Improving Physical Map and Sequence of the Wheat 7DS Arm Through a BioNano Map

Hana Šimková

PAG XXIII, 10-14th January 2015, San Diego

CHALLENGES IN PLANT GENOME SEQUENCING

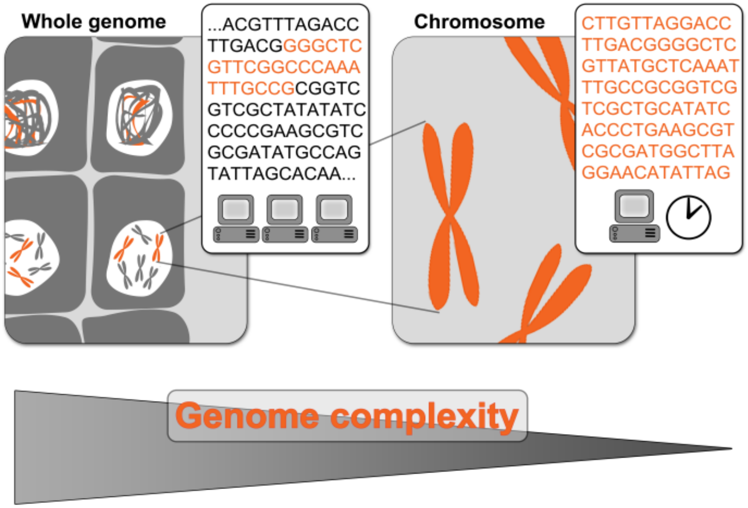
De novo genome assemblies using only short read data of NGS technologies are generally incomplete and highly fragmented due to

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

- Large genome size (~17 Gb)
- Polyploidy (AABBDD genome)

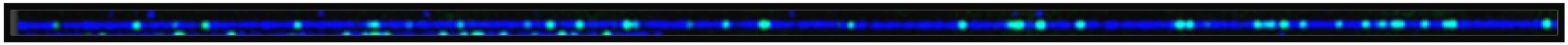
IWGSC

Chromosomal approach



SOLUTIONS FOR THE REPEATS

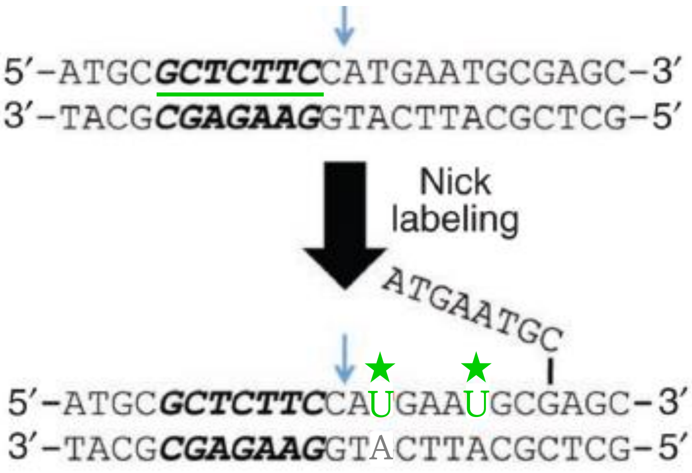
- **Long mate-pair reads**
- **Long read technologies** – PacBio, Oxford Nanopore
- **Optical mapping/genome mapping in nanochannel arrays**
    - Single-molecule mapping of genomic DNA **hundreds of kilobases to several megabases in size**
    - Creates **sequence-motif maps**, which provide long-range template for ordering genomic sequences
    - **Visualisation of reality** “Seeing is Believing”



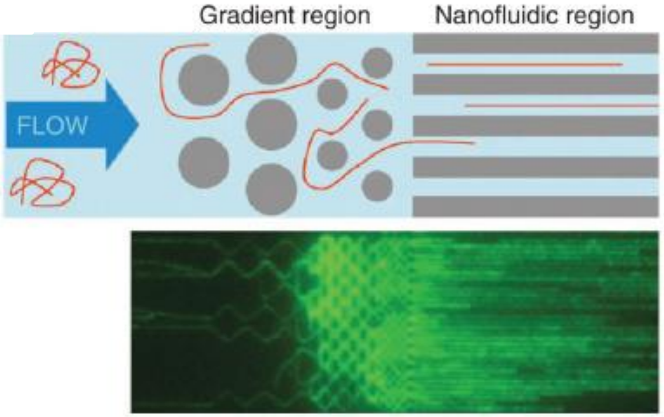
BIONANO GENOME MAPPING ON NANOCHANEL ARRAYS

1 Sequence-specific labeling

Nickase (Nt.BspQI)



