BAC-pool sequencing of the physical map from wheat chromosome-arm 3DS

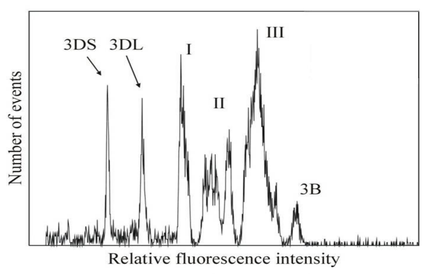
Kateřina Cviková[1], Jan Bartoš[1], Hanka Šimková[1], Jan Šafář[1], Nils Stein[2], Federica Cattonaro[3], Hélène Bergès[4], Klaus Mayer[5], Jaroslav Doležel[1]

1) Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Šlechtitelů 31, 783 71 Olomouc, Czech Republic
2) Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Corrensstrasse 3, 06466 Gatersleben, Germany
3) Instituto di Genomica Applicata, Via J. Linussio 51, 33100 Udine, Italy
4) INRA, National Resources Centre for Plant Genomics, F-31326 Castanet Tolosan Cedex, France
5) Institute for Bioinformatics and Systems Biology (MIPS), Helmholtz Center f. Health and Environment, Ingolstadter Landstr. 1, D-85764 Neuherberg, Germany

BACKGROUND

Construction of a physical map using BAC library from individual chromosome or chromosome-arm after flow-sorting is an important step towards sequncing the complex genome of hexaploid wheat. For sequencing, an order and true orientation of contigs is necessary to avoid distortions on reference sequence. Usually, screening BAC library through pools with available and newly developed molecular markers is used for contig anchoring. We coupled screening of BAC library pools and next-generation sequencing to anchor and organize contigs of the physical map in a high-throughput manner. Minimum Tilling Path of the physical map of wheat chromosome-arm 3DS was used for preparation of pools that were sequenced by Illumina Hiseq2000. Reads gained with this sequencing technology were aligned to reference sequences represented by *in silico* markers. BAC addresses of positive clones were deconvoluted from positive pools for each sequence and over 21,000 sequences of intra- and inter- specific origin were anchored to the physical map. These sequences were used for contig ordering supported by genetic maps and collinearity with related grass species.

BAC library prepared from flow-sorted chromosome arm

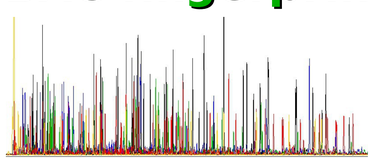


Histogram of relative fluorescence intensity obtained from 3DS/3DL double ditelosomic line of wheat cultivar Chinese Spring

- Size of 3DS: 321 Mb
- BAC library parameters

  Number of clones: 36,864
  Average insert size: 110 kb
  Chromosome coverage: 11.0x

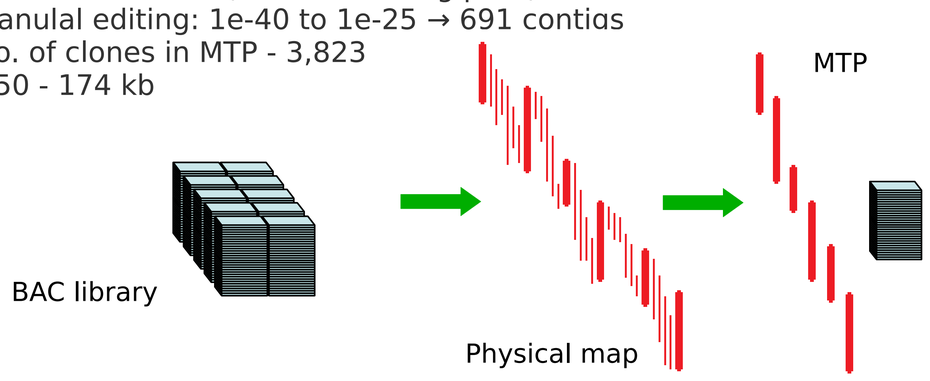
BAC fingerprinting using SnaPshot method



- 27,880 (75.6%) high qualiy fingerprints after contamination removal

Contig assembly by FPC (FingerPrinted Contigs software)

- 19,002 useful fingerprints
- Cutoff 1e-75 to 1e-45 → 1,360 contigs
- Selection of **MTP** (Minimum tilling path)
- Manulal editing: 1e-40 to 1e-25 → 691 contigs
- No. of clones in MTP - 3,823
- N50 - 174 kb



Sequencing of 3D pools of MTP

- Isolation of 3-dimensional pools (10 plate, 16 row and 24 column pools)
- 50 pools from 3,823 BACs were sequenced by Illumina HiSEQ2000

