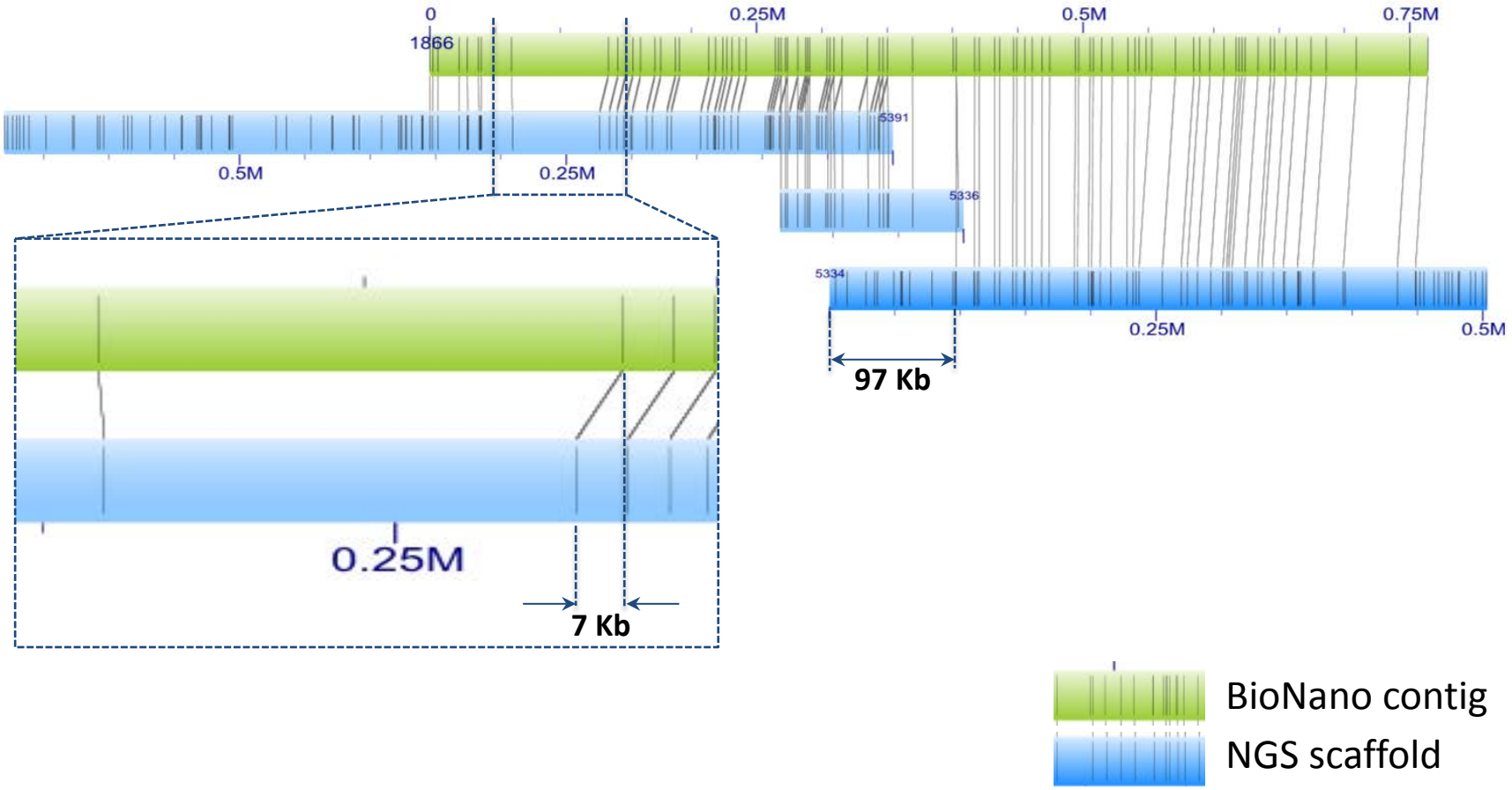
Detection of contentious regions: extra sequence



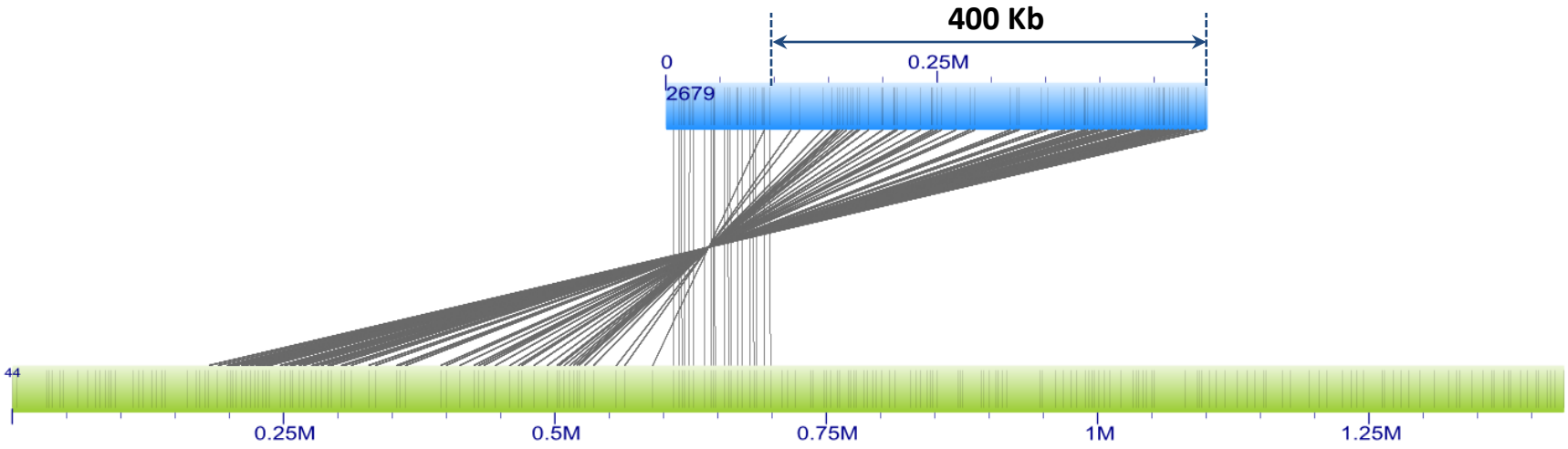
BioNano contig

NGS scaffold

Detection of contentious regions: mis-assembly



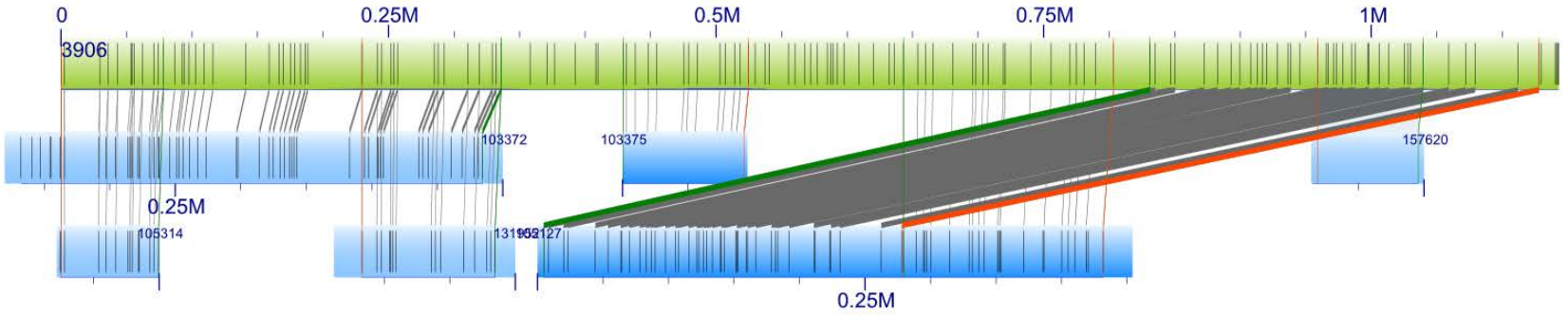
Detection of contentious regions: error in scaffolding



