



**Structural & Comparative  
Genomics in Bread Wheat**  
**TriAnnotPipeline**  
**A LifeGrid Project based on  
AUVERGRID**

**3rd EGEE  
User Forum**

**February 12th, 2008**

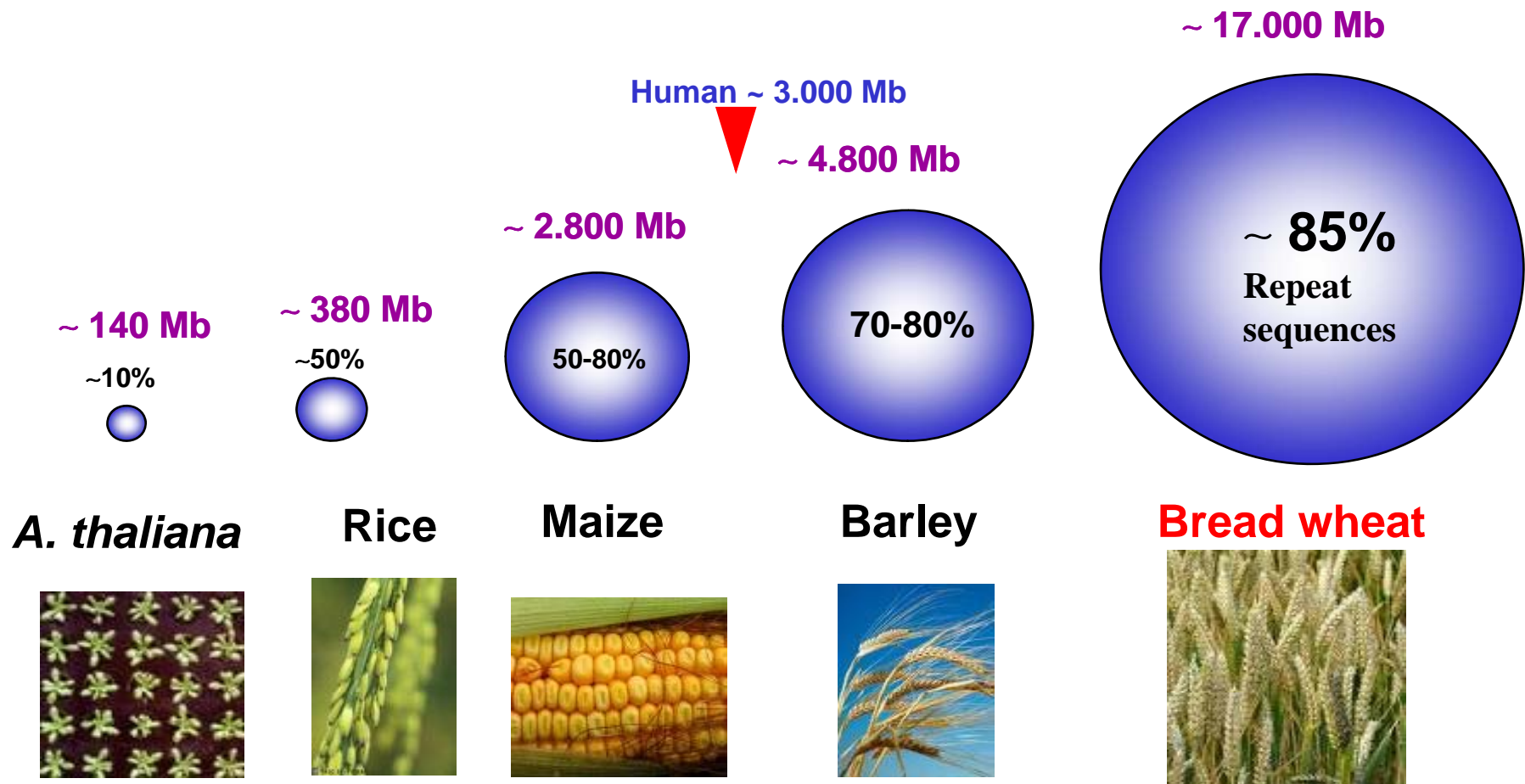
**F. Giacomoni, M. Reichstadt, P. Leroy**

