# BioNano Genome Map of Bread Wheat 7DS Arm Supports Sequence Assembly and Analysis

Hana Šimková
Institute of Experimental Botany
Olomouc
Czech Republic

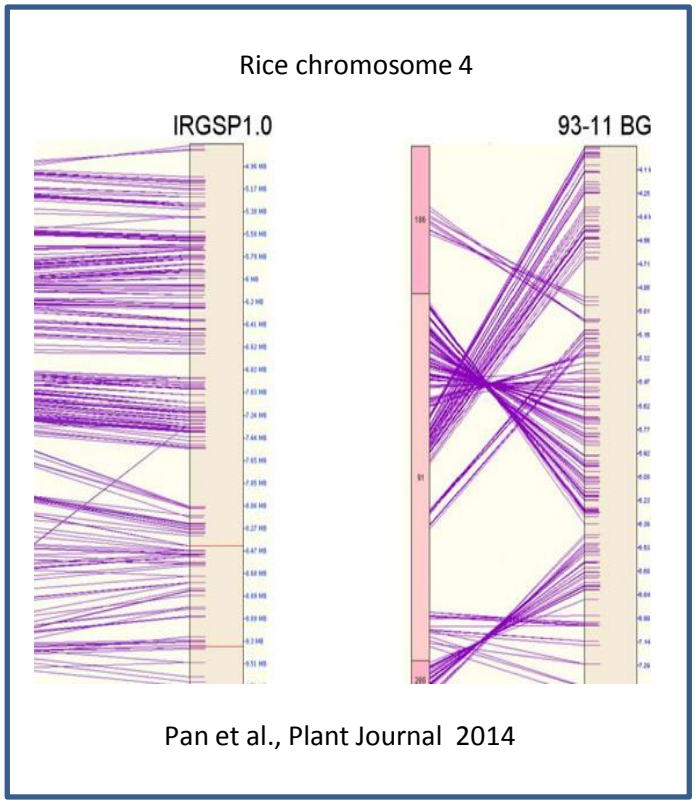
PAG XXIV, IWGSC workshop, San Diego, 9th January 2016

# OUTLINE

- Why BioNano mapping
- Why chromosomal approach

# QUALITY OF GENOME SEQUENCES

- The accepted genome quality has been decreasing
  - The consequence of a massive use of next generation shotgun sequencing
- The bottleneck:
  - NOT the sequencing capacity
  - BUT the ability to assemble many short reads with prevalence of repeated DNA (… and polyploidy)
- Promising:
  - long read technologies (PacBio, Oxford Nanopore, …)



Pan et al., Plant Journal 2014

- BUT read lengths exceeding 200 kb will be needed to unpick long arrays of repeats and segmental duplications

Marx, Nature 501:263, 2013
Shangguan et al., PLOS One 8:e69890, 2013
Pan et al., Plant J. 77 (5): 795-805, 2014
Chaisson et al., Nature Rev. Genet. 16: 627-640, 2015

# How To Make Next Generation Sequences Better?

## Use next generation mapping!

- **Single-molecule mapping** of genomic DNA hundreds to thousands kilobases in size
- Creates **sequence-motif maps**, which provide long-range template for ordering genomic sequences
- **Visualisation of reality**

# GENOME MAPPING ON NANOCHANEL ARRAYS

BIONANO GENOMICS

## Nick labelling

5´-ATGC**GCTCTTC**CATGAATGCG-3´
3´-TACG**CGAGAAG**GTACTTACGC-5´

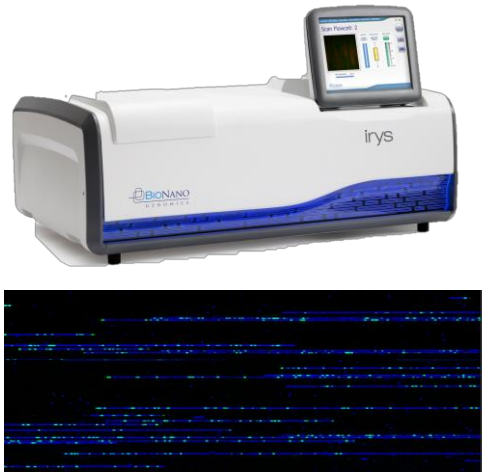
Nt.BspQ1 nicking site

ATGAATGC

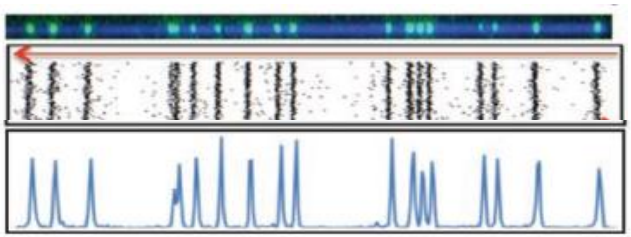
5´-ATGC**GCTCTTC**CAUGAAUGCG-3´
3´-TACG**CGAGAAG**GTACTTACGC-5´

