

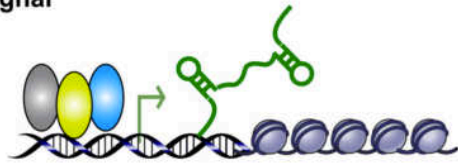
# Long non-coding RNA in the pea aphid; identification and comparative expression in sexual and asexual embryos

Fabrice Legeai, Thomas Derrien, Valentin Wucher, Audrey David, Gael Le Trionnaire and Denis Tagu

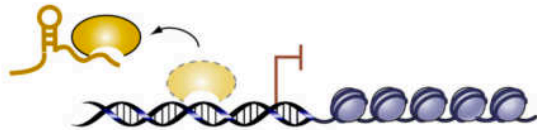
# Role of the long non coding RNAs

Cis

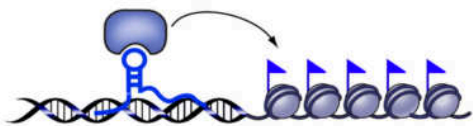
I. Signal



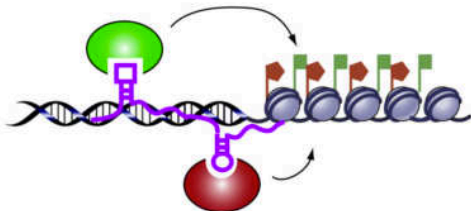
II. Decoy



III. Guide



IV. Scaffold

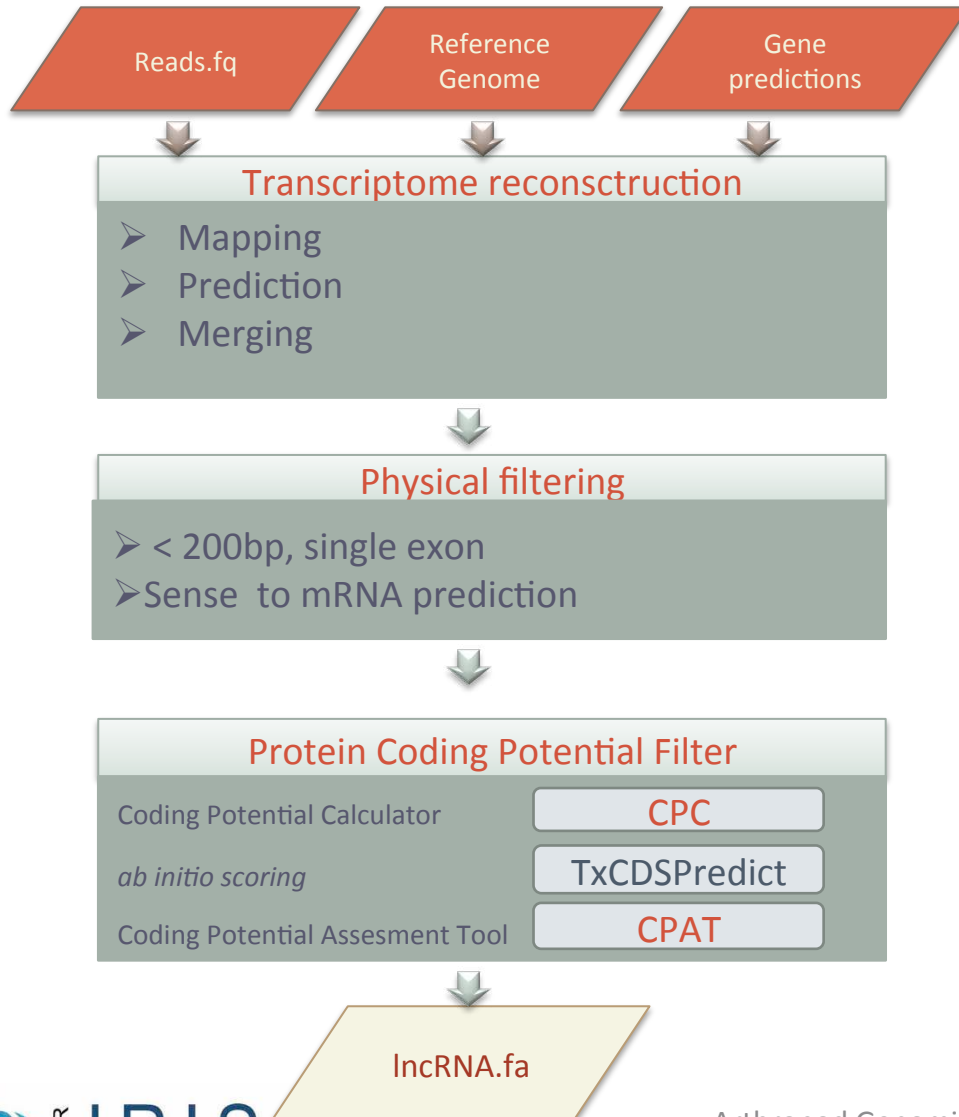


- **HOTAIR** (HOX antisense intergenic RNA) regulates the transcription of HOXD transcription factor by modifying the methylation of the chromatin.
- **lincRNA-Cox2** : regulates immunity genes
- **roX1** and **roX2** play a key role in the dosage compensation in dropsohila as they are involve in the transcriptional upregulation of genes on the single X chromosome in males to match the expression from the two X chromosomes
- **COOLAIR** antisense RNAs are upregulated in response to cold temperatures and inactives arabidopsis flowering genes

