



GENOME ANNOTATION

INTRODUCTION TO CONCEPTS AND METHODS

Olivier GARSMEUR

Training course in Bioinformatics applied to *Musa* genome
18-22 November 2013

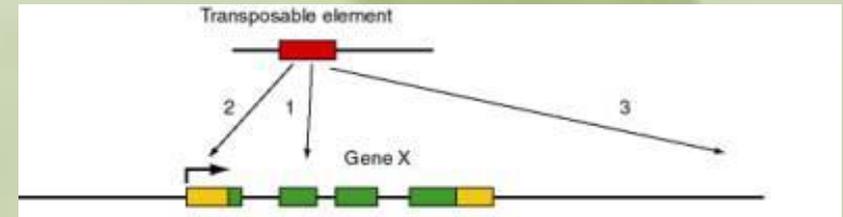


Introduction

two main concepts:

Identify the different elements of the genome, (location and structure) :

structural annotation



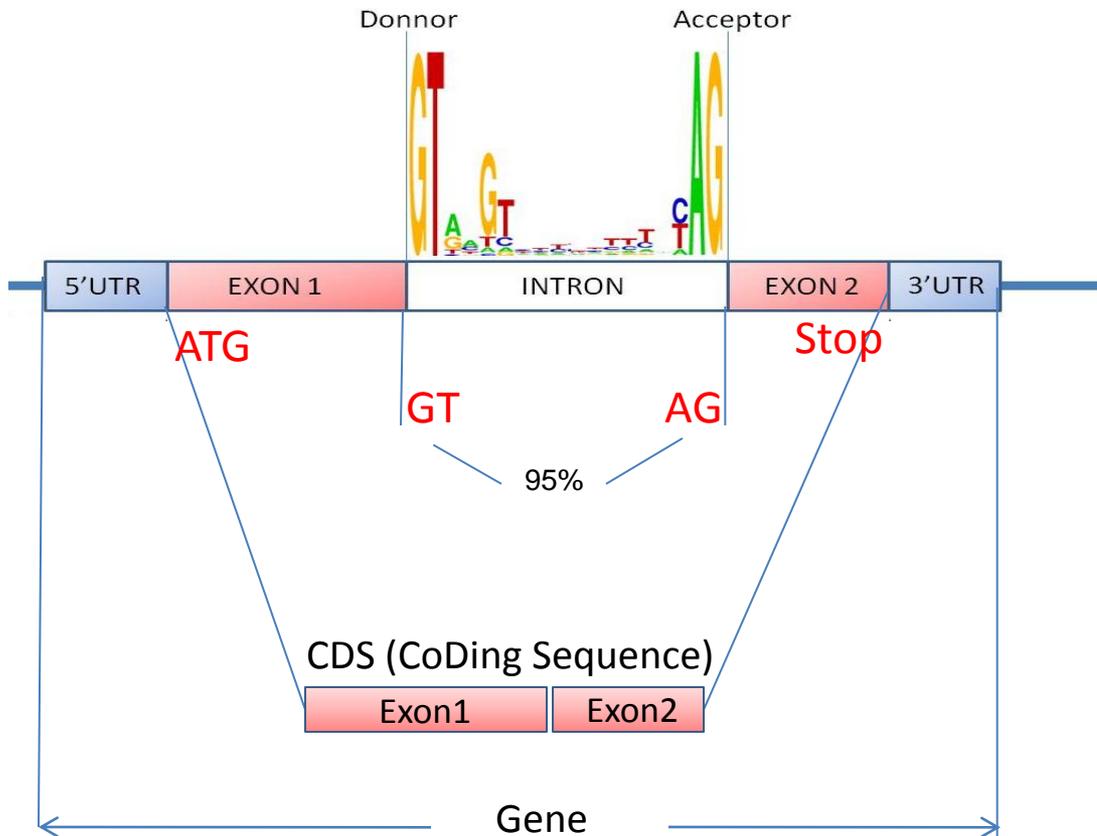
Attribute a biological information to these elements : **functional annotation**

Amino acid biosynthesis	Cellular processes	DNA metabolism
Biosynthesis of cofactors, prosthetic groups, and carriers	Central intermediary metabolism	Energy metabolism
Cell envelope	Disrupted reading frame	Fatty acid and phospholipid metabolism
Hypothetical proteins	Pathogen responses	Purines, pyrimidines, nucleosides, and nucleotides
Hypothetical proteins (conserved)	Protein fate	Regulatory functions
Mobile and extrachromosomal element functions	Protein synthesis	Signal transduction
Transcription	Transport and binding proteins	Unclassified
Unknown function	Viral functions	

Gene structure

Where are the genes on the sequence ?

Genes predictions (gene-finder softwares) are based on the structure (intron, exon, splice site, UTR).



Predicted functional gene

→ Structure is complete

Starts with M (ATG)

Stop codon TAA, TAG or TGA

GT (GC) / AG splicing site

No stop in exon frame

Pseudogene (not functional)

→ Structure is not complete

missing_acceptor

missing_donor

missing_start_codon

missing_stop_codon

multiple_stop_in_frame

Intrinsic methods

(ab-initio)



- Only based on computational analysis using statistical models.
- Probabilistic models like *Hidden Markov models* (HMM) for discriminating coding and non-coding region of the genome.

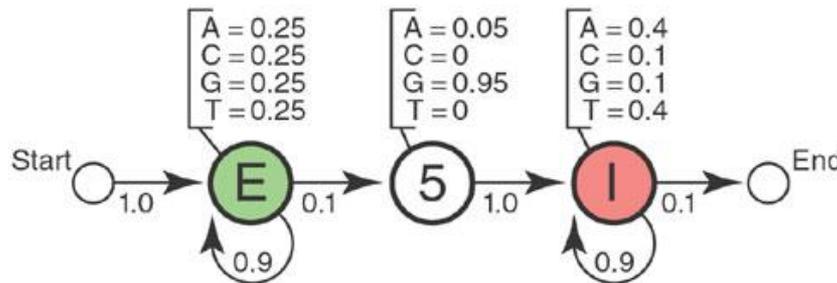
- Need a training set genes. (learning)

The learning set is composed of several hundred of gene-sequences manually annotated derived from cDNAs / genomic alignments. Ideally, these genes represent the diversity of the genes that can be found in the genome.

Intrinsic methods : Hidden Markov Model = HMM

The sequences of exons, splice sites and introns have different statistical properties, such as GC%. Introns are AT rich and splicing site consensus is almost GT (95%)

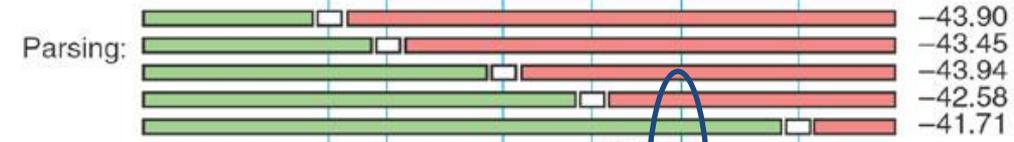
Example: Where is the 5' splicing site (5) corresponding to the transition between the exon (E) and the intron (I). ➡ 3 states : E, I, and 5



Sequence: **CTTCATGTGAAAGCAGACGTAAGTCA**

State path: **EEEEEEEEEEEEEEEEEE5IIIIII**

$\log P$
-41.22



Probabilities of moving from a state to another state

The transition probabilities describe the linear order in which we expect the states to occur: Exons, Introns, Splicing site.

HMM based gene structure prediction : FGENESH

<http://linux1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind>



HOME ALL SOFTWARE PRODUCTS NEW PRODUCTS SERVICES MANAGEMENT TEAM CORPORATE PROFILE CONTACT

TEST ON LINE

- GENE FINDING in Eukaryota
- GENE FINDING WITH SIMILARITY
- OPERON AND GENE FINDING IN BACTERIA
- GENE FINDING IN VIRUSES
- NEXT GENERATION
- ALIGNMENT /Sequences&genomes
- GENOME EXPLORER /Infogene
- SEARCH FOR MOTIFS /promoters&functional
- PROTEIN LOCATION /patterns/Epitops
- RNA STRUCTURE COMPUTING
- PROTEIN STRUCTURE
- PROTEIN / DNA 3D-Visual Works
- SEQMAN
- MULTIPLE ALIGNMENTS
- ANALYSIS OF EXPRESSION DATA
- PLANT PROMOTERS DATABASE
- REPEATS /find&map repeats
- SNP Extracting known SNPs
- Proteomics

FGENESH

Reference: Solovyev V, Kosarev P, Seledsov I, Vorobyev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* 2006,7, Suppl 1: P. 10.1-10.12.

HMM-based gene structure prediction (multiple genes, both chains). The Fgenesh gene-finder was selected as the most accurate program for plant gene identification. *Plant Molecular Biology* (2005), 57, 3, 445-460: "Five ab initio programs (FGENESH, GeneMark.hmm, GENSCAN, GlimmerR and Grail) were evaluated for their accuracy in predicting maize genes. FGENESH yielded the most accurate and GeneMark.hmm the second most accurate predictions" (FGENESH identified 11% more correct gene models than GeneMark on a set of 1353 test genes).

Example: Search in -chain

Paste nucleotide sequence here:

```
>0
TCACTGCAAGACTCTCCTTTCTTGTACACTCATCAAGTAAGAAGAGGGCCAGTTGGTGG
CAAAGGCCAAAGAGTTAAGATGGGAGACGCAGCATTGTAGGCTGTGGTTAGAGCCCTCGC
TATTAGAGA.AAGGGGATTCTACGGGGATCAGAACTTGGTCCCTCGCCCATGAGTTT
AAGGACGGCAAAGTTGTAAGTGGCAGCAATCATGAAGGTGAGCTGTAGGGAACAGAAAG
AGAATGTGAATGAGAGTAGA.AAAGAAAGCACTCCCTCCTCAAATCCCTTTGGGAATGAGG
AAGGGTAGAAATA.AAAATCATA.ACTAAATAAGGGAAAAACCCATGCTCTGCA.AAAAGCAGT
ACATCCCAGGCTTGGCTTTCTCCATATACACTTGAAGTGGAAAGTCCGATCTCTGAGGCA
TCATGGGCTCATGGCCAGGCCAGAGAATCA.AATCATA.AAGATTATTCACACCTCT
GTTGATCTAGCACTGAGTTTTCTCAGGGAAGTGAACAATATTTGGCTTGGCCACATCA
ATATGTCGAGTGACTTAGAGTCATAGGGTGCCTTGTCTGTAGACACACAGCCACATCATGC
AGTAGAGCTCTCCCTGGTTAATGAATCA.ATTCTGTTTTACAAGGTGAGAAATACCTTGAC
TCTGTATGAAACCTCAGCAGTAGGGGCCAGGAGTTCTGTGAGCTCATCTTGTCTGATCA
CTGCTATA.AAATA.AACTTCAGATA.AATCACTTCAACCTTCAGTGCCTTAGGTTTTTAATC
TGCACAATGAGCACACTACATTACAGGGCAGCTGTAGCACTATCAGCAAGACCTGGGACA
TAGCATTGTATTATAATCATA.ACAATTTCTAATAATGTTAGGAGGTCCCTCCAGTATTTCC
ATATGAGTA.ACTAAATCCACTGGAGAGTATCCA.AAGTGGATTTCATATGATAGTATATAG
A.ATTCCAGGCTATAGCAAGTCATGTTTCGTGATGTTTCAGTGTTTTGTCACTGATCTGAT
CA.AAATGATTGAGA.ATTAACCGTA.ATTTATGACTGCA.AAGTACTTGGCAACATTACATG
```