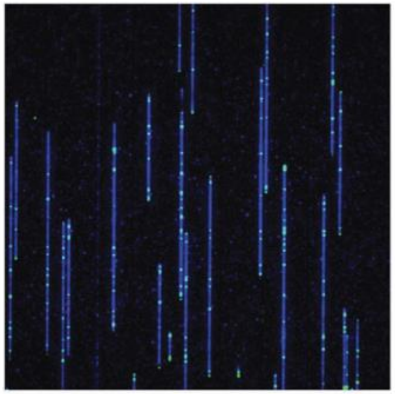
2 DNA linearization



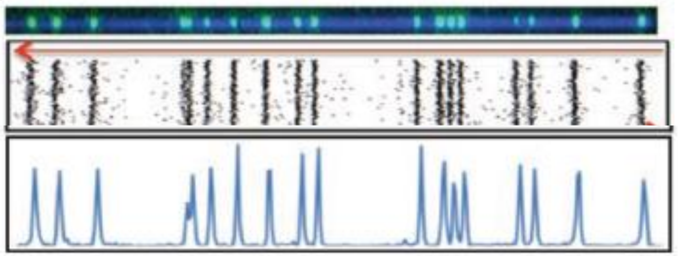
Lam et al., Nat. Biotechnol. 30(8) 2012

Irys

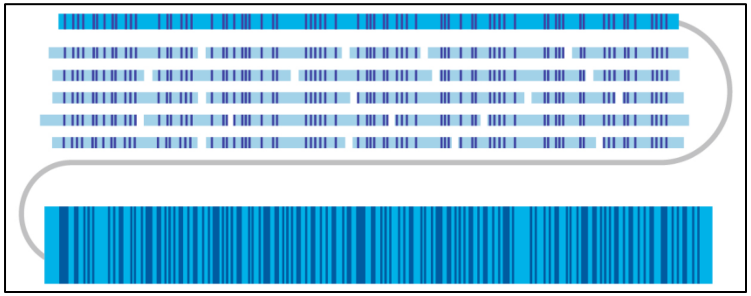
3 Fluorescence imaging



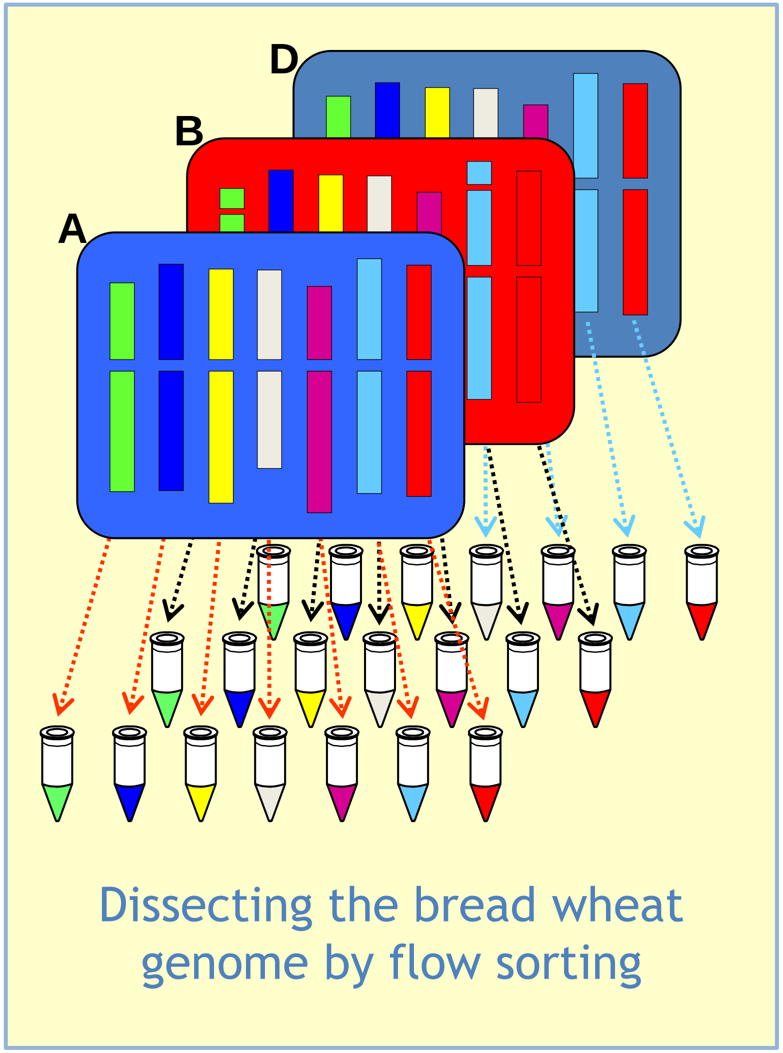
4 Map construction



5 Building consensus map

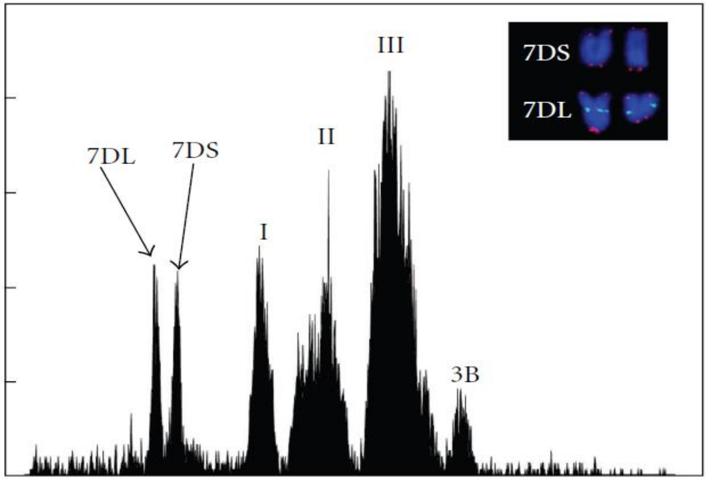


CHROMOSOME MAPPING ON NANOCHANNEL ARRAYS



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