## DNA linearization

## Fluorescence imaging

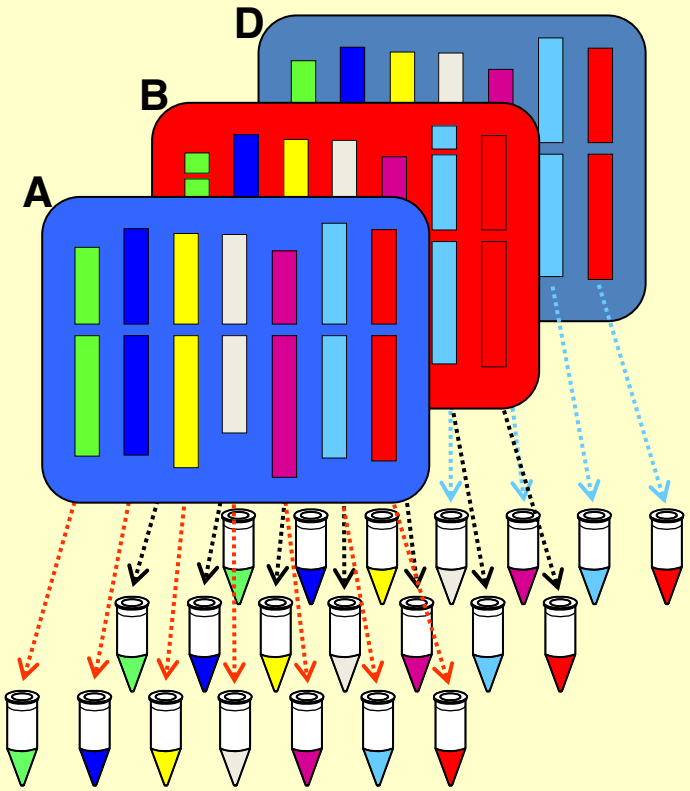



## Single molecule maps



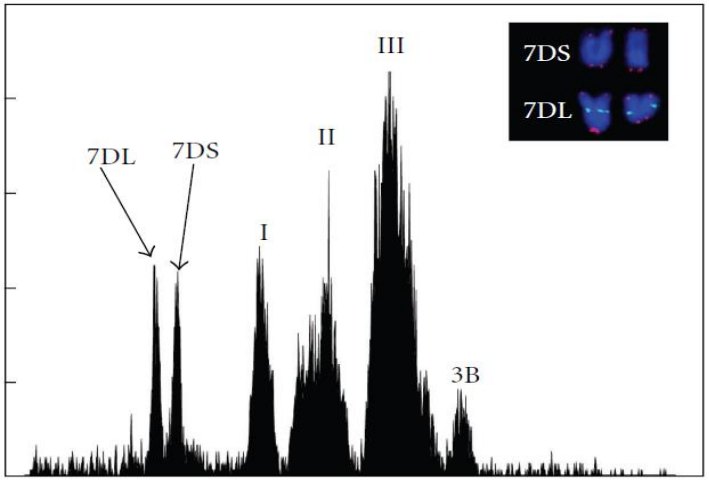
## Consensus genome maps

# Chromosome Mapping on Nanochannel Arrays




Dissecting the bread wheat genome by flow sorting

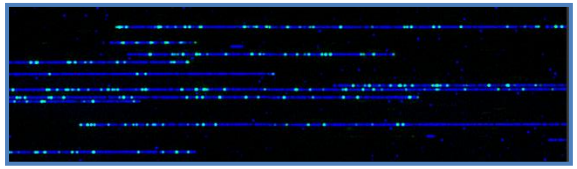
- Pilot study on wheat **7DS** chromosome arm (381 Mb, 2.25% wheat genome)
  - Purified as telocentric chromosome by flow cytometric sorting




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

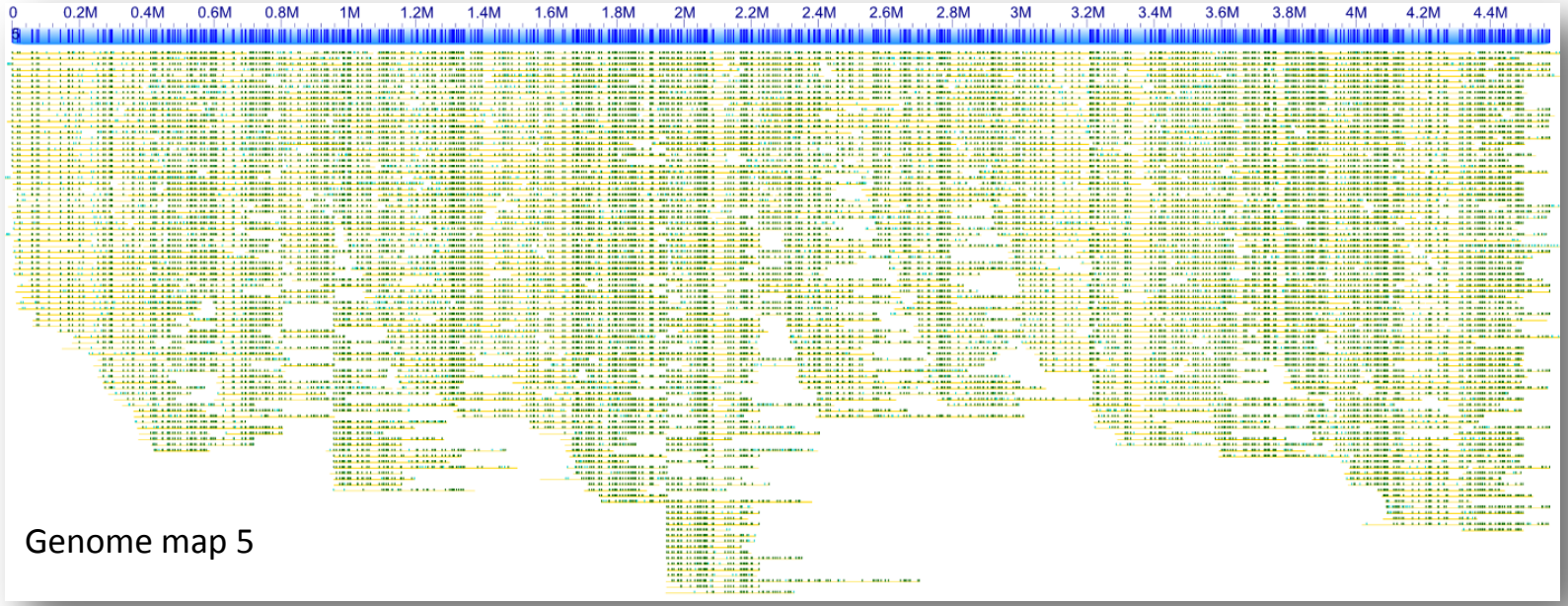
# *De Novo* Assembly of 7DS BioNano Map

- 1.6 million flow-sorted 7DS arms of 'Chinese Spring'



