IPK
GATERSLEBEN

The Whole Genome Profiling (WGP ™) for physical mapping of wheat chromosome 6A

Naser Poursarebani, Thomas Nussbaumer, Hana Šimková, Jan Šafář, Hanneke Witsenboer, Jan van Oeveren, IWGSC, Jaroslav Dolezel, Klaus Mayer, Nils Stein, Thorsten Schnurbusch

Zeev Frenkel, Abraham Korol

Bundesministerium
für Bildung
und Forschung

Outline:

Introduction and methodology

Why 6A chromosome?

BAC-based Physical map; Basic principles and the new technology

The strategies for 6A: WGP technology; FPC & LTC assembly tools

Results:

Physical map assemblies

Physical map anchoring

Comparison of the assemblies developed by the two tools

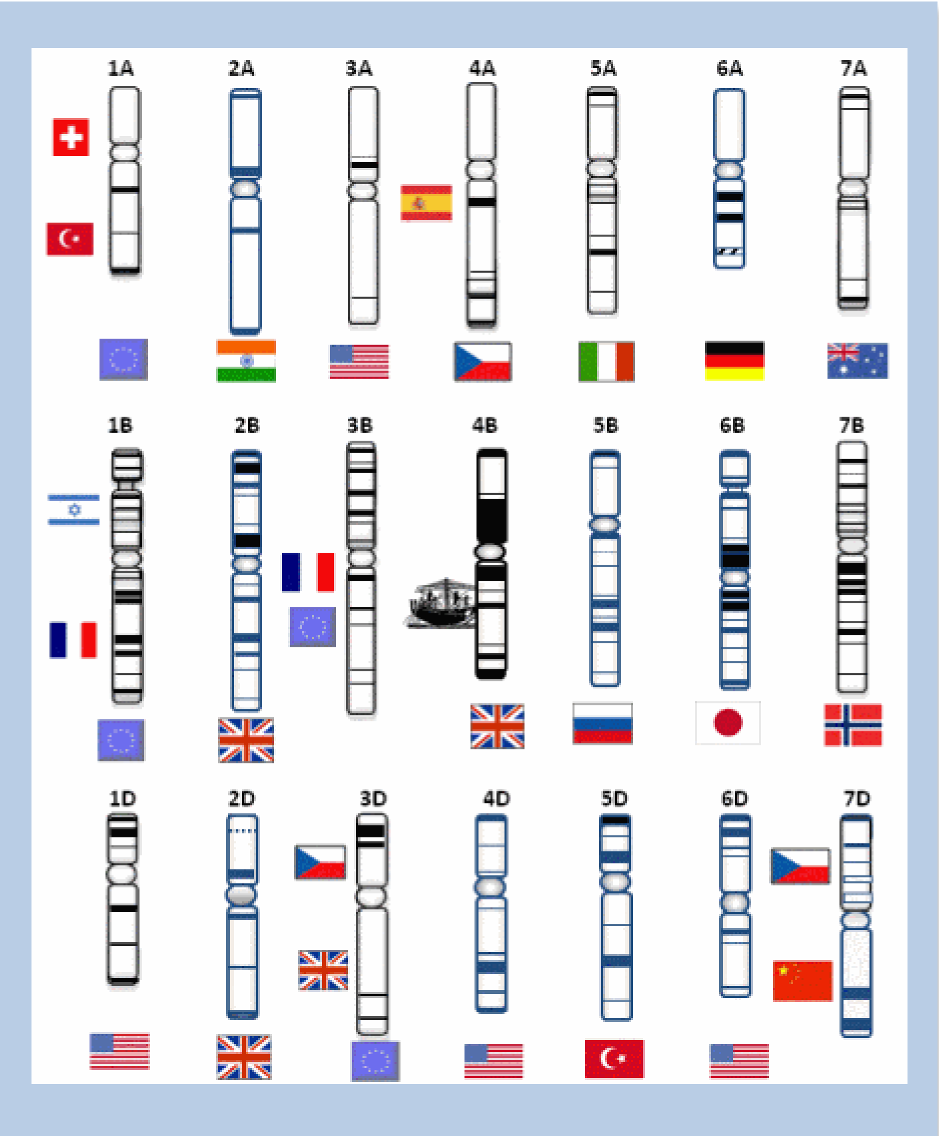
Recombination estimation and physical map gene decoration

Synteny with other grass genomes

6A Browser

Summary

Why 6A chromosome?

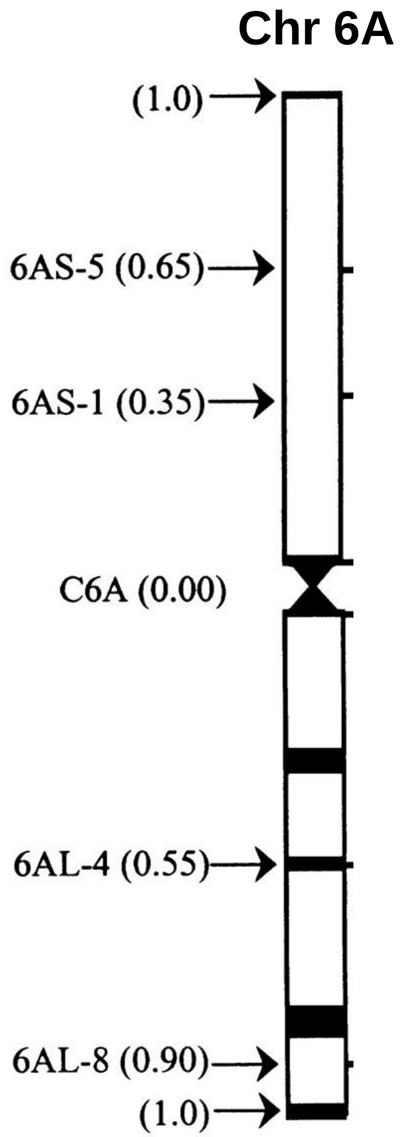


- make a German contribution to the international consortium

- ***last-remained chromsome***

IWGSC **= International Wheat Genome Sequencing Consortium,** http://www.wheatgenome.org

Selected Traits associated with Chromosome 6A



Yield

-QTL for Yield and TKW

(ca. 10-15%; Snape et al. 2007; Röder et al. unpubli., centromere)

Development

-Seedling vigor (Spielmeyer et al. 2007)