BioNano contig

NGS scaffold

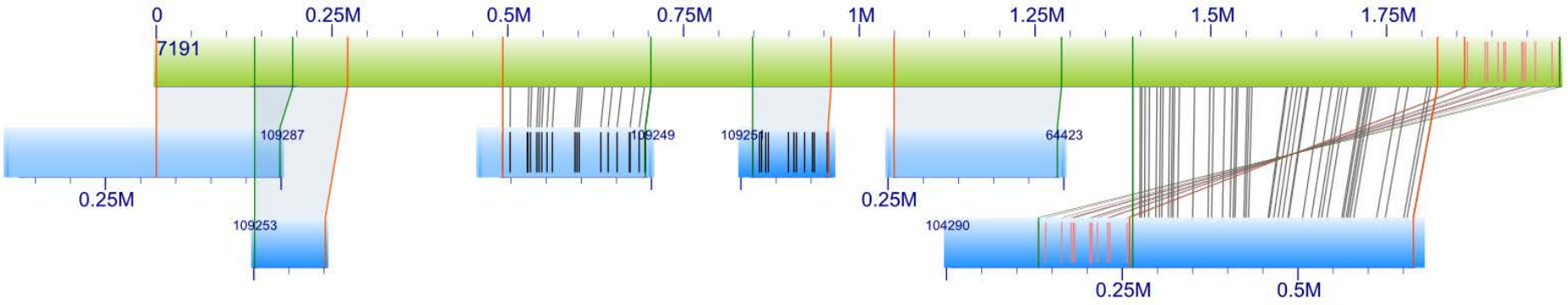
Detection of contentious regions: error in scaffolding



BioNano contig

NGS scaffold

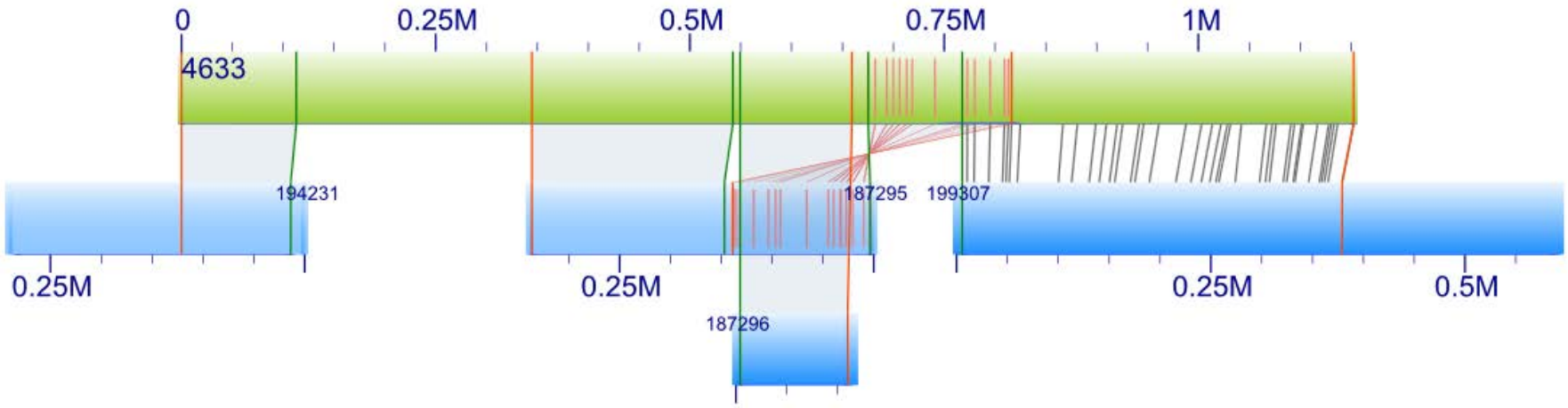
Detection of contentious regions: error in scaffolding



BioNano contig

NGS scaffold

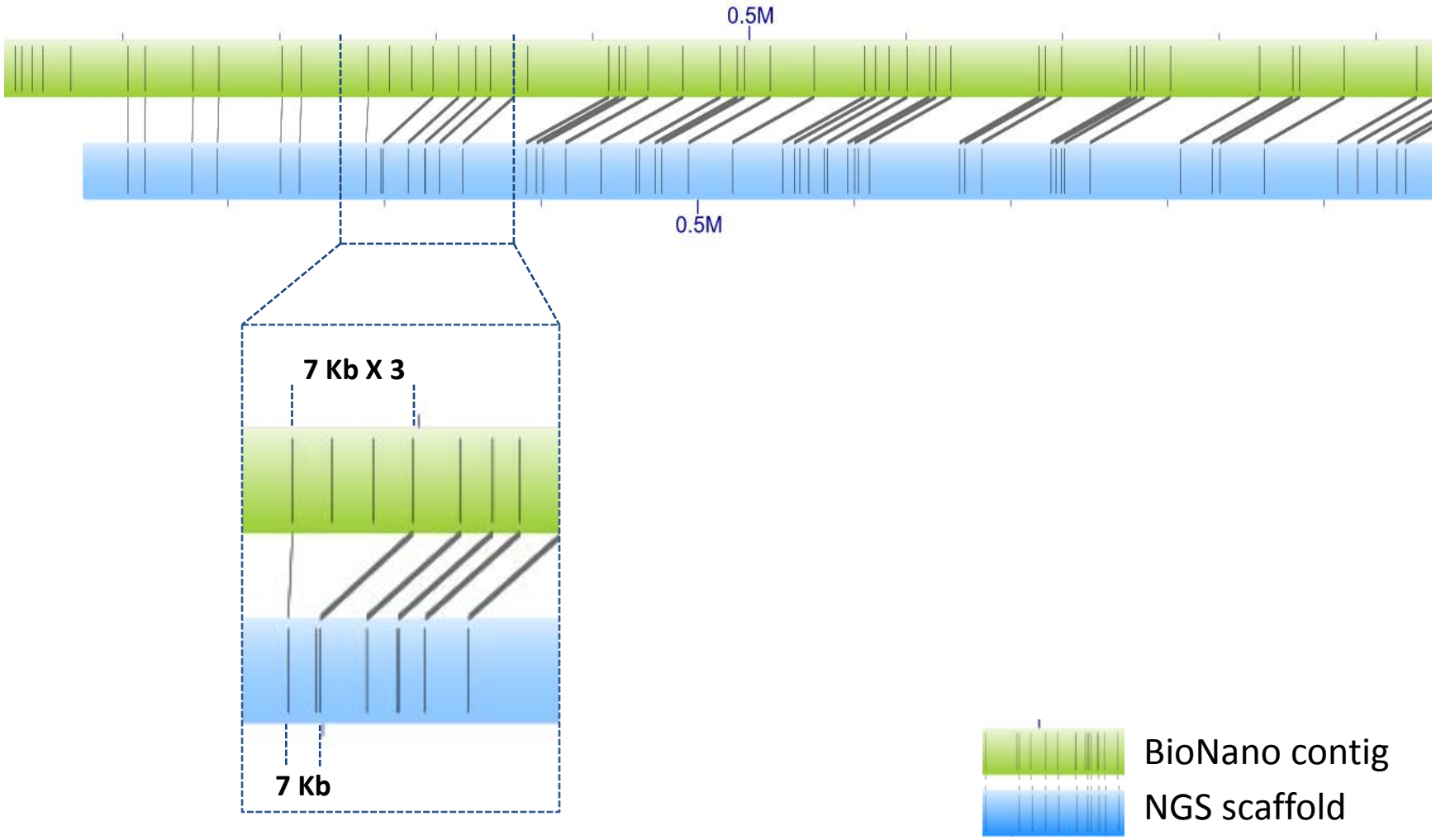
Detection of contentious regions: error in scaffolding