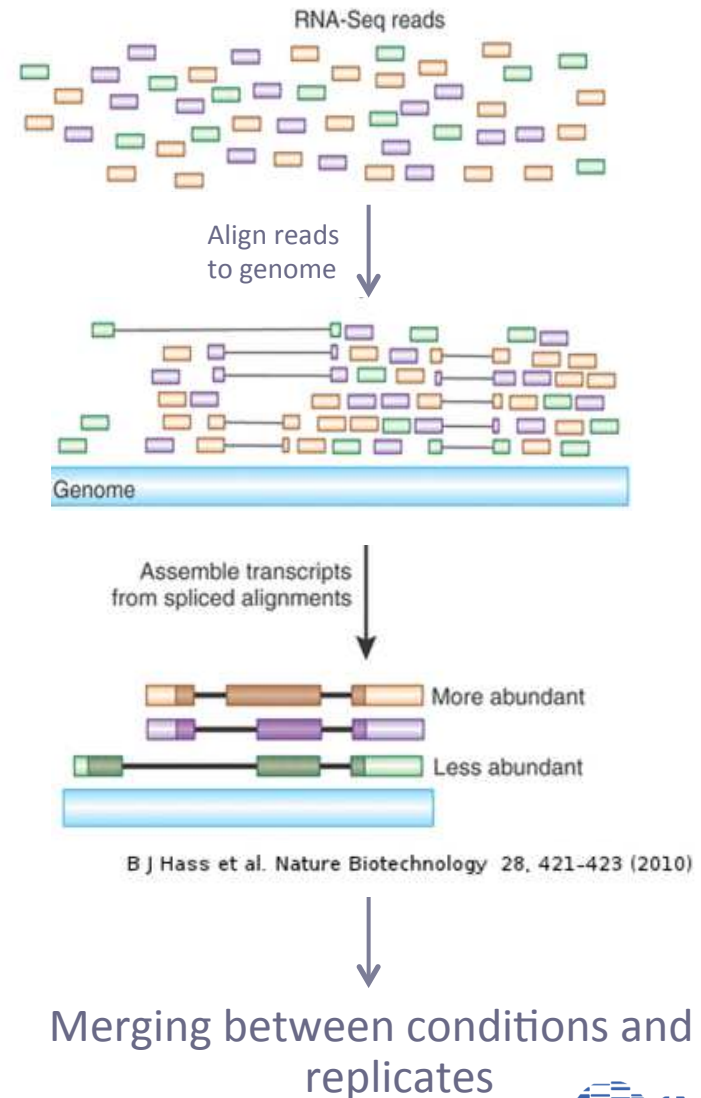
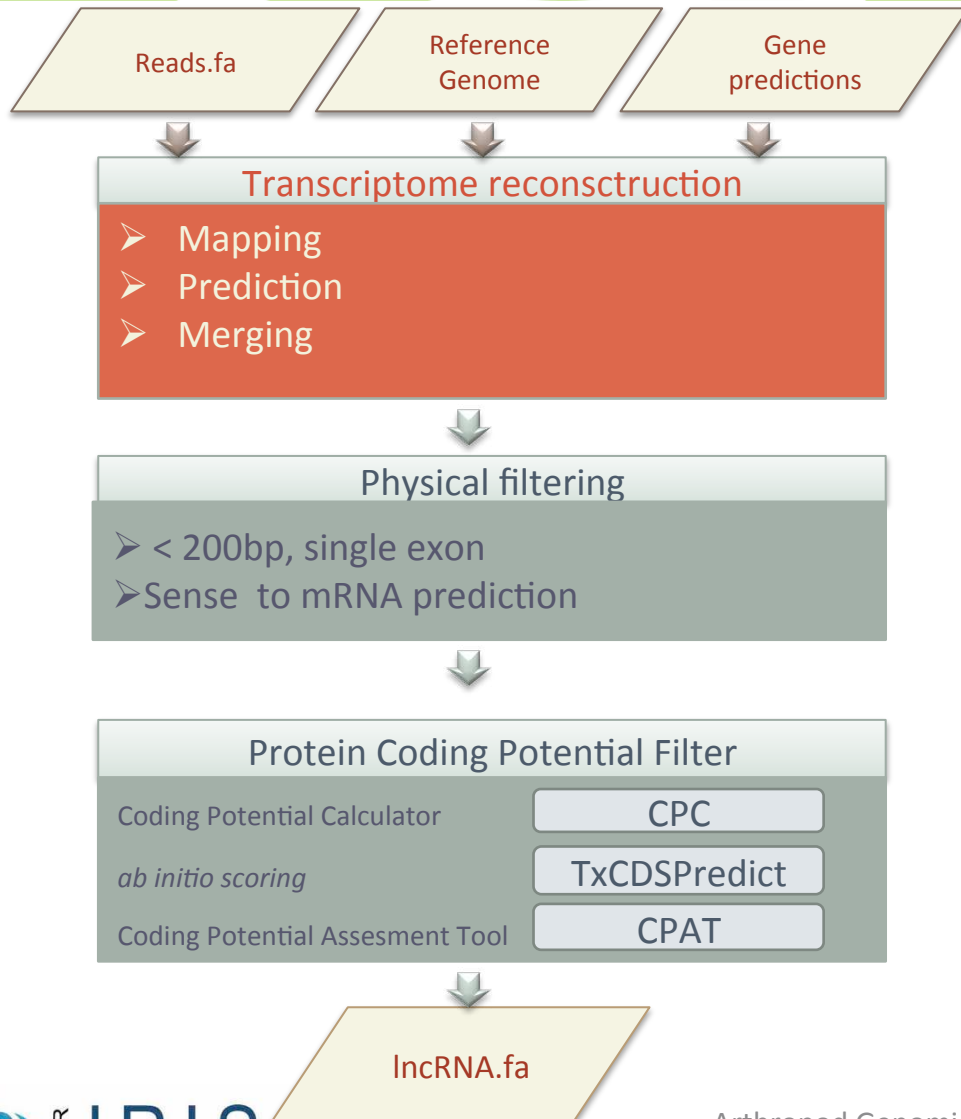
Trans

