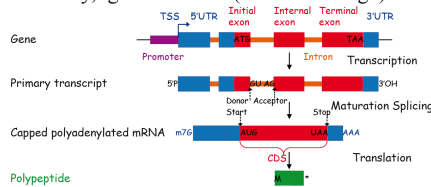


## Purpose:

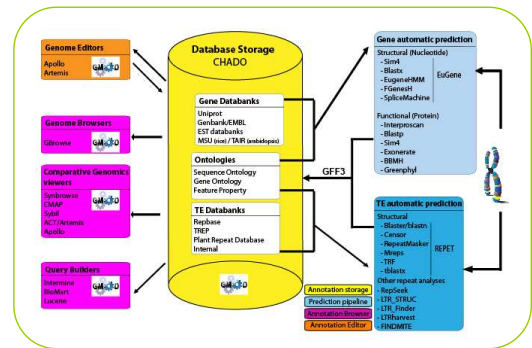
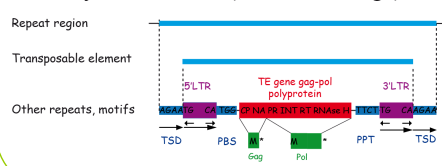
The *Musa* genome sequence will be released soon (MusaTract project<sup>3</sup>) and a robust system is being put in place and will contain in due course the *Musa* sequence predictions. We aim at developing the Community Annotation System (GNPAnnot project<sup>4</sup>) for structural and functional annotation of monocotyledon genomic sequences supported by results of comparative genomics. It will allow both automatic predictions and manual curation of genes and transposable elements.

## In theory, gene features (3 kb in average)



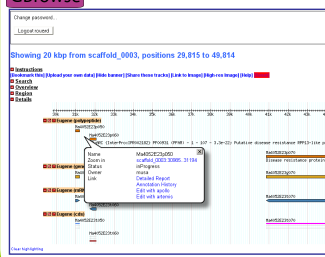
**Figure 1:** Examples of Genomic features that are predicted on genomic sequences (chromosome, BACs) using our pipeline of analyses. The main features to be annotated are genes and transposable elements. Genomic information and associated cross-references (UniProt, InterPro, ESTs, GO etc.) are inserted in a public database.

## In theory, TE features (10 kb in average)

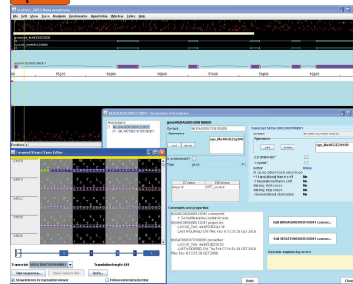


**Figure 2:** GNPAnnot system: diagram of the different components (principally GMOD) such as the database, viewers, annotation editors and pipelines of genomic object predictions (genes, transposable elements, SSR, ESTs etc.)

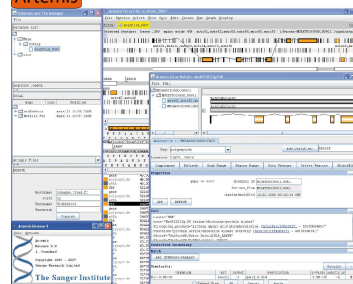
## GBrowse



## Apollo



## Artemis



## History



**Figure 3:** Genomic data can be visualized using a genome browser. By selecting a genomic feature, registered users can launch online editors (Apollo or Artemis) to edit for instance the boundaries of the genes or to add controlled vocabulary. History of annotation can be tracked down in the database to follow the several versions of annotation made by the same or by different annotators. Access to work in progress may be restricted up to the publication of data. Based on the analysis of 64 BACs (~7 Mbp), 1508 genes and 3140 TEs have been predicted (GCP Projects)<sup>5</sup>.

## Future collaborations:

The participation of the Global *musa* genomics consortium members will be needed to improve the automated annotation and to ensure that all the data is centralized in a single place. This would allow to release new versions of the genome annotation, similar to what is done by others consortia (e.g. TAIR - The Arabidopsis Information Resource).

## References:

- <http://gnpannot.musagenomics.org>
- <http://www.gnpannot.org/>
- <http://gmod.org>

## Contacts

Mathieu Rouard [m.rouard@cgiar.org](mailto:m.rouard@cgiar.org)  
Stéphanie Sidibe-bocs [stephanie.sidibe-bocs@cirad.fr](mailto:stephanie.sidibe-bocs@cirad.fr)

<sup>3</sup> ANR-08-GENM-V019 2009-2010

<sup>4</sup> This work was supported by The French National Research Agency, grant number ANR-07-GPLA-004 (2008-2010)

<sup>5</sup> *Musa* genome frame-map construction and connection with the rice sequence (GCP-SP2-project #15)