Disease Resistances

-Powdery mildew, *Pm21* (Muranty et al. 2009)

-Stem rust, *Sr8*, *Sr13 (Ug99)*, *Sr26,* (McIntosh et al. 1995/ Belayneh Admassu, 2010)

-Aphid resistance (Castro et al. 2005; centromere)

-*Septoria tritici*, *Stb15* (Arraiano et al. 2007)

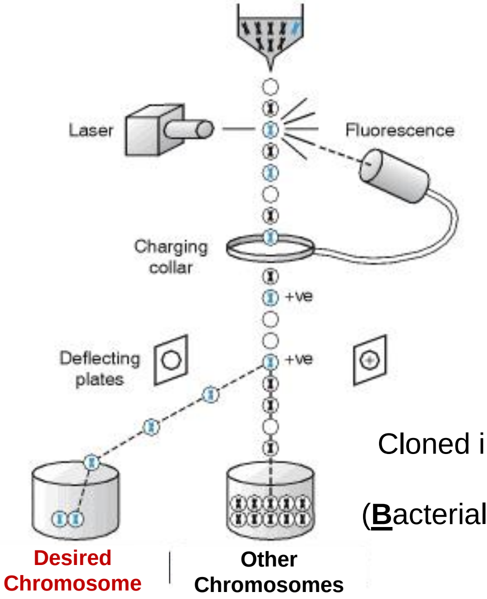
-Fusarium head blight, major QTL (Holzapfel et al. 2008)

BAC-based physical map

Basic principles and the new technology

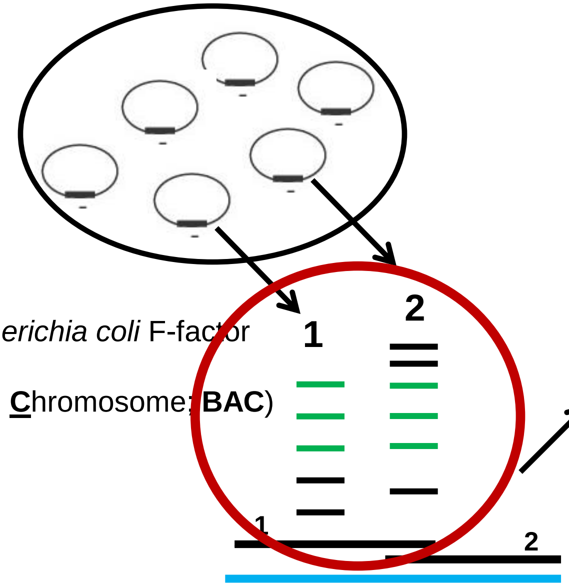
An anchored physical map:
A model of a “Reconstructed Chromosome”

Chromosome Sorting by Flow Cytometry
(Wheat Genome; Hexaploid, ~ 17Gb)

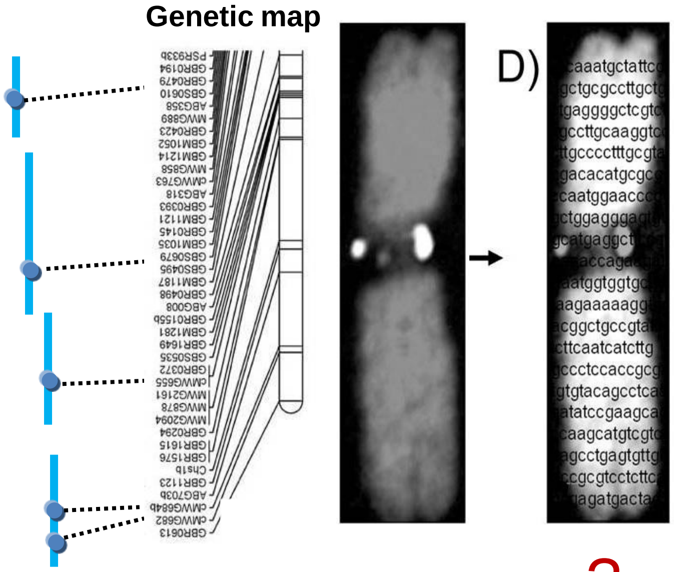


Cloned into *Escherichia coli* F-factor

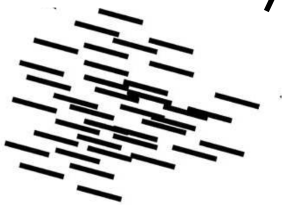
(**B**acterial **A**rtificial **C**hromosome; **BAC**)



Anchoring to the chromosome



Fragmented
Chromosomal DNA



A FP contig (FingerPrinted contig)

with two Clones

Keygene, Wageningen

IPK & MIPS, Germany

?

UEB, Czech Republic

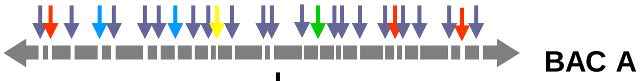
Modified from Nils Stein; 2007

Fingerprinting strategies

High Information Content Fingerprinting(HICF)

(Luo et al., 2003)

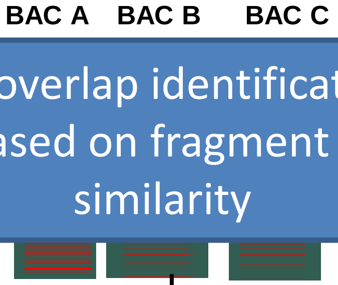
I: Multiple RE digestion



II: Enzymatic digested fragments: BAC fingerprints

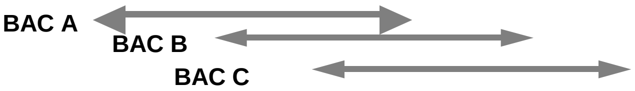
III: Fragment labeling using fluorescent SNaPshot chemistry: required for size calling

IV: BAC fragment based profile: Between 40 to 250 bans per BAC



BAC overlap identification;
Only based on fragment length similarity

IV: BAC overlap identification and contig assembly

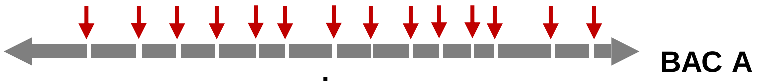


Whole Genome Profiling (**WGP**™)

(van Oeveren et al., 2011)

