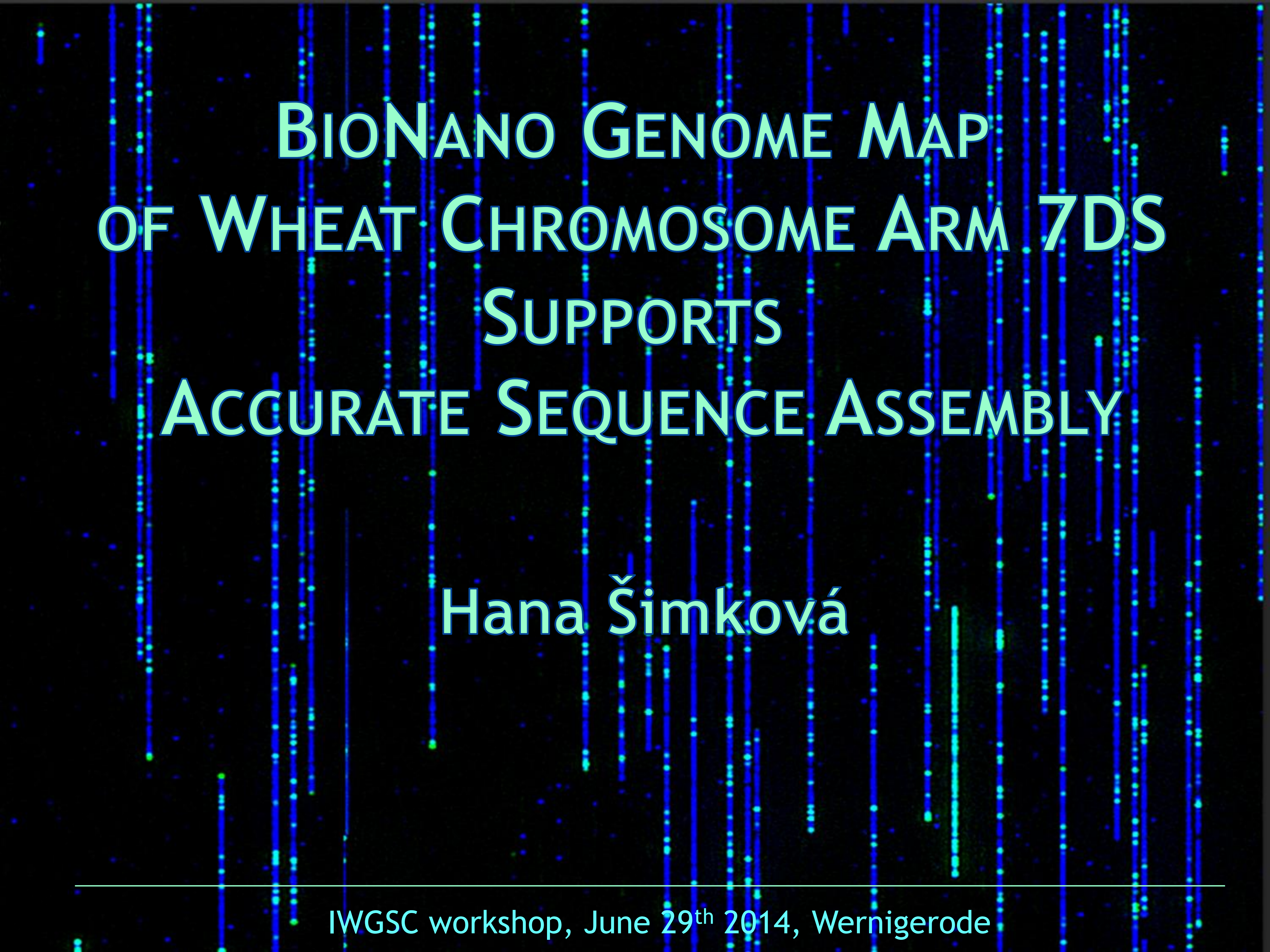


BioNano Genome Map of Wheat Chromosome Arm 7DS Supports Accurate Sequence Assembly

Hana Šimková

IWGSC workshop, June 29th 2014, Wernigerode

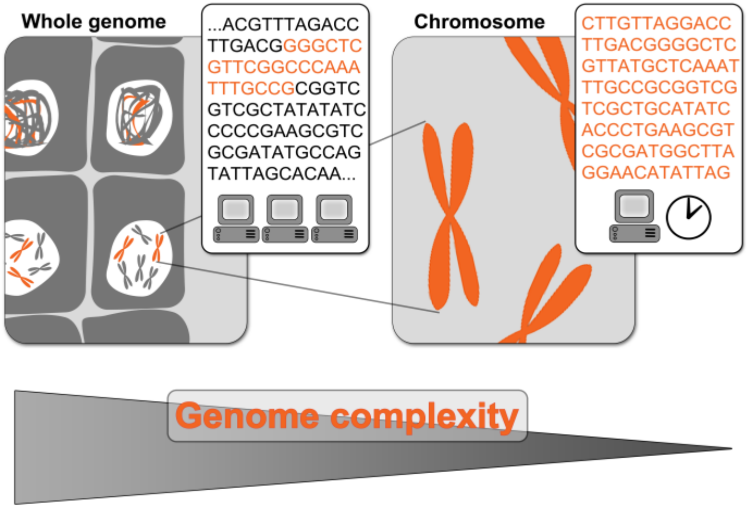
CHALLENGES IN PLANT GENOME SEQUENCING

De novo genome assemblies using only short read data of NGS technologies are generally incomplete and highly fragmented due to

- Large duplications - chromosomal approach, BAC-by-BAC sequencing
- High proportion of repetitive DNA - challenge!

- Large genome size
- Polyploidy

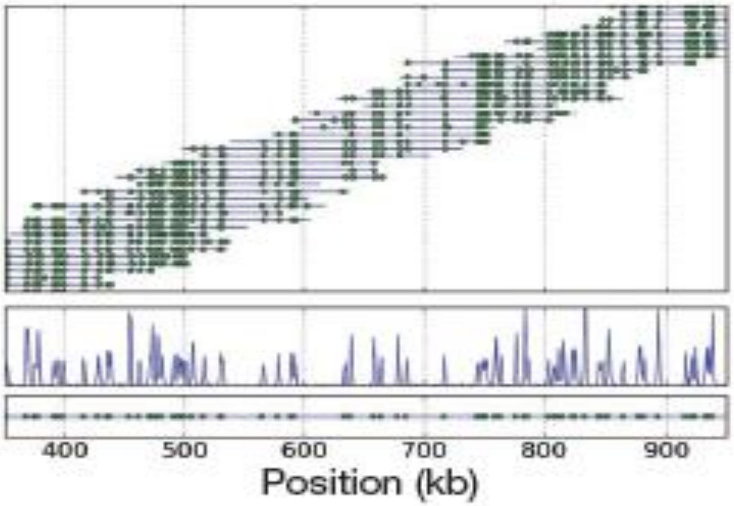
Chromosomal approach



SOLUTIONS FOR THE REPEATS

- **Long mate-pair reads**
- **Long read technologies** – PacBio, Moleculo, Oxford Nanopore
- **Optical mapping**/genome mapping in nanochannel arrays (BioNano Genomics, Irys platform)