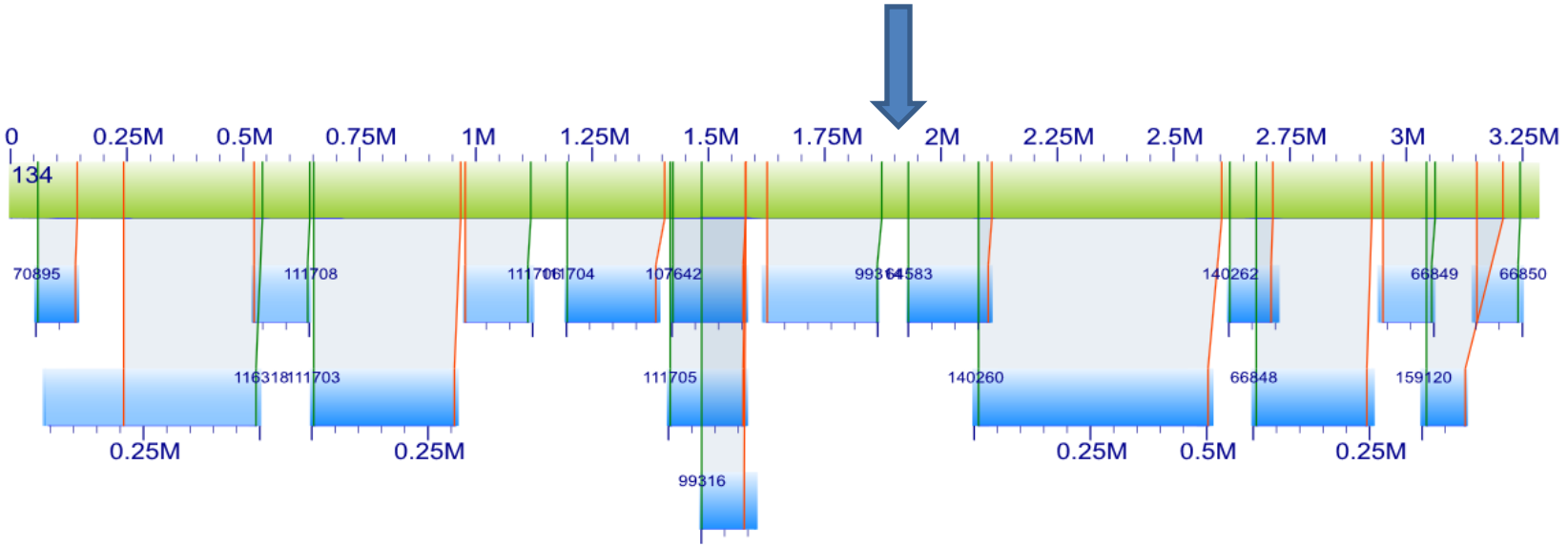
BioNano contig

NGS scaffold

Detection of contentious regions: repeats



Locating unanchored BAC contigs/scaffolds



ctg5752 – ctg1331

2D at 108 cM

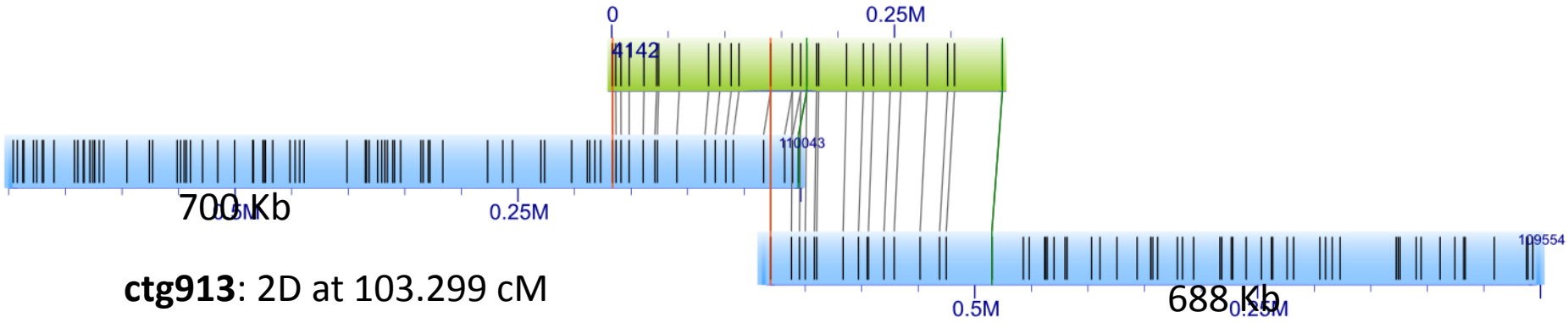
ctg11306 – ctg1606 – ctg10317

Un-anchored

BioNano contig

NGS scaffold

Link BAC contigs



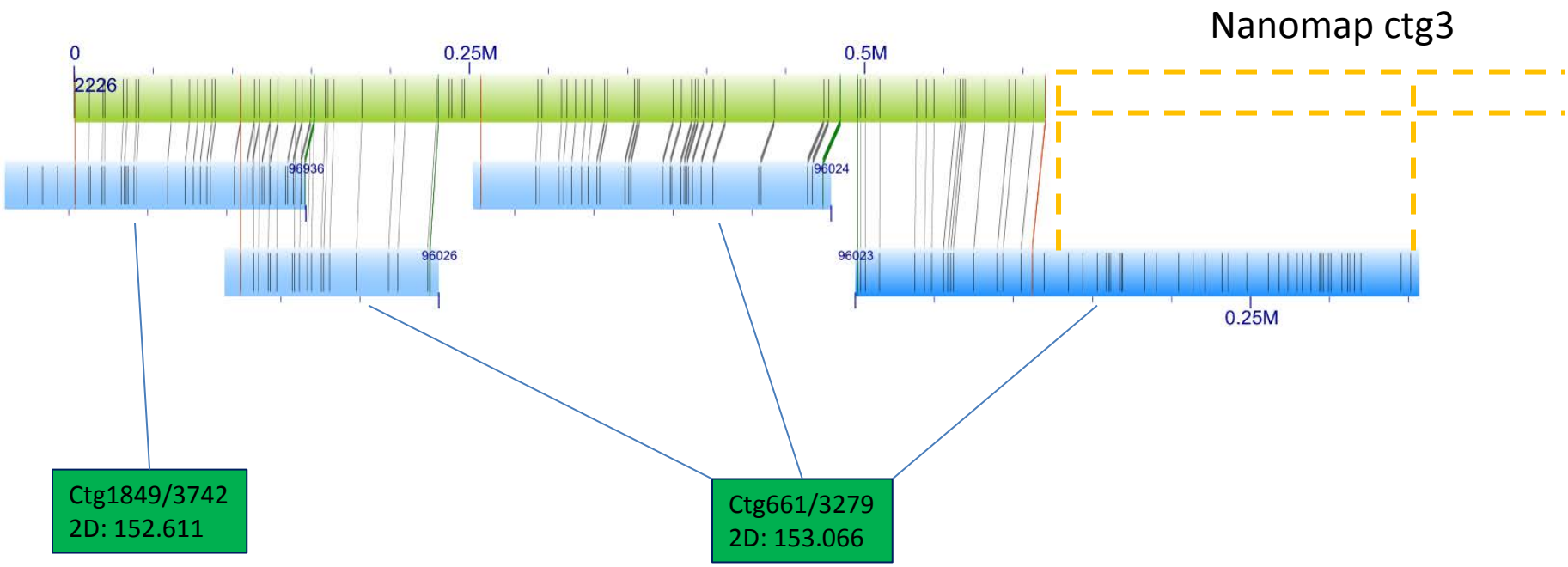
ctg913: 2D at 103.299 cM

ctg1885: 2D at 103.299 cM

BioNano contig

NGS scaffold

Merge of nanomap contigs

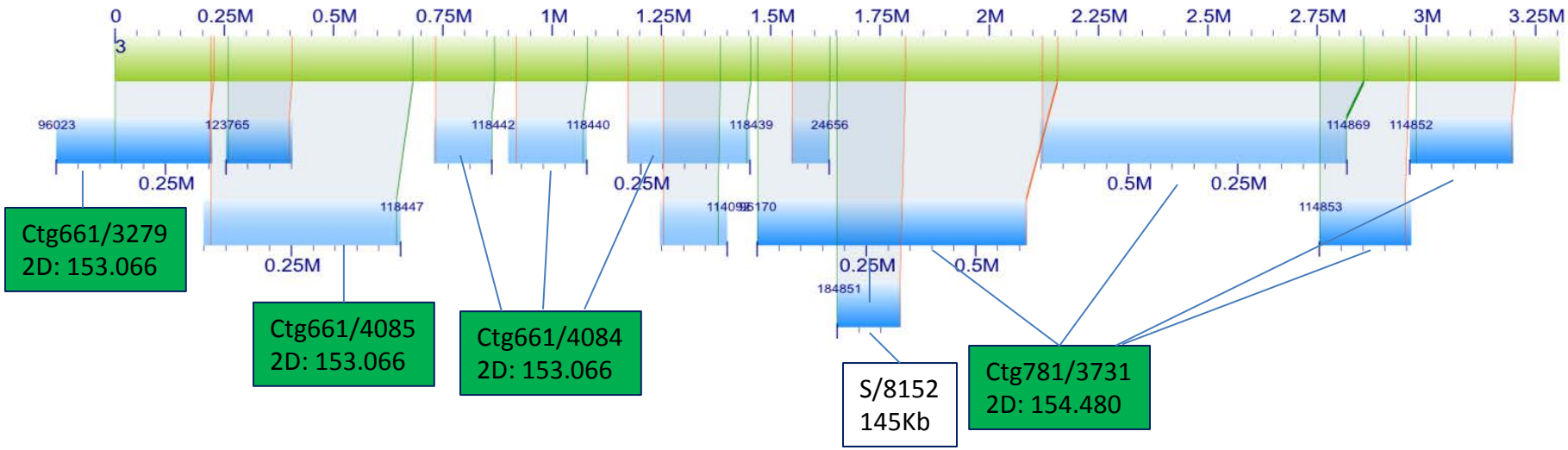


Ctg1849/3742
2D: 152.611

BAC contig ID: **ctg1849**; MTP pool no.: **3742**
On the chromosome **2D** at **152.611** cM

