All in all, more than one billion genotyping datapoints have been produced (Figure 5) and are currently being analyzed in the framework of the project to address BreedWheat's main challenge: breeding for economically and environmentally sustainable wheat varieties.

Figure 5. Genotyping data of 367 lines representing the worldwide diversity – a chromosome 3B region. Each line represents a wheat line and each column a SNP.

A part of the international effort

The BreedWheat project is tightly linked to other international initiatives all aiming at improving wheat production to keep pace with the human demand in the next decades.

For the whole community to benefit from BreedWheat's results, a part of the TaBW420K genotyping chip representing 280,000 SNPs has been made available and can be used by third parties for both research and breeding purposes. A genetic map comprising about 83,000 SNPs will also be released.

In addition, beside SNP discovery and genotyping, the BreedWheat project also aims at continuing the French effort towards the sequencing of the complex wheat genome. In close collaboration with the IWGSC, BreedWheat partners will sequence chromosome 1B and provide useful data, such as whole-genome genetic and linkage disequilibrium maps, to anchor and order other chromosome sequences.

For more information on this topic, please contact :
Etienne PAUX from INRA
etienne.paux@clermont.inra.fr

More information www.breedwheat.fr

Coordinator: **Dr. Jacques Le Gouis**, UMR GDEC - jacques.legouis@clermont.inra.fr

Project manager: **Emmanuelle Lagendijk**, INRA Transfert - emmanuelle.lagendijk@paris.inra.fr

Communication manager: **Grégoire-Yves Berthe**, Céréales Vallée
gregoire.berthe@cereales-vallee.org

This project receives funding from the French Government managed by the Research National Agency (ANR) in the framework of the Investments for the Future (ANR-10-BTBR-03), France Agrimer and the French Fund to support Plant Breeding (FSOV).

Photos : © P. Desray, S. Roche, E. Paux - INRA

High throughput SNP discovery and genotyping in bread wheat

Sept. 2014

Marker-assisted selection (MAS) can improve the accuracy of breeding by: (1) permitting selection for a trait based on a single plant; (2) facilitating selection and fixation of several genes at the same time; (3) enabling selection of dominant and recessive genes without applying phenotypic tests at each generation; and (4) maximizing or minimizing the diversity of the parents chosen for crossing. In wheat, the widespread application of MAS has long been limited by the lack of markers and cost-efficient genotyping platforms.

However, the recent advent of high-throughput sequencing and genotyping technologies led to a paradigm shift in SNP discovery and genotyping, therefore opening the way to efficient MAS in wheat.

One of the objectives of the BreedWheat project is to make use of the most up-to-date technologies to identify SNPs covering the whole genome and use them to genotype wheat lines to conduct genome-wide association studies, to characterize genetic resources and develop new breeding methodologies.

SNP discovery in the bread wheat genome

Because of its size (17,000 Mb), allohexaploid nature (ABD-genome) and high repeat content (more than 85%), the wheat genome is one of the most complex among plant species. As a consequence, SNP discovery has long been hampered in wheat.

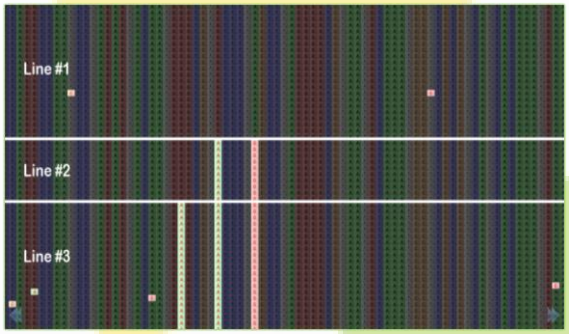

Figure 1. Sequence comparison for SNP discovery between three wheat lines. Each line represents a sequence, each column a nucleotide. Sequences are grouped by line and aligned on the reference sequence. SNPs are highlighted.

Indeed, SNP discovery relies on the comparison of sequences from the same region between several accessions in order to identify nucleotidic variations (Figure 1). In the recent years, technological advances have opened new perspectives and two main approaches can now be considered: (1) whole genome sequencing and (2) partial sequencing through random or targeted complexity reduction approaches.

In the framework of the BreedWheat project, these two approaches have been combined to identify a large number of SNPs in a set of selected lines corresponding to European elite material as well as more exotic lines.

Targeted resequencing was conducted through sequence capture of genic and intergenic regions. To mine SNPs in transposable elements, 275,000 Insertion Site-Based Polymorphism (ISBP) markers were targeted and resequenced in a set of 96 wheat accessions. More than 350,000 were identified. Similarly, 2,000 candidate genes potentially involved in traits such as yield or stress resistance, were resequenced in 16 lines, leading to the discovery of roughly 10,000 SNPs

In parallel, and in collaboration with the International Wheat Genome Sequencing Consortium (IWGSC), whole genome resequencing data from eight wheat varieties were used to mine for SNPs in both genic and intergenic regions. Eventually, more than three million SNPs were identified at the whole genome level, the largest SNP collection ever produced in wheat so far.

