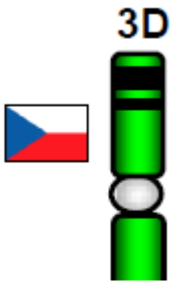
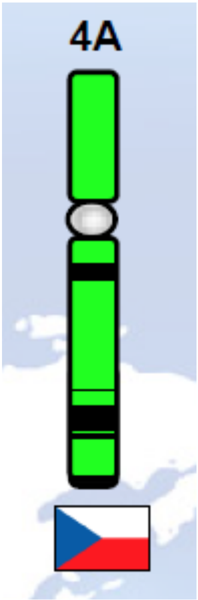
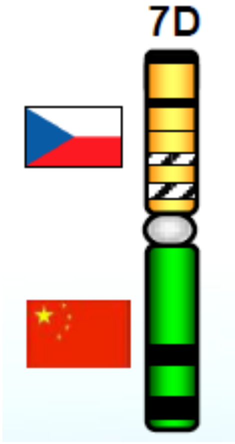
# Towards the reference sequences of four chromosome arms of bread wheat
## 3DS, 4AS, 4AL, 7DS, and 7DL



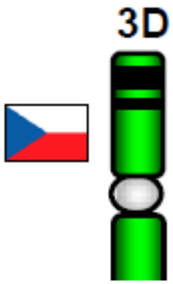




**Miroslav Valárik**

**Fargo, 2012**

# Chromosome arms 3DS



321 Mb

## Team members and collaborators:

IEB team:

Jan Bartos
Katerina Cvikova
Jan Safar
Hana simkova
Jaroslav Dolezel

Collaborators:

| | |
|---|---|
| Federica Cattonaro & Co. | IGA, Italy |
| Nils Stein & Co. | IPK, Germany |
| Catherine Feuillet & Co. | INRA, France |
| Čestmír Vlček & Co. | IMG, Czech Republic |
| Klaus Mayer & Co. | MIPS, Germany |
| Mingcheng Luo & Co. | USA |
| Jan Dvorak | USA |

| **Chromosome:** | **3DS (321 Mb)** | |
|---|---|---|
| **Leader: Jan Bartoš** | | |
| **BAC Library** | TaaCsp3DShA | |
| Number of Clones | 36 864 | |
| Insert Size | 110kb | |
| **Chromosome Coverage** | **11.0x** | |
| **Fingerprint of BAC Library** | | |
| HICF | 36 864 | |
| High Quality FP | 27.880 | |
| BAC Pools resources | 3D MTP pools | |
| Number of 3D/6D pools for the MTP | 10/16/24 | |
| **Physical Map Results** | **FPC** | **LTC** |
| Useful HICFs | 27.880 | |
| **Contigs** | **945 (25)** | **587 (15-45)** |
| **N50** | **219 kb** | |
| Total size (% of chromosome) | 96% | |
| **MTP** | **3 827** | **4 360** |
| **anchoring** | **750 contigs (70 %)** | |
| **Sequencing:** | | |
| Number of Mb Sequenced | 718.2 (454) | |
| BAC-end sequences | 4.3 | |
| BACs (Mb) | 2.3 | |
| GenomeZipper | ~715 genes | |
| **MTP 3D pools** | **All Illumina sequenced (~36x)** | |
| **Mapping resources** | **305 contigs ordered** | |
| 3D Triticum aestivum - Syntetic/Opata - GPW, 3D Triticum aestivum - SSR-2004-3, 3D Wheat-Composite2004-3D | 36 | |
| GenomeZipper | 156 | |
| 7,152 SNPs genetically mapped to chromosome arm 3DS in Ae. Tauschii | 207 | |
| 3DS- bin map | 39 | |
| **3DS low paht sequencing - GenomeZipper** | **46,297 454 (2,5x 3DS) ~ 20,899 BAC hits = 594 contig** | |

