

of the wheat genome (cv. Chinese Spring) that is anchored to the genetic and phenotypic maps to provide high resolution linkages between the traits and the underlying variations in sequence and polymorphisms. A first goal has been to obtain physical maps of the individual chromosomes/arms. To anchor the physical contigs, survey sequences were achieved using NGS technologies for a majority of the Chinese Spring chromosomes. The IWGSC then launched a short-term initiative to provide survey sequences of all 21 chromosomes. Sponsored by industry and government partners, the aim is to generate sequence and virtual order for most wheat genes. The sequences generated for each chromosome arm are assembled using the latest software tools. A first pass annotation is implemented over the draft assembly sequences where comparative genomics and colinearity with other grass genomes is used to derive a virtual gene order with an account of non-syntenic genes and pseudogenes. As sequences are generated independently for each chromosome, a second aim of this project is to characterize genomic variation between gene homoeologues (ie orthologous genes placed in different subgenomes), regulatory elements, and repeat content. An update of this initiative will be presented.

Poster 3. TriAnnot: a high performance pipeline for the automated structural and functional annotation of plant genomes - new developments.

Philippe Leroy ¹, Nicolas Guilhot ¹, Isabelle Lesur-Kupin ⁴, Patricia Faivre ⁵, Sébastien Theil ¹, Frédéric Choulet ¹, Hiroaki Sakai ³, Michael Alaux ², Takeshi Itoh ³, Hadi Quesneville ², Christophe Plomion ⁴, and Catherine Feuillet ¹.

¹ INRA-UBP UMR1095, Clermont-Ferrand, France; ² INRA UR1164, Versailles, France; ³ NIAS, Tsukuba, Japan; ⁴ INRA UMR1202, Cestas, France; and ⁵ INRA UMR1165, Evry, France.

A versatile, easy-to-use, online, automated annotation pipeline, TriAnnot (Leroy et al. 2012. *Frontiers in Plant Sciences* 3:1-14; <http://www.clermont.inra.fr/triannot>), has been developed under the umbrella of the International Wheat Genome Sequencing Consortium, the TriticeaeGenome, and 3BSEQ projects to obtain a reference sequence of the bread wheat genome. Its modular architecture allows for the annotation and masking of transposable elements, the structural and functional annotation of protein-coding genes with an evidence-based quality indexing, and identification of conserved noncoding sequences and molecular markers. The performance of TriAnnot was evaluated in terms of sensitivity and specificity using curated reference sequence sets from rice (IRGSP build5, August 2010) and wheat (Choulet et al. 2010. *Plant Cell* 22:1686-1701). In less than 8 hours, TriAnnot was able to predict more than 83% of the 3,748 CDS from rice chromosome 1 with a fitness of 67.4%. On a set of 12 reference Mb-sized contigs from wheat chromosome 3B, TriAnnot predicted and annotated 93.3% of the genes among which 54% were perfectly identified in accordance with the reference annotation. It also allowed the curation of 12 genes based on new biological evidences, increasing the percentage of perfect gene predictions to 63%. TriAnnot systematically showed a higher fitness than other annotation pipelines that are not improved for wheat. The TriAnnot pipeline is parallelized on a 712 CPU computing cluster that can run a 1-Gb sequence annotation in 26 hours. It is accessible through a web interface for small-scale analyses and/or through a server for large scale annotations. For the later, the pipeline is launched automatically using a PERL script. After completion, the structural and functional annotation can be viewed through an online GBrowse and can be manually curated using Artemis and GenomeView graphical editors. As it is easily adaptable to the annotation of other plant genomes, TriAnnot is currently improved for other plant genomes annotations such as barley, rice, maize, and oak species, and should become a useful resource for the annotation of large and complex genomes in the future. Release 3.5 is on-line since January 2012, and a new release 3.6 is underway.

Poster 4. Current status of physical mapping on wheat chromosome 6B.

F. Kobayashi ¹, S. Katagiri ¹, W. Karasawa ¹, Y. Hanawa ¹, S. Kaneko ², S. Nasuda ², K. Hayakawa ³, H. Fujisawa ¹, Y. Ito ¹, Y. Mukai ¹, J. Dolezel ⁴, T. Matsumoto ¹, J. Wu ¹, and H. Handa ¹.