*Génétique, Diversité & Ecophysiologie des Céréales - Clermont-Ferrand, France*



# Wheat as a challenge for Genomics

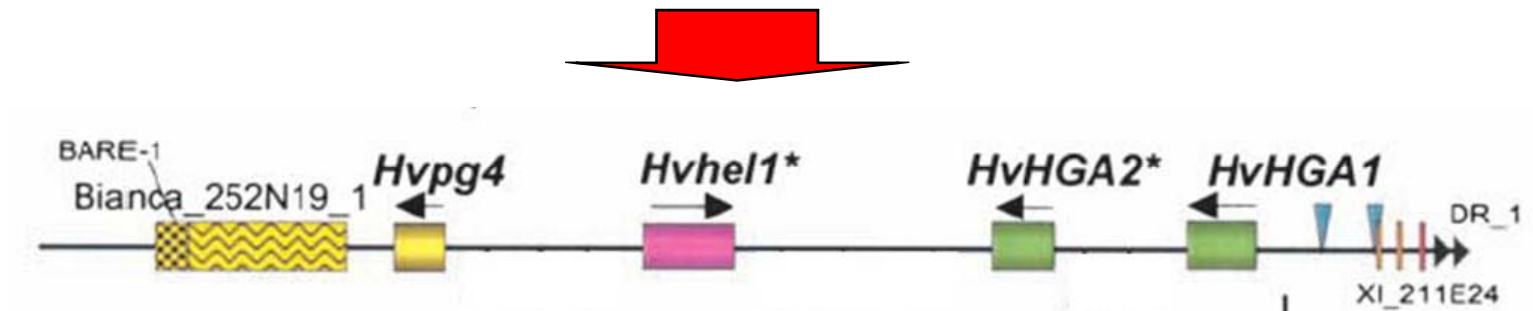
- Important Economic Crop
- Large Genome size



# I.N.R.A. Work on the Wheat Genome

- Sequencing
