



HAL
open science

Sequencing, annotation and characterization of 17 Mb of chromosome 3B contigs provide novel insights into the wheat genome organization and evolution

Frédéric Choulet, Etienne Paux, Jerome J. Salse, Philippe Leroy, Ghislaine Magdelenat, Sylvie Bernard-Samain, Valérie Barbe, Catherine C. Feuillet

► To cite this version:

Frédéric Choulet, Etienne Paux, Jerome J. Salse, Philippe Leroy, Ghislaine Magdelenat, et al.. Sequencing, annotation and characterization of 17 Mb of chromosome 3B contigs provide novel insights into the wheat genome organization and evolution. 19. International Triticeae Mapping Initiative - 3rd COST Tritigen, 2009, Clermont-Ferrand, France. hal-02818223

HAL Id: hal-02818223

<https://hal.inrae.fr/hal-02818223>

Submitted on 6 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Abstract reference: ITMI2009_071

Sequencing, annotation and characterization of 17 Mb of chromosome 3B contigs provide novel insights into the wheat genome organization and evolution

Type of presentation: Oral presentation

Session: Session 4 [Genome Structure and Evolution]

Presenting author is member of COST Action Tritigen FAO604

Choulet Frédéric¹, Paux Etienne¹, Salse Jérôme¹, Leroy Philippe¹, Magdelenat Ghislaine², Samain Sylvie², Barbe Valérie², Feuillet Catherine¹

¹ INRA-UBP UMR 1095 Genetics, Diversity and Ecophysiology of Cereals 234 avenue du Brezet 63100 Clermont-Ferrand France, ² Génoscope, Institut de Génomique, CEA 2 rue Gaston Crémieux CP5706 91057 Evry France

Contact author:

Choulet Frédéric, fchoulet@clermont.inra.fr

Our current knowledge of the wheat genome structure is limited to the analysis of individual BAC sequences and of random BAC end and plasmid sequences. To improve our understanding of the composition and evolution of this large and complex genome, we have sequenced and completely annotated 136 BAC clones assembled into 12 contigs of 0.5 to 3.1 Mb in size originating from different regions of chromosome 3B. This first set of large wheat contiguous sequences account for 17 Mb and contains 128 genes, 21 pseudogenes and 14 gene relics as well as 2963 transposable elements, representing 83.5% of the sequences. The average gene density is of 1 gene/114 kb with important local variations ranging from 0 to 10 genes per BAC. Gene distribution is not random with 77% of the genes being clustered into several small islands containing 3 genes in average and a slight gradient of gene density towards the telomeric regions. Interestingly, all Mb-sized contigs carry genes irrespective of their chromosomal location. Only 32% of the BACs are devoid of genes and the largest gene-free region is 709 kb indicating that gene islands are small and scattered all along the chromosome, thus supporting the idea that the whole genome needs to be sequenced to ensure access to the entire gene space. Only half of the predicted wheat genes are orthologous and syntenic with rice chromosome 1. The other half consists of genes present in rice but on different chromosomes. Non syntenic genes are found in almost all (9/10) sequenced gene-containing regions with large variations depending on their composition and chromosomal location. The highest level of rearrangement was observed for a disease resistance region with 2/3 of the genes at non orthologous positions whereas other regions exhibit >90% of syntenic genes. We also precisely annotated the 2963 transposable elements and identified 271 different families, including 155 new ones (at least 2 members), and 356 unknown single copy elements. TE composition shows large variations along the chromosomes with some families strictly found in centromeric or telomeric regions. Compared to isolate BAC sequencing, analyses of Mb-sized contigs give new insights into the genome structure by revealing, for instance, that most of TEs are complete and fragmented by nested insertions that can reach more than 200 kb.

This project is supported by grants from the Agence Nationale de la Recherche (ANR-GPLA06001G SMART) and the Genoscope.