Wang K C, Chang, H Y Mol Cell 2011

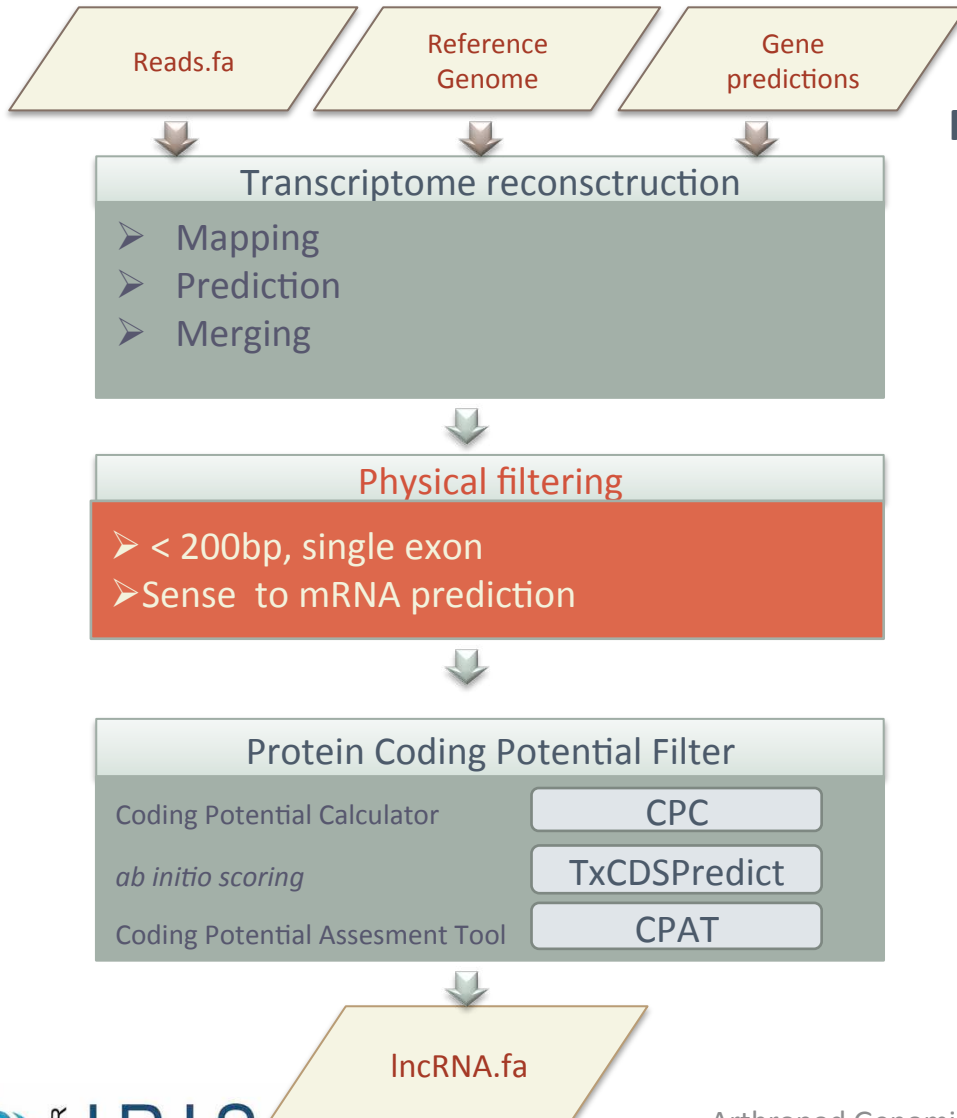
# lncRNA Prediction workflow



# IncRNA Prediction workflow



# lncRNA Prediction workflow



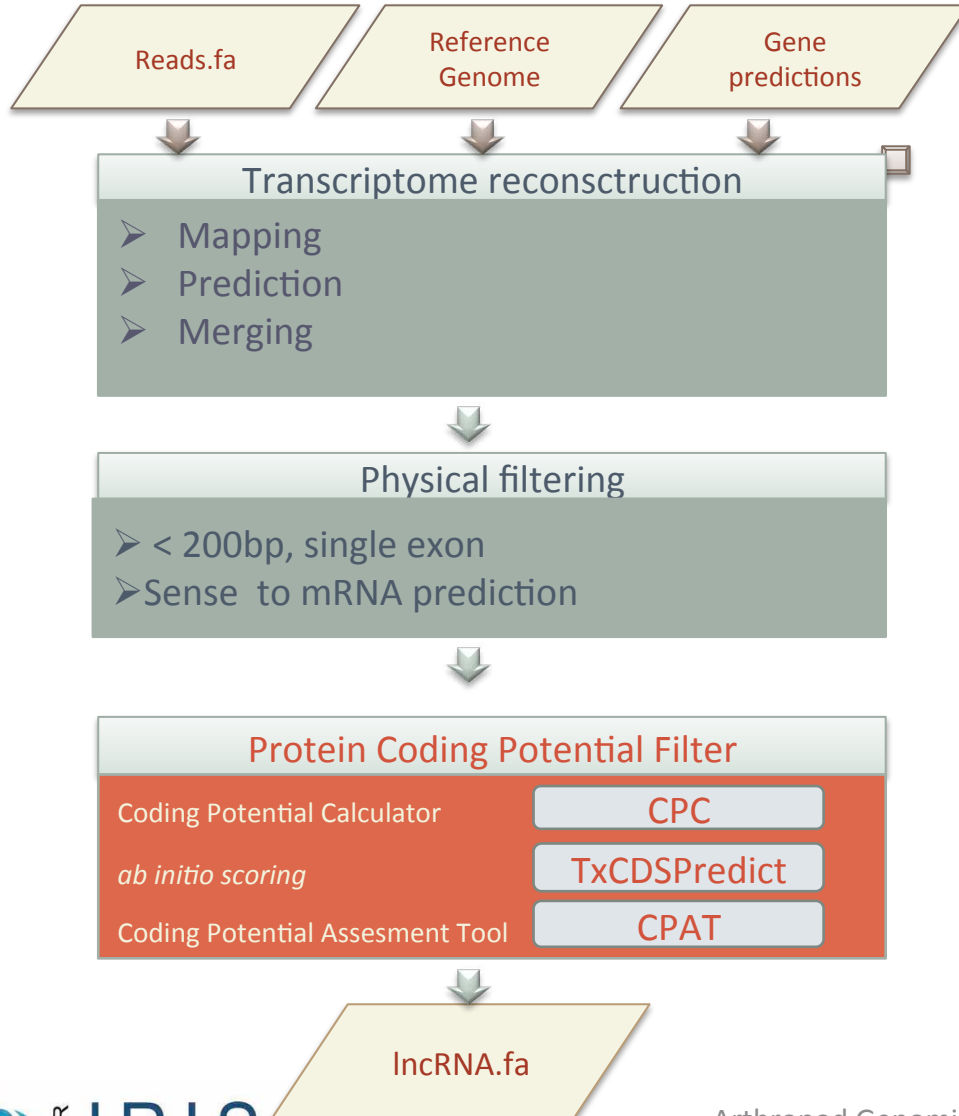
## Physical filtering

Everything that is overlapping a protein coding gene in same strand is not removed from the set.

Only RNA larger than 200 bp are taken into account.

Due to the tophat/cufflinks over-prediction of genes, only genes with more than one exon are considered