Alternatively, load a local file with sequence in Fasta format:

Local file name:

Select organism specific gene-finding parameters (from 115 model organisms):

[\[Help\]](#) [\[Show advanced options\]](#) [\[Program's page\]](#)
[\[Example: Homo sapiens genomic beta globin region \(HBB@\) on chromosome 11\]](#)

Return to page with other programs of group: [Gene finding](#)

Automatic annotation : tools

Example of prediction softwares

Plants



GeneMark.HMM

GeneFinder

Eugene-HMM

FgeneSH

GlimmerA

Augustus



Animals



Genie

HMMgene

MagPie

GeneID

Grail

Human



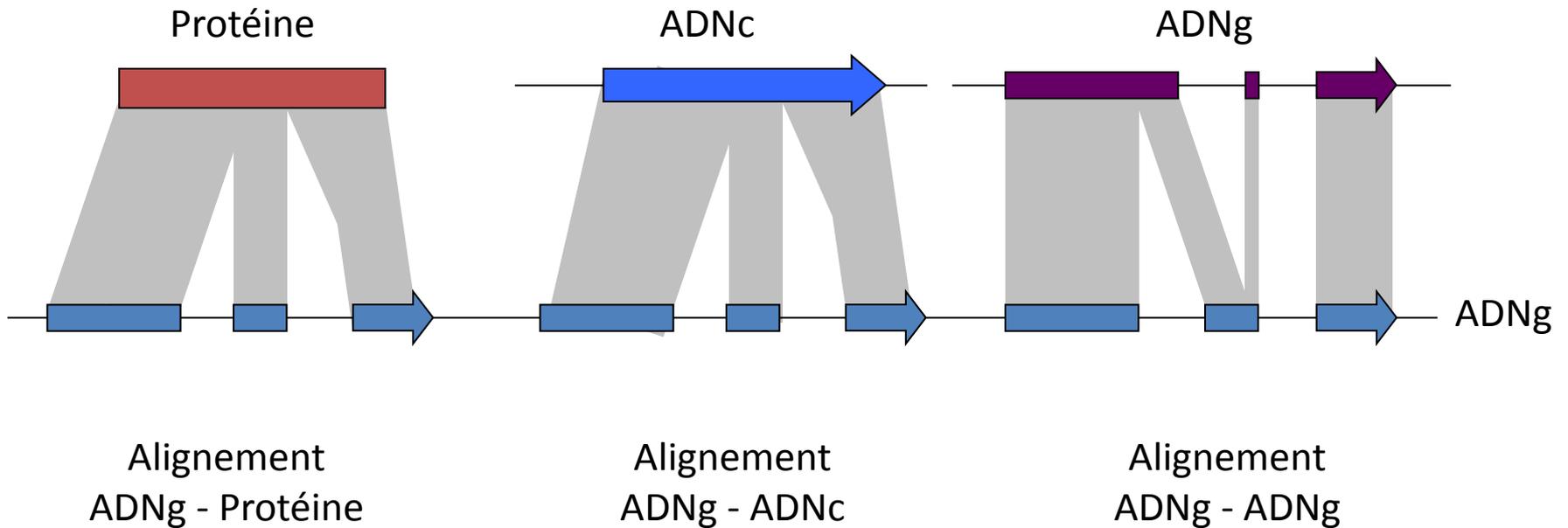
GeneScan

GeneFinder

GeneWise

Extrinsic methods

Comparative approach based on sequence similarities.
The sequence to annotate is compared with databases.



Extrinsic methods

	sequence (query)	target (Subject)	database
BLAST ^N	nucleotide	nucleotide	NR , EST, genomes
BLAST ^X	translated nucleotide	protein	Swissprot-Treml
BLAST ^P	protein	protein	Swissprot-Treml
TBLAST ^X	translated nucleotide	translated nucleotide	NR , EST, genomes
TBALST ^N	protein	translated nucleotide	NR , EST, genomes

Comparisons with existing genes or proteins is helpful to refine the structure of predictions.

The Functional annotation will be always deduced of the similarities (homology) of the predicted elements with databases.

Other alignment tools

Genome Threader

Use a similarity-based approach where cDNA/EST and/or protein sequences are compared to the genomic sequence

Useful to predict the splice sites

www.genomethreader.org/

Exonarate

This is a generic tool for pairwise sequence comparison using a many alignment models

This tool is used to refine the structure of gene models (based on cDNA alignments)

<http://www.ebi.ac.uk/~guy/exonarate/>

Search for conserved protein domains

INTERPROSCAN

<http://www.ebi.ac.uk/interpro/>

EMBL-EBI EB-eye Search All Databases Enter Text Here Go Reset Advanced Search Give us feedback

Databases Tools EBI Groups Training Industry About Us Help Site Index

- InterPro home
- Text Search
- InterProScan

EBI > Tools > Protein Functional Analysis

InterProScan Sequence Search

conserved protein domains = signatures

database of predictive protein "signatures" can be used for the classification and automatic annotation of proteins.

Interproscan classifies sequences at superfamily, family and subfamily levels, predicting the occurrence of functional domains and important sites.

Domain databases used by interproscan:

Prosite patterns	GENE3D
Pfam	HAMAP
ProDom	PANTHER
Superfamily	PIRSF
TIGRFAMs	

database limits

Proteins databases are almost derived from the automatic translation of nucleic sequences. For example, TrEMBL is the automatic translation of the nucleic EMBL (NR) database. (No biological evidence)

→ SWISSPROT, is another protein database that contains less sequences than TrEMBL, but sequences are manually curated by biologists.

ESTs sequences are often produced from single strand sequencing and contain mistakes.

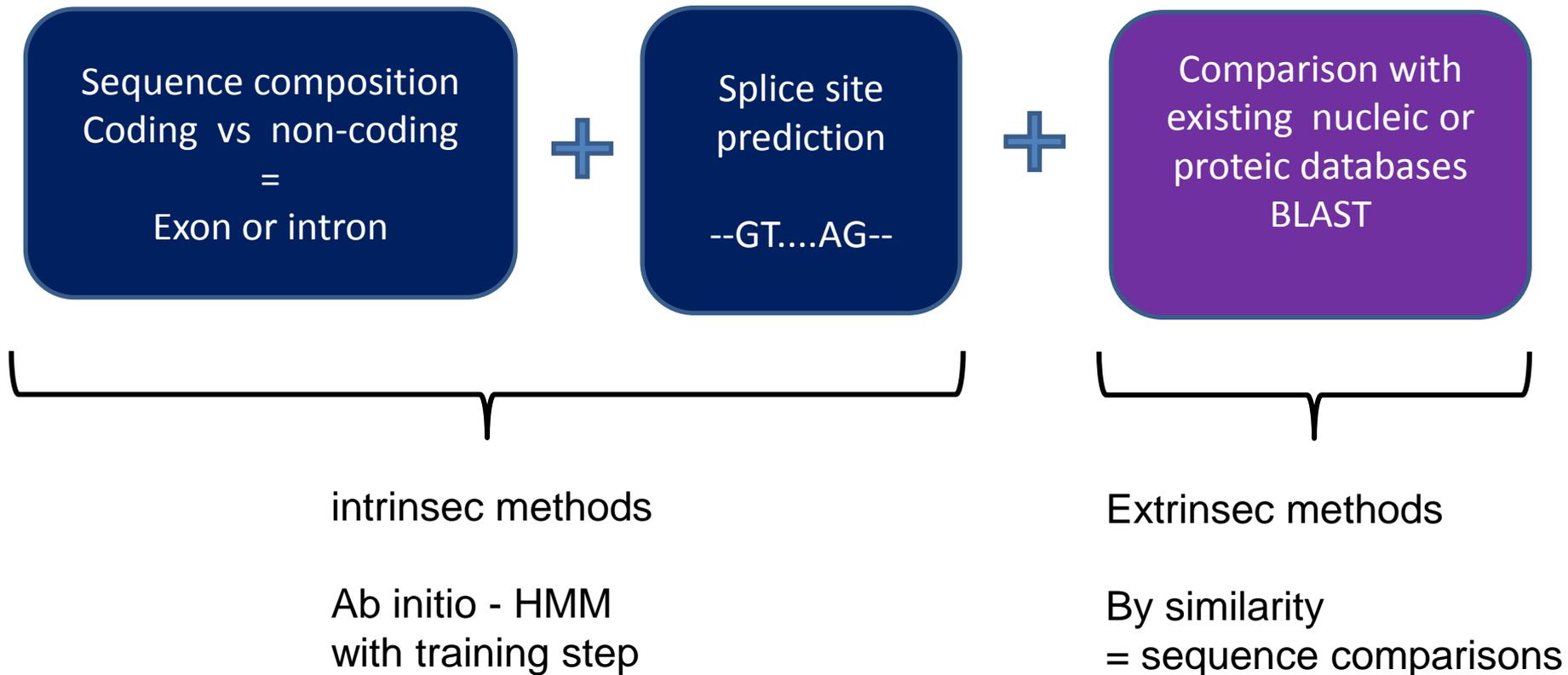
→ Cluster of ESTs are available on a lot of species and represent genes that are really expressed.

Genome sequences are useful, but gene predictions are based on automatic annotations. Taking in account genome evolutions, only close relative species can be used to transfer the annotation.

→ For sugarcane annotation, the sorghum sequence is available and represents a good model for comparison.



