

[Home/Search](#)[Browse by Day](#)[Browse by Type](#)[Author Index](#)[Poster Categories](#)

IWGSC - Standards and Protocols

Organizers: *Catherine Feuillet(INRA) and Rudolf Appels(Murdoch University)*

Date: Tuesday, January 17, 2012

Time: 1:30 PM-6:00 PM

Room: Golden Ballroom

Moderator: *Nicolas Guilhot*

1:30 PM

W421

Using LTC software to assemble physical maps in complex genomes such as wheat
Zeev Frenkel, University of Haifa; Etienne Paux, INRA; David Mester, University of Haifa;
Catherine Feuillet, INRA; **Abraham Korol**, University of Haifa

3:15 PM

Break

3:45 PM

W422

The TriAnnot Automated Annotation Pipeline: Making Sense of the Output Files and Information - a Case Study
Nicolas Guilhot, INRA GDEC; **Philippe Leroy**, INRA GDEC; Sébastien Theil, INRA GDEC;
Frederic Choulet, INRA GDEC; Sébastien Reboux, INRA - URGI, Research Unit Genomic-Info; Michael Alaux, INRA - URGI, Research Unit Genomic-Info; Matthieu Reichstadt, UMR1019 Unité de Recherche en Nutrition Humaine, Institut National de la Recherche Agronomique; Hadi Quesneville, INRA - URGI, Research Unit Genomic-Info; Catherine Feuillet, INRA GDEC

5:30 PM

W423

In-silico mapping of my favorite gene using the survey sequence of wheat
Michael Alaux, INRA-URGI; Françoise Alfama, INRA-URGI; Véronique Jamilloux, INRA-URGI; Sébastien Reboux, INRA-URGI; Baptiste Brault, INRA-URGI; Daphné Verdelet, INRA-URGI; Aminah Keliet, INRA-URGI; Delphine Steinbach, INRA-URGI; Hadi Quesneville, INRA-URGI

The TriAnnot Automated Annotation Pipeline: Making Sense of the Output Files and Information - a Case Study

W422

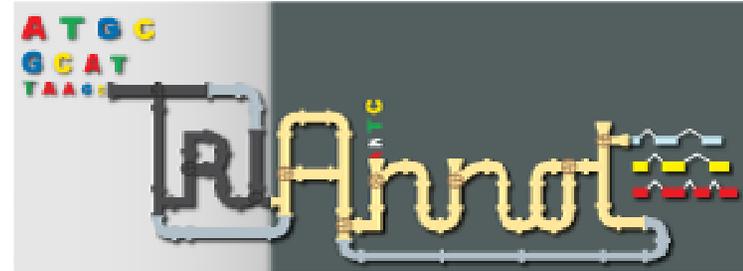


V3.5

3:45 – 4:05	Introduction	<i>P. Leroy</i>
4:05 – 4:25	New Interface Download New GBrowse	<i>N. Guilhot</i>
4:25 – 5:25	How to use output files	<i>P. Leroy & N. Guilhot</i>
5:25 – 5:30	Conclusions	<i>P. Leroy & N. Guilhot</i>

3:45 – 4:05 Introduction

P. Leroy



V3.5



Versatility

*Small & Large scale analysis
URGI Information System (IS)*



Architecture

TriAnnot is modular & paralyzed



Color code system

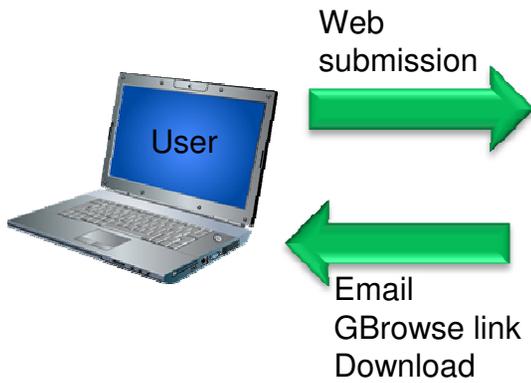
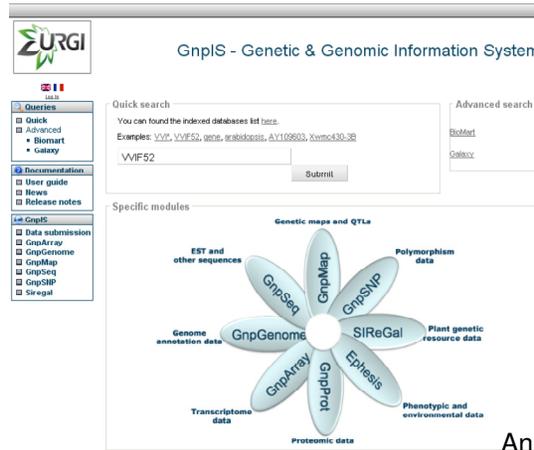
Confidence of structural gene annotation



Web pages

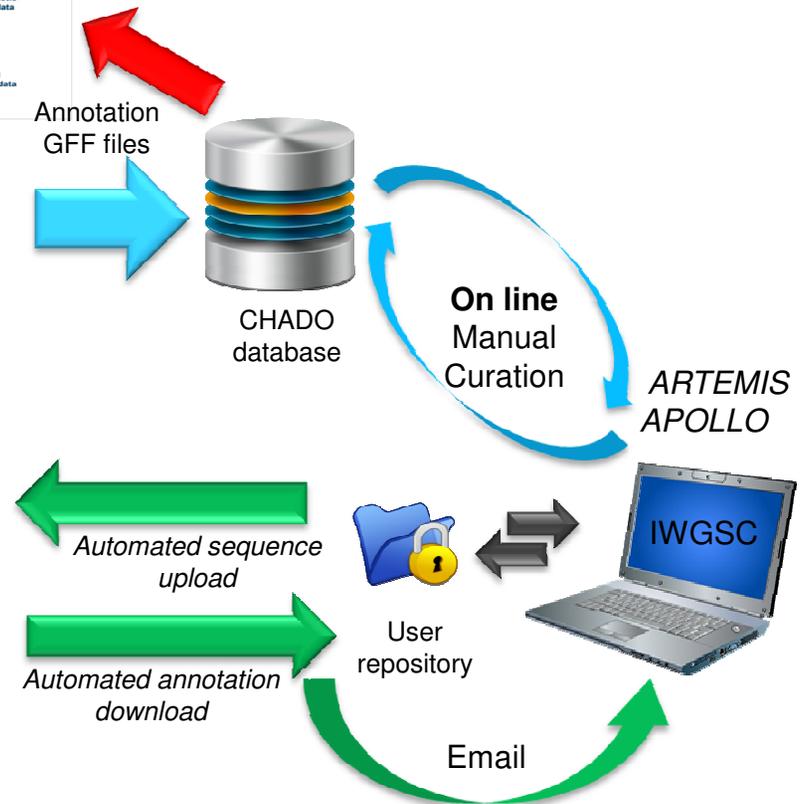
Login / password, availability, reference, intellectual property, etc ...

Versatility



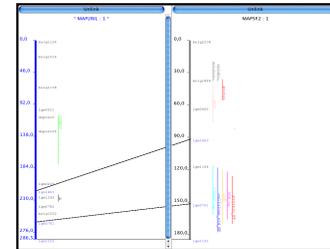
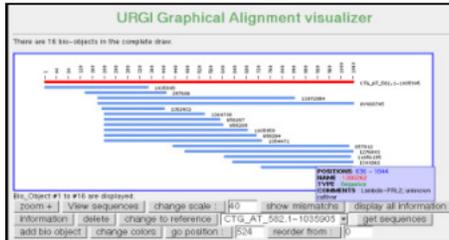
712 cores – 50 Tb – 8.5 Tflop

Small scale analysis



Large scale analysis

Integrative Databases for interoperability



Genetic maps and QTLs

EST and other sequences

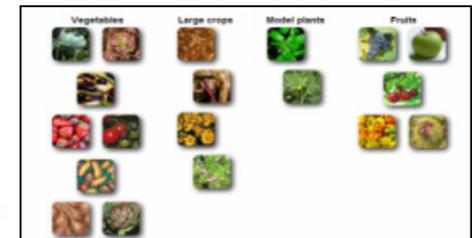
Polymorphism data

2	35	37	44	67	68	73	76	84	85	110	112	120	121	129	162	190	192	211	223	236	234	235	250	291	297	
C	T	T	A	T	-	T	C	T	T	G	G	T	C	G	C	G	T	C	C	T	A	C	C	T		
C	T	T	A	T	-	T	C	T	T	G	G	T	C	G	C	G	T	C	C	T	A	C	C	T		
C	T	T	A	T	-	T	C	T	T	G	G	T	C	G	A	G	T	C	C	T	A	C	C	T		

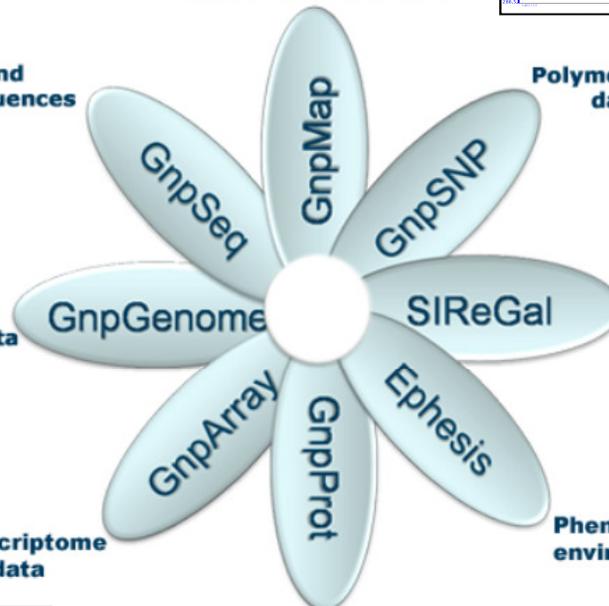


Genome annotation data

Plant genetic resource data



Phenotypic and environmental data

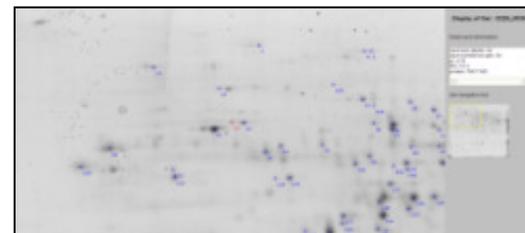


Transcriptome data

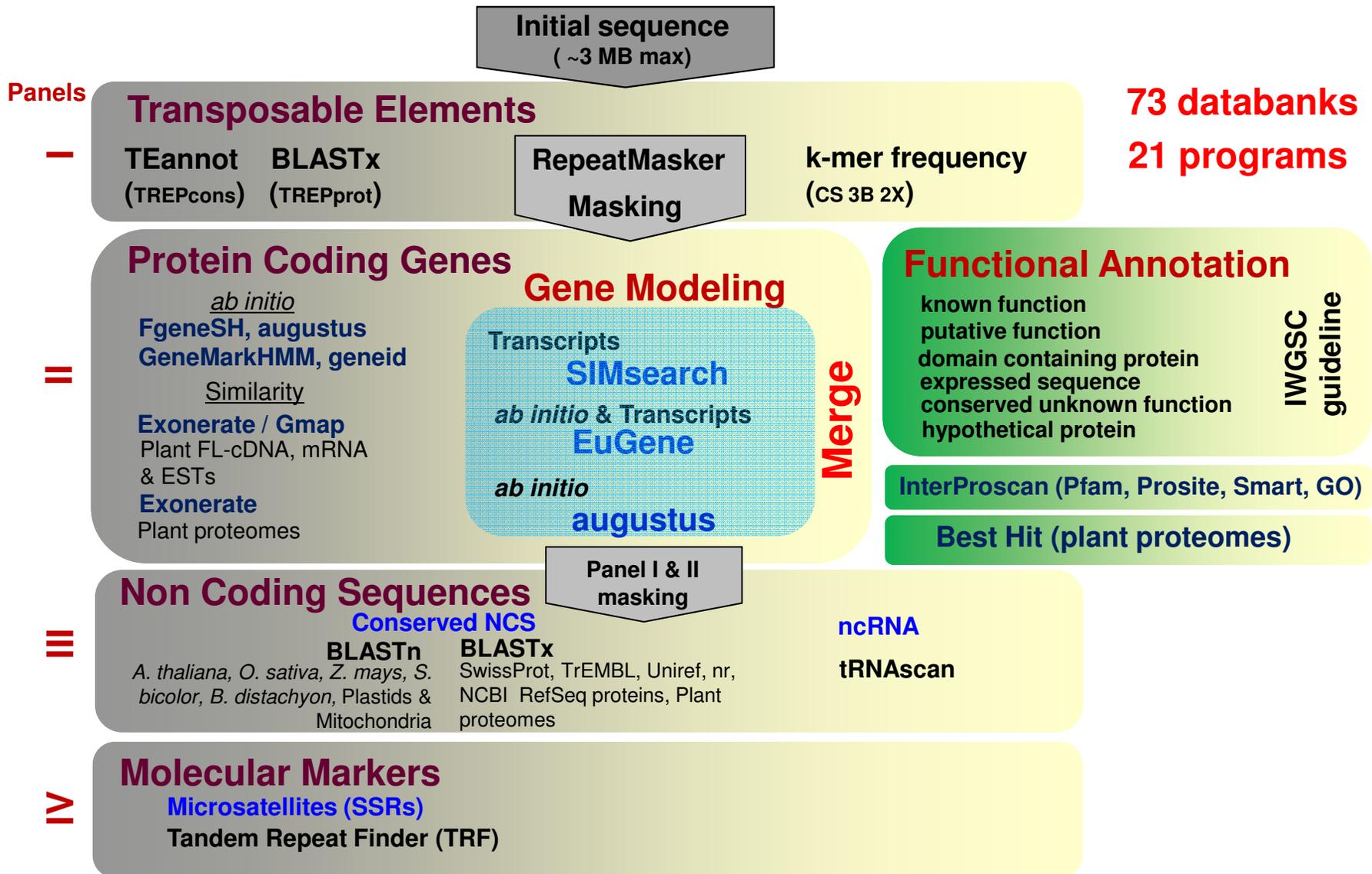
Proteomic data

Hybridization results

Result	Hybridization date	Description	Experiment	Gene name	Design name	Site number	Product	Image	File
1	2008-08-15	Hybridization	Hybridization	Gene	Design	Site	Product	Image	File
2	2008-08-15	Hybridization	Hybridization	Gene	Design	Site	Product	Image	File
3	2008-08-15	Hybridization	Hybridization	Gene	Design	Site	Product	Image	File
4	2008-08-15	Hybridization	Hybridization	Gene	Design	Site	Product	Image	File
5	2008-08-15	Hybridization	Hybridization	Gene	Design	Site	Product	Image	File
6	2008-08-15	Hybridization	Hybridization	Gene	Design	Site	Product	Image	File
7	2008-08-15	Hybridization	Hybridization	Gene	Design	Site	Product	Image	File
8	2008-08-15	Hybridization	Hybridization	Gene	Design	Site	Product	Image	File
9	2008-08-15	Hybridization	Hybridization	Gene	Design	Site	Product	Image	File
10	2008-08-15	Hybridization	Hybridization	Gene	Design	Site	Product	Image	File



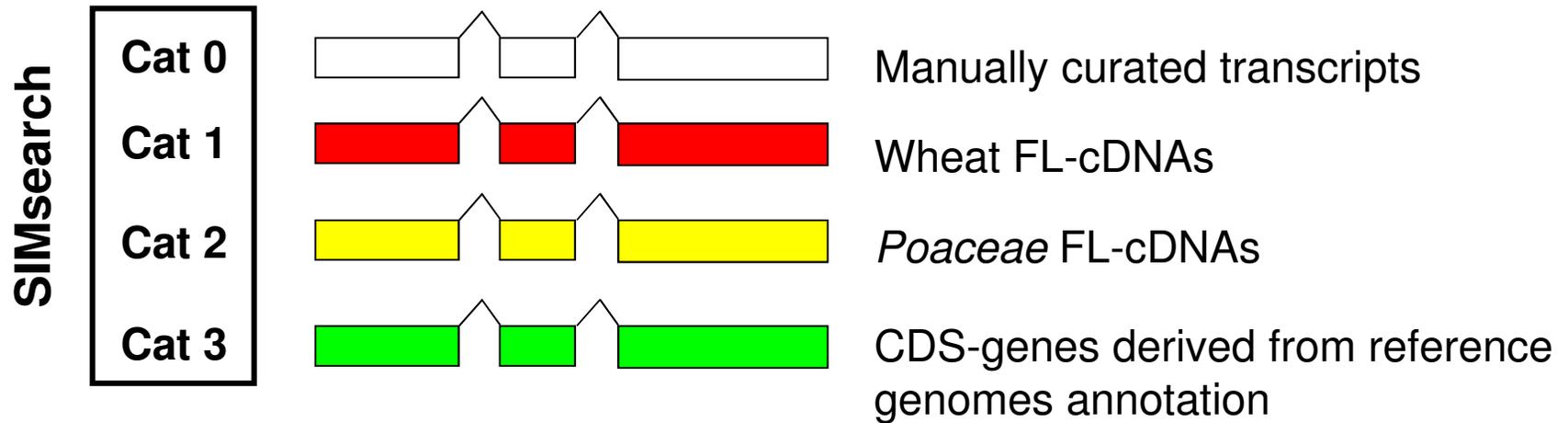
TriAnnot Architecture



Structural Annotation – 6 categories

Use a color code system to assess the confidence of the structural annotation

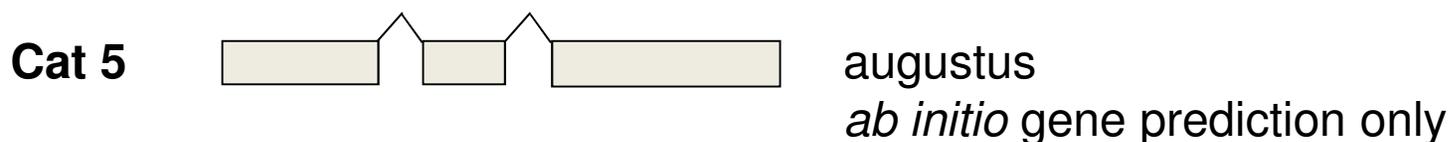
Similarity



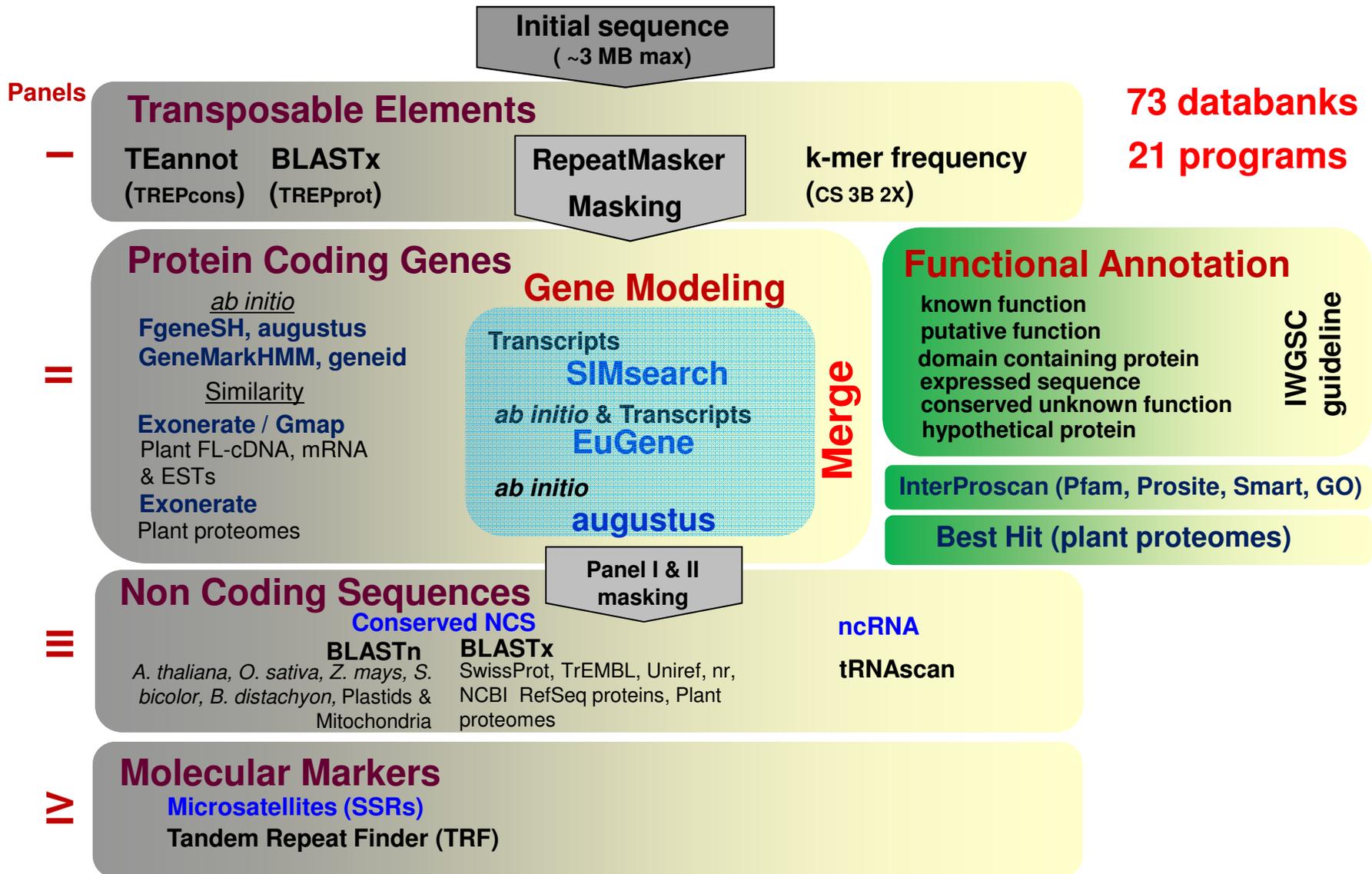
ab initio & Similarity



ab initio



TriAnnot Architecture



Functional Annotation – 5 Classes

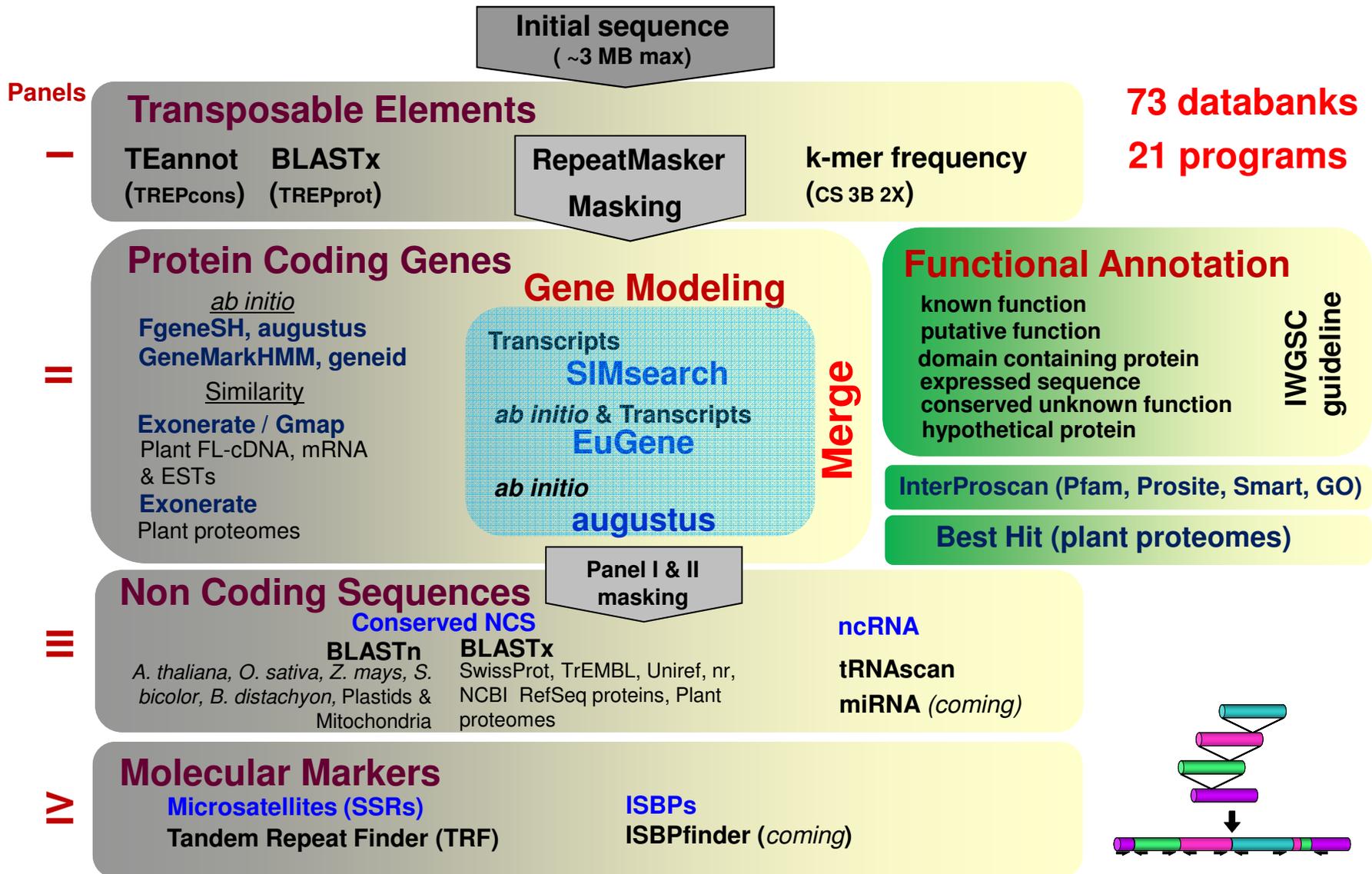
<http://urgi.versailles.inra.fr/Species/Wheat/Triannot-Pipeline/Architecture>