# Anchoring of sequences to BAC clones

- **Novel *in silico* approach based on sequencing of MTP pools (made by Jan Bartoš)**
- **Fifty three-dimensional MTP pools were sequenced by Illumina**

  6 – 166X coverage (mean 35X; median 23.5X)

- **Three sequence datasets were mapped to physical map**
  1. GenomeZipper fragments (594 sequences)
  2. *Ae. tauschii* sequences used for SNP mapping (7,152 sequences; ≈ 500 3DS ≈ 204 ctgs)
  3. Sequences assembled from 3DS low-pass 454 sequencing (46,297 seq )
- **Reads of individual pools were aligned to sequences using Mosaik**

  http://bioinformatics.bc.edu/marthlab/Mosaik

# Identification of positive pools

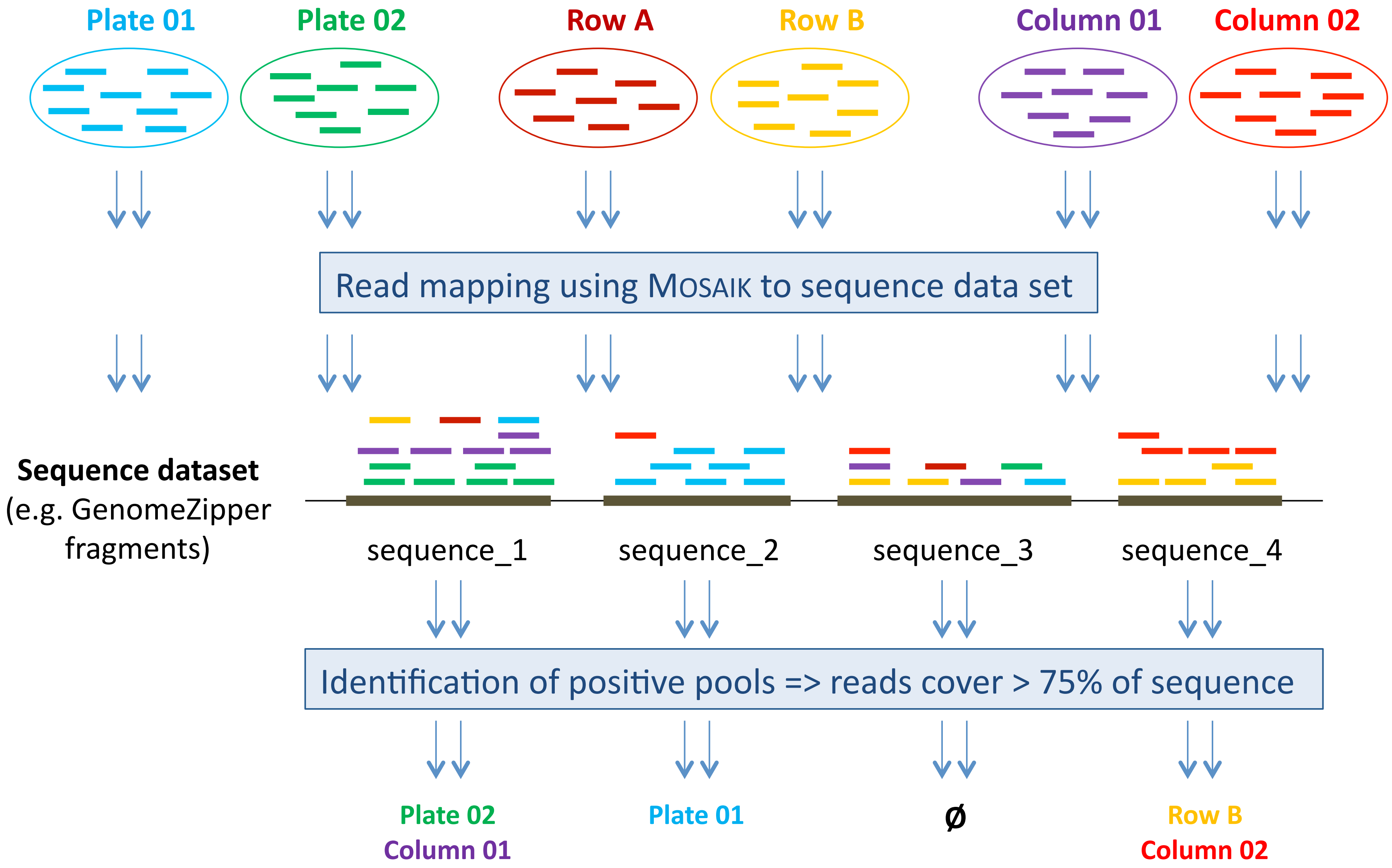


Plate 01 Plate 02 Row A Row B Column 01 Column 02

Read mapping using MOSAIK to sequence data set

**Sequence dataset**
(e.g. GenomeZipper
fragments)

sequence_1 sequence_2 sequence_3 sequence_4

Identification of positive pools => reads cover > 75% of sequence

Plate 02
Column 01