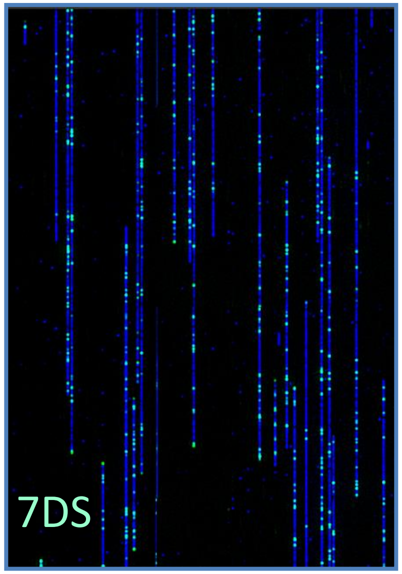


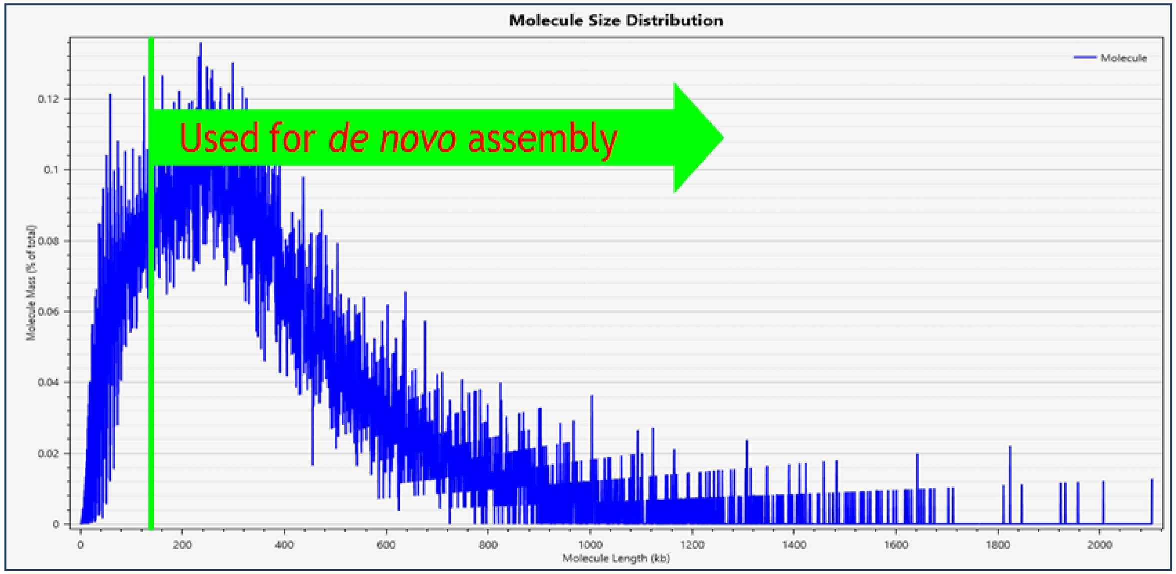
- *In silico* analysis (7DS CSS sequence) for chromosome mapping
  - Nt.BspQI ~13 sites per 100kb
  - Nb.BbvCI ~7 sites per 100kb

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** (GCTCTTC motif)
- Collecting data from **one version-2 chip**

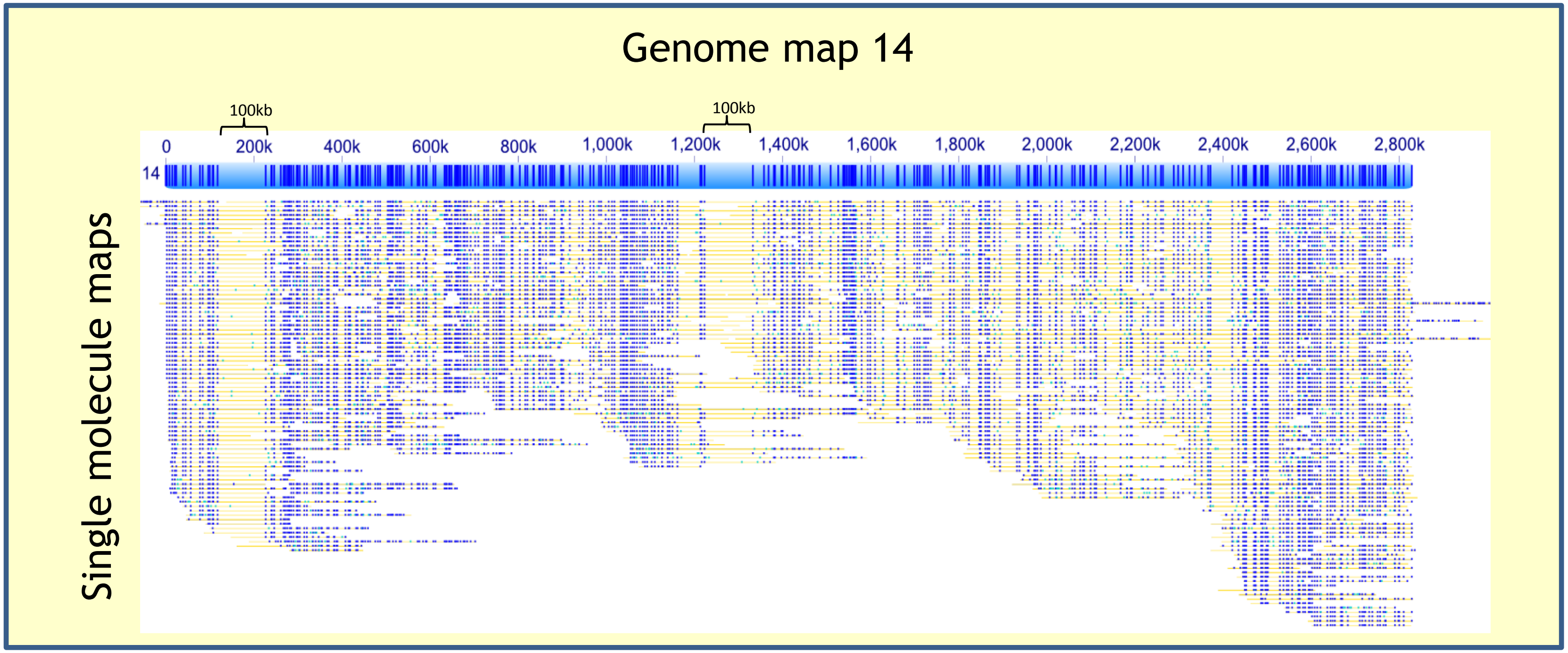
Length treshold	**Total coverage**	**n50**
150kb	200x	344kb





