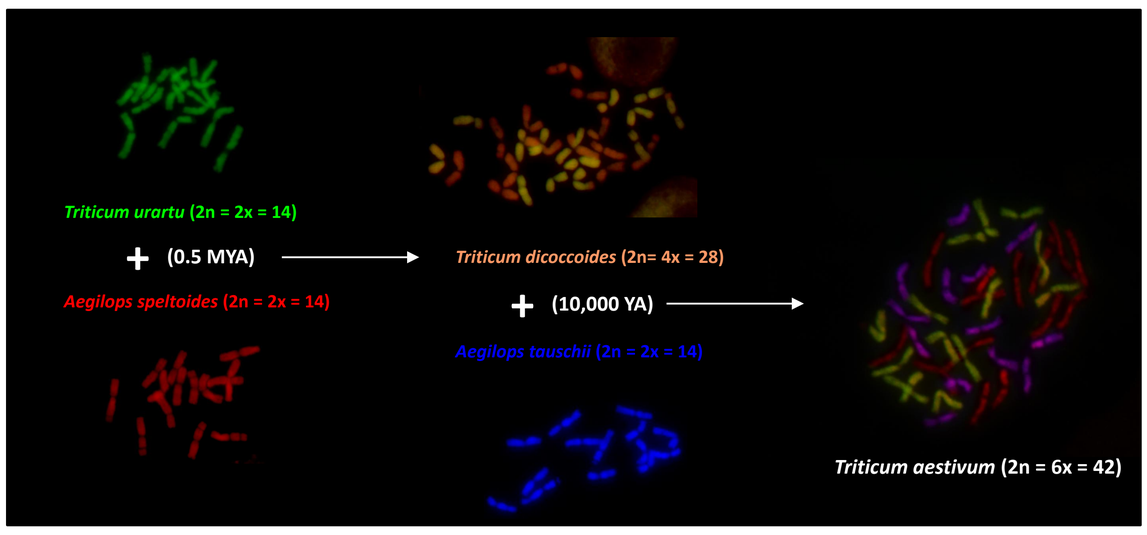
Comparative Sequence Analysis Reveals Similar Rates of Non-Collinear Gene Insertion in the B and D Genomes of Hexaploid Wheat