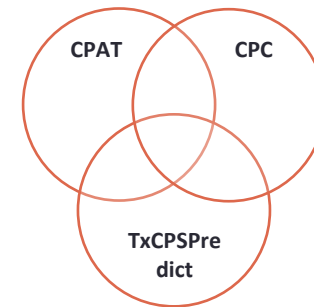
# lncRNA Prediction workflow



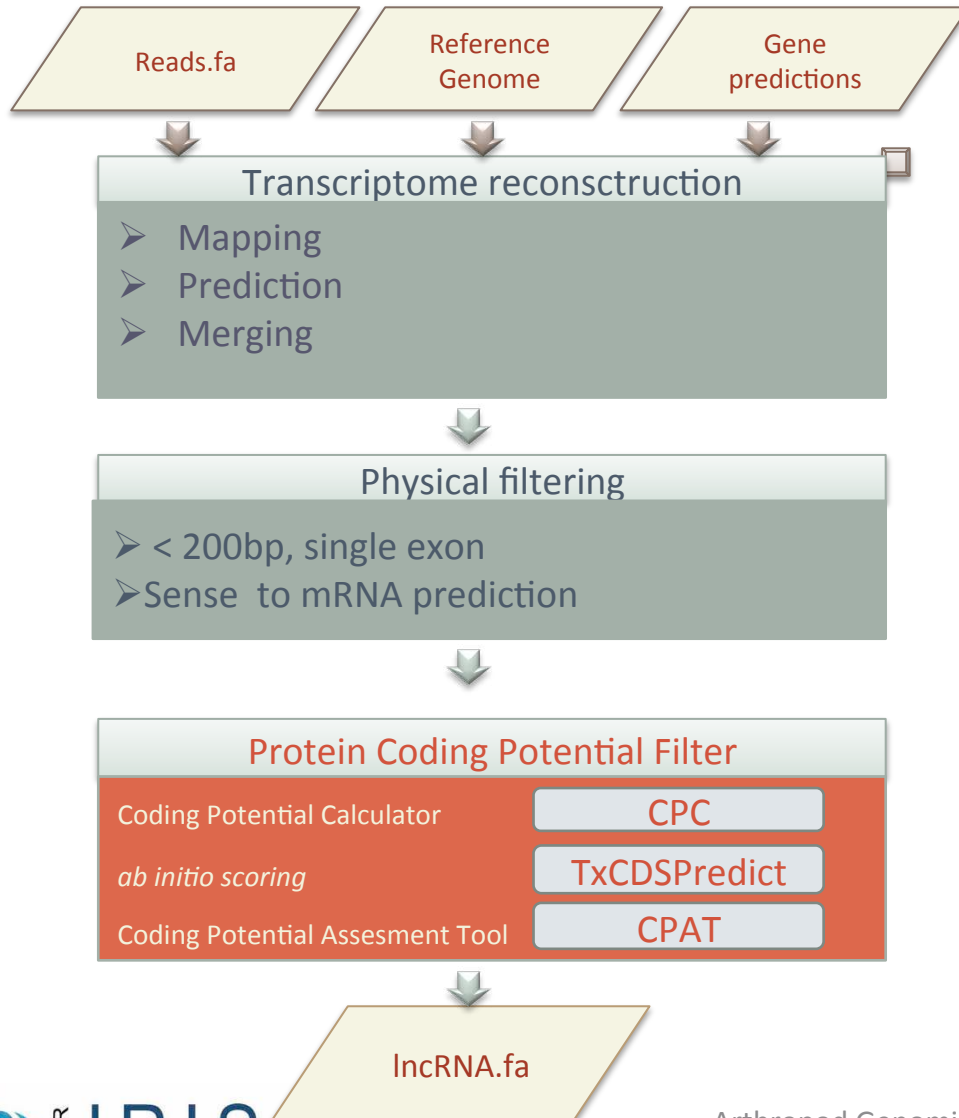
## Coding Potential Calculation

Probability that a RNA is coding for a protein