Plate 01

ø

Row B
Column 02

# BAC clone addressing

1) **One positive pool in each dimension (1 – 1 – 1)**

   --> direct BAC clone identification

   *Plate07 – RowC – Column18 --> TaaCsp3DShA_007C18*

2) **Multiple positive pools in at least one dimension (e.g. 2 – 2 – 2)**

   --> Identification of all candidate BAC clones

   a) Determine contig for all clones

      --> two clones belonging to one contig --> positive clones

   b) Check EndToEnd information for all clones

      --> two clones have significant match --> positive clones --> contig merge

3) **Sequence not anchored if:**

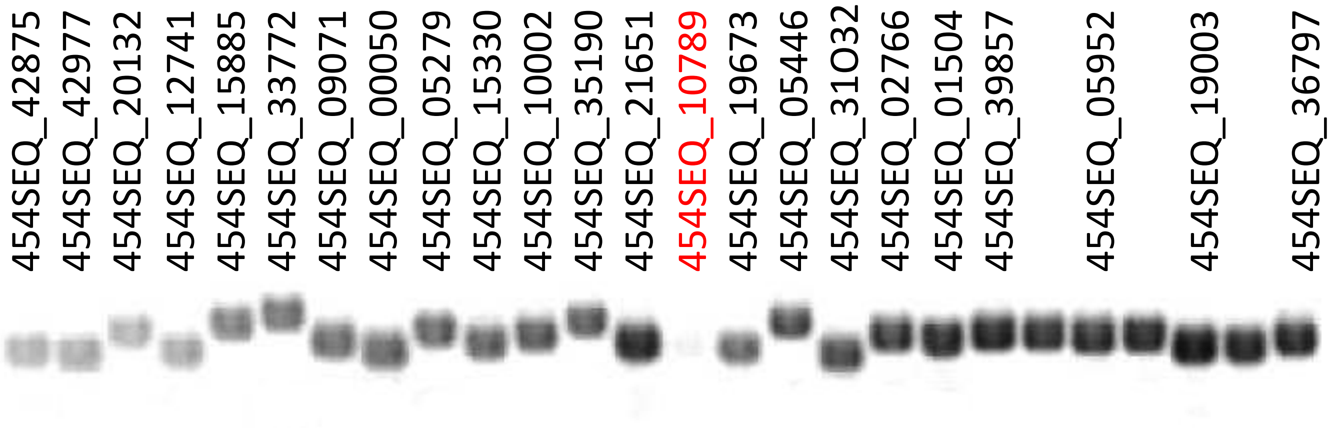
   a) positive pool is missing for at least one dimension.

   b) five or more positive pools in at least one dimension.

   c) no positive clone identified

# Evaluation of the method

- Primers were designed for 60 sequences anchored *in silico* to the physical map
- Primers were used to check presence of the sequence in particular BAC clone(s)

