Bayer CropScience

Research

creating the Future of Agriculture

Bayer CropScience - Belgium

June 17th, 2009

GBrowse: lessons learned and statement of interest

Erick Antezana
Frederic Potier

Who are we?

Research
creating the Future of Agriculture

- Working at Research Centre of Bayer CropScience
- Fungicides, herbicides, insecticides
- ~18’000 world wide,
- ~250 Ghent, Belgium
- Bayer BioScience
  - Biotech company
  - Dealing with: crops, cereals, vegetables, …
- GMOD
  - GBrowse 1.70 and 2.0
  - CMap
  - Galaxy
  - ERGATIS (tigr-workflow)
  - ...

Bayer CropScience

Outline

- A bit of history
- Current Bayer GBrowse infrastructure
  - Public Genome Annotations
  - Private Genome Annotations
- In house developed components
- Requirements/Needs
- Conclusion/Discussion

Outline

- A bit of history
- Current Bayer GBrowse infrastructure
  - Public Genome Annotations
  - Private Genome Annotations
- In house developed components
- Requirements/Needs
- Conclusion/Discussion

A bit of history

Research
creating the Future of Agriculture

- GBrowse utilised since 2004
- Tested most of the versions and the available adaptors
  - Currently: **GBrowse 2** and mainly **Bio::DB::GFF**
- Mainly focus on plant genomes (*e.g.* rice)

Lots of :

  - Publicly available plant genome sequences
  - Private genomes
  - Annotation release updates are more and more frequent
- Requirements:
  - Minor data reformatting
  - Fast data loading
  - Fast querying
  - Highly customizable application
  - High level of integrity in our bioinformatics platform

Bayer CropScience

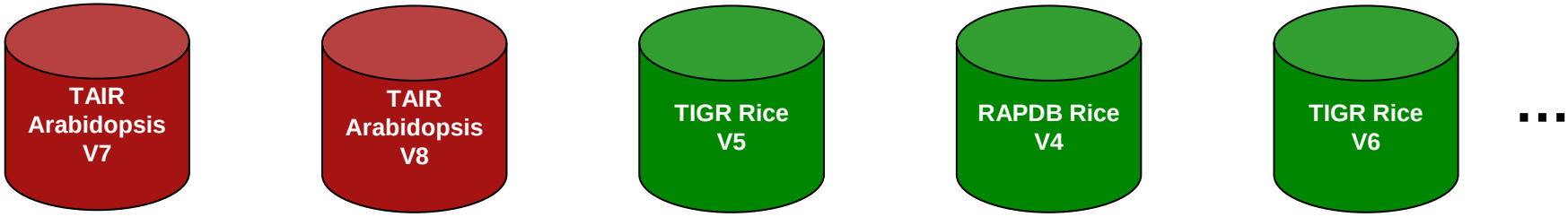
Outline

- A bit of history
- Current Bayer GBrowse infrastructure
  - Public Genome Annotations
  - Private Genome Annotations
- In house developed components.
- Requirements/Needs
- Conclusion/Discussion