- **Panel II - Step9 - Functional Annotation**

- Putative function for the gene models are assigned via a combination of similarity search (BLASTP) against several protein databanks and against the Pfam (*Sammut et al., 2008 Brief Bioinform. 9, 210-219; Finn et al., 2010 Nucleic Acids Res. 38, D211-222*) protein domain collection with HMMER 3.0 (<http://hmmer.janelia.org/software>). TriAnnot follows a nomenclature based on the guideline established in 2006 by the IWGSC annotation working group (<http://www.wheatgenome.org/tools.php>):

- “**known-function**”: when >80% identity over >80% of the protein length is found with a known protein in UniProtKB/Swiss-Prot. This category reflects the highest quality for functional annotation.
- “**putative-function**”: when >45% similarity over >50% of the protein length is found with a known protein in UniProtKB/Swiss-Prot and UniProtKB/TrEMBL.
- “**domain-containing-protein**”: when there is no significant BLASTP hit with a known or putative function in the previous steps, but one or more Pfam domains (*Sammut et al., 2008; Finn et al., 2010*) are identified.
- “**expressed-sequence**”: based on TBLASTN against plant EST databanks with >45% identity and >50% coverage.
- “**conserved-unknown-function**”: when no expressed sequence is found, and when >45% similarity over >50% of the protein length is found only with an unknown function (i.e. a protein annotated as “putative” or “hypothetical”) in UniProtKB/Swiss-Prot and UniProtKB/TrEMBL.
- “**hypothetical-protein**”: when no similarity is found, either in UniProtKB/Swiss-Prot or UniProtKB/TrEMBL, or Pfam domain or ESTs.

TriAnnot Architecture



http://www.clermont.inra.fr/triannot



triannot

Environ 2 060 résultats (0,17 secondes)

- Tout
- Images
- Vidéos

[Triannot Pipeline - URGI](#) - [Traduire cette page]

19 Nov 2010 ... A first version of **TriAnnot** has been deposited on January 31th, 2006 to the French APP ("Agence pour la Protection des Programmes") ...

[urgi.versailles.inra.fr/.../Triannot-Pipeline](#) - En cache - Pages similaires

FEEDBACK | CONTACT | SITE MAP | ABOUT US



PLANT AND FUNGI DATA INTEGRATION

Register

Platform Research Projects Data Tools Species

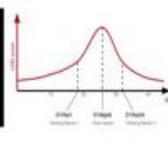
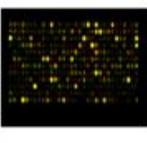
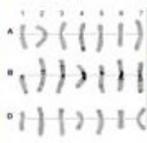
SEARCH OK

Species

You are here : [Home](#) / [Home URGI](#) / [Species](#) / [Wheat](#) / Triannot Pipeline

Triannot Pipeline

Together with rice and maize, wheat provides more than 60 % of the calories and proteins for our daily life. Among the grasses, bread wheat (*Triticum aestivum* L., 2n=6x=42, AABBDD) has one of the largest genome size with 17,000 Mb (about 45-fold larger than the rice genome). It is an allohexaploid consisting of seven groups of chromosomes, each group containing a set of three homoeologous chromosomes belonging to the A, B and D genomes. The A genome was contributed by *Triticum urartu*, a diploid wheat ancestor, and the B genome by an unknown close relative of *Aegilops speltoides*, another diploid wheat ancestor. About 10,000 years ago, the D genome from *Aegilops tauschii* was added to the AB genome tetraploid *Triticum turgidum* resulting in bread wheat. Thus, in addition to its socio economic importance, its recent history makes wheat one of the best species to study the evolution of polyploids.



Genome sequences hold the key for understanding the molecular basis of phenotypic traits and variation and provide a framework for rapid varietal development through the utilization of marker-assisted selection, marker-assisted recurrent selection, genome wide selection, and molecular breeding. Despite the recognition that genome sequencing is critical for crop improvement, the size and complexity of the wheat genome has been perceived as an

- Vitis
- Wheat
- Sequence Repository
- Projects
- Data
- Tools
- Triannot Pipeline**
 - Help
 - Versions
 - Architecture
 - Usage
 - Softwares
 - Databanks
 - Defaults
 - Links
 - Presentation in congress
 - Run Pipeline
 - Acknowledgments
- Deletion Bins
- Publications

Triannot Pipeline

[Help](#)

[Versions](#)

[Architecture](#)

[Usage](#)

[Softwares](#)

[Databanks](#)

[Defaults](#)

[Links](#)

[Presentation in congress](#)

[Run Pipeline](#)

[Acknowledgments](#)

How to cite TriAnnot

References

Leroy P, Guilhot N, Sakai H, Bernard A, Choulet F, Theil S, Reboux S, Amano N, Flutre T, Pelegriin C, Ohyanagi H, Seidel M, Giacomoni F, Reichstadt M, Alaux M, Gicquello E, Legeai F, Cerutti L, Numa H, Tanaka T, Mayer K, Itoh T, Quesneville H, Feuillet C (2011) TriAnnot: a versatile and high performance pipeline for the automated annotation of plant genomes *Frontiers in Plant Genetics and Genomics* (in review).

**Paper accepted by Frontiers in Plant Genetics and Genomics
January 4th - edited by Dr. Takuji Sasaki**

Triannot Pipeline

Help

Versions

Architecture

Usage

Softwares

Databanks

Defaults

Links

Presentation in congress

Run Pipeline

Acknowledgments

The pipeline has many dependences.

Here is a non exhaustive list: Oracle, Grid Engine, MySQL database, **73** databanks, Python, Perl, BiopPerl, **21** bioinformatics softwares that are used (ie. *blast*, *RepeatMasker*, *exonerate*, *FGeneSH* (license), *GTallymer*, ...).

TriAnnot uses also other pipelines such as **TEannot** and **EuGene** that will be quite difficult to install as well.

In addition, please note that TriAnnot has been developed on the cluster (8.5 Tflop) at URGI and as not yet been installed in another environment.

Availability

The code of TriAnnot will be available soon upon request (triannot-support@clermont.inra.fr) and groups can choose to install the program in-house instead of running the analysis on the URGI server. However, such installation may require extensive skills in informatics and bioinformatics. INRA will not be able to provide any technical support for the installation except in the framework of formal collaborations.

Triannot Pipeline

[Help](#)

[Versions](#)

[Architecture](#)

[Usage](#)

[Softwares](#)

[Databanks](#)

[Defaults](#)

[Links](#)

[Presentation in congress](#)

[Run Pipeline](#)

[Acknowledgments](#)

Intellectual Property

A first version of TriAnnot has been deposited on January 31th, 2006 to the French APP ("Agence pour la Protection des Programmes") repository for intellectual property (APP # IDDN.FR.001.05008.000.R.C.2006.000.31235).

→ New deposit underway for TriAnnot v3.5

Triannot Pipeline

Help

Versions

Architecture

Usage

Softwares

Databanks

Defaults

Links

Presentation in congress

Run Pipeline

Acknowledgments

In order to use the TriAnnot pipeline, you need :

1. An INRA URGI personal Secure Account (login / password).

This procedure is necessary to protect your data from third party.

To obtain this **Secure Account** click on this link: [online form](#) .

Please select the **project "TriAnnot - M Alaux"** .

2. to sign the "Agreement & Access Rights" .

You can easily download this document [AgreementFileV05.pdf](#) (50.34 kB) sign it, and sent it either :

- by email (scan) at [leroy\[at\]clermont.inra.fr](mailto:leroy[at]clermont.inra.fr)
- by Fax at **+33 4 73 62 44 53** - P. Leroy
- by Post Office at :
P. Leroy,
INRA, UMR 1095
234 Avenue du Brézet
F-63100 Clermont-Ferrand, France

Register

With this form you ask for an [URGI account](#).

This account grants you access to private tools, software and data, signaled by this icon: 

All fields are mandatory. Please wait for an email with your credential.

You can upgrade your account later [upon request](#) to access more services (secure shell, electronic document management, apollo software and more).

You

First name	<input type="text"/>
Last name	<input type="text"/>
Institution	<input type="text"/>
Lab	<input type="text"/>
Email	<input type="text"/>

I agree with the [INRA.IT charter](#).

Your manager

First name	<input type="text"/>
Last name	<input type="text"/>

Our project

Project [Not found?](#)

Are you human?

	
<input type="text"/>	stop spam. read books.

Triannot Pipeline

Help

Versions

Architecture

Usage

Softwares

Databanks

Defaults

Links

Presentation in congress

Run Pipeline

Acknowledgments

In order to use the TriAnnot pipeline, you need :

1. An INRA URGI personal Secure Account (login / password).

This procedure is necessary to protect your data from third party.

To obtain this **Secure Account** click on this link: [online form](#) .

Please select the **project "TriAnnot - M Alaux"** .

2. to sign the "Agreement & Access Rights" .

You can easily download this document [AgreementFileV05.pdf](#) (50.34 kB) sign it, and sent it either :

- by email (scan) at [leroy\[at\]clermont.inra.fr](mailto:leroy[at]clermont.inra.fr)
- by Fax at **+33 4 73 62 44 53** - P. Leroy
- by Post Office at :
P. Leroy,
INRA, UMR 1095
234 Avenue du Brézet
F-63100 Clermont-Ferrand, France

Triannot Pipeline

Help

Versions

Architecture

Usage

Softwares

Databanks

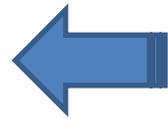
Defaults

Links

Presentation in congress

Run Pipeline

Acknowledgments



Later we will add the history of the TriAnnot pipeline

Versions

25 Nov 2011 TriAnnot v3.5

With this version the TriAnnot pipeline have been considerably improved:

- Parallelization of the PERL code to be able to use the pipeline on a cluster and speed up considerably the annotation process
- Complete new web interface for sequences submission Multi-fasta file allowed (up to ten)
- Sequence size increase up to 3Mb
- New interface to allow the users to manage their analysis and download all the output files
- Exonerate used for spliced transcript alignments
- K-mer composition using an Illumina read sample of sorted wheat chromosome 3B representing 2x coverage
- Augustus wheat matrix for *ab initio* gene prediction
- EuGene wheat matrix for *ab initio* gene prediction
- New GBrowse graphical display
- New EMBL file to be able to use GenomeView for manual curation
- New default analysis (step.xml) for wheat
- Databanks update

06 Dec 2010 TriAnnot v2.1

- Important improvement of the **NIAS-search** module, much more efficient (Developed by H. Sakai, N. Amano from NIAS and P. Leroy from GDEC).
- Databanks :
 - New proteome databanks: CDS derived peptides from genome model and addition of FMBI proteomes for seve

Triannot Pipeline

[Help](#)

[Versions](#)

[Architecture](#)

[Usage](#)

[Softwares](#)

[Databanks](#)

[Defaults](#)

[Links](#)

[Presentation in congress](#)

[Run Pipeline](#)

[Acknowledgments](#)

Species

Vitis

Wheat

Sequence Repository

Projects

Data

Tools

Triannot Pipeline

[Help](#)

[Versions](#)

[Architecture](#)

[Usage](#)

[Softwares](#)

[Databanks](#)

[Defaults](#)

[Links](#)

[Presentation in congress](#)

[Run Pipeline](#)

[Acknowledgments](#)

Deletion Bins

Publications

Links

Botrytis

Leptosphaeria

Microbotryum

Venturia

Arabidopsis

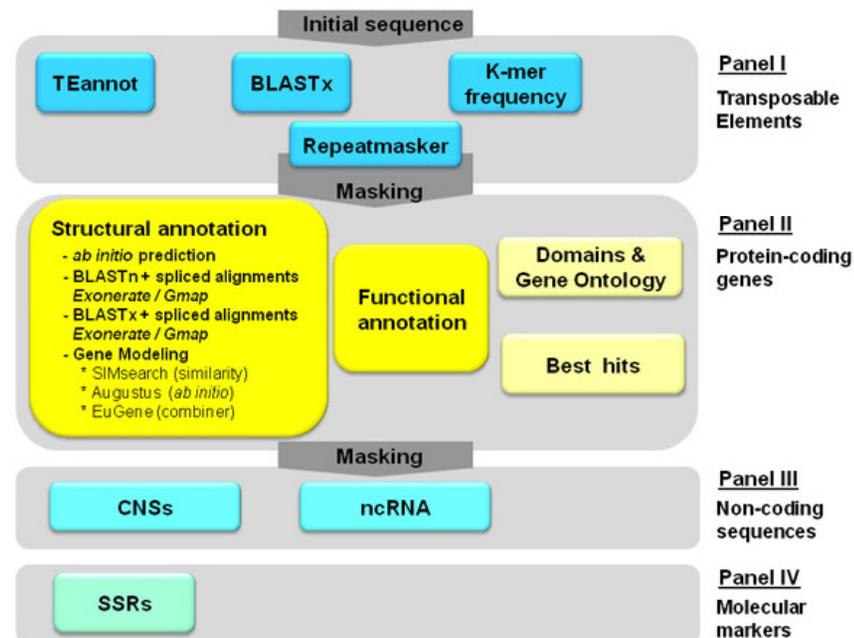
Oryza

Populus

Maize

You are here : [Home](#) / [Home URGI](#) / [Species](#) / [Wheat](#) / [Triannot Pipeline](#) / [Architecture](#)

Architecture



The structural & functional automatic annotation pipeline TriAnnot is divided into three main panels. However, in practice and for description convenience, the TriAnnot pipeline is divided within several blocks or modules. Depending of the block used, you can choose to work on the initial or masked sequence, and modify several parameters.

Below are given the block number and the block name within the 'step.xml' a file needed by the pipeline and which represents the receipt corresponding to a full analysis. Then, we give a short explanation. For details concerning software used and databanks see respectively [Softwares](#) and [Databanks](#). A default receipt is also available [see](#)

Triannot Pipeline

[Help](#)

[Versions](#)

[Architecture](#)

[Usage](#)

[Softwares](#)

[Databanks](#)

[Defaults](#)

[Links](#)

[Presentation in congress](#)

[Run Pipeline](#)

[Acknowledgments](#)

Species

Vitis

Wheat

Sequence Repository

Projects

Data

Tools

Triannot Pipeline

[Help](#)

[Versions](#)

[Architecture](#)

[Usage](#)

[Softwares](#)

[Databanks](#)

[Defaults](#)

[Links](#)

[Presentation in congress](#)

[Run Pipeline](#)

[Acknowledgments](#)

Deletion Bins

Publications

Links

Botrytis

Leptosphaeria

Microbotryum

Venturia

Arabidopsis

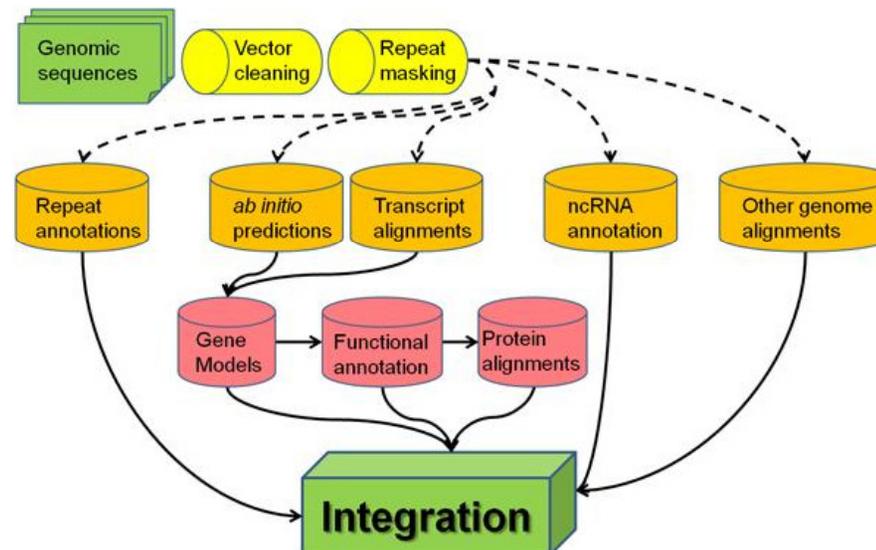
Oryza

Populus

Maize

You are here : [Home](#) / [Home URGI](#) / [Species](#) / [Wheat](#) / [Triannot Pipeline](#) / [Usage](#)

Usage



How to use the pipeline?

Official URLs

Home:

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

Direct access to the submission window with **login/password** :

- <http://uroj.versailles.inra.fr/triannot/>

Triannot Pipeline

Help

Versions

Architecture

Usage

Softwares

Databanks

Defaults

Links

Presentation in congress

Run Pipeline

Acknowledgments

21 programs

Species

You are here : [Home](#) / [Home URGI](#) / [Species](#) / [Wheat](#) / [Triannot Pipeline](#) / [Softwares](#)

Vitis

Wheat

Sequence Repository

Projects

Data

Tools

Triannot Pipeline

Help

Versions

Architecture

Usage

Softwares

Databanks

Defaults

Links

Presentation in congress

Run Pipeline

Acknowledgments

Deletion Bins

Publications

Links

Botrytis

Leptosphaeria

Microbotryum

Venturia

Arabidopsis

Oryza

Populus

Softwares

The program versions used by the pipeline, for a specific analysis, are provided within the first line of the GFF3 output files.

- **SIMsearch** - Bioinformatic Group at [NIAS](#) laboratory, Japan
 - **Version AUG2011**
 - Perl programs developed by H. Sakai, N. Amano, H. Numa, T. Tanaka & T. Itoh
- **BLAST** [[Altschul et al., 1990 J. Mol. Biol 215:403-410](#)]
 - **Version 2.2.21**
 - Basic Local Alignment and Search Tool which finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families. see description at [NCBI](#)
- **RepeatMasker** [[Smit 1993 Nucleic Acids Res. 21, 1863-1872](#)]
 - **Version 3-2-6**
 - RepeatMasker screens DNA sequences in FASTA format against a library of repetitive elements and returns a masked query sequence ready for database searches. RepeatMasker also generates a table annotating the masked regions.
- **wu-blast** - BLAST 2.0 from Washington University (used by RepeatMasker)
 - **Version 2.2.6**
 - See description at Washington University in St Louis.
 - Today, Rights to BLAST 2.0 (WU-BLAST) have been acquired by Advanced Biocomputing, LLC. All interested parties are hereby referred to: [Advanced Biocomputing](#) , LLC
- **cross-match** - [Phil Green Group](#) at University of Washington, US (used by RepeatMasker)
 - **Version 0.990329**
 - cross_match is a general purpose utility for comparing any two DNA sequence sets using a 'banded' version of swat.