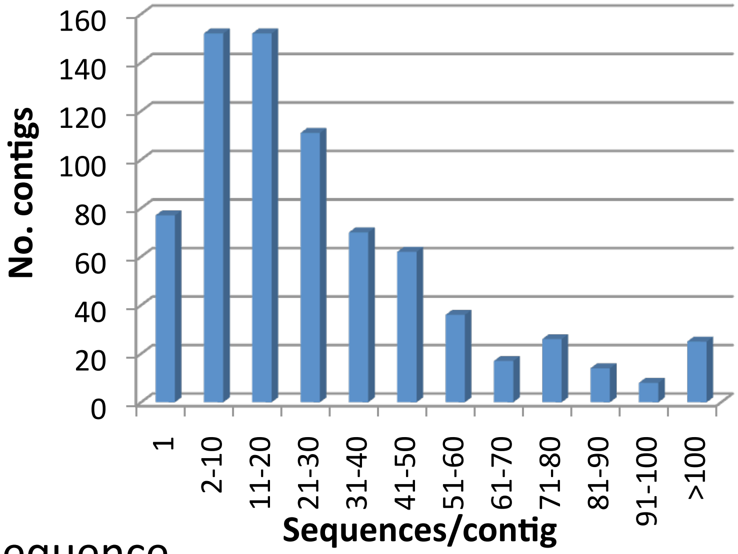


- **More than 95% of sequences were mapped correctly**

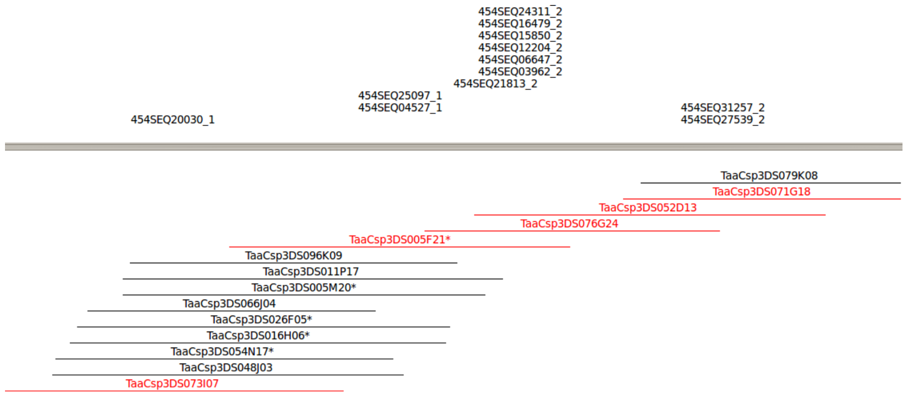
# Results

**Sequences anchored to contigs**

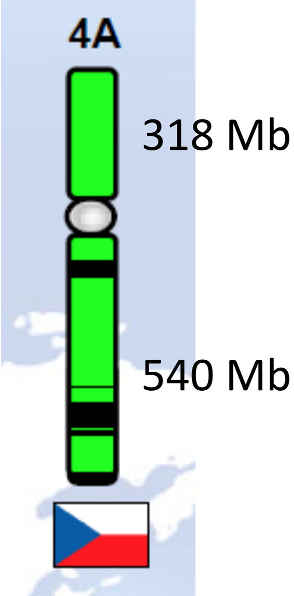
- 224 GenomeZipper fragments ≈ 156 ctgs
- 231 *Ae. tauschii* SNPs ≈ 207 ctgs
- 20,899 anonymous sequences



- 22 sequences per contig on average
- 750 (79.4%) contigs anchored to at least one sequence



# Chromosome arms 4AS and 4AL



Team members and collaborators:

IEB team:

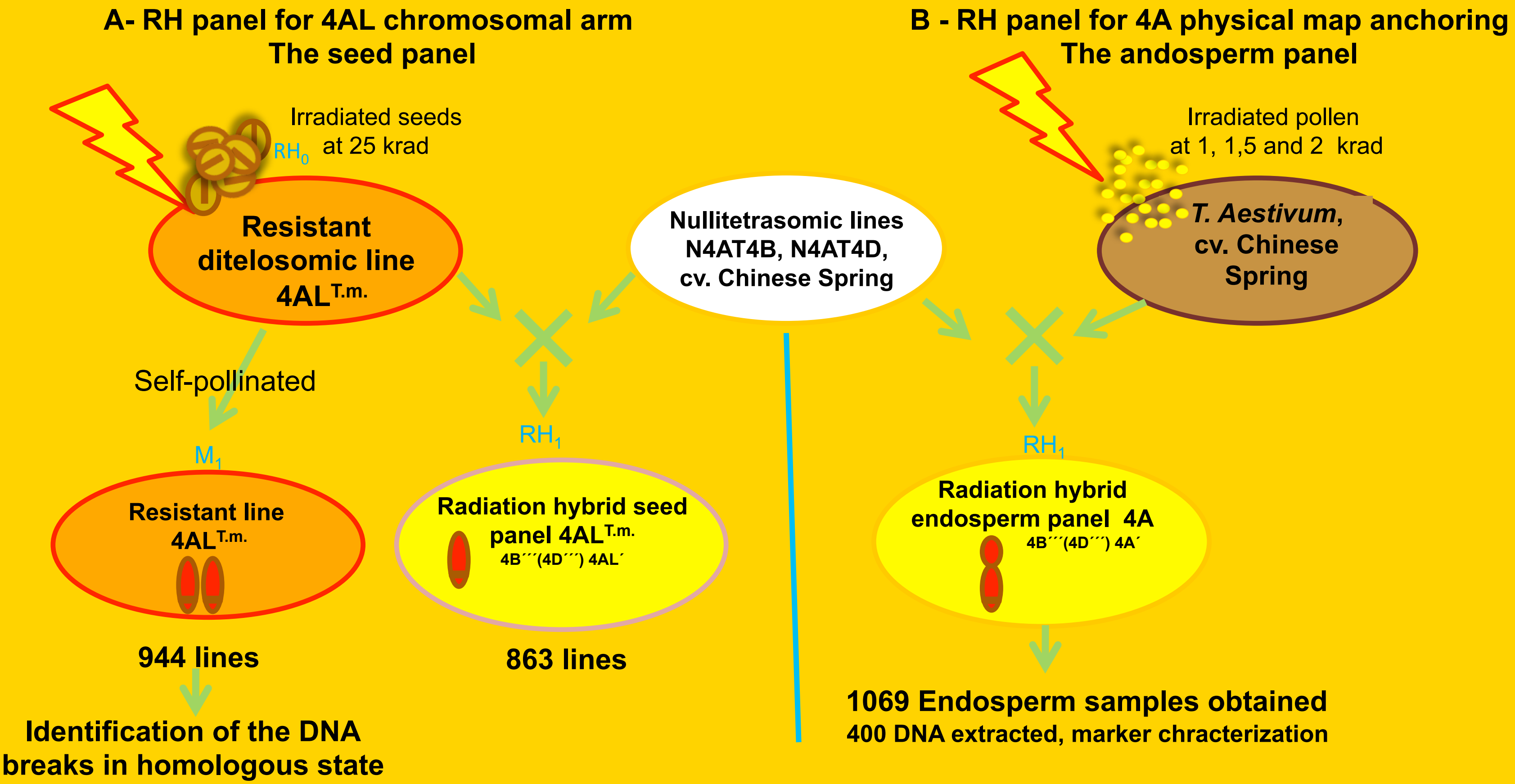
Jaroslav Doležel
Hana Šimková
Jan Šafář
Barbora Klocová