GBrowse infrastructure: Public Data

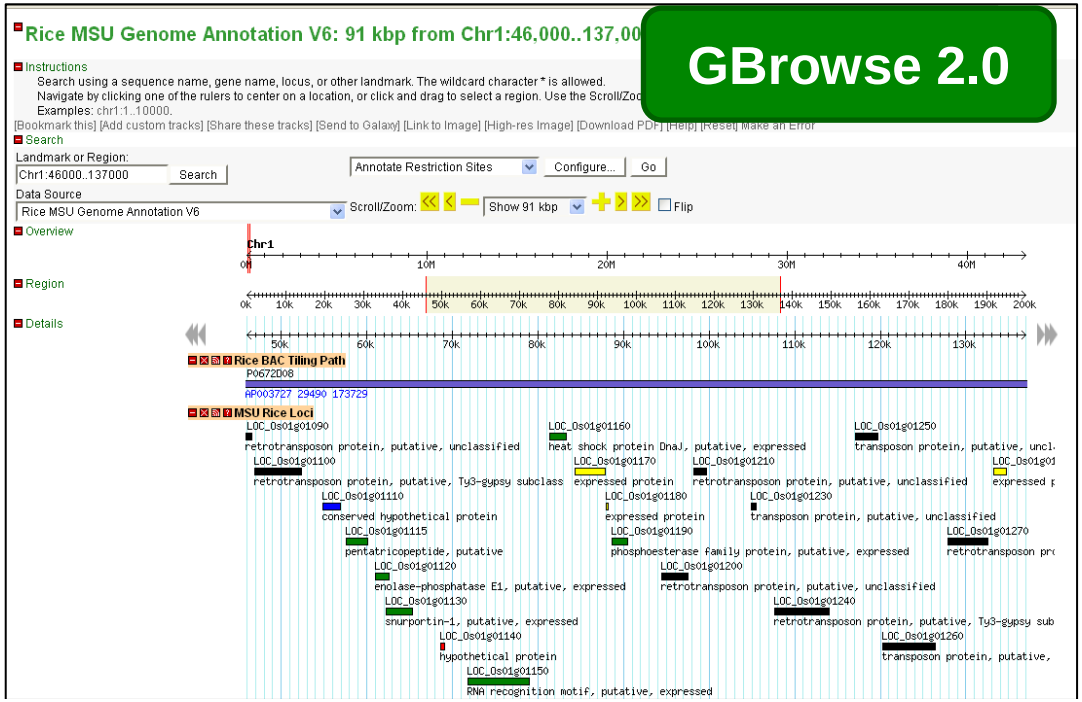
Research
creating the Future of Agriculture

One MySQL database per Genome Annotation Version



Connection to MySQL using Bio::DB::GFF adaptor

- More than 30 databases
- Around 30 GB of data

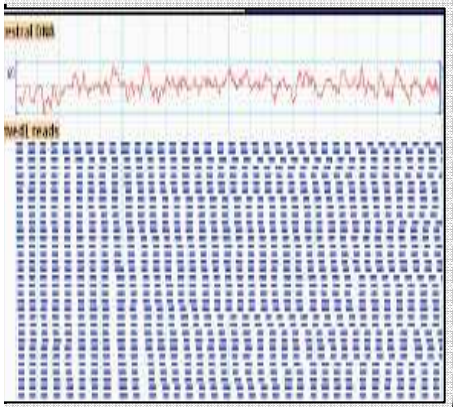


GBrowse infrastructure: Private Data

Research
creating the Future of Agriculture

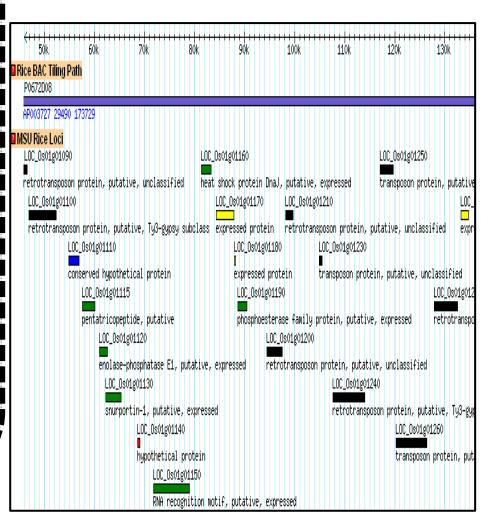
NGS Mapping/ Coverage

Bio-SamTools



Genome Annotation

Bio::DB::GFF



Molecular Mapping

CMap

User Annotation /Manual Curation

Chado ??