- Annotating
  - Discover Genes
  - Find Transposable Elements
  - Study other biological components

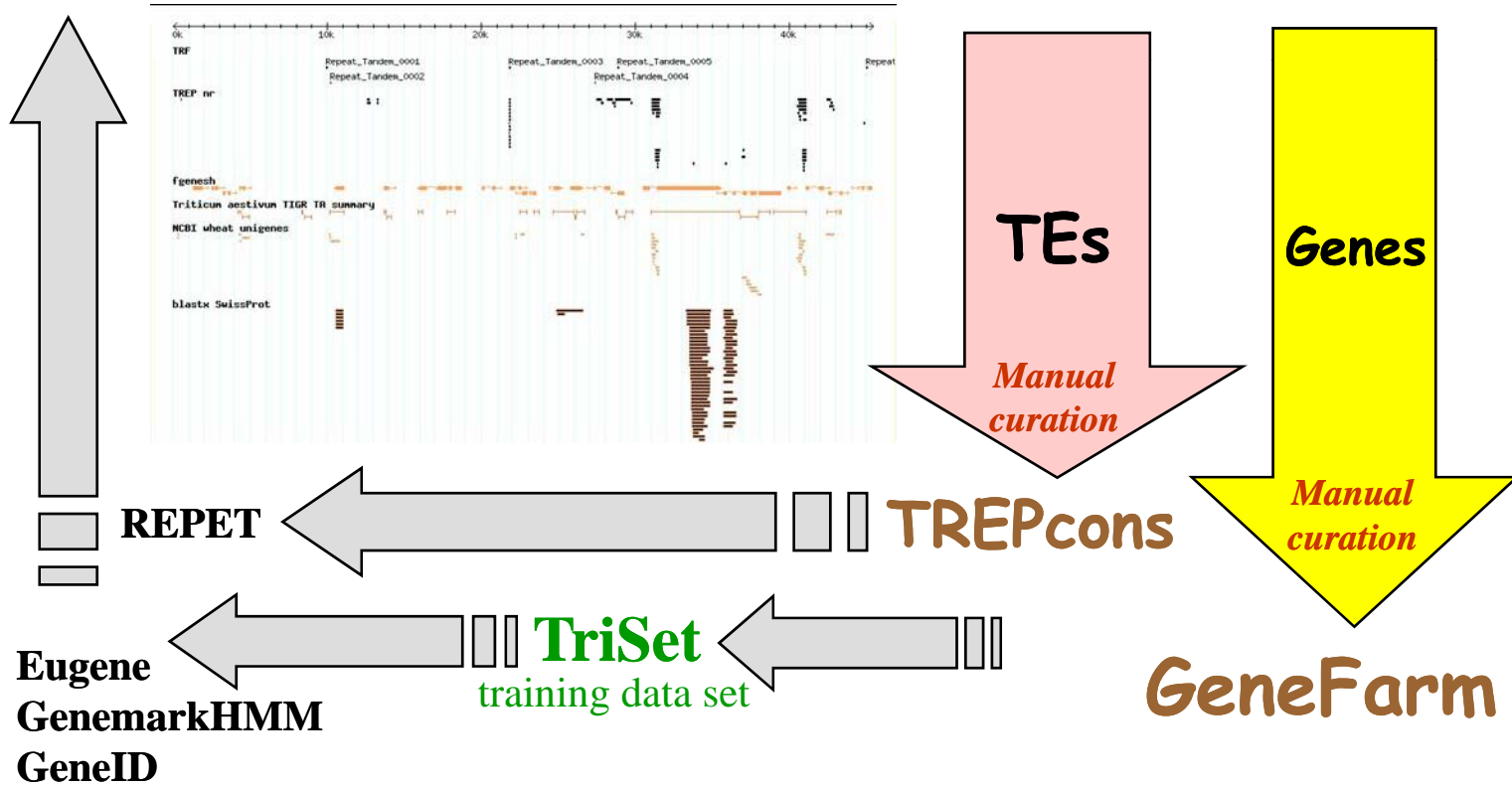
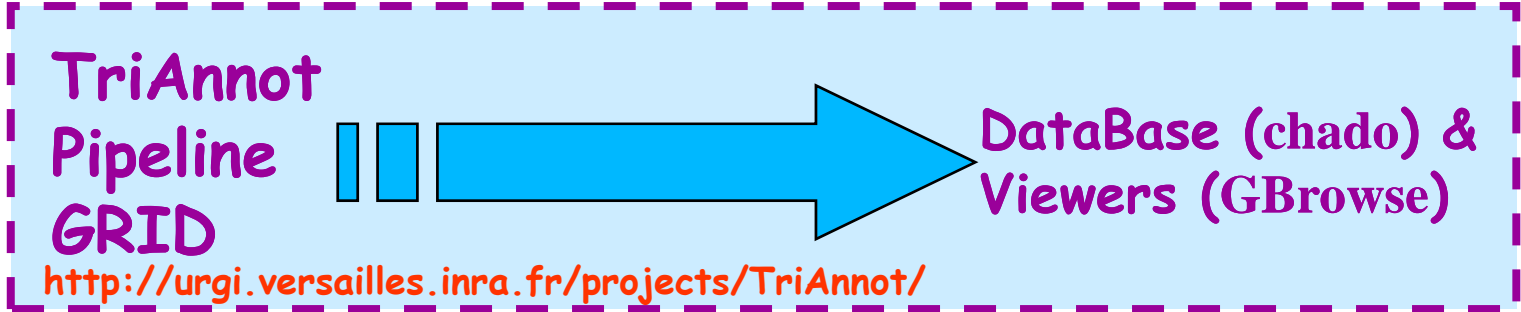
AAAATCGATATAGAGTATGTAGACAAATTTTAAACCCGGGGGAGAGAGAGA **DNA sequence**