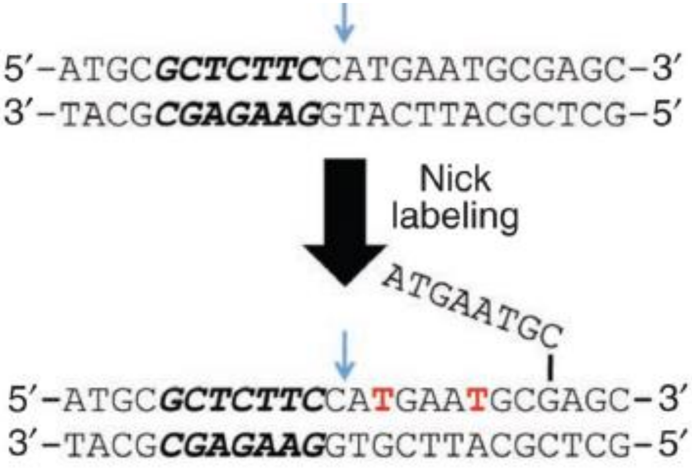
Single-molecule mapping of genomic DNA hundreds of kilobases in size



BioNano Genome mapping on Nanochanel Arrays

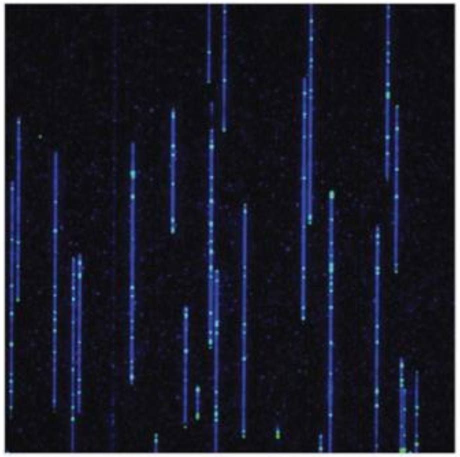
1 Sequence-specific labeling

Nickase (Nt.BspQI)

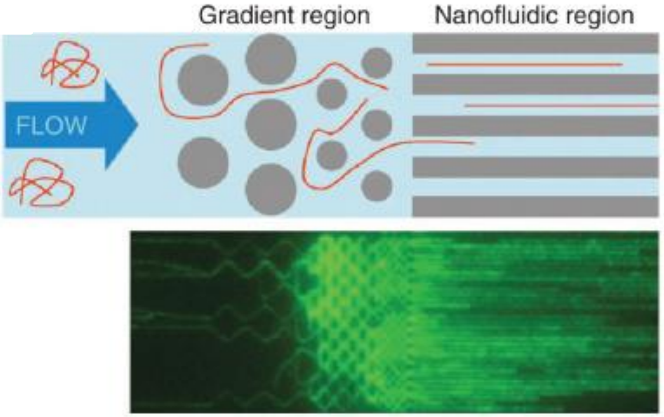

5′–ATGC***GCTCTTC***CATGAATGCGAGC–3′
3′–TACG***CGAGAAG***GTACTTACGCTCG–5′

5′–ATGC***GCTCTTC***CATGAATGCGAGC–3′
3′–TACG***CGAGAAG***GTGCTTACGCTCG–5′

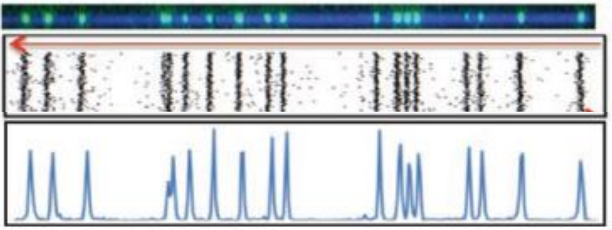
3 Fluorescence imaging



2 DNA linearization



4 Map construction

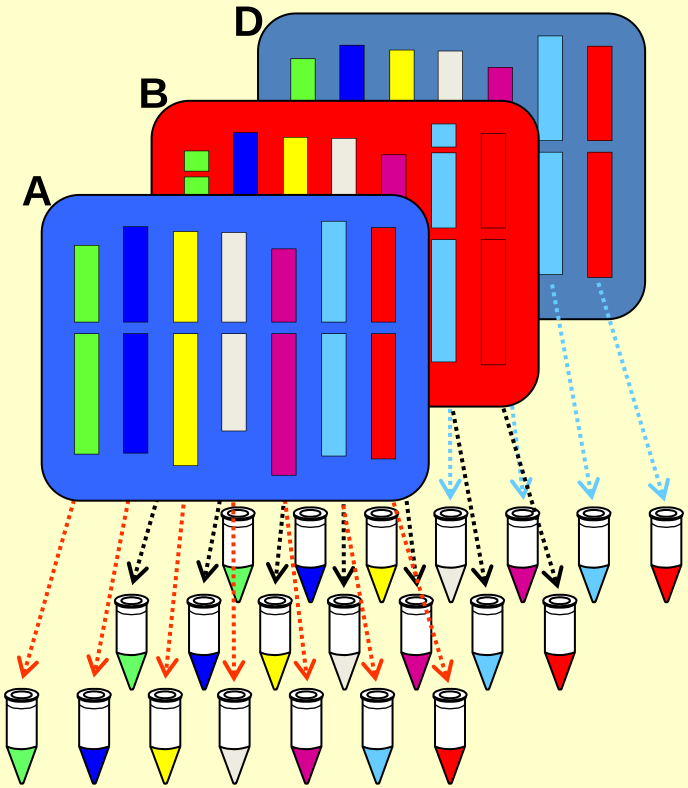


Output:

sequence motif (GCTCTTC) map along hundreds-kb to megabase DNA stretches

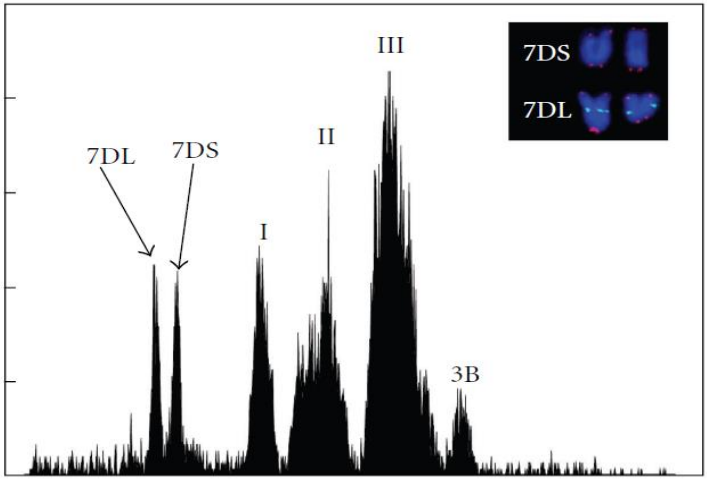
Lam et al., Nat. Biotechnol. 30(8) 2012

CHROMOSOME MAPPING ON NANOCHANNEL ARRAYS



Dissecting the bread wheat genome by flow sorting

- Pilot study on wheat **7DS** chromosome arm (381 Mb, 2.25% wheat genome)
  - Purified as telocentric chromosome by flow cytometric sorting