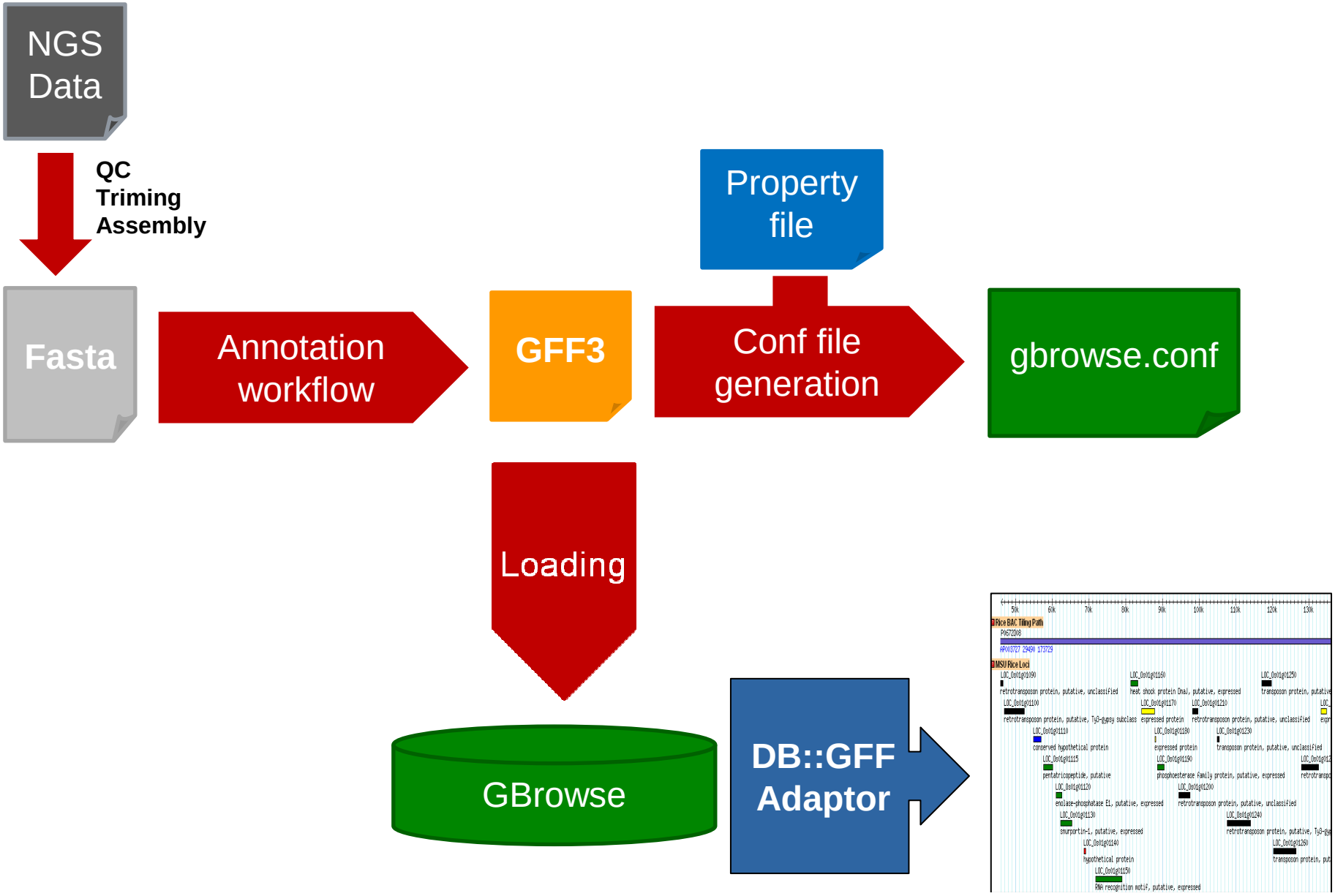
Artemis
Apollo

...

Bayer CropScience

Automated Annotation workflow

Research
creating the Future of Agriculture



Outline

- A bit of history
- Current Bayer GBrowse infrastructure
  - Public Genome Annotations
  - Private Genome Annotations
- In house developed components
- Requirements/Needs
- Conclusion/Discussion

