





Case work of scaffolding (TaaCsp6BLhA183B09 (CTG356, 9 assemblies))



# Scaffolding of individual BAC clones

#### Case work of scaffolding (TaaCsp6BLhA183B09 (CTG356, 9 assemblies))

# Step 1. Identify both ends of each BAC clone

- from summary of result of assemble.
  based on result of SSPACE with mate-pair library.
- 3. by reading BAC end sequence actually.



Step 2. Decide order and orientation of internal assemblie	s
Construction of mate-pair library and sequencing	
BAC clones (7.573clones)	
library: Nextera mate-pair sample prep kit (Illumina) fragment size: <b>3kh, 9kh</b> Scale of pooled library: <b>96 clones/Tag, 24Tags/Run</b>	>scoffold91 size37287 tigs1 fTA.TosCsp68LhA195D28_contig88882
Sequencing system: MiSeq reg. Ver.2 (250x2 cycles) Software for scaffolding: SSPACE	>ecaffold921size458431tigs2 rTA,TaaCsp60LhA165L14_contig00005 2339 fTA,TaaCsp60LhA165L14_contig00003
Example of scaffolding (183B09/CTG356, 9 assembled sequence contigs) Output of SSPACE	acaffold93 size90185 tigs7_ 
1	001A.TeaCsp68LhA103089.Lcont.ig80889 1808 1A.TeaCsp68LhA183899.Lcont.ig80899 1A.TeaCsp68LhA183899.Lcont.ig80898 001A.TeaCsp68LhA183899.Lcont.ig80888
secoffold#8 size3#23 t3 t1gel	TA. ToaCap68LhA183899 cont. 1089885 1619 TA. ToaCap68LhA103899 cont. 1089892
contigi contigé contigé contigé contigé contigé	Section 6 (1944) 512636949 11 (gri F TA. Tostopfelluki28115, contrig09882

Case work of scaffolding (TaaCsp6BLhA183B09 (CTG356, 9 assemblies))

Advantage of

**Our Sequencing System** 

-- To escape misassembly due to

multiple repeat sequence --

## Step 3. Identification of overlapped sequence on both ends of BAC clone



## Overlap check of neighboring BAC clones

Case work of scaffolding (TaaCsp6BLhA183B09 (CTG356, 9 assemblies))



	055	OBL	whole	
BAC clones with assembled sequence	2,719	4,061	6,780	
Overlapped site of BAC clones on the physical map	2,478	3,782	6,260	
Overlapped site confirmed by BAC sequences	2,406	3,681	6,087	
Conformability (%)	97.09	97.33	97.24	



BAC clone; TaaCsp6BShA0079D16, Threshold ; 1,000

Detected with miropeats 2.02 (http://www.littlest.co.uk/software/bioinf/old\_packages/miropeats/)





Contig ; CTG171, size 1.08 Mb, including TaaCsp6BShA0079D16, Threshold; 1,000

Distribution of Repeated sequences 3



Contig ; CTG171, size 1.08 Mb, including TaaCsp6BShA0079D16, Threshold; 150

### Constructed Pseudomolecule (ver. 1.0) of chromosome 6B of wheat (Chinese spring)

	Short arm	Long arm	Whole
Clones	2,846	4,240	7,086
(unmapped clones)	(165)	(179)	(344)
Scafolds			
a <b>ll</b> (> 0.5kb)	3,774	3,861	7,635
>1.5kb	2,483	2,908	5,391
>10.0kb	1,580	1,698	3,278
Ave. No. Sca./BAC	1.33	0.91	1.08
Total	302,926,340	384,629,617	687,555,957
coverage	72.99%	77.23%	75.31%
average	83,102	105,489	94,394
max	2,810,317	3,140,367	3,140,367
min	500	500	500
N50	403,001	526,758	459,245
	(186th)	(195th)	(403th)

# **Pseudomolecule Construction**

## Why the coverage of 6BS is low



We found map-Contigs on the border of Nor region in 6BS. But no map-contigs corresponding to the internal part of Nor region.

## Estimation of length for Nor region in 6BS.





Condition of qPCR Marker, TNAC8298 (Control), S5.8, S18, S26 Instulment; LightCycler® Nano (Roche) Regents; KAPA Library Quantification Kits

### Number of rDNA sequence in wheat genome

		Chinese Spring	N1BT1A	N1BT1D
Nullisomic chr. 18S 26S		-	1B	1B
		7,970	5,508	5,152
		6,357	4,037	3,748
	5.8S	3,859	2,493	2,164
NOR		3,859	2,493	2,164
	Length (Mb)	36.895	23.831	20.690

The coverage of chromosome 6B

No.	6BS	6BL	Whole	
1 Estimated chromosome length (Mb)	415.00	498.00	913.00	
2 Physical map (Mb)	331.34	459.54	790.88	
3 Pseudomolecle (ver 1.0 (Mb)	) 282.46	365.84	648.30	Nor-re
The ratio of 1 and 2 (The ratio of Mapping)	79.84%	92.28%	86.62%	
The ratio of 2 and 3 (The ratio of <mark>Sequencing</mark> )	85.25%	79.61%	81.97%	]
The ratio of 1 and 3	68.06%	73.46%	71.01%	

The coverage of 6BS is less than 70% because of Genomic regions carrying the rDNA genes (Nor-region). (red circle) Over 80% region of Chr. 6B was sequenced based on estimates length in construction physical map. (blue square)



#### Summary

- Overlap check was completed; 97.86% of 6,260 sites (6,780 clones) was correctly overlapped.
  - This result indicates one of advantages of our sequencing system (individual BAC sequencing with the 96-tag) for the wheat genome rich in repetitive sequences.
- Mate-pair libraries of MTP BAC clones were constructed, sequenced and used for scaffolding of the assembled sequences of each BAC clone.
   Manual curation (checking overlapped sequences and mapping mate-pair reads of BAC clones) was essential and effective for construction of pseudomolecule.
- The pseudomolecule (ver. 1.0) of wheat chromosome 6B was constructed.
  This pseudomolecule is made up of 7086 BAC clones.
  The number of scaffolds for each BAC clone is 1.02 (ave.)
  The maximum size of scaffold is about 3.1 Mb.
- The pseudomolecule of wheat chromosome 6B will be further improved.
  Add Illumina short reads to improve sequence accuracy.
  Search BAC clones covering *Nor*-region to improve sequence coverage.