Collaborators:

| | |
|---|---|
| Pilar Hernandez | IAS (CSIC), Spain |
| Mingcheng Luo | UC Davis, California, USA |
| Klaus Mayer | MIPS, Germany |

| **Chromosome:** | **4AL** | | **4AS** | |
|---|---|---|---|---|
| **Leader: Miroslav Valárik** | | | | |
| **BAC Library** | TaaCsp4ALhA | | TaaCsp4AShA | |
| Number of Clones | 92 160 | | 49152 | |
| Insert Size | 126kb | | 131kb | |
| Chromosome Coverage | 17.3x | | 16.6x | |
| **Fingerprints** | | | | |
| HICF | 63 174 | | 36 864 | |
| High Quality FP | 60 144 | | 32 944 | |
| **BAC Pools** | | | | |
| Method of pooling | **plate pools** | | | |
| Number of plates pooled | **220** | | | |
| **Physical Map** | **FPC** | **LTC** | **FPC** | **LTC** |
| Useful HICFs | 60 144 | 60 144 | 32 944 | 32 944 |
| **Contigs** | **1 230** | **882 (15-43)** | **372** | **431 (15-43)** |
| **N50 (kB)** | **613** | | **1669** | |
| Total size (% of chromosome) | 106 | | 104 | |
| **MTP** | **7268** | **7684** | **4 433** | **4304** |
| Number of Contigs anchored | 11 | | | |
| **Sequencing** | | | | |
| BAC-end sequences | 500 | | | |
| Low pass sequences (Mb) | 900 (454) | | 707 (454) | |
| Identified Genes | 1182 | | 1751 | |
| | | | | |
| **Mapping resources** | | | | |
| 4AL radiation hybrid panel | 863 lines | | | |
| **4A radiation hybrid panel** | **1069 lines** | | | |
| ITMI mapping pop. F8 | 93 lines | | | |
| Chinese Spring x Renan F8 | 200 lines | | | |
| *T. monococcum* mapping pop F8 | 86 lines 628 markers | | | |

# Radiation Hybrid Panel Preparation for Wheat 4A Chromosome



**A- RH panel for 4AL chromosomal arm**
**The seed panel**

Irradiated seeds
RH$_0$ at 25 krad

**Resistant ditelosomic line 4AL$^{T.m.}$**

Self-pollinated

M$_1$

**Resistant line 4AL$^{T.m.}$**

**944 lines**

**Identification of the DNA breaks in homologous state**

**Nullitetrasomic lines N4AT4B, N4AT4D, cv. Chinese Spring**

RH$_1$

**Radiation hybrid seed panel 4AL$^{T.m.}$**
**4B´´´(4D´´´) 4AL´**

**863 lines**

**B - RH panel for 4A physical map anchoring**
**The andosperm panel**

Irradiated pollen
at 1, 1,5 and 2 krad

***T. Aestivum*, cv. Chinese Spring**

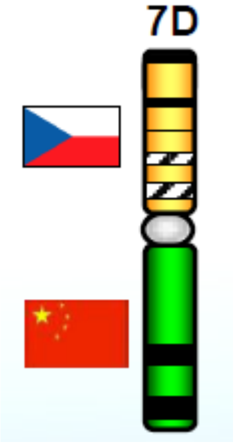
RH$_1$

**Radiation hybrid endosperm panel 4A**
**4B´´´(4D´´´) 4A´**

**1069 Endosperm samples obtained**
**400 DNA extracted, marker chracterization**

NDSU **Shahryar F. Kianian, Fargo, USA**

# Chromosome arm 7DS



347 Mb

| | | |
|---|---|---|
| **Chromosome:** | **7DS** | |
| **Leader: Hana Simkova** | | |
| **BAC Library** | **TaaCsp7DShA** | |
| Number of Clones | 49 152 | |
| Insert Size | 113kb | |
| Chromosome Coverage | **12.1x** | |
| **Fingerprint of BAC Library** | | |
| HICF | 49 152 | |
| High Quality FP | 39 765 | |
| **BAC Pools resources** | | |
| Method of pooling | plate pools, 3D pools | |
| Number of plates pooled | 128 PP, 768 3D pools | |
| **3D pools for the MTP** | **52 3D** | |
| Completion Date | 2013 | |
| **Physical Map Results** | | |
| Useful HICFs | 29 850 | |
| **Contigs** | **896** | |
| **N50** | **528** | |
| Total size (% of chromosome) | 94% | |
| **MTP** | **4608** | |
| **Sequencing:** | | |
| **Low pass sequences** | **27,700 Mbp (72x)** | done by David Edwards - assembled sequence |
| BACs | 0.45 Mbp | |