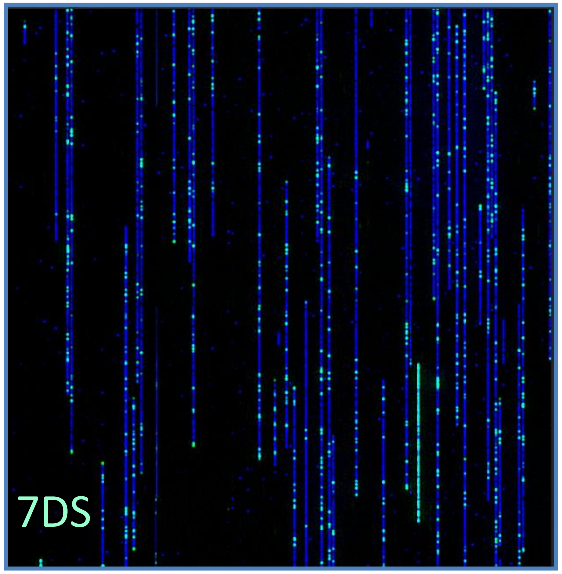


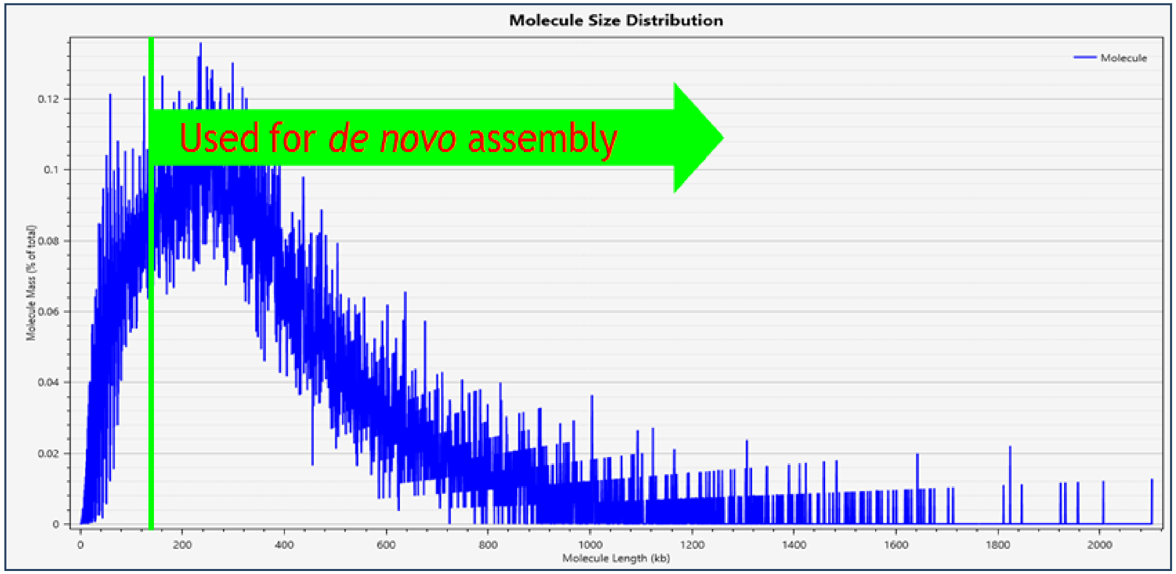
- *In silico* analysis (7DS CSS sequence) for chromosome mapping
  - Nt.BspQI ~13 sites per 100kb
  - Nb.BbvCI ~7 sites per 100kb

BIONANO MAP OF 7DS: DATA ACQUISITION

- Three miniplugs from flow-sorted 7DS chromosome arm:
  - flow sorted equivalent of 950 ng, recovered **575 ng** at 25ng/µl
- Labelling - **Nt.BspQI**
- Collecting data from **one version-2 chip**