DE NOVO ASSEMBLY OF A 7DS MAP

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


Genome map 14

7DS SEQUENCING STRATEGY

- BAC-by-BAC sequencing based on 7DS physical map, sequencing contigs ≥ 3 BAC clones
- 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

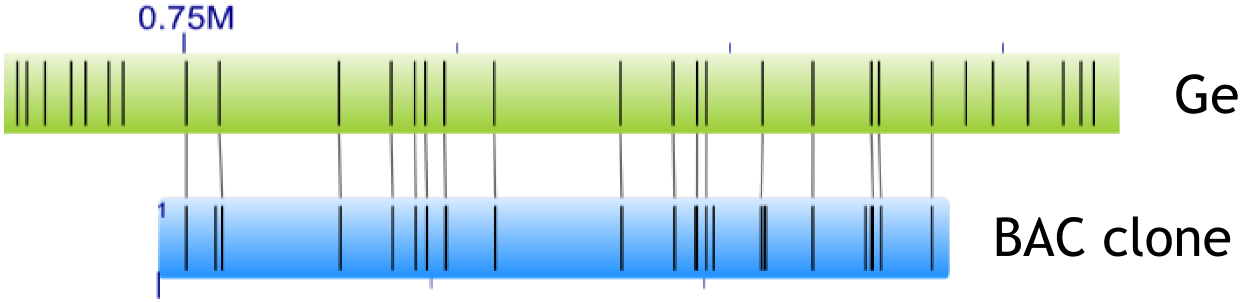
⇩

  - **1-20 contigs per BAC clone, median 3.8**
  - **average contig size 24.3 kb, N50: 65 kb**

- 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 genome map

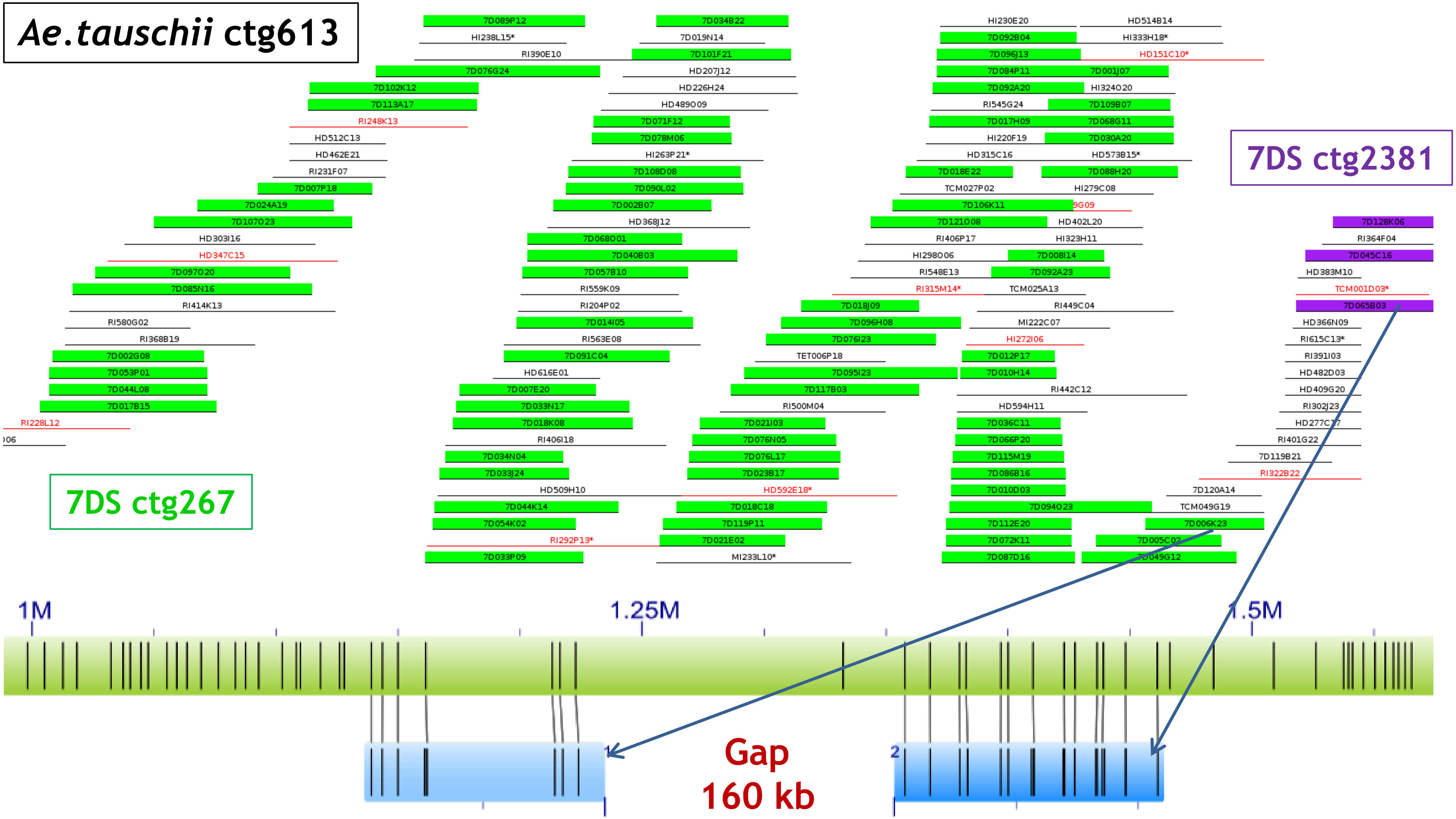
COMBINING BIONANO MAP WITH THE 7DS SEQUENCE