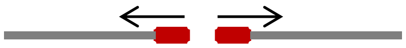
KeyGene

I: Single RE digestion



II: Sequencing adaptor ligation

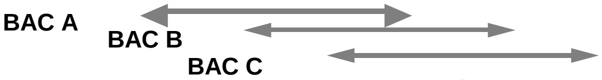
III: End sequencing of enzymatic digested fragments



IV: BAC sequence based profile: on average 30 sequence tags of ~50bp per BAC (126kb, 1HindIII site/2.5Kb)

BAC overlap identification;
based on tag length and sequence similarity

IV: BAC overlap identification and contig assembly



Assembly tools and their parameters

FPC (FingerPrinted Contigs)

(Soderlund et al., 1997)

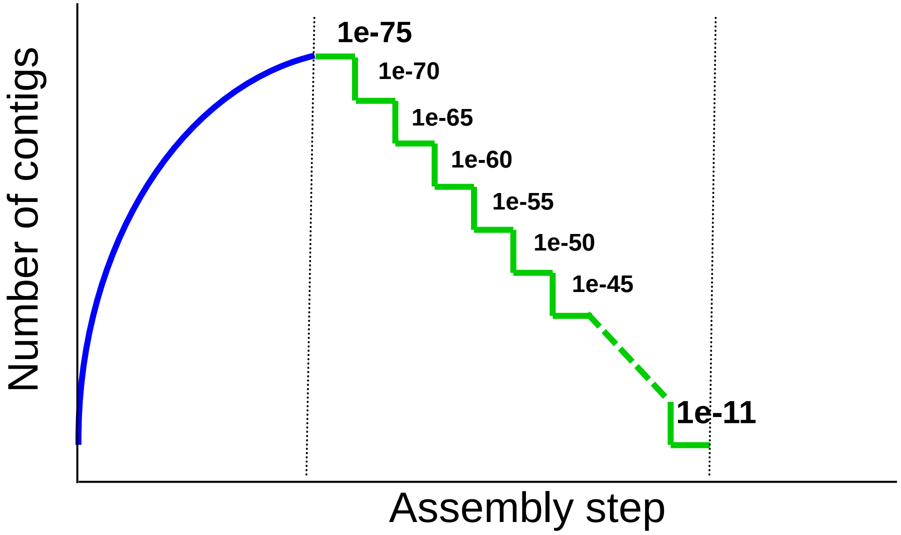
LTC (Linear Topological Conig)

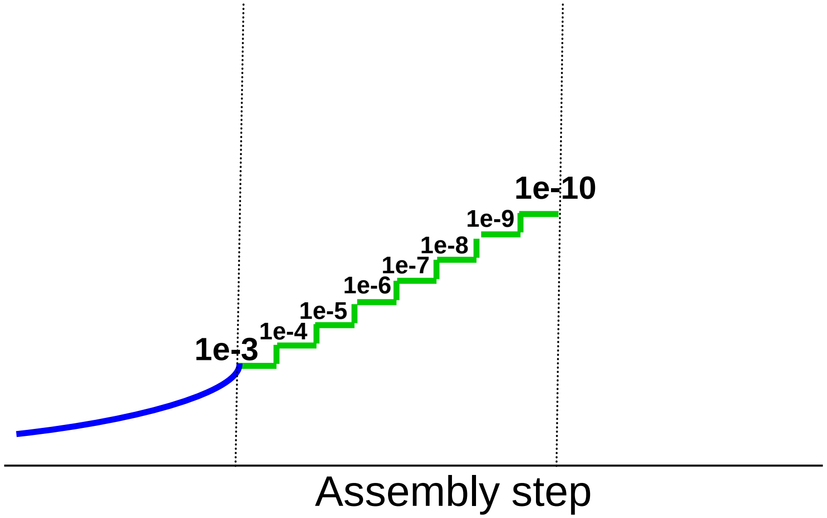
(Frenkel et al., 2010)

$$\sum_{m=M}^{nL}\left[\binom{nL}{m}\left((1-p)^{m}p^{nL-m}\right)\right]$$

Sulston score:

- ➢nL= minimum of bands
- ➢nH=maximum number of bands
- ➢ M=number of shared bands
- ➢Gellen=possible value for bands





- Automated Assembly (Single to end & end to end Merging , DQing)

The strategies for 6A:

WGP technology

+

FPC & LTC assembly tools

Results

Physical map assembly

The BAC library resources utilized

The DNA obtained by flow cytometry of ditelosomic wheat lines

(Dr. Jaroslav Dolezel; Czech Republic)