| Length threshold | Total coverage | Molecule N50 |
|---|---|---|
| 150kb | 180x | 344kb |

- A total of **371 genome maps** were *de novo* assembled
- Total assembly length 350 Mb (**92% of estimated 7DS size**)
- **Average map size 0.9 Mb**
- **Map N50 is 1.3 Mb**



# Co-assembly of 7DS and 7AS Data

| | 7AS | 7DS | Theoretical 7AS + 7DS | Co-assembly (% vs. theoretical) |
|---|---|---|---|---|
| **Molecule Coverage** | 194x | 206x | | |
| **Molecule N50 (kb)** | 206 | 219 | | |
| **No. Genome Maps** | 783 | 468 | 1251 | 2043 (163%) |
| **Avg. Genome Map Length (kb)** | 571 | 765 | 0,643 | 0,488 (76%) |
| **Genome Map N50 (kb)** | 1,553 | 1,355 | 1,454 | 1,053 (70%) |

- More fragmented genome maps
- Several percentage of chimeric maps – merged through shared repeats

# 7DS Physical Map and Sequence

## Physical map

- Automatic assembly using FPC
- Manual end-merging of contigs based on integration of *Aegilops tauschii* whole-genome map with the map of the bread wheat **7DS** chromosome arm
- Verification of the assembly using LTC

| | |
|---|---|
| No. contigs > 2 BAC clones | 904 |
| No. contigs > 5 clones | 652 |
| Assembly length | 360 Mb |
| 7DS arm coverage | 95% |
| Contig N50 | 548 kb |
| No. MTP BAC clones | 4,608 |

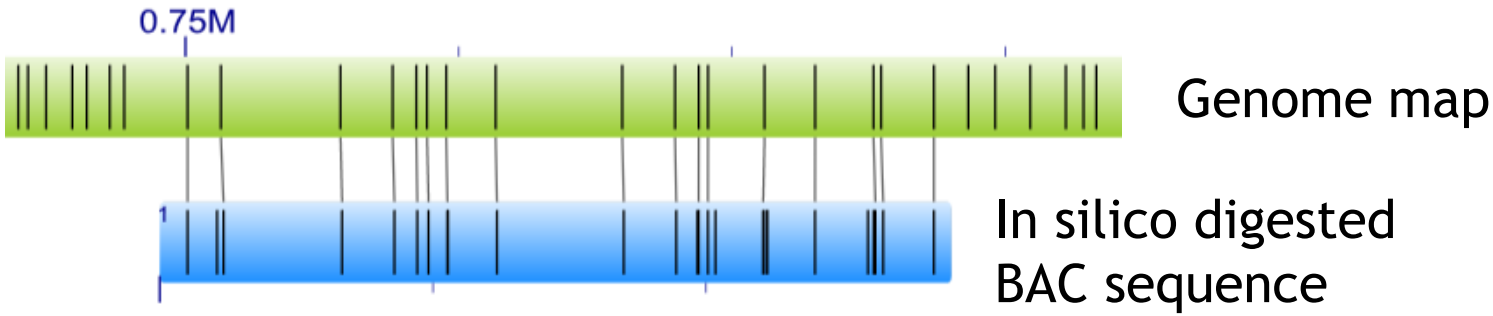
## Sequence

- Sequencing pools of four non-overlapping MTP BAC clones
- Illumina pair-end sequencing - 550bp fragment size, 100 bp read length, coverage >500x → Sassy
- Deconvolution through BAC end sequences
- Scaffolding through mate-pair data obtained from MTP-plate pools (384 clones/pool) → SSPACE

| | |
|---|---|
| Contigs per BAC | 1 - 17 |
| Average No. contigs/BAC | 1.9 |
| Median No. contigs/BAC | 1.5 |
| Average scaffold size | 56 kb |
| Scaffold N50 | 116 kb |
| % N in assembly | 2.9 % |

