

TriAnnot

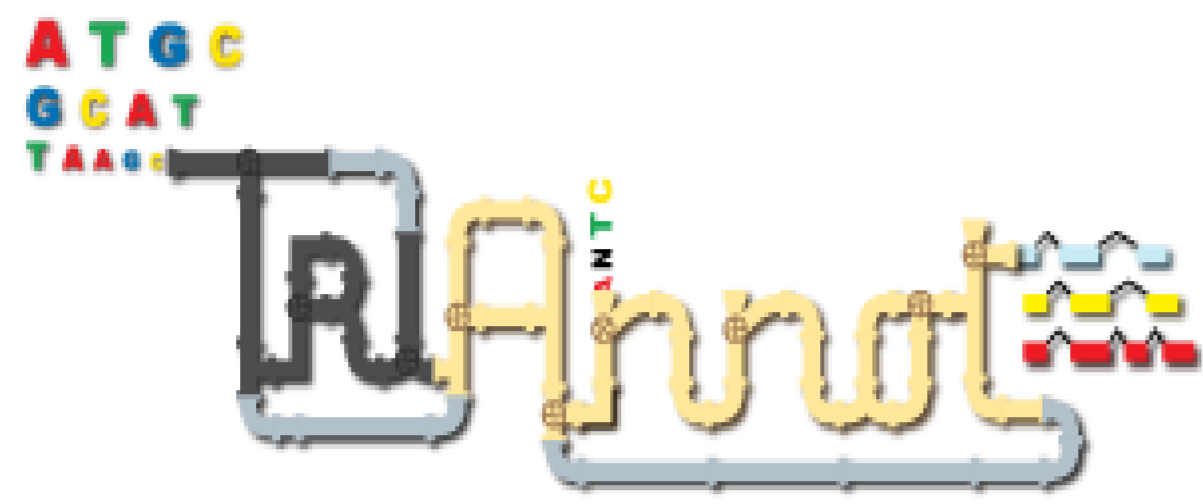
A high performance pipeline for the automated structural and functional annotation of plant genome – new developments

Philippe Leroy¹, Nicolas Guilhot¹, Sébastien Theil¹, Frédéric Choulet¹, Hiroaki Sakai³, Michael Alaux², Takeshi Itoh³, Hadi Quesneville², Catherine Feuillet¹

¹Inra, UBP-UMR1095, Clermont-Ferrand, France. ²Inra UR1164, Versailles, France. ³NIAS, Tsukuba, Japan.

Web: <http://www.clermont.inra.fr/triannot>

Contact: triannot-support@clermont.inra.fr



The TriAnnot pipeline offers an integrated environment allowing structural and functional annotations of plant genomes:

- * runs a variety of *ab initio* gene prediction programs,
- * displays gene models with a color code system to assess the confidence of the model (5 categories)
- * gives a functional annotation with 6 levels (classes) of confidence.

