http://urgi.versailles.inra.fr

The URGI bioinformatic platform: Focus on genomic annotation database and tools

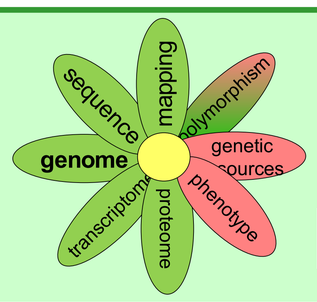
Michaël Alaux, Joëlle Amselem, Cyril Pommier, Isabelle Luyten, Sébastien Reboux, Benoit Hilselberger, Delphine Steinbach and Hadi Quesneville
INRA Unité de Recherche en Génomique-Info, Centre de Versailles, FRANCE

Abstract: The INRA URGI is a bioinformatics unit dedicated to plant genomes and their bio-agressors. We develop GnpIS, an information system dedicated to genomic and genetic data. We present here GnpGenome (http://urgi.versailles.inra.fr/gbrowse) our genomic annotation database and the "roundtrip" annotation process:
GnpGenome system relies on the well known GMOD tools (http://gmod.org): Apollo, Chado and GBrowse. Apollo is the graphical interface for visualization and annotation edition allowing curators to edit their genes according to evidences (transcript and protein similarity, comparative genomics). Manual annotations (gene curation validated/in progress) are saved in a dedicated Chado database and shared at the same time with other community annotation members. Validated genes/pseudogenes are then commited in a second Chado database accessible by GBrowse.

GnpIS description

The GnpIS information system is composed of 8 modules: 5 genomic modules and 3 genetic modules

1) GnpSeq, the EST, mRNA sequence database which contains clusters, contigs and annotations
2) GnpMap, the mapping database
3) **GnpGenome, a multispecies database containing genomic sequences and their structural annotations**
4) GnpArray, the MIAME compliant expression database
5) GnpProt, the proteomic database
6) GnpSNP, the polymorphism (SNP/DIP/STR) database
7) SiReGal, the genetic resources database for accessions collections
8) Ephesis (development in progress), the genotypes, environment and experimentation database



URGI Contact: urgi-contact@versailles.inra.fr
INRA URGI Web site: http://urgi.versailles.inra.fr

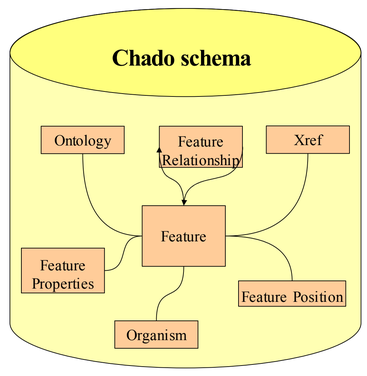
GnpGenome data submission

GFF3 file containing annotations for a specific genome/genomic sequence

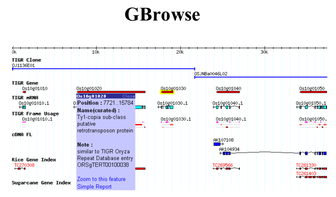
```
##gff-version 3
Vivi_v1_Chr1  Genoscope    chromosome  1     15630816  .  .  .  ID=Vivi_v1_Chr1

###
Vivi_v1_Chr1  translation  gene        1579  2025      .  .  .  ID=GSVIVG00029992001;Parent=Vivi_v1_Chr1
Vivi_v1_Chr1  translation  mRNA        1579  2025      3.98  +  0  ID=GSVIVT00029992001;Parent=GSVIVG00029992001
Vivi_v1_Chr1  translation  exon        1579  2025      3.98  +  0  ID=GSVIVE00029992001.1;Parent=GSVIVT00029992001
Vivi_v1_Chr1  translation  polypeptide  1579  2025      .  .  .  ID=GSVIVP00029992001;Derives_from=GSVIVT00029992001
```

Integration into the genomic database thanks to an automatic loader

BulkLoad



Visualization of the annotations in the Genome Browser



Technology:
- RDBMS: PostGreSQL 8.2
- Upload and integration tool: Perl - BioPerl
- Web interface: Perl - BioPerl