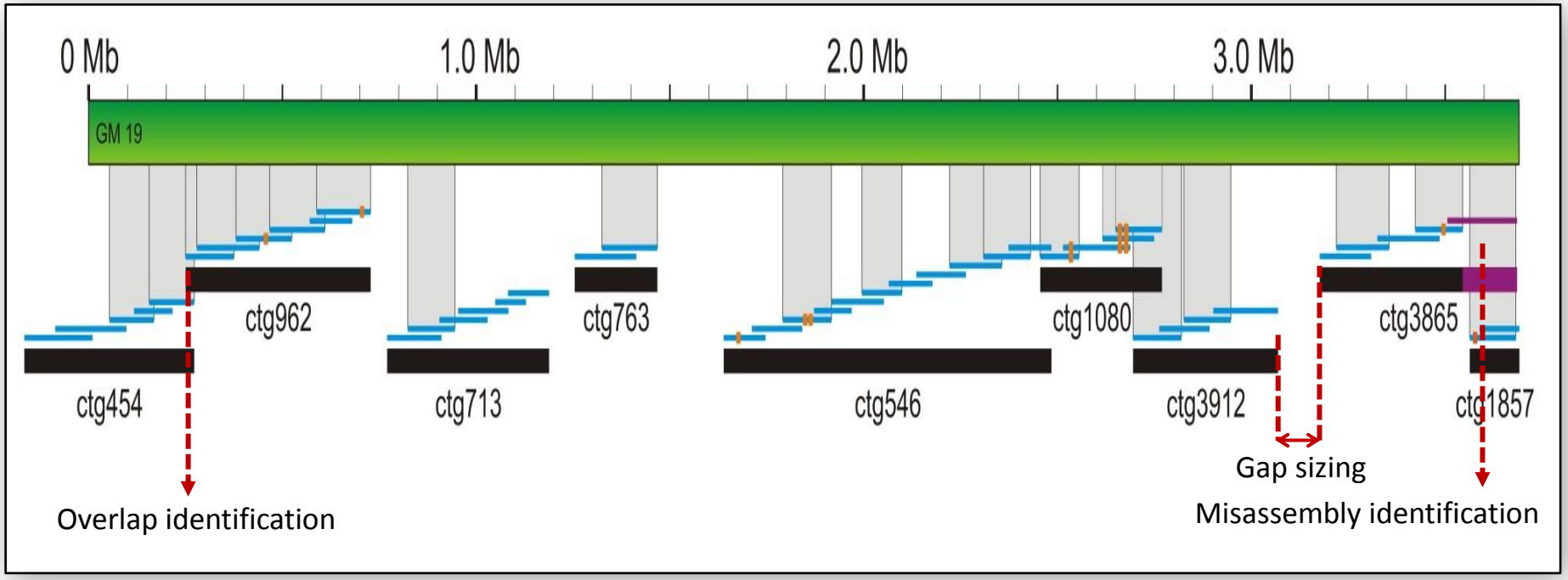
# COMBINING BIONANO MAP WITH THE 7DS SEQUENCE

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



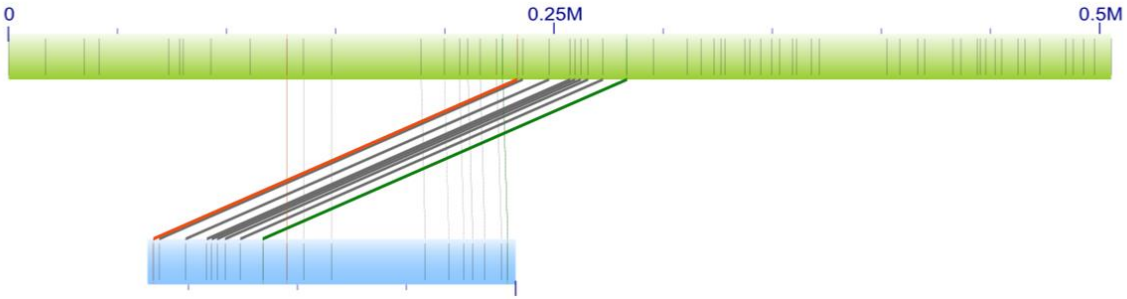
# BioNano Map for Physical Map Improvement

- Enables scaffolding of BAC contigs
  - Reduces demand on number of markers
  - Integrates multiple genetic maps
  - Useful in non-recombing regions
- Sizes gaps
- Identifies and corrects mis-assemblies

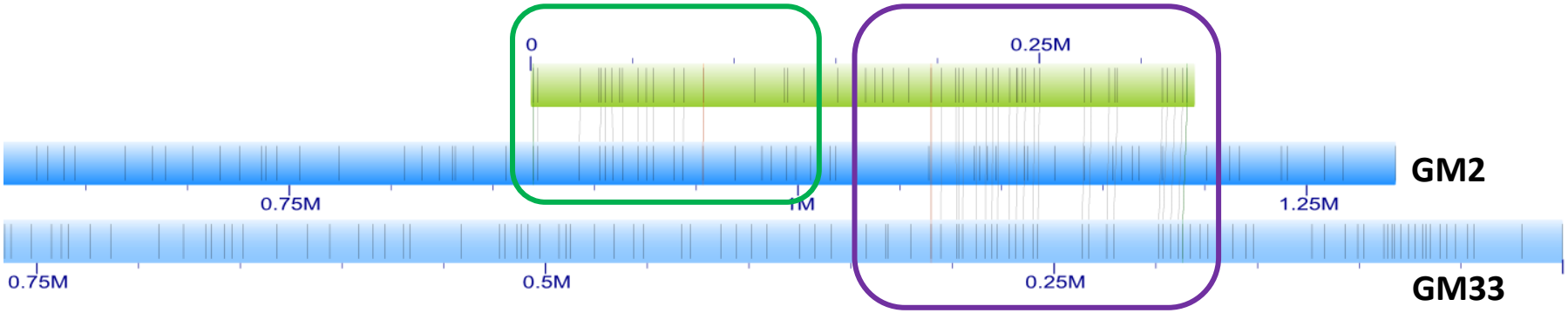


# BioNano Map for Sequence Improvement

- Identification of scaffold mis-assemblies
  - Due to mis-orientation contigs in a scaffold