Triannot Pipeline

Help

Versions

Architecture

Usage

Softwares

Databanks

Defaults

Links

Presentation in congress

Run Pipeline

Acknowledgments

73 databanks

Updated twice a year

Species

Vitis

Wheat

Sequence Repository

Projects

Data

Tools

Triannot Pipeline

Help

Versions

Architecture

Usage

Softwares

Databanks

Defaults

Links

Presentation in congress

Run Pipeline

Acknowledgments

Deletion Bins

Publications

Links

Botrytis

Leptosphaeria

Microbotryum

Venturia

Arabidopsis

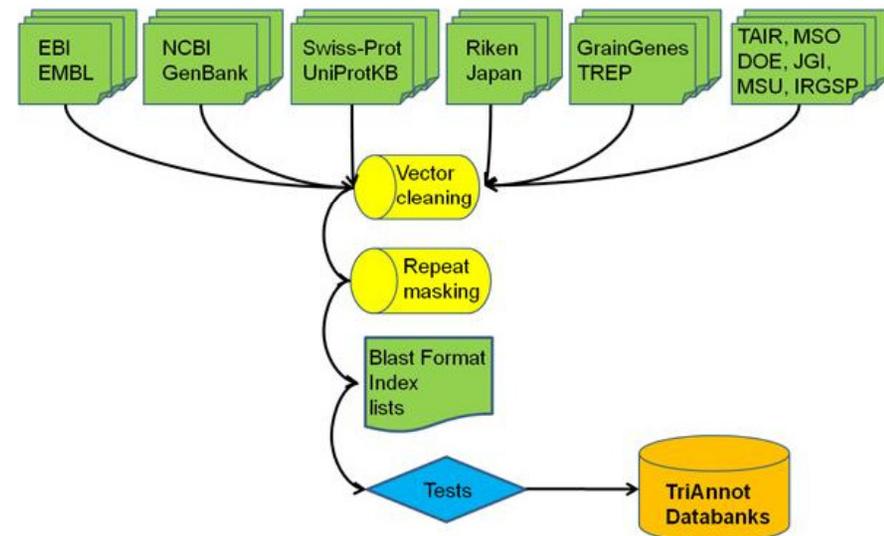
Oryza

Populus

Maize

You are here : [Home](#) / [Home_URGI](#) / [Species](#) / [Wheat](#) / [Triannot Pipeline](#) / [Databanks](#)

Databanks



The databank versions used by the pipeline for a given analysis are displayed within the first line of the GFF3 output files. Below, the TriAnnot databank names are written in **orange**. This name is used to identify TriAnnot pipeline output gff and/or embi files.

Transposable Elements and repeat databanks for annotation and masking

- **TREP** - TREPnr, TREPprot, TREPtotal
 - **Version 10**
 - The *Triticaceae* Repeat Sequence Databases from Thomas Wicker *et al.* [Wicker *et al.*, 2002 *Trends in Plant Science* 7:561-562]
 - **TREP_nr; TREP_prot; TREP_total**

Triannot Pipeline

Help

Versions

Architecture

Usage

Softwares

Databanks

Defaults

Links

Presentation in congress

Run Pipeline

Acknowledgments

Wheat

➔ Barley, Rice, Maize, Sugarcane, Oak

Species

Vitis

Wheat

Sequence Repository

Projects

Data

Tools

Triannot Pipeline

Help

Versions

Architecture

Usage

Softwares

Databanks

Defaults

Links

Presentation in congress

Run Pipeline

Acknowledgments

Deletion Bins

Publications

Links

Botrytis

Leptosphaeria

Microbotryum

Venturia

Arabidopsis

Oryza

Populus

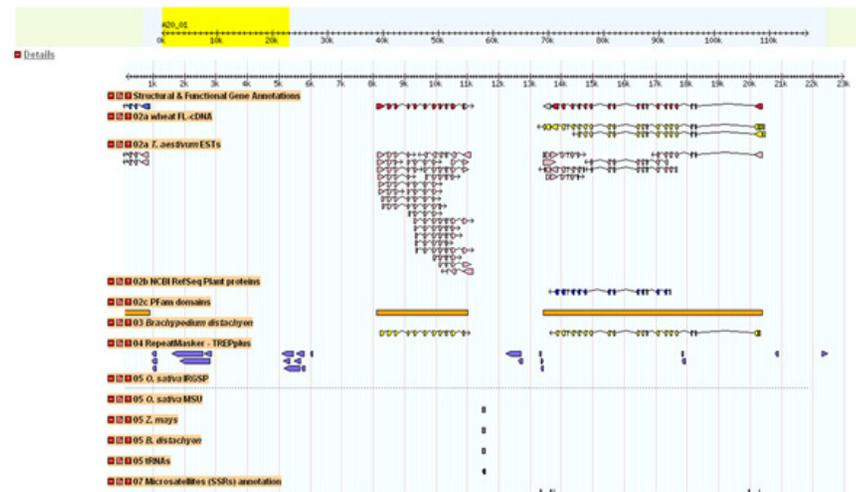
Maize

Medicago

Pisum

You are here : [Home](#) / [Home URGI](#) / [Species](#) / [Wheat](#) / [Triannot Pipeline](#) / Defaults

Defaults



Default parameters are used to launch an automatic analysis using the TriAnnot pipeline (for a full description of all steps see the pipeline [architecture](#)). These default parameters can be changed, time by time, to improve the quality of the annotation. These default parameters should be, at the end, defined by the [IWGSC bioinformatics working group](#) . Default parameters are set, at present, as follow for wheat sequence annotations:

- **Panel I - Step1 - RepeatMasker**
 - Initial sequence masking using RepeatMasker against TREPplus as external libraries. [cutoff 225; 'qq' parameter (rush job) about 10% less sensitive, 4 to 10 times faster than default repeat options; masked with lower-case and Ns]
- **Panel I - Step2 - BLASTx**
 - Initial sequence annotation of Transposable Elements using BLASTx against TREPprot (proteins) [Evalue 10-5]
- **Panel II - Step5 - *ab initio* Prediction**

Triannot Pipeline

[Help](#)

[Versions](#)

[Architecture](#)

[Usage](#)

[Softwares](#)

[Databanks](#)

[Defaults](#)

[Links](#)

[Presentation in congress](#)

[Run Pipeline](#)

[Acknowledgments](#)

Species

[Vitis](#)

[Wheat](#)

[Sequence Repository](#)

[Projects](#)

[Data](#)

[Tools](#)

[Triannot Pipeline](#)

[Help](#)

[Versions](#)

[Architecture](#)

[Usage](#)

[Softwares](#)

[Databanks](#)

[Defaults](#)

[Links](#)

[Presentation in congress](#)

[Run Pipeline](#)

[Acknowledgments](#)

[Deletion Bins](#)

[Publications](#)

[Links](#)

[Botrytis](#)

[Leptosphaeria](#)

[Microbotryum](#)

[Venturia](#)

[Arabidopsis](#)

[Oryza](#)

[Populus](#)

[Maize](#)

[Medicago](#)

[Pisum](#)

[Flax](#)

You are here : [Home](#) / [Home URGI](#) / [Species](#) / [Wheat](#) / [Triannot Pipeline](#) / [Links](#)

Links

Wheat useful links

- [INRA Wheat URGI portal](#)
- [Guidelines for Annotating Wheat Genomic Sequences: IWGSC_AnnotationGuidelines_ReleaseV1.pdf](#) - 23.19 ko

Institutions

- [INRA - Institut National de la Recherche Agronomique](#)

Research Centre

- [UMR INRA-UBP 1095 - Génétique, Diversité & Ecophysiologie des Céréales – GDEC](#)

Networks

- [Triticeae Genomics Vision - 2006](#)
- [IWGSC - International Wheat Genome Sequencing Consortium](#)
- [ITMI - International Triticeae Mapping Initiative](#)
- [ETGI - European Triticeae Genomics Initiative](#)
- [ITEC - International Triticeae EST Cooperative](#)
- [US Wheat Genome Project](#)
- [IBSC - International Barley Sequencing Consortium](#)

Databases

- [GrainGenes](#)
- [TREP - Triticeae Repeat Sequence Database @ GrainGenes](#)
- [Wheat at NCBI - National Center for Biotechnology Information](#)
- [GnpSeq query - the URGI EST wheat database informations](#)

Pipelines for genome annotation

- [RiceGAAS](#)
- [fpdp - Flowerin Plant Gene Picker](#)
- [MAKER](#)
- [DNA subway](#)
- [DAWGPAWS - sourceforge](#)

Triannot Pipeline

[Help](#)

[Versions](#)

[Architecture](#)

[Usage](#)

[Softwares](#)

[Databanks](#)

[Defaults](#)

[Links](#)

[Presentation in congress](#)

[Run Pipeline](#)

[Acknowledgments](#)

Since 2006 ...

Species

Vitis

Wheat

Sequence Repository

Projects

Data

Tools

Triannot Pipeline

[Help](#)

[Versions](#)

[Architecture](#)

[Usage](#)

[Softwares](#)

[Databanks](#)

[Defaults](#)

[Links](#)

[Presentation in congress](#)

[Run Pipeline](#)

[Acknowledgments](#)

Deletion Bins

Publications

Links

Botrytis

Leptosphaeria

Microbotryum

Venturia

Arabidopsis

Oryza

Populus

Maize

Medicago

Pisum

You are here : [Home](#) / [Home URGI](#) / [Species](#) / [Wheat](#) / [Triannot Pipeline](#) / Presentation in congress

Presentation in congress

2011 - Leroy P, Guilhot N, Sakai H, Bernard A, Choulet F, Pelegrin C, Reboux S, Flutre T, Amano N, Seidel M, Ohyanagi H, Alaux M, Numa H, Tanaka T, Mayer K, Itoh T, Quesneville H, Feuillet C.

TriAnnot: a user friendly web interface for structural and functional automatic annotation of plant genomes.

In: Plant & Animal Genome XIX, January 15-19, 2011, San Diego, CA (USA).

2011 - Leroy P, Guilhot N, Sakai H, Bernard A, Choulet F, Pelegrin C, Reboux S, Flutre T, Amano N, Seidel M, Ohyanagi H, Alaux M, Numa H, Tanaka T, Mayer K, Itoh T, Quesneville H, Feuillet C.

TriAnnot: a user friendly web interface for structural and functional automatic annotation of plant genomes.

In: Comparative & Regulatory Genomics in Plants, April 11-12, 2011, Gent, Belgium.

2011 - Leroy P.

Nouveau paradigme pour l'analyse des données biologique.

In: Journées GRISBI 2011, Mai 26, 2011, PRABI-HBCP, Lyon, France.

2011 - Leroy P, Bernard A, Guilhot N, Theil S, Alaux M, Reboux S, Inizean O, Choulet F, Sakai H, Tanaka T, Itoh T, Quesneville H, Feuillet C.

TriAnnot: a high performance pipeline for the automated structural and functional of plant genomes.

In: JOBIM, June 28 – July 1er, 2011, Paris, France, pp 217. Eds E. Barillot, C. Froidevaux & E Rocha.

2011 - Leroy P.

Tentative de mise sur Grille de Calcul du pipeline d'annotation automatique structural et fonctionnel TriAnnot dans le cadre du projet Européen LifeGrid 2006-2008.

In: Journée de sensibilisation au calcul intensif en biologie intégrative et écologies, Septembre 14, 2011, Bordeaux, France.

2011 - Leroy P.

TriAnnot - un pipeline parallélisé (cluster) pour l'annotation structurel et fonctionnel des génomes de plantes.

In: TWIST meeting, Septembre 19, 2011, Paris, France.

2011 - Leroy P, Bernard A, Guilhot N, Theil S, Alaux M, Reboux S, Inizean O, Choulet F, Sakai H, Tanaka T, Itoh T, Quesneville H, Feuillet C

TriAnnot: a high performance pipeline for the automated structural and functional of plant genomes.

In: Anais 28° encontro sobre temas de genetic e malhoramento, Vol 28, "Genomica e bioinformatica", Outubro 10 – 11, 2011, Piracicaba, SP, Brazil, pp 31. Eds C.B. Monteiro-Vitorello, J.B. Pinheiro, M.L. Carneiro Vieira, G. Bandel, A. De Mello Varani & M. Mendes Brandao.

Triannot Pipeline

Help

Versions

Architecture

Usage

Softwares

Databanks

Defaults

Links

Presentation in congress

Run Pipeline

Acknowledgments

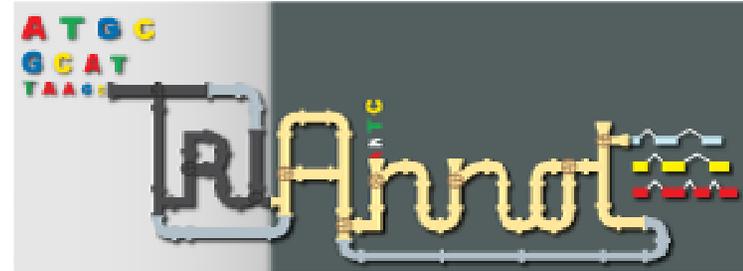
Authentification requise ✕

 Le site <http://urgi.versailles.inra.fr> demande un nom d'utilisateur et un mot de passe. Le site indique : « URGI »

Utilisateur :

Mot de passe :

4:05 – 4:25 New Interface *N. Guilhot*
Download
GBrowse



V3.5

- X** **New Interface**
- X** **Download TriAnnot output files**
- X** **New GBrowse**



Fill user profile on first connexion

TriAnnot Pipeline

My Analyses

My profile

About TriAnnot

Welcome to TriAnnot Web Interface

We are pleased to welcome you on the TriAnnot Pipeline web interface. As this is your first visit, we need you to fill in your TriAnnot user profile. Please double check your email address as it will be used to send you your analysis results.

Login name: *nguillhot*

First name:

Last name:

Email address:

Company/Institution:

Country:

Submit



Editing User Profile

TriAnnot Pipeline

My Analyses

My profile

Admin

About TriAnnot

Edit TriAnnot user profile

Login name: *nguilhot*

First name:

Last name:

Email address:

Company/Institution:

Country:

Submit



Submitting an analysis

TriAnnot Pipeline

My Analyses

My profile

Admin

About TriAnnot

TRIANNOT PIPELINE ANALYSIS SUBMISSION

Analysis parameters

Analysis title [?]

Pipeline template [?]

Wheat IWGSC Annotation ▾

Enter Query Sequence

Enter FASTA sequence (> 1000 bp and < 3 Mbp) [?]

Or, upload file [?]

Parcourir...

Clear

Submit analysis



Submitting an analysis

TriAnnot Pipeline

My Analyses

My profile

Admin

About TriAnnot

TRIANNOT PIPELINE ANALYSIS SUBMISSION

Analysis parameters

Analysis title 

Pipeline template 

Wheat IWGSC Annotation ▾

Enter Query Sequence

Enter FASTA sequence (> 1000 bp and < 3 Mbp) 

You can enter a single nucleotide sequence in fasta format or up to 10 sequences in multifasta format. The length of each sequence must be between 1000 bp and 3 Mbp. Sample input:

```
>Contig385B22 from BAC T. aestivum BAC library Pool A
ATGCTTCGGGGAACCTGTCGAAATTT
ATTCCGTTCCAATTGGTTACGATTGAG
ATTCCGAATCCGGAATTGGCCAATGC
```

Or, upload file 

Parcourir...

Clear

Submit analysis

Analysis Template file overview

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<analysis triannot_version="3.5" description="Wheat_default.xml -
Default Analysis - October 21th, 2011 - Ph Leroy">

  <!-- Transposable Annotation and Masking -->
  <program id="1" step="1" type="RepeatMasker" sequence="initial">
    <parameter name="database">TREP_plus</parameter>
  </program>