Plant data summary

- Wheat 3B FPC physical map
- TriAnnot pipeline BAC

- Grapevine genome browser

- Maize FPC physical map
- BAC Maize in progress
- ZMS1P3 BAC annotation

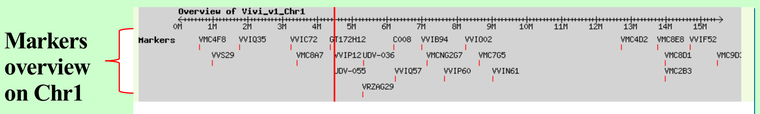
- TAIR v7
- Ath v5 and Rice v3 (CATMA, TIGR, SNP, DIP)
- Arabidopsis chloroplast

- Oryza sativa ssp japonica TIGR v4 annotations
- Oryza sativa ssp indica
- Oryza sativa pseudo-chromosomes v3

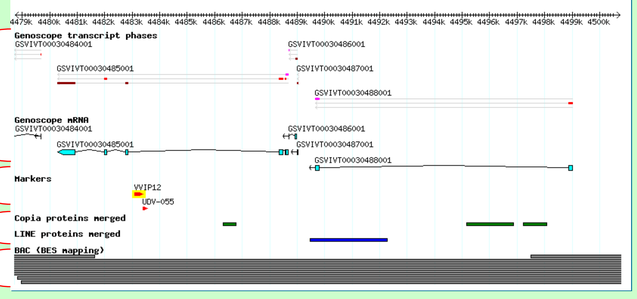
GnpGenome browser

Grapevine genome

http://urgi.versailles.inra.fr/gbrowse/grape



Markers overview on Chr1

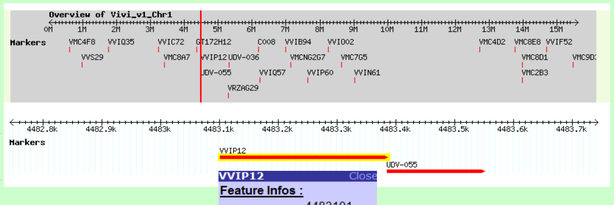


mRNA and frames (Génoscope)

Markers

Transposable elements (IGA)

BACs from BES



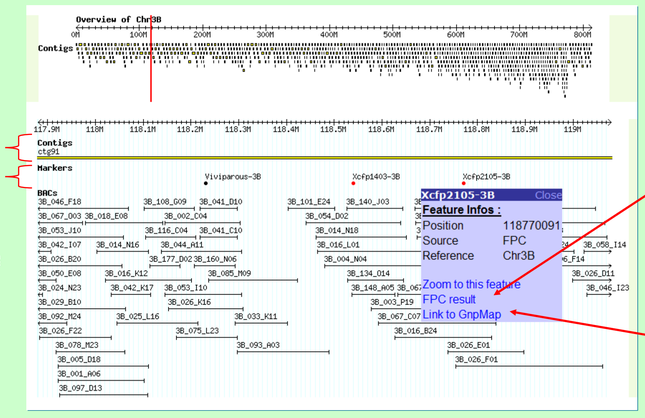
Markers

GBrowse detailed report

```
Name:       VVIP12
Class:      STS:marker
Type:       STS
Source:     marker
Position:   Vivi_v1_Chr1:4483101..4483386 (+ strand)
Length:     286
category:   marker
dbxref:     GFF_source:marker
            GnpMap:VVIP12
group:      Markers:VVIP12
label:      Markers:VVIP12
```