Integrative method – The combiner

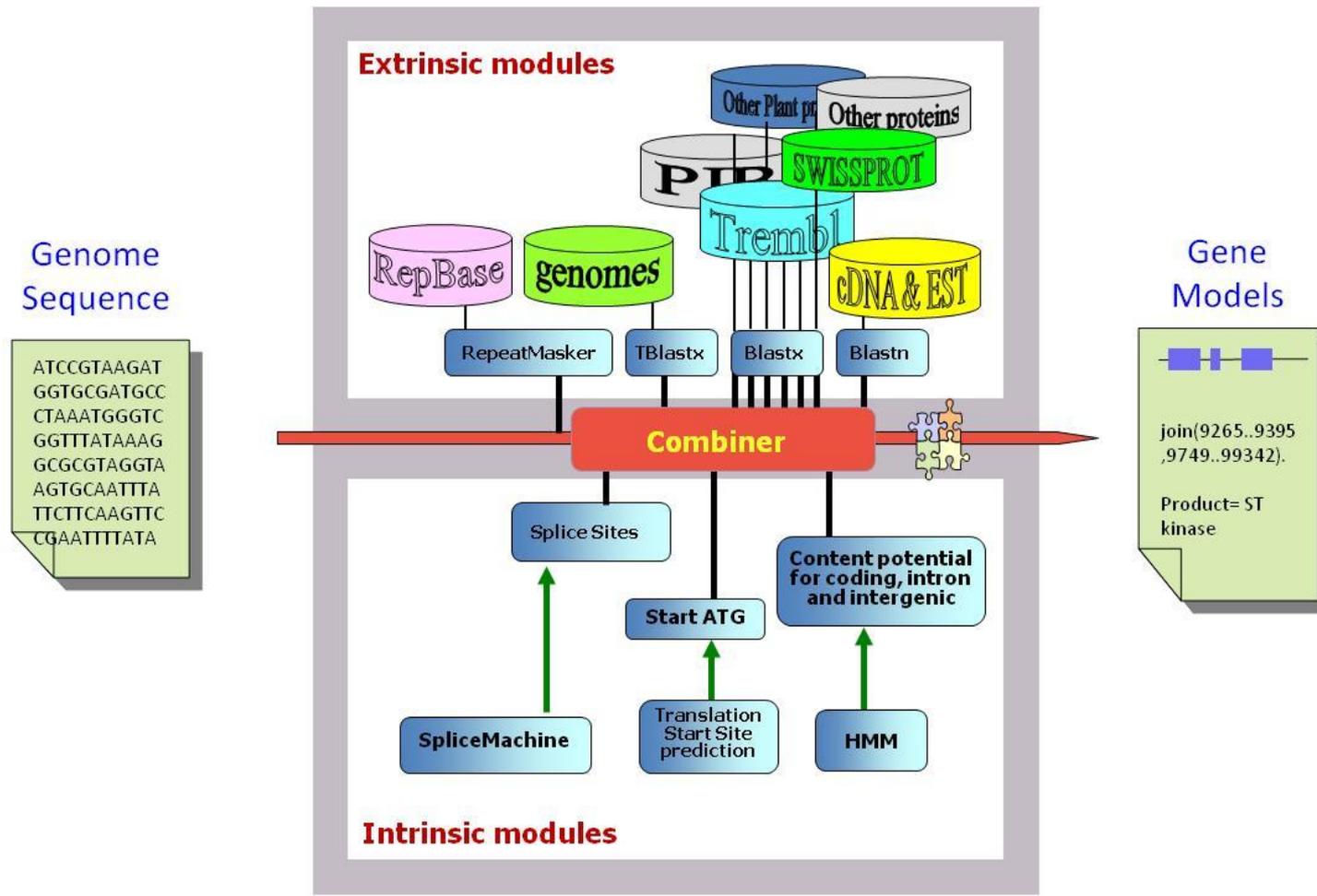


Integrative method – The combiner

Integrative methods = ab-initio + comparative approaches

Predictions of ab-initio gene finders combined with the database similarities improve significantly the annotation.

Intrinsic and extrinsic method complement each other.



Automatic annotation tool : Eugene (combiner)

<http://eugene.toulouse.inra.fr/>

EuGène : A Gene Finding Software

[P. BARDOU](#) [M.J. CROS](#) [S. FOISSAC](#) [J. GOUZY](#) [A. MOISAN](#) [T. SCHIEX](#)

Home

Plugins

Organisms

References

Documentation

Download

GForge

Links - Contacts

Organisms

Arabidopsis thaliana :



<i>Arabidopsis thaliana</i> (L.) Heynh. mouseear cress	
Group	Dicot
Family	Brassicaceae
Growth Habit	Forb/herb
Duration	Annual
Links	
USDA, NRCS. 2002. The PLANTS Database, Version 3.5.	
ARS Germplasm Resources Information Network (GRIN)	



[An evaluation of EuGène and several gene prediction programs was performed.](#)

Oryza sativa :

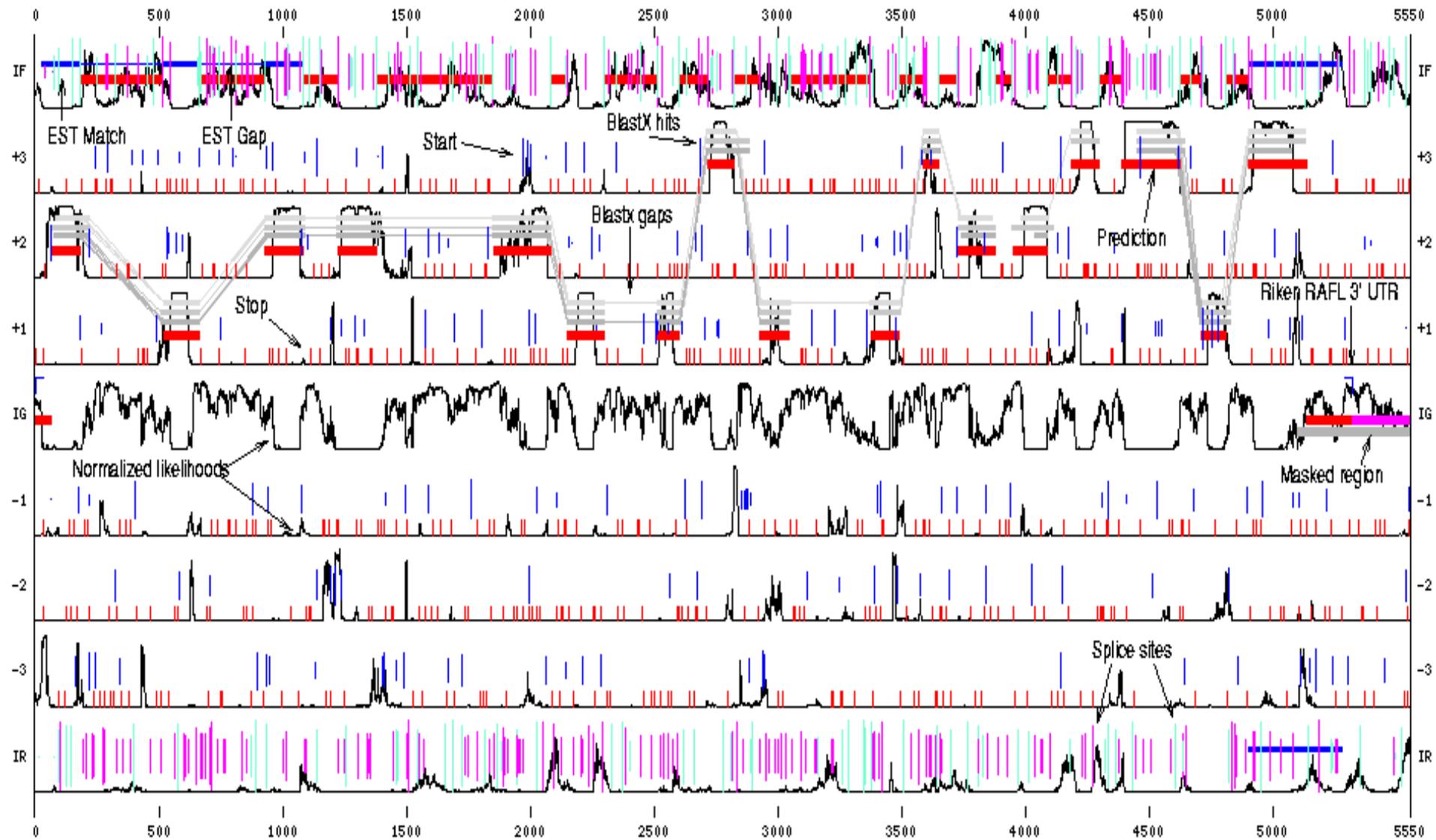


<i>Oryza sativa</i> L. rice	
Group	Monocot
Family	Proceae
Growth Habit	Graminoid
Duration	Annual
Links	
USDA, NRCS. 2002. The PLANTS Database, Version 3.5.	
ARS Germplasm Resources Information Network (GRIN)	