  <program id="2" step="1" type="SequenceMasker" sequence="initial">
    <dependences>
      <dependence id="1" type="RepeatMasker"/>
    </dependences>
    <parameter name="masked_sequence">TE_N_masked.seq</parameter>
    <parameter name="masking_mode">N</parameter>
    <parameter name="input_type">XM</parameter>
  </program>
  ...
```



Analyses overview

[TriAnnot Pipeline](#)
[My Analyses](#)
[My profile](#)
[Admin](#)
[About TriAnnot](#)

MY TRIANNOT PIPELINE ANALYSES

[Pending \(0\)](#)
[Running \(0\)](#)
[Finished \(4\)](#)
[Failed \(2\)](#)
[All \(6\)](#)

Search:

Show entries

<input type="checkbox"/>	Status	Submitted	Progress	Title	Sequence	Sequence Length	Started	Finished	Pipeline Template	TriAnnot version
<input type="checkbox"/>	✗	12-01-12 10:33	0%	Just a test on a short sequence	ANF4514	4367	12-01-12 10:35	12-01-12 10:45	Wheat IWGSC Annotation	3.5
<input type="checkbox"/>	✗	12-01-12 10:33	0%	Another test	ANF4514	1257	12-01-12 10:35	12-01-12 10:45	Wheat IWGSC Annotation	3.5
<input type="checkbox"/>	✓ 📄	12-01-03 11:18	100%	ctg0464b_000000001_000200000	ctg0464b_000000001_000200000	200000	12-01-03 11:20	12-01-03 12:02	Wheat IWGSC Annotation	3.5
<input type="checkbox"/>	✓ 📄	11-12-31 12:14	100%	BACsynth12	Synth12	127860	11-12-31 12:15	11-12-31 13:33	Wheat IWGSC Annotation	3.5
<input type="checkbox"/>	✓ 📄	11-12-20 12:18	100%	Scaffolds A20	A20	117015	11-12-20 12:20	11-12-20 13:09	Wheat IWGSC Annotation	3.5
<input type="checkbox"/>	✓ 📄	11-12-07 11:51	100%	Test_3Mb	ctg954_3Mb	2999940	11-12-23 18:15	11-12-24 11:39	Wheat IWGSC Annotation	3.5

Showing 1 to 6 of 6 entries

With selected rows: [Download Results](#) [Delete](#)

Analyses is waiting in the queue for a free slot

Analyses is in progress

Analyses has been processed successfully

Analyses has failed

View results with GBrowse

Download results file

Email sent when analysis is done

Dear Philippe Leroy,

Results for the following TriAnnot analysis are available for download:

16 - "Scaffolds A20" submitted on "2011-12-20 13:18:01"

You can download results files from "My Analyses" section on TriAnnot pipeline web interface.

<http://urgi.versailles.inra.fr/triannot/?results>

Download output files

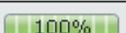
ATGCGCATTAA... **TriAnnot** Version 3.5 BETA

TriAnnot Pipeline | **My Analyses** | My profile | Admin | About TriAnnot

MY TRIANNOT PIPELINE ANALYSES

Pending (0) | Running (0) | Finished (2) | Failed (0) | **All (2)**

Search: Show entries

<input type="checkbox"/>	Status	Submitted	Progress	Title	Sequence	Sequence Length	Started	Finished	Pipeline Template	TriAnnot version
<input type="checkbox"/>	 	11-12-20 12:18		Scaffolds A20	A20	117015	11-12-20 12:20	11-12-20 13:09	Wheat IWGSC Annotation	3.5
<input type="checkbox"/>	 	11-12-07 11:51		Test_3Mb	ctg954_3Mb	2999940	11-12-23 18:15	11-12-24 11:39	Wheat IWGSC Annotation	3.5

Showing 1 to 2 of 2 entries

With selected rows:

 Analysis is waiting in the queue for a free slot
 Analysis is in progress
 Analysis has been processed successfully
 Analysis failed

 Click on this icon to download zipped result files
 Click on this icon to view results with GBrowse

[Contact us](#)
Copyright ©2011 INRA GDEC

Sequences

- [-] TriAnnotResults_00016
 - EMBL
 - GFF
 - Other files
 - sequences



Name ▲	Size	Type
 initial.seq	116 KB	SEQ File
 proteins.seq	3 KB	SEQ File
 TE_and_Genes_lower_case_masked.seq	116 KB	SEQ File
 TE_and_Genes_N_masked.seq	116 KB	SEQ File
 TE_lower_case_masked.seq	116 KB	SEQ File
 TE_N_masked.seq	116 KB	SEQ File



proteins.seq

>A20_NIAS_13417_20382_gene_0001_mRNA_0001

MGEYCAPAAAAEAEGERPVAVSVLPLPPVAAAGRYLYGDYDRCSTKQVFD
NLHGNIISLDPLAREFVDTEEFQRLRLDLKQLGLTYLVFPGAVHTRFEHSLG
VYRLAGEAMNNLQKYQGNELGIDRIDVQTVKLAGLLHDIGHGPFSHLFEH
EFLPRVHPGSTWSHEHMSALLLDSIVDKHAIDIEPDYLVKVIKEMIVASSD
VSTAEGVKEKRFLYDIVANGRNGIDVDKFDYIDRDCRACGIGSNFQHWRL
LEGMRVMGDEICYPADYLSIHKLFTTRADLHRTVYTHAKVKAVELMLVD
ALVEANEYLGISLHADDPEDFWKLDDTIVKSIETAPNDELKKAKEIIQRI
RRRELYKFCNQYSVPKDKLDHFKNITAQDIVCSQITSKVLLKEEDVAVSN
VKIDLTRGKDNPLESSVMPGFRMRREVDPNGRAREPPAARVQPGQDREG
VRQEAGAGGGGVGGV*

>A20_AUGUSTUS_8122_11021_gene_0002_mRNA_0001

MTQLQAAPLKTLSFSGRRSAVAAARPRHATATFRCSDAARSYIITLLPGD
GIGPEVVAVAKDVLSAAGAKEGVLELRFREMLMGGAALDAVGVPLPDETLA
AAQASDAILLGAIGGYKWSNEKHLKPETGLLNIRAGLGVFANLRPATVL
PQLVDASTLKKEVAEGVDIMVVRELTGGIYFGQPRGFGTNDMGEEIGFNT
EIYSVSEIDRIARVAFEVARKRGGKLCSDKANVLEASMLWRKRVTAIAS
EFPDVELSHMYVDNASMQLVRNPKQFDTIVTNNIFGDILSDEASMITGSI
GMLPSASVGESGPGFLFEP IHGSAPDIAGQDKANPLATILSAAMLLKYGLG
AETAAKRIETAVTETLDNGFRTGDIYSPGTTLVGCKRMGEEVLKALESQK
*

>A20_EUGENE_16_854_gene_0001_mRNA_0001

MKHAALLLILAAAHVLLPWAQCHRLNPEIHDAGGLSRQGFPEGFVFGT
AASAYQVEGMAEQGGRGPSIWDAFIKIPGTIAGNGTADVAVDEYHRYKED
VDIMKNMGFDAYRESISWSRIFPDGTGKVNWEGVDYYNRLIDYMLQQGNW
NFYTRQEA*

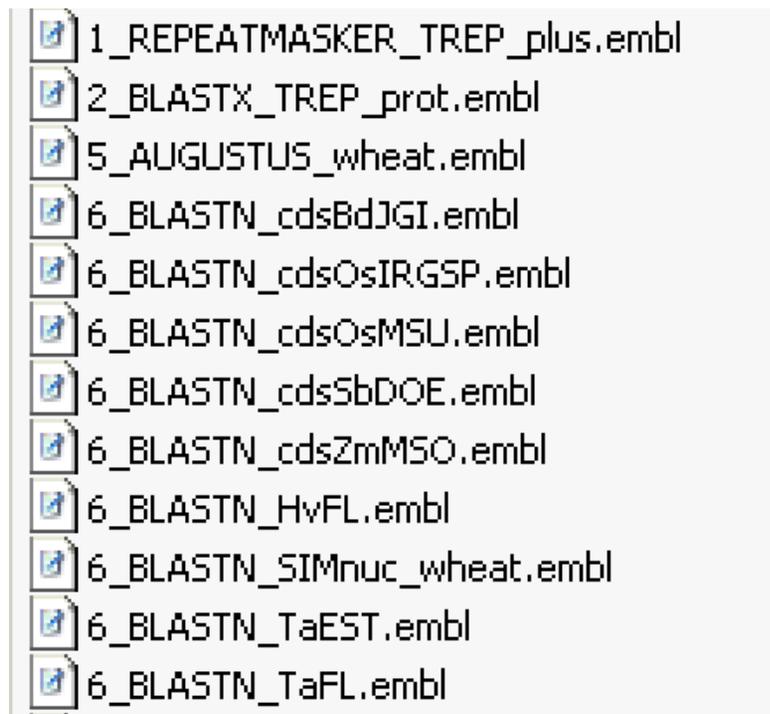
>A20_EUGENE_50225_50590_gene_0005_mRNA_0001

METQAPPASLDLSLALATMPQPLPPPAAAAPPLSLQAAGDAVSSAVAGAG
WKVFSCLFYEKKFLKSQALGGHQNAHRKDRGAAGWNASLYLPAADRPWPP
TTATSHPEIGDENQLDLSLKL*

Result files in GFF and EMBL format



In addition the TriAnnot pipeline provides a file which can be used with the graphical viewer GBrowse and the appropriate configuration file.
16_Selector_A20.gff



Rules of GFF & EMBL files naming

<http://urgi.versailles.inra.fr/Species/Wheat/Triannot-Pipeline/Usage>

- **EMBL** folder

- In each folder files are tagged and follows the following rules:

- A number related to the step number
- The type of programs *i.e.* REPEATMASKER; AUGUSTUS; BLASTX; BLASTN;BLASTP; EXONERATE; EUGENE; GENEMODEL; BESTHIT; TRNASCAN-SE; TRF.
- The databank used (see [databanks](#)). When no databank is used, just the type of program is displayed *i.e.* TRF. For *ab initio* gene prediction programs the matrix used is displayed *i.e.* AUGUSTUS_wheat.
- Extension **.embl**

- **Few examples**

- Step1 - *Transposable Elements annotation & masking*

- 1_REPEATMASKER_TREP_plus.embl

- Step2 - *BLASTx against TREPprot*

- 2_BLASTX_TREP_prot.embl

2_BLASTX_TREP_prot.embl

- Step5 - *ab initio gene prediction*

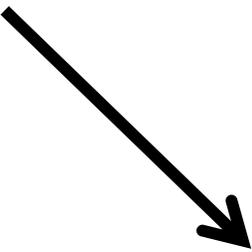
- 5_AUGUSTUS_wheat.embl

- Step6 - *BLASTn / Exonerate*

- 6_BLASTN_cdsBdJGI.embl
- 6_BLASTN_cdsOsIRGSP.embl
- *etc.*

Protein alignments

- [-] TriAnnotResults_00016
 - EMBL
 - GFF
 - Other files
 - sequences



- 10_BESTHIT_BLASTP_nr.align
- 10_BESTHIT_BLASTP_protBdJGI.align
- 10_BESTHIT_BLASTP_prothOR.align
- 10_BESTHIT_BLASTP_protOsIRGSP.align
- 10_BESTHIT_BLASTP_protPlant.align
- 10_BESTHIT_BLASTP_protZmMSO.align
- Global_XM_for_TE_N_masked.xml

Gbrowse selecting tracks

Browser | **Select Tracks** | Snapshots | Custom Tracks | Preferences

<< [Back to Browser](#) | [Show Favorites Only](#) ★ | [Clear All Favorites](#) ☆

■ **Tracks**

■ **01 - Gene Models** All on All off

- ★ ✓ Structural & Functional Gene Annotations [?]

■ **02 - Biological Evidences**

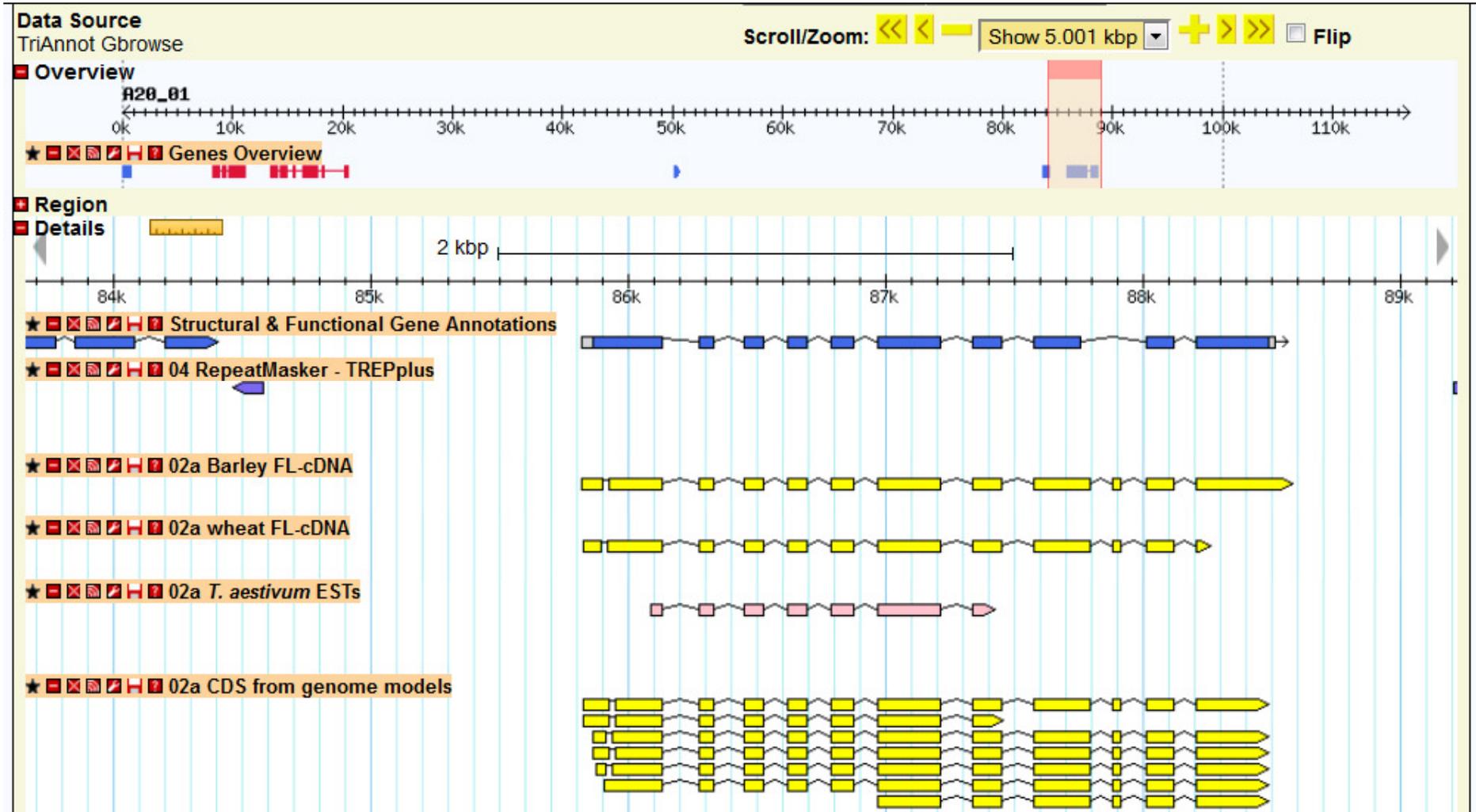
■ **A - Transcripts (BLASTn)** All on All off

- ★ ✓ 02a CDS from genome models [?]
- ★ ✓ 02a Barley FL-cDNA [?]
- ★ ✓ 02a wheat FL-cDNA [?]
- ★ ✓ 02a *T. aestivum* ESTs [?]

■ **B - Proteins (BLASTx)** All on All off

- ★ ✓ 02b UniProtKB/Swiss-Prot [?]
- ★ ✓ 02b Protein derived rice (MSU) genome annotation [?]
- ★ ✓ 02b wheat proteome [?]
- ★ ✓ 02b NCBI RefSeq Plant proteins [?]
- ★ ✓ 02b Protein derived *B. distachyon* genome annotation [?]
- ★ ✓ 02b Protein derived rice (IRGSP) genome annotation [?]
- ★ ✓ 02b *Hordeum* proteome [?]

Viewing TriAnnot results with GBrowse



Viewing TriAnnot results with GBrowse

Data Source
TriAnnot Gbrowse

Scroll/Zoom: << < - Show 5.001 kbp + > >> Flip

Overview

R20_01

0k 10k 20k 30k 40k 50k 60k 70k 80k 90k 100k 110k

★ Genes Overview

Region

Details

2 kbp

86k 87k 88k 89k 90k 91k

★ Structural & Functional Gene Annotations

★ 04 RepeatMasker - TREPplus

★ 02a Barley FL-cDNA

★ 02a wheat FL-cDNA

★ 02a *T. aestivum* ESTs

★ 02a CDS from genome models

Feature type
mRNA

Available annotation
HMMScan - domain_containing_protein --> PF00232.11 - Glycosyl hydrolase family 1 (7.8e-155)

Please, click on the feature to get more informations

Viewing feature details with GBrowse

Functional Gene Annotations
er - TREP plus

DNA
DNA
ESTs
genome

Global informations

Type :	mRNA
Start position :	8122
End position :	11021
Length :	2900
Strand :	+

Additional informations

Function coverage :	90.12
Function target :	P93832 (35..399)
Function identity :	81.92
Structure target :	Cat01_tplb0041 24
Annotation :	known_function - LEU32_ARATH SwissProt_DB 3-isopropylmalate dehydrogenase 2 chloroplastic OS Arabidopsis thaliana GN IMDH2 PE 2 SV 1
Parent feature :	A20_01_NIAS_8122_11021_gene_0002

Sub features

Sub structure : This mRNA feature contains 11 exon(s) - (5'UTR region(s): 1 - 3'UTR region(s): 1)

Get complete CDS : [Display](#) / [Download](#)

Get protein : [Display](#) / [Download](#)

Usefull links

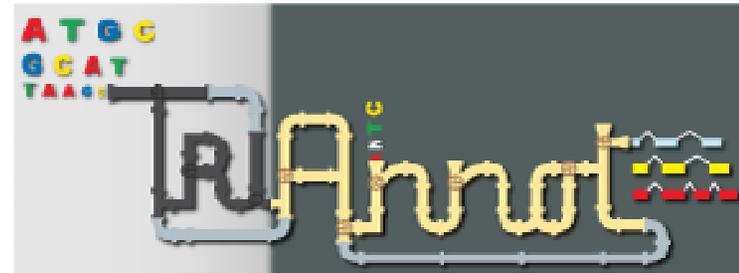
Zoom in : [Zoom on the feature region](#)

Gbrowse details page : [Show all details](#)

Get BAC segment : [Display genomic sequence](#) / [Download genomic sequence](#)

Copyright 2012 INRA GDEC - Contact us

4:25 – 5:25 How to use *P. Leroy & N. Guilhot*
with your own computer



V3.5

This is not a tutorial about how to annotate a sequence 😊



GenomeView (1838)

<http://genomeview.org/>

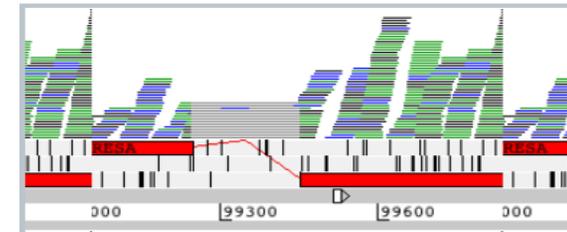
Abeel *et al.* (2011) *Nucleic Acids Res.* doi: 10.1093/nar/gkr995



Artemis (13.2.0)

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

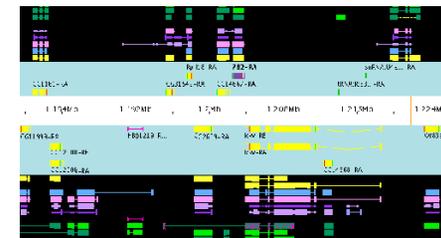
Carver *et al.* (2008) *Bioinformatics* 24, 2672-2676



Apollo (1.11.6)

<http://apollo.berkeleybop.org/current/index.html>

Lewis *et al.* (2002) *Genome Biology* 3, research0082



- 1_REPEATMASKER_TREP_plus.embl
- 2_BLASTX_TREP_prot.embl
- 5_AUGUSTUS_wheat.embl ←
- 6_BLASTN_cdsBdJGI.embl
- 6_BLASTN_cdsOsIRGSP.embl
- 6_BLASTN_cdsOsMSU.embl
- 6_BLASTN_cdsSbDOE.embl
- 6_BLASTN_cdsZmMSO.embl
- 6_BLASTN_HvFL.embl
- 6_BLASTN_SIMnuc_wheat.embl
- 6_BLASTN-TaEST.embl
- 6_BLASTN-TaFL.embl
- 6_EXONERATE_cdsBdJGI.embl
- 6_EXONERATE_cdsOsIRGSP.embl
- 6_EXONERATE_cdsOsMSU.embl
- 6_EXONERATE_cdsSbDOE.embl
- 6_EXONERATE_cdsZmMSO.embl
- 6_EXONERATE_HvFL.embl
- 6_EXONERATE_SIMnuc_wheat.embl
- 6_EXONERATE-TaEST.embl
- 6_EXONERATE-TaFL.embl
- 7_BLASTX_protBdJGI.embl
- 7_BLASTX_protHOR.embl
- 7_BLASTX_protOsIRGSP.embl
- 7_BLASTX_protOsMSU.embl
- 7_BLASTX_protSbDOE.embl
- 7_BLASTX_protTRI.embl
- 7_BLASTX_protZmMSO.embl
- 7_BLASTX_refSeqPlnProt.embl
- 7_BLASTX_SIMprot_wheat.embl
- 7_BLASTX_uniprot_sprot.embl
- 7_EXONERATE_protBdJGI.embl
- 7_EXONERATE_prothOR.embl

- 7_EXONERATE_protOsIRGSP.embl
- 7_EXONERATE_protOsMSU.embl
- 7_EXONERATE_protSbDOE.embl
- 7_EXONERATE_protTRI.embl
- 7_EXONERATE_protZmMSO.embl
- 7_EXONERATE_refSeqPlnProt.embl
- 7_EXONERATE_SIMprot_wheat.embl
- 7_EXONERATE_uniprot_sprot.embl
- 8_EUGENE.embl ←
- 9_GENEMODEL.embl ←
- 10_BESTHIT_BLASTP_nr.embl
- 10_BESTHIT_BLASTP_protAtTAIR.embl
- 10_BESTHIT_BLASTP_protBdJGI.embl
- 10_BESTHIT_BLASTP_protHOR.embl
- 10_BESTHIT_BLASTP_protOsIRGSP.embl
- 10_BESTHIT_BLASTP_protOsMSU.embl
- 10_BESTHIT_BLASTP_protPlant.embl
- 10_BESTHIT_BLASTP_protSAC.embl
- 10_BESTHIT_BLASTP_protSbDOE.embl
- 10_BESTHIT_BLASTP_protTRI.embl
- 10_BESTHIT_BLASTP_protZmMSO.embl
- 10_EXONERATE_nr.embl
- 10_EXONERATE_protAtTAIR.embl
- 10_EXONERATE_protBdJGI.embl
- 10_EXONERATE_protHOR.embl
- 10_EXONERATE_protOsIRGSP.embl
- 10_EXONERATE_protOsMSU.embl
- 10_EXONERATE_protPlant.embl
- 10_EXONERATE_protSAC.embl
- 10_EXONERATE_protSbDOE.embl
- 10_EXONERATE_protTRI.embl
- 10_EXONERATE_protZmMSO.embl
- 11_INTERPROSCAN.embl ←

- 12_BLASTN_genoAtTAIR.embl
- 12_BLASTN_genoBdJGI.embl
- 12_BLASTN_genoOsIRGSP.embl
- 12_BLASTN_genoOsMSU.embl
- 12_BLASTN_genoSbDOE.embl
- 12_BLASTN_genoZmMSO.embl
- 12_BLASTN_refSeq_chloro.embl
- 12_BLASTN_refSeq_mito.embl
- 13_BLASTX_protAtTAIR.embl
- 13_BLASTX_protBdJGI.embl
- 13_BLASTX_protOsIRGSP.embl
- 13_BLASTX_protOsMSU.embl
- 13_BLASTX_protSbDOE.embl
- 13_BLASTX_protZmMSO.embl
- 14_TRNASCAN-SE.embl
- 15_TRF.embl

Gene Annotation

● **GenomeView (1838)**

<http://genomeview.org/>

Abeel *et al.* (2011) **Nucleic Acids Res.** doi: 10.1093/nar/gkr995



[1_REPEATMASKER_TREP_plus.embl](#)
[2_BLASTX_TREP_prot.embl](#)
[5_AUGUSTUS_wheat.embl](#)
[6_BLASTN_cdsBdJGI.embl](#)
[6_BLASTN_cdsOsIRGSP.embl](#)
[6_BLASTN_cdsOsMSU.embl](#)
[6_BLASTN_cdsSbDOE.embl](#)
[6_BLASTN_cdsZmMSO.embl](#)
[6_BLASTN_HvFL.embl](#)
[6_BLASTN_SIMnuc_wheat.embl](#)
[6_BLASTN-TaEST.embl](#)
[6_BLASTN-TaFL.embl](#)
[6_EXONERATE_cdsBdJGI.embl](#)
[6_EXONERATE_cdsOsIRGSP.embl](#)
[6_EXONERATE_cdsOsMSU.embl](#)
[6_EXONERATE_cdsSbDOE.embl](#)
[6_EXONERATE_cdsZmMSO.embl](#)
[6_EXONERATE_HvFL.embl](#)
[6_EXONERATE_SIMnuc_wheat.embl](#)
[6_EXONERATE-TaEST.embl](#)
[6_EXONERATE-TaFL.embl](#)
[7_BLASTX_protBdJGI.embl](#)
[7_BLASTX_prothOR.embl](#)
[7_BLASTX_protOsIRGSP.embl](#)
[7_BLASTX_protOsMSU.embl](#)
[7_BLASTX_protSbDOE.embl](#)
[7_BLASTX_protTRI.embl](#)
[7_BLASTX_protZmMSO.embl](#)
[7_BLASTX_refSeqPlnProt.embl](#)
[7_BLASTX_SIMprot_wheat.embl](#)
[7_BLASTX_uniprot_sprot.embl](#)
[7_EXONERATE_protBdJGI.embl](#)
[7_EXONERATE_prothOR.embl](#)

[7_EXONERATE_protOsIRGSP.embl](#)
[7_EXONERATE_protOsMSU.embl](#)
[7_EXONERATE_protSbDOE.embl](#)
[7_EXONERATE_protTRI.embl](#)
[7_EXONERATE_protZmMSO.embl](#)
[7_EXONERATE_refSeqPlnProt.embl](#)
[7_EXONERATE_SIMprot_wheat.embl](#)
[7_EXONERATE_uniprot_sprot.embl](#)
[8_EUGENE.embl](#)
[9_GENEMODEL.embl](#) 
[10_BESTHIT_BLASTP_nr.embl](#)
[10_BESTHIT_BLASTP_protAtTAIR.embl](#)
[10_BESTHIT_BLASTP_protBdJGI.embl](#)
[10_BESTHIT_BLASTP_prothOR.embl](#)
[10_BESTHIT_BLASTP_protOsIRGSP.embl](#)
[10_BESTHIT_BLASTP_protOsMSU.embl](#)
[10_BESTHIT_BLASTP_protPlant.embl](#)
[10_BESTHIT_BLASTP_protSAC.embl](#)
[10_BESTHIT_BLASTP_protSbDOE.embl](#)
[10_BESTHIT_BLASTP_protTRI.embl](#)
[10_BESTHIT_BLASTP_protZmMSO.embl](#)
[10_EXONERATE_nr.embl](#)
[10_EXONERATE_protAtTAIR.embl](#)
[10_EXONERATE_protBdJGI.embl](#)
[10_EXONERATE_prothOR.embl](#)
[10_EXONERATE_protOsIRGSP.embl](#)
[10_EXONERATE_protOsMSU.embl](#)
[10_EXONERATE_protPlant.embl](#)
[10_EXONERATE_protSAC.embl](#)
[10_EXONERATE_protSbDOE.embl](#)
[10_EXONERATE_protTRI.embl](#)
[10_EXONERATE_protZmMSO.embl](#)
[11_INTERPROSCAN.embl](#)

[12_BLASTN_genoAtTAIR.embl](#)
[12_BLASTN_genoBdJGI.embl](#)
[12_BLASTN_genoOsIRGSP.embl](#)
[12_BLASTN_genoOsMSU.embl](#)
[12_BLASTN_genoSbDOE.embl](#)
[12_BLASTN_genoZmMSO.embl](#)
[12_BLASTN_refSeq_chloro.embl](#)
[12_BLASTN_refSeq_mito.embl](#)
[13_BLASTX_protAtTAIR.embl](#)
[13_BLASTX_protBdJGI.embl](#)
[13_BLASTX_protOsIRGSP.embl](#)
[13_BLASTX_protOsMSU.embl](#)
[13_BLASTX_protSbDOE.embl](#)
[13_BLASTX_protZmMSO.embl](#)
[14_TRNASCAN-SE.embl](#)
[15_TRF.embl](#)

9_GENEMODEL.embl