Intersection of 3 methods



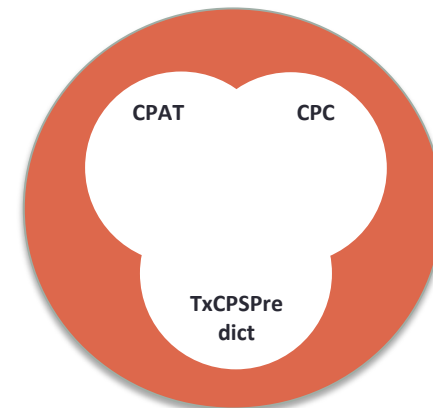
# lncRNA Prediction workflow



## Coding Potential Calculation

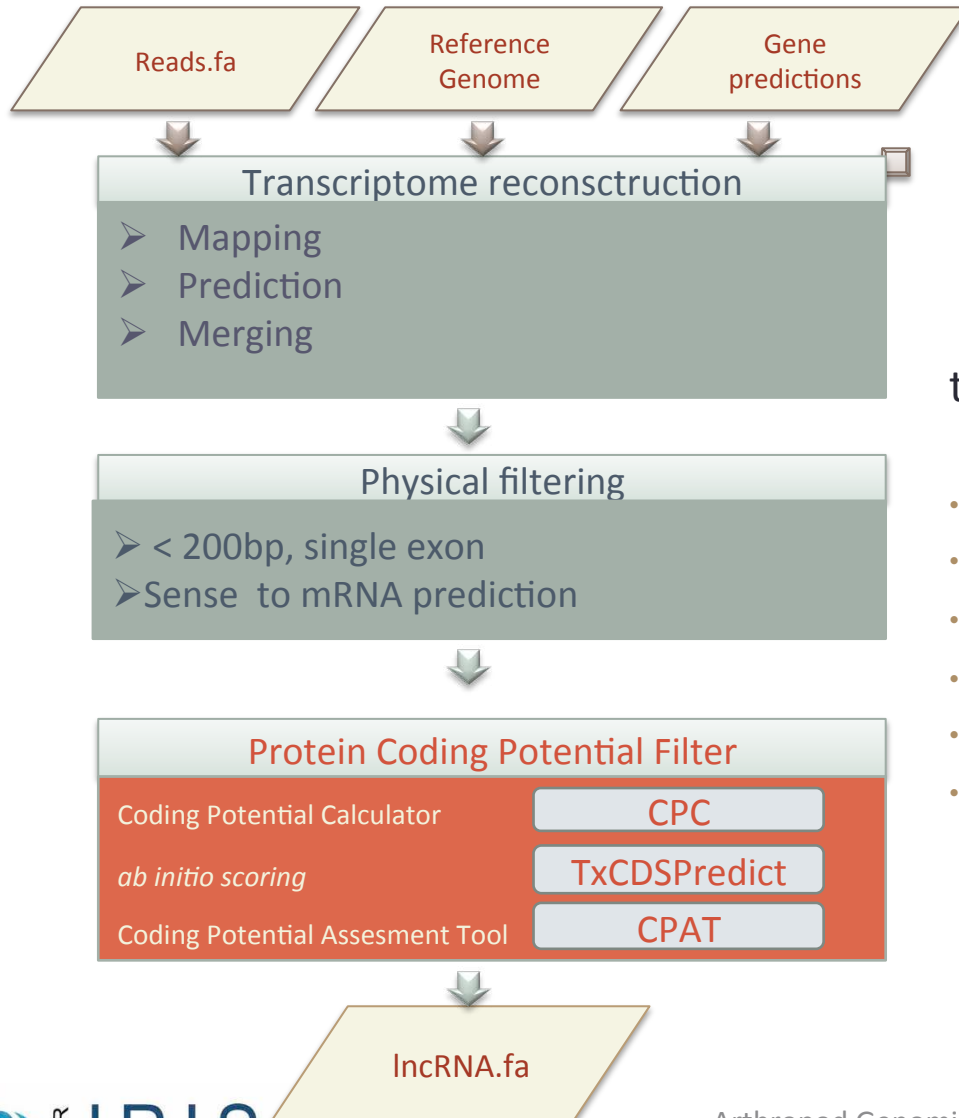
Probability that a RNA is coding for a protein