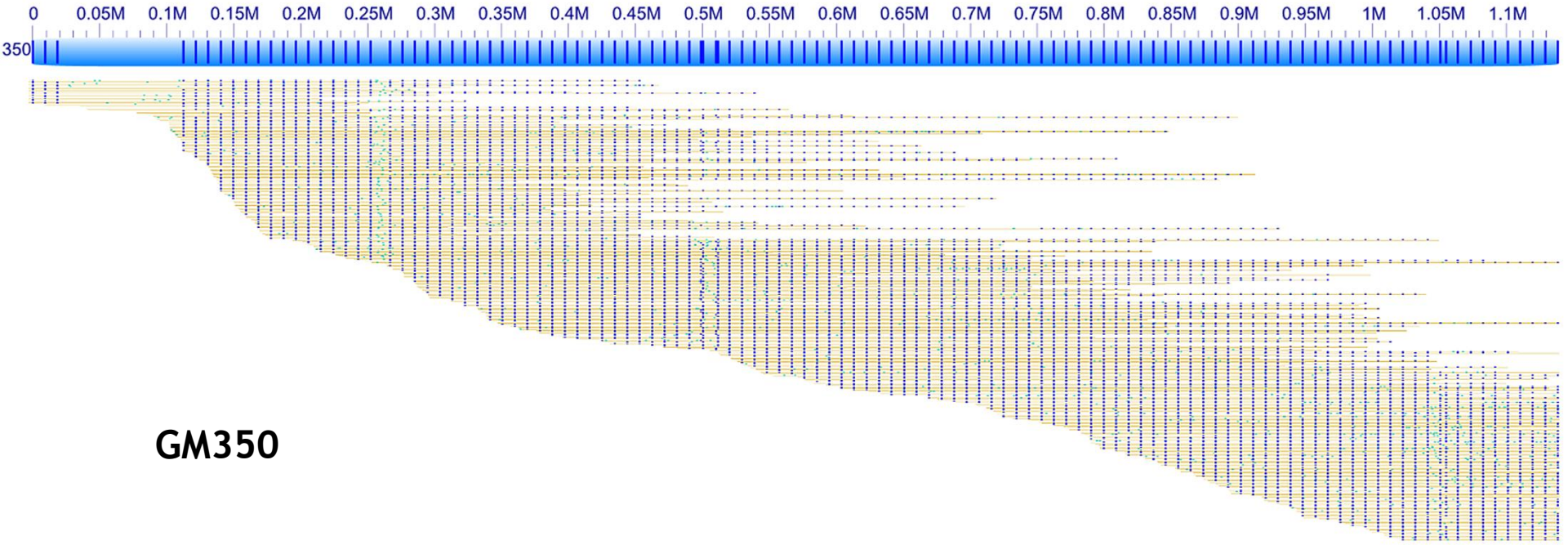



  - Due to mis-joining of contigs within a sequencing pool or cross-contamination



# Large Array of Tandem Repeats in the 7DS

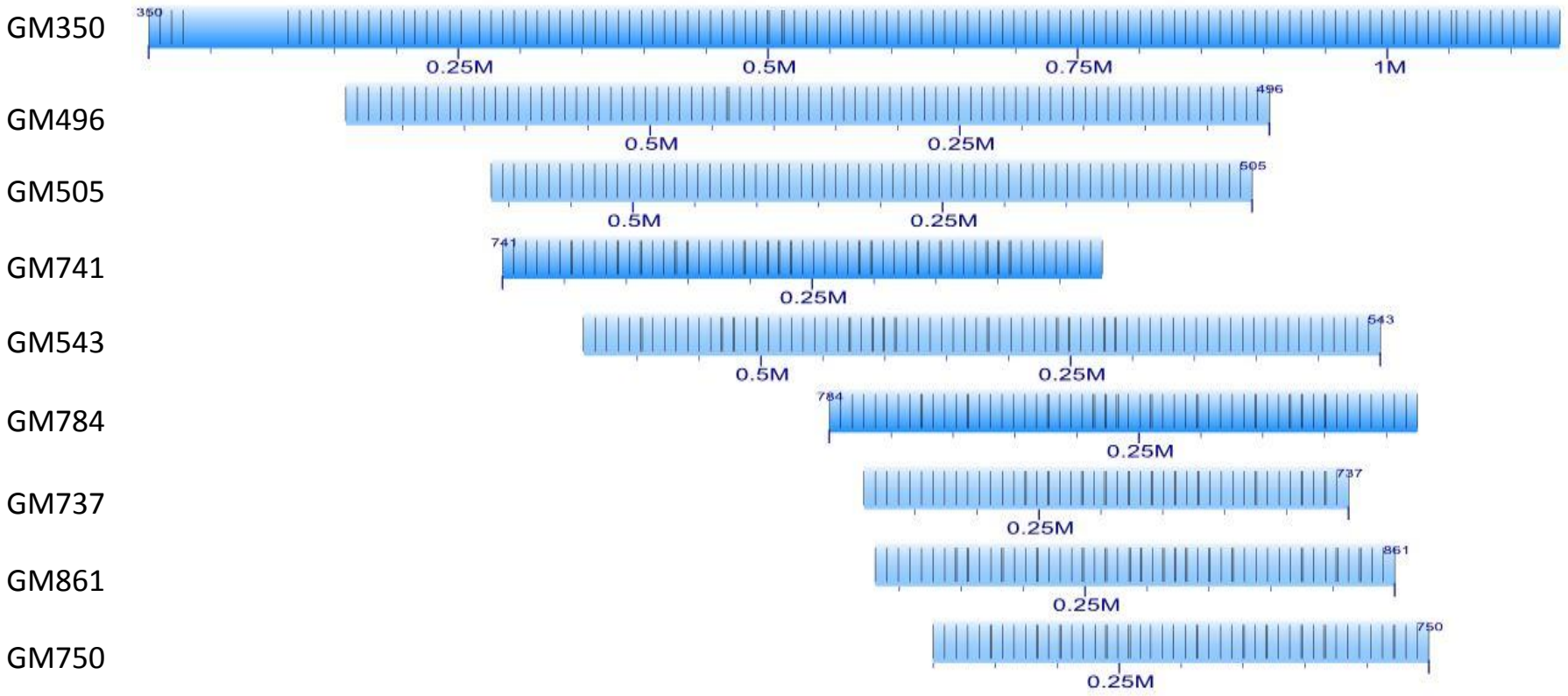
- ~1 Mb-long array of tandem repeats was identified in the genome map No. 350 (longest single molecule ~ 800 kb)
- Unit size ~9.3 kb