```
ID unknown; SV 1; linear; unassigned DNA; STD; UNC; 117015 BP.
XX
AC unknown;
XX
XX
FH Key Location/Qualifiers
FH
FT gene complement(13417..20382)
FT /locus_tag="A20_CAT01_1"
FT /color="223 38 56"
FT /label=CAT01
FT /id="A20_NIAS_13417_20382_gene_0001"
FT mRNA complement(13417..20382)
FT /locus_tag="A20_CAT01_2"
FT /color="223 38 56"
FT /label=CAT01
FT /id="A20_NIAS_13417_20382_gene_0001_mRNA_0001"
FT exon complement(13417..13642)
FT /locus_tag="A20_CAT01_3"
FT /color="223 38 56"
FT /label=CAT01
FT /id="A20_NIAS_13417_13642_gene_0001_mRNA_0001_exon_0001"
FT /note="Ontology_term: SO:0000202"
FT exon complement(13660..13894)
FT /locus_tag="A20_CAT01_4"
FT /color="223 38 56"
FT /label=CAT01
FT /id="A20_NIAS_13660_13894_gene_0001_mRNA_0001_exon_0002"
FT /note="Ontology_term: SO:0000004"
FT exon complement(13984..14111)
FT
FT exon complement(20152..20382)
FT /locus_tag="A20_CAT01_21"
FT /color="223 38 56"
FT /label=CAT01
FT /id="A20_NIAS_20152_20382_gene_0001_mRNA_0001_exon_0019"
FT /note="Ontology_term: SO:0000200"
FT CDS complement(join(13869..13894,13984..14111,14318..14409,
FT 14507..14587,14693..14794,14883..14975,15469..15542,
FT 15652..15707,16392..16454,16555..16622,16706..16744,
FT 17016..17106,17298..17382,17450..17501,17598..17704,
FT 18140..18162,18260..18297,20152..20331))
FT /locus_tag="A20_CAT01_22"
FT /color="223 38 56"
FT /label=CAT01
FT /id="A20_NIAS_13417_20382_gene_0001_mRNA_0001_joinedCDS"
FT /note="Similar_to: HMMScan - domain_containing_protein -->
FT PF01966.15 - HD domain (1.6e-13)"
FT /note="Structure_target: Cat01_tplb0006o12"
FT gene 8122..11021
```

Functional
annotation





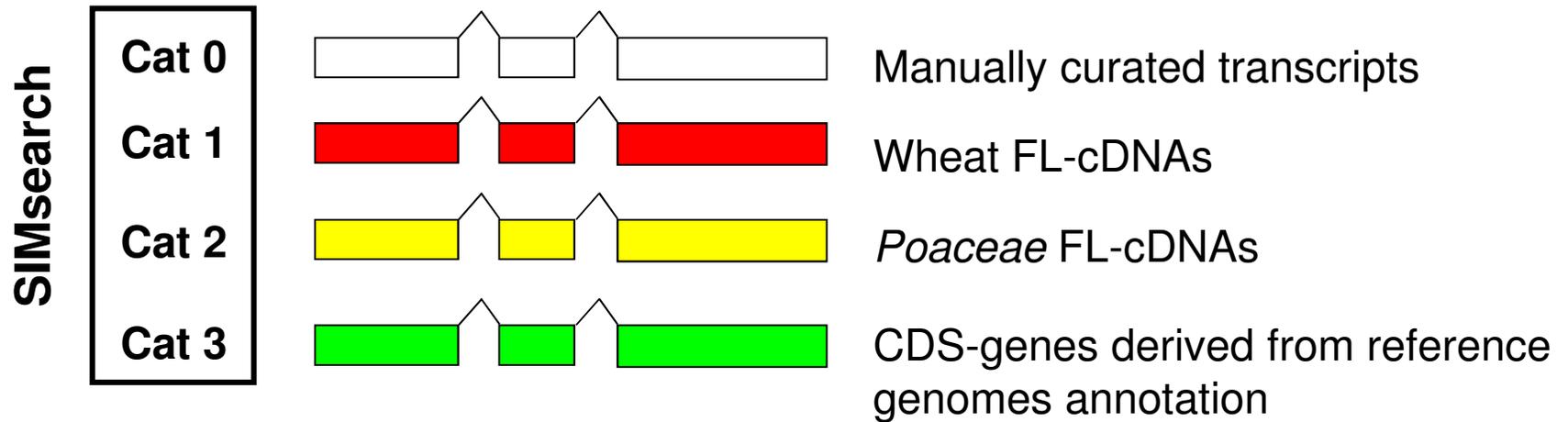
Ctg954 (3 Mb)
Gene 2: 39717..43017

- File / Load data... / local file/9_GENEMODEL.embl
 - Zoom and move
 - Features (list Name)
 - Exon/CDS/Gene Structure (hidden)
 - Clic on exon / CDS → shadow bands
 - CTRL E on exon (window on the right)
 - **Class of the Structural annotation**
 - see Structure_target

Structural Annotation

Use a color code system to assess the confidence of the structural annotation

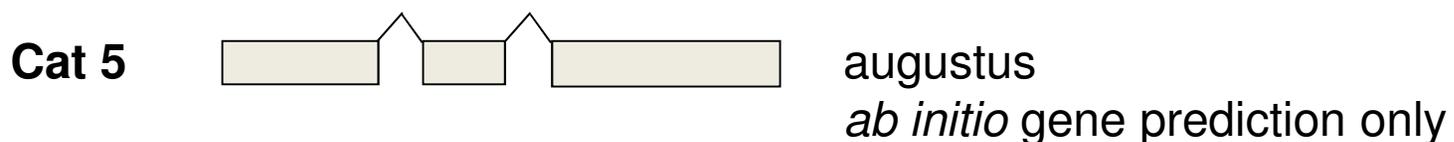
Similarity



ab initio & Similarity



ab initio





Ctg954 (3 Mb)
Gene 2: 39717..43017

- File / Load data... / local file/9_GENEMODEL.embl
 - Zoom and move
 - Features (list Name)
 - Exon/CDS/Gene Structure (hidden)
 - Clic on exon / CDS → shadow bands
 - CTRL E on exon (window)
 - Class of the Structural annotation
 - see Structure_target
 - **Class of the Functional annotation**
 - see Function_target

Functional Annotation

<http://urgi.versailles.inra.fr/Species/Wheat/Triannot-Pipeline/Architecture>

- **Panel II - Step9 - Functional Annotation**

- Putative function for the gene models are assigned via a combination of similarity search (BLASTP) against several protein databanks and against the Pfam (*Sammut et al., 2008 Brief Bioinform. 9, 210-219; Finn et al., 2010 Nucleic Acids Res. 38, D211-222*) protein domain collection with HMMER 3.0 (<http://hmmer.janelia.org/software>). TriAnnot follows a nomenclature based on the guideline established in 2006 by the IWGSC annotation working group (<http://www.wheatgenome.org/tools.php>):

- “**known-function**”: when >80% identity over >80% of the protein length is found with a known protein in UniProtKB/Swiss-Prot. This category reflects the highest quality for functional annotation.
- “**putative-function**”: when >45% similarity over >50% of the protein length is found with a known protein in UniProtKB/Swiss-Prot and UniProtKB/TrEMBL.
- “**domain-containing-protein**”: when there is no significant BLASTP hit with a known or putative function in the previous steps, but one or more Pfam domains (*Sammut et al., 2008; Finn et al., 2010*) are identified.
- “**expressed-sequence**”: based on TBLASTN against plant EST databanks with >45% identity and >50% coverage.
- “**conserved-unknown-function**”: when no expressed sequence is found, and when >45% similarity over >50% of the protein length is found only with an unknown function (i.e. a protein annotated as “putative” or “hypothetical”) in UniProtKB/Swiss-Prot and UniProtKB/TrEMBL.
- “**hypothetical-protein**”: when no similarity is found, either in UniProtKB/Swiss-Prot or UniProtKB/TrEMBL, or Pfam domain or ESTs.



Ctg954 (3 Mb)
Gene 2: 39717..43017

- File / Load data... / local file/9_GENEMODEL.embl
 - Zoom and move
 - Features (list Name)
 - Exon/CDS/Gene Structure (hidden)
 - Clic on exon / CDS → shadow bands
 - CTRL E on exon (window)
 - Class of the Structural annotation
 - see Structure_target
 - Class of the Functional annotation
 - see Function_target
 - **Zoom**
 - **Meth – stop codon**
 - **Intron / Exon junctions (Do/Ac)**
 - **Blue box on top – expand / move**



- **Rename the Label:** fructose-bisphosphate aldolase 3
- **See CDS Name on the right**
- **See Query: Google, NCBI, Ensembl, Plaza**

Ctg954 (3 Mb)
Gene 2: 39717..43017

1_REPEATMASKER_TREP_plus.embl
2_BLASTX_TREP_prot.embl
5_AUGUSTUS_wheat.embl
6_BLASTN_cdsBdJGI.embl
6_BLASTN_cdsOsIRGSP.embl
6_BLASTN_cdsOsMSU.embl
6_BLASTN_cdsSbDOE.embl
6_BLASTN_cdsZmMSO.embl
6_BLASTN_HvFL.embl
6_BLASTN_SIMnuc_wheat.embl
6_BLASTN-TaEST.embl
6_BLASTN-TaFL.embl
6_EXONERATE_cdsBdJGI.embl
6_EXONERATE_cdsOsIRGSP.embl
6_EXONERATE_cdsOsMSU.embl
6_EXONERATE_cdsSbDOE.embl
6_EXONERATE_cdsZmMSO.embl
6_EXONERATE_HvFL.embl
~~6_EXONERATE_SIMnuc_wheat.embl~~
6_EXONERATE-TaEST.embl
6_EXONERATE-TaFL.embl
7_BLASTX_protBdJGI.embl
7_BLASTX_prothOR.embl
7_BLASTX_protOsIRGSP.embl
7_BLASTX_protOsMSU.embl
7_BLASTX_protSbDOE.embl
7_BLASTX_protTRI.embl
7_BLASTX_protZmMSO.embl
7_BLASTX_refSeqPlnProt.embl
7_BLASTX_SIMprot_wheat.embl
7_BLASTX_uniprot_sprot.embl
7_EXONERATE_protBdJGI.embl
7_EXONERATE_prothOR.embl

7_EXONERATE_protOsIRGSP.embl
7_EXONERATE_protOsMSU.embl
7_EXONERATE_protSbDOE.embl
7_EXONERATE_protTRI.embl
7_EXONERATE_protZmMSO.embl
7_EXONERATE_refSeqPlnProt.embl
7_EXONERATE_SIMprot_wheat.embl
7_EXONERATE_uniprot_sprot.embl
8_EUGENE.embl
9_GENEMODEL.embl
10_BESTHIT_BLASTP_nr.embl
10_BESTHIT_BLASTP_protAtTAIR.embl
10_BESTHIT_BLASTP_protBdJGI.embl
10_BESTHIT_BLASTP_prothOR.embl
10_BESTHIT_BLASTP_protOsIRGSP.embl
10_BESTHIT_BLASTP_protOsMSU.embl
10_BESTHIT_BLASTP_protPlant.embl
10_BESTHIT_BLASTP_protSAC.embl
10_BESTHIT_BLASTP_protSbDOE.embl
10_BESTHIT_BLASTP_protTRI.embl
10_BESTHIT_BLASTP_protZmMSO.embl
10_EXONERATE_nr.embl
10_EXONERATE_protAtTAIR.embl
10_EXONERATE_protBdJGI.embl
10_EXONERATE_prothOR.embl
10_EXONERATE_protOsIRGSP.embl
10_EXONERATE_protOsMSU.embl
10_EXONERATE_protPlant.embl
10_EXONERATE_protSAC.embl
10_EXONERATE_protSbDOE.embl
10_EXONERATE_protTRI.embl
10_EXONERATE_protZmMSO.embl
11_INTERPROSCAN.embl

12_BLASTN_genoAtTAIR.embl
12_BLASTN_genoBdJGI.embl
12_BLASTN_genoOsIRGSP.embl
12_BLASTN_genoOsMSU.embl
12_BLASTN_genoSbDOE.embl
12_BLASTN_genoZmMSO.embl
12_BLASTN_refSeq_chloro.embl
12_BLASTN_refSeq_mito.embl
13_BLASTX_protAtTAIR.embl
13_BLASTX_protBdJGI.embl
13_BLASTX_protOsIRGSP.embl
13_BLASTX_protOsMSU.embl
13_BLASTX_protSbDOE.embl
13_BLASTX_protZmMSO.embl
14_TRNASCAN-SE.embl
15_TRF.embl





9_GENEMODEL.embl

- **FL-cDNA**
 - **Wheat / Barley**
 - delete wheat FL features
 - hidden Barley
 - case of BLASTn (*Ta*) then delete
 - TEs
 - extension of 5' exon/mRNA/gene start (copy/paste FL 5' position)

BLASTN
6_EXONERATE

Ctg954 (3 Mb)
Gene 2: 39717..43017

1_REPEATMASKER_TREP_plus.embl
2_BLASTX_TREP_prot.embl
5_AUGUSTUS_wheat.embl
6_BLASTN_cdsBdJGI.embl
6_BLASTN_cdsOsIRGSP.embl
6_BLASTN_cdsOsMSU.embl
6_BLASTN_cdsSbDOE.embl
6_BLASTN_cdsZmMSO.embl
6_BLASTN_HvFL.embl
6_BLASTN_SIMnuc_wheat.embl
6_BLASTN-TaEST.embl
6_BLASTN-TaFL.embl
6_EXONERATE_cdsBdJGI.embl
6_EXONERATE_cdsOsIRGSP.embl
6_EXONERATE_cdsOsMSU.embl
6_EXONERATE_cdsSbDOE.embl
6_EXONERATE_cdsZmMSO.embl
6_EXONERATE_HvFL.embl
~~6_EXONERATE_SIMnuc_wheat.embl~~
6_EXONERATE-TaEST.embl
6_EXONERATE-TaFL.embl
7_BLASTX_protBdJGI.embl
7_BLASTX_prothOR.embl
7_BLASTX_protOsIRGSP.embl
7_BLASTX_protOsMSU.embl
7_BLASTX_protSbDOE.embl
7_BLASTX_protTRI.embl
7_BLASTX_protZmMSO.embl
7_BLASTX_refSeqPlnProt.embl
7_BLASTX_SIMprot_wheat.embl
7_BLASTX_uniprot_sprot.embl
7_EXONERATE_protBdJGI.embl
7_EXONERATE_prothOR.embl

7_EXONERATE_protOsIRGSP.embl
7_EXONERATE_protOsMSU.embl
7_EXONERATE_protSbDOE.embl
7_EXONERATE_protTRI.embl
7_EXONERATE_protZmMSO.embl
7_EXONERATE_refSeqPlnProt.embl
7_EXONERATE_SIMprot_wheat.embl
7_EXONERATE_uniprot_sprot.embl
8_EUGENE.embl
9_GENEMODEL.embl
10_BESTHIT_BLASTP_nr.embl
10_BESTHIT_BLASTP_protAtTAIR.embl
10_BESTHIT_BLASTP_protBdJGI.embl
10_BESTHIT_BLASTP_prothOR.embl
10_BESTHIT_BLASTP_protOsIRGSP.embl
10_BESTHIT_BLASTP_protOsMSU.embl
10_BESTHIT_BLASTP_protPlant.embl
10_BESTHIT_BLASTP_protSAC.embl
10_BESTHIT_BLASTP_protSbDOE.embl
10_BESTHIT_BLASTP_protTRI.embl
10_BESTHIT_BLASTP_protZmMSO.embl
10_EXONERATE_nr.embl
10_EXONERATE_protAtTAIR.embl
10_EXONERATE_protBdJGI.embl
10_EXONERATE_prothOR.embl
10_EXONERATE_protOsIRGSP.embl
10_EXONERATE_protOsMSU.embl
10_EXONERATE_protPlant.embl
10_EXONERATE_protSAC.embl
10_EXONERATE_protSbDOE.embl
10_EXONERATE_protTRI.embl
10_EXONERATE_protZmMSO.embl
11_INTERPROSCAN.embl

12_BLASTN_genoAtTAIR.embl
12_BLASTN_genoBdJGI.embl
12_BLASTN_genoOsIRGSP.embl
12_BLASTN_genoOsMSU.embl
12_BLASTN_genoSbDOE.embl
12_BLASTN_genoZmMSO.embl
12_BLASTN_refSeq_chloro.embl
12_BLASTN_refSeq_mito.embl
13_BLASTX_protAtTAIR.embl
13_BLASTX_protBdJGI.embl
13_BLASTX_protOsIRGSP.embl
13_BLASTX_protOsMSU.embl
13_BLASTX_protSbDOE.embl
13_BLASTX_protZmMSO.embl
14_TRNASCAN-SE.embl
15_TRF.embl





9_GENEMODEL.embl

- Wheat ESTs
 - confirm the 5' end and intron junctions

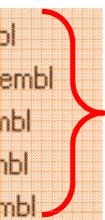


Ctg954 (3 Mb)
Gene 2: 39717..43017

1_REPEATMASKER_TREP_plus.embl
2_BLASTX_TREP_prot.embl
5_AUGUSTUS_wheat.embl
6_BLASTN_cdsBdJGI.embl
6_BLASTN_cdsOsIRGSP.embl
6_BLASTN_cdsOsMSU.embl
6_BLASTN_cdsSbDOE.embl
6_BLASTN_cdsZmMSO.embl
6_BLASTN_HvFL.embl
6_BLASTN_SIMnuc_wheat.embl
6_BLASTN-TaEST.embl
6_BLASTN-TaFL.embl
6_EXONERATE_cdsBdJGI.embl
6_EXONERATE_cdsOsIRGSP.embl
6_EXONERATE_cdsOsMSU.embl
6_EXONERATE_cdsSbDOE.embl
6_EXONERATE_cdsZmMSO.embl
6_EXONERATE_HvFL.embl
~~6_EXONERATE_SIMnuc_wheat.embl~~
6_EXONERATE-TaEST.embl
6_EXONERATE-TaFL.embl
7_BLASTX_protBdJGI.embl
7_BLASTX_prothOR.embl
7_BLASTX_protOsIRGSP.embl
7_BLASTX_protOsMSU.embl
7_BLASTX_protSbDOE.embl
7_BLASTX_protTRI.embl
7_BLASTX_protZmMSO.embl
7_BLASTX_refSeqPlnProt.embl
7_BLASTX_SIMprot_wheat.embl
7_BLASTX_uniprot_sprot.embl
7_EXONERATE_protBdJGI.embl
7_EXONERATE_prothOR.embl

7_EXONERATE_protOsIRGSP.embl
7_EXONERATE_protOsMSU.embl
7_EXONERATE_protSbDOE.embl
7_EXONERATE_protTRI.embl
7_EXONERATE_protZmMSO.embl
7_EXONERATE_refSeqPlnProt.embl
7_EXONERATE_SIMprot_wheat.embl
7_EXONERATE_uniprot_sprot.embl
8_EUGENE.embl
9_GENEMODEL.embl
10_BESTHIT_BLASTP_nr.embl
10_BESTHIT_BLASTP_protAtTAIR.embl
10_BESTHIT_BLASTP_protBdJGI.embl
10_BESTHIT_BLASTP_prothOR.embl
10_BESTHIT_BLASTP_protOsIRGSP.embl
10_BESTHIT_BLASTP_protOsMSU.embl
10_BESTHIT_BLASTP_protPlant.embl
10_BESTHIT_BLASTP_protSAC.embl
10_BESTHIT_BLASTP_protSbDOE.embl
10_BESTHIT_BLASTP_protTRI.embl
10_BESTHIT_BLASTP_protZmMSO.embl
10_EXONERATE_nr.embl
10_EXONERATE_protAtTAIR.embl
10_EXONERATE_protBdJGI.embl
10_EXONERATE_prothOR.embl
10_EXONERATE_protOsIRGSP.embl
10_EXONERATE_protOsMSU.embl
10_EXONERATE_protPlant.embl
10_EXONERATE_protSAC.embl
10_EXONERATE_protSbDOE.embl
10_EXONERATE_protTRI.embl
10_EXONERATE_protZmMSO.embl
11_INTERPROSCAN.embl

12_BLASTN_genoAtTAIR.embl
12_BLASTN_genoBdJGI.embl
12_BLASTN_genoOsIRGSP.embl
12_BLASTN_genoOsMSU.embl
12_BLASTN_genoSbDOE.embl
12_BLASTN_genoZmMSO.embl
12_BLASTN_refSeq_chloro.embl
12_BLASTN_refSeq_mito.embl
13_BLASTX_protAtTAIR.embl
13_BLASTX_protBdJGI.embl
13_BLASTX_protOsIRGSP.embl
13_BLASTX_protOsMSU.embl
13_BLASTX_protSbDOE.embl
13_BLASTX_protZmMSO.embl
14_TRNASCAN-SE.embl
15_TRF.embl





9_GENEMODEL.embl

- **Wheat ESTs**
 - confirm the 5' end and intron junctions
- **CDS other species**
 - keep the best one for biological evidences
 - move track list

Ctg954 (3 Mb)
Gene 2: 39717..43017

BLASTN
6_EXONERATE

1_REPEATMASKER_TREP_plus.embl
2_BLASTX_TREP_prot.embl
5_AUGUSTUS_wheat.embl
6_BLASTN_cdsBdJGI.embl
6_BLASTN_cdsOsIRGSP.embl
6_BLASTN_cdsOsMSU.embl
6_BLASTN_cdsSbDOE.embl
6_BLASTN_cdsZmMSO.embl
6_BLASTN_HvFL.embl
6_BLASTN_SIMnuc_wheat.embl
6_BLASTN-TaEST.embl
6_BLASTN-TaFL.embl
6_EXONERATE_cdsBdJGI.embl
6_EXONERATE_cdsOsIRGSP.embl
6_EXONERATE_cdsOsMSU.embl
6_EXONERATE_cdsSbDOE.embl
6_EXONERATE_cdsZmMSO.embl
6_EXONERATE_HvFL.embl
~~6_EXONERATE_SIMnuc_wheat.embl~~
6_EXONERATE-TaEST.embl
6_EXONERATE-TaFL.embl
7_BLASTX_protBdJGI.embl
7_BLASTX_prothOR.embl
7_BLASTX_protOsIRGSP.embl
7_BLASTX_protOsMSU.embl
7_BLASTX_protSbDOE.embl
7_BLASTX_protTRI.embl
7_BLASTX_protZmMSO.embl
7_BLASTX_refSeqPlnProt.embl
7_BLASTX_SIMprot_wheat.embl
7_BLASTX_uniprot_sprot.embl
7_EXONERATE_protBdJGI.embl
7_EXONERATE_prothOR.embl

7_EXONERATE_protOsIRGSP.embl
7_EXONERATE_protOsMSU.embl
7_EXONERATE_protSbDOE.embl
7_EXONERATE_protTRI.embl
7_EXONERATE_protZmMSO.embl
7_EXONERATE_refSeqPlnProt.embl
~~7_EXONERATE_SIMprot_wheat.embl~~
7_EXONERATE_uniprot_sprot.embl
8_EUGENE.embl
9_GENEMODEL.embl
10_BESTHIT_BLASTP_nr.embl
10_BESTHIT_BLASTP_protAtTAIR.embl
10_BESTHIT_BLASTP_protBdJGI.embl
10_BESTHIT_BLASTP_prothOR.embl
10_BESTHIT_BLASTP_protOsIRGSP.embl
10_BESTHIT_BLASTP_protOsMSU.embl
10_BESTHIT_BLASTP_protPlant.embl
10_BESTHIT_BLASTP_protSAC.embl
10_BESTHIT_BLASTP_protSbDOE.embl
10_BESTHIT_BLASTP_protTRI.embl
10_BESTHIT_BLASTP_protZmMSO.embl
10_EXONERATE_nr.embl
10_EXONERATE_protAtTAIR.embl
10_EXONERATE_protBdJGI.embl
10_EXONERATE_prothOR.embl
10_EXONERATE_protOsIRGSP.embl
10_EXONERATE_protOsMSU.embl
10_EXONERATE_protPlant.embl
10_EXONERATE_protSAC.embl
10_EXONERATE_protSbDOE.embl
10_EXONERATE_protTRI.embl
10_EXONERATE_protZmMSO.embl
11_INTERPROSCAN.embl

12_BLASTN_genoAtTAIR.embl
12_BLASTN_genoBdJGI.embl
12_BLASTN_genoOsIRGSP.embl
12_BLASTN_genoOsMSU.embl
12_BLASTN_genoSbDOE.embl
12_BLASTN_genoZmMSO.embl
12_BLASTN_refSeq_chloro.embl
12_BLASTN_refSeq_mito.embl
13_BLASTX_protAtTAIR.embl
13_BLASTX_protBdJGI.embl
13_BLASTX_protOsIRGSP.embl
13_BLASTX_protOsMSU.embl
13_BLASTX_protSbDOE.embl
13_BLASTX_protZmMSO.embl
14_TRNASCAN-SE.embl
15_TRF.embl





9_GENEMODEL.embl

**BLASTN
6_EXONERATE**

**Ctg954 (3 Mb)
Gene 2: 39717..43017**

**BLASTX
7_EXONERATE**

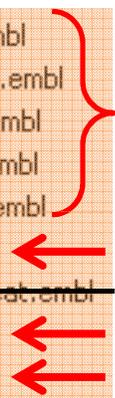
● **exonerate-BLASTX other species**

- hidden CDS
- hidden RefSeq
- keep the best biological evidences
- SwissProt
 - CTRL E => functional annotation - Q9ZU52

1_REPEATMASKER_TREP_plus.embl
2_BLASTX_TREP_prot.embl
5_AUGUSTUS_wheat.embl
6_BLASTN_cdsBdJGI.embl
6_BLASTN_cdsOsIRGSP.embl
6_BLASTN_cdsOsMSU.embl
6_BLASTN_cdsSbDOE.embl
6_BLASTN_cdsZmMSO.embl
6_BLASTN_HvFL.embl
6_BLASTN_SIMnuc_wheat.embl
6_BLASTN-TaEST.embl
6_BLASTN-TaFL.embl
6_EXONERATE_cdsBdJGI.embl
6_EXONERATE_cdsOsIRGSP.embl
6_EXONERATE_cdsOsMSU.embl
6_EXONERATE_cdsSbDOE.embl
6_EXONERATE_cdsZmMSO.embl
6_EXONERATE_HvFL.embl
~~6_EXONERATE_SIMnuc_wheat.embl~~
6_EXONERATE-TaEST.embl
6_EXONERATE-TaFL.embl
7_BLASTX_protBdJGI.embl
7_BLASTX_prothOR.embl
7_BLASTX_protOsIRGSP.embl
7_BLASTX_protOsMSU.embl
7_BLASTX_protSbDOE.embl
7_BLASTX_protTRI.embl
7_BLASTX_protZmMSO.embl
7_BLASTX_refSeqPlnProt.embl
7_BLASTX_SIMprot_wheat.embl
7_BLASTX_uniprot_sprot.embl
7_EXONERATE_protBdJGI.embl
7_EXONERATE_prothOR.embl

7_EXONERATE_protOsIRGSP.embl
7_EXONERATE_protOsMSU.embl
7_EXONERATE_protSbDOE.embl
7_EXONERATE_protTRI.embl
7_EXONERATE_protZmMSO.embl
7_EXONERATE_refSeqPlnProt.embl
~~7_EXONERATE_SIMprot_wheat.embl~~
7_EXONERATE_uniprot_sprot.embl
8_EUGENE.embl
9_GENEMODEL.embl
10_BESTHIT_BLASTP_nr.embl
10_BESTHIT_BLASTP_protAtTAIR.embl
10_BESTHIT_BLASTP_protBdJGI.embl
10_BESTHIT_BLASTP_prothOR.embl
10_BESTHIT_BLASTP_protOsIRGSP.embl
10_BESTHIT_BLASTP_protOsMSU.embl
10_BESTHIT_BLASTP_protPlant.embl
10_BESTHIT_BLASTP_protSAC.embl
10_BESTHIT_BLASTP_protSbDOE.embl
10_BESTHIT_BLASTP_protTRI.embl
10_BESTHIT_BLASTP_protZmMSO.embl
10_EXONERATE_nr.embl
10_EXONERATE_protAtTAIR.embl
10_EXONERATE_protBdJGI.embl
10_EXONERATE_prothOR.embl
10_EXONERATE_protOsIRGSP.embl
10_EXONERATE_protOsMSU.embl
10_EXONERATE_protPlant.embl
10_EXONERATE_protSAC.embl
10_EXONERATE_protSbDOE.embl
10_EXONERATE_protTRI.embl
10_EXONERATE_protZmMSO.embl
11_INTERPROSCAN.embl

12_BLASTN_genoAtTAIR.embl
12_BLASTN_genoBdJGI.embl
12_BLASTN_genoOsIRGSP.embl
12_BLASTN_genoOsMSU.embl
12_BLASTN_genoSbDOE.embl
12_BLASTN_genoZmMSO.embl
12_BLASTN_refSeq_chloro.embl
12_BLASTN_refSeq_mito.embl
13_BLASTX_protAtTAIR.embl
13_BLASTX_protBdJGI.embl
13_BLASTX_protOsIRGSP.embl
13_BLASTX_protOsMSU.embl
13_BLASTX_protSbDOE.embl
13_BLASTX_protZmMSO.embl
14_TRNASCAN-SE.embl
15_TRF.embl





9_GENEMODEL.embl

**BLASTN
6_EXONERATE**

Ctg954 (3 Mb)
Gene 2: 39717..43017

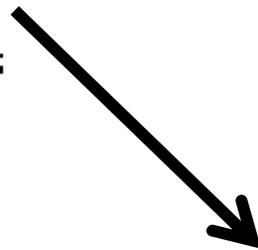
**BLASTX
7_EXONERATE**

- Hidden BLASTX
- Best hits

**BLASTP
10_EXONERATE**

Protein alignments

- [-] TriAnnotResults_00016
 - EMBL
 - GFF
 - Other files
 - sequences



Name ▲	Size	Type
 10_BESTHIT_BLASTP_nr.align	3 KB	ALIGN File
 10_BESTHIT_BLASTP_protBdJGI.align	3 KB	ALIGN File
 10_BESTHIT_BLASTP_prothOR.align	2 KB	ALIGN File
 10_BESTHIT_BLASTP_protOsIRGSP.align	2 KB	ALIGN File
 10_BESTHIT_BLASTP_protPlant.align	2 KB	ALIGN File
 10_BESTHIT_BLASTP_protZmMSO.align	1 KB	ALIGN File
 Global_XM_for_TE_N_masked.xm	8 KB	Winamp media f

>ctg954_3Mb_NIAS_39717_43017_gene_002_mRNA_0001

Bradi2g01350.1

Coverage: 100.00%

Identity: 93.57%