Length treshold	Total coverage	n50
150kb	200x	344kb

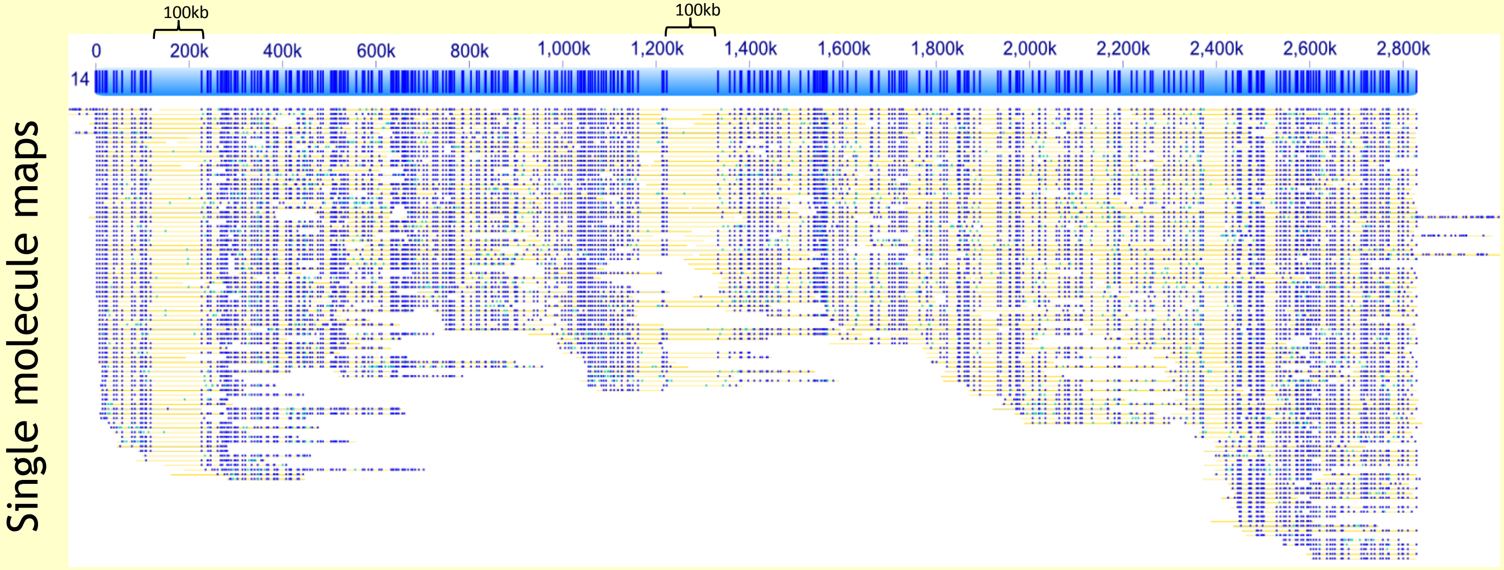




DE NOVO ASSEMBLY OF 7DS

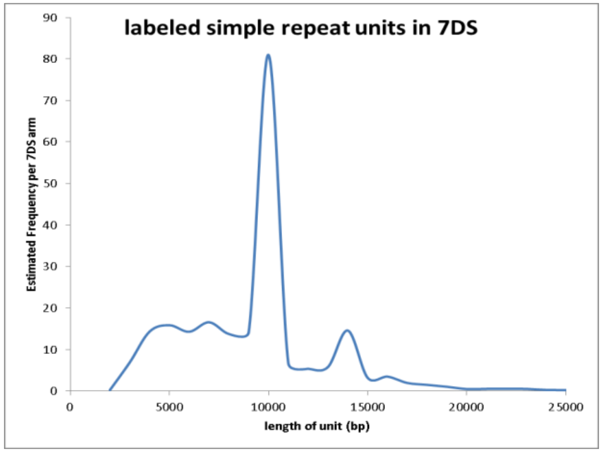
- A total of **371 genome maps** were *de novo* assembled
- Total assembly length is 350Mb (**92% of estimated 7DS size**)
- **Average map size is 0.9Mb**
- **n50 is 1.3 Mb**

Genome map 14

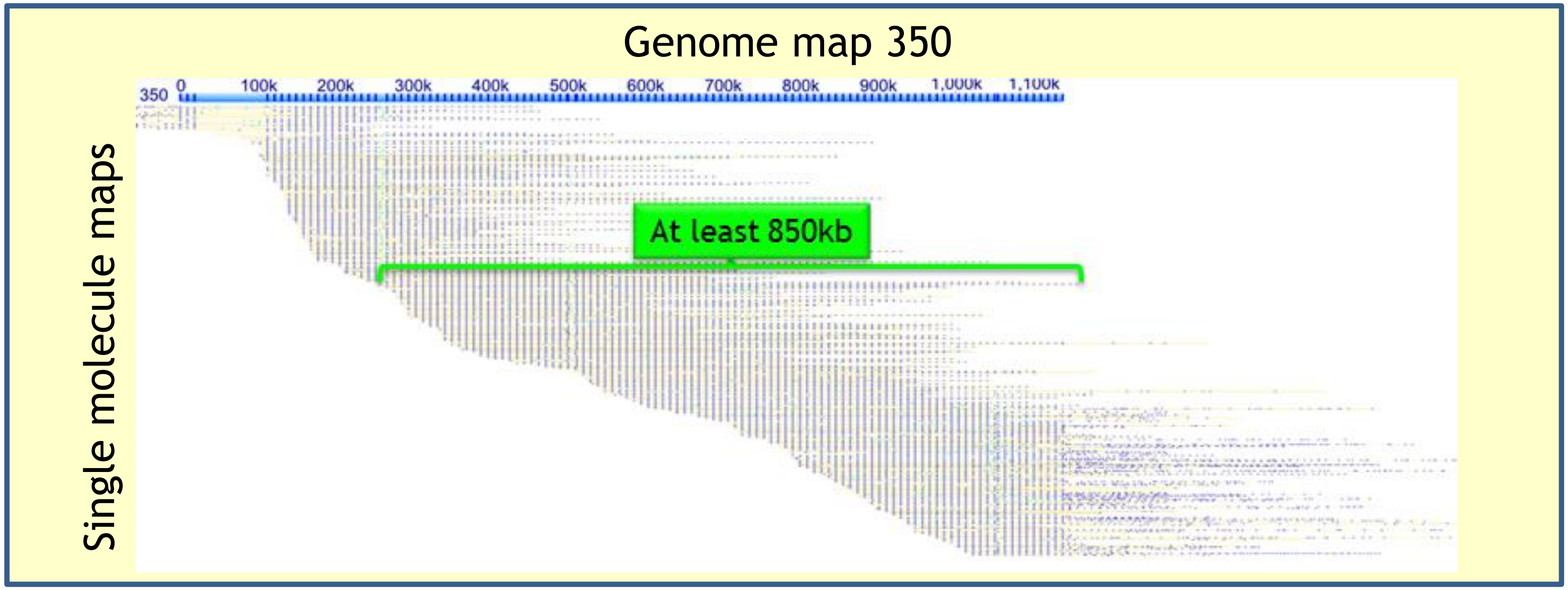


REPEAT UNITS IN 7DS

- Labeled repeat occurrence can be measured from single molecules
- Based on the quantitation of repeat units in single molecules in the whole sample and the longest single array in a molecule, it appears that this repeat is likely contained in a **single array**



Genome map 350



7DS SEQUENCING STRATEGY

- 4608 MTP clones → 1152 pools of four non-overlapping BAC clones
- Illumina pair-end seuqencing - 550bp fragment size, 96 pools per lane of HiSeq, 100bp read length, coverage ~500x
- Assembler – Sassy (Kazakoff *et al*. 2012)
- Deconvolution through BAC end sequences, inner contigs unresolved

⇩

  - **1-20 contigs per BAC clone, median 3.8**
  - **average contig size 24.3 kb**

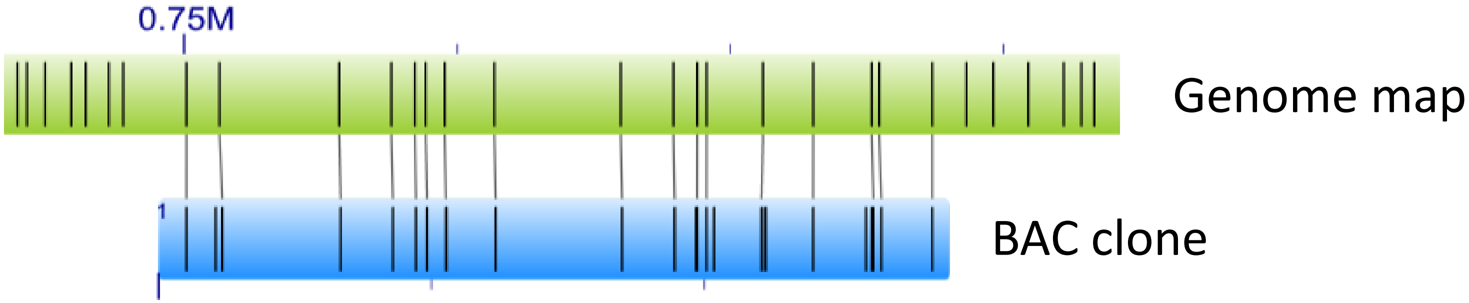
Poster
No. 45

- Assignment of inner contigs based on
  - mate-pair data obtained from MTP-plate pools (384 clones)
  - information from overlapping BAC clones (BLAST on BAC pools)
  - BioNano mapping ?