**GM350**

# Large Arrays of Tandem Repeats

+ several other maps with identical/similar motif




No high-confidence match in 7DS BAC sequence assemblies

# LARGE ARRAYS OF TANDEM REPEATS

## rDNA locus?

- Corresponds to the size of wheat 25S-18S rDNA unit: ~9 kb (Gerlach and Bedbrook, 1979)
- Major loci on 1B, 6B and 5D

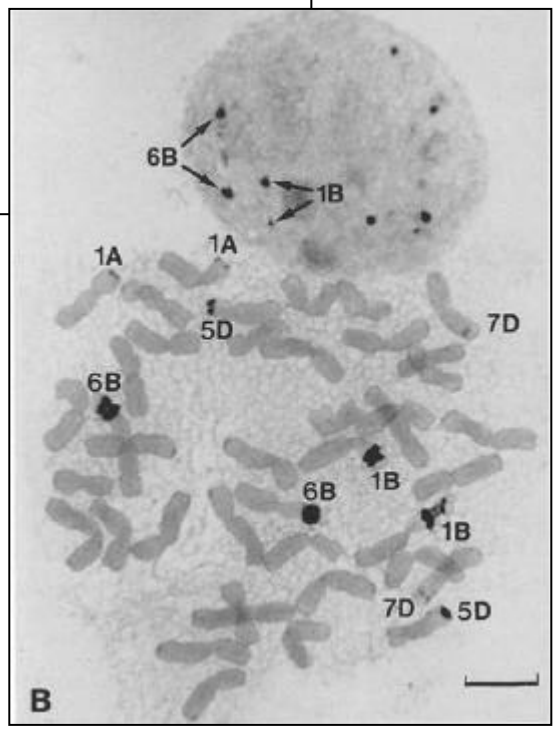
Chromosoma (1991) 100:71–78

CHROMOSOMA
© Springer-Verlag 1991

**Physical mapping of the 18S.26S rRNA multigene family in common wheat: Identification of a new locus**

Y. Mukai*, T.R. Endo**, and B.S. Gill

- Minor locus identified on the 7DS arm



# Studying Local Structural Variation

- Project on cloning a Russian wheat aphid resistance gene *Dn2401*
- Region of 0.81 cM spanned by five BAC clones (~350 kb)

Theor Appl Genet (2015) 128:1373–1383
DOI 10.1007/s00122-015-2512-2

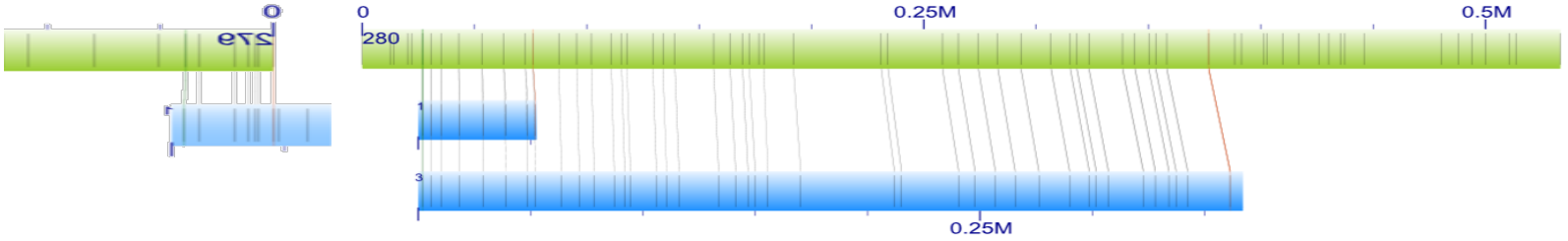
ORIGINAL PAPER

**Chromosomal genomics facilitates fine mapping of a Russian wheat aphid resistance gene**

Helena Staňková[1] · Miroslav Valárik[1] · Nora L. V. Lapitan[2,3] · Paul J. Berkman[4] · Jacqueline Batley[5] · David Edwards[5,6] · Ming-Cheng Luo[7] · Zuzana Tulpová[1] · Marie Kubaláková[1] · Nils Stein[8] · Jaroslav Doležel[1] · Hana Šimková[1]

- BAC clones sequenced, candidate genes identified

CS



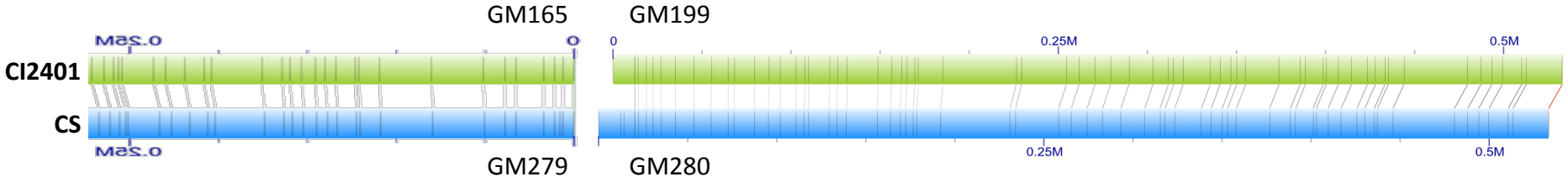

- Sequence available for cv. Chinese Spring (susceptible), while the gene has been mapped in a resistant line CI2401