Reference sequences for anchoring

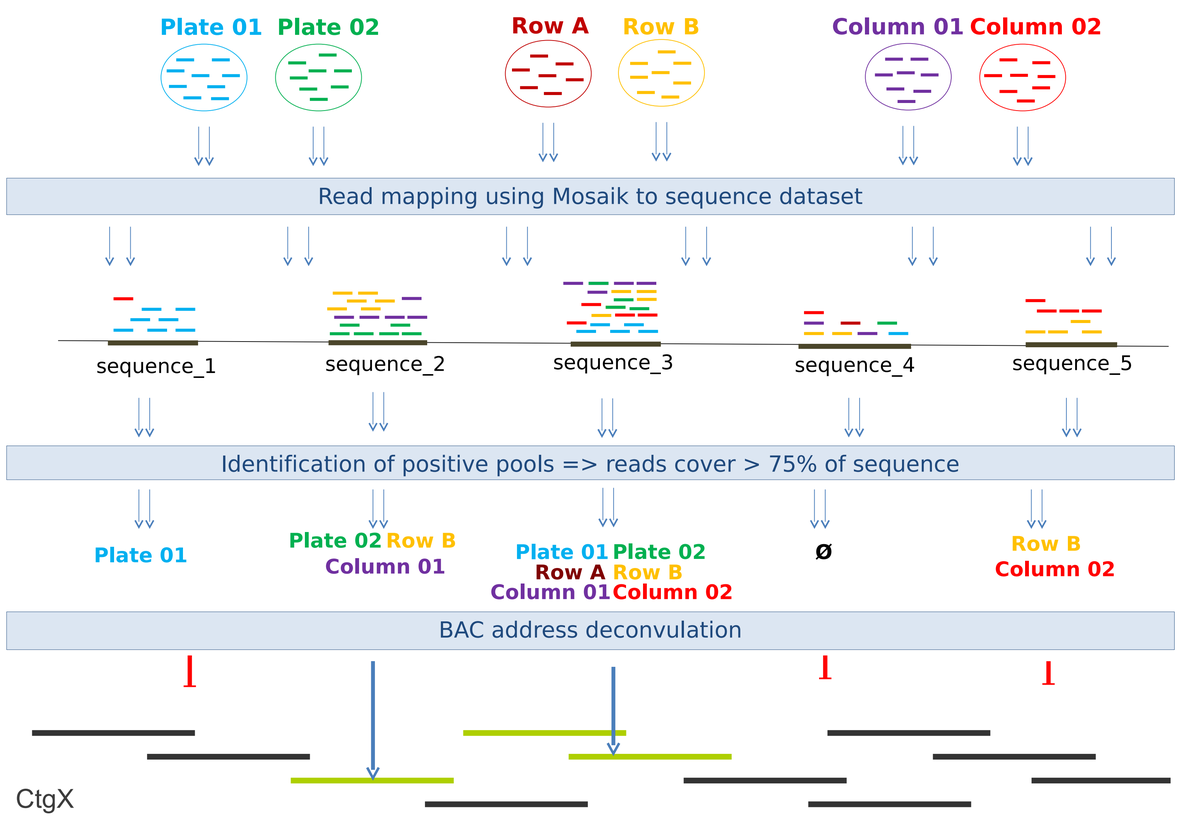
1) Sequence contigs
   - Chromosome arm 3DS was sequenced with low coverage (3x) using Roche 454 GS-FLX system
   - 2,081,761 reads were assembled using Newbler 2.5.3 to gain 46,297contigs (42 Mbp), which were used for anchoring

2) GenomeZipper
   - Gene fragments in 454 reads were identified and ordered virtually according to collinearity with Brachypodium, rice and Sorghum syntenic genome regions
   - 594 genes at collinear possition with related genomes were found and used for anchoring

3) SNP sequences
   - 7,152 sequences mapped in *Ae.tauschii* were used for anchoring

Read alignment & BAC address deconvulation

- Reads of individual BAC pools were aligned using Mosaik 1.1.0021 to each sequence set described above separately
- Hash size, aligmnment candidate treshhold and number of mismatches were optimalized according to time of analysis and type of reference sequence
- Non-uniquely mapped reads were resolved and filtered
- Coverage statistics for each sequence in data set were calculated
- Reference sequences with coverage at least 75 % were used for further analysis
- Positive pools were deconvoluted using perl script and new markers were anchored to the physical map



CONCLUSION

- BAC library and physical map from chromosome arm 3DS were prepared
- We present new time-saving and cost-effective aproach for BAC library screening based on next-generation sequencing of 3-dimensional pools
- Three data sets were used as the reference and more than 21,000 markers were anchored to the BAC contigs:
  1) 20,899 markers from sequence contigs
  2) 224 from GenomeZipper
  3) 231 SNPs genetically mapped in *Ae. tauschii*

This work was supported by European Seventh Framework Project FP7-212019 and IGA PrF/2012/001