Intersection of 3 methods





# lncRNA Prediction workflow



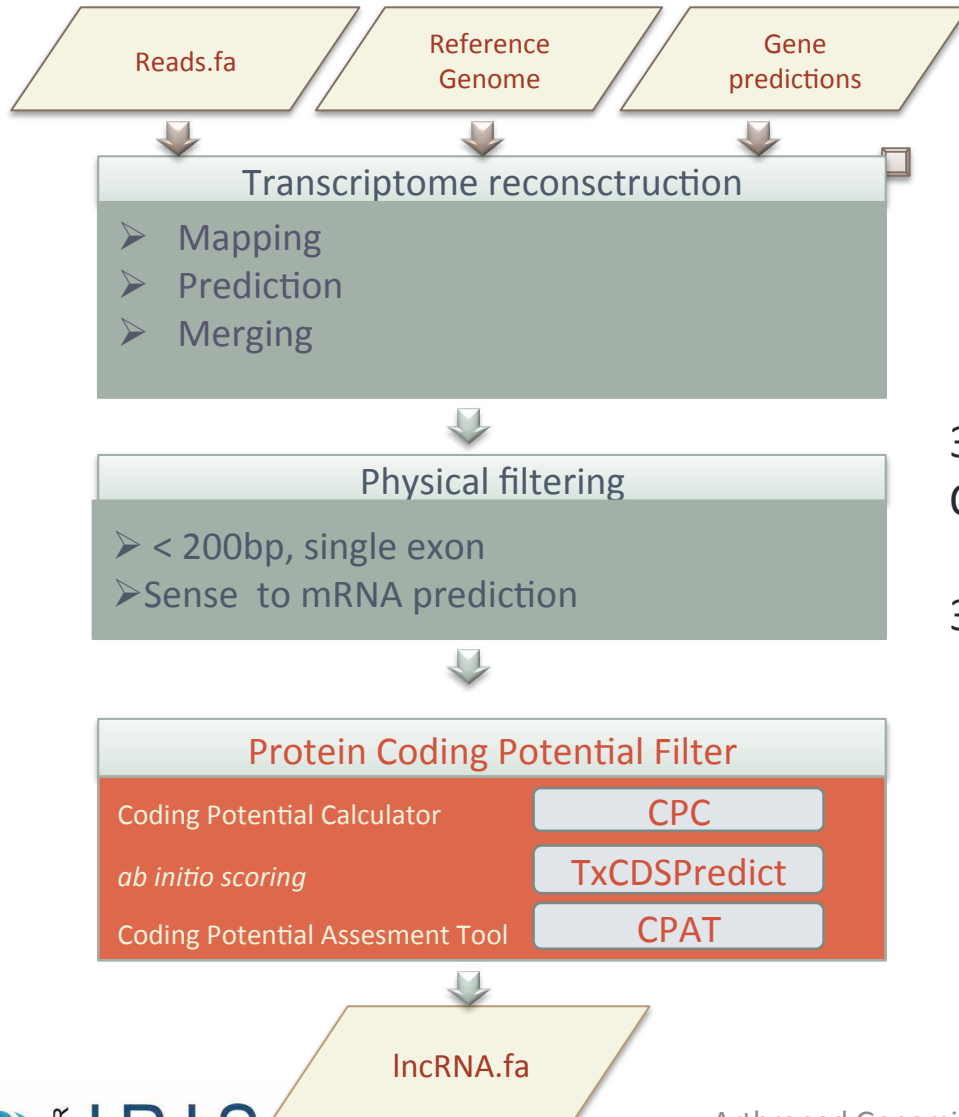
## Coding Potential Calculation

**TxCDSPredict** : Scoring and ranking of the RNA with criteria:

- ORF size
- Presence of an ATG
- Presence of a stop codon
- Number of upstream nucleotides
- Kozak sequence
- Stop in the last exon



# lncRNA Prediction workflow



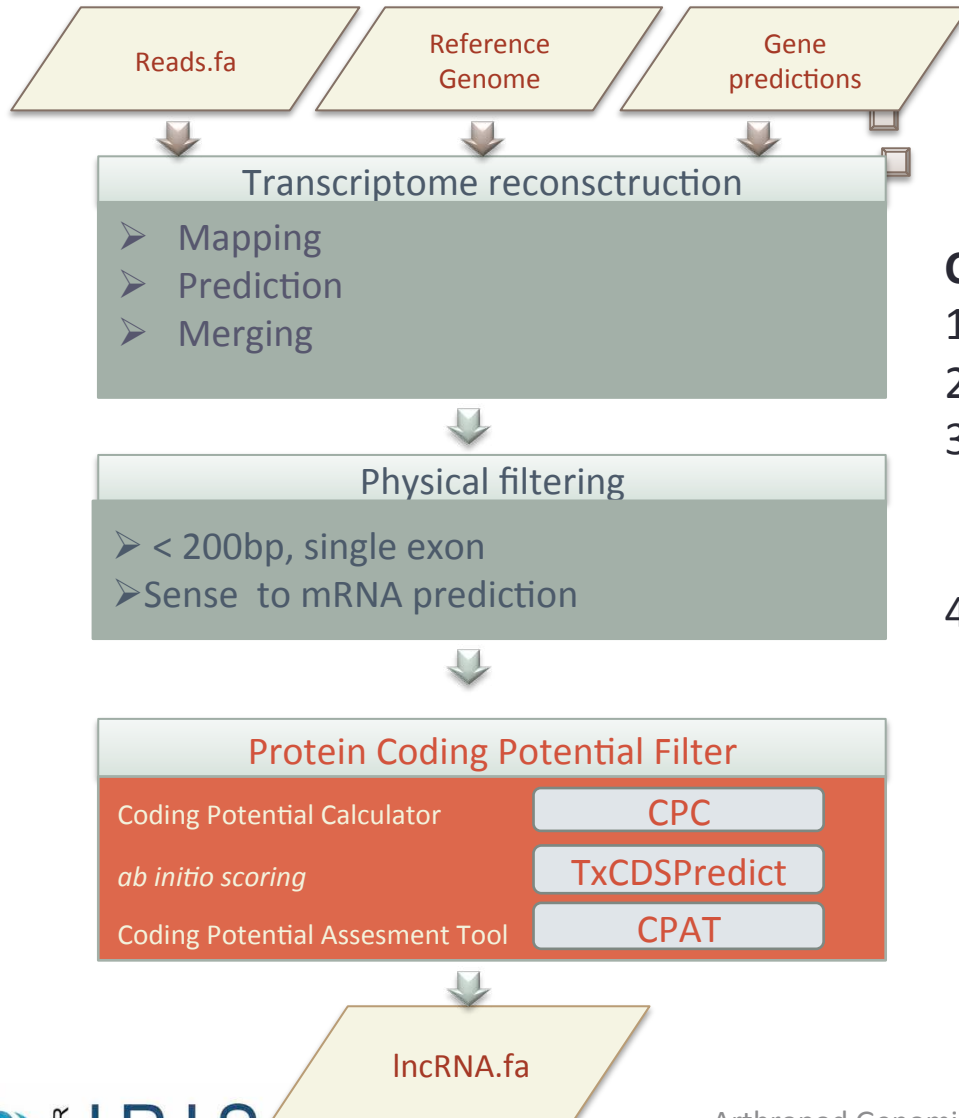
## Coding Potential Calculation

**CPC** : SVM based on 6 parameters

3 criteria based on the quality of the largest ORF (size, coverage, ATG)