1. Load genomic sequence

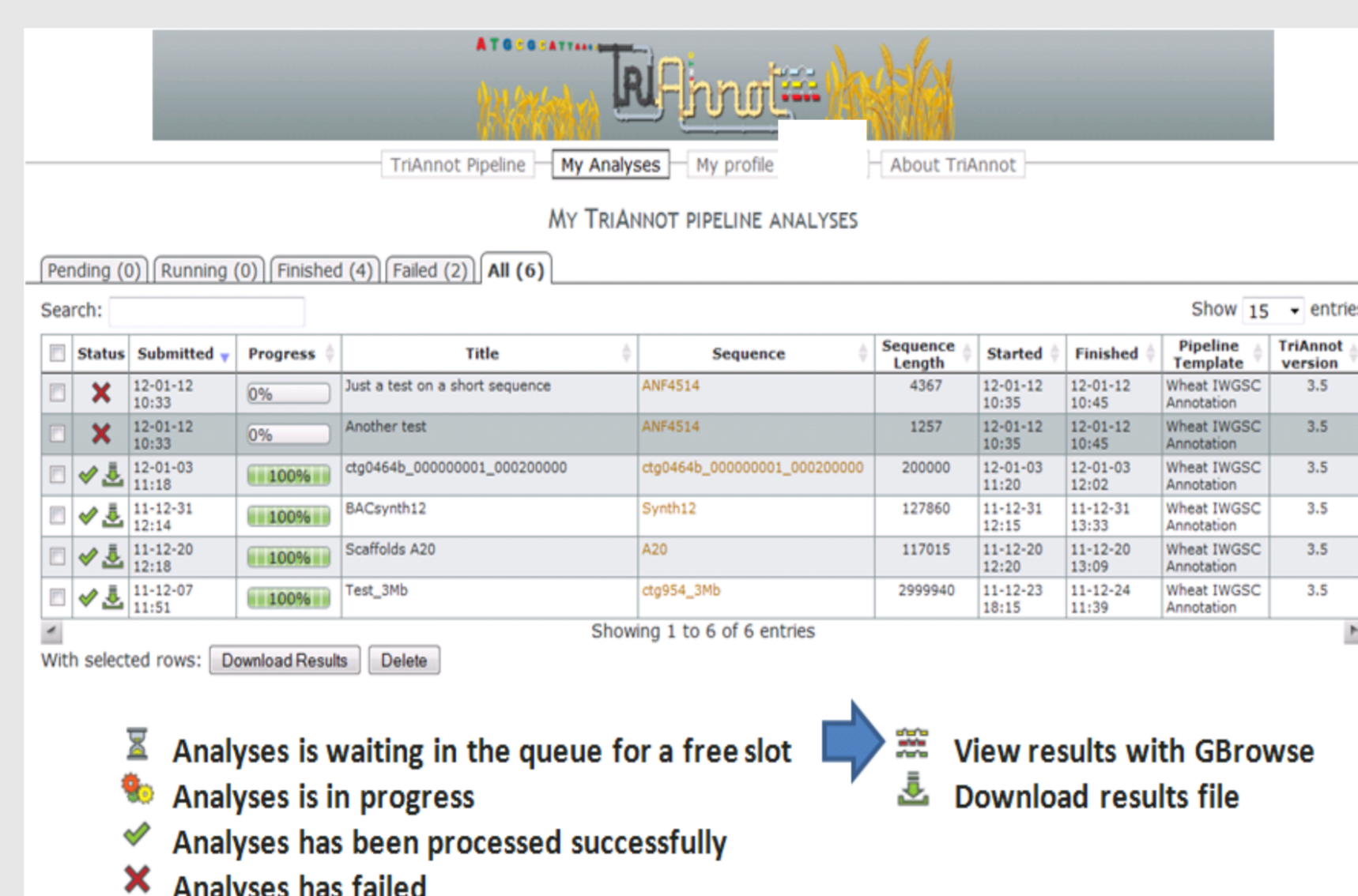


Optimized for wheat

