



HAL
open science

Sequencing and analysis of megabase-sized genomic regions from wheat

Frédéric Choulet, Thomas Wicker, Etienne Paux, Cyrille C. Saintenac, Pierre Sourdille, Beat Keller, Rudi Appels, Catherine C. Feuillet

► **To cite this version:**

Frédéric Choulet, Thomas Wicker, Etienne Paux, Cyrille C. Saintenac, Pierre Sourdille, et al.. Sequencing and analysis of megabase-sized genomic regions from wheat. 16. Plant and Animal Genome Conference, Jan 2008, San Diego, Californie, United States. hal-02820795

HAL Id: hal-02820795

<https://hal.inrae.fr/hal-02820795>

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.

Sequencing and analysis of megabase-sized genomic regions from wheat

Frédéric Choulet¹, Thomas Wicker², Etienne Paux¹, Cyrille Saintenac¹, Pierre Sourdille¹, Beat Keller², Rudi Appels³, Catherine Feuillet¹

¹ UMR 1095 INRA/UBP, Amélioration et Santé des Plantes, 63100 Clermont-Ferrand, France

² Institute of Plant Biology, University Zurich, Zollikerstrasse 107, 8008 Zurich, Switzerland

³ Molecular Plant Breeding CRC, State Agriculture Biotechnology Centre, Murdoch University, Western Australian Department of Agriculture and Food, South Street, Perth, WA 6150 Australia

Our current knowledge of the wheat genome composition is limited to the analysis of small segments (120-150 kb). Using the physical map of chromosome 3B, we have selected and sequenced two large BAC contigs that cover the *Rph7* (3.0 Mbp, 8x coverage) and *Sr2* (1.8 Mbp, 6x coverage) loci on the short chromosomal arm.

The *Sr2* locus has a high content (~85%) of transposable elements (TEs) which were the main source of problems during assembly and finishing. Interestingly, most gaps were found in very narrow and specific regions of some TE families and not, as was expected, caused by mis-assembly of different repeat units. Genic regions were mostly covered evenly and posed no major problems in the finishing phase. This indicates that the gene-containing portions of the wheat genome will be relatively easy to obtain but a complete finishing of TE-rich regions will often be necessary to confirm the integrity of large BAC contigs.

In contrast, with less than 70% of TEs, the *Rph7* locus is more gene-rich. To date, 41 putative genes have been predicted, representing an average gene density of 1 gene/75 kb. A particular effort was made in the development of bioinformatics tools dedicated to scaffolding, annotation and the automated design of markers. Genetic to physical map comparisons reveal a strong variation of the recombination rate between the two loci. This study provides new insights into the composition of the wheat genome and promotes the development of efficient tools and methods for sequencing and annotation.