TESTING BIONANO MAP ON 7DS SEQUENCE

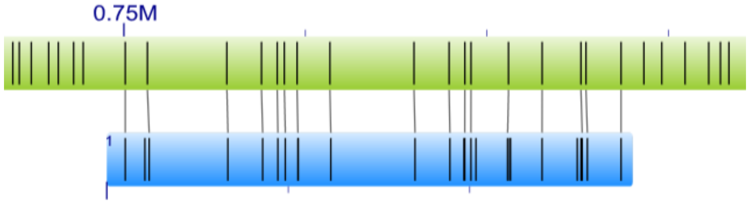
By aligning BAC clone sequences to the BioNano genome map through **IrysView sofware**

- 10 BAC clones assembled as one contig
- ctg 783 of the 7DS physical map
  - 8 BAC clones
  - 700 kb



TESTING BIONANO MAP ON 7DS SEQUENCE

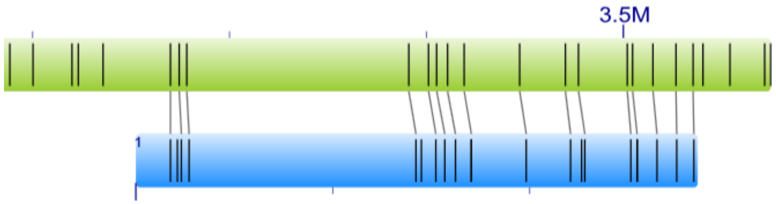
BAC 01



147,078 bp

genome map 62
conf.: 17,79

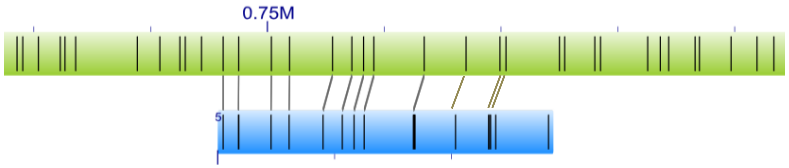
BAC 02



145,818 bp

genome map 21
conf.: 11.33

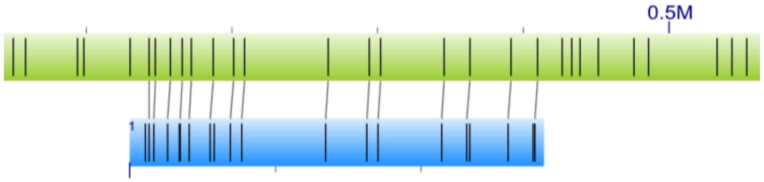
BAC 05



146,641 bp

genome map 249
conf.: 10.13

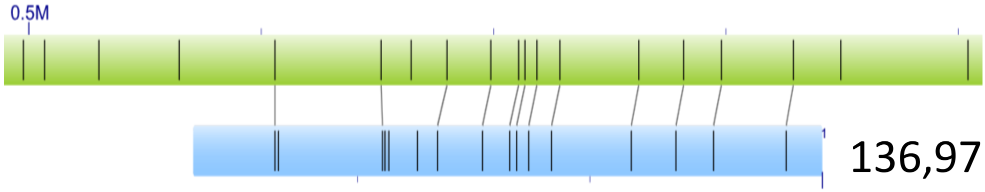
BAC 06



144,521 bp

genome map 26
conf.: 13.25

BAC 08

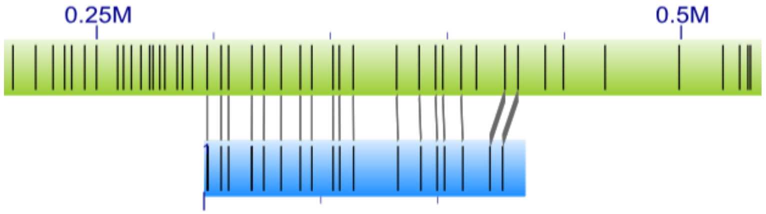


136,972 bp

genome map 167
conf.: 12.94

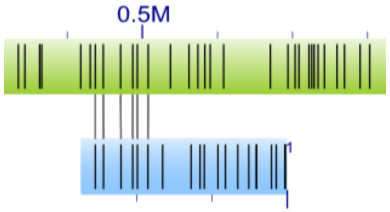