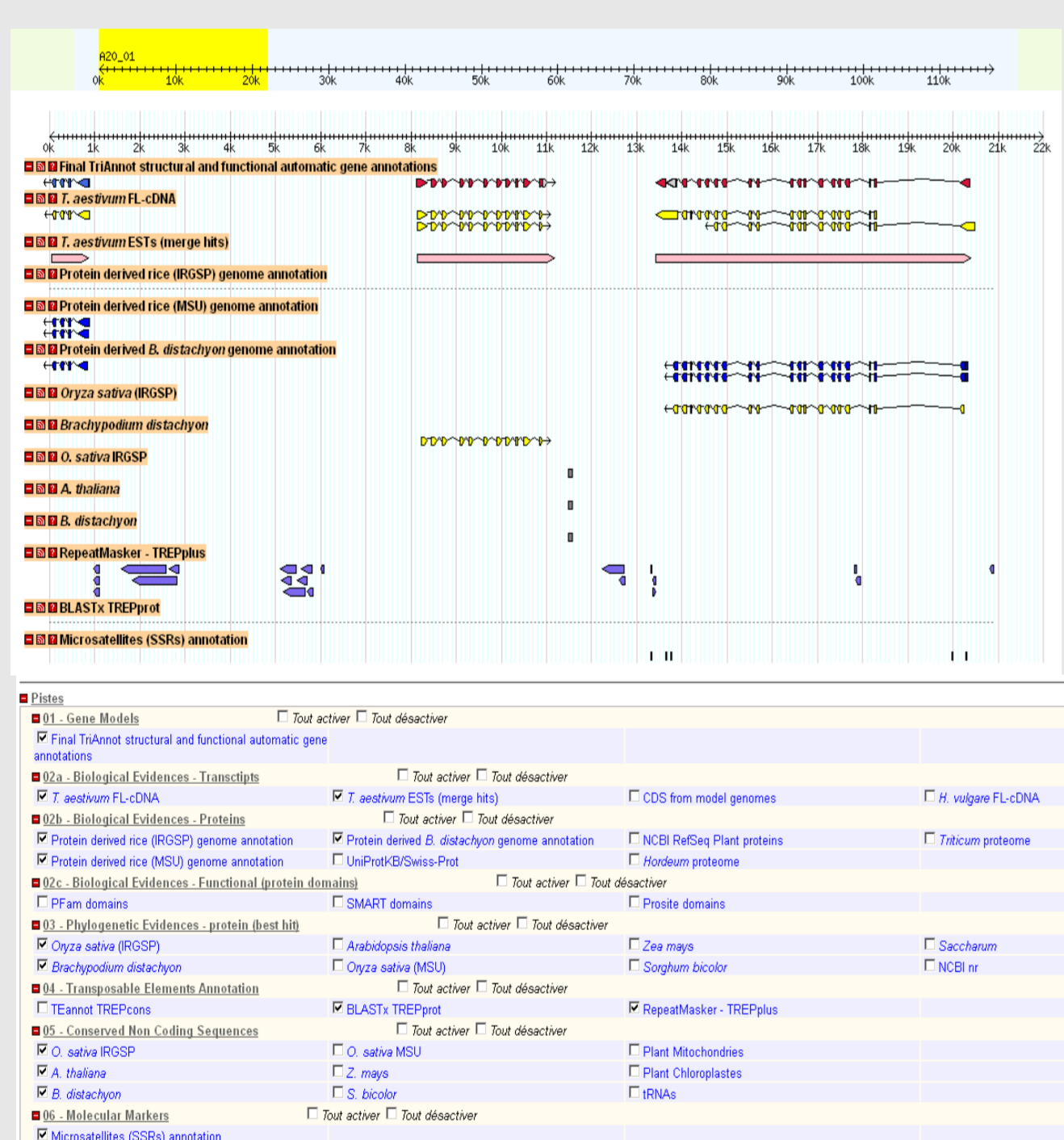
Under optimization for Oak, barley, rice and maize