BioNano contig
NGS scaffold

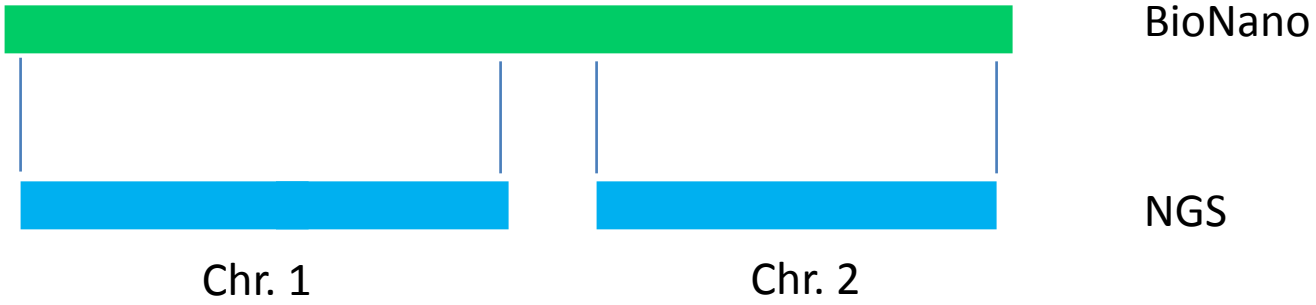
Merge of nanomap contigs



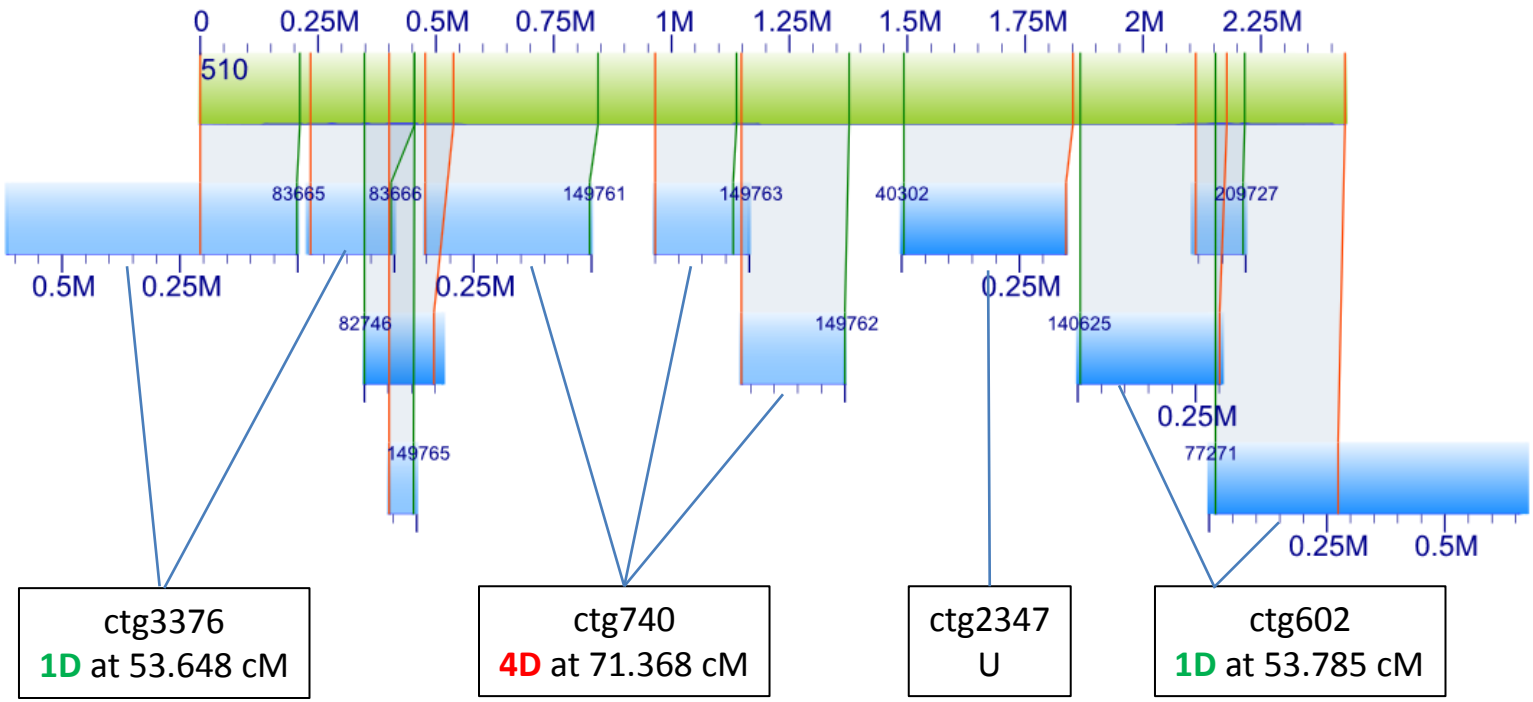
Merge of nanomap contigs

	No. of contigs	Average length (Mb)	N50 (Mb)
Initial assembly	4,072	1.09	1.56
1st round merge	3,665	1.21	1.74

Problem in BNG or NGS?