Towards a ultra-high throughput genotyping array in bread wheat

Concomitantly to the advent of SNPs, several genotyping technologies have been developed over the past years to allow for the scoring of one to several thousand markers simultaneously. Among these, the BreedWheat partners decided to work in collaboration with Affymetrix company to develop a high-throughput Axiom SNP array. With almost four million SNPs available, a selection was made to keep the best and more informative markers, based on several criteria, including polymorphism rate and distribution on the wheat chromosome (Figure 2).

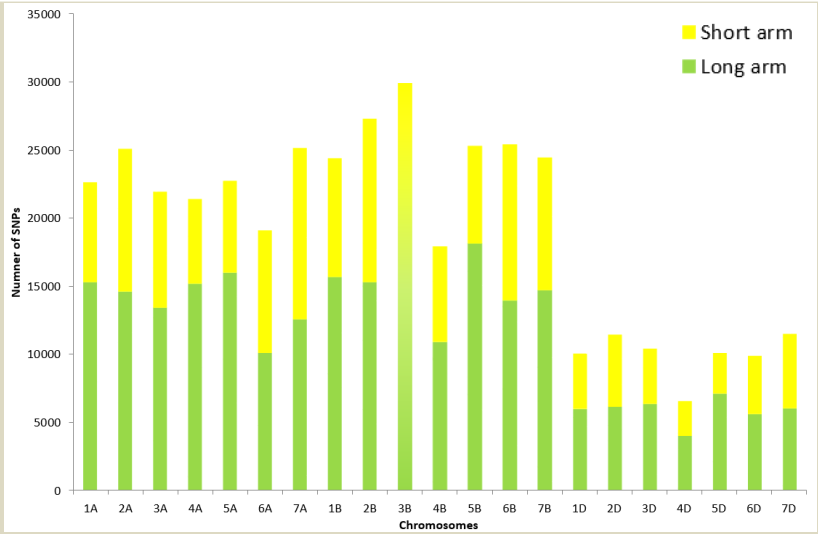

Figure 2. SNP distribution on the 21 bread wheat chromosomes. For chromosome 3B, SNPs are not assigned to a specific arm and the data are represented at the whole chromosome level.

To ensure the best genome coverage possible and avoid any bias, SNPs were selected in genes, intergenic non-repetitive as well as repetitive regions, as follows:

- 105,000 ISBP-derived SNPs
- 140,000 intergenic SNPs
- 150,000 genic SNPs, including 10,000 candidate gene-derived polymorphisms.

In addition, roughly 30,000 additional SNPs coming from the literature or brought to BreedWheat by partners or international initiatives as enabling technologies were added. These SNPs will allow to link different projects that have been led by the international wheat research community in the past years.

Altogether, a set of 423,385 SNPs was selected to design an ultra-high-throughput Axiom SNP array, one of the largest genotyping chip ever developed in wheat (Figure 3).

Figure 3. The BreedWheat Axiom 420K SNP chip (TaBW420K)

A billion datapoints produced and analysed

The TaBW420K Axiom array has been used in the framework of the BreedWheat project to genotype more than 7,000 wheat accessions comprising elite varieties as well as more exotic lines.

For genome-wide association studies, two panels were genotyped: (1) a 220-line elite variety panel and (2) 367 accessions representing the worldwide diversity. Combined with phenotypic data, these genotyping data allow for the identification of genomic regions involved in the genetic control of traits of interest such as yield, disease resistance or drought tolerance (Figure 4).

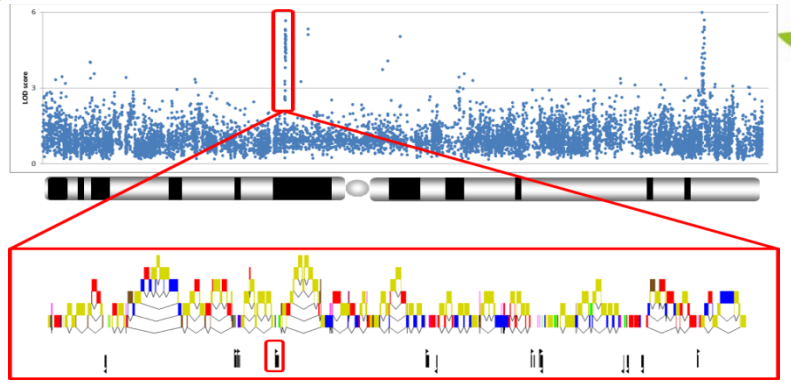

Figure 4. Genome-wide association study on dough quality – identification of a candidate gene on chromosome 3B.

In addition, the TaBW420K chip was used to genotype 4,600 out of 12,000 wheat accessions hosted at the genetic resources center of INRA Clermont-Ferrand. These data are currently being used to better characterize these exotic accessions in order to identify new and original alleles that will contribute to enrich the genetic diversity that was reduced in the elite wheat gene pool during domestication and selection.

Finally, the TaBW420K chip is also being used in the framework of a genomic selection. To this aim, a training population comprising about 500 lines has been genotyped to train the genomic selection prediction model, in combination with phenotyping data. This model will be used to estimate breeding value of a breeding panel of about 1,000 lines that has also been genotyped.