Today the pipeline uses, as biological evidences, RNA-Seq assemblies for wheat, oak and barley.

2. Management screen



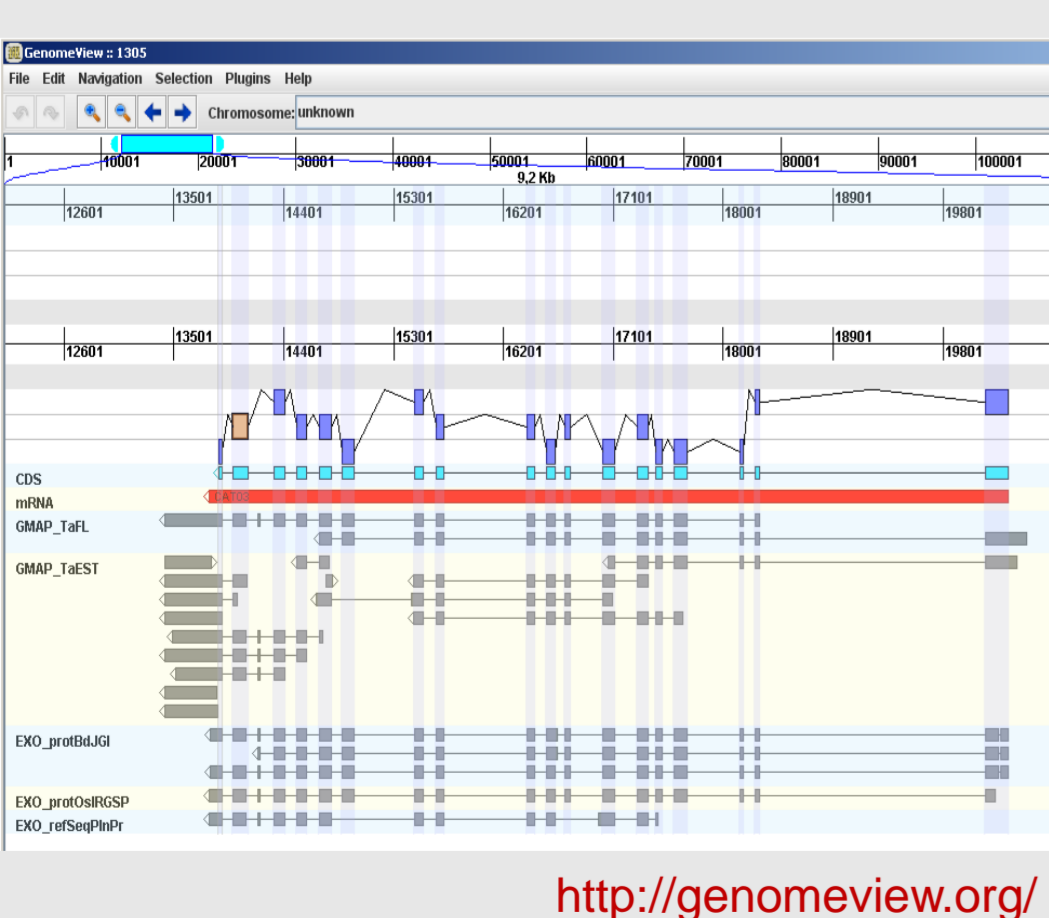
3. Gbrowse viewer



4. Manual curation

GenomeView

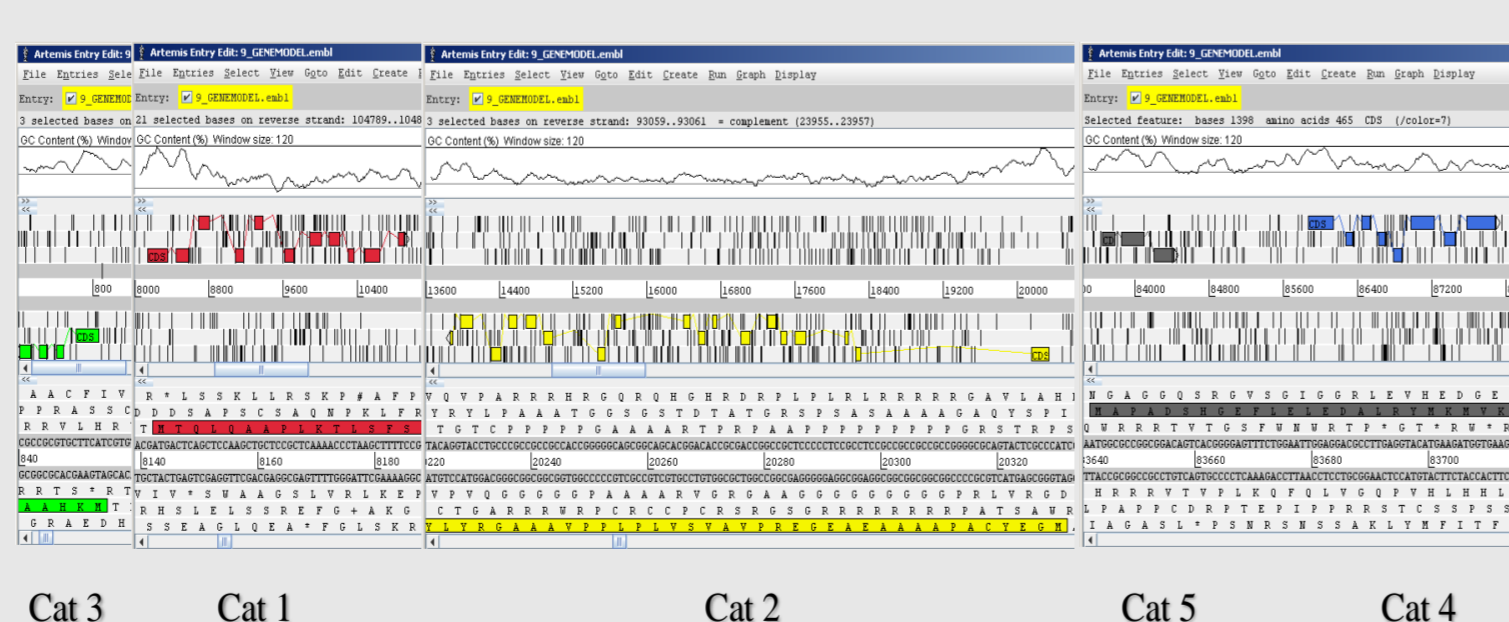
Abeel *et al.* (2012) Nucleic Acids Res. 40:e12