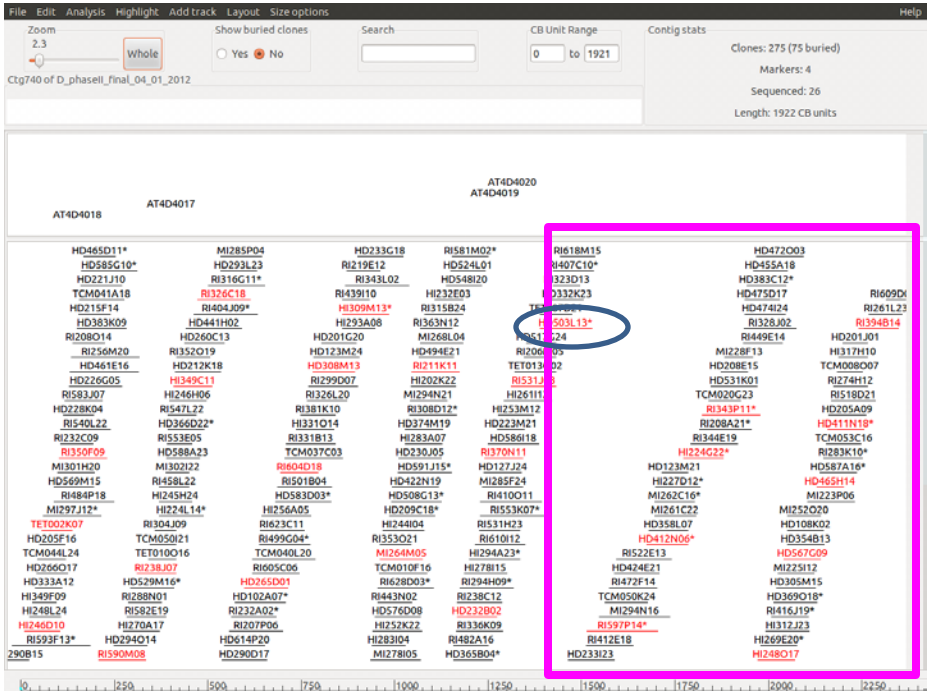


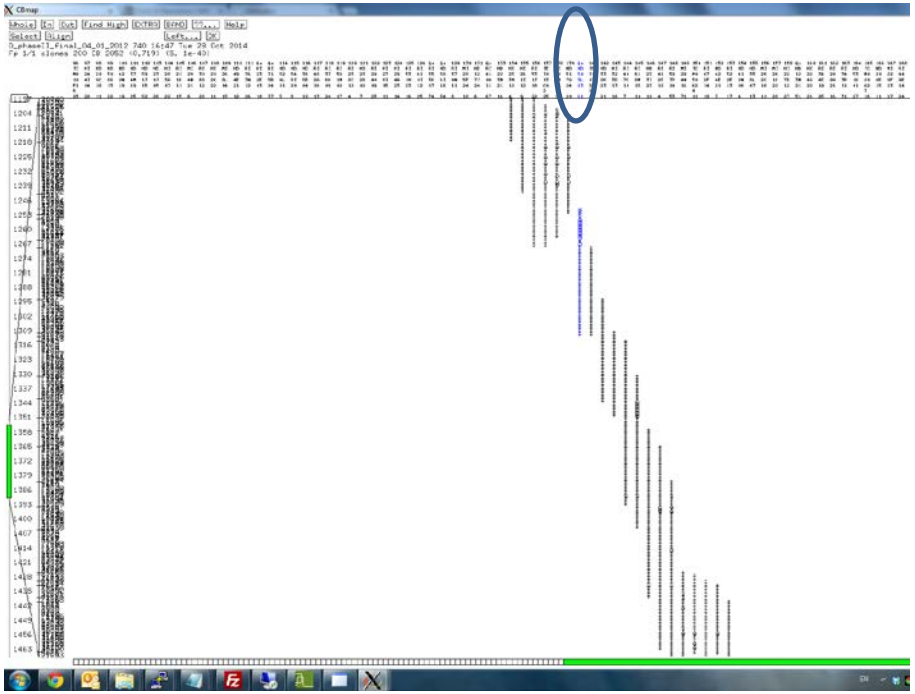
Problem in BNG or NGS?



Problem in BNG or NGS?

ctg740
4D at 71.368 cM

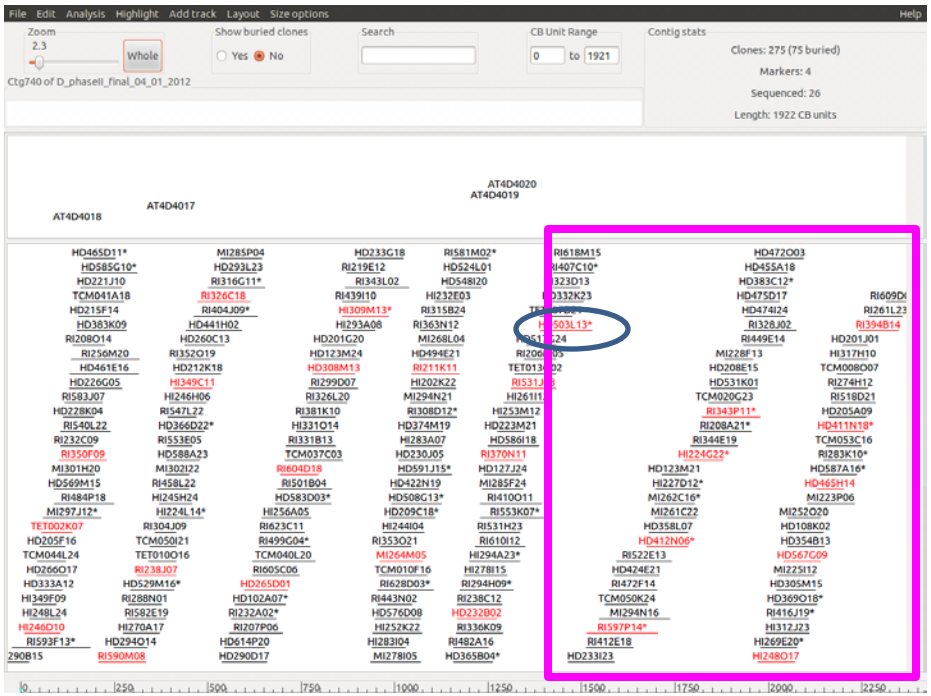


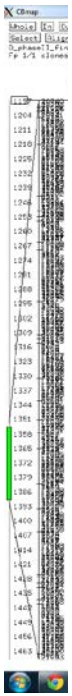


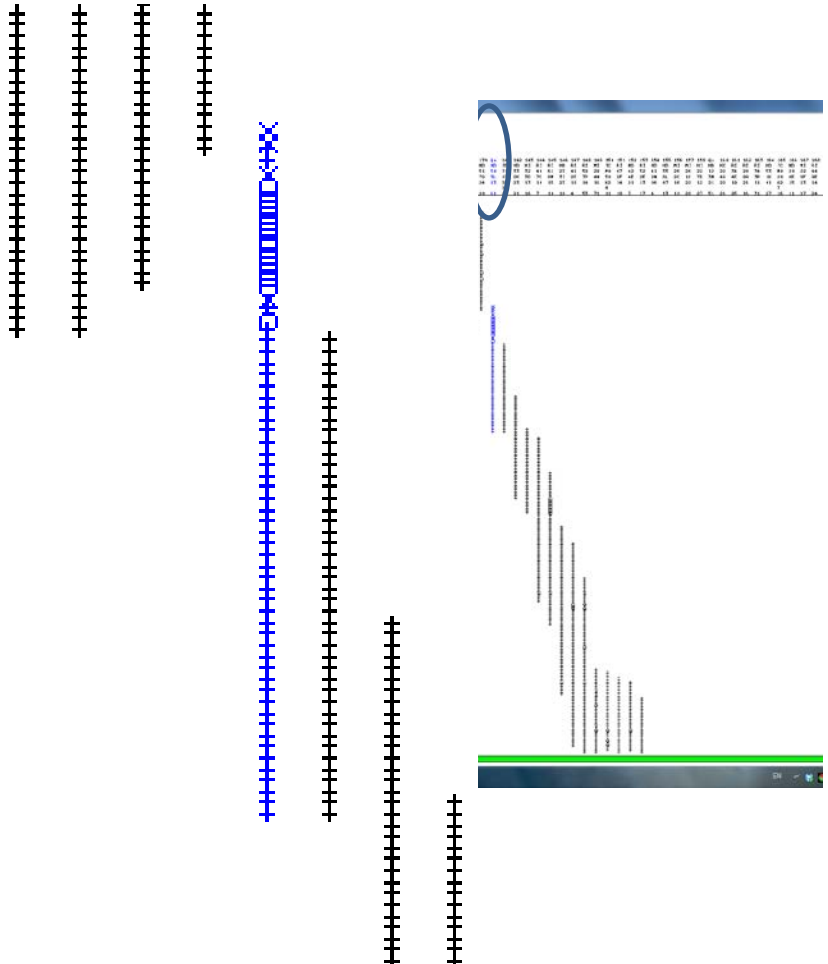
Problem in BNG or NGS?

ctg740
4D at 71.368 cM

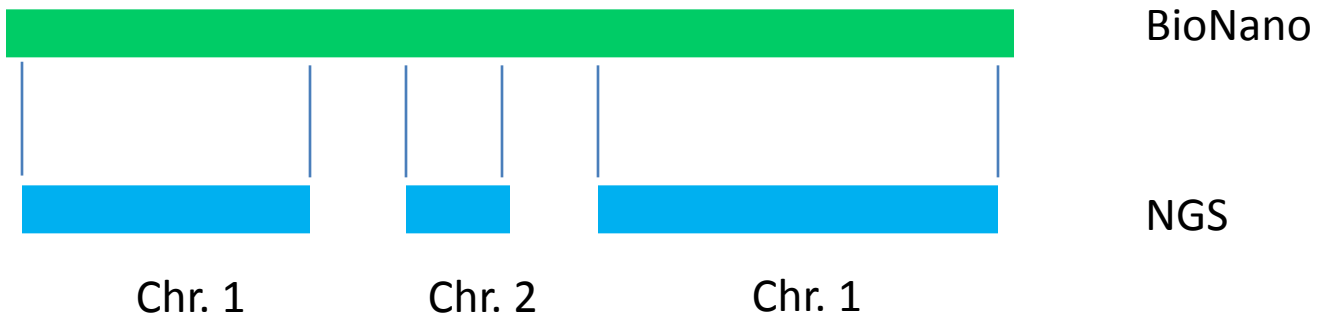
Chimeric BAC contig!