Testing BioNano Map on 7DS Sequence

BAC 09

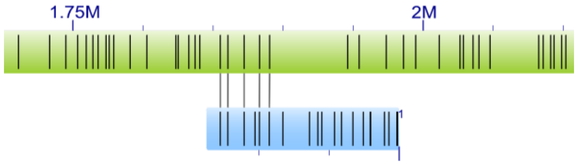


140,150 bp

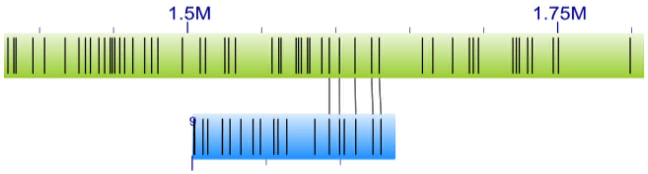
genome map 99
conf.: 19.68



genome map 50
conf.: 6.49

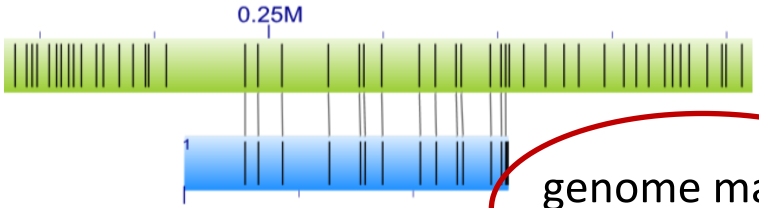


genome map 21
conf.: 6.17



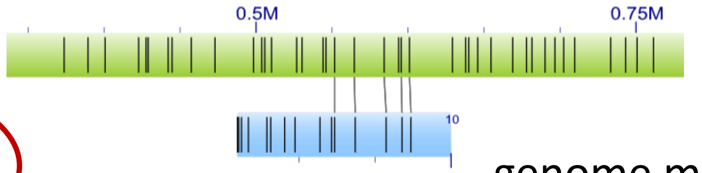
genome map 12
conf.: 6.13

BAC 10



143,891 bp

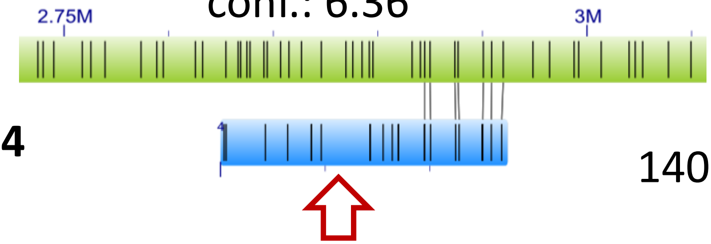
genome map 140
conf.: 16.69



genome map 3
conf.: 6.01

BIONANO MAP FOR IDENTIFYING MISASSEMBLIES

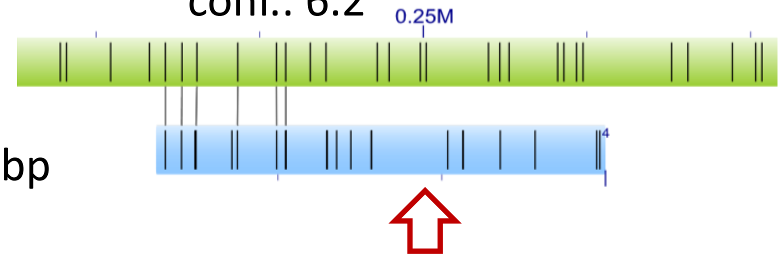
genome map 15
conf.: 6.36



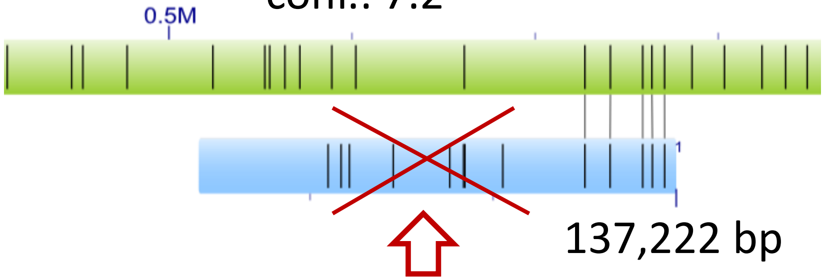
BAC 04

140,129 bp

genome map 249
conf.: 6.2



genome map 46
conf.: 7.2

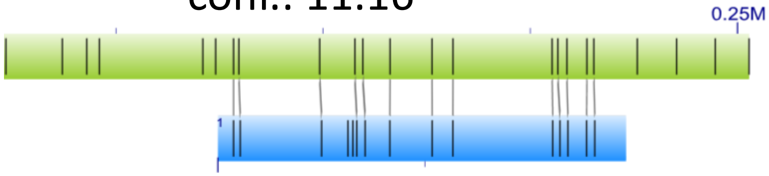


BAC 07

137,222 bp

Overlapping clone from MTP located in a different genome map