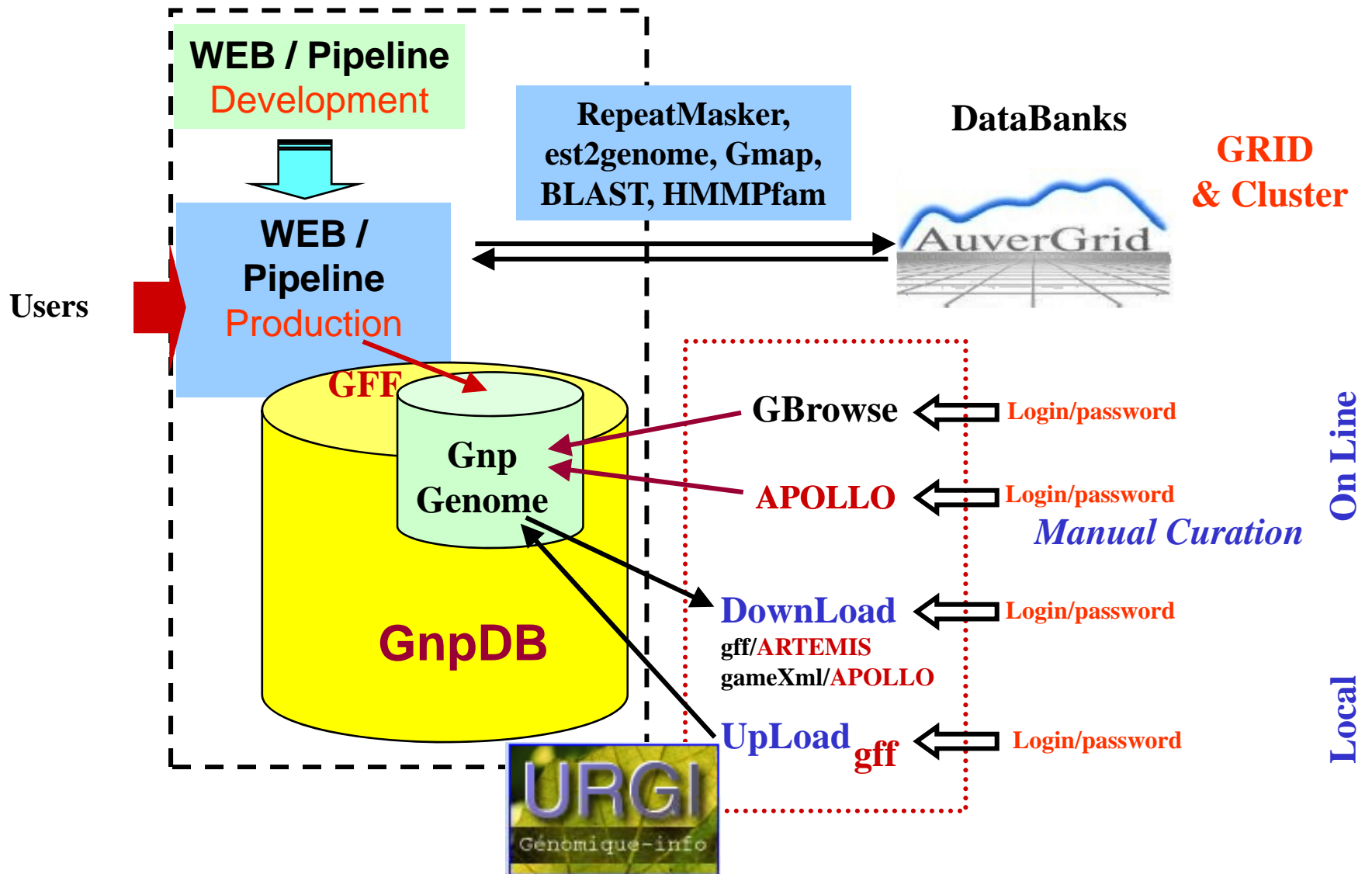
**Results after Annotation of the DNA Sequence**