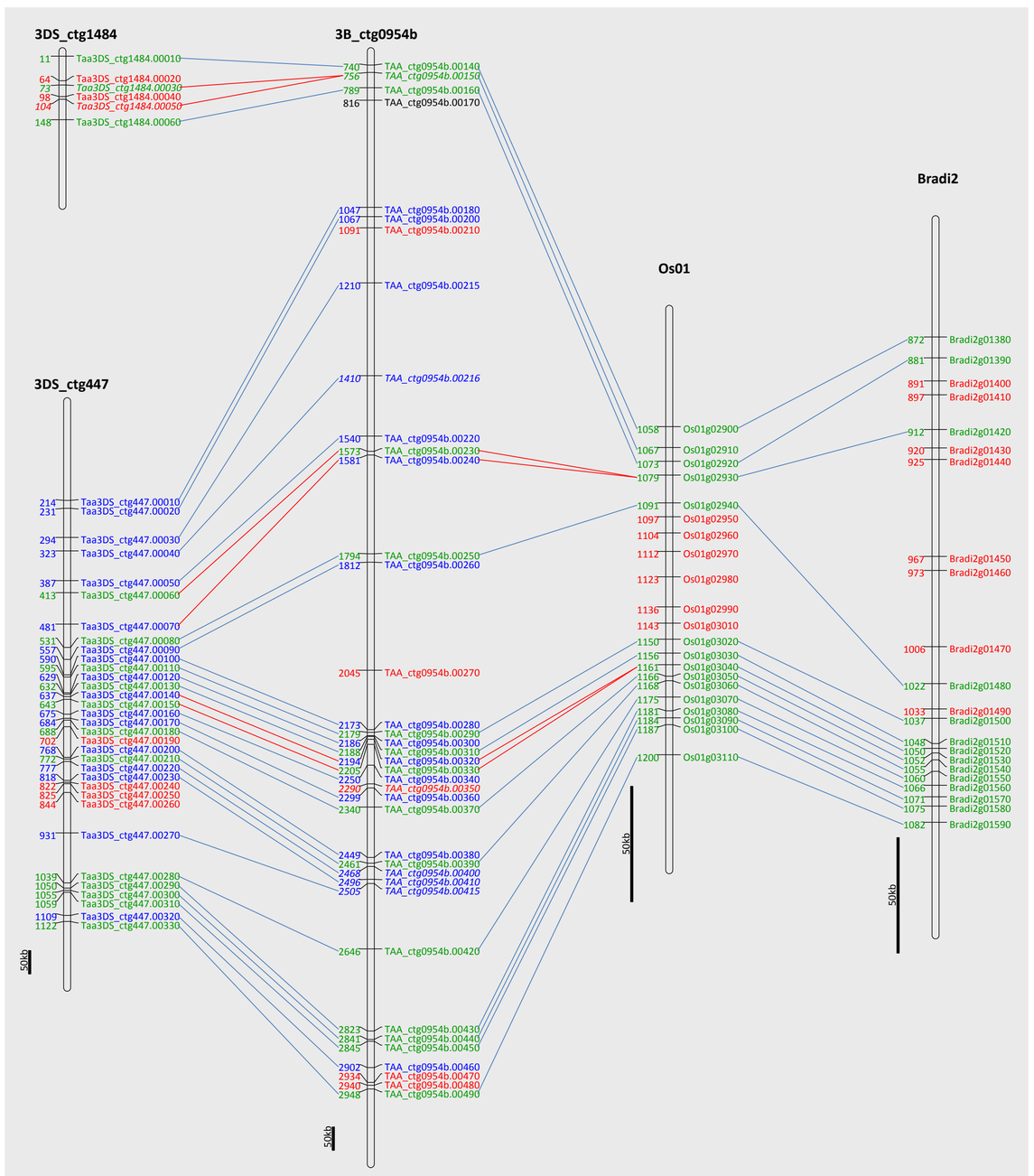
Jan Bartoš[1], Čestmír Vlček[2], Frédéric Choulet[3], Mária Džunková[2], Kateřina Cviková[1], Jan Šafář[1], Hana Šimková[1], Jan Pačes[2], Hynek Strnad[2], Pierre Sourdille[3], Hélène Bèrges[4], Federica Cattonaro[5], Catherine Feuillet[3] & Jaroslav Doležel[1]

[1] *Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Sokolovská 6, CZ-77200 Olomouc, Czech Republic*
[2] *Institute of Molecular Genetics, Vídeňská 1083, Praha, CZ-14220, Czech Republic*
[3] *INRA University Blaise Pascal UMR 1095 Genetics Diversity Ecophysiology of Cereals, F-63100 Clermont-Ferrand, France*
[4] *INRA, National Resources Centre for Plant Genomics, F-31326 Castanet Tolosan Cedex, France*
[5] *Instituto di Genomica Applicata, Via J. Linussio 51, I-33100 Udine, Italy*

Background

Polyploidization is considered one of the main mechanisms of plant genome evolution. The presence of multiple copies of the same gene reduces selection pressure and allows sub-functionalization and neo-functionalization leading to plant diversification, adaptation and speciation. In hexaploid wheat, polyploidization as well as the prevalence of transposable elements resulted in massive gene duplication and movement. As a result, the number of genes which are non-collinear to genomes of related species seems strongly increased. However, it is not clear whether rates of non-collinear gene insertion are similar or different among homoeologous genomes. We have identified and sequenced Hga locus (1.6 Mb) on chromosome 3DS corresponding to part of the contig ctg0954b on homoeologous chromosome 3BS and compared gene content, gene by gene.