Wheat chromosome 6A arm-sorted BAC libraries

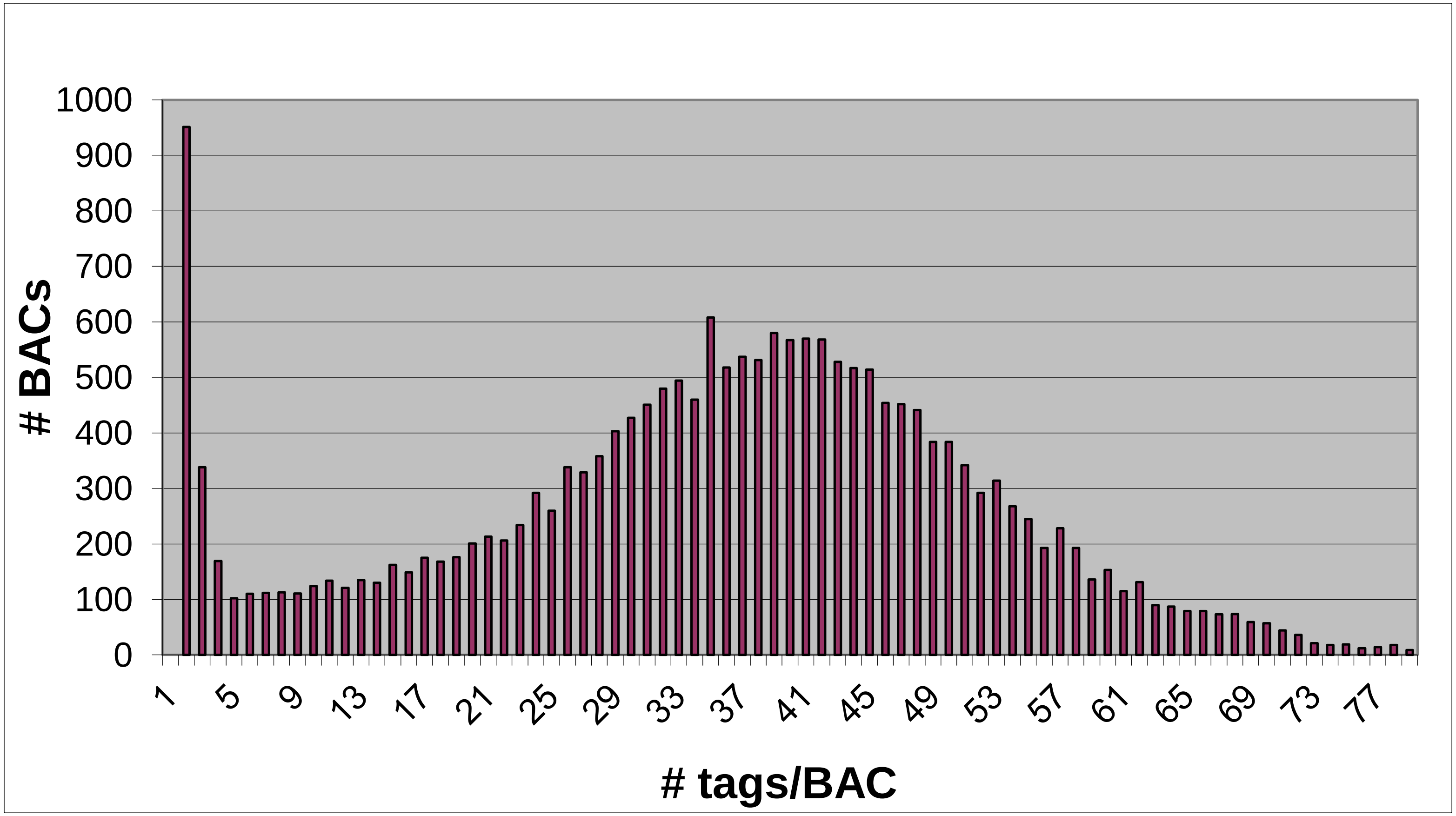
No of BAC clone	Long Arm (6AL)	Short Arm (6AS)	Library coverage
Estimated size of chromosome	369 Mbp	336 Mbp	-
Complete BAC library (1)	49.152	55.296	~16x
for WGP™	24.576	22.656	~8x
as WGP™ Output	18,660	19,289	~7x
WGP™ Output used for BAC assembly (2)	17,309	18,820	~5x (6AL) & ~7x (6AS)

Of WGP-based output

General WGP parameters and sequence data processing

Chromosome arm	Long arm	Short arm
No of tagged BACs	18,660	19,289
% tagged BACs	75,9%	85%
WGP tag length	50 nt	50 nt
average No of WGP tags/ BAC	27	29.2

Distribution of #tags/BAC



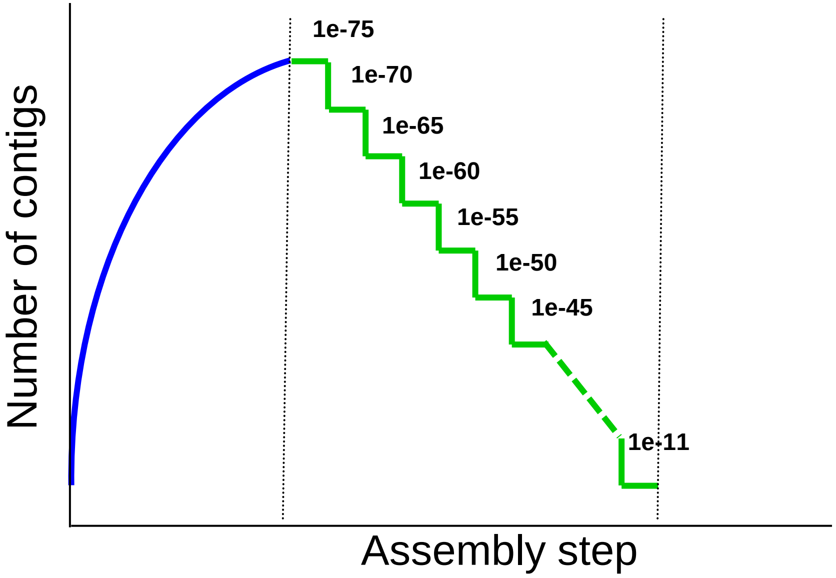
Entered to the assembly (Clones with 6 to 68 tag)

Short Arm BACs: 18,820, 7x

Long Arm BACS: 17,309, 5 x

Wheat 6A physical map assembly via FPC

FPC (FingerPrinted Contigs)

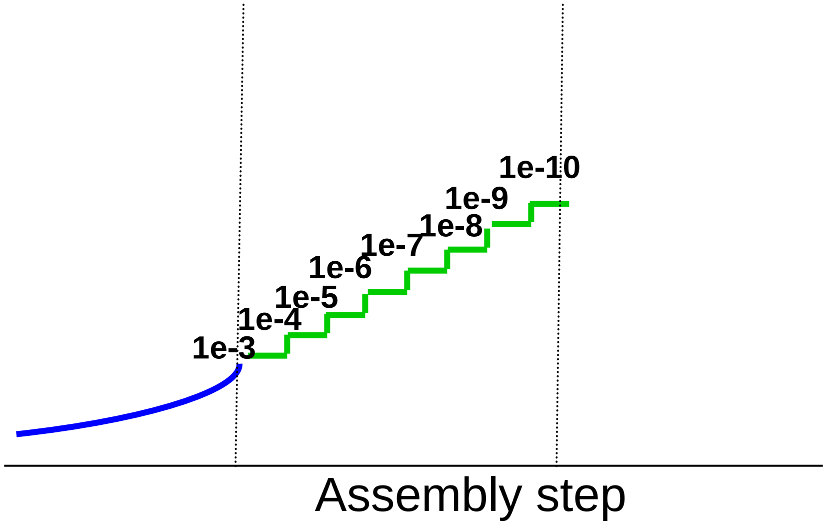


Chromosome 6A arm separated assemblies via FPC (1e-11)

	Short arm	Long arm
No of contig	640	620
No of singleton	5045	3560
Average Kb	706	780
L50 Kb	2223	2693

Wheat 6A physical map assembly via LTC

LTC (Linear Topological Contig)