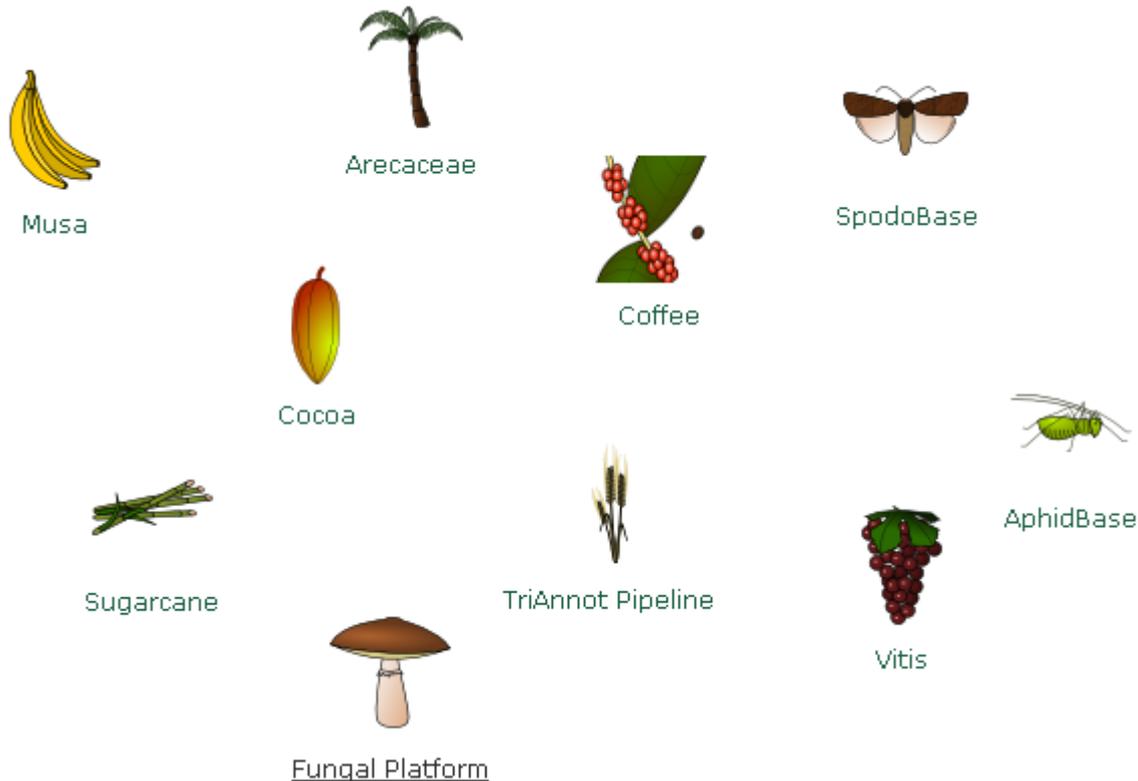


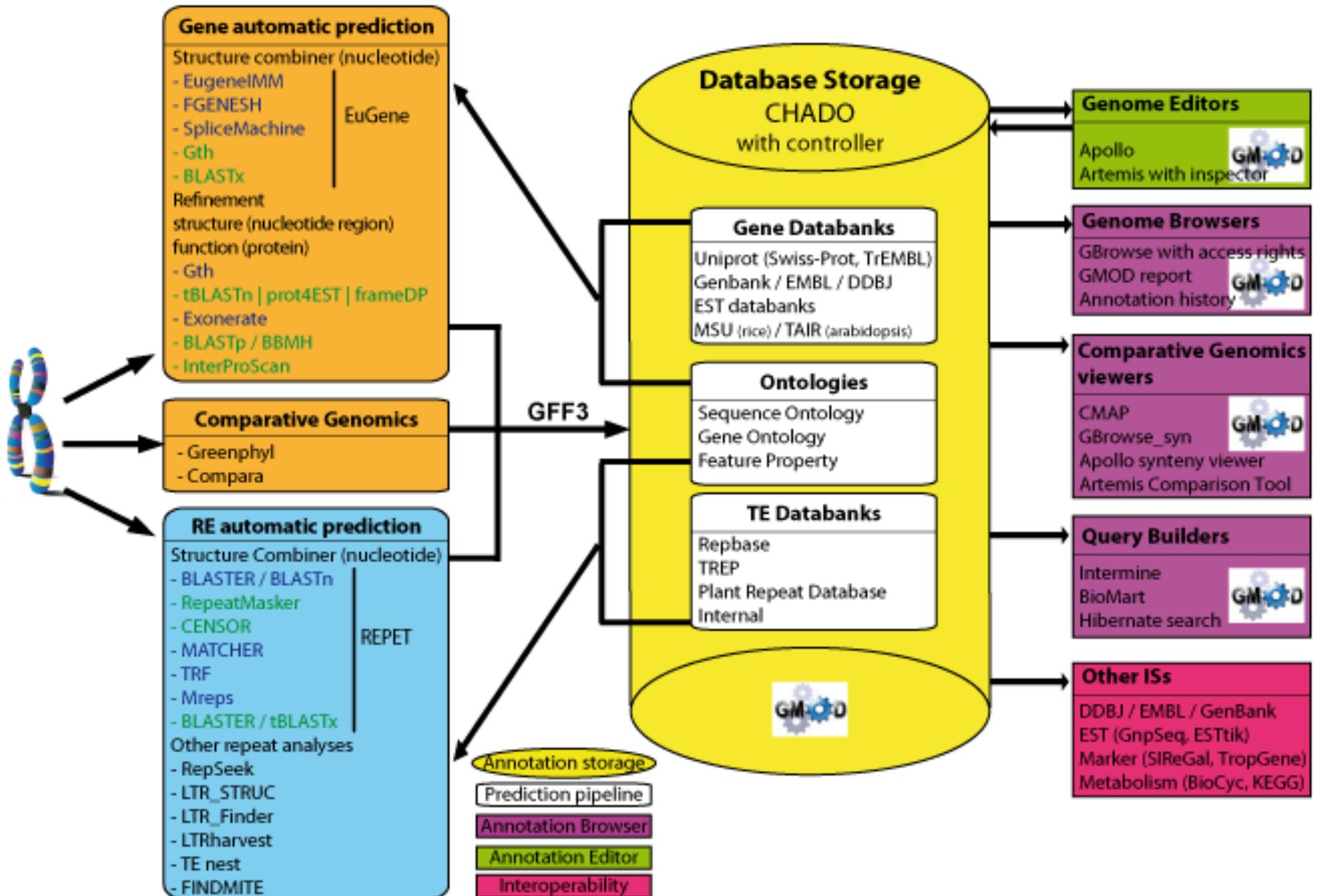
Medicago truncatula :

Eugène graphical output



GNPAnnot is a community system for structural and functional annotation dedicated to plants, insects and fungus genomes allowing both automatic predictions and manual curations of genomic objects.





Gene annotation Pipeline

Genomic sequence



1

Launch intrinsic + extrinsic modules

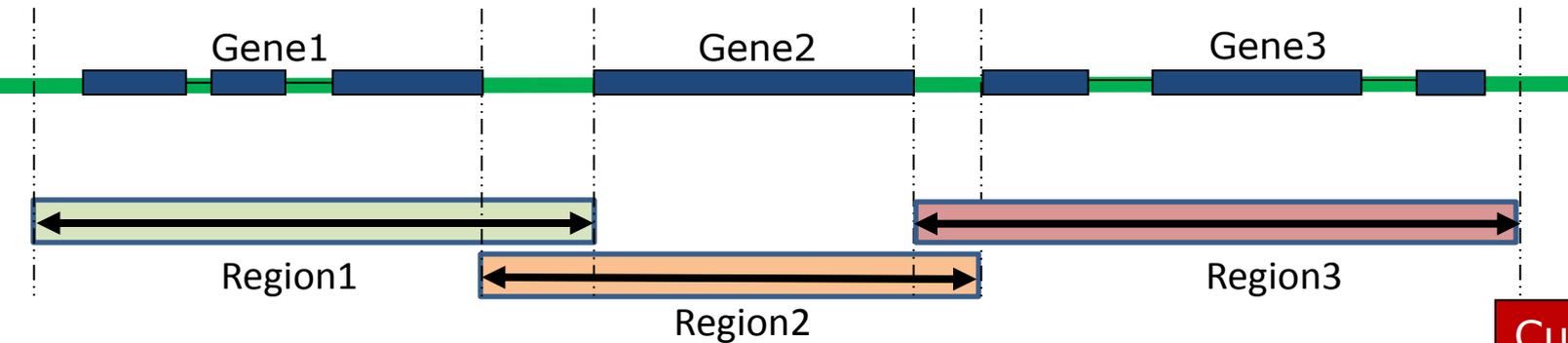


2

Integration in Eugene combiner



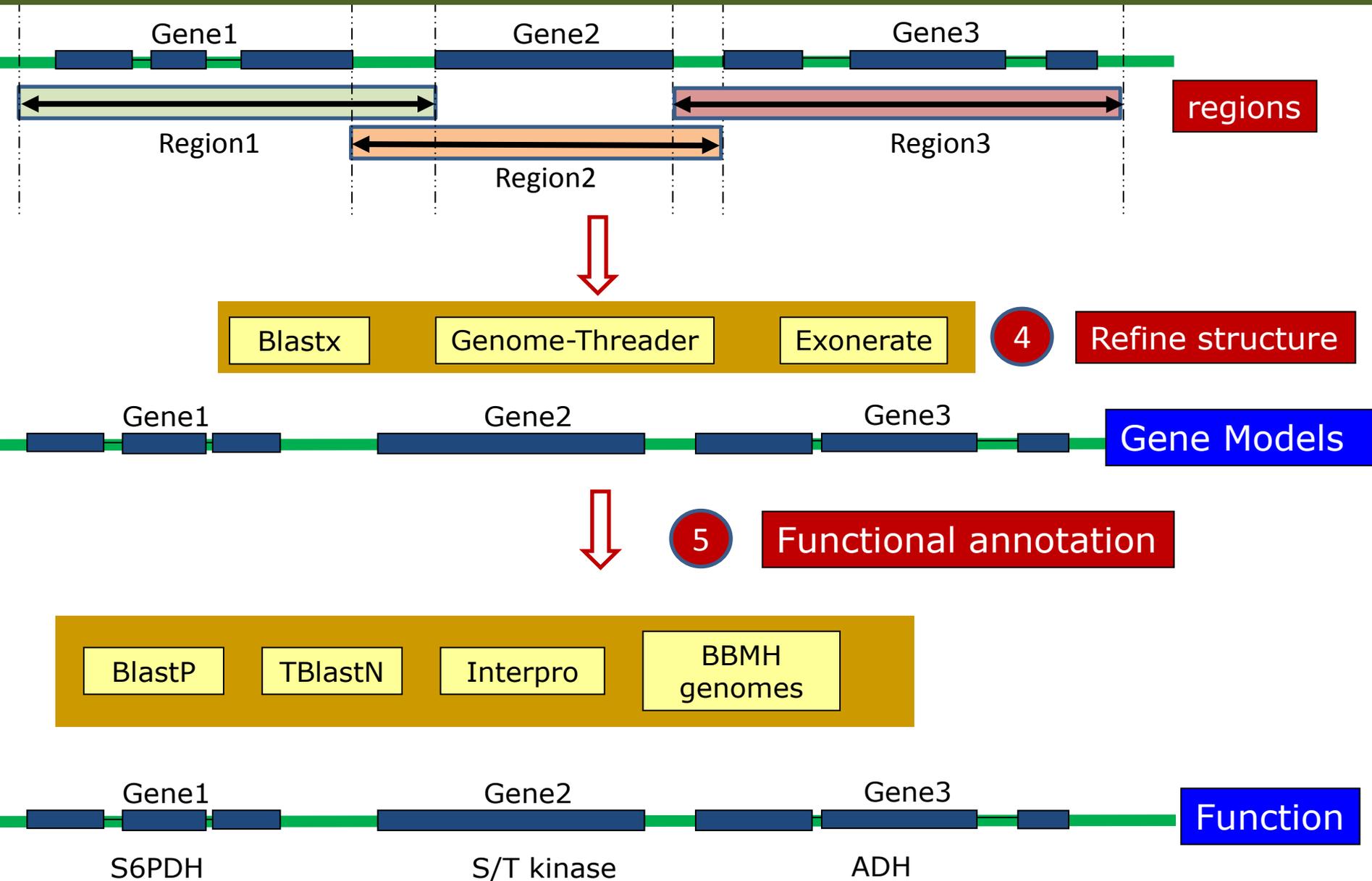
Structure prediction



3

Cutting regions

Annotation Pipeline



Evaluation of the relevance of annotations

VP (true positives) = predicted genes that are really presents = good prediction

FP (False positives) = predicted genes that are normally absents = over prediction

FN (False negatives) = genes that are missed by the automatic annotation = under prediction

VN (True negatives) = genes not predicted that are really absents = good prediction