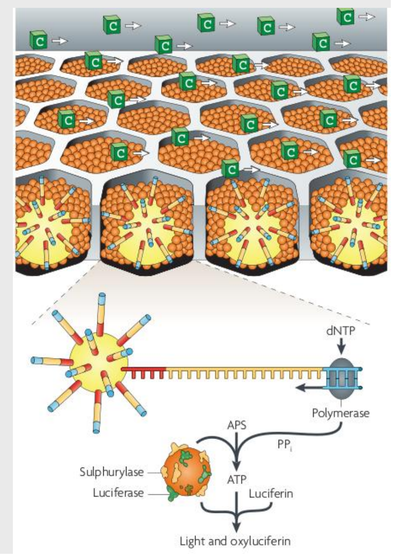
Comparison of the gene content at the Hga locus on 3DS with its homoeologous sequence on wheat chromosome 3B and homologous regions in rice (Os01) and Brachypodium (Bradi2) genome. Genes shared at the two wheat loci and at least one model genome at collinear positions are labelled in green. Genes shared between the homoeologous chromosomes 3D and 3B only are labelled in blue. Genes unique to one of the compared regions are labelled red. Pseudogenes and gene fragments are marked by an asterisk. Duplications are highlighted by red lines associating both copies with particular orthologous gene.

BAC clone selection and sequencing

BAC clones corresponding to 3B contig ctg0954b were selected from BAC library specific for chromosome arm 3DS using molecular markers developed from EST. By PCR screening, 8 contigs (represented by 24 clones) were identified.

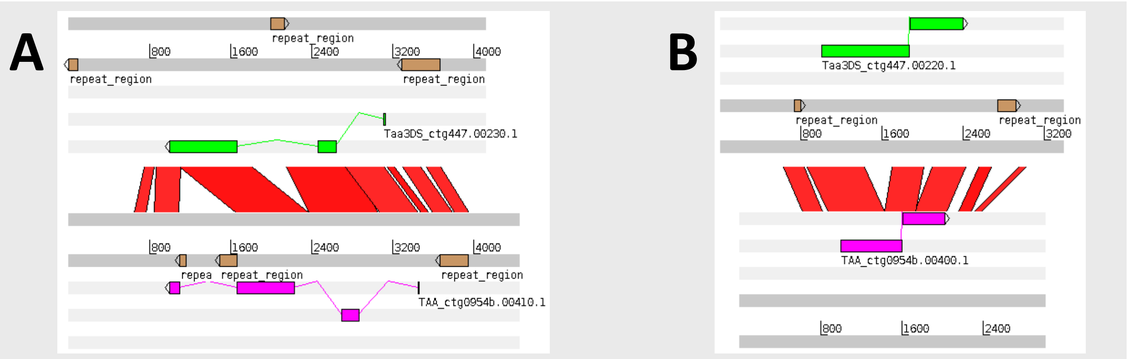
3-kb and 8-kb paired-end libraries were created from pooled DNA representing all 24 BAC clones. Libraries were prepared using 454-Roche Titanium standard paired-end sequencing kits and sequenced on 454-Roche GS FLX sequencer.

- 3 kb library: 278,412 reads; 50,780,080 bp
- 8 kb library: 387,392 reads; 82,540,639 bp
- coverage 88 x



Sequence assembly

- done by Newbler (version 2.0)
- 1 to 15 contigs per BAC; ordered in scaffolds
- 24 clones organized in two scaffolds only
- scaffold A: 1,264,820 bp; 39 gaps
- scaffold B: 333,768 bp; 21 gaps



Comparison of gene – pseudogene orthologous pairs. Intact coding sequences (all from the 3DS locus) are in upper trace highlighted in green colour. Pseudogenes (from 3B locus) are displayed in the lower trace in pink. Sequence features shown in brown represent repetitive elements. Red connections between traces represent Blastn hits between sequences. A) Insertion of repetitive element into the last exon of TAA_ctg0954b.00410 gene caused its pseudogenization. B) TAA_ctg0954b.00400 lost its function by partial deletion.

Conclusion

- Sequence assembly of 24 BAC clones resulted in two scaffolds 334 and 1,265 kb long
- Direct comparison with 3B homoeologous region revealed similar rate of non-collinear gene insertions
- Few pseudogenes were found among non-collinear genes
- Some evidences were found for gene erosion along 3B locus

This work was supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under the grant agreement FP7-212019 (Triticeae Genome) and by the Ministry of Education, Youth and Sports of the Czech Republic and the European Regional Development Fund (Operational Programme Research and Development for Innovations No. ED0007/01/01).