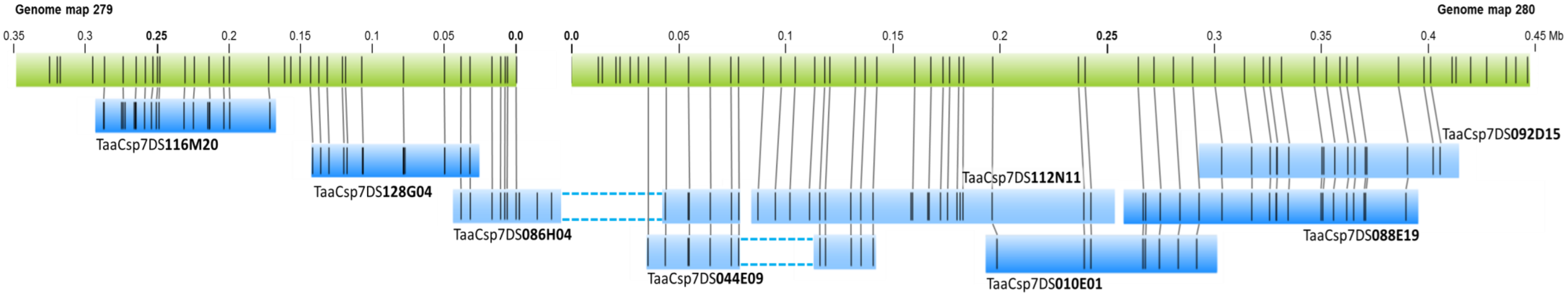
genome map 352
conf.: 11.16



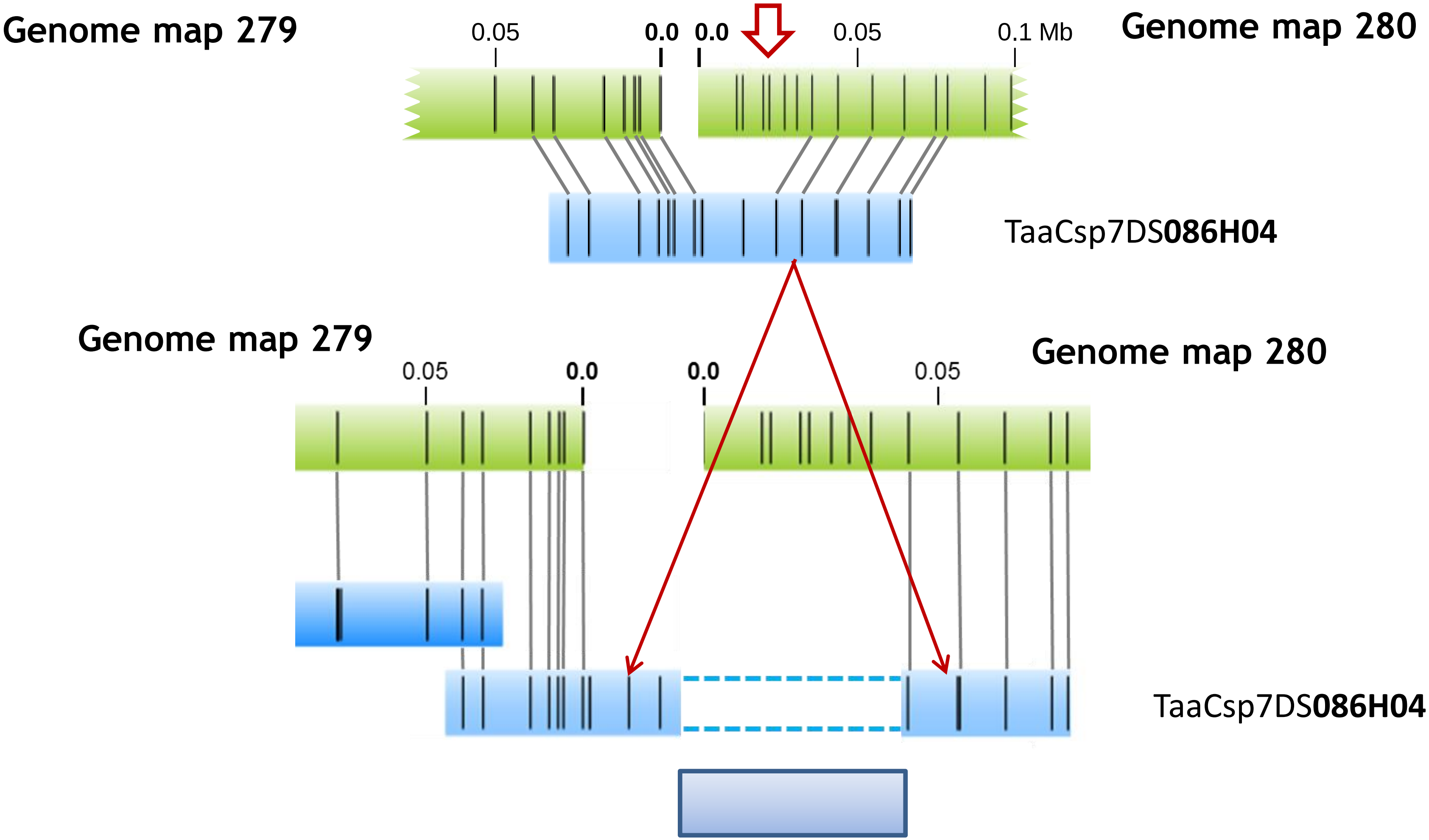
99,861 bp

BAC 03 **No hit → totally misassmbled or missing in the BioNano map**

BIONANO MAP FOR IDENTIFYING AND CORRECTING MISASSEMBLIES

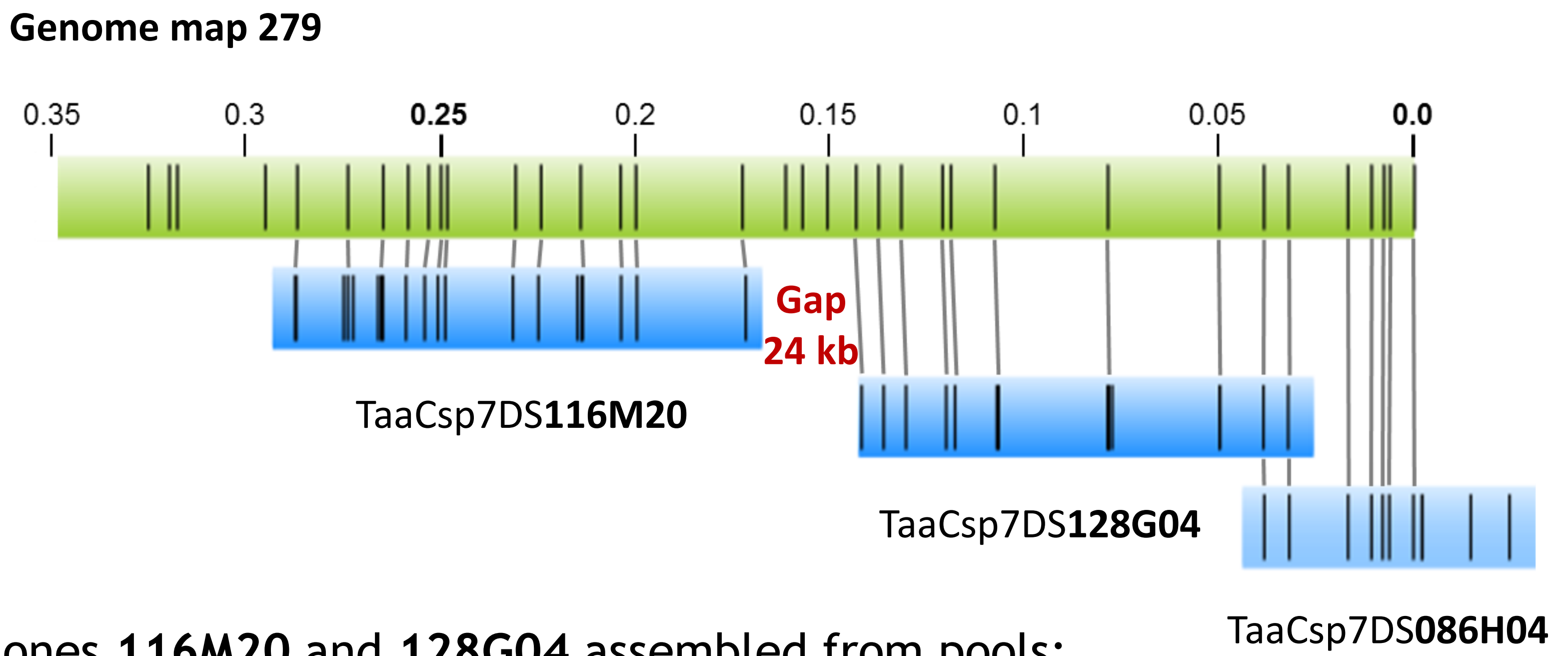


BIONANO MAP FOR IDENTIFYING AND CORRECTING MISASSEMBLIES



7DS086H04-Ctg2 – match in size (19 kb)
mismatch in sequence

BioNano Map for Pool Deconvolution



Genome map 279

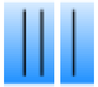
TaaCsp7DS**116M20**

TaaCsp7DS**128G04**

TaaCsp7DS**086H04**

Clones **116M20** and **128G04** assembled from pools:

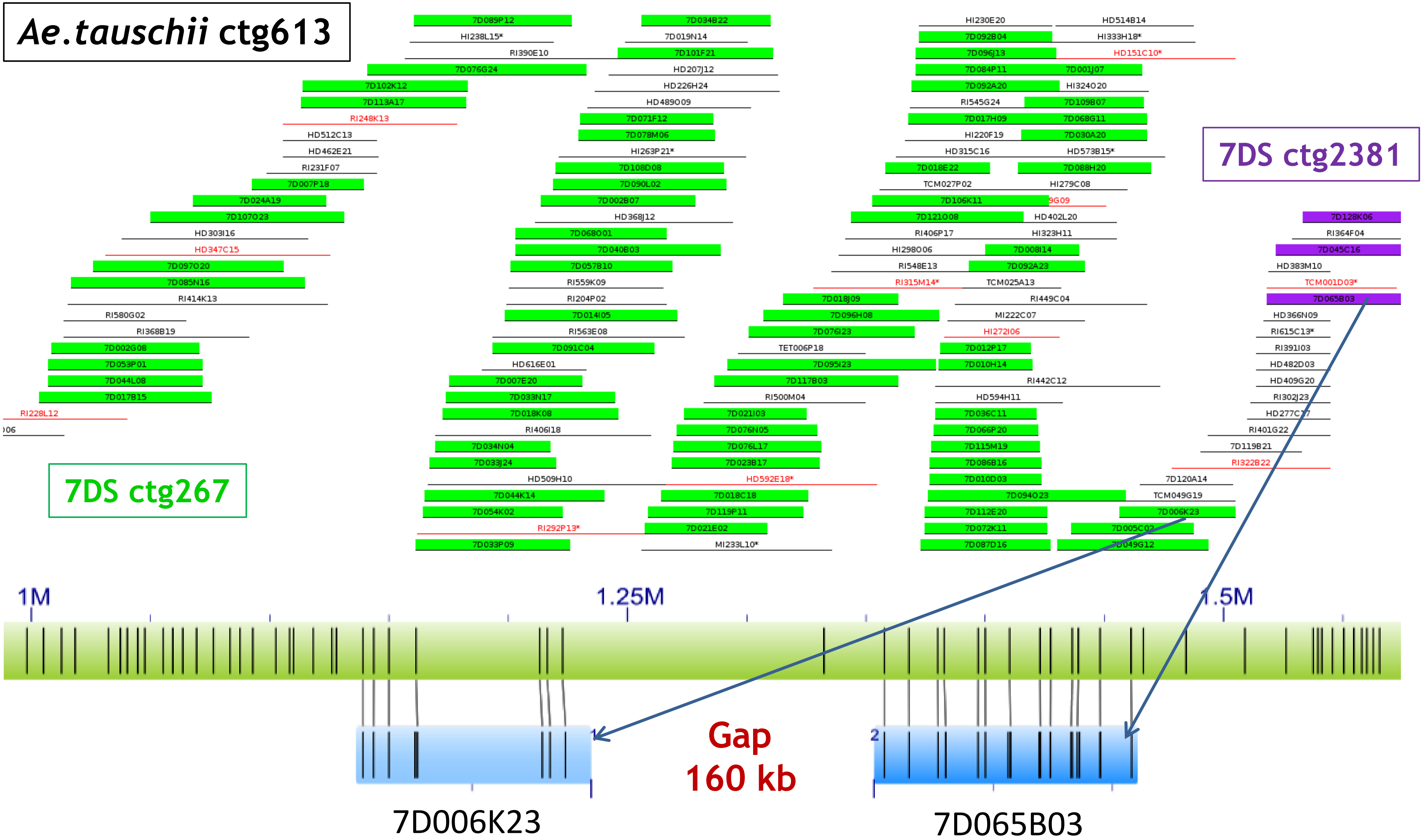
- inner contigs unresolved
- BLASTing the pools against each other indicated two contigs (16 and 8 kb) shared between the pools