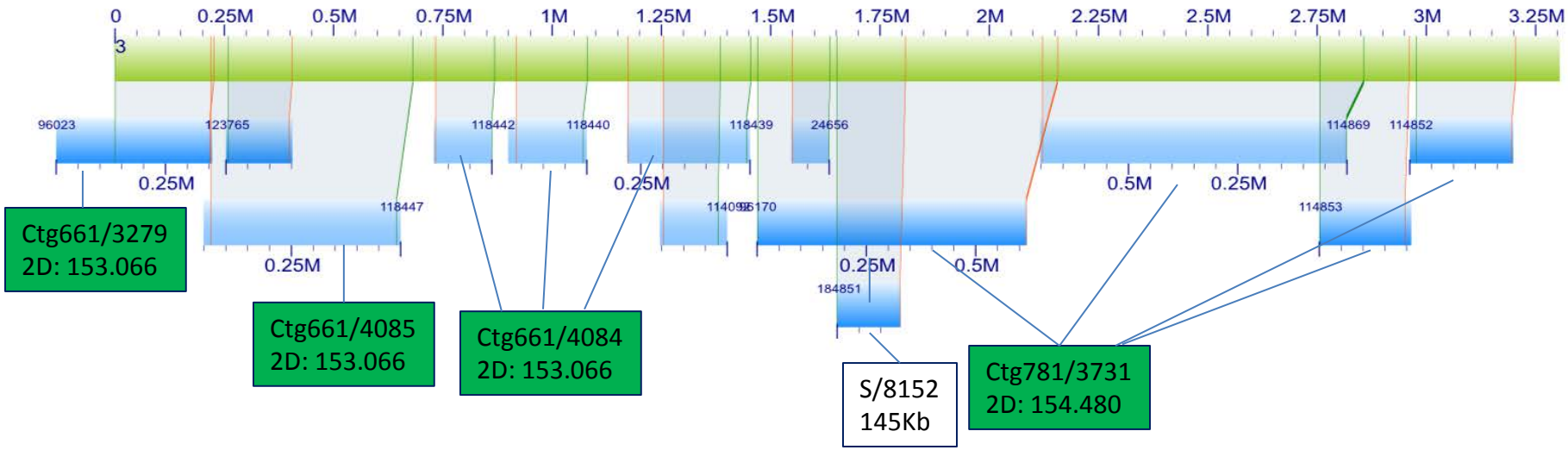


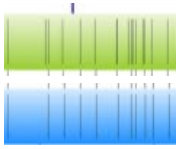


Problem in BNG or NGS?



Merge of nanomap contigs

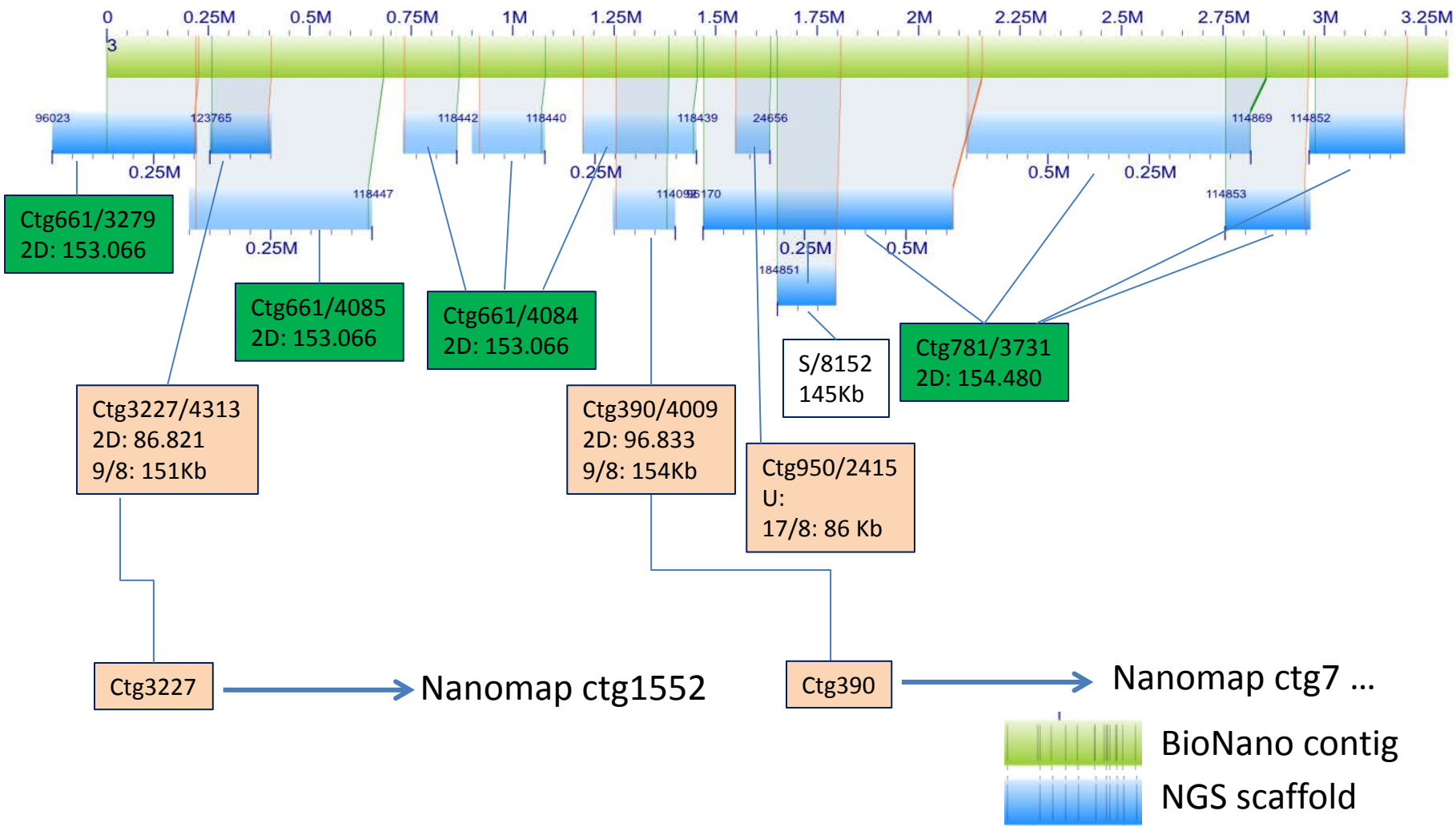




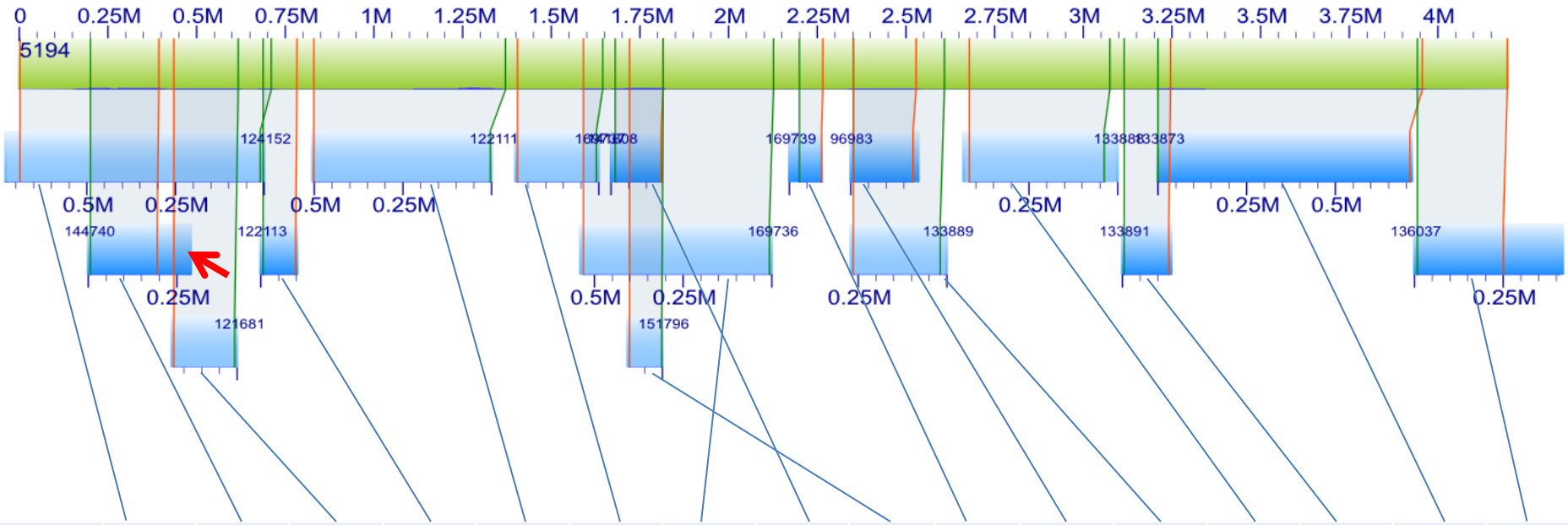
BioNano contig

NGS scaffold

Cross-contamination during NGS



Cross-contamination during NGS



NGS ID	124152	144740	121681	122113	122111	169737	169736	141808	151796	169739	96983	133889	133888	133891	133873	136037
Scaff. Length	729,584	288,847	185,166	102,312	505,262	235,860	540,245	142,784	98,256	90,789	190,546	271,199	435,311	137,494	715,521	418,277
BAC ctg#	ctg160	ctg372	ctg2901	ctg160	ctg160	ctg1202	ctg1202	ctg4835	ctg48	ctg1202	ctg683	ctg1202	ctg1202	ctg1202	ctg1202	ctg1918
Chrom. #	4D	5D	4D	4D	4D	4D	4D	5D	4D	4D	1D	4D	4D	4D	4D	4D
cM	60.6	156.897	121.334	60.6	60.6	60.555	60.555	39.518	58.827	60.555	103.519	60.555	60.555	60.555	60.555	60.373
BES count	9	8	9	12	12	12	12	20	14	12	11	10	10	10	14	8
Pool size	8	8	6	7	7	10	10	6	8	10	8	8	8	8	8	8