In house developments

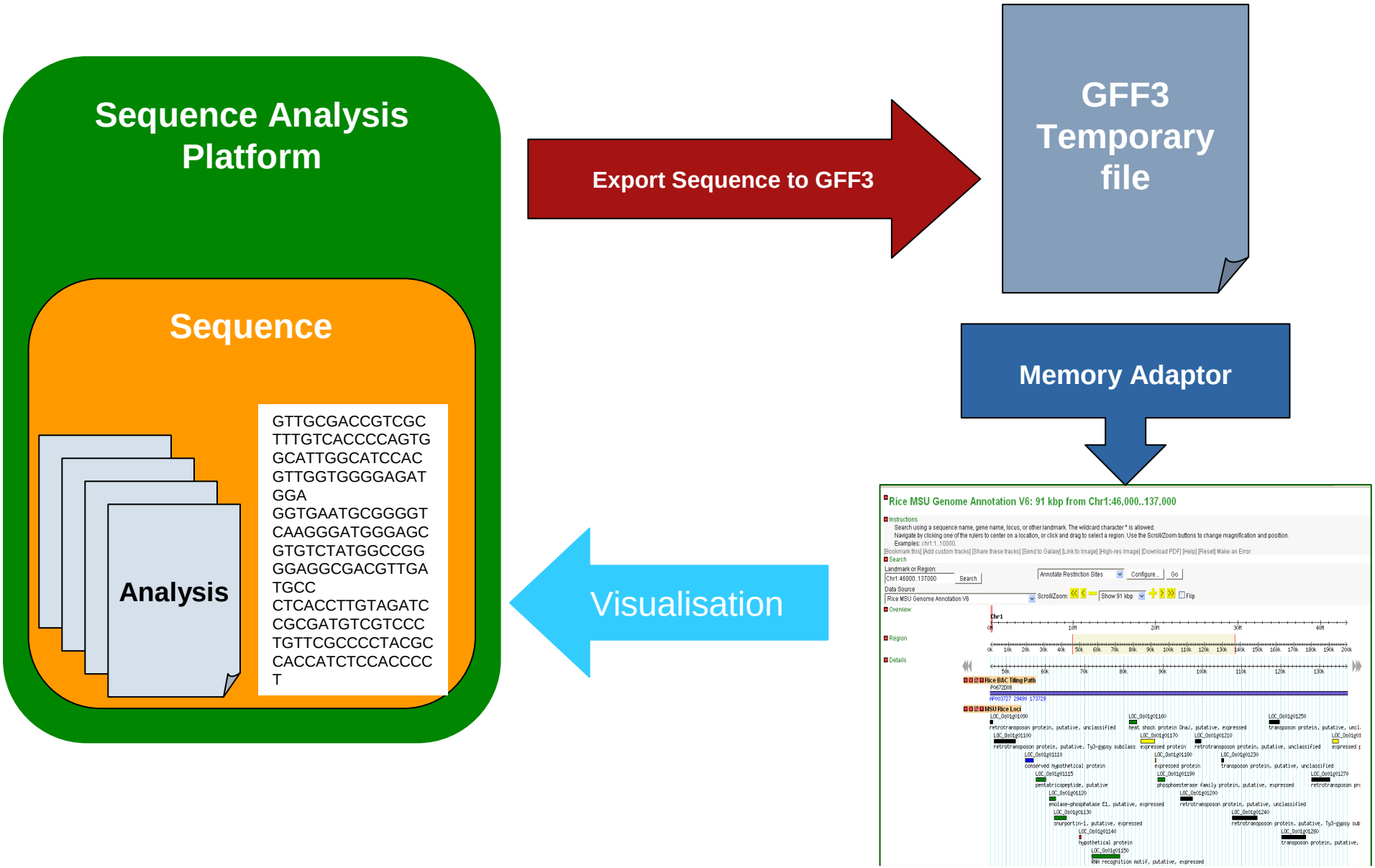
Research
creating the Future of Agriculture

- Authentication system
  - track of user sessions
  - storage of the user annotation on the server
  - So, activate user access rights
- GFF3 files on-the-fly visualization.
- Blast anchoring/Sequence homology search
  - blast homologies are uploaded as user annotations
- Plugins
  - data export
  - links to in house applications
- In house keyword search engine
  - fast search utility
  - cross databases search
- Gateway
  - centralised access point