- they match the size of the gap and comprise the recognition sites predicted from the genome map

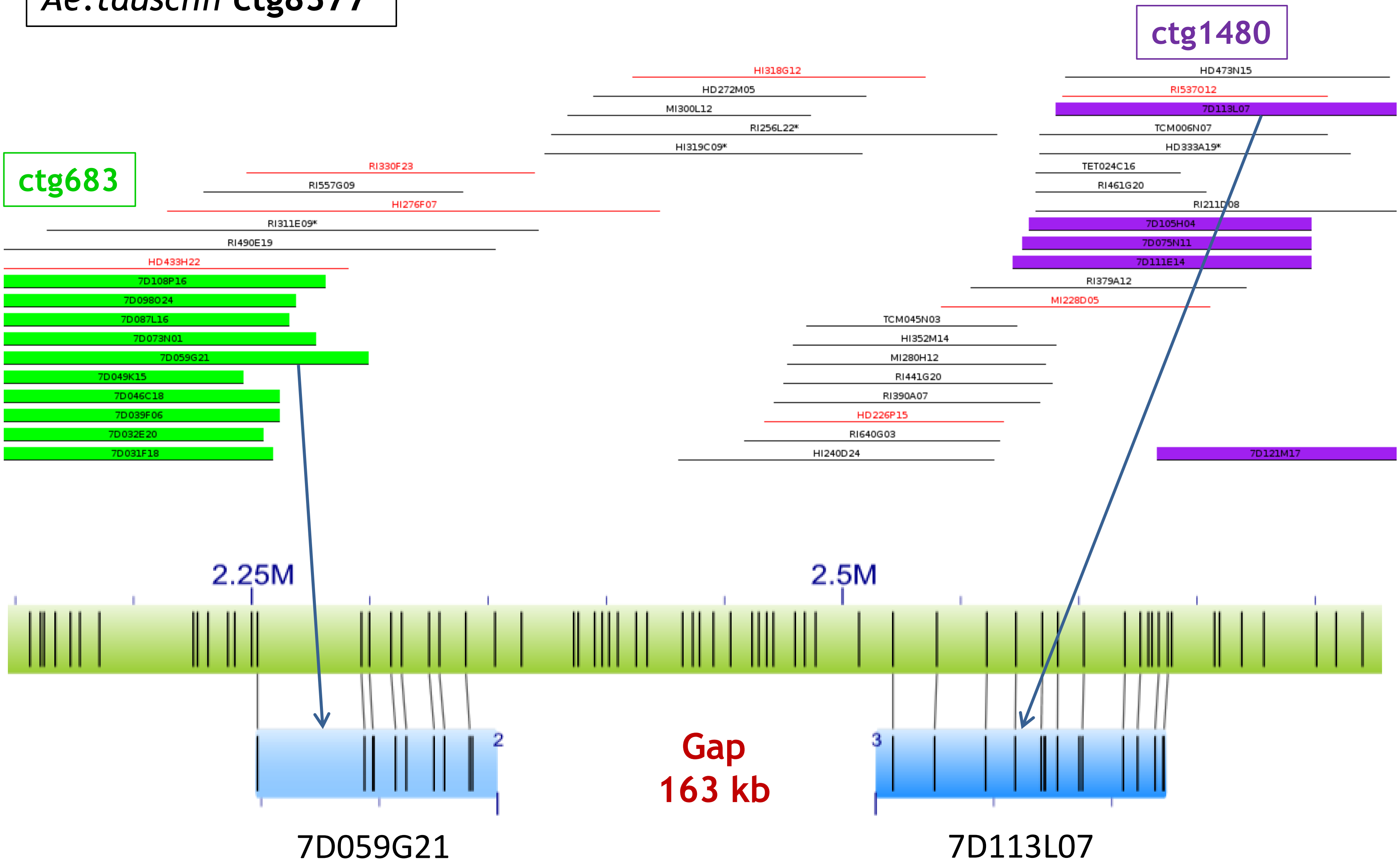
BIONANO MAP FOR PHYSICAL MAP IMPROVEMENT

Co-assembly of 7DS with *Ae. tauschii*



BIONANO MAP FOR PHYSICAL MAP IMPROVEMENT

***Ae.tauschii* ctg8377**



CONCLUSIONS

- Coupling chromosome sorting with BioNano technology enables producing quality *de novo* genome maps for particular chromosomes/arms
- Size estimation is very precise (error 0.3%)

The genome map showed useful for

- Studying distribution of large DNA repeats
- Genome sequence assembling (deconvolution of BAC pools, identifying misassemblies, sizing gaps, assembly improvement)
- Improving physical maps, orienting contigs, scaffolding

ACKNOWLEDGEMENTS

IEB

Helena Staňková
Jan Vrána
Marie Kubaláková
Jaroslav Doležel

BioNano Genomics

Alex Hastie
Han Cao

The University of Queensland Australia

David Edwards
Paul Visendi
Jacqueline Batley
Satomi Hayashi

UC Davis University of California

University of California, Davis
Mingcheng Luo

Kansas State University
Bikram Gill
Bernd Friebe

IWGSC

Akademie věd České republiky

GAČR

MŠMT

European Union
European Regional Development Fund
Investing in your future

OP Research and Development for Innovation