The relevance is estimated with the ratio Sensibility/spécificity: **Sn/Sp**

The calculation is performed with a set of manually annotated genes

$$\text{Sensibility} = \frac{VP}{VP + FN}$$

$$\text{Spécificité} = \frac{VN}{VN + FP}$$

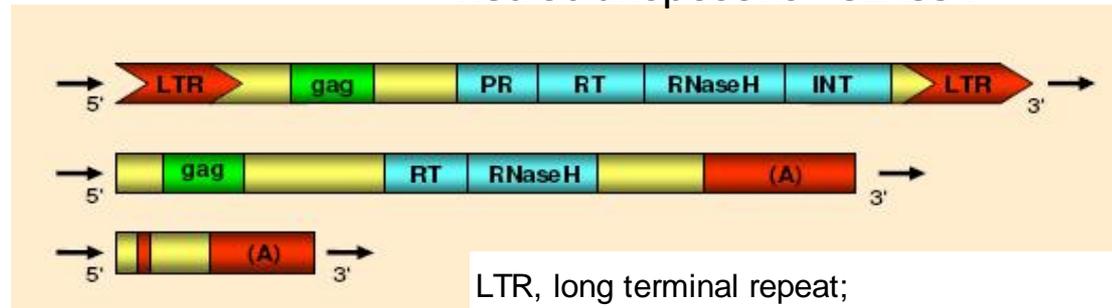
Gene finder	Sn N	Sp N	Sn E	Sp E	Sn G	Sp G
GenScan+	83.2	98.2	69.6	78	25.8	29
GenMarkHMM	89.9	94.8	73.1	76.6	32.4	31.6
FgenesH-At	95.1	93	85.3	81.4	47	46.5
FgenesH-Mt	97.6	92.1	85.1	80.7	52.8	47.8
EGN	93.7	95	84.7	85.4	55.5	50.5
EGN+FgenesH	97.8	94.2	90	86.9	63.2	56.4
EGN+FH+AA	98.6	93.9	92.4	88	69.2	61.8
EGN+FH+AA+ EST	98.2	99.9	94.4	94.6	80.2	79.4

Transposable Elements (TEs)

Genes represent only a little part of the genome. Some regions can be gene-rich but some other can contain a majority of repeated elements.

Retrotransposons CLASS I

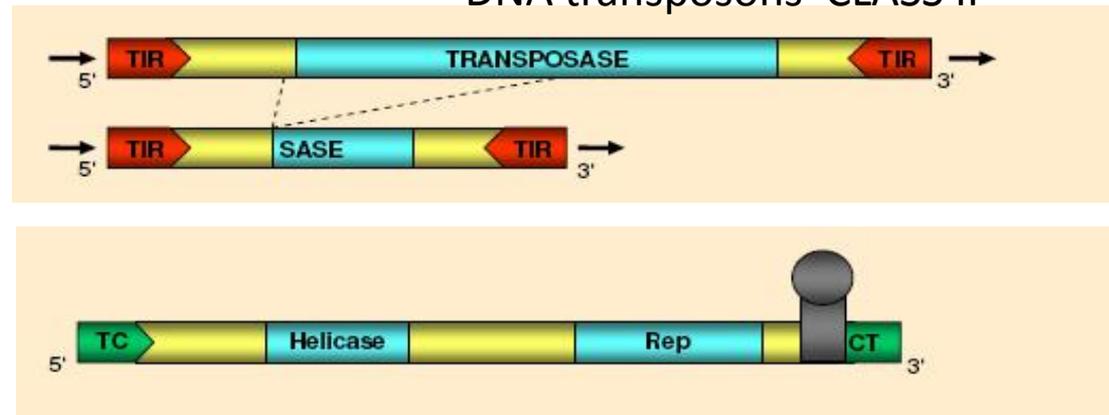
LTR
Non-LTR (LINEs)
SINEs



LTR, long terminal repeat;
LINE, long interspersed nuclear element;
SINE, short interspersed nuclear element;

DNA transposons CLASS II

TIR
MITEs
Helitrons



TIR, terminal inverted repeat.
MITE, Miniature Inverted Transposable Element

Transposable Elements

Classification from Wicker et al (2007). A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*, 8, 973-982.

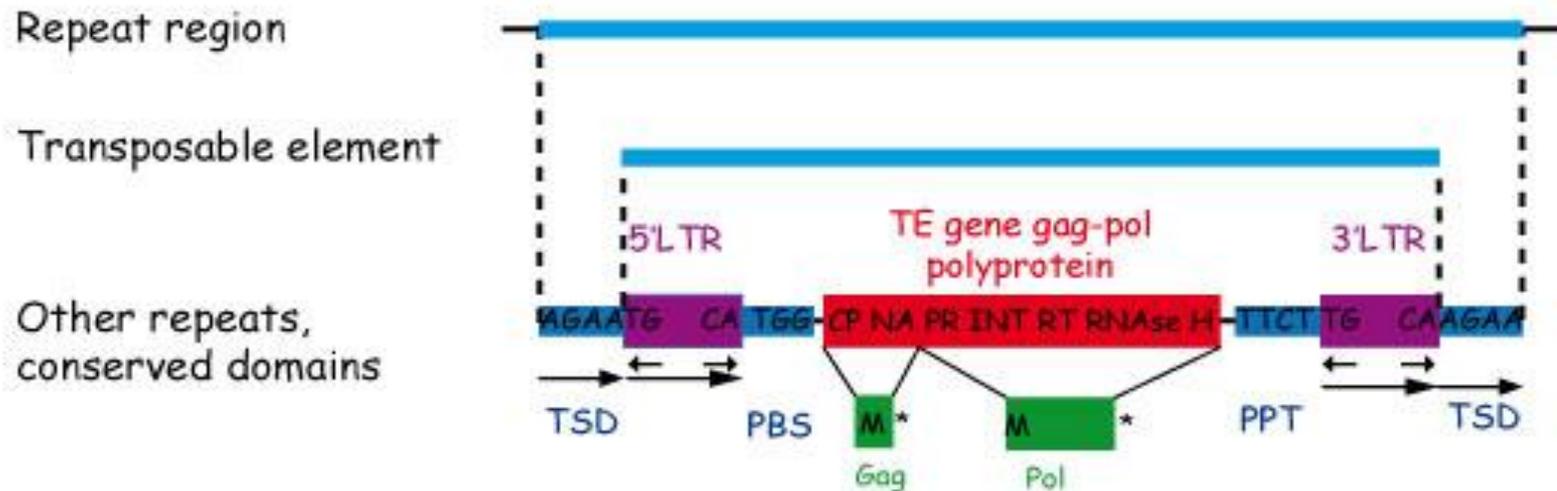
Class	Order	Superfamily	Family	Code / Label
Retrotransposon	LTR	Copia	opie	RLC
		Gypsy	maggy	RLG
		Unclassified		RLX
	LINE	L1		RIL
		Unclassified		RIX
	SINE	Alu		RSA
Unclassified			RSX	
DNA transposons	TIR	CACTA		DTC
		Mutator		DTM
	MITE	Stowaway		DTT
		Tourist		DTH
	Helitron	Helitron		DHH

Retrotransposon, LTR, Copia, opie = RLC_opie

Classification of TEs based on sequence similarities : the rule 80-80-80

LTR-Retrotransposons, Ty1 (*copia*) & Ty3 (*gypsy*)

In plants, the most frequent mobile elements are LTR retrotransposons
2 super families of LTR-retrotransposons : *Copia* and *Gypsy*

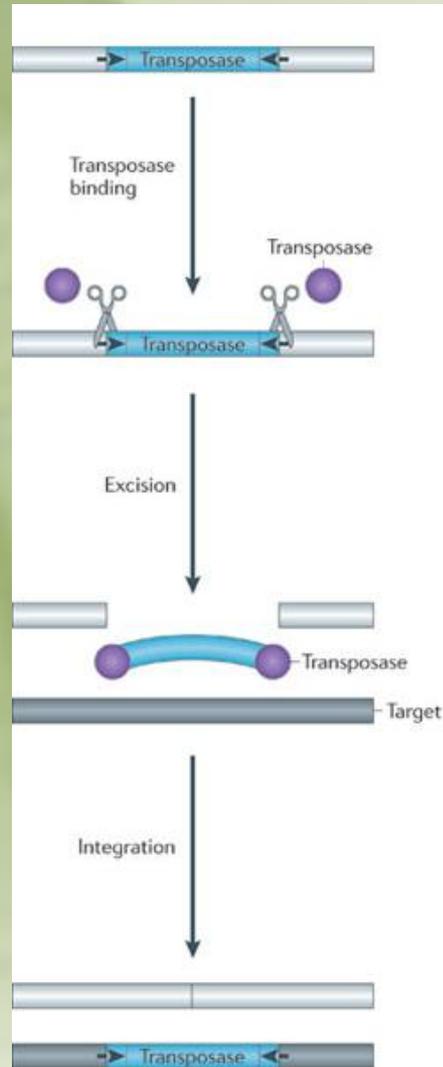


Example of LTR retrotransposon Ty1-copia

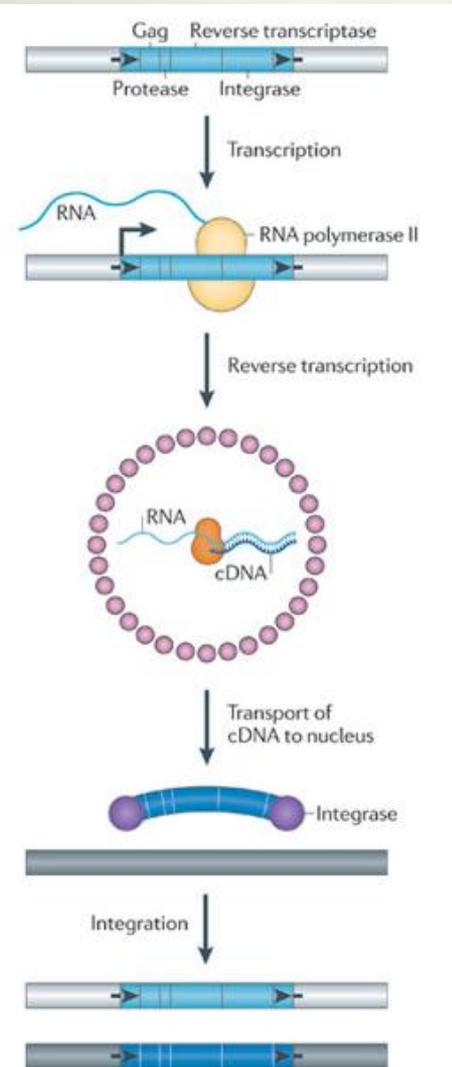
In Ty3-gypsy superfamily, the order of the protein domain differ: CP-NA-PR-RT-RNAseH-INT
 LTR, long terminal repeat; TSD, target site duplication; PBS, primer binding site;
 PPT, polypurine tract (RR tract); CP, capsid-like protein; NA, nucleic acid binding moiety;
 PR, protease; INT, integrase; RT, reverse transcriptase.