3 criteria based on a result of a comparison to a protein databases (NR, Swissprot, ...) (number of hits, mean of the e-value, frameshifts)

# lncRNA Prediction workflow

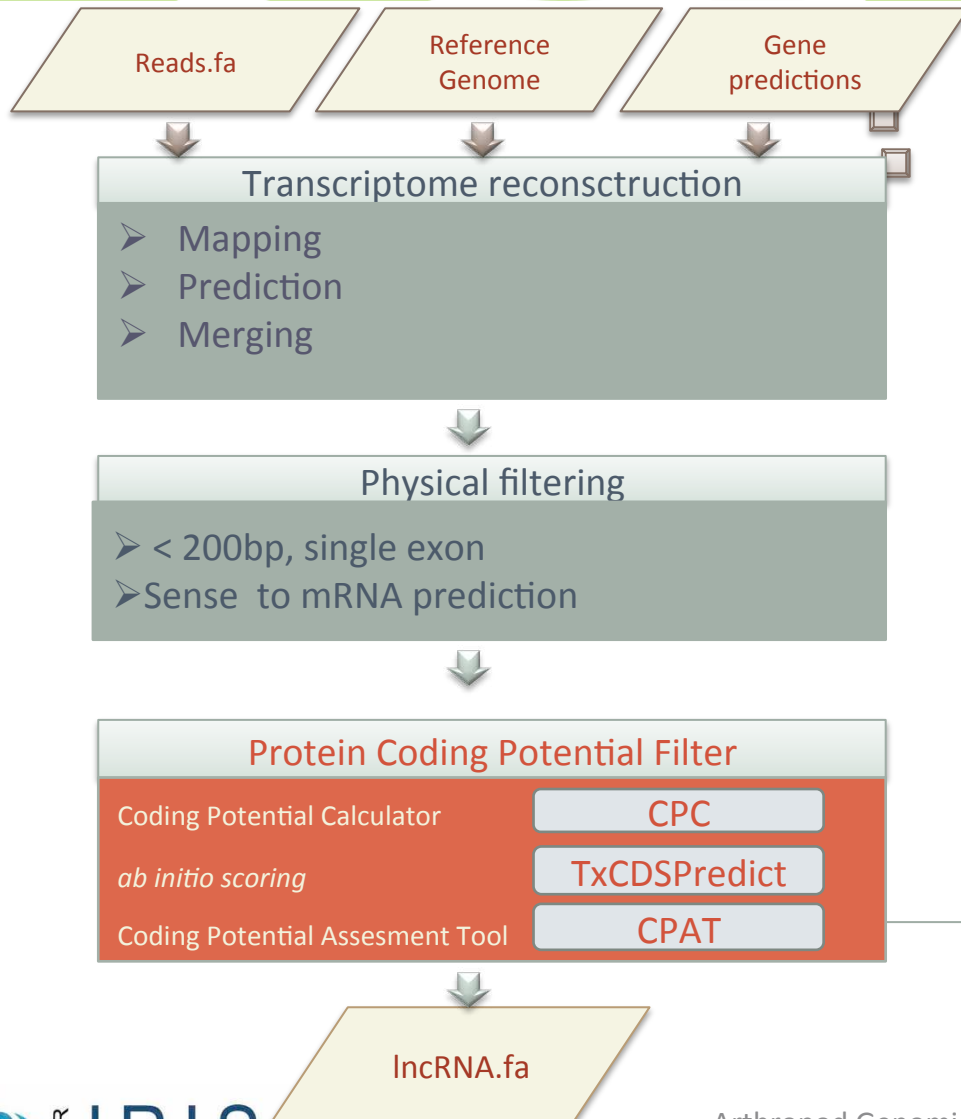


## Coding Potential Calculation

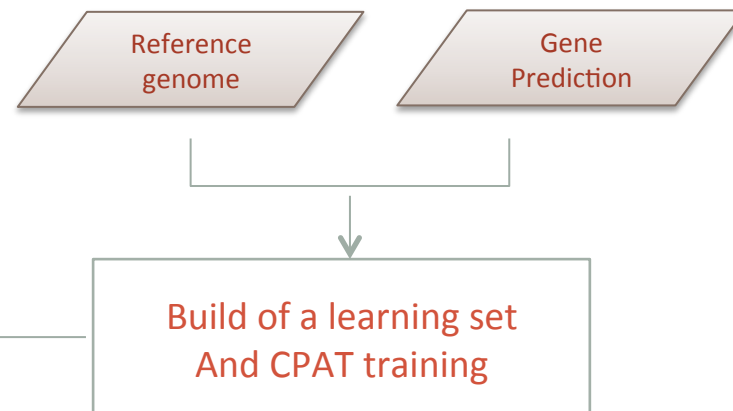
**CPAT** : logistic regression

1. ORF size
2. Coverage (ORF size/transcript size)
3. Fickett score (combination of the nucleotide composition and codon usage)
4. Hexamer score : composition of adjacent amino acids