Bayer CropScience

GBrowse for on-the-fly visualisation

Research
creating the Future of Agriculture

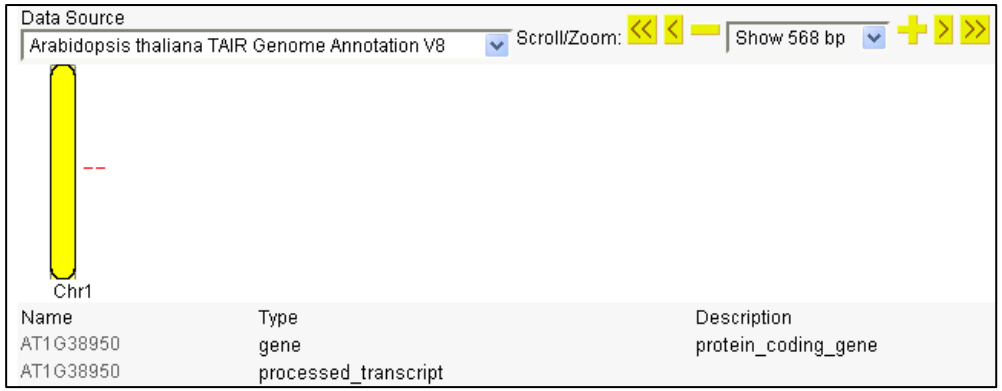


BLAST anchoring*

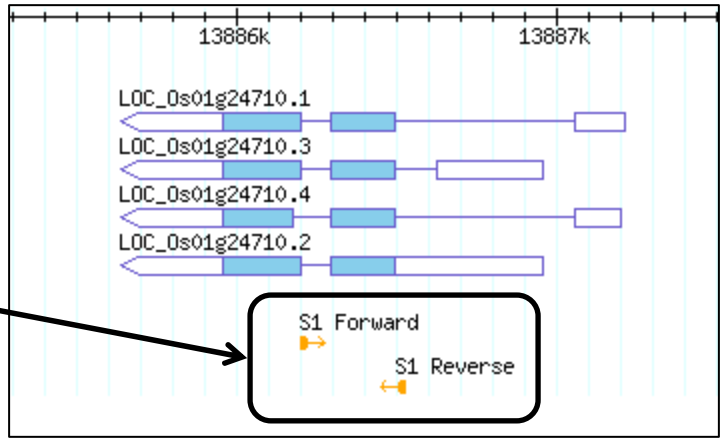
Research
creating the Future of Agriculture

```
>Fasta
AGGAAGAAAA TAGGGAAAAA
AAAGGAGAGA GAATATTATG
AATTATTCTT TGCTTGAGCT
CAGAAACAGT TCTTCTTCTG
CTTCTTCGAC TTCTTTTCTC
TGTCTTTCTT CTTTATGCTT
AGTGCTAAAT CACTCGTTTA
CTTGTGAAGA TTATGGATCT
CTGATTAAAG TTTGTTTCTC
GTATTTATTC CAAGGTTGCT
TCTTCTTTTT CTCAATTGGA
TCTTTTAATT TTTGTTTTTC
```

BLAST



User Annotation



* under development

Bayer CropScience

GGB Generic Genome Browser - Mozilla Firefox

http://bioim.bioscience.bayercropscience/Tools/GeneticAndGenomic/GGB/

You are here: Home > BioIM Tools > Genetics and Genomics > GGB

News | Workspaces | BioIM Tools | Documentation | BioIM staff | Contact Us

- News
- Workspaces
- BioIM Tools
- Documentation

- BioIM Tools
  - Genetics and Genomics
    - GGB

<< back

Quick links

- Blast2
- GGB Gateway
- CMAP Gateway
- Hecate
- jCeasar
- jCeasar citrix
- Mobman
- MarkerDB
- Rlims
- Resman
- Strains
- Textual Searching

Gateway to GGB

[Generic Genome Browser]

BIOIM Entrez Search Engine dedicated to GGB Go Clear Help

Internal browsers | **External browsers**

Rice | Other Gramineae | **Other Brassicaceae** | Other Plants

ORGANISM	DESCRIPTION	GENERATED AND ANNOTATED BY	
Arabidopsis thaliana	Complete genome annotation V7	CIRAD	
Arabidopsis thaliana	AtiDB: Arabidopsis thaliana Integrated Database with Brassica sequence homologies	AtiDB	
Brassica rapa	Pseudomolecule annotation	BrGP	
Brassica rapa	BAC annotation	BrGP	
Arabidopsis thaliana	Arabidopsis thaliana Small RNA Project	ASRP	