```
10_BESTHIT_BLA...IP_protBdJGI.align - WordPad
File Edit View Insert Format Help

>ctg954_3Mb_NIAS_2642016_2649390_gene_0001_mRNA_0001 | lcl|Bradi2g01550.1 | Coverage: 81.59% | Identity: 89.06% | Positive: 95.77%
Query: MSTFAKPENALKRAEBELIHVGQKQAALQALHDLITSKRYRSWQKPLEKIMMKYVELCVDLRKGRFAKDGLIQRIVCCQVNVVSSLEEVIKHFMQLSNEKALEARNQAQALDVEDLEADKR
Align: MSTFAKPENALKRAEBELIHVGQKQAALQALHDLITSKRYRSWQKPLEKIMMKYVELCVDLRKGRFAKDGLIQRIVCCQVNVVSSLEEVIKHFMQLSNEKALEEA++Q++ALE+ALDVEDLEADKR
Sbjct: MSTFAKPENALKRAEBELIHVGQKQAALQALHDLITSKRYRSWQKPLEKIMMKYVELCVDLRKGRFAKDGLIQRIVCCQVNVVSSLEEVIKHFMQLSNEKALEEA++Q++ALE+ALDVEDLEADKR

>ctg954_3Mb_NIAS_39717_43017_gene_0002_mRNA_0001 | lcl|Bradi2g01350.1 | Coverage: 100.00% | Identity: 93.57% | Positive: 95.63%
Query: MAMVTAKLSSPAAARLAPSAGSARRASRVR----VRASGGSYADELVSTAKTVASPGRGILAIDESSATCGKRLASIGLDNTEVNRQAYRQLLLTTAGLGEYISGAILFEETLYQSTTDGKTFV
Align: MAMVTAKLSSPAA+L+G R++R VR+GSYADELVSTAK+VASPGRGILAIDES+ATCGKRLASIGLDNTEVNRQAYRQLLLTTAGLGEYISGAILFEETLYQSTTDGKTFV
Sbjct: MAMVTAKLSSPAAASWL--TGGRTHRSAPTRRATLVRAAAGSYADELVSTAKSVASPGRGILAIDESNATCGKRLASIGLDNTEVNRQAYRQLLLTTAGLGEYISGAILFEETLYQSTTDGKTFV

>ctg954_3Mb_NIAS_2460244_2463150_gene_0004_mRNA_0001 | lcl|Bradi2g01540.1 | Coverage: 95.26% | Identity: 77.48% | Positive: 87.40%
Query: MDARKFLQMVEDKKRLLLEKKEAPLKWQKLEAAIKATEEKEKKLKSRRHRRSYSSSESSESDSDRKHRRKRDRKRHRKHGHSDSGARRRKRHRKRSSGSSDESDESDEYDGESEEEERR
Align: MDARKFLQMV+K+K+RLL+KKEAPLKWQKLEAAIKATEEKEKKLKS+KHRRR SSSESDS+SESD DRKHRRKRDR+RH++HGHS+SD ARR K RSKRRSS S+DESDESDEY SE ERR
Sbjct: MDARKFLQMV+K+K+RLL+KKEAPLKWQKLEAAIKATEEKEKKLKSRRHRRRGDSSSESDESDES+SDRKHRRKRDRRHRKHGHSSESDDARRRKRHRKRSSDSTDESDESDEYDGESEEEERR

>ctg954_3Mb_NIAS_1538098_1540961_gene_0005_mRNA_0001 | lcl|Bradi2g0168880.2 | Coverage: 100.00% | Identity: 92.92% | Positive: 97.35%
Query: MATSKKVVTRREEWEGKLRDVKIRKEDMNRLVMNFLVTEGPFVDAADRFRVSESGTQPEIDLATITDRMEVKKAVQSGNVQEAIERINDLNPTILDTPNPIYFHLQKQKLIELIRVKGIEALEFAQ
Align: MA+SKKVVTR+EWE KLRDVKIRKEDMNRLVMNFLVTEGPFVDAAD+FR+ESGTQP+IDLATITDRMEVKKAVQSGNVQEAIE+INDLNPTILDTPN++YFHLQKQKLIELIRVKGIE+ALEFAQ
Sbjct: MASSKKVVTRDEWERKLRDVKIRKEDMNRLVMNFLVTEGPFVDAADKFRVSESGTQPEIDLATITDRMEVKKAVQSGNVQEAIEKINDLNPTILDTPNPLYFHLQKQKLIELIRVKGIEALEFAQ

>ctg954_3Mb_NIAS_1207699_1212541_gene_0007_mRNA_0001 | lcl|Bradi2g013280.1 | Coverage: 74.49% | Identity: 72.48% | Positive: 79.91%
Query: MDAEQESRWAAAQIGIG--EDLVPAALRHLEFLAAVDRRRWLYDGPLLHRAIRRYKACWLPPLAKHTEATVADDEPLVVPLDCEWIWHCHRLNPTRYIKDCKRLYGRILDCKNVRSSIQAKSK
Align: MDAEQESRWAAAQIGIG+G E+LVPAALR LEFLAAVDRRRWLY+GPLLHRAIRRYKACWLPPLAKHTEA V D PL+VPLDCEWIWHCHRLNP +YIKDCKRLYGRILD K V SSIQ KSK
Sbjct: MDAEQESRWAAAQIGIGVGVGVEELVPAALRQLEFLAAVDRRRWLYEGPLLHRAIRRYKACWLPPLAKHTEAAVVDG-PLIVPLDCEWIWHCHRLNPAQYIKDCKRLYGRILDNKYVESSIQVQSK

>ctg954_3Mb_NIAS_2839551_2843427_gene_0014_mRNA_0001 | lcl|Bradi2g01570.1 | Coverage: 99.70% | Identity: 87.21% | Positive: 90.41%
Query: MAGIRWPPEDPEMPFTRMLGTGVWGGPPAAAAGGGGGPPGEMASDDDRSVAADSWSIKSDYGSTLDDEQRYADTAEVLLASCPSSASSSSAASAAAPSACSSSLSAHSSDF3FDKDVDPVVPFM
Align: MAGIRWPPEDPEMPFTRMLG+ AG G G P GEMASDDDRSVAADSWSIKSDYGSTLDDEQRYADTAEVLLAS S++S++ A SA S S SDF3FDKDVDPVVPFM
Sbjct: MAGIRWPPEDPEMPFTRMLGS-----GVWAGAGAGAPGEMASDDDRSVAADSWSIKSDYGSTLDDEQRYADTAEVLLASSCSASSASSSSVVAPSSASSLSAH--SDF3FDKDVDPVVPFM

