Construction and characterization of wheat 4A chromosome specific physical map as a base step for the chromosome sequencing.

Barbora Klocová[1], Monika Kladivová[1], Zeev Frenkel[2], Ajay Kumar[3], Shahryar F. Kianian[3], Hana Šimková[1], Jan Šafář[1], Yuqin Hu[4], Yong Zhang[4], Frank M. You[4], Mingcheng Luo[4], Abraham Korol[2], Jaroslav Doležel[1], Miroslav Valárik[1]

[1]Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Šlechtitelů 31, Olomouc – Holice 783 71, Czech Republic
[2]Institute of Evolution, University of Haifa, Haifa 31905, Israel
[3]Department of Plant Sciences, North Dakota State University, Fargo, Loftsgard Hall 470G,ND 58108, USA
[4]Department of Plant Sciences, University of California, Davis, CA 95616, USA

Introduction

***Triticum astivum* L.** (bread wheat) has large and complex genome. The genome size is approximately 17 Gb with 80% of repetitive sequences and is composed of three homeologous genomes A, B and D. All of this makes sequencing, mapping, and **Marker Assisted Breeding** (**MAB**) a challenging task. Also, anchoring and orienting contigs of the physical maps are difficult.
To overcome all of these problems, the genome of cv. Chinese Spring is dissected to individual chromosomes and chromosomal arms by flow cytometry and chromosome specific BAC libraries are constructed. This allows to coordinate the collaboration of several laboratories from around the world in the framework of the **International Wheat Genome Sequencing Consortium** (**IWGSC**, http://www.wheatgenome.org) to succeed in wheat genome sequencing.
Our laboratory is coordinating development of sequence-ready chromosome arm specific physical maps for both arms of chromosome 4A.

Results

Chromosome arm specific BAC libraries were constructed for wheat chromosome 4A. Both libraries were fingerprinted and assembled into contigs using both, **LTC** (**Linear Topology Contig**) and **FPC** (**FingerPrint Contigs**) software. The **LTC** assembly was enhanced by newly developed algorithm for super-scaffolding (Frenkel et al., 2010). Applying the super-scaffolding resulted in significant improvement of contig length (Table 1).
The **4AS** library was assembled into **416** contigs and **MTP** (**Minimum Tiling Path**) consist of **4 422** clones. After superclustering, there were 99 supercontigs with more than 100 clones in one contig. The longest supercontig comprised **582** clones.

Also, **4AL** physical map was assembled into **1129** contigs and **MTP** contains of **8 369** clones. After additional superclustering there were **115** supercontigs with more than 100 clones in one contig. The longest supercontig comprised **997** clones.
Moreover, the supercontigs of both physical maps contains more than half of informative clones. The newly established connections between contigs in supercontigs were evaluated by molecular markers (Figure 1).
3D pools of **MTP** clones were prepared to simplify marker anchoring to the maps. To the physical map of 4AL were anchored 20 markers, four of this markers were found in one supercontig.

Table 1.: Summary of 4AS and 4AL arm specific physical maps

Chromosome	4AS			4AL		
assembly method	FPC	LTC	LTC (super-scaffolding)	FPC	LTC	LTC (super-scaffolding)
Number of all useful BACs	32 944			60 140		
Number of BACs in contigs > 5 clones	24 633	24 040	24 040	39 285	41 097	41 254
Number of clusters >100 clones	76	73	**99**	75	105	**115**
Number of clusters 100- 50 clones	171	106	31	258	151	64
Number of clusters < 50 clones	203	238	118	972	872	745
Number of clones in clusters > 50 clones	20 212	18 563	**21 544**	23 103	29 634	**32 856**
Number of clones in clusters < 50 clones	4 421	5 477	2 496	16 182	11 463	8 398
Number of singletons	8 311	8 904	8 904	20 855	19 043	18 886
MTP by LTC	4 433	4 422	4 422	7 291	8 369	8 526
Estimated coverage (%)[a,b]	83	94	**94**	72	87	**87**

[a] based on the expected size of chromosome arms (318 and 540 Mb for 4AS and 4AL arms, respectively)
[b] based on the total number of MTP clones and expected overlap 30% and 50% for the LTC and FPC, respectively and average insert size 131 and 128 kb for 4AS and 4AL BAC libraries, respectively

Conclusion and prospects

- **4AS** physical map has 248 contigs after super-scaffolding. This contigs consist of 24 040 clones.
- **4AL** physical map has 924 contigs after super-scaffolding. This contigs consist of 41 254 clones.
- **4AS** and **4AL MTP** clones covers **94 %** and **87 %** length of arm, respectively.
- After sequencing of 3D pools of **MTP** BAC clones high-capacity physical map anchoring will be performed *in silico*.
- Endosperm radiation hybrid panel comprising 1 100 lines will be used for contig ordering and orientation.