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



BIONANO MAP FOR PHYSICAL MAP IMPROVEMENT

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



***Ae.tauschii* ctg613**

7DS ctg2381

7DS ctg267

Gap
160 kb

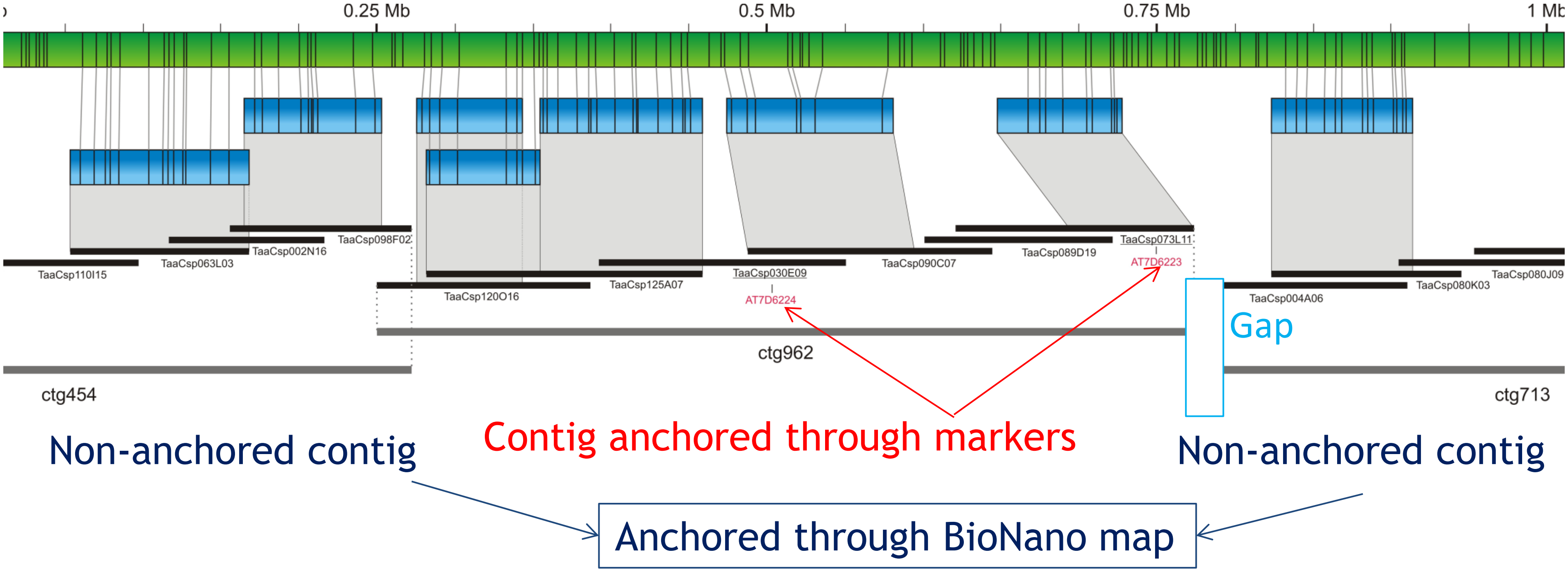
7D006K23

7D065B03

BIONANO MAP FOR PHYSICAL MAP IMPROVEMENT

7DS physical map

- 931 BAC contigs → reduced hitherto to 905 using information from the BioNano map
- 65% anchored through markers
- the rest might be anchored through the BioNano map