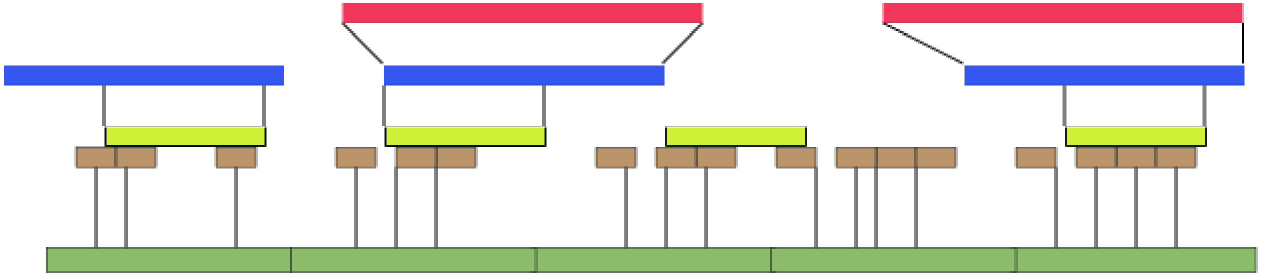
Chromosome 6A arm separated assemblies via LTC

	Short arm	Long arm
No of contig	1217	1113
No of singleton	3136	2581
Average Kb	428	488
L50 Kb	1090	945

in silico genetic anchoring of the 6A physical map

sequnce enrichment of the physical map

🟥 *Aegilops tauschii*
🟦 *Triticum urartu*
🟨 6A WCS contigs
🟫 WGP tags
🟩 6A clones



Average amount of sequence information per contig: 2144 bp ⟶ **11067 bp**

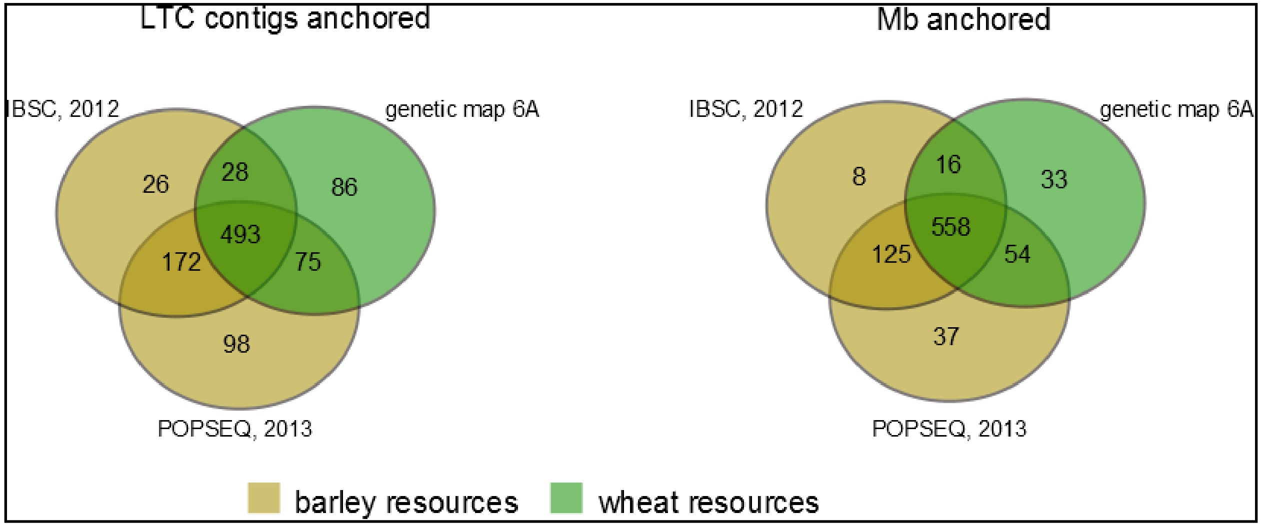
Provided the basis for sequence homology search against genetic markers with known sequences!

in silico genetic anchoring of the 6A physical map

sequence homology searches against genetic marker with known sequences

Wheat and barley genetic maps with markers of known sequences

Wheat map developed by Poland et al., 2012;
Wheat map developed by Cavanagh et al., 2013;
Barley genomic resources, (IBSC, 2012; Mascher et al., 2013)



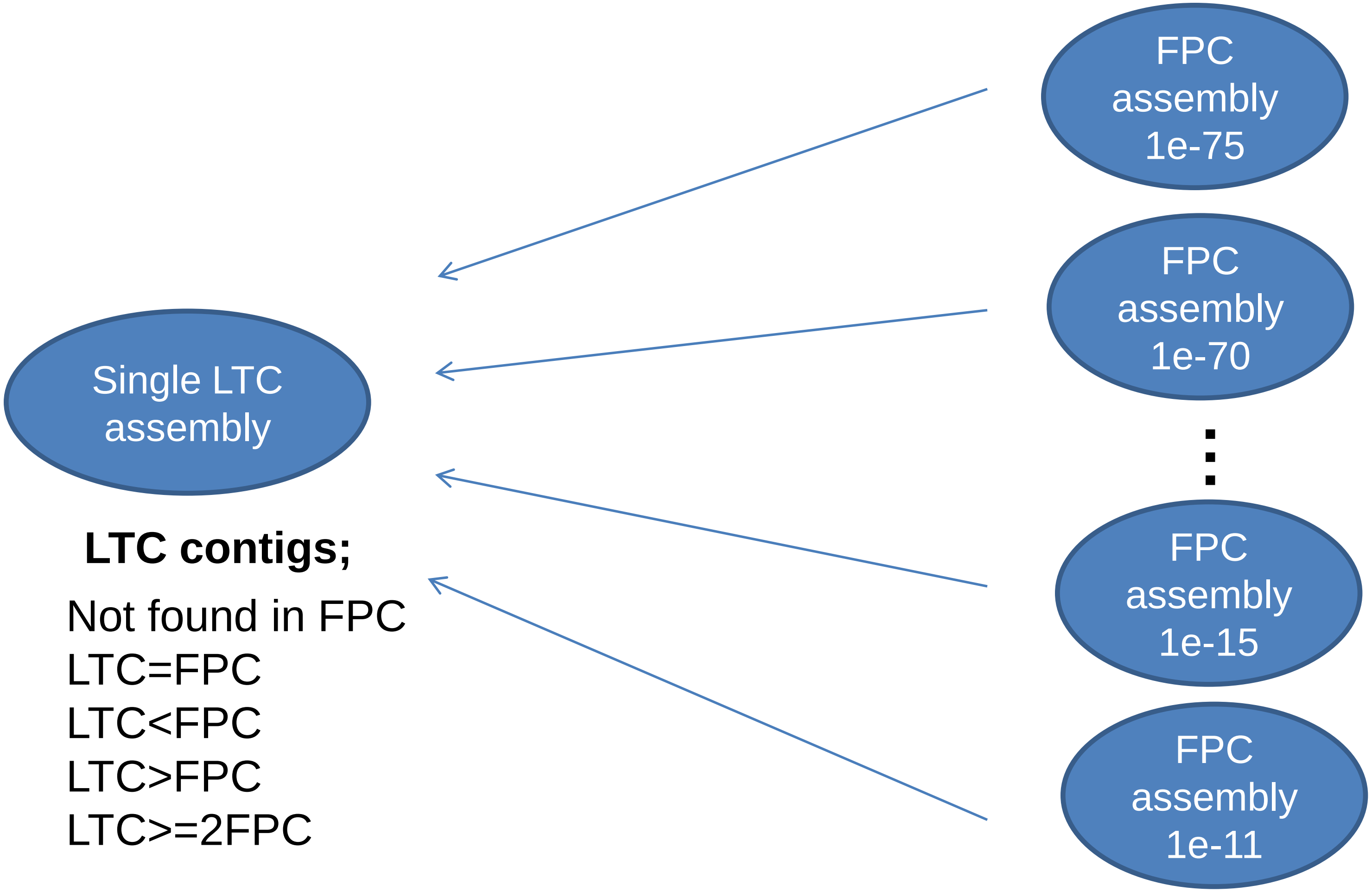
Overall, 831 Mb (or 79%) of 6A physical contigs could be genetically anchored!