Dynamics of TEs

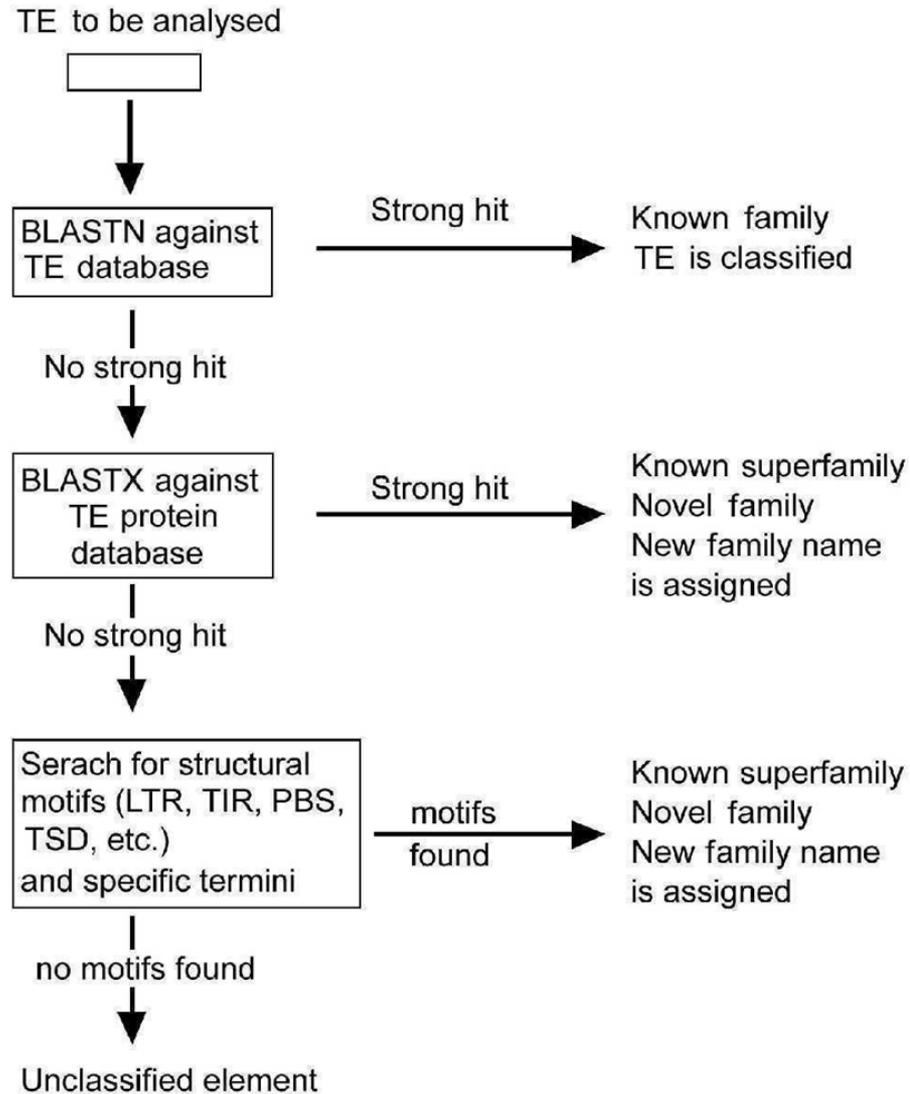
DNA transposons



LTR retrotransposons

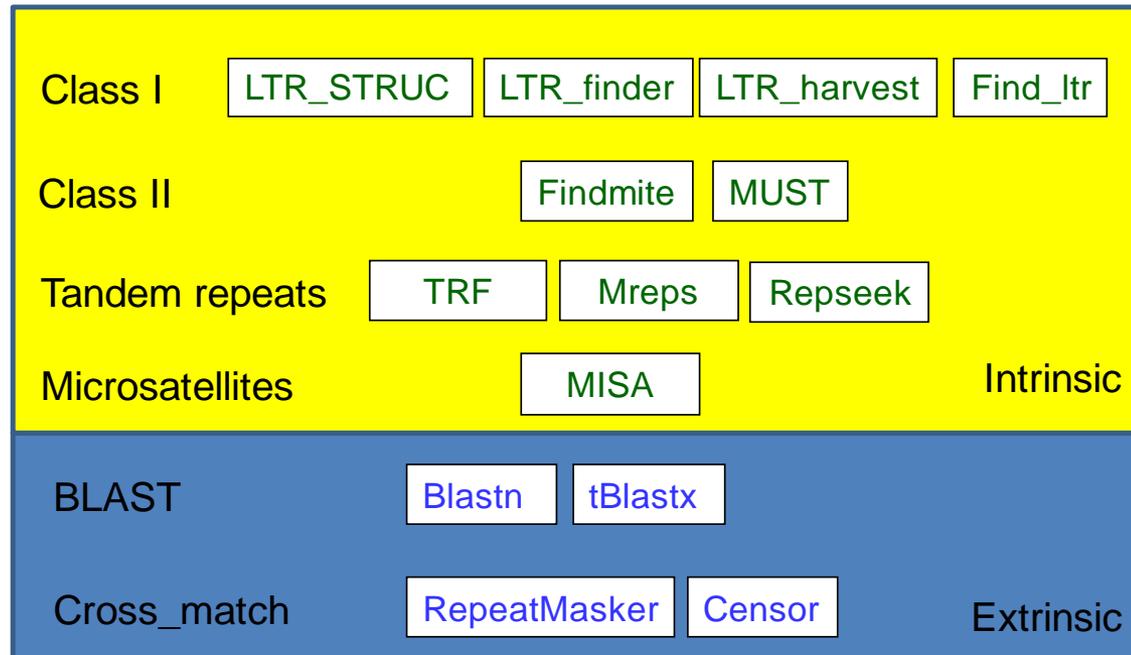


Annotation of TE, general approach

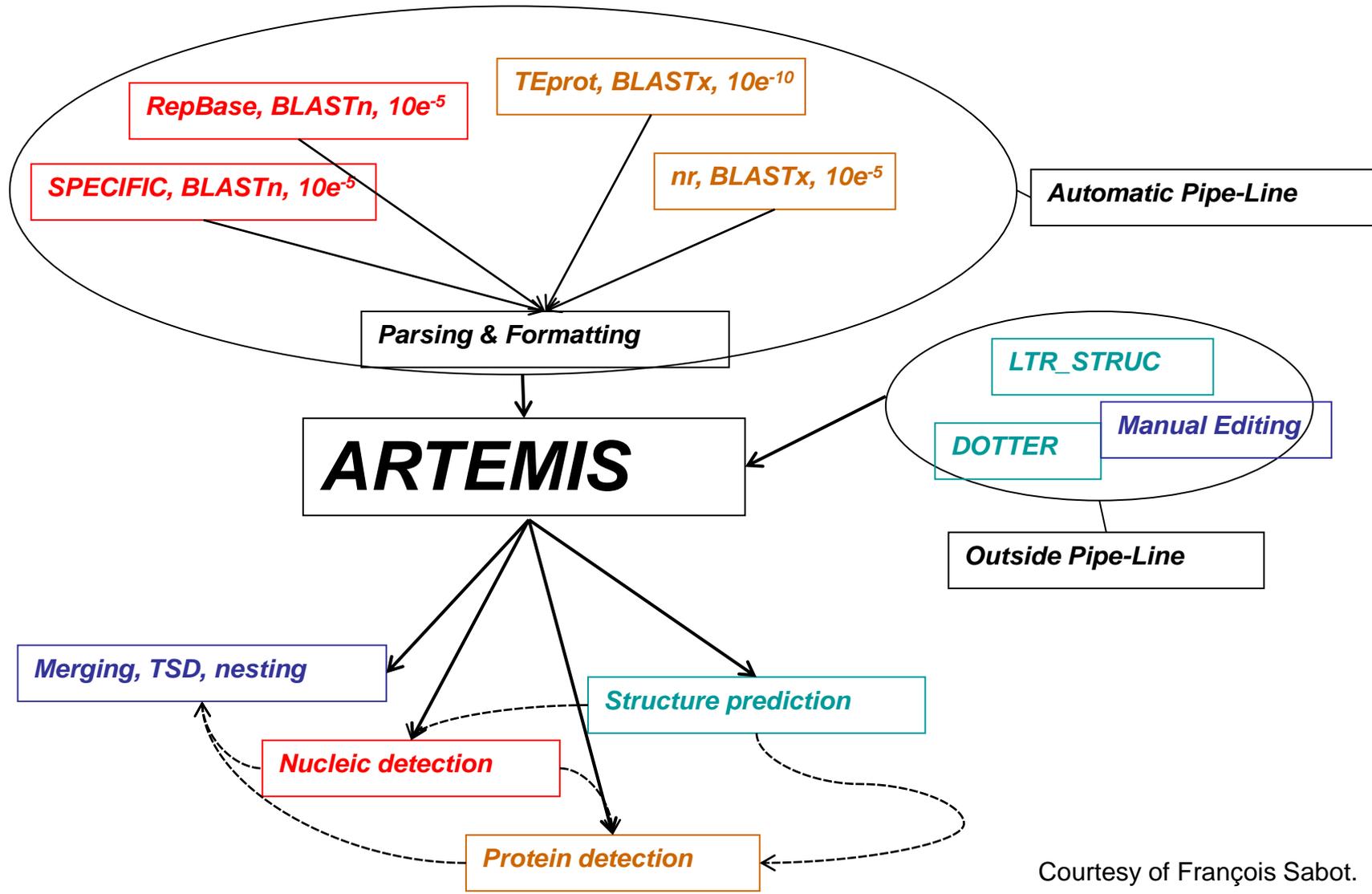


Annotation of transposable elements : tools

Several ab-initio programs can be used to detect structures of TEs
Comparisons with databases can be used to classify the elements.



Annotation of transposable elements : methods



Refine the structure of TE search for LTR and IR

DOT PLOT

<http://www.helmholtz-muenchen.de/icb/gepard>

Gepard

Gepard (German: "cheetah", Backronym for "GEnome PAir - Rapid Dotter") allows the calculation of dotplots even for large sequences like chromosomes or bacterial genomes.