Gap

Non-anchored contig

Contig anchored through markers

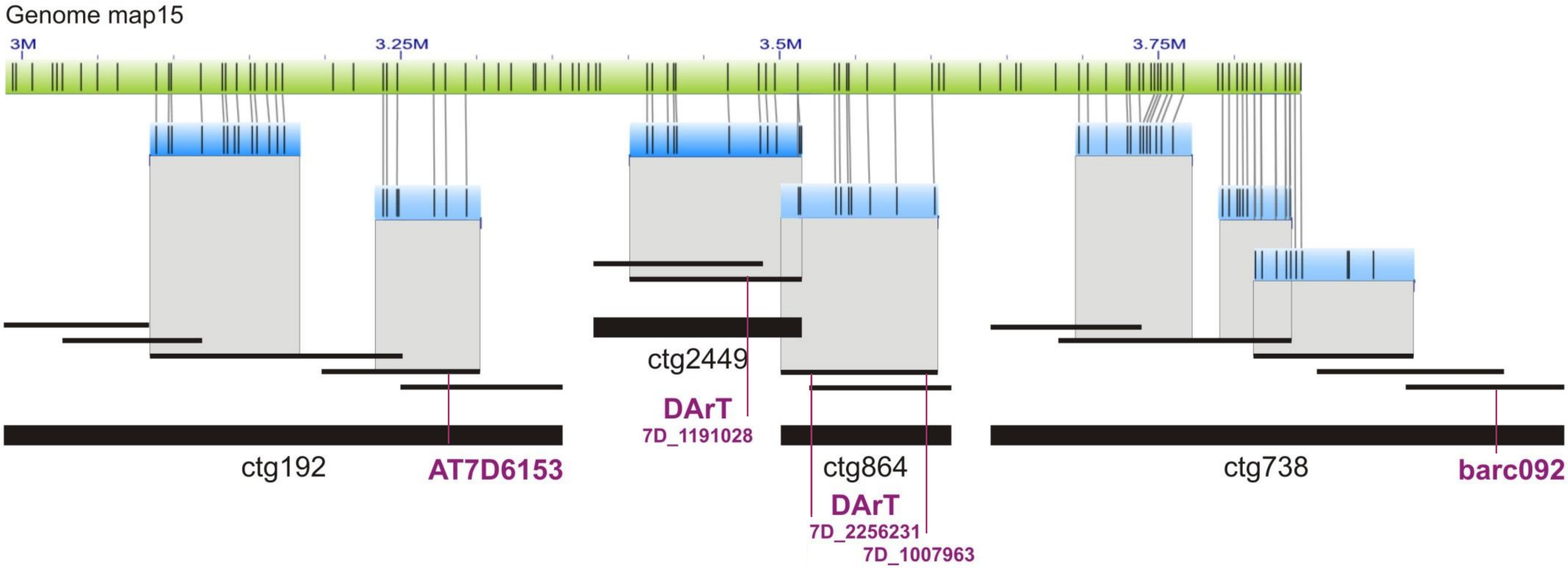
Non-anchored contig

Anchored through BioNano map

BIONANO MAP FOR PHYSICAL MAP IMPROVEMENT

7DS physical map anchored by

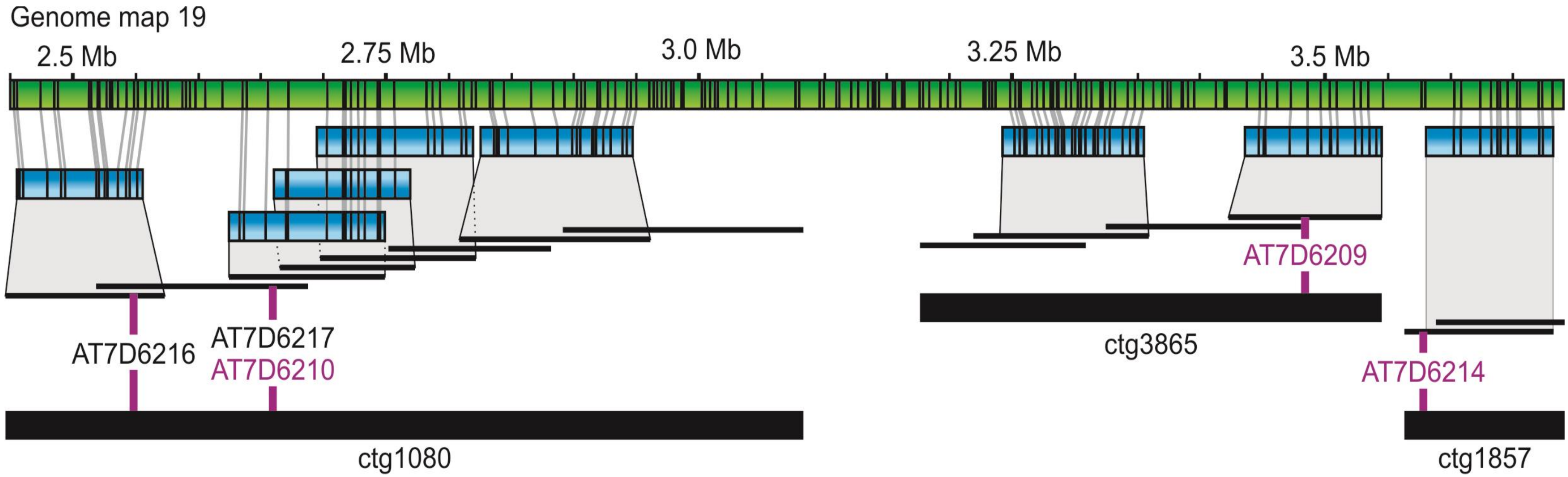
- 583 *Ae. tauschii* SNP markers
- 30 STS markers from *Ae. tauschii* RH map
- 134 SNP markers from CS RH map
- 76 DArT markers
- 23 SSR markers and 11 STS markers



- BioNano map enables integrating various genetic/radiation hybrid maps used for landing BAC contigs on chromosomes