Comparison of FPC against LTC

Different FPC assemblies obtained at different stringencies

Against

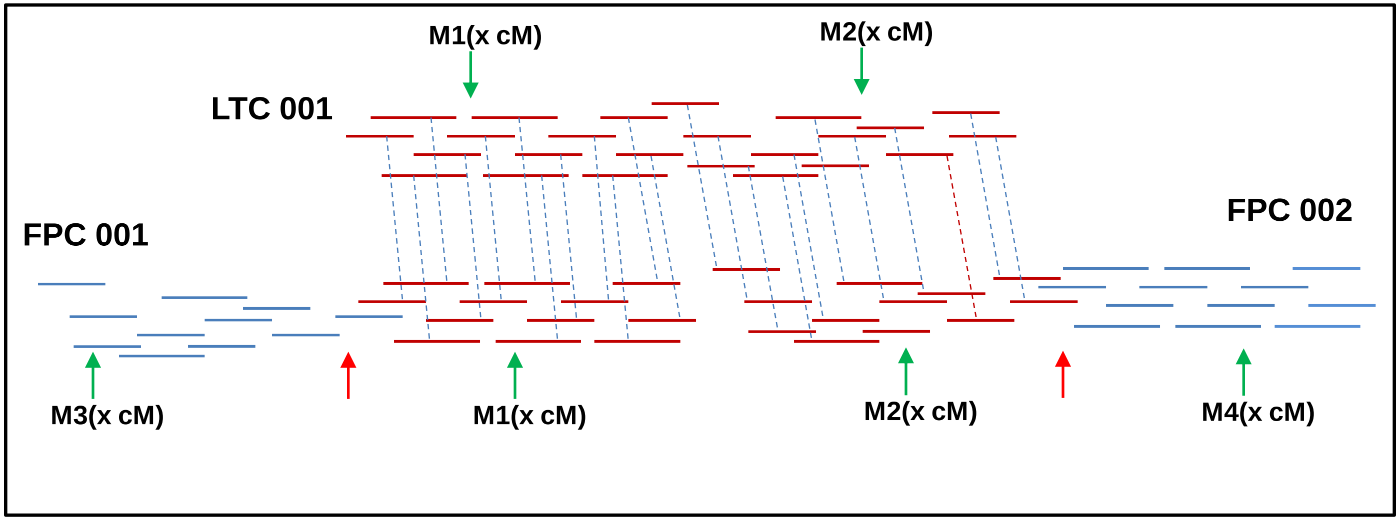
single assembly of LTC as reference



LTC contigs;

Not found in FPC
LTC=FPC
LTC<FPC
LTC>FPC
LTC>=2FPC

Comparison of FPC against LTC (LTC>=2FPC, 1e-11)

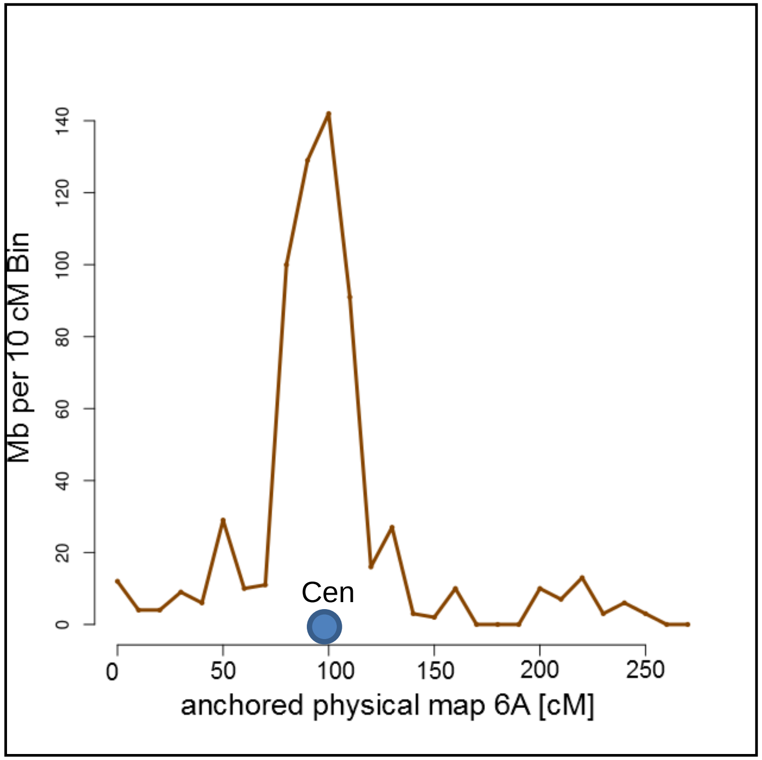


Cases of conflicts (LTC>=2 FPC) for which marker data allowed further confirmation of the contigs structure		
Chromosome arm	Short arm	Long arm
No of conflicts	72	143
No of conflicts for which informative markers (1) were available	22	26
Engaged FPC contig (s) is chimeric	17	20
Engaged LTC contig is chimeric	4	4
The LTC contig corresponds to the ends of the respective FPC contigs(2)	1	2

using WGP data, LTC showed a better performance than FPC!