<http://genomeview.org/>

Artemis

Carver *et al.* (2008) Bioinformatics 24, 2672-2676

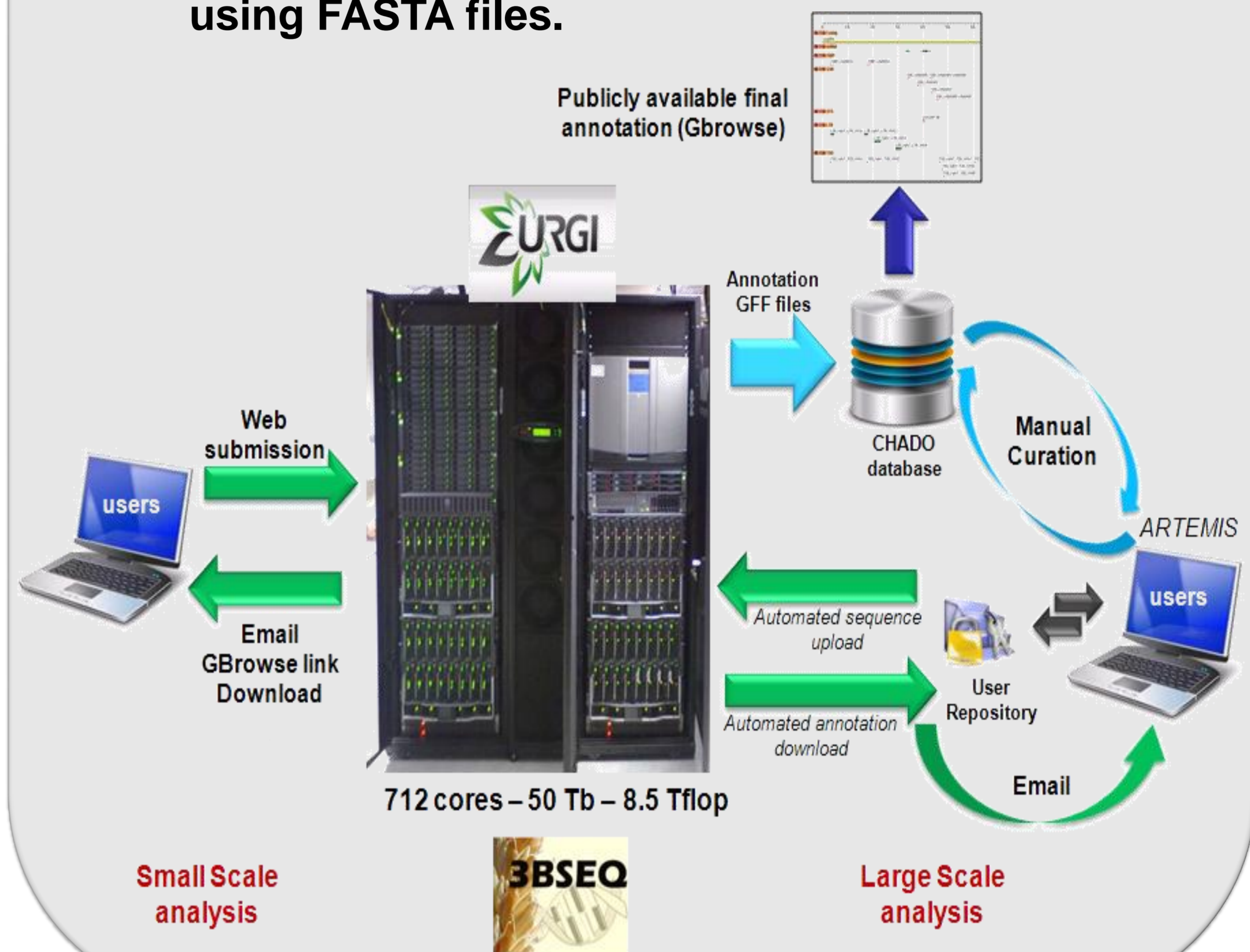


<http://www.sanger.ac.uk/resources/software/artemis/>

The TriAnnot pipeline is parallelized on a 712 CPU computing cluster hosted at INRA URGI.

It is accessible:

- * through a web interface for small scale analyses
 - * through a server for large scale analyses.
- For the later, the pipeline is launched automatically using FASTA files.



Small Scale analysis



Large Scale analysis

- In less than 8 hours, TriAnnot v3.5 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 63% were perfectly identified in accordance with the reference annotation,
- The 1 Gb of wheat chromosome 3B have been annotated in 26 hours,
- As it is easily adaptable to the annotation of other plant genomes, TriAnnot should become a useful resource for the annotation of large and complex genomes in the future,

• v3.6 is underway.

Perspectives

- A complete update of web pages
- A simplify procedure for login/password request
- Update of InterProScan from EBI
- Development of post-processing program to validate gene models (S. Theil) : High Confidence gene; Low Confidence gene; Pseudogenes; Fragmented-gene
- Implementation of *rnspace* for ncRNA annotation
rnspace is developed by INRA Toulouse M-J Cros *et al.*
- Implementation of ISBP, SSR and SNP markers identification with PCR-primer files.

Leroy *et al.* (2012) Frontiers in Plant Sciences, 3:1-14