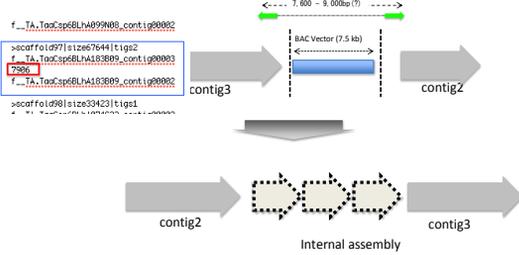




Case work of scaffolding (TaaCsp6BLhA183B09 (CTG356, 9 assemblies))

**Step 1. Identify both ends of each BAC clone**

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



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

**Step 2. Decide order and orientation of internal assemblies**

Construction of mate-pair library and sequencing

BAC clones (7,573clones)

Library: Nextera mate-pair sample prep kit (Illumina)  
 fragment size: **3kb, 9kb**  
 Scale of pooled library: **96 clones/Tag, 24Tags/Run**  
 Sequencing system: MiSeq reg. Ver.2 (250x2 cycles)  
 Software for scaffolding: **SSPACE**

Example of scaffolding (183B09/CTG356, 9 assembled sequence contigs)

Output of SSPACE

f...TA.TaaCsp6BLhA099908\_cont1.gp0002

scaffol04971size76441t1gs2

f...TA.TaaCsp6BLhA183809\_cont1.gp0003

7906

f...TA.TaaCsp6BLhA183809\_cont1.gp0002

scaffol04981size334231t1gs1

contig1

contig6

contig9

contig4

contig8

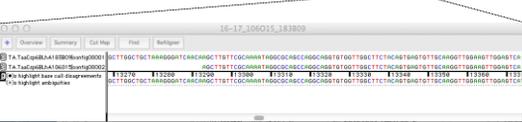
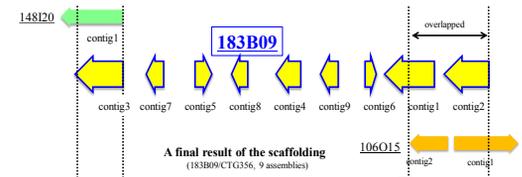
contig5

contig7

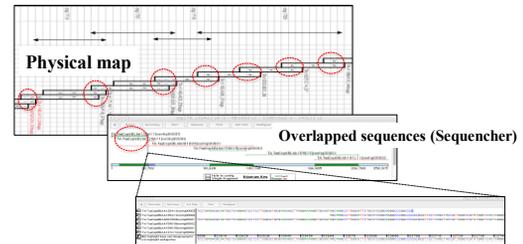


Case work of scaffolding (TaaCsp6BLhA183B09 (CTG356, 9 assemblies))

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



Overlap check of neighboring BAC clones

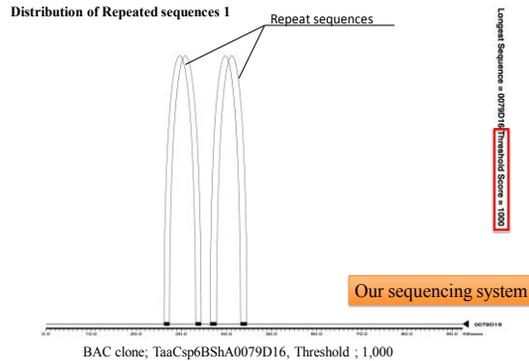


Result of overlap check between neighboring BAC clones

	6BS	6BL	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

**Advantage of Our Sequencing System**

- To escape misassembly due to multiple repeat sequence -



Detected with mirapeats 2.02 ([http://www.litltest.co.uk/software/bioinf/old\\_packages/mirapeats/](http://www.litltest.co.uk/software/bioinf/old_packages/mirapeats/))



## 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	Pseudomolecule (ver.1.0) (Mb)	282.46	365.84	648.30 <i>Nor-region?</i>
	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 Sequencing)	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.

## Contributors



## Acknowledgements



This work was supported by the grants from the Ministry of Agriculture, Forestry and Fisheries of Japan (NGB1003) and funding from Nissin Flour Milling Inc..