Estimated recombination frequency along wheat chromosome 6A

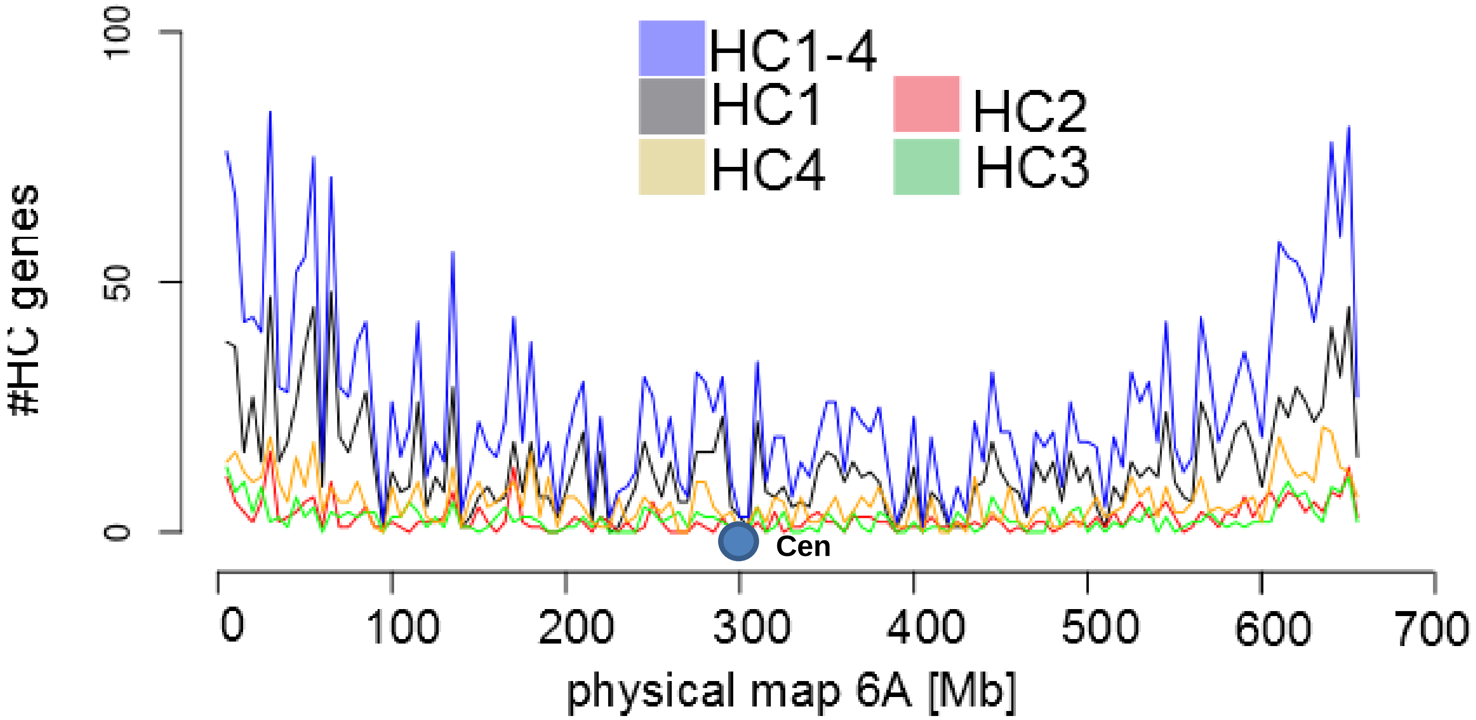
The underlying genetic map used for 6A genetic anchoring was divided into 26 bins, each 10 cM in size.



Centromeric area of the chromosome was suppressed for recombination!

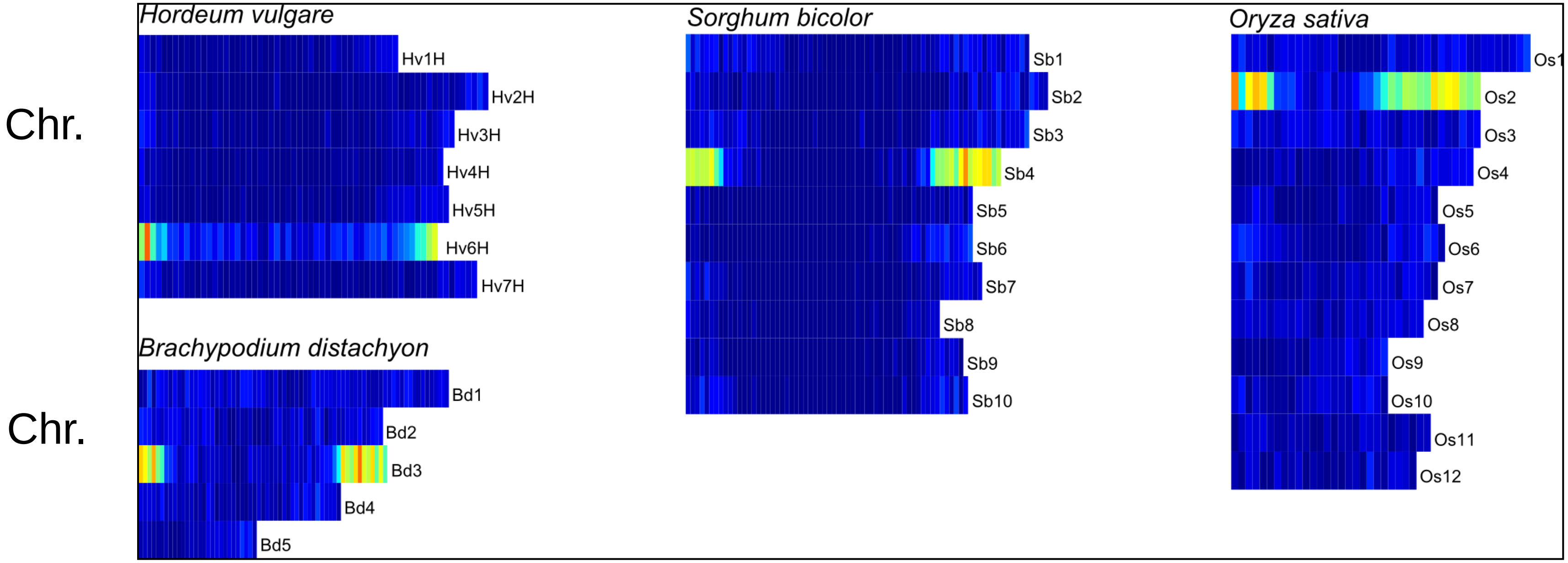
Estimated gene distribution along wheat chromosome 6A