Wheat 3B FPC physical map

http://urgi.versailles.inra.fr/gbrowse/Wheat_FPC



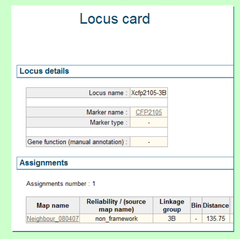
Contigs

Markers

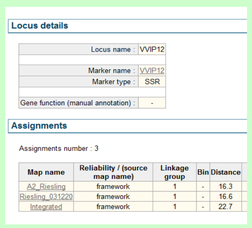
BACs

Link to FPC result

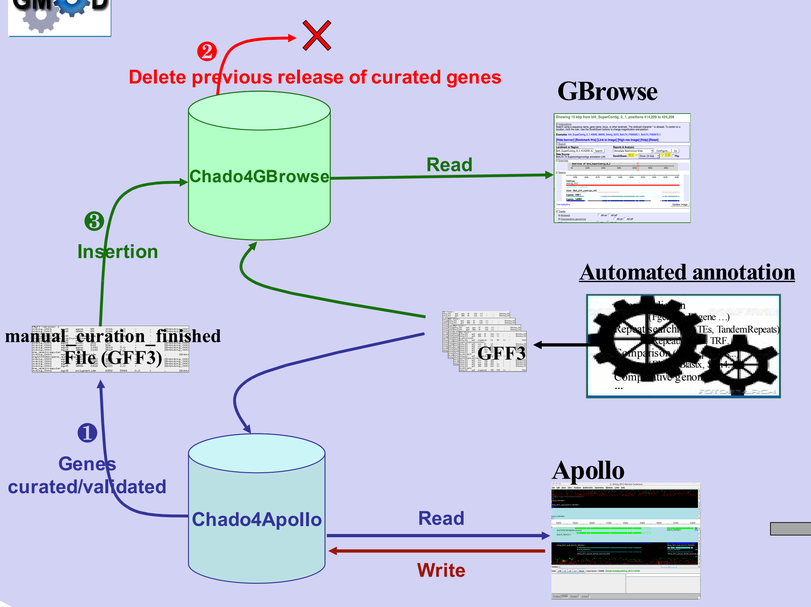
Link to GnpMap

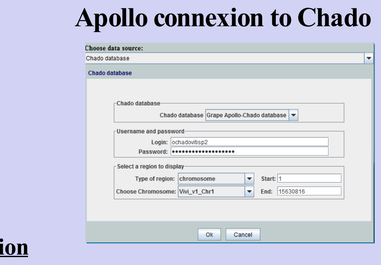


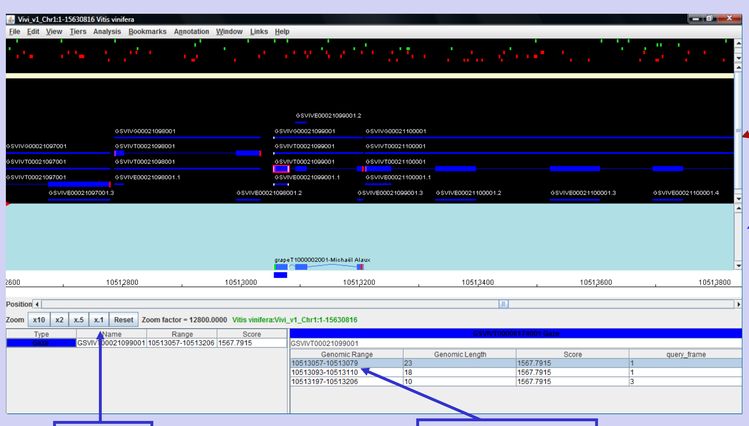
Link to GnpMap

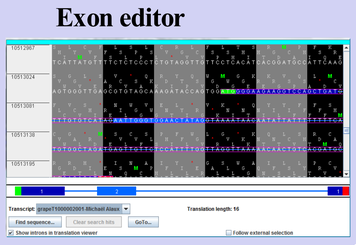


Roundtrip: Apollo - Chado - GBrowse









Related publications:

• A physical map of the 1-gigabase bread wheat chromosome 3B
Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Kianian S, Spielmeyer W, Lagudah E, Somers D, Kilian A, Alaux M, Vautrin S, Bergès H, Eversole K, Appels R, Safar J, Simkova H, Dolezel J, Bernard M, Feuillet C.
Science. 2008 Oct 3;322(5898):101-4.

• The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla
Jaillon O et al; French-Italian Public Consortium for Grapevine Genome Characterization.
Nature. 2007 Sep 27;449(7161):463-7. Epub 2007 Aug 26.

Acknowledgements:
Thanks to INRA partners and especially Anne-Françoise Adam-Blondon for grape, Catherine Feuillet, Etienne Paux, Philippe Leroy for wheat.
Thanks to former URGI team members involved in GnpGenome development and especially Fabrice Legeai, Isabelle Blanc-Lenfle, Olivier Arnaiz, Frédéric Sapet.
Thanks to GnpAnnot project partners and GMOD consortium.
Thanks to plant data partners (Génoscope, INRA teams, Biogemma, TAIR, TIGR).