## Team members and collaborators:

IEB team:

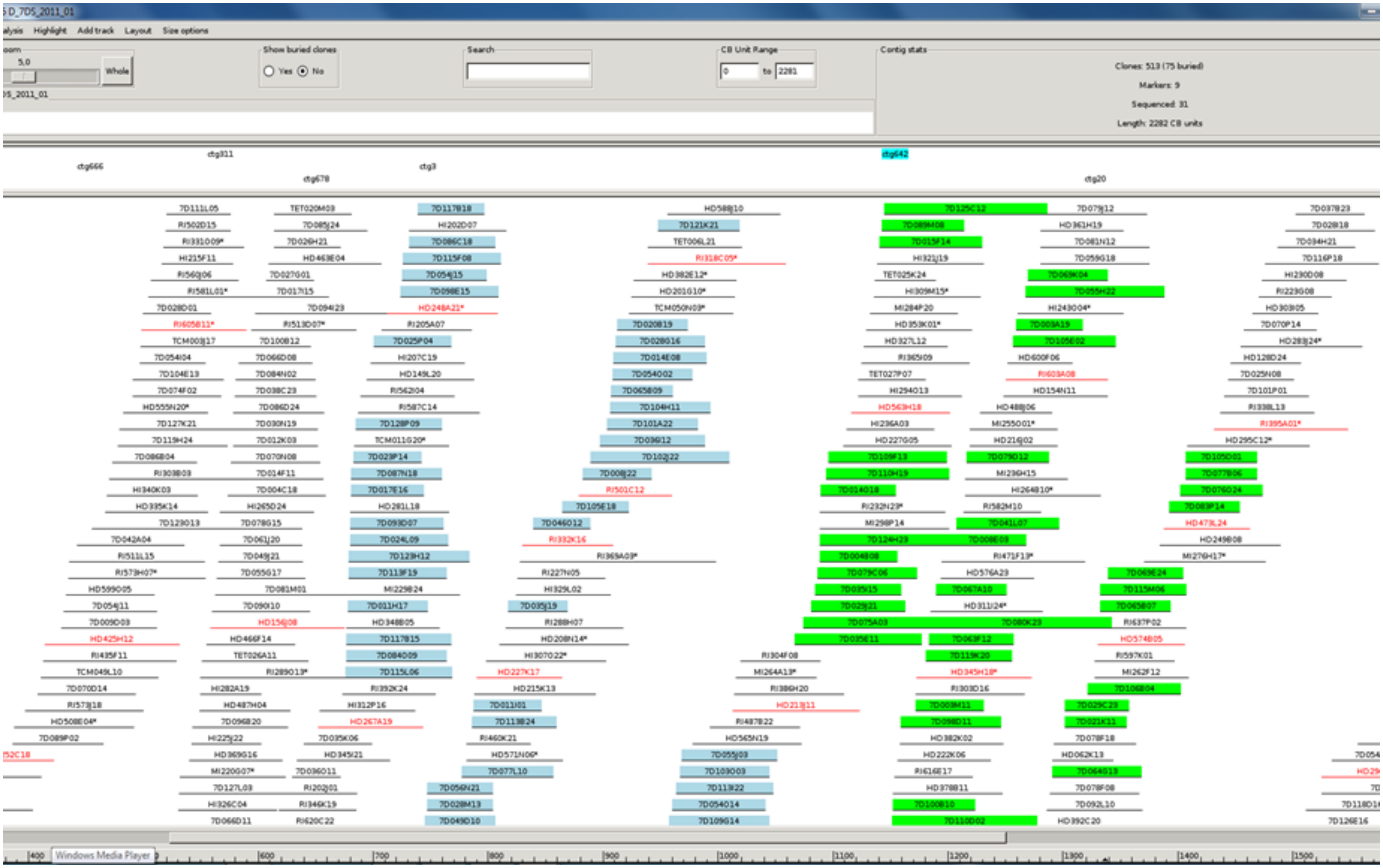
- Hana Simkova
- Jan Bartos
- Helena Stankova
- Jaroslav Dolezel
- Nicolas Blavet
- Jan Safar
- Marie Kubalakova

