

# TriAnnotPipelineGRID, a tool for the automated annotation of wheat BAC sequences

*Tuesday, 12 February 2008 14:20 (20 minutes)*

A long term project of the IWGSC is to sequence the wheat genome to decipher the chromosomal location and biological function of all genes. This knowledge should enhance the understanding of the biology of the wheat plant and create a new paradigm for the improvement of this major crop. Because of the genetic and metabolic conservation among species in the grass family, efforts to decipher gene function in wheat and its close relatives will work synergistically with similar efforts in maize, rice, sorghum, barley and other grasses, for a global understanding of the function, structure, and evolution of the grass genomes.

The aim of the project is to provide a wheat automated annotation system to annotate new BAC sequences and regularly updating previous BAC annotations. It also provides a GBrowse graphical viewer. The pipeline integrates programs for prediction and analysis of protein-coding gene structure, as well as the identification of transposable elements and repeats in BAC sequences.

### 3. Impact

The project aims are:

1. Improving time calculation through the use of the grid technology. BLAST, RepeatMasker, HMMPFAM, est2genome, Gmap have been already adapted to the grid (AUVERGRID),
2. Store all the output within a CHADO data base for improving Gbrowse graphical display,
3. Adding new modules such as a gene modelling and new prediction tools

The project is performed through a strong collaboration between the INRA URGI bioinformatics platform at Evry, France; the NIAS at Tsukuba, Japan; The Broad Institut, US ; and Iowa State University

This program also benefits from a Genoplante project (GNPannot 2008-2010 WP5) to allow an on line BAC annotation curation using Apollo; A FP7 European project (TriticeaeGenome 2008-2010 WP5) to add the REPET pipeline for Transposable Element annotation; and a Generation Challenge Programme (2008-2009) for developing a web service on GreenPhyl (CIRAD).

### URL for further information:

<http://urgi.versailles.inra.fr/projects/TriAnnot/pipeline.php>

### 4. Conclusions / Future plans

The main goal of the TriAnnotPipelineGRID project is to provide to the scientific wheat community, and especially to the IWGSC, new resources for efficient BAC sequence analysis, as well as a platform for the re-annotation of BAC sequences as knowledge of the wheat genome sequence is increasing and new genes, transposable elements and new biological targets are continually identified.

### Provide a set of generic keywords that define your contribution (e.g. Data Management, Workflows, High Energy Physics)

wheat, genome sequencing, bac sequence analysis, grid, IWGSC, LifeGrid

### 1. Short overview

The main goal of the TriAnnotPipelineGrid project is to provide to the scientific wheat community, and especially to the IWGSC (International Wheat Genome Sequencing Consortium), new resources for efficient BAC sequence analysis, as well as a platform for the re-annotation of BAC sequences as knowledge of the wheat genome sequence is increasing and new genes, transposable elements and new biological targets are continually identified.

**Primary authors:** Mr REICHSTADT, Matthieu (Auvergrid); Mr LEROY, Philippe (INRA Clermont-Ferrand)

**Co-authors:** Mr CLAUDE, Alexandre (Lifegrid); Mr CHARPENTIER, Christian (INRA Clermont-Ferrand); Mr GIACOMONI, Franck (INRA Clermont-Ferrand); Mrs LIAUZU, Marianne (Lifegrid); Mr ALAUX, Michael (INRA URGI); Mr REBOUX, Sebastien (INRA URGI); Mr ITOH, Takeshi (National Institut of Agrobiological Sciences); Mr TANAKA, Tsuyoshi (National Institut of Agrobiological Sciences)

**Presenters:** Mr REICHSTADT, Matthieu (Auvergrid); Mr LEROY, Philippe (INRA Clermont-Ferrand)

**Session Classification:** Life Sciences

**Track Classification:** Application Porting and Deployment