# General Pipeline Structure of TriAnnot

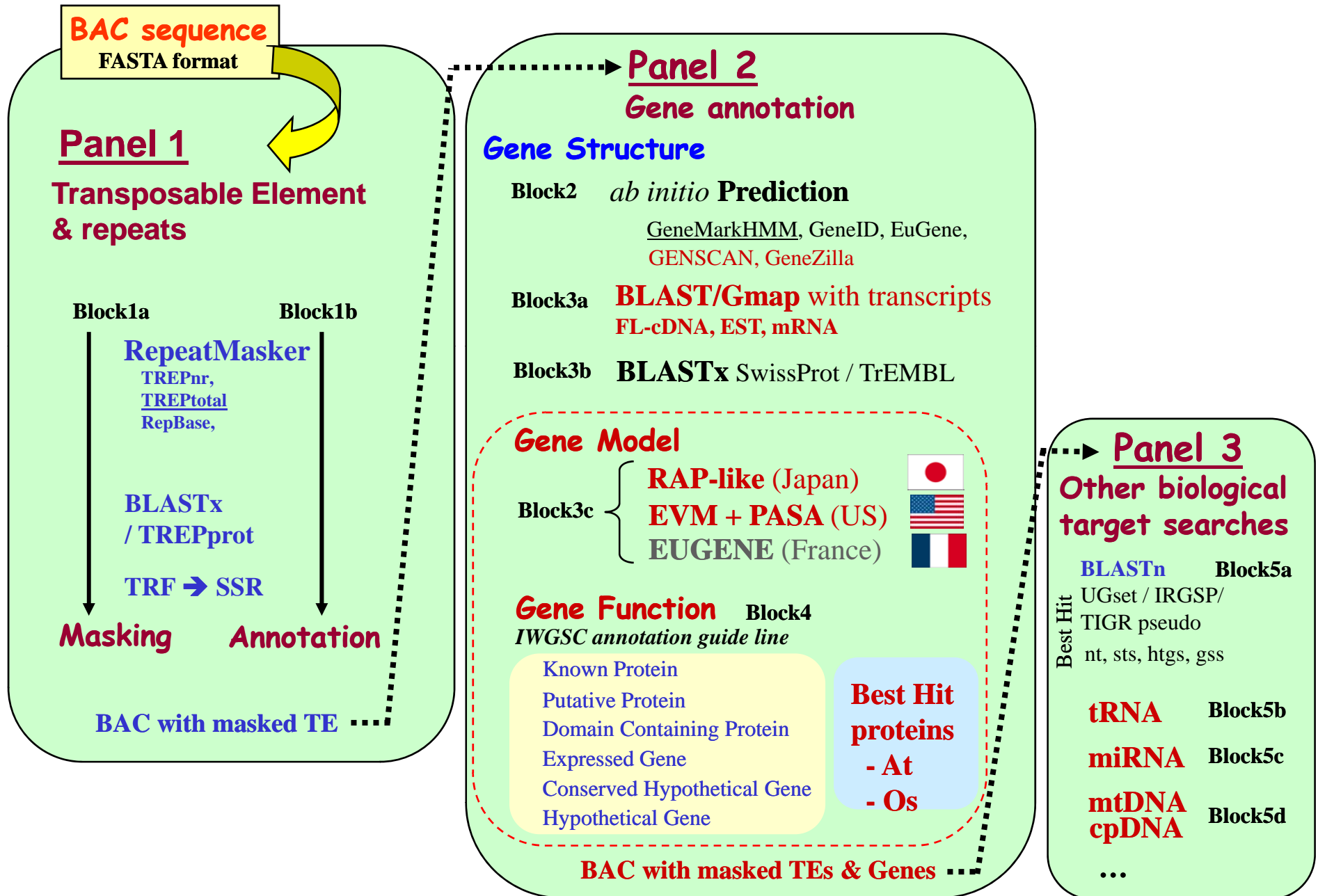
DNA  
sequences



# TriAnnotPipelineGRID Architecture



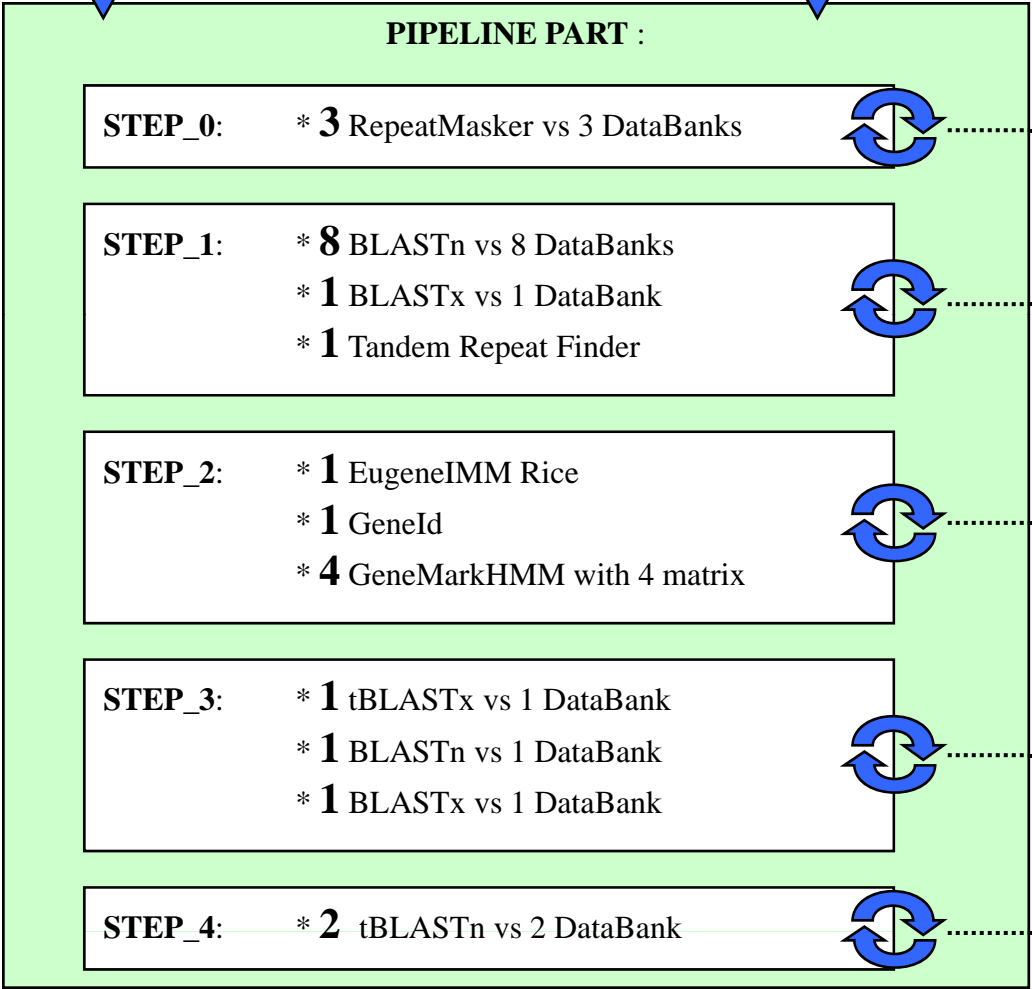
# TriAnnotPipelineGRID Detailed Architecture