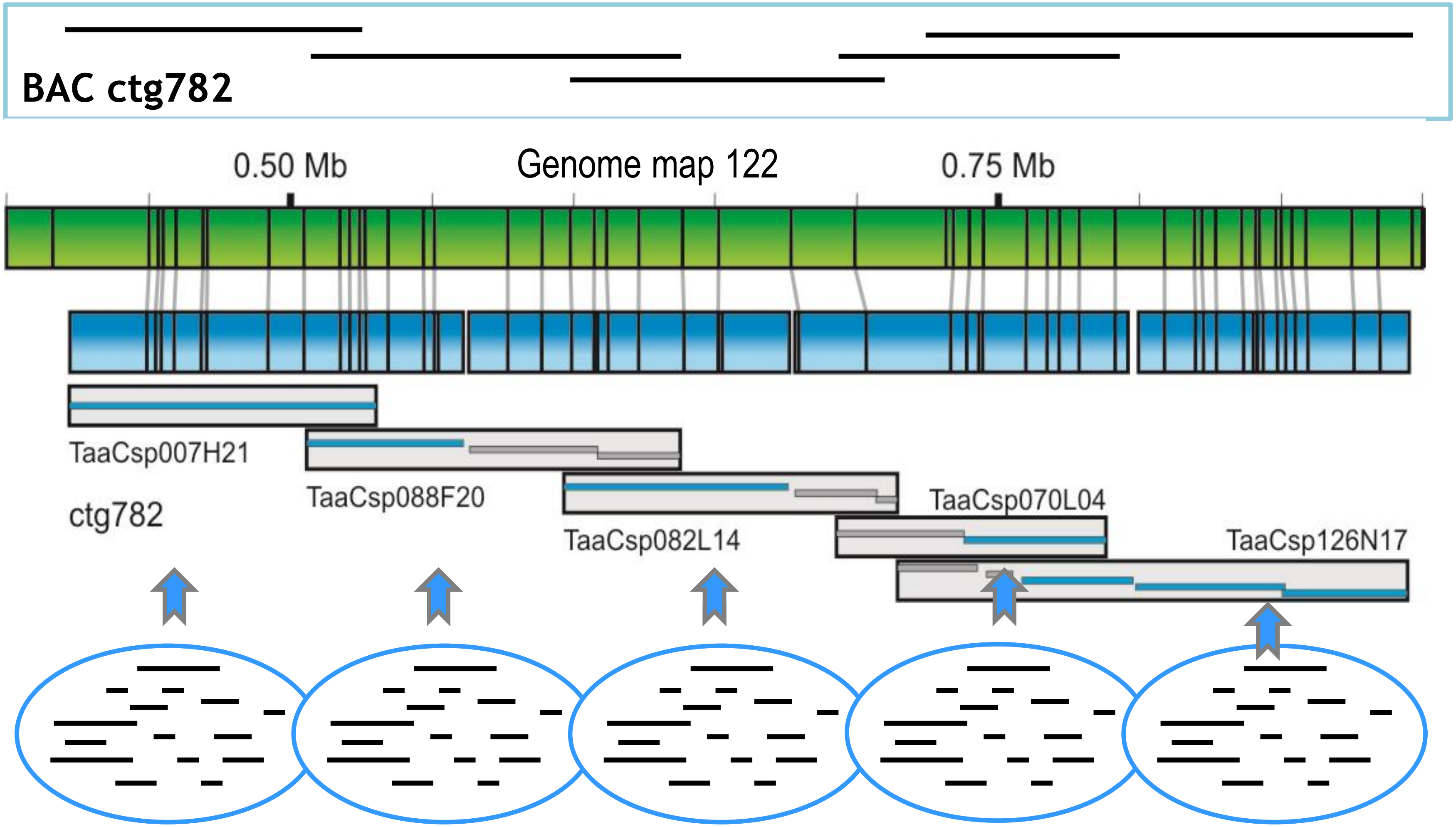
BIONANO MAP FOR PHYSICAL MAP IMPROVEMENT

- Helps ordering contigs in regions with clustering genetic markers



- Genome maps for centromeric region available but are relatively short

BIONANO MAP FOR POOL DECONVOLUTION

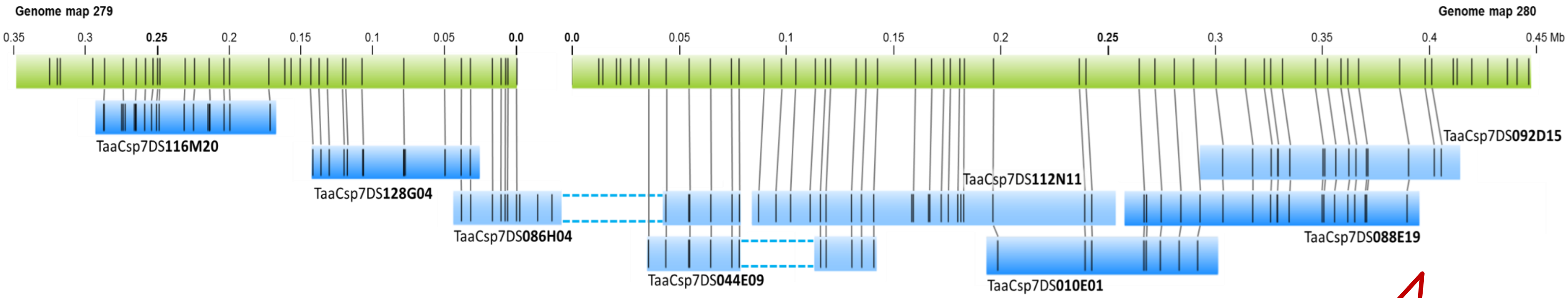


1) Sequences of five BAC pools aligned to the 7DS BioNano map
→ 1-3 contigs per clone anchored to genome map 122