Cross-contamination during NGS

Scaffold#	Scaffold size (bp)	Chrom. by pool#	cM by pool#	BNG map#	BNG_Chrom. Assignment	Scaffold derived marker	Chrom. mapped	Approx. cM
5136.1	367,728	5D	44	126	5D at 44 cM	RJM5136.1F1R1	5D	44

Cross-contamination during NGS

Scaffold#	Scaffold size (bp)	Chrom. by pool#	cM by pool#	BNG map#	BNG_Chrom. Assignment	Scaffold derived marker	Chrom. mapped	Approx. cM
5136.1	367,728	5D	44	126	5D at 44 cM	RJM5136.1F1R1	5D	44
3539.3	78,963	1D	58	46	6D at 96-99 cM	RJM3539.3F1R1	6D	96

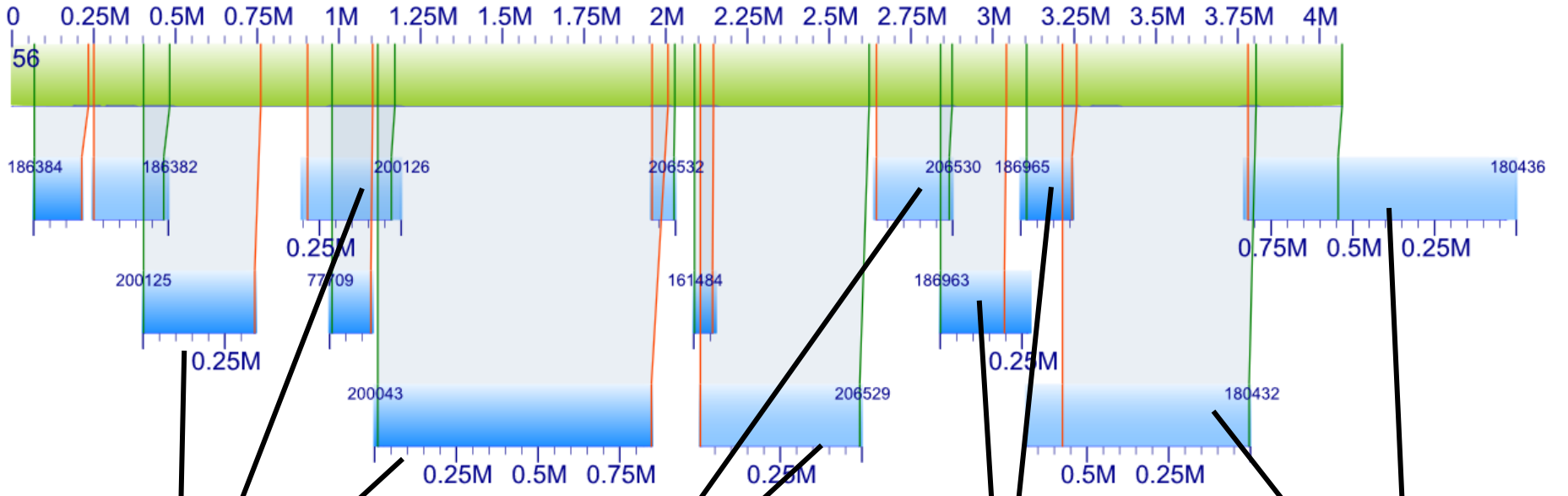
Cross-contamination during NGS

Scaffold#	Scaffold size (bp)	Chrom. by pool#	cM by pool#	BNG map#	BNG_Chrom. Assignment	Scaffold derived marker	Chrom. mapped	Approx. cM
2.1	307,372	3DS	77	28A	3DS at 77 cM	RJM2.1F2R2	3D	77
5136.1	367,728	5D	44	126	5D at 44 cM	RJM5136.1F1R1	5D	44
57.5	144,366	3DS	80	1	3DL at 94 cM	RJM57.5F1R1	3D	94
262.2	147,939	3DS	37/63	1	3DL at 94 cM	RJM262.2F1R1	3D	92
2080.1	365,412	3DL	88	28B	3DL at 88 cM	RJM2080.1F1R1	3D	88
2172.1	396,715	3DL	93	1	3DL at 93 cM	RJM2172.1F2R2	3D	93
3539.3	78,963	1D	58	46	6D at 96-99 cM	RJM3539.3F1R1	6D	96
3691.3	187,367	1D	55	8367	1D at 55 cM	RJM3691.3F1R1	1D	55
3947.4	147,347	2D	110	9	2D at 160-163 cM	RJM3947.4F1R1	2D	165
4121.3	149,318	2D	234	46	6D at 96-99 cM	RJM4121.3F1R1	6D	99

Cross-contamination during NGS

Scaffold#	Scaffold size (bp)	Chrom. by pool#	cM by pool#	BNG map#	BNG_Chrom. Assignment	Scaffold derived marker	Chrom. mapped	Approx. cM
2.1	307,372	3DS	77	28A	3DS at 77 cM	RJM2.1F2R2	3D	77
5136.1	367,728	5D	44	126	5D at 44 cM	RJM5136.1F1R1	5D	44
57.5	144,366	3DS	80	1	3DL at 94 cM	RJM57.5F1R1	3D	94
262.2	147,939	3DS	37/63	1	3DL at 94 cM	RJM262.2F1R1	3D	92
2080.1	365,412	3DL	88	28B	3DL at 88 cM	RJM2080.1F1R1	3D	88
2172.1	396,715	3DL	93	1	3DL at 93 cM	RJM2172.1F2R2	3D	93
3539.3	78,963	1D	58	46	6D at 96-99 cM	RJM3539.3F1R1	6D	96
3691.3	187,367	1D	55	8367	1D at 55 cM	RJM3691.3F1R1	1D	55
3947.4	147,347	2D	110	9	2D at 160-163 cM	RJM3947.4F1R1	2D	165
4121.3	149,318	2D	234	46	6D at 96-99 cM	RJM4121.3F1R1	6D	99