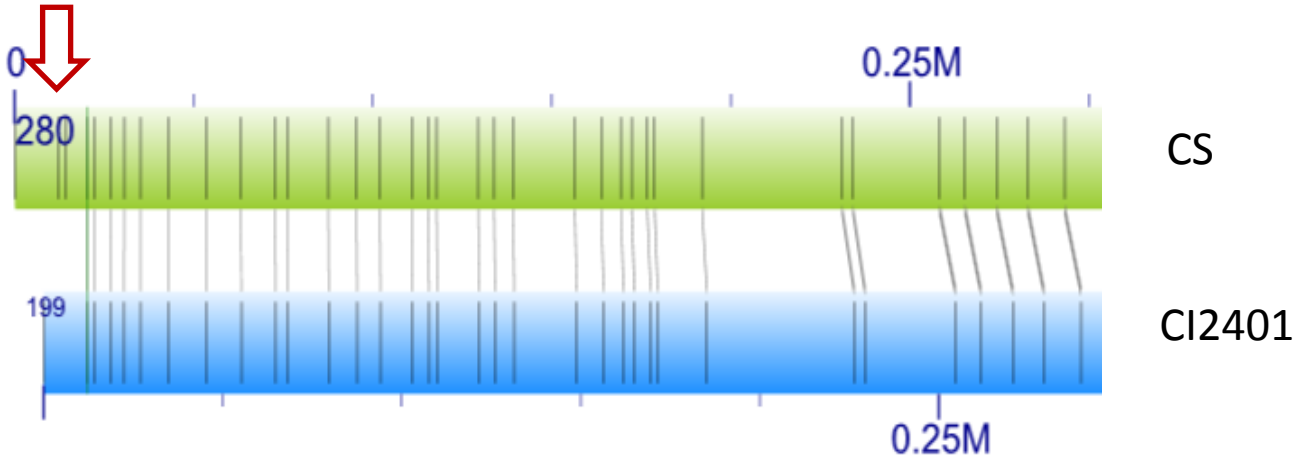
**Do we miss anything when analyzing the CS sequence?**

# Studying Local Structural Variation

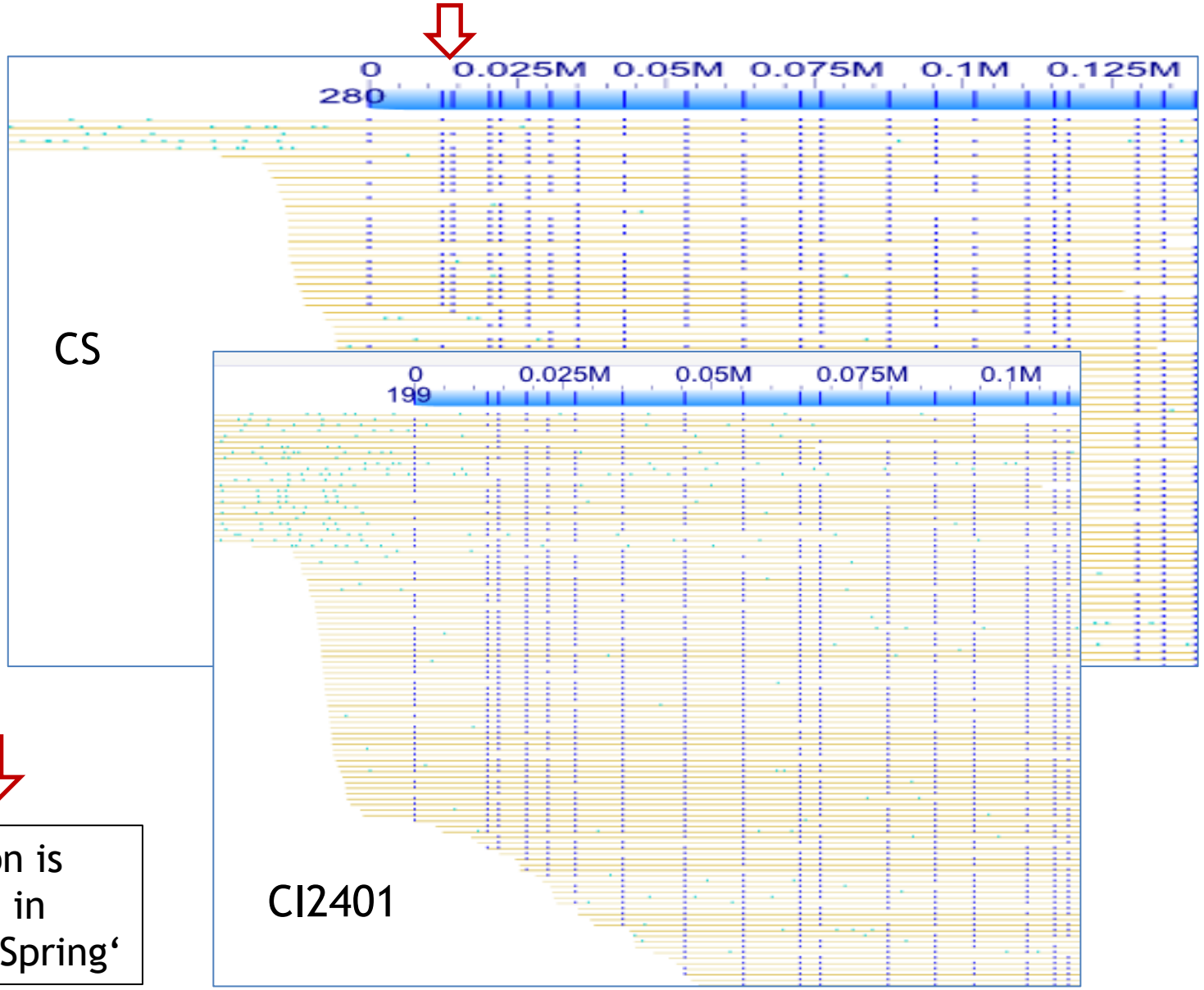
- Construction of BioNano map from 7DS arm of CI2401 (7DS ditelosomic line newly created from CI2401)