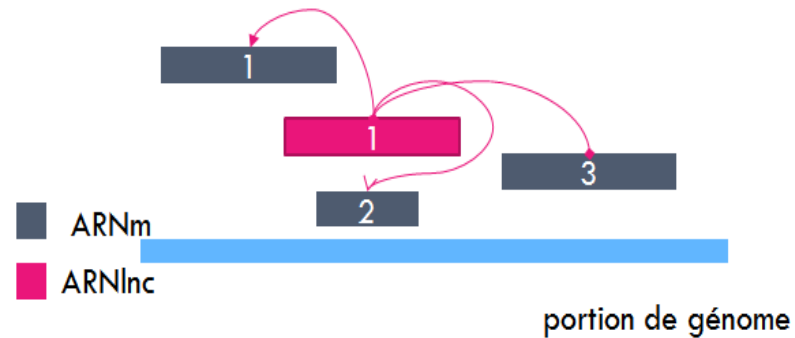
# IncRNA Prediction workflow



## Coding Potential Calculation



# lncRNA – mRNA Interactions

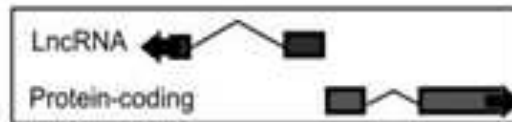


## Intergenic lncRNA

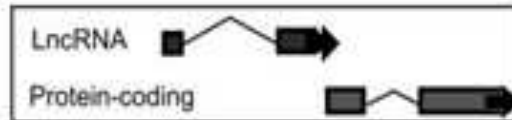
Convergent



Divergent



Same strand

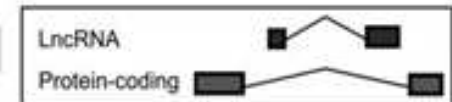


## Genic lncRNA

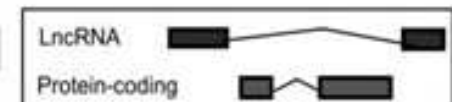
Exonic



Intronic



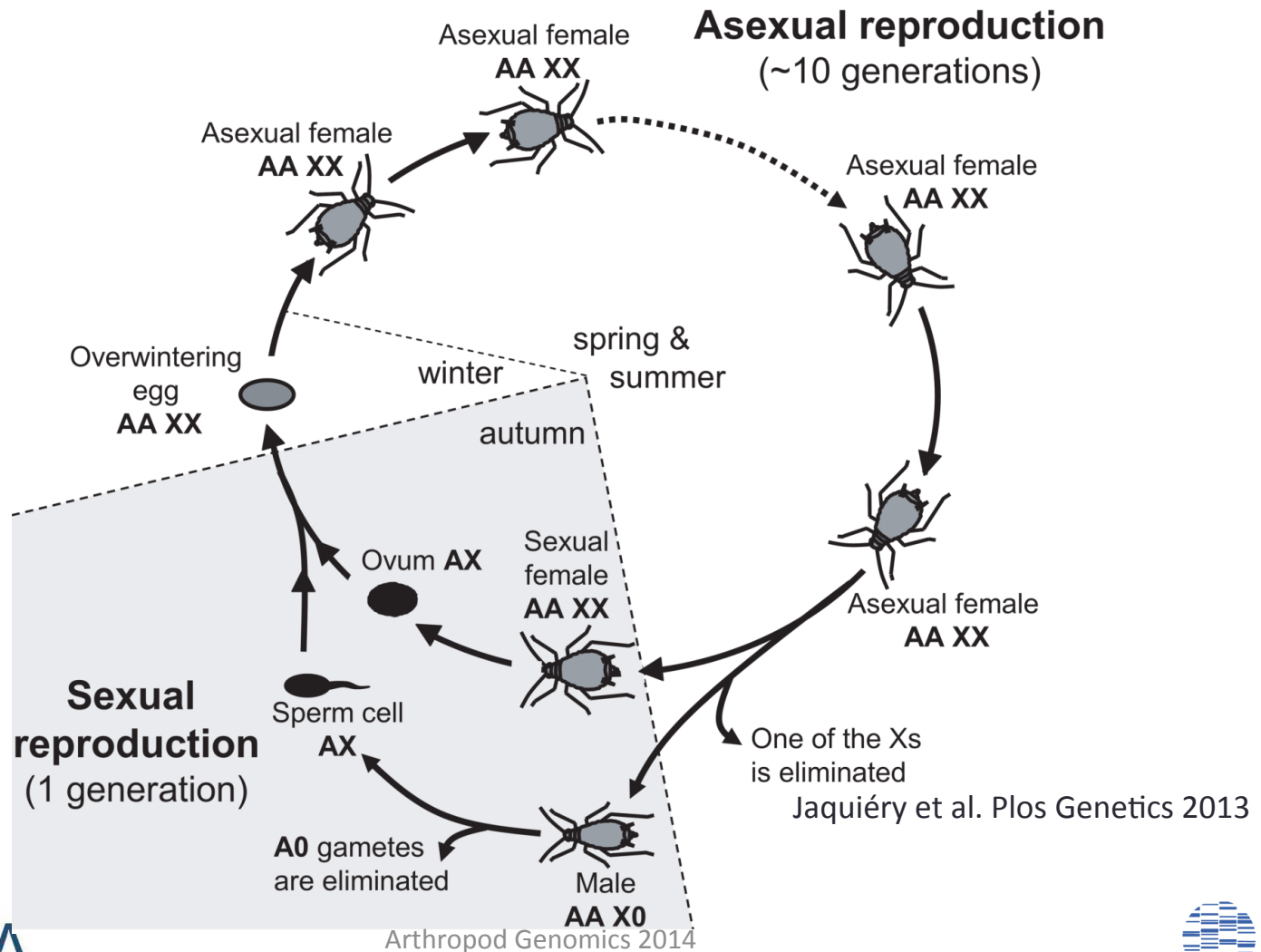
Overlapping



Derrien et al. Genome Research 2012

