Towards sequencing the 1 Gb wheat chromosome 3B

Frédéric Choulet, Etienne Paux, Philippe Leroy, Pierre Sourdille, Stéphane
Schlub, Dominique D. Brunel, Valérie Barbe, Patrick Wincker, Catherine C.
Feuillet

Towards sequencing the 1 Gb wheat chromosome 3B

Frédéric Choulet[1*], Etienne Paux[1], Philippe Leroy[1], Pierre Sourdille[1], Stéphane Schlub[2], Dominique Brunel[2], Valérie Barbe[3], Patrick Wincker[3], Catherine Feuillet[1]

[1]*INRA-UBP UMR 1095 Genetics, Diversity and Ecophysiology of Cereals, F-63100 Clermont-Ferrand, France*
[2]*INRA, UR1279 EPGV, CEA/IG/CNG, F-91057 Evry, France*
[3]*Génoscope, Institut de Génomique, CEA, F-91057 Evry, France*
* presenting author, frederic.choulet@clermont.inra.fr

* BACKGROUND: Wheat is the most widely grown crop worldwide but so far, the access to its genome sequence has been hampered by its size (17 Gb) and highly repetitive structure. We plan to sequence and annotate the largest bread wheat chromosome, 3B (1 Gb i.e. 2.5x the rice genome) using a hybrid approach combining Roche454 Titanium sequencing of BACs and Whole Chromosome Shotgun (WCS) with Illumina/Solexa GAII PE reads. In a pilot project, we sequenced and annotated 13 large contigs (18 Mb) from different regions of chromosome 3B and produced 2 Gb (2X coverage) of WCS by Solexa GAI.

* RESULTS: Genic regions represent 3% of the 18 Mb sample whereas 81% are composed of transposable elements (TEs) with the largest gene-free region being 709 kb. With 1 gene/104 kb on average, the genes are not randomly distributed and 75% of them are clustered into small islands that are scattered along the entire chromosome. This supports the idea that the whole genome needs to be sequenced to ensure access to the entire gene space. A Mathematically Defined Repeats (MDR) index was defined from the 2 Gb of Solexa reads and used to support the annotation and estimate the relative proportion of TEs.

* CONCLUSIONS: This pilot project allowed us to get new insights into the wheat genome organization and establish novel bioinformatics tools and methods in preparation of the sequencing and annotation of chromosome 3B (2009-2010). This will lay the foundation for in-depth analyses of the structure, function and evolution of the wheat genome and the development of state-of-the art tools for genomics breeding in wheat.