**WEB INTERFACE PART with:**  
Upload of BAC FASTA format sequence  
Programming parameters of the Annotation with 5 blocks

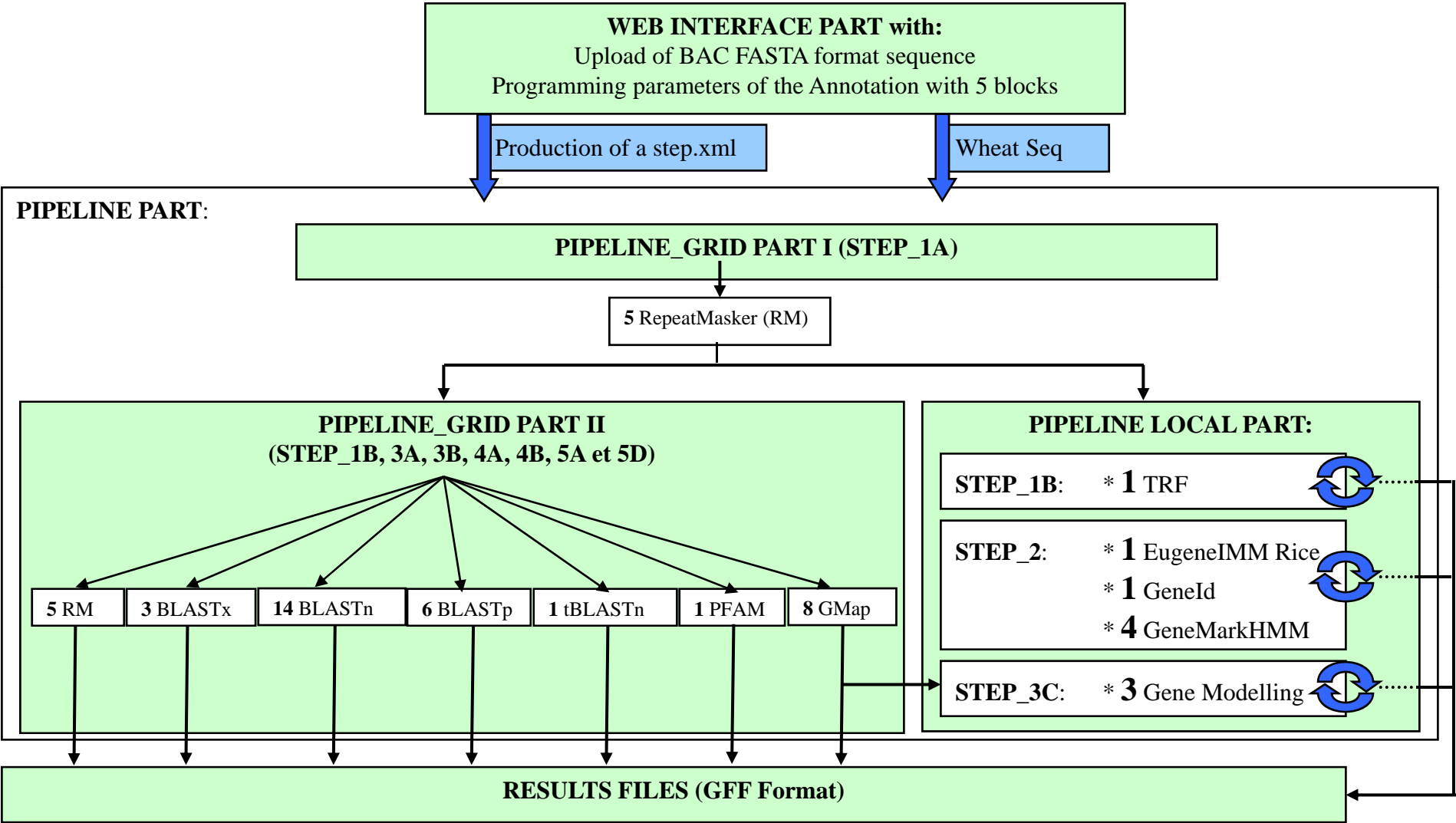