Ordering and orienting scaffolds & estimating gaps



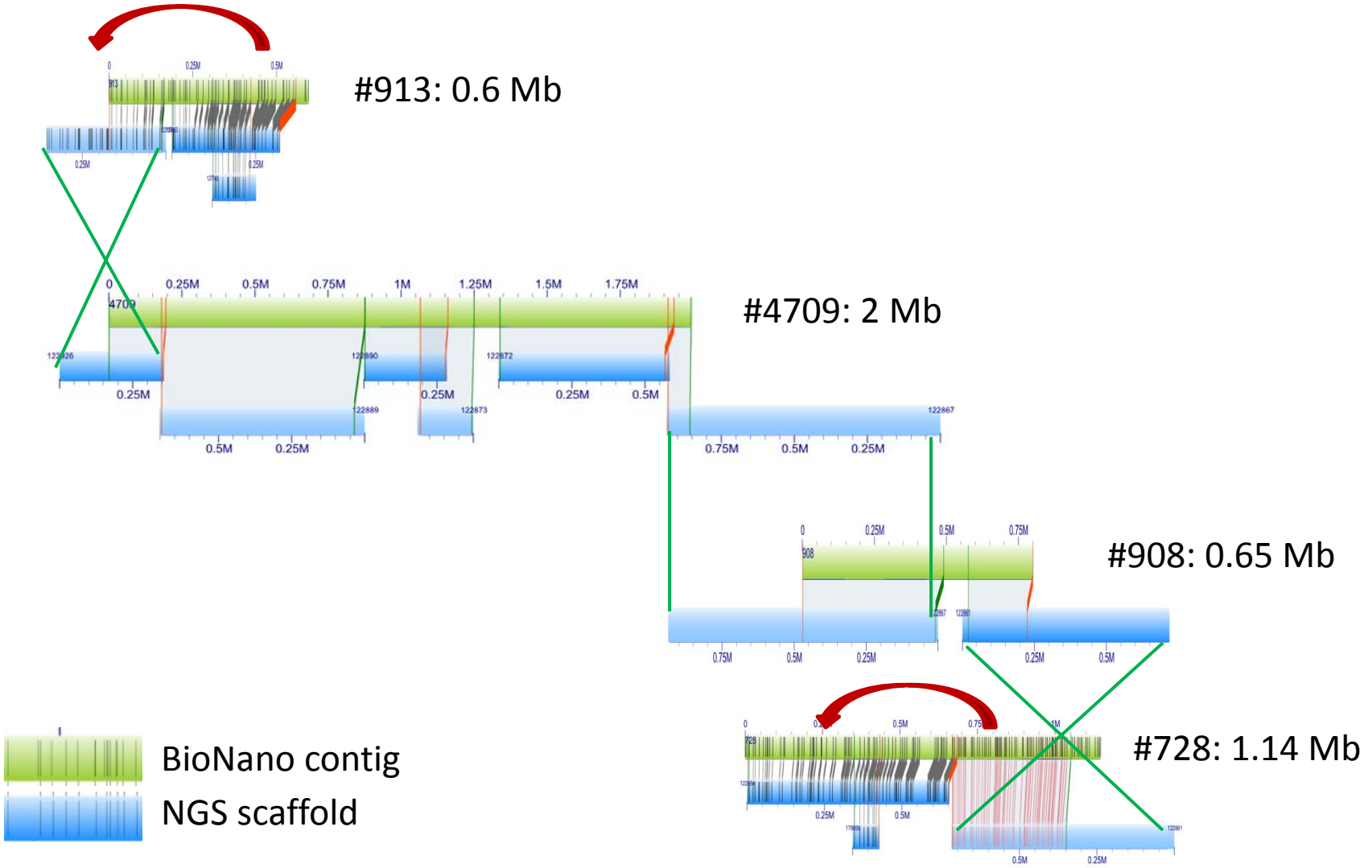
ctg1715
6D at 76.097 cM

ctg12344
6D at 76.097 cM

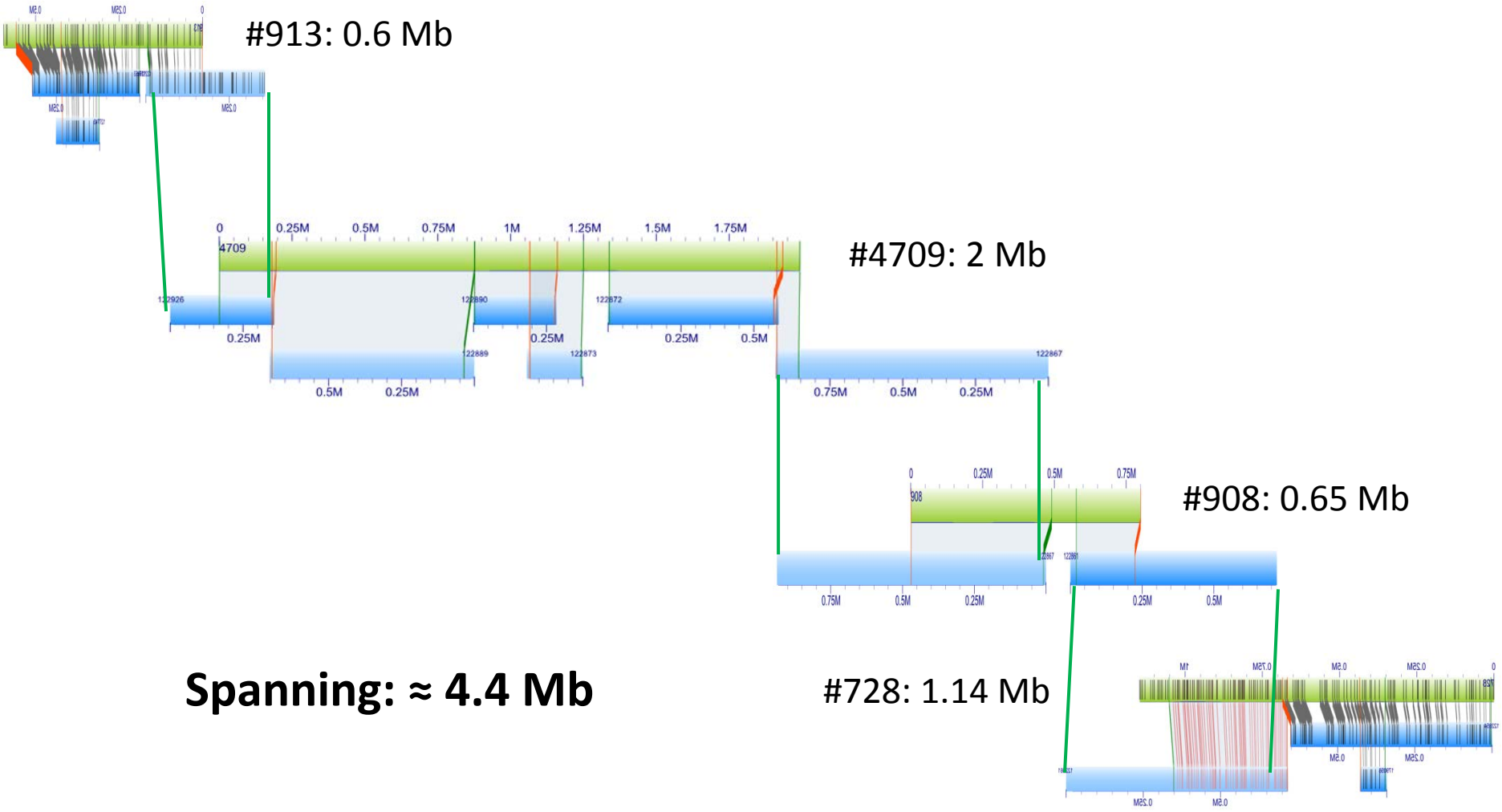
ctg6115
6D at 76.097 cM

ctg195
6D at 75.686 cM

Super-scaffolding to pesudomolecules



Super-scaffolding to pesudomolecules



Nanomap ctg27 (7.4 Mb)

Mapped at centromere region of 3D

In progress

- **Improve the current nanomap**
- **Establish a pipeline to generate AGP file to guide error correction of NGS assembly and scaffolding**
- **Construct pseudomolecule for each of the seven chromosomes.**

From our experiences

- It is possible to construct whole-genome nano maps for large and complex genomes.
- Contamination is inevitable if using clone-based approach; but nano maps will place unintended clone to its right location.
- Whole genome shotgun sequencing will be better off if assembly is good enough to align them on nano maps.
- HMW DNAs with high quality are essential.
- Consistence of data quality is critical (in contrast to sequencing; 4700 Gb vs. 901 Gb).
- We do see chimeric BioNano contigs, 11 among ca. 3000 contigs examined (0.4%).
- BAC singletons won't contribute much to fill gaps.

Acknowledgements

UC DAVIS UNIVERSITY OF CALIFORNIA

Karin R. Deal
Armond Murray
Sonny Van
Tanh Ngo
Scott Liu
Lichan Xiao
Hai Long
Juan Rodriguez
Naxin Huo
Luxia Yuan
Luis Curiel
Yi Wang
Patrick McGuire
Jan Dvorak

UC DAVIS UNIVERSITY OF CALIFORNIA USDA ARS

Yong Q. Gu
Olin D. Anderson

JOHNS HOPKINS UNIVERSITY

Daniela Puiu
Geo Pertea
Steven Salzberg

中國農業大學 China Agricultural University

Shuhong Ouyang
Yong Liang
Zhenzhong Wang
Zhiyong Liu
Qixin Sun

Nanjing Agricultural University

Zhengqiang Ma

中国农业科学院 CHINESE ACADEMY OF AGRICULTURAL SCIENCES

Long Mao

BIONANO GENOMICS

Alex Hastie
Andrew Anfora
Palak Sheth

NSF

Financial support:
PGRP/NSF, USA