Collaborators:

| | |
|---|---|
| Ming-Cheng Luo | UC Davis, USA |
| David Edwards | University of Queensland, Australia |
| Shahryar Kianian | NDSU, USA |
| Andrzej Kilian | Diversity Arrays Technology, Australia |

# 7DS Physical map enhancement using *T. tauschii* physical map

**Integration of the 7DS and Ae. tauschii physical maps resulted in 25% decrease of contig number ~ from 1197 to 896**



Zoom 5.0

Show buried clones ○ Yes ● No

CB Unit Range 0 to 2281

Clones: 513 (75 buried)

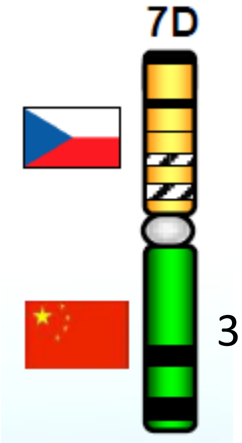
Markers: 9

Sequenced: 31

Length: 2282 CB units

Intergrated by M. Luo UC Davis, California, USA

# Chromosome arm 7DL



381 Mb

Song Weining & Group

**Northwest A&F University**

Collaborators:

| | |
|---|---|
| Ming-Cheng Luo | UC Davis, USA |
| David Edwards | University of Queensland, Australia |
| Jaroslav Dolezel & Group | IEB, Czech Republic |
| Rudi Appels & Group | Murdoch University, Australia |
| Dave Edwards & Group | University of Queensland |

| **Chromosome:** | **7DL** | |
|---|---|---|
| **Leader: Song Weining** | | |
| **BAC Library** | **TaaCsp7DLhA** | |
| Number of Clones | 50,304 | |
| Insert Size | 115kb | |
| Chromosome Coverage | **14,9x** | |
| **Fingerprint of BAC Library** | | |
| HICF | 50, 304 | |
| High Quality FP | 48, 452 | |
| **BAC Pools resources** | | |
| Method of pooling | | |
| Number of plates pooled | | |
| **Physical Map Results** | | |
| **Contigs** | **1614** | |
| **Average contigs length(kb)** | **653** | |
| Total size (% of chromosome) | 92% | |
| **MTP** | **4472** | |
| **Sequencing:** | | |
| **Low pass sequences** | | |
| low copy regions | 238 Mbp | |
| Genes | 1,659 | |
| SSRs | 23, 272 (40 valid.) | |

- An assembly of 238 Mbp representing unique and low copy regions produced
- 1,659 confident genes identified
- 62% of which could be ordered and orientated within a syntenic build
- Annotated assemblies are available on www.wheatgenome.info (Lai *et al*. 2012, Plant Cell Physiol. 53:1–7)

- A total of 23, 272 simple sequence repeats were identified in the 7DL survey sequence
- 40 SSR markers designed and validated
- Nie *et al.*, *American Journal of Botany* in press

- A number of ISBP markers were also identified in the survey sequence
- pre-miRNAs, miRNAs, transposable elements and other repetitive sequences were identified in the survey sequence
- The integration of the 7DL physical map with its genetic counterpart is in progress

THANK YOU