Production of a step.xml

Wheat Seq



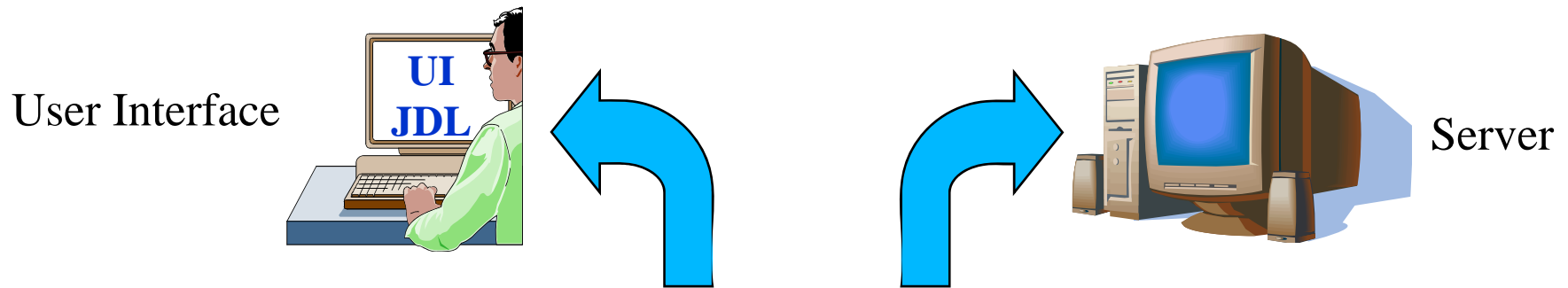
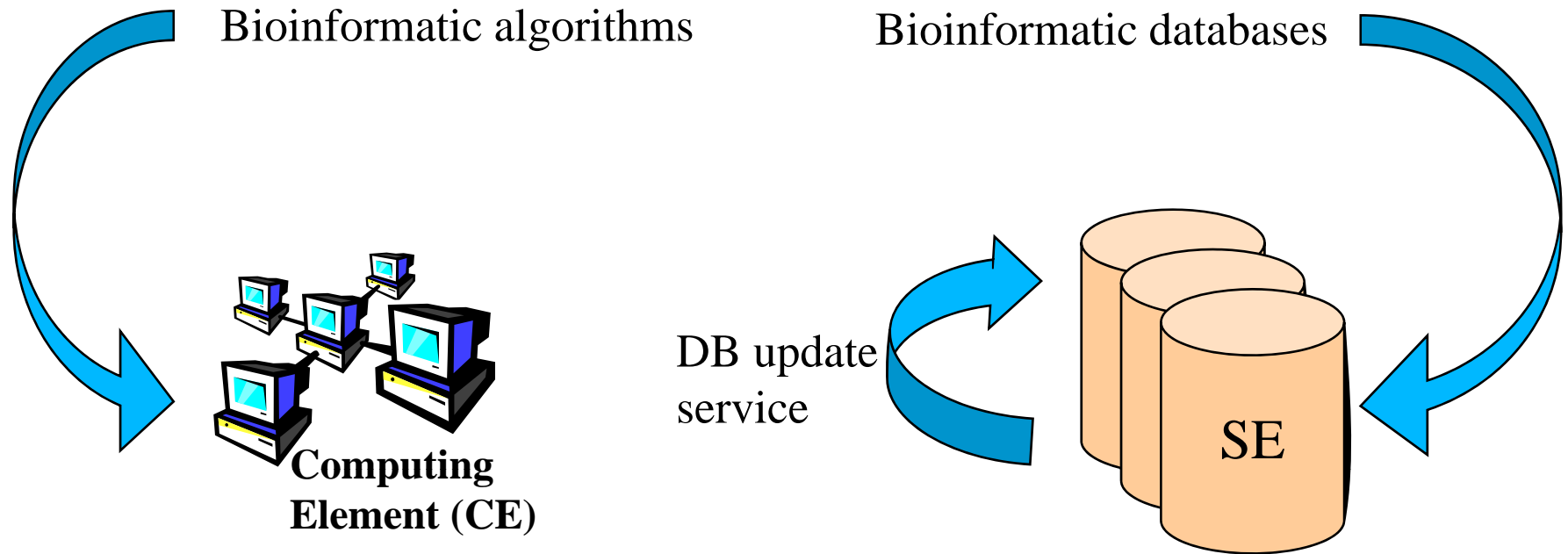
**RESULTS FILES (GFF Format)**

