



Sequence analyses of Megabase-sized contigs of the wheat genome

Frédéric Choulet, Philippe Leroy, Franck Giacomoni, Jerome J. Salse, Etienne Paux, Catherine C. Feuillet

► To cite this version:

Frédéric Choulet, Philippe Leroy, Franck Giacomoni, Jerome J. Salse, Etienne Paux, et al.. Sequence analyses of Megabase-sized contigs of the wheat genome. 2. Workshop on Triticum. COST Action FA604, Apr 2008, Albena, Bulgaria. hal-02818229

HAL Id: hal-02818229

<https://hal.inrae.fr/hal-02818229>

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.

Sequence analyses of Megabase-sized contigs of the wheat genome

Frédéric Choulet, Philippe Leroy, Franck Giacomoni, Jérôme Salse, Etienne Paux, Catherine Feuillet

UMR INRA/UBP 1095, Genetic Diversity and Ecophysiology of Cereals, F-63100 Clermont-Ferrand, France

fchoulet@clermont.inra.fr

With 17 Gb and an hexaploid composition, molecular characterization of the bread wheat (*Triticum aestivum* L.) genome is a challenge and its genome structure and composition remain largely unknown. Our current knowledge is limited to the analysis of individual BAC sequences and of random BAC end and plasmid sequences. To improve our knowledge on the structure and evolution of the wheat genome, we have sequenced a dozen of Mb-sized contigs (up to 3.2 Mb; 8-12x coverage) originating from different regions of chromosome 3B for which a physical map has been recently established in our laboratory. Altogether, these data account for 16.5 Mb of sequences and represent the first set of large contiguous wheat genome sequences to be analyzed.

Automatic annotation has been performed using the web-based TriAnnot pipeline (<http://urgi.versailles.inra.fr/projects/TriAnnot/>; EU-Auvergne Lifegrid project) that is under final stage of development. TriAnnot allows direct submission of BAC sequences and uses *Triticeae* specific libraries, algorithms and combiners to provide gene models. Similarity-based identification of the transposable elements is also computed and a more sophisticated approach for characterization of nested elements is under development. TriAnnot benefits from a user friendly interface and from a modular grid-based architecture.

The gene density over the 16.5 Mb is 1 gene per 139 kb. Locally, important variations of the gene content is detected, from 0 to 10 genes per BAC, revealing a non-random distribution of the genes along the chromosome. Intergenic region as small as 500 bps were observed while a region of 620 kb without genes has also been found. Curation of the TE annotation has been performed over the larger contig (3.2 Mb). With 60 different TE families detected in the region, it revealed that the wheat genome is characterized by an important diversity of TEs. However, a few number of elements are over-represented suggesting that wheat genome expansion has occurred by a massive amplification of a small number of specific TEs. Finally comparative analyses has been performed with other available cereal genomes. First comparisons reveal that, for some regions, wheat may have undergone many DNA rearrangements that led to numerous micro-synteny breaks when compared to the other cereal genomes that exhibit a more ancestral genome structure. The latest results on gene and TE annotation, on the architecture of the annotation pipeline and on comparative analyses will be presented.

This project is supported by the Agence Nationale de la Recherche (ANR-Genoplante-SMART).