>ctg954_3Mb_NIAS_259875_261449_gene_0016_mRNA_0001 | lcl|Bradi2g025170.1 | Coverage: 98.30% | Identity: 72.96% | Positive: 82.40%
Query: MAGGGDEEVTLLGHMGSAAYVTRVRLALHLKGVRYEYVEEDLRNKSDDLRCNPVHRAVPVLIHNGRPVCSQIILQYIDQAFGPPPTLLPADPHERAVARFWAAYADDVLGAPWEKAYRARTEE
Align: MAGG DE V LLGHMGSA+VTRV+LA HLKGV YEY EEDL NKS+LLLR NPVH+AVPVLIHNGRP+CESQ+ILQYID+ F LLPADPHERAVARFWAAY D+ +GAPWEKAYRART E
Sbjct: MAGGVADE-VKLLGHMGSPFVTRVRLALHLKGVSYEYAEEDLHNSKSELLLRNPVHRAVPVLIHNGRPICESQVILQYIDETFDGNLLLPADPHERAVARFWAAYIDNEMGAPWEKAYRARTAE
```



Ctg954 (3 Mb)
Gene 2: 39717..43017

- How to get the protein
- BLAST @ NCBI
 - select all
- Alignments @ NCBI
 - select 11 bests
- Re-alignment @ NCBI
- Tree @ NCBI
 - sequence label
 - show sub tree on monocots | 4 leaves

BLAST @ NCBI

 **BLAST**[®] *Basic Local Alignment Search Tool* My NCBI
[\[Sign In\]](#) [\[Re...](#)

[Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

▸ [NCBI/ BLAST/ blastp suite](#) **Standard Protein BLAST**

[blastn](#) **[blastp](#)** [blastx](#) [tblastn](#) [tblastx](#)

BLASTP programs search protein databases using a protein query. [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)  [Clear](#) **Query subrange** 

MAMVTAKLSSPAAARLAPSAGSARRASRVVRASGGSYADELVSTAKTVAS PGRGILAIDESSATC
GKRLASIGLDNTEVNRQAYRQLLLTTAGLGEYISGAILFEETLYQSTTDGKTFVDVLKDQNI MPGI
KVDKGLVPLPGSNNESWCQGLDGLASRCAEYYKQGARFAKWR TVVSI PCGPTALAVKEAAMGLARY
AAIAQDNGLVPIVEPEILLDGDHGIERTLEVAEKVWSEVFFYLAENNVLEFEGILLKPSMVT PGAEH
KEKASPEAIKNTLTMLRRRVPPAVPGIMFLSGGQSELEATMNLNAMNQSANPWHVSE SYARALQN

Or, upload file [Parcourir...](#) 

Job Title
Enter a descriptive title for your BLAST search 

Align two or more sequences 

Choose Search Set

Database 

Organism

50 best-hits Evalue = 1e-5

BLAST result @ NCBI

Query ID |cd|16982
Description GVquery
Molecule type amino acid
Query Length 385

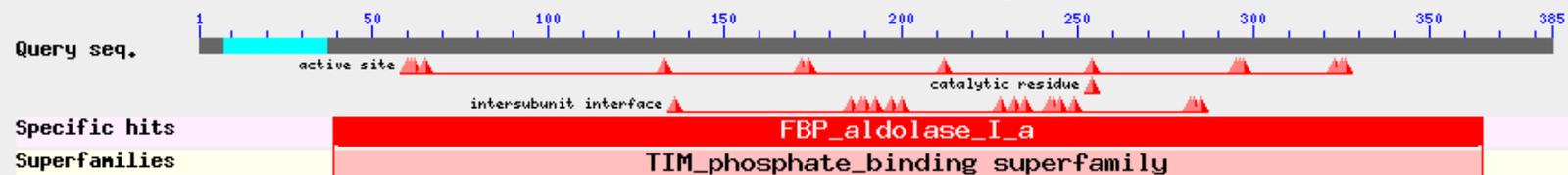
Database Name nr
Description All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Program BLASTP 2.2.26+ [Citation](#)

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [Multiple alignment](#)

Graphic Summary

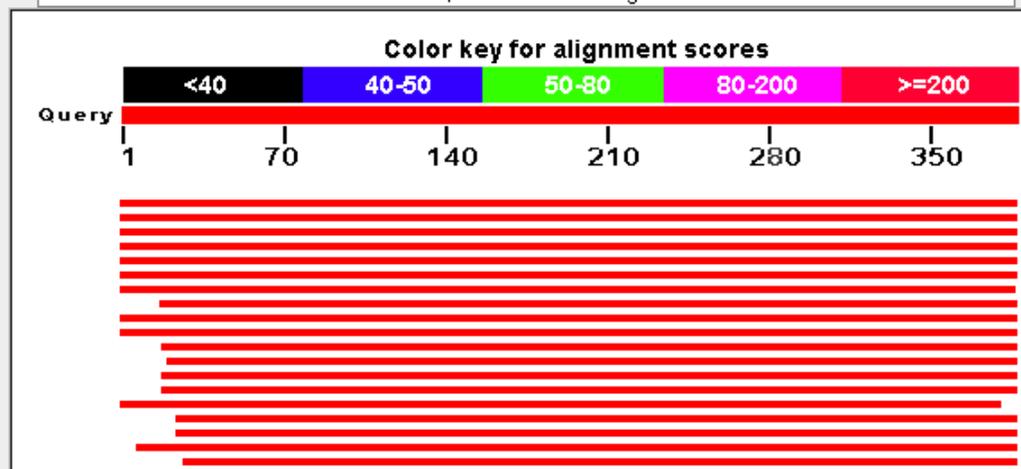
Show Conserved Domains

Putative conserved domains have been detected, click on the image below for detailed results.



Distribution of 50 Blast Hits on the Query Sequence

Mouse-over to show define and scores, click to show alignments



Multiple alignments @ NCBI

▼ **Alignments** Select All Mouse over the sequence identifier for sequence title

<input checked="" type="checkbox"/>	16982	1	MA-----MV-T---AKLSSP--AA-ARLA--P--SAGSA-RRAS-----RVR---VRASGGSYADELVSTAKTVASP	52
<input checked="" type="checkbox"/>	CBH32613	1	MA-----MV-T---AKLSSP--AA-ARLA--P--SAGSA-RRAS-----RVR---VRASGGSYADELVSTAKTVASP	52
<input checked="" type="checkbox"/>	BAJ98432	1	MA-----MV-T---AKLSSP--AT-ARLG--PVRSAQPA-RRAS-----LVR---VRASGGSYADELVSTAKTVASP	54
<input checked="" type="checkbox"/>	XP_003567597	1	MA-----MV-T---AKLSSP--AA-SWLT--G----GRT-HRSA-----PTRRATLVRAAAGSYADELVSTAKSVASP	54
<input checked="" type="checkbox"/>	XP_002455301	1	MA-----MA-T---AKLSSP--AT-SLLA--G----GRT-RRSA-----PARRATVIRAAAGSYDELVSTAKSVASP	54
<input checked="" type="checkbox"/>	NP_001041857	1	MA-----ML-T---AKLTSPPAAT-TWLP--G----G-G-RRSA-----PPRRATVIRAAAVSYADELVSTAKSVASP	55
<input checked="" type="checkbox"/>	NP_001152410	1	MA-----MA-T---AKLNMP--AT-SLVA--G----GLT-RRSA-----PARCTTVIRAAAGSYDELISTAKSVASP	54
<input checked="" type="checkbox"/>	ACG47669	1	MA-----MA-T---AKLNMP--AT-SLVA--G----GRT-RRSA-----PARCTTVIRAAAGSYDELISTAKSVASP	54
<input type="checkbox"/>	XP_002526308	1	MA-----CASF---AKLNAA--SS-TWIG--QQS-FGQR-PGSSSARFATRRVSLPIRA--SSYKDELVQTAKTVASP	61
<input type="checkbox"/>	NP_001242172	1	MA-----MA-S---AKLNLT--SS-QWIA--HNS-FSPR-RRSS-----SRRVSLPIRA--SSYQHELVTAKSIASP	55
<input type="checkbox"/>	XP_002282753	1	MA-----SVTF---AKFNAS--SS-QWIG--QQS-FSQR-QGSS-ARFPARRVSVPIRA--GSYSEELVQTAKTVASP	60
<input type="checkbox"/>	XP_002311480	1	MA-----CASF---VKLNAA--SS-SWTG--QKSSFGKRSPGSS-----TRRVFSIRA--SSYDELVQTAKLIASP	58
<input type="checkbox"/>	AEH04657	1	MA-----CSSL---VKLNAS--SS-QWIG--QQS-FNQR-PGSS-CRFASPRVSV-VRA--GSYDELVKTAKSIASP	59
<input type="checkbox"/>	XP_002315908	1	MA-----CANL---VKLNAA--SS-SWIG--QKSPFGQRSQGSS-----TRRVFSIRA--NSYDELVQTAKTIASP	58
<input type="checkbox"/>	ABK94899	1	MA-----CANL---VKLNAA--SS-SWIG--QKSPFGQRSQGSS-----TRRVFSIRA--NSYDELVQTAKTIASP	58
<input type="checkbox"/>	EEE53770	1	-----MAMLTA--KLT-SPPAATTWLPGGGR--RSAPPRRATVIRAAAVSYADELVSTAKSVASP	55
<input type="checkbox"/>	ABK24568	1	MA-----AS-V---IKAGVRLGSS-QWTG--QSLTQNID-SHKQ-----PKAQRVSMPIRAGSYAEELVQTAKTVASP	60
<input type="checkbox"/>	ABK24286	1	MA-----AS-V---LKAGVRLGSS-QWTG--QSLTQNID-SHKQ-----PKAQRVSMPIRAGSYAEELVQTAKTVASP	60
<input type="checkbox"/>	NP_178224	1	MA-----SASF---VKPNTL--SS-PWIG--QRS-FAHT-SASS---SPPPRVFAIRA--GAYSDELVKTAKSIASP	58
<input type="checkbox"/>	XP_002876733	1	MA-----SASF---VKPNTL--SS-PWIG--QRS-FAHT-SAS-----PPRVFAIRA--GAYSDELVKTAKSVASP	55
<input type="checkbox"/>	ABK25226	1	MA-----AS-V---LKAGVRLGSS-QWTG--QSLTQNID-SHKQ-----PKAQRVSMPIRAGSYAEELVQTAKTVASP	60
<input type="checkbox"/>	XP_003536659	1	MA-----MAAS---AKLNLT--SSSQWIA--HNS-FSPR-RGSS-----SRRVSLPIRA--SSYQHELVTAKSIASP	57
<input type="checkbox"/>	XP_003540103	1	MA-----MA-S---AKLNLT--SS-QWIA--HNS-FSPR-RGSS-----SRRVSLPIRA--SSYQHELVTAKSIASP	55
<input type="checkbox"/>	ADD52195	1	MASA--SLLKTSP-VLDKSEFV--KG-QALLLRQPSV--AVVRCQP-----VAPSGVTVRA--S-YADELVKTAKTIASP	64
<input type="checkbox"/>	AAM46780	1	MASA--SLLKSSP-VLDRSEFV--KG-QSL--RQPSA--SIIRCHP-----ASSSALTILA--GSYADELVKTAKTVASP	63
<input type="checkbox"/>	AER26531	1	MASA--SFLKSSP-VLDKSEWI--KG-QTL--RQPSV--SVVRCQP-----VAPSSLLVRA--SSYADELVKTAKTIASP	63
<input type="checkbox"/>	XP_002305462	1	MASA--SFLKSSP-VLDKSEFV--KG-QTL--RLPSA--SIVRCRS-----TAPSALTVRA--GSYAEELVKTAKTIASP	63
<input type="checkbox"/>	ARK96397	1	MASA--SFLKSSP-VLDKSEFV--KG-QTL--RLPSA--SIVRCRS-----TAPSALTVRA--GSYAEELVKTAKTIASP	63

Re-align @ NCBI

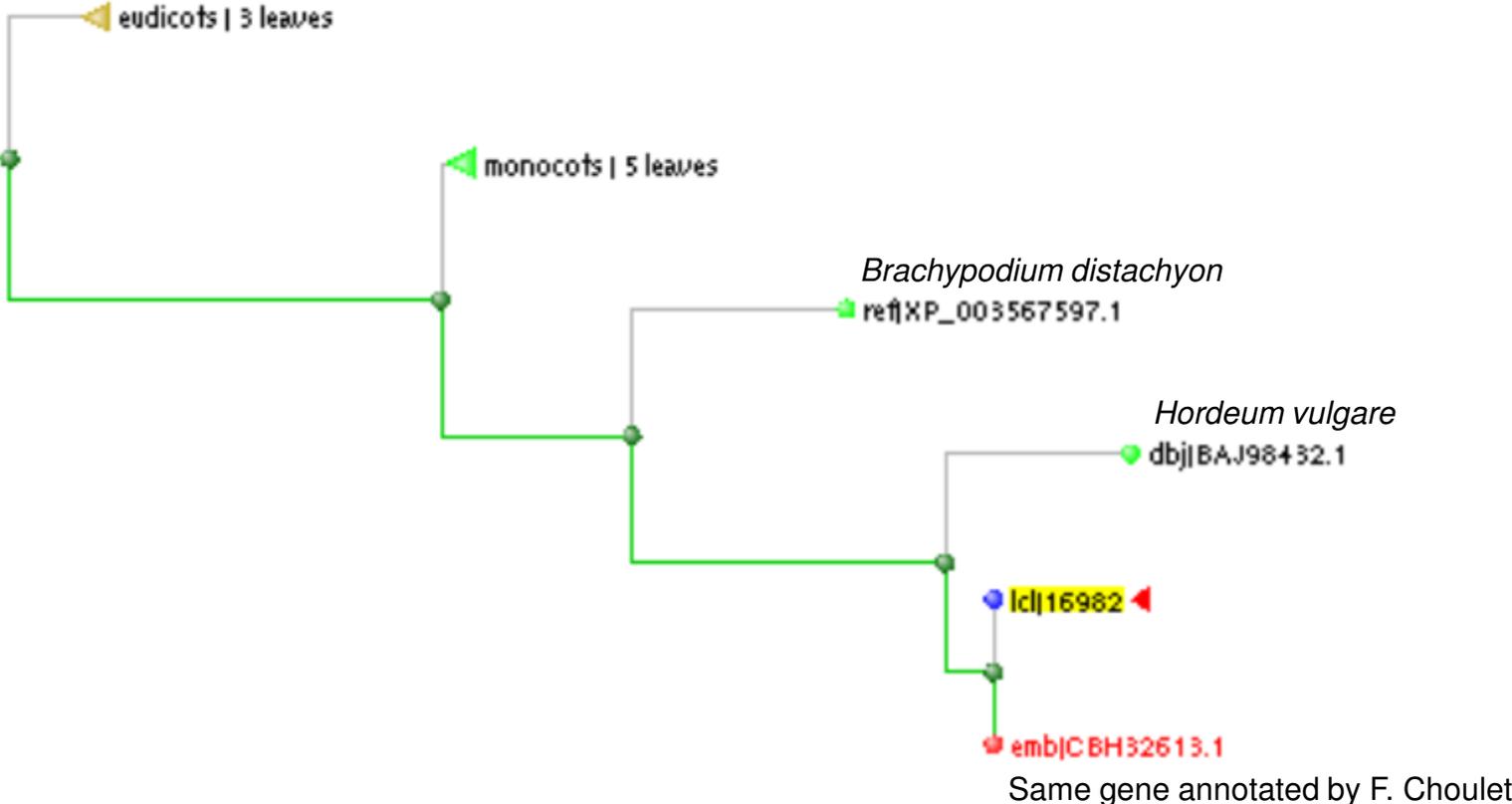
Alignments Select All Re-align Mouse over the sequence identifier for sequence title

<input checked="" type="checkbox"/>	16982	1	MAMVTAKLSSP--AAARLAP--SAGSARRASRVR----VRASGGSYADELVSTAKTVASPGRGILAIDESSATCGKRLAS	72
<input checked="" type="checkbox"/>	CBH32613	1	MAMVTAKLSSP--AAARLAP--SAGSARRASRVR----VRASGGSYADELVSTAKTVASPGRGILAIDESSATCGKRLAS	72
<input checked="" type="checkbox"/>	BAJ98432	1	MAMVTAKLSSP--ATARLGPVRSAPARRASLVR----VRASGGSYADELVSTAKTVASPGRGILAIDESSATCGKRLAS	74
<input checked="" type="checkbox"/>	XP_003567597	1	MAMVTAKLSSP--AASWL----TGGRTHRSAPTRRATLVRAAAGSYADELVSTAKSVASPGRGILAIDESNATCGKRLAS	74
<input checked="" type="checkbox"/>	XP_002455301	1	MAMATAKLSSP--ATSLV----AGGRTRRSAPARRATVIRAAAGSYADELVSTAKSVASPGRGILAIDESNATCGKRLSS	74
<input checked="" type="checkbox"/>	NP_001041857	1	MAMLTAKLTSPPAATTWL----PGG-GRRSAPRRATVIRAAAVSYADELVSTAKSVASPGRGILAIDESNATCGKRLAS	75
<input checked="" type="checkbox"/>	NP_001152410	1	MAMATAKLNSP--ATSLV----AGGLTRRSAPARCTTVIRAAAGSYDELISTAKSVASPGRGILAIDESNATCGKRLSS	74
<input checked="" type="checkbox"/>	ACG47669	1	MAMATAKLNSP--ATSLV----AGGRTRRSAPARCTTVIRAAAGSYDELISTAKSVASPGRGILAIDESNATCGKRLSS	74
<input checked="" type="checkbox"/>	16982	73	IGLDNTEVNRQAYRQLLLTTAGLGEYISGAILFEETLYQSTTDGKTFVDVLKDNIMPQIKVDKGLVPLPGSNNESWCQG	152
<input checked="" type="checkbox"/>	CBH32613	73	IGLDNTEVNRQAYRQLLLTTAGLGEYISGAILFEETLYQSTTDGKTFVDVLKDNIMPQIKVDKGLVPLPGSNNESWCQG	152
<input checked="" type="checkbox"/>	BAJ98432	75	IGLDNTEVNRQAYRQLLLTTAGLGEYISGAIRFEETLYQSTTDGKTFVDVLKDNIMPQIKVDKGLVPLPGSNNESWCQG	154
<input checked="" type="checkbox"/>	XP_003567597	75	IGLDNTEVNRQAYRQLLLTTAGLGEYISGAILFEETLYQSTTDGKTFVDCLKDNIMPQIKVDKGLVPLPGSNNESWCQG	154
<input checked="" type="checkbox"/>	XP_002455301	75	IGLDNTEPNRQAYRQLLLTTAGLGEYISGAILFEETLYQSTTDGKFFVDCLKEQNIMPQIKVDKGLVPLPGSNNESWCQG	154
<input checked="" type="checkbox"/>	NP_001041857	76	IGLDNTEVNRQAYRQLLLTTAGLGEYISGAILFEETLYQSTTDGKFFVDCLKDNIMPQIKVDKGLVPLPGSNNESWCQG	155
<input checked="" type="checkbox"/>	NP_001152410	75	IGLDNTEVNRQAYRQLLLTTAGLGEYISGAILFEETLYQSTTDGKFFVDCLKDNIMPQIKVDKGLVPLPGSNNESWCQG	154
<input checked="" type="checkbox"/>	ACG47669	75	IGLDNTEVNRQAYRQLLLTTAGLGEYISGAILFEETLYQSTTDGKFFVDCLKDNIMPQIKVDKGLVPLPGSNNESWCQG	154
<input checked="" type="checkbox"/>	16982	153	LDGLASRCAEYKQGARFAKWRVTVVSI PCGPTALAVKEAAWGLARYAAIAQDNGLVPIVEPEILLDGDHGIERTLEVAEK	232
<input checked="" type="checkbox"/>	CBH32613	153	LDGLASRCAEYKQGARFAKWRVTVVSI PCGPTALAVKEAAWGLARYAAIAQDNGLVPIVEPEILLDGDHGIERTLEVAEK	232
<input checked="" type="checkbox"/>	BAJ98432	155	LDGLASRCAEYKQGARFAKWRVTVVSI PCGPTALAVKEAAWGLARYAAIAQDNGLVPIVEPEILLDGDHGIERTLEVAEK	234
<input checked="" type="checkbox"/>	XP_003567597	155	LDGLASRCAEYKQGARFAKWRVTVVSI PCGPTALAVKEAAWGLARYAAIAQDNGLVPIVEPEILLDGDHGIERTLEVAEK	234
<input checked="" type="checkbox"/>	XP_002455301	155	LDGLASRCAEYKQGARFAKWRVTVVSI PCGPTALAVKEAAWGLARYAAIAQDNGLVPIVEPEILLDGDHGIERTLEVAEK	234
<input checked="" type="checkbox"/>	NP_001041857	156	LDGLASRCAEYKQGARFAKWRVTVVSI PCGPTALAVKEAAWGLARYAAIAQDNGLVPIVEPEILLDGDHAIERTLEVAEK	235
<input checked="" type="checkbox"/>	NP_001152410	155	LDGLASRCAEYKQGARFAKWRVTVVSI PCGPTALAVKEAAWGLARYAAIAQDNGLVPIVEPEILLDGDHIEGAELEVAEK	234
<input checked="" type="checkbox"/>	ACG47669	155	LDGLASRCAEYKQGARFAKWRVTVVSI PCGPTALAVKEAAWGLARYAAIAQDNGLVPIVEPEILLDGDHIEGAELEVAEK	234

Brachypodium

>ctg954_3Mb_NIAS_39717_43017_gene_0002_mRNA_0001 | lcl|Bradi2g01350.1 | Coverage: 100.00% | Identity: 93.57% | Positive: 95.63%
Query: MAMVTAKLSSPAAARLAPSAGSARRASRVR----VRASGGSYADELVSTAKTVASPGRGILAIDESSATCGKRLASIGLDNTEVNRQAYRQLLLTTAGLGEYISGAILFEETLYQSTTDG
Align: MAMVTAKLSSPAA+ L + G R++ R VRA+ GSYADELVSTAK+VASPGRGILAIDES+ATCGKRLASIGLDNTEVNRQAYRQLLLTTAGLGEYISGAILFEETLYQSTTDG
Sbjct: MAMVTAKLSSPAAASWL--TGGRTHRSAPTRRATLVRAAAGSYADELVSTAKSVASPGRGILAIDESNATCGKRLASIGLDNTEVNRQAYRQLLLTTAGLGEYISGAILFEETLYQSTTDG

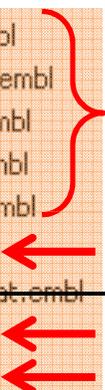
Tree @ NCBI



1_REPEATMASKER_TREP_plus.embl
2_BLASTX_TREP_prot.embl
5_AUGUSTUS_wheat.embl
6_BLASTN_cdsBdJGI.embl
6_BLASTN_cdsOsIRGSP.embl
6_BLASTN_cdsOsMSU.embl
6_BLASTN_cdsSbDOE.embl
6_BLASTN_cdsZmMSO.embl
6_BLASTN_HvFL.embl
6_BLASTN_SIMnuc_wheat.embl
6_BLASTN-TaEST.embl
6_BLASTN-TaFL.embl
6_EXONERATE_cdsBdJGI.embl
6_EXONERATE_cdsOsIRGSP.embl
6_EXONERATE_cdsOsMSU.embl
6_EXONERATE_cdsSbDOE.embl
6_EXONERATE_cdsZmMSO.embl
6_EXONERATE_HvFL.embl
~~6_EXONERATE_SIMnuc_wheat.embl~~
6_EXONERATE-TaEST.embl
6_EXONERATE-TaFL.embl
7_BLASTX_protBdJGI.embl
7_BLASTX_prothOR.embl
7_BLASTX_protOsIRGSP.embl
7_BLASTX_protOsMSU.embl
7_BLASTX_protSbDOE.embl
7_BLASTX_protTRI.embl
7_BLASTX_protZmMSO.embl
7_BLASTX_refSeqPlnProt.embl
7_BLASTX_SIMprot_wheat.embl
7_BLASTX_uniprot_sprot.embl
7_EXONERATE_protBdJGI.embl
7_EXONERATE_prothOR.embl

7_EXONERATE_protOsIRGSP.embl
7_EXONERATE_protOsMSU.embl
7_EXONERATE_protSbDOE.embl
7_EXONERATE_protTRI.embl
7_EXONERATE_protZmMSO.embl
7_EXONERATE_refSeqPlnProt.embl
~~7_EXONERATE_SIMprot_wheat.embl~~
7_EXONERATE_uniprot_sprot.embl
8_EUGENE.embl
9_GENEMODEL.embl
10_BESTHIT_BLASTP_nr.embl
10_BESTHIT_BLASTP_protAtTAIR.embl
10_BESTHIT_BLASTP_protBdJGI.embl
10_BESTHIT_BLASTP_prothOR.embl
10_BESTHIT_BLASTP_protOsIRGSP.embl
10_BESTHIT_BLASTP_protOsMSU.embl
10_BESTHIT_BLASTP_protPlant.embl
10_BESTHIT_BLASTP_protSAC.embl
10_BESTHIT_BLASTP_protSbDOE.embl
10_BESTHIT_BLASTP_protTRI.embl
10_BESTHIT_BLASTP_protZmMSO.embl
10_EXONERATE_nr.embl
10_EXONERATE_protAtTAIR.embl
10_EXONERATE_protBdJGI.embl
10_EXONERATE_prothOR.embl
10_EXONERATE_protOsIRGSP.embl
10_EXONERATE_protOsMSU.embl
10_EXONERATE_protPlant.embl
10_EXONERATE_protSAC.embl
10_EXONERATE_protSbDOE.embl
10_EXONERATE_protTRI.embl
10_EXONERATE_protZmMSO.embl
11_INTERPROSCAN.embl

12_BLASTN_genoAtTAIR.embl
12_BLASTN_genoBdJGI.embl
12_BLASTN_genoOsIRGSP.embl
12_BLASTN_genoOsMSU.embl
12_BLASTN_genoSbDOE.embl
12_BLASTN_genoZmMSO.embl
12_BLASTN_refSeq_chloro.embl
12_BLASTN_refSeq_mito.embl
13_BLASTX_protAtTAIR.embl
13_BLASTX_protBdJGI.embl
13_BLASTX_protOsIRGSP.embl
13_BLASTX_protOsMSU.embl
13_BLASTX_protSbDOE.embl
13_BLASTX_protZmMSO.embl
14_TRNASCAN-SE.embl
15_TRF.embl





**BLASTN
6_EXONERATE**

**Ctg954 (3 Mb)
Gene 2: 39717..43017**

**BLASTX
7_EXONERATE**

**BLASTP
10_EXONERATE**

11_INTERPROSCAN

- Hidden Best hits
- InterProScan
 - misc_feature
 - CTRL E



Ctg954 (3 Mb)
Gene 7: 1207699..1212541

- Create features from selection (Maj mouse)
 - delete
- Create features from coordinates (*Ta*-FL)
- Merge CDS (CTRL-MAJ)
- Remove selected feature (or del – the all transcript)
- Remove selected location (for one exon)
- Clone selected feature (no for exon/gene/mRNA)
- Split features between two selected locations (on CDS only)
 - merge again

1_REPEATMASKER_TREP_plus.embl
2_BLASTX_TREP_prot.embl
5_AUGUSTUS_wheat.embl
6_BLASTN_cdsBdJGI.embl
6_BLASTN_cdsOsIRGSP.embl
6_BLASTN_cdsOsMSU.embl
6_BLASTN_cdsSbDOE.embl
6_BLASTN_cdsZmMSO.embl
6_BLASTN_HvFL.embl
6_BLASTN_SIMnuc_wheat.embl
6_BLASTN-TaEST.embl
6_BLASTN-TaFL.embl
6_EXONERATE_cdsBdJGI.embl
6_EXONERATE_cdsOsIRGSP.embl
6_EXONERATE_cdsOsMSU.embl
6_EXONERATE_cdsSbDOE.embl
6_EXONERATE_cdsZmMSO.embl
6_EXONERATE_HvFL.embl
6_EXONERATE_SIMnuc_wheat.embl
6_EXONERATE-TaEST.embl
6_EXONERATE-TaFL.embl
7_BLASTX_protBdJGI.embl
7_BLASTX_protHOR.embl
7_BLASTX_protOsIRGSP.embl
7_BLASTX_protOsMSU.embl
7_BLASTX_protSbDOE.embl
7_BLASTX_protTRI.embl
7_BLASTX_protZmMSO.embl
7_BLASTX_refSeqPlnProt.embl
7_BLASTX_SIMprot_wheat.embl
7_BLASTX_uniprot_sprot.embl
7_EXONERATE_protBdJGI.embl
7_EXONERATE_protHOR.embl

7_EXONERATE_protOsIRGSP.embl
7_EXONERATE_protOsMSU.embl
7_EXONERATE_protSbDOE.embl
7_EXONERATE_protTRI.embl