References

- Frenkel, Z.; Paux, E.; Mester, D.; et al., BMC Bioinformatics S (2010) DOI: 10.1186/1471-2105-11-584
- Xue S., Zhang Z., Lin F., et al., Theor Appl Genet (2008) DOI 10.1007/s00122-008-0764-9

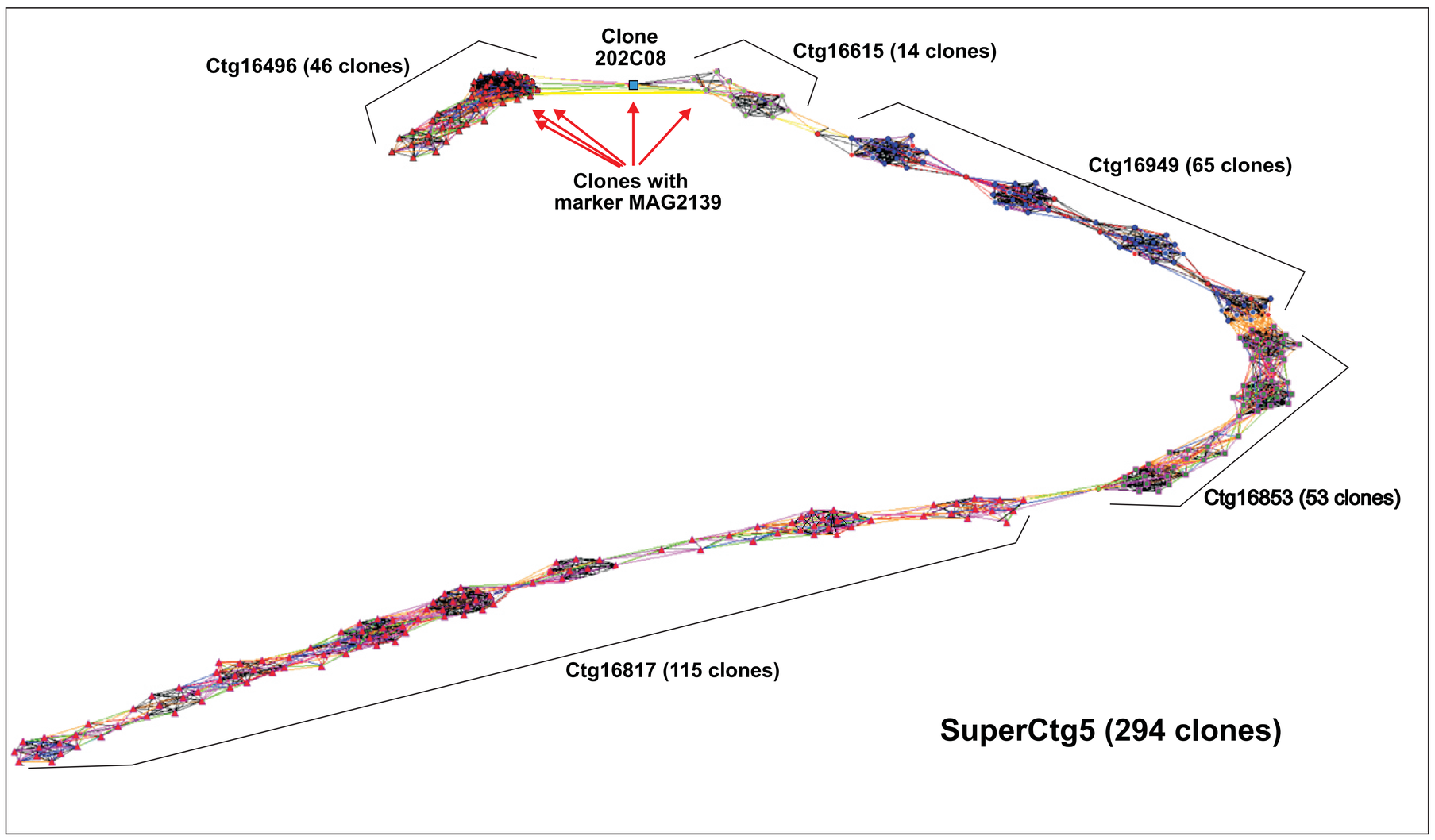

Figure 1: Super-contig algorithm verification

The super-contig SuperCtg5 comprises 294 BAC clones and represents about 3.6 Mb of chromosome 4A. It was put together by end-to-end merging of five standard LTC contigs and one singleton (202C08) via lower significant clone overlaps (without using markers).

The nodes correspond to clones; edges correspond to significant clone overlaps; colors of edges reflect p-value; nodes corresponding to the same contig have the same color and geometric shape.

Connecting clone 202C08 was not included into initial assembly because it is not proven by overlaps of parallel clones (probably putative chimerical). More detailed analysis detected low significant overlap between clones from the ends of Ctg16496 and Ctg16615 (yellow edge) proving this clone (202C08).

The connection between these two contigs was verified by molecular marker MAG2139 (Xue et al., 2008). Nodes corresponding to clones carrying this marker are marked by red arrows.

This work has been supported by the Czech Science Foundation (P501/10/1740), MSMT CR and EU (Operational Programme Research and Development for Innovations No. ED0007/01/01) and Internal Grant Agency PrF-2012-003