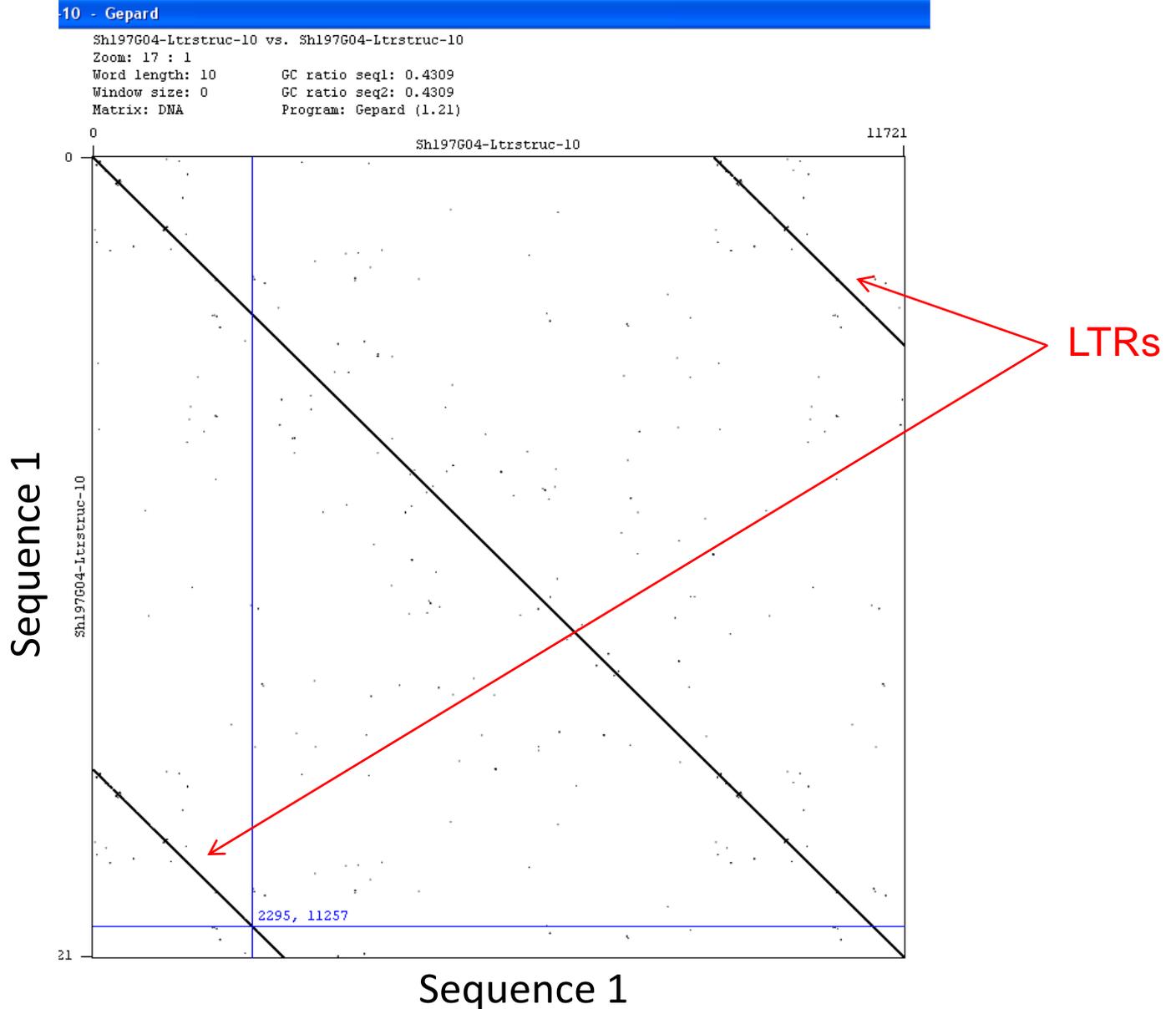


Reference:
Krumstiek J, Arnold R, Rattei T. Gepard: A rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 2007; 23(8): 1026-8. PMID: 17309896

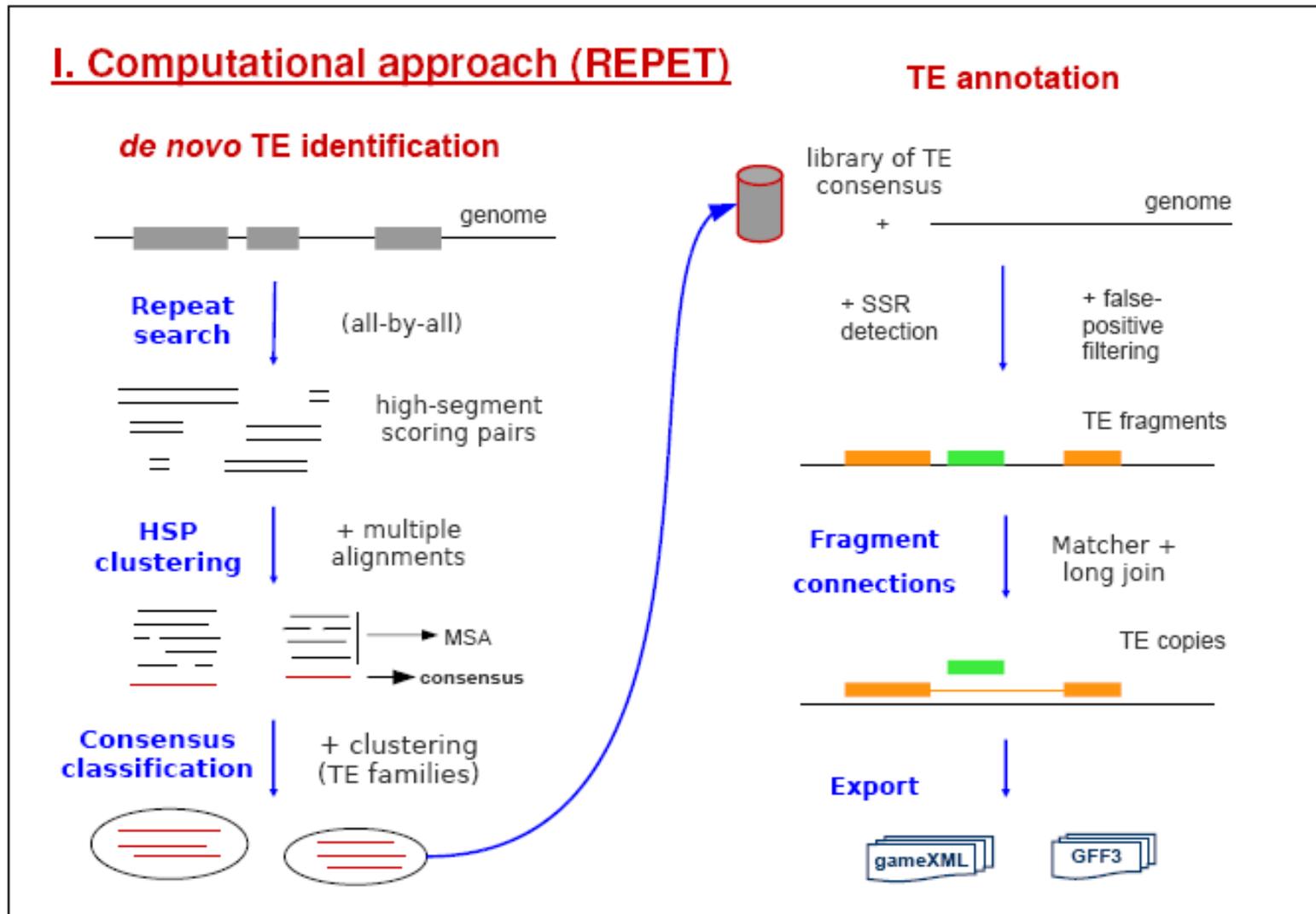
- Use cases +
- Features +
- Screenshot +
- System requirements +
- Download +
- Bugs +
- Source code +
- Tutorial +
- Method +
- Memory issues / Vmatch support +
- Contact +

Refine the structure of TE, e.g. LTR-retrotransposon

Gepard dotter



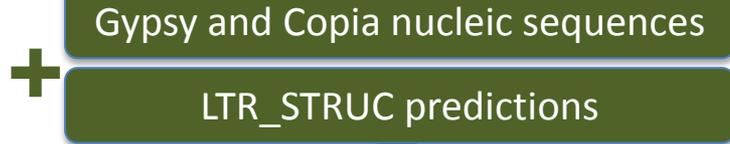
Pipeline for Classification, Screening and Mapping of TEs



80 – 80 rule = 80% identity on 80% of the length

Classification, screening and mapping of TEs onto the genome

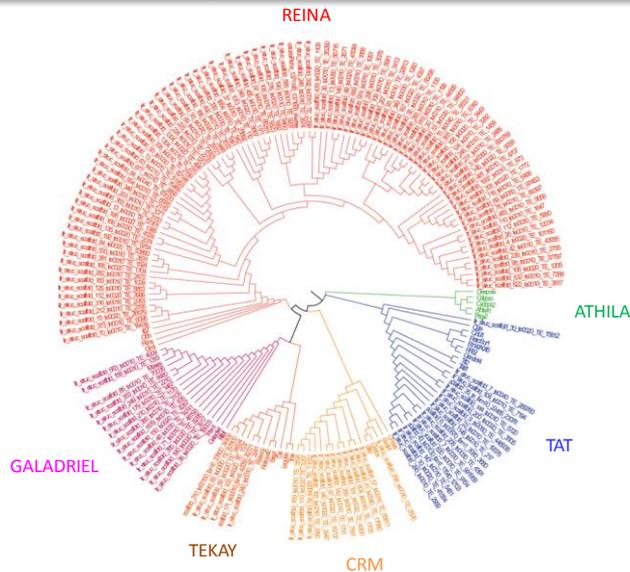
STEP 1 : continued



Clustering
80% id
80%
coverage

TGICL
Tclust
CAP3

Classification phylogenetic analysis of RT domains



Gypsy and Copia => TE reference library

STEP 2: Screening with REPET (TEannot)

TE reference library
- LTR retrotransposons
- Non-LTR retrotransposons
- DNA transposons

Screening
of the assembly

REPET
(TEannot)

Priority rule

DTA > DTX >
RLC > RLG > RLX
> RPP > RIL >
RIX > DHH

NESTED
OVERLAPPING TEs

TE Mapping

Before gene annotation, mask repeat sequences

Gene modelling is generally done after filtering out repeats because this makes gene finding even more difficult :

TEs may look like protein encoding genes (transposase and remains)

TEs perturb the structure of gene models, e.g. by inserting in introns (long insertion, and gene model is cut in many parts)

→ Repeat masker is the tool the mostly used to mask repeats.

 RepeatMasker Web Server

[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.

Reference: A.F.A. Smit, R. Hubley & P. Green, unpublished data. Current Version: open-3.2.9 (RMLib: 20090604)

[Check Current Queue Status](#)

Basic Options

or

Sequence:

Search Engine: abblast rmbblast cross_match

Speed/Sensitivity: rush quick default slow

DNA source:

Return Format: html tar file

Return Method: html email

Terminé

Select a sequence file to process or paste the sequence(s) in [FASTA format](#). [Large sequences](#) will be queued, and may take a while to process.

Select the search engine to use when searching the sequence. *Cross_match* is slower but often more sensitive than the other engines. *ABlast* (formally known as *WUblast*) is very fast with a slight cost of sensitivity. *RMBlast* is a RepeatMasker compatible version of the *NCBI Blast* tool suite.