[Green links are internal] [Red Links are external] [: Access restricted]

Last updated Tue, 30 Jun 2009, 14:18

Done

Maintenant : Mostly Cloudy, 27 °C | Wed : 28 °C | Thu : 28 °C

Outline

- A bit of history
- Current Bayer GBrowse infrastructure
  - Public Genome Annotations
  - Private Genome Annotations
- In house developed components
- Requirements/Needs
- Conclusion/Discussion

Statement of interest: DB adaptors

Research
creating the Future of Agriculture

- **NGS adaptor**

  Key priority
- **Memory adaptor**

  To be able to specify a file name or a complete path via a parameter

  so, the adaptor doesn't need to load all the GFF files in the directory
- **Chado adaptor**

  - Portability to Oracle

  - To store user annotation and manual curation

  - Including a system track versions and history of the annotations

  - Management of user access rights
- **SeqFeature::Store**

  Portability to Oracle (c.f. user access rights via VPD)

  Improve loading process: time issues
- **Compatibility with other genome browsers databases**

  For instance: ensembl databases?

Bayer CropScience

Statement of interest: User Interaction

Research
creating the Future of Agriculture

- **Authentication**
  - To track user sessions
  - To enable user access rights management
- **User Annotation Management**
  - To store the user annotations in a database or in a file on the server

    Thus the users will be able to get their annotations while getting connected to different machines
  - To send automatically user's annotations to GBrowse via a URL parameter

- **Integration with CMap**

Bayer CropScience

Statement of interest: Gbrowse.conf

Research
creating the Future of Agriculture

- Issues with the conf file format:
  - Error prone
  - Difficult to debug
  - Steep learning curve
  - Time consuming to maintain
  - …
- **Solution**: automatic conf file generation for instance
- **Ideal solution**: better representation of the configuration
  - Use XML for instance
- Configuration of the global layout to enable/disable components thereof:
  - Disable the custom tracks component
  - Disable the display settings component
  - …

Bayer CropScience

Statement of interest: *data_source*.conf

Research
creating the Future of Agriculture

- Genome annotation metadata
  - Species information
  - Assembly and Annotation version

```
################################
# database definitions
################################
[TAIR_Arabidopsis_V8:database]
db_adaptor    = Bio::DB::GFF
db_args       = -adaptor DBI::mysql
                -dsn dbi:mysql:TAIR_Arabidopsis_V8
species            = Arabidopsis thaliana
assembly.source    = TAIR
assembly.version   = 8
annotation.source  = TAIR
annotation.version = 8
```

Bayer CropScience

Statement of interest: web services

Research
creating the Future of Agriculture

- Querying/Reporting tool on metadata
  - List of reference sequences
  - Annotation version
  - Assembly version
  - List of available feature types
  - Suggestion:

```
<browser>
  <species>Arabidopsis</species>
  <assembly>bayer</assembly>
  <annotation>1.0</annotation>
  <reference-sequence>chr1</reference-sequence>
  <reference-sequence>chr2</reference-sequence>
  <feature-type>fgenesh:mRNA</feature-type>
  <feature-type>splign:mRNA</feature-type>
</browser>
```

Bayer CropScience

Outline

- A bit of history
- Current Bayer GBrowse infrastructure
  - Public Genome Annotations
  - Private Genome Annotations
- In house developed components.
- Requirements/Needs
- Conclusion/Discussion

Conclusion / Discussion

Research
creating the Future of Agriculture

- GBrowse 2 is a tool that can be used in a production environment
  - Performance (rendering farm)
  - Various DB's
- Intensively used within the Bayer Bioinformatics platform:
  - Facilitate data integration
  - High level of integration
  - Easy to maintain
- Our priorities for further developments:
  - Adaptors performance
  - Need to focus on user interaction
  - GBrowse.conf representation
  - Native integration of other GMOD tools (e.g. CMap)

Bayer CropScience

Bayer CropScience

Research

creating the Future of Agriculture

Thank you for your attention