2) The remaining contigs resolved through BLASTing BAC pools against each other

BIONANO MAP FOR IDENTIFYING AND CORRECTING MISASSEMBLIES

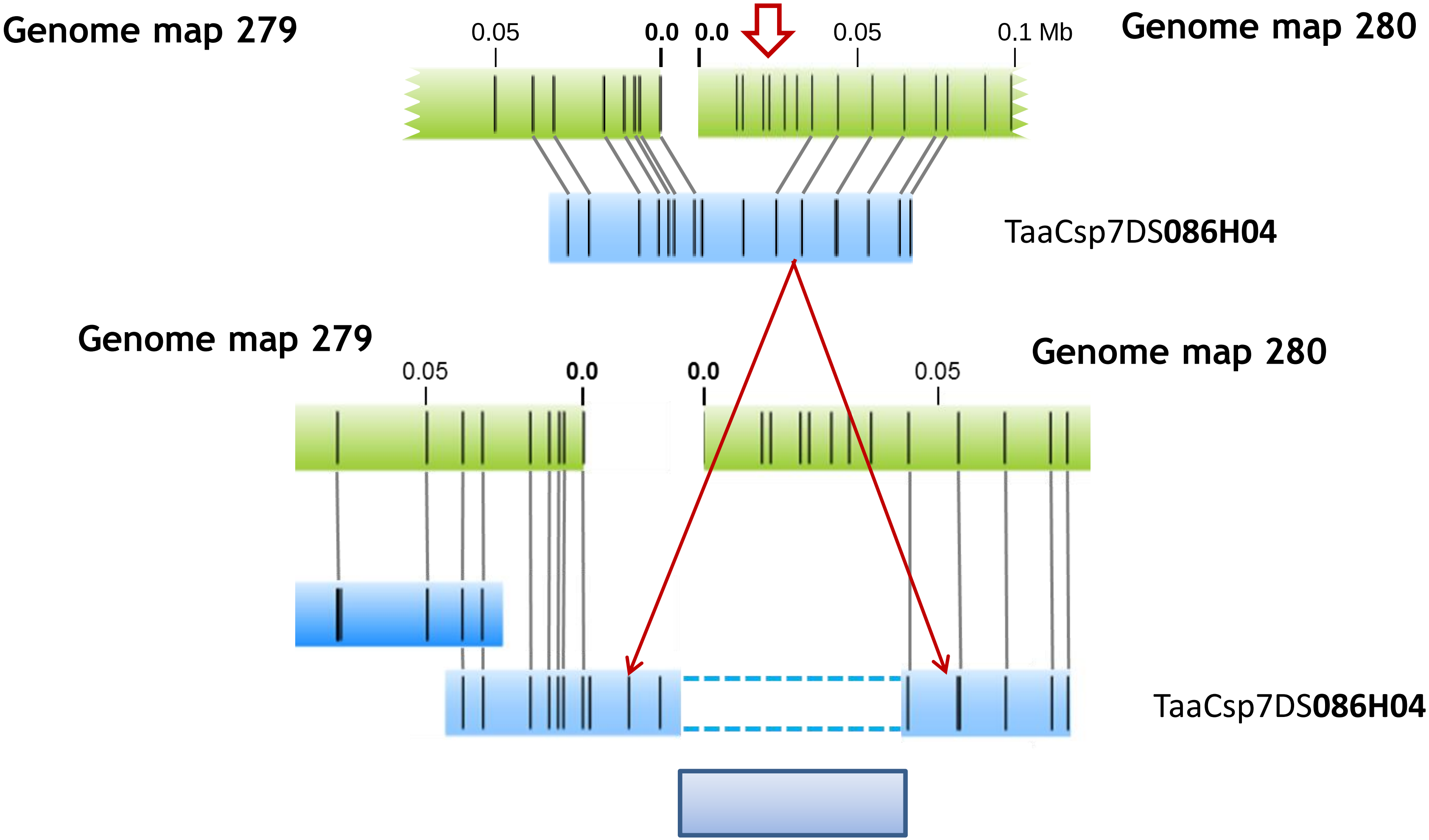
~700 kb



Poster
No. P590

http://olomouc.ueb.cas.cz/

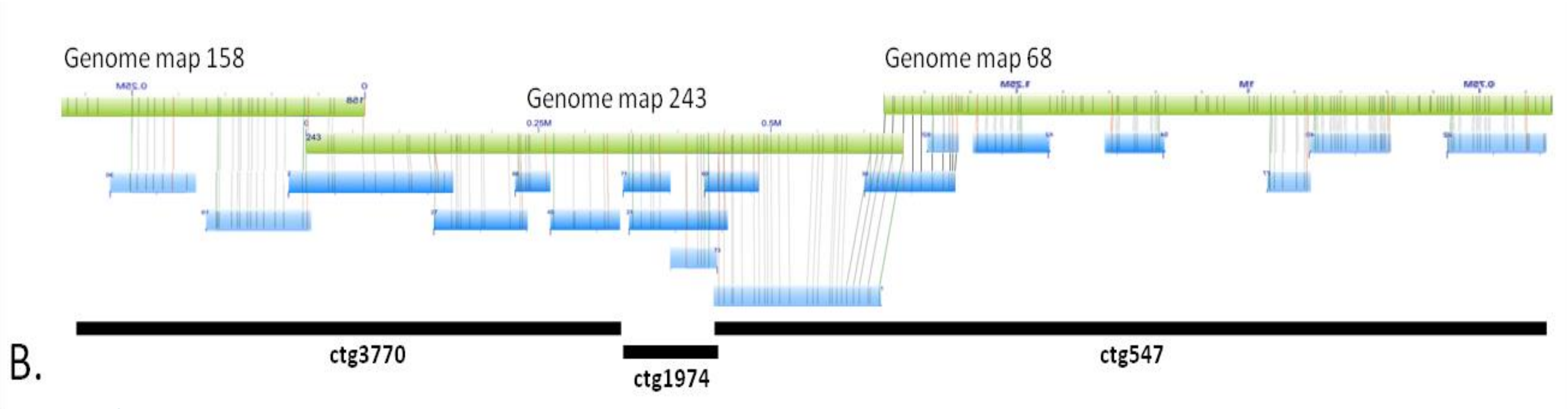
BIONANO MAP FOR IDENTIFYING AND CORRECTING MISASSEMBLIES



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

WHAT HAS NOT BEEN MENTIONED YET…

- Size estimation is very precise (± 3/1000 bp)
- Parametres of the BioNano map can be further improved through alignment with BAC contig map and sequence



BAC clone sequences

Physical map contigs

CONCLUSIONS

- Coupling chromosome sorting with BioNano technology enables producing quality *de novo* genome maps for particular wheat chromosomes/arms

The genome map shows big potential for

- Improving physical maps - anchoring and orientating BAC contigs, scaffolding, validation of the assembly. The map reduces # markers needed for anchoring, facilates ordering contigs in non-recombining regions
- Genome sequence assembling – identifying misassemblies, deconvolution of sequence contigs, sizing gaps, assembly improvement, building pseudomolucule

ACKNOWLEDGEMENTS

IEB

Helena Staňková
Zuzana Tulpová
Jan Vrána
Marie Kubaláková
Jaroslav Doležel

BIONANO GENOMICS

Alex Hastie
Saki Chan
Han Cao

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

David Edwards
Paul Visendi
Jacqueline Batley
Satomi Hayashi

UCDAVIS 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