Select the sensitivity of your search. The more sensitive the longer the processing time.

Select a species from the drop down box or select "Other.." and enter a species name in the text box. Try the [protein based repeatmasker](#) if the repeat database for your species is small.

Select the format for the results of your search. The "tar" option will return the results as a compressed archive file, and "html" will present the results as a summary web page with links to the individual data files.

The "HTML" return method will run RepeatMasker on your sequence and return the results immediately to your web browser, provided your sequences are short. The "email" return method will email you when your results are ready.

The background of the slide features a soft-focus photograph of green leaves and a white flower. The leaves are in various shades of green, with some showing prominent veins. The white flower is partially visible in the upper right corner. A dark green horizontal band is overlaid across the middle of the image, containing the text.

Manual curation of annotations

artemis : to see, to check, to correct, to validate, to store

Artemis Entry Edit: AM403006.update.embl

File Entries Select View Goto Edit Create Run Graph Display

Entry: AM403006.update.embl

Nothing selected

repeat_region repeat_region repeat_region

0 52000 52800 53600 54400 55200 56000 56800 57600 58400 59200 60000 60800 61600 62400 63200 64000

Sh51L01g_70

Sh51L01g_60

R N H S A R R T A C Q S I C D S T N H F R G N + D S V I A L P L H I C E T N I V L I L K F N N H R Q G S H L Q Q K E W K C L N V L E R K
E T I Q Q G E L P A K A S A I A P T I L E V T R T Q # L H C L Y I Y V K Q I L F # Y # N S I T T D R E V I Y S K R N G S A * T C W K G
. K P F S K E N C L P K H L R + H Q P F + R # L G L S N C I A S T Y M * N K Y C F N I K I Q # P Q T G K S F T A K G M E V L E R V G K E
CGAAACCATTACGCAAGGAGAACTGCCTGCCAAAGCATCTGCGATAGCACCACCAACCATTTAGAGGTAAGTACTAGGACTCAGTAATTGCATTGCCTACATATATGTGAAACAATAATTGTTTAAATATTAAAATCAATAACCACAGAGGGAAGTCATTTACAGCAAAAAGGAATGGAAGTGCCTTGAACGTGTTGAAAAGGAA
20 40 60 80 100 120 140 160 180 200
GCTTTGGTAAGTCGTTCTCTTCGACGGACGGTTTCGTAGACGCTATCGTGGTTGGTAAAATCTCCATTGATCCTGAGTCATTAACTAACGGAGATGTATATACACTTTGTTTATAACAAAATATAATTTTAAGTTATTGGTGTCTGTCCCTTCAGTAAATGTGTTTTCCTTACCTCAGCAACTTGCACAACCTTTCCTT
S V M * C P S S G A L A D A I A G V M K S T V L V * Y N C Q R + M Y T F C I N N # Y # F E I V V S L S T M # L L L F P L A Q V H Q F P F
F G N L L S F Q R G F C R R Y C W G N # L Y S P S L L Q M A E V Y I H F L Y Q K L I L I * Y G C V P P F D N V A F P I S T S S R T P F S
. F W E A L L V A Q W L M Q S L V L W K L P L + S E T I A N G R C I H S V F I T K I N F N L L W L C P L * K C C F S H F H K F T N S L F

source 1 97616
CDS 2 1713 protein disrupted by cloning site
repeat_region 1725 7141 c LINE
CDS 9888 12765
repeat_region 10638 10871 LINE
CDS 13165 13576
repeat_region 14188 14508 MITE: tourist
repeat_region 15650 20602 c LINE
repeat_region 20629 20825 MITE: tourist
CDS 24023 34867 c
repeat_region 30261 30427 MITE
repeat_region 32656 32923 MITE: tourist
repeat_region 38814 43729 retrotransposon: solo LTR, levithan
repeat_region 46640 46920 MITE: tourist
CDS 47410 50653

démarrer Microsoft P... Saccharum ... Artemis Rel... File Manager Artemis Ent... 08:51

GNPAnnot on the web : <http://southgreen.cirad.fr>

The image shows the homepage of the South Green bioinformatics platform. At the top, there is a navigation bar with links for 'ABOUT US', 'SITE MAP', and 'CONTACT'. Below this is the main header with the 'South Green bioinformatics platform' logo. A secondary navigation bar contains links for 'Home', 'Tools', 'Databases', 'Developer tools', 'Species', 'Projects', 'Teaching', and 'Platform', along with a search box. The main content area features a grid of tool logos: Galaxy, SNIPlay, GreenPhyl 3.0, The Banana Genome Hub, TropGENE, OryGenesDB, Oryza Tag Line, EURGEN, GNPAnnot (circled in red), EST TDI, HaploPhyle, and GnDIVERSITY. A 'CocoaGen DB' logo is also present. On the right side, there is a 'NEWS' section with three entries: 'Welcome to our new website' (02/10/2013), 'Publication of SNIploid' (04/10/2013), and 'Training course in Bioinformatics applied to Musa genome, 18-22 November 2013' (17/11/2013). At the bottom, there are sections for 'Partners' (including cirad, Bioersity, INRA, IRD, SupAgro), 'Funding partners' (including arcad), and 'Networks' (including IBISA, ATGC, and ReNaBi).

South Green is a bioinformatics platform applied to the genomic resource analysis of southern and Mediterranean plants. ([read more](#))

NEWS

Welcome to our new website
02/10/2013

Publication of SNIploid
04/10/2013

Training course in Bioinformatics applied to Musa genome, 18-22 November 2013
17/11/2013

Annotation tool

Manual curation of annotations

Banana Genome Hub

HOME GENOME DETAILS TOOLS GBROWSE BLAST DOWNLOAD RESEQUENCING DOCUMENTATION

File Help

Musa acuminata Pahang: 50 kbp from chr1:9,245,587..9,295,586

Browser Select Tracks Snapshots Custom Tracks Preferences

Search

Landmark or Region:

chr1:9,245,587..9,295,586

Search

Design PCR primers

Configure...

Go

Save Snapshot

Load Snapshot

Examples: GSMUA_Achr1P12150_001, GSMUA_Achr5P02570_001

Data Source

Musa acuminata Pahang

Scroll/Zoom: << < - Show 50 kbp + > >> Flip

Overview

chr1

Region

Details

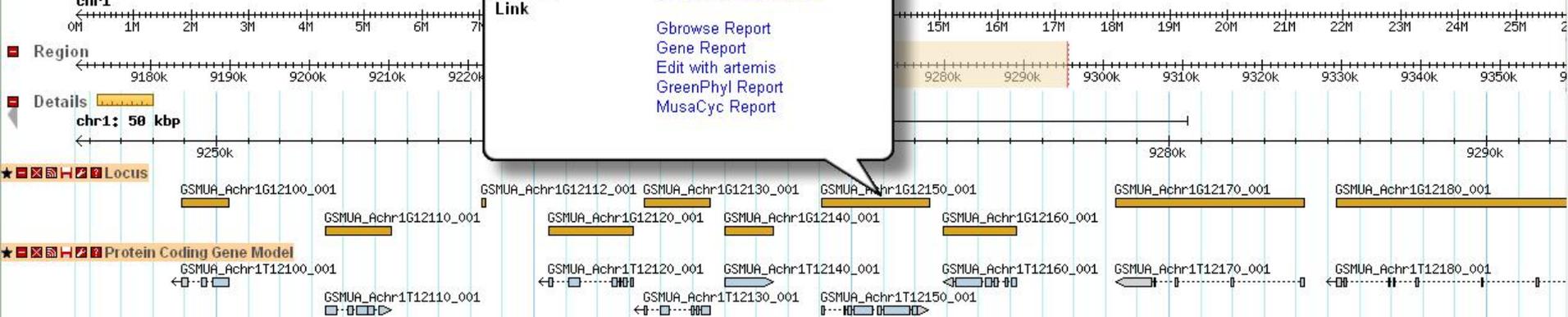
chr1: 50 kbp

Locus

Protein Coding Gene Model

Name GSMUA_Achr1G12150_001
Zoom in Link chr1:9269057..9272476

- Gbrowse Report
- Gene Report
- Edit with artemis
- GreenPhyl Report
- MusaCyc Report



Gbrowse is connected to artemis through the chado database

The screenshot shows a Mozilla Firefox browser window displaying the Gbrowse interface for the Saccharum hybrid cultivar. The browser address bar shows the URL: `gnpannot.cirad.fr/cgi-bin/gbrowse/sugarcane/#search`. The Gbrowse interface displays genomic tracks for contigs and genes, with a scale from 0k to 30k. A dialog box titled "Enter Database Address" is overlaid on the Gbrowse page, showing the following details:

- Server: medoc.cirad.fr
- Port: 5432
- Database: gnpannot_sugarcane
- User: ogarsmeur
- Password: [masked]
- Available databases: Default

Red text annotations point to these fields: "Database", "User", and "Password". Below the dialog box, the Artemis application window is visible, showing the Sanger Institute logo and the text "Artemis Release 13.2.8". A red arrow points to the Artemis window with the text "Artemis annotation tool".

The Artemis window displays the following information:

- File Options Windows
- wellcome trust sanger institute pathogen genomics group
- Artemis Release 13.2.8
- 1. Standard
- Connecting ...

The Artemis window also shows a list of tracks for manual and automatic annotation, including Eugene (cds), Eugene (gene), Eugene (mRNA), Eugene (polypeptide), and Repet Manual. The "Eugene (polypeptide)" track is checked.

Tutorials available on the portal

[e.g. tutoriel_gnpannotmonocot_gene](#)

The screenshot shows the GNPAnnot Portal website. The main content area is titled 'Monocots tutorials' and includes a search bar and a list of tutorials. The left sidebar contains a navigation menu with 'Monocots e-Learning Courses' circled in red. The bottom of the page features logos for CIRAD, INRA, and Bioversity International.

Monocots tutorials | GNPAnnot Portal - Mozilla Firefox

www.gnpannot.org/content/monocots-tutorials

Home > Documentation

Monocots tutorials **View** Revisions

Search this site:

- Home
- Documentation
 - Monocots e-Learning Courses**
 - Monocots Tutorials
 - Chado Controller
 - Publications
- Development
- Resources
 - ▷ Dicots Statistics
 - ▷ Monocots Statistics
 - Arecaceae Statistics
 - Musaceae Statistics
 - Musaceae Statistics (2)
 - Musaceae Statistics (3)
 - Poaceae Statistics
 - Sandbox
- Downloads
- Community

Online GNPAnnot Monocots tutorials:

1. Launch the CAS
2. Annotate a gene
3. Annotate a transposable element

Download GNPAnnot Monocots tutorials:

1. Launch the CAS
2. Annotate a gene
3. Annotate a transposable element

démarrer Microsoft PowerPoint ... Diaporama PowerPoin... Monocots tutorials | ... 08:20

ARTEMIS



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