| 7DS | CS | CI2401 |
|---|---|---|
| **Molecule Coverage** | 180x | 206x |
| **Molecule N50 (kb)** | 344 | 219 |
| **Genome Map N50 (Mb)** | 1.3 | 1.35 |

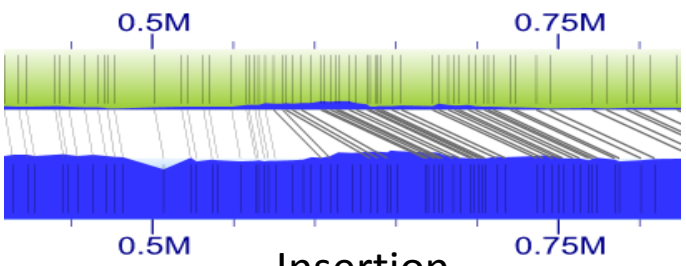


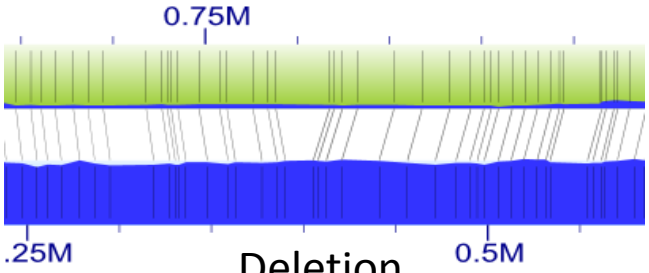
# Studying Local Structural Variation



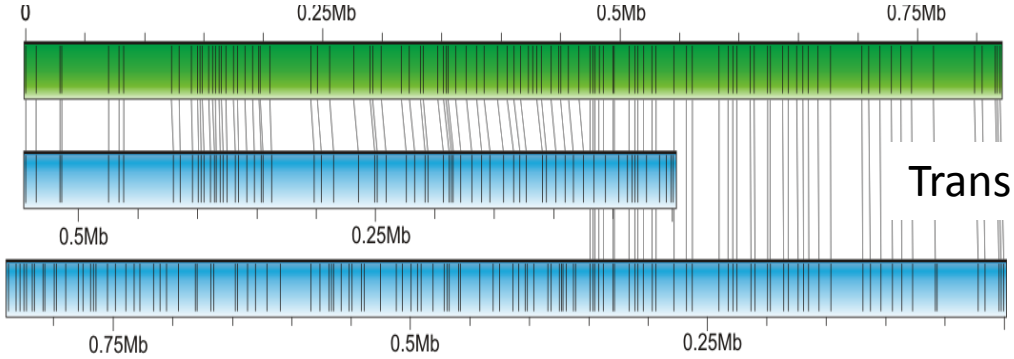
# Studying Structural Variation on Chromosome arm Level




Insertion




Deletion




Translocation

| | No. cases | Total length (% CS arm length) |
|---|---|---|
| Deletions in CI2401 vs. CS | 28 | 1.2 Mbp (0.31%) |
| Insertions in CI2401 vs. CS | 109 | 1.7 Mbp (0.44%) |
| Variable map ends | 73 | 22.8 Mbp (6%) |

- Significant reduction of the long tandem repeat array

# CONCLUSIONS

**The BioNano genome map is useful for**

- Validating and improving physical map and sequence (scaffolding contigs, sizing gaps, identification and correction of misassemblies)
- Identification of large tandem repeats intractable to current sequencing technologies
- Structural variation detection (insertions, deletions, translocations, copy number variations)

**Chromosomal approach**

- Enables higher-quality genome map assembly
- Resolves repeats on chromosomal level
- Focused approach to studying structural variation

# ACKNOWLEDGEMENTS


IEB

Helena Staňková
Zuzana Tulpová
Jan Vrána
Marie Kubaláková
David Kopecký
Jaroslav Doležel

BIONANO GENOMICS

Alex Hastie
Saki Chan

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

David Edwards
Paul Visendi
Jacqueline Batley
Satomi Hayashi

Colorado State University

Nora Lapitan

UC RIVERSIDE UNIVERSITY OF CALIFORNIA

Adam Lukaszewski

UC DAVIS UNIVERSITY OF CALIFORNIA

Mingcheng Luo

Bikram Gill
Bernd Friebe

IWGSC

AKADEMIE VĚD ČESKÉ REPUBLIKY

GA ČR

MŠMT

EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND
INVESTING IN YOUR FUTURE

OP Research and Development for Innovation