(of the 5024 genes of 6A, 3359 genes mapped to the anchored physical map)



Gene frequency per physical unit was increased from centromere towards telomeres

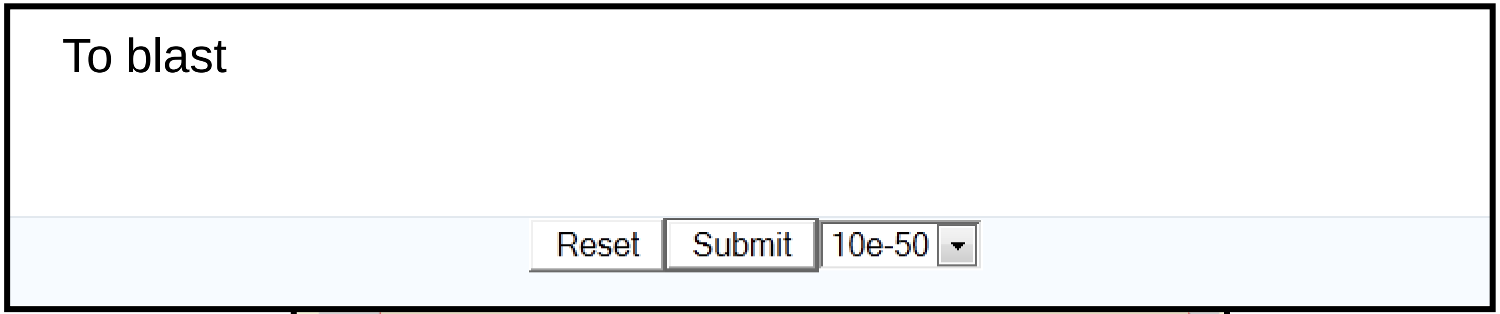
Synteny of wheat 6A physical map with other grass genomes

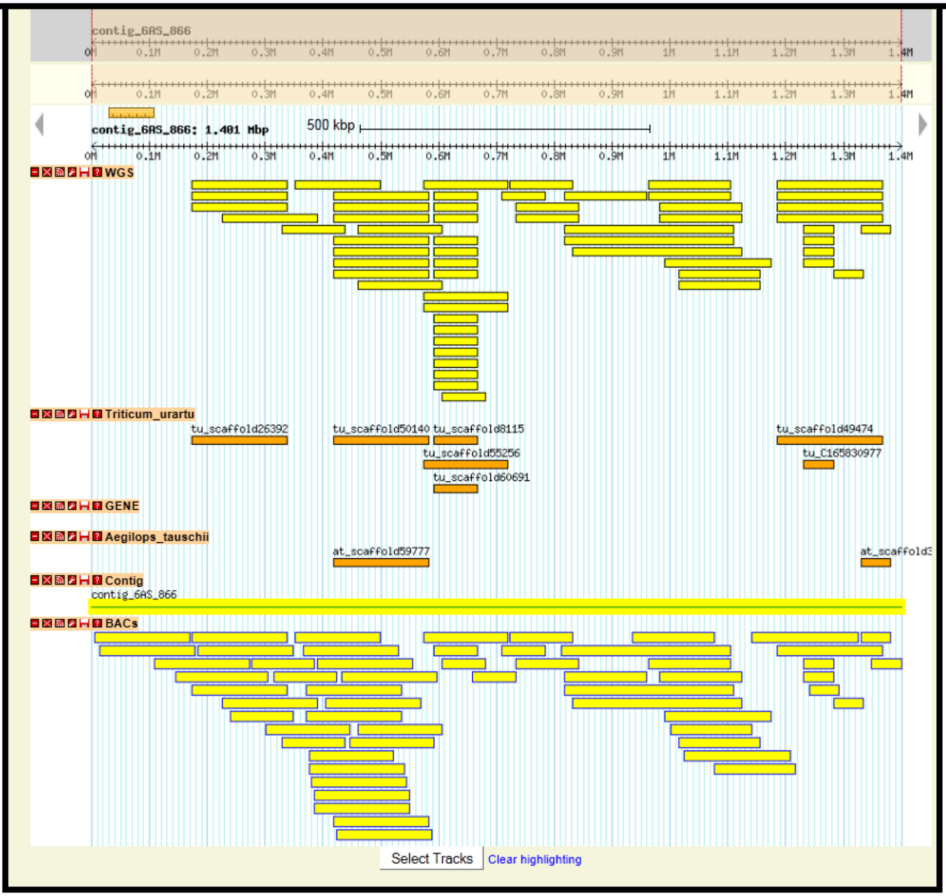


The wheat chromosome 6A physical map re-confirmed the known syntenic relations to the other grass genomes

Wheat chromosome 6A physical map DataBase!

http://seacow.helmholtz-muenchen.de/cgi-bin/gb2/gbrowse/Wheat_PhysMap_6A


To blast



Summary:

❖ WGP technology and LTC revealed a highly robust physical map of wheat chromosome 6A.

❖ In silico based anchoring delivered 79% of the physical map integrated to the corresponding genetic maps.

❖ Recombination rate and the gene distribution was estimated that confirmed the general pattern in large grass genomes.

❖ We were the last chromosome that started, however, we are the sixth to be finalized and reported so far!

❖ This physical map will serve as a framework for sequencing of this wheat chromosome

❖ And is of immediate use for map-based isolation of agronomically important genes/QTL located on this chromosome.

Thank you so much for your

kind attention!