# TriAnnotPipelineGRID Architecture





# Bioinformatic algorithms

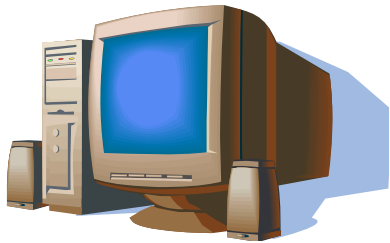


Grid part      Server part

Bioinformatic package

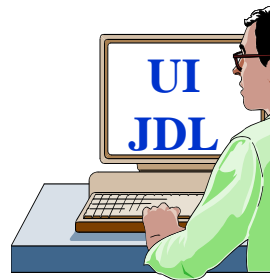
# Bioinformatic algorithms

## Server



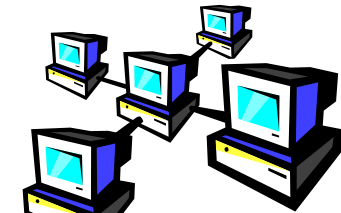
Get the parameter  
Create the XML step file  
Get the input (sequence) file  
Create the grid environment  
(JDL, shellscripts)  
Mask the repeated sequences  
RepeatMasker/Blast/  
GMap/HMMer  
Retrieve the output  
Fill the database

## UI



Get the parameter  
Create the XML step file  
Get the input (sequence) file  
Create the grid environment  
(JDL, shellscripts)  
Mask the repeated sequences  
RepeatMasker/Blast/  
GMap/HMMer  
Retrieve the output  
Fill the database

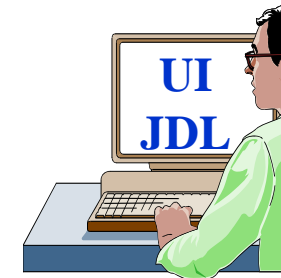
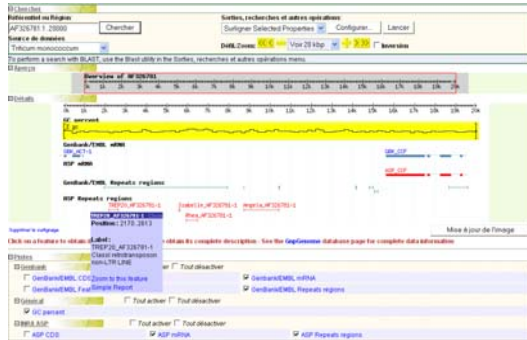
## CE



## Computing Element (CE)

Get the parameter  
Create the XML step file  
Get the input (sequence) file  
Create the grid environment  
(JDL, shellscripts)  
Mask the repeated sequences  
RepeatMasker/Blast/  
GMap/HMMer  
Retrieve the output  
Fill the database

# Bioinformatic algorithms



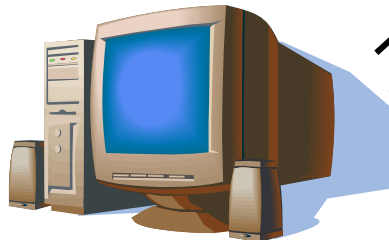
4-Creation environment

1-Parameters  
+ input file

3-copy  
input files

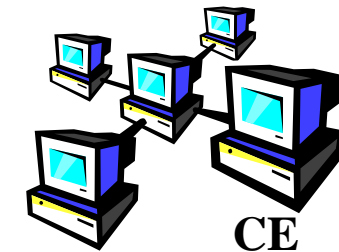
5-job submission

7- job output



2-Creation XML file  
9-DB filling

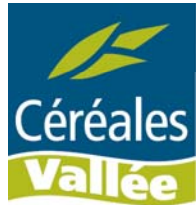
8-output transfer



6-job running (BLAST/  
HMMer/RepeatMasker/GMap)

2007-2008

## TriAnnotPipelineGRID Partners



**F. Giacomoni**  
**C. Charpentier**  
**N. Guilhot**  
**F. Choulet**  
**P. Leroy**  
**C. Feuillet**

**M. Reichstadt**  
**A. Claude**  
**M. Liauzu**  
**A. Mahul**

**M. Alaux**  
**T. Flutre**  
**I. Blanc-Lenfle**  
**S. Reboux**  
**H. Quesneville**



**T. Tanaka**  
**H. Ikawa**  
**H. Numa**  
**T. Itoh**



**F. Legeai**



**B. Haas**



**B. Kronmiller**