¹ National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan; ² Laboratory of Plant Genetics, Graduate School of Agriculture, Kyoto University, Kyoto, Japan; ³ Nisshin Flour Milling Inc, Tsukuba, Ibaraki, Japan; and ⁴ Institute of Experimental Botany, Olomouc, Czech Republic.

For the purpose of better understanding the genome structure in wheat and accelerating the development of DNA markers for gene isolations and future breeding, the Japanese research consortium, as a member of IWGSC, is now conducting a project for the physical mapping and genomic sequencing of Chinese Spring chromosome 6B (914 Mb). With a

sequence-based finger printing method (Whole Genome Profiling; Amplicon Express, KeyGene Inc.), physical maps for the short and long arms of chromosome 6B have been successfully established using the 6BS- and 6BL-specific BAC libraries, respectively, which consist of 2,667 and 1,842 BAC contigs. The estimated chromosomal coverage is more than 80% of both arms. Overlap analysis between the neighboring clones within the BAC contigs resulted in a total number of 5,079 and 4,889 MTP (Minimal Tiling Path) BACs on 6BS and 6BL, respectively. For the confirmation and the chromosomal assignment of the BAC contigs onto their corresponding genomic regions, we currently are developing a large number of 6B-specific DNA markers using the public marker resources, available EST databases, and the 6B survey sequence. To date, among 3,743 markers tested for their availability, 2,480 were found to be useful for the PCR screening of BACs, and 451 BAC contigs has been anchored by 627 markers to the specific genomic regions on chromosome 6B presumed on the basis of genetic maps, deletion maps and/or Genome Zipper. In parallel, a high-resolution, radiation hybrid map and a genetic map using recombinant inbred lines are under development for further improvement of physical maps. This work was supported by grants from the Ministry of Agriculture, Forestry and Fisheries of Japan (Genomics Agricultural Innovation; KGS-1001 and KGS-1003) and funding from Nisshin Flour Milling Inc.

Poster 5. The distribution, duplication, pseudogenization, and expression of noncollinear genes of wheat chromosome 3B.

Natasha Glover, Frédéric Choulet, Sébastien Theil, Josquin Daron, Lise Pingault, Etienne Paux, and Catherine Feuillet. INRA-University Blaise Pascal Joint Research Unit 1095 Genetics, Diversity and Ecophysiology of Cereals, Clermont-Ferrand 63100, France.

3BSeq is a large, ANR (Agence Nationale de la Recherche)-funded project developed under the umbrella of the IWGSC. It aims to sequence and characterize the largest wheat chromosome (3B) at the molecular level. The complete chromosome sequence is being produced by combining Roche/454 sequencing of BAC pools with whole chromosome shotgun using Illumina technology. Preliminary studies showed that about 50% of the genes identified in wheat are not collinear with the other grass genomes, raising several questions about their origin, fate, and expression in the wheat genome.

A set of 3,005 collinear genes and 2,432 noncollinear genes was defined using the annotation of chromosome 3B and orthologous genes from *Brachypodium*, rice, and sorghum. While waiting for the establishment of the 3B pseudomolecule, deletion bin information was used to determine the gene density per deletion bin and the relative gene order along the chromosome. Collinear genes show a 2-fold increase in gene density from the centromeric regions to the telomeric regions, whereas noncollinear genes have five-fold increase. In addition, noncollinear genes are more likely to be resulting from gene duplication events, with approximately 10% more of the noncollinear genes having duplicates on 3B in comparison to the collinear genes. The same pattern was found for pseudogenes. RNASeq data, collected from 15 different RNA samples, was used to determine that a higher percentage of collinear genes are expressed in comparison to noncollinear genes.

These results provide evidence that noncollinear genes have a higher probability than collinear genes to be duplicated, revert to pseudogenes, and subsequently become nonfunctional. The primary mechanism for this gene duplication and movement on 3B remains to be determined. We found 50% more CACTA superfamilies in the regions surrounding noncollinear genes than collinear genes, suggesting a major role for transposable elements in noncollinear gene movement. These results will shed some light on the evolutionary forces involved in the shuffling of the gene content on 3B and more generally in the accelerated evolution of the wheat genome. Functional analyses are underway to determine the fate of noncollinear duplicated genes in wheat.