7_EXONERATE_protZmMSO.embl
7_EXONERATE_refSeqPlnProt.embl
7_EXONERATE_SIMprot_wheat.embl
7_EXONERATE_uniprot_sprot.embl
8_EUGENE.embl
9_GENEMODEL.embl
10_BESTHIT_BLASTP_nr.embl
10_BESTHIT_BLASTP_protAtTAIR.embl
10_BESTHIT_BLASTP_protBdJGI.embl
10_BESTHIT_BLASTP_protHOR.embl
10_BESTHIT_BLASTP_protOsIRGSP.embl
10_BESTHIT_BLASTP_protOsMSU.embl
10_BESTHIT_BLASTP_protPlant.embl
10_BESTHIT_BLASTP_protSAC.embl
10_BESTHIT_BLASTP_protSbDOE.embl
10_BESTHIT_BLASTP_protTRI.embl
10_BESTHIT_BLASTP_protZmMSO.embl
10_EXONERATE_nr.embl
10_EXONERATE_protAtTAIR.embl
10_EXONERATE_protBdJGI.embl
10_EXONERATE_protHOR.embl
10_EXONERATE_protOsIRGSP.embl
10_EXONERATE_protOsMSU.embl
10_EXONERATE_protPlant.embl
10_EXONERATE_protSAC.embl
10_EXONERATE_protSbDOE.embl
10_EXONERATE_protTRI.embl
10_EXONERATE_protZmMSO.embl
11_INTERPROSCAN.embl

12_BLASTN_genoAtTAIR.embl
12_BLASTN_genoBdJGI.embl
12_BLASTN_genoOsIRGSP.embl
12_BLASTN_genoOsMSU.embl
12_BLASTN_genoSbDOE.embl
12_BLASTN_genoZmMSO.embl
12_BLASTN_refSeq_chloro.embl
12_BLASTN_refSeq_mito.embl
13_BLASTX_protAtTAIR.embl
13_BLASTX_protBdJGI.embl
13_BLASTX_protOsIRGSP.embl
13_BLASTX_protOsMSU.embl
13_BLASTX_protSbDOE.embl
13_BLASTX_protZmMSO.embl
14_TRNASCAN-SE.embl
15_TRF.embl

Panel III

**Conserved Non-coding
Sequences - CNSs**



Ctg954 (3 Mb)

- • 12_BLASTN - Plastid
 - Extension blue bar / right arrow
 - Goto position **714001**
 - **Check**
 - FL-cDNA
 - ESTs
 - CDS
 - augustus
 - EuGene

1_REPEATMASKER_TREP_plus.embl
2_BLASTX_TREP_prot.embl
5_AUGUSTUS_wheat.embl
6_BLASTN_cdsBdJGI.embl
6_BLASTN_cdsOsIRGSP.embl
6_BLASTN_cdsOsMSU.embl
6_BLASTN_cdsSbDOE.embl
6_BLASTN_cdsZmMSO.embl
6_BLASTN_HvFL.embl
6_BLASTN_SIMnuc_wheat.embl
6_BLASTN-TaEST.embl
6_BLASTN-TaFL.embl
6_EXONERATE_cdsBdJGI.embl
6_EXONERATE_cdsOsIRGSP.embl
6_EXONERATE_cdsOsMSU.embl
6_EXONERATE_cdsSbDOE.embl
6_EXONERATE_cdsZmMSO.embl
6_EXONERATE_HvFL.embl
6_EXONERATE_SIMnuc_wheat.embl
6_EXONERATE-TaEST.embl
6_EXONERATE-TaFL.embl
7_BLASTX_protBdJGI.embl
7_BLASTX_prothOR.embl
7_BLASTX_protOsIRGSP.embl
7_BLASTX_protOsMSU.embl
7_BLASTX_protSbDOE.embl
7_BLASTX_protTRI.embl
7_BLASTX_protZmMSO.embl
7_BLASTX_refSeqPlnProt.embl
7_BLASTX_SIMprot_wheat.embl
7_BLASTX_uniprot_sprot.embl
7_EXONERATE_protBdJGI.embl
7_EXONERATE_prothOR.embl

7_EXONERATE_protOsIRGSP.embl
7_EXONERATE_protOsMSU.embl
7_EXONERATE_protSbDOE.embl
7_EXONERATE_protTRI.embl
7_EXONERATE_protZmMSO.embl
7_EXONERATE_refSeqPlnProt.embl
7_EXONERATE_SIMprot_wheat.embl
7_EXONERATE_uniprot_sprot.embl
8_EUGENE.embl
9_GENEMODEL.embl
10_BESTHIT_BLASTP_nr.embl
10_BESTHIT_BLASTP_protAtTAIR.embl
10_BESTHIT_BLASTP_protBdJGI.embl
10_BESTHIT_BLASTP_prothOR.embl
10_BESTHIT_BLASTP_protOsIRGSP.embl
10_BESTHIT_BLASTP_protOsMSU.embl
10_BESTHIT_BLASTP_protPlant.embl
10_BESTHIT_BLASTP_protSAC.embl
10_BESTHIT_BLASTP_protSbDOE.embl
10_BESTHIT_BLASTP_protTRI.embl
10_BESTHIT_BLASTP_protZmMSO.embl
10_EXONERATE_nr.embl
10_EXONERATE_protAtTAIR.embl
10_EXONERATE_protBdJGI.embl
10_EXONERATE_prothOR.embl
10_EXONERATE_protOsIRGSP.embl
10_EXONERATE_protOsMSU.embl
10_EXONERATE_protPlant.embl
10_EXONERATE_protSAC.embl
10_EXONERATE_protSbDOE.embl
10_EXONERATE_protTRI.embl
10_EXONERATE_protZmMSO.embl
11_INTERPROSCAN.embl

12_BLASTN_genoAtTAIR.embl
12_BLASTN_genoBdJGI.embl
12_BLASTN_genoOsIRGSP.embl
12_BLASTN_genoOsMSU.embl
12_BLASTN_genoSbDOE.embl
12_BLASTN_genoZmMSO.embl
12_BLASTN_refSeq_chloro.embl
12_BLASTN_refSeq_mito.embl
13_BLASTX_protAtTAIR.embl
13_BLASTX_protBdJGI.embl
13_BLASTX_protOsIRGSP.embl
13_BLASTX_protOsMSU.embl
13_BLASTX_protSbDOE.embl
13_BLASTX_protZmMSO.embl
14_TRNASCAN-SE.embl
15_TRF.embl

← Panel IV ●

Microsatellite markers SSRs

1_REPEATMASKER_TREP_plus.embl
2_BLASTX_TREP_prot.embl
5_AUGUSTUS_wheat.embl
6_BLASTN_cdsBdJGI.embl
6_BLASTN_cdsOsIRGSP.embl
6_BLASTN_cdsOsMSU.embl
6_BLASTN_cdsSbDOE.embl
6_BLASTN_cdsZmMSO.embl
6_BLASTN_HvFL.embl
6_BLASTN_SIMnuc_wheat.embl
6_BLASTN-TaEST.embl
6_BLASTN-TaFL.embl
6_EXONERATE_cdsBdJGI.embl
6_EXONERATE_cdsOsIRGSP.embl
6_EXONERATE_cdsOsMSU.embl
6_EXONERATE_cdsSbDOE.embl
6_EXONERATE_cdsZmMSO.embl
6_EXONERATE_HvFL.embl
6_EXONERATE_SIMnuc_wheat.embl
6_EXONERATE-TaEST.embl
6_EXONERATE-TaFL.embl
7_BLASTX_protBdJGI.embl
7_BLASTX_prothOR.embl
7_BLASTX_protOsIRGSP.embl
7_BLASTX_protOsMSU.embl
7_BLASTX_protSbDOE.embl
7_BLASTX_protTRI.embl
7_BLASTX_protZmMSO.embl
7_BLASTX_refSeqPlnProt.embl
7_BLASTX_SIMprot_wheat.embl
7_BLASTX_uniprot_sprot.embl
7_EXONERATE_protBdJGI.embl
7_EXONERATE_prothOR.embl

Panel I

7_EXONERATE_protOsIRGSP.embl
7_EXONERATE_protOsMSU.embl
7_EXONERATE_protSbDOE.embl
7_EXONERATE_protTRI.embl
7_EXONERATE_protZmMSO.embl
7_EXONERATE_refSeqPlnProt.embl
7_EXONERATE_SIMprot_wheat.embl
7_EXONERATE_uniprot_sprot.embl
8_EUGENE.embl
9_GENEMODEL.embl
10_BESTHIT_BLASTP_nr.embl
10_BESTHIT_BLASTP_protAtTAIR.embl
10_BESTHIT_BLASTP_protBdJGI.embl
10_BESTHIT_BLASTP_prothOR.embl
10_BESTHIT_BLASTP_protOsIRGSP.embl
10_BESTHIT_BLASTP_protOsMSU.embl
10_BESTHIT_BLASTP_protPlant.embl
10_BESTHIT_BLASTP_protSAC.embl
10_BESTHIT_BLASTP_protSbDOE.embl
10_BESTHIT_BLASTP_protTRI.embl
10_BESTHIT_BLASTP_protZmMSO.embl
10_EXONERATE_nr.embl
10_EXONERATE_protAtTAIR.embl
10_EXONERATE_protBdJGI.embl
10_EXONERATE_prothOR.embl
10_EXONERATE_protOsIRGSP.embl
10_EXONERATE_protOsMSU.embl
10_EXONERATE_protPlant.embl
10_EXONERATE_protSAC.embl
10_EXONERATE_protSbDOE.embl
10_EXONERATE_protTRI.embl
10_EXONERATE_protZmMSO.embl
11_INTERPROSCAN.embl

12_BLASTN_genoAtTAIR.embl
12_BLASTN_genoBdJGI.embl
12_BLASTN_genoOsIRGSP.embl
12_BLASTN_genoOsMSU.embl
12_BLASTN_genoSbDOE.embl
12_BLASTN_genoZmMSO.embl
12_BLASTN_refSeq_chloro.embl
12_BLASTN_refSeq_mito.embl
13_BLASTX_protAtTAIR.embl
13_BLASTX_protBdJGI.embl
13_BLASTX_protOsIRGSP.embl
13_BLASTX_protOsMSU.embl
13_BLASTX_protSbDOE.embl
13_BLASTX_protZmMSO.embl
14_TRNASCAN-SE.embl
15_TRF.embl

**Transposable Elements
TEs**

Transposable Elements “code”

Wicker *et al.* (2007)

A unified classification system for eukaryotic transposable elements.

Nature Reviews Genetics
8:973-982



Classification		Structure	TSD	Code	Occurrence
Order	Superfamily				
<i>Class I (retrotransposons)</i>					
LTR	<i>Copia</i>	→ GAG AP INT RT RH →	4-6	RLC	P, M, F, O
	<i>Gypsy</i>	→ GAG AP RT RH INT →	4-6	RLG	P, M, F, O
	<i>Bel-Pao</i>	→ GAG AP RT RH INT →	4-6	RLB	M
	<i>Retrovirus</i>	→ GAG AP RT RH INT ENV →	4-6	RLR	M
	<i>ERV</i>	→ GAG AP RT RH INT ENV →	4-6	RLE	M
DIRS	<i>DIRS</i>	→ GAG AP RT RH YR ←	0	RYD	P, M, F, O
	<i>Ngaro</i>	→ GAG AP RT RH YR → → →	0	RYN	M, F
	<i>VIPER</i>	→ GAG AP RT RH YR → → → →	0	RYV	O
PLE	<i>Penelope</i>	← RT EN →	Variable	RPP	P, M, F, O
LINE	<i>R2</i>	RT EN	Variable	RIR	M
	<i>RTE</i>	APE RT	Variable	RIT	M
	<i>Jockey</i>	ORF1 APE RT	Variable	RIJ	M
	<i>L1</i>	ORF1 APE RT	Variable	RIL	P, M, F, O
	<i>I</i>	ORF1 APE RT RH	Variable	RII	P, M, F
SINE	tRNA		Variable	RST	P, M, F
	7SL		Variable	RSL	P, M, F
	5S		Variable	RSS	M, O
<i>Class II (DNA transposons) - Subclass 1</i>					
TIR	<i>Tc1-Mariner</i>	Tase*	TA	DTT	P, M, F, O
	<i>hAT</i>	Tase*	8	DTA	P, M, F, O
	<i>Mutator</i>	Tase*	9-11	DTM	P, M, F, O
	<i>Merlin</i>	Tase*	8-9	DTE	M, O
	<i>Transib</i>	Tase*	5	DTR	M, F
	<i>P</i>	Tase	8	DTP	P, M
	<i>PiggyBac</i>	Tase	TTAA	DTB	M, O
	<i>PIF-Harbinger</i>	Tase* ORF2	3	DTH	P, M, F, O
	<i>CACTA</i>	Tase ORF2	2-3	DTC	P, M, F
Crypton	<i>Crypton</i>	YR	0	DYC	F
<i>Class II (DNA transposons) - Subclass 2</i>					
Helitron	<i>Helitron</i>	RPA Y2 HEL	0	DHH	P, M, F
Maverick	<i>Maverick</i>	C-INT ATP CYP POL B	6	DMM	M, F, O



Ctg954 (3 Mb)

- Save Session
- Search / Sequence search (aa)
- Export image
- Configuration

- Save annotation
 - Keep: Gene/mRNA/Exon/CDS/
 - no sequence
 - the organization is different as a standard embl file
 - see embl file



In summary for GenomeView you need the correspondence between the embl file name and the feature key.

1_REPEATMASKER_TREP_plus → **repeat_region**

2_BLASTX_TREP_prot → **BLASTX_TREP_pro**

5_AUGUSTUS_wheat → **AUGUSTUS_wheat**

6_EXONERATE_cdsBdJGI → **EXO_N_cdsBdJGI**

6_EXONERATE_cdsOsIRGSP → **EXO_N_cdsOsIRGSP**

6_EXONERATE_cdsOsMSU → **EXO_N_cdsOsMSU**

6_EXONERATE_cdsSbDOE → **EXO_N_cdsSbDOE**

6_EXONERATE_cdsZmMSO → **EXO_N_cdsZmMSO**

6_EXONERATE_HvFL → **EXO_N_HvFL**

6_EXONERATE_TaEST → **EXO_N_TaEST**

6_EXONERATE_TaFL → **EXO_N_TaFL**

7_EXONERATE_protBdJGI → **EXO_X_protBdJGI**

7_EXONERATE_protHOR → **EXO_X_protHOR**

7_EXONERATE_protOsIRGSP → **EXO_X_protOsIRGSP**

7_EXONERATE_protOsMSU → **EXO_X_protOsMSU**

7_EXONERATE_protSbDOE → **EXO_X_protSbDOE**

7_EXONERATE_protTRI → **EXO_X_protTRI**

7_EXONERATE_protZmMSO → **EXO_X_protZmMSO**

7_EXONERATE_refSeqPlnProt → **EXO_X_refSeqPln**

7_EXONERATE_uniprot_sprot → **EXO_X_uniprot_s**

8_EUGENE → **EUGENE**

9_GENEMODEL → **gene/mRNA/exon/CDS**

10_EXONERATE_nr → **EXO_P_nr**

10_EXONERATE_protAtTAIR → **EXO_P_protAtTAIR**

10_EXONERATE_protBdJGI → **EXO_P_protBdJGI**

10_EXONERATE_protHOR → **EXO_P_protHOR**

10_EXONERATE_protOsIRGSP → **EXO_P_protOsIRGSP**

10_EXONERATE_protOsMSU → **EXO_P_protOsMSU**

10_EXONERATE_protPlant → **EXO_P_protPlant**

10_EXONERATE_protSAC → **EXO_P_protSAC**

10_EXONERATE_protSbDOE → **EXO_P_protSbDOE**

10_EXONERATE_protTRI → **EXO_P_protTRI**

10_EXONERATE_protZmMSO → **EXO_P_protZmMSO**

11_INTERPROSCAN → **misc_feature !!!!**

12_BLASTN_genoAtTAIR → **BLASTN_genoAtTA**

12_BLASTN_genoBdJGI → **BLASTN_genoBdJG**

12_BLASTN_genoOsIRGSP → **BLASTN_genoOsIR**

12_BLASTN_genoOsMSU → **BLASTN_genoOsMS**

12_BLASTN_genoSbDOE → **BLASTN_genoSbDO**

12_BLASTN_genoZmMSO → **BLASTN_genoZmMS**

12_BLASTN_refSeq_chloro → **BLASTN_refSeq_c**

12_BLASTN_refSeq_mito → **BLASTN_refSeq_m**

13_BLASTX_protAtTAIR → **BLASTX_genoAtTA**

13_BLASTX_protBdJGI → **BLASTX_genoBdJG**

13_BLASTX_protOsIRGSP → **BLASTX_genoOsIR**

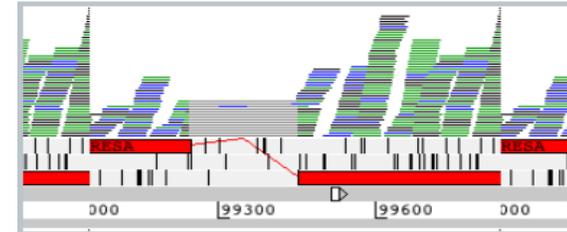
13_BLASTX_protOsMSU → **BLASTX_genoOsMS**

13_BLASTX_protSbDOE → **BLASTX_genoSbDO**

13_BLASTX_protZmMSO → **BLASTX_genoZmMS**

14_TRNASCAN-SE → **tRNA**

15_TRF → **tandem_repeat**



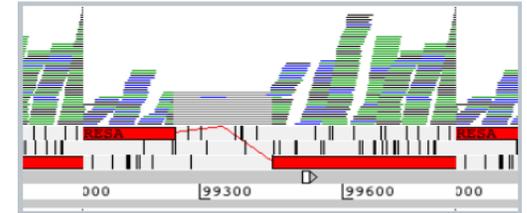
Artemis (13.2.0)

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

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

Artemis (13.2.0)

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



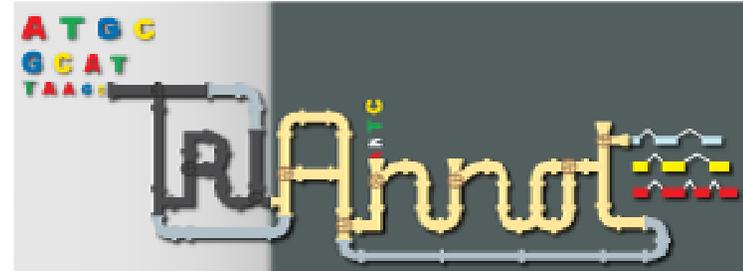
Ctg954 (3 Mb)
Gene 2: 39717..43017



- File/Open initial.seq
- File/Read an entry into (GenomeView EMBL)
 - **TEs**
 - Selected
 - copy to initial
 - Save as ...
 - check
 - **Wheat FL-cDNA**
 - **Wheat ESTs**
 - **Proteins (SwissProt)**
 - **Best-hits (Prot Plant)**
 - **Delete entries**
- **Subsequence**
 - **Save as ...**

5:25 – 5:30 Conclusions

*P. Leroy &
N. Guilhot*

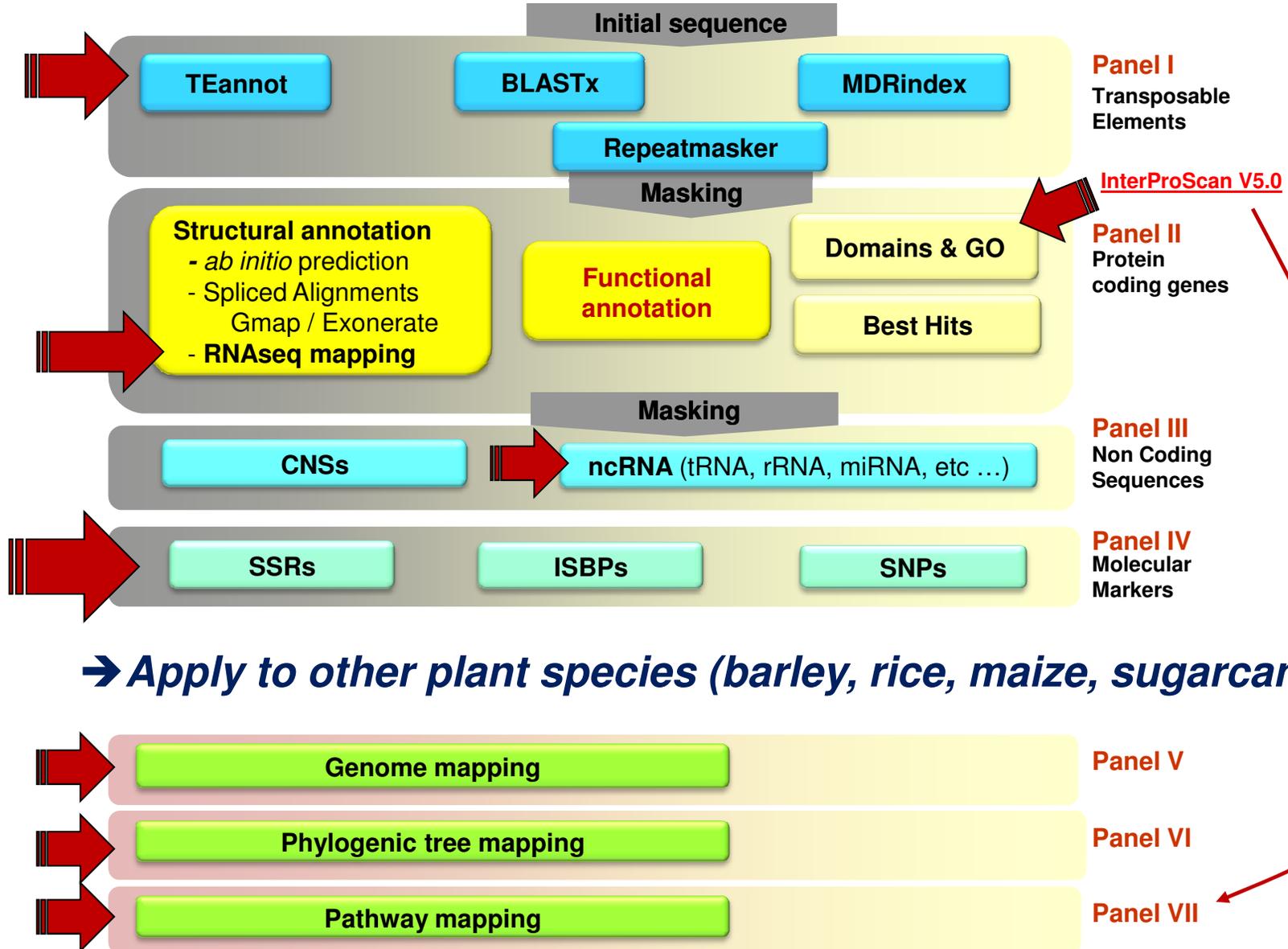


V3.5

X **Perspectives**
New modules, new species

X **Acknowledgements**

Perspectives



Triannot Pipeline

[Help](#)

[Versions](#)

[Architecture](#)

[Usage](#)

[Softwares](#)

[Databanks](#)

[Defaults](#)

[Links](#)

[Presentation in congress](#)

[Run Pipeline](#)

[Acknowledgments](#)

Acknowledgments

The French INRA TriAnnot pipeline has been created on **March 2005** by **P. Leroy & F. Legeai**. A major release has been made in **2011** thanks to **N. Guilhot**.

This project is supported by :

- The "Département de Génétique et d'Amélioration des Plantes" of [INRA](#) (*National Institute of Agronomic Research*),
- The French "Programme Fédérateur en Génomique Végétale" - [Genoplante](#) ,
- The "Pôle de Compétitivité" - [Céréales Vallée](#) ,
- The "Programme Régional d'Actions Innovatrices" (PRAI) e-nnovergne [LifeGrid](#) through a European project "Actions innovatrices du FEDER" (2006-2008) coordinated by the "[Région Auvergne](#)" .
- [TriticeaeGenome](#) , a european EC 7th Framework Program (Food, Agriculture and Fisheries) project (2008 – 2011) lead by **C. Feuillet**,
- [3BSeq](#) , a flagship project funded by the [ANR](#) and [France Agrimer](#) for a duration of 3 years (2010-2013) and lead by **C. Feuillet**,
- *The International Wheat Genome Sequencing Project* - [IWGSC](#)
- *The International Barley Sequencing Project* - [IBSC](#)
- [GrainGenes](#)

Coordinators :

- Philippe LEROY & Catherine FEUILLET - [UMR 1095 INRA-UBP Génétique, Diversité, Ecophysiologie des Céréales \(GDEC\), Clermont-Ferrand](#).
- Michael ALAUX & Hadi QUESNEVILLE - [INRA Unité de Recherche Génomique-Info \(URGI\), Versailles](#).

We would like to thank the following persons for their contribution in this project:

Contribution	Persons
TriAnnot pipeline architecture	Nicolas GUILHOT, Aurélien BERNARD, Franck GIACOMONI, Frédéric CHOULET, Matthieu REICHSTADT, Olivier INIZAN, Fabrice LEGEAI, Emmanuelle GICQUELOT, Lorenzo CERUTTI, Bastien LAUBIN, Philippe LEROY
Databanks update	Philippe LEROY, Nicolas LAPALU, Sébastien REBOUX, Christophe CARON, Matthieu REICHSTADT
NIASsearch "Gene modeling" - Collaboration with NIAS at Tsukuba, Japan	Hiroaki SAKAI, Tsuyoshi TANAKA, Naoki AMANO, Hisataka NUMA, Franck GIACOMONI, Takeshi ITOH, Philippe LEROY



Acknowledgements



N. Guilhot
S. Theil
F. Choulet
P. Leroy
C. Feuillet



M. Alaux
S. Reboux
H. Quesneville



MIPS
M. Seidel
K. Mayer



H. Sakai
N. Amano
H. Numa
T. Tanaka
T. Itoh



RiceGAAS
H. Ohyanagi



Previous Collaborators

P. Dufour
L. Cerutti
B. Laubin
F. Sabot
F. Legeai
E. Gicquelo
F. Giacomoni
C. Caron
I. Blanc-Lenfle
A. Mahul
A. Claude
M. Liauzu
T. Flutre
I. Luyten
C. Pelegrin
A. Bernard
O. Inizan
M. Reichstadt

Thank You very much



Massif du Sancy
Joël Damase



Muroi Castle

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

philippe.leroy@clermont.inra.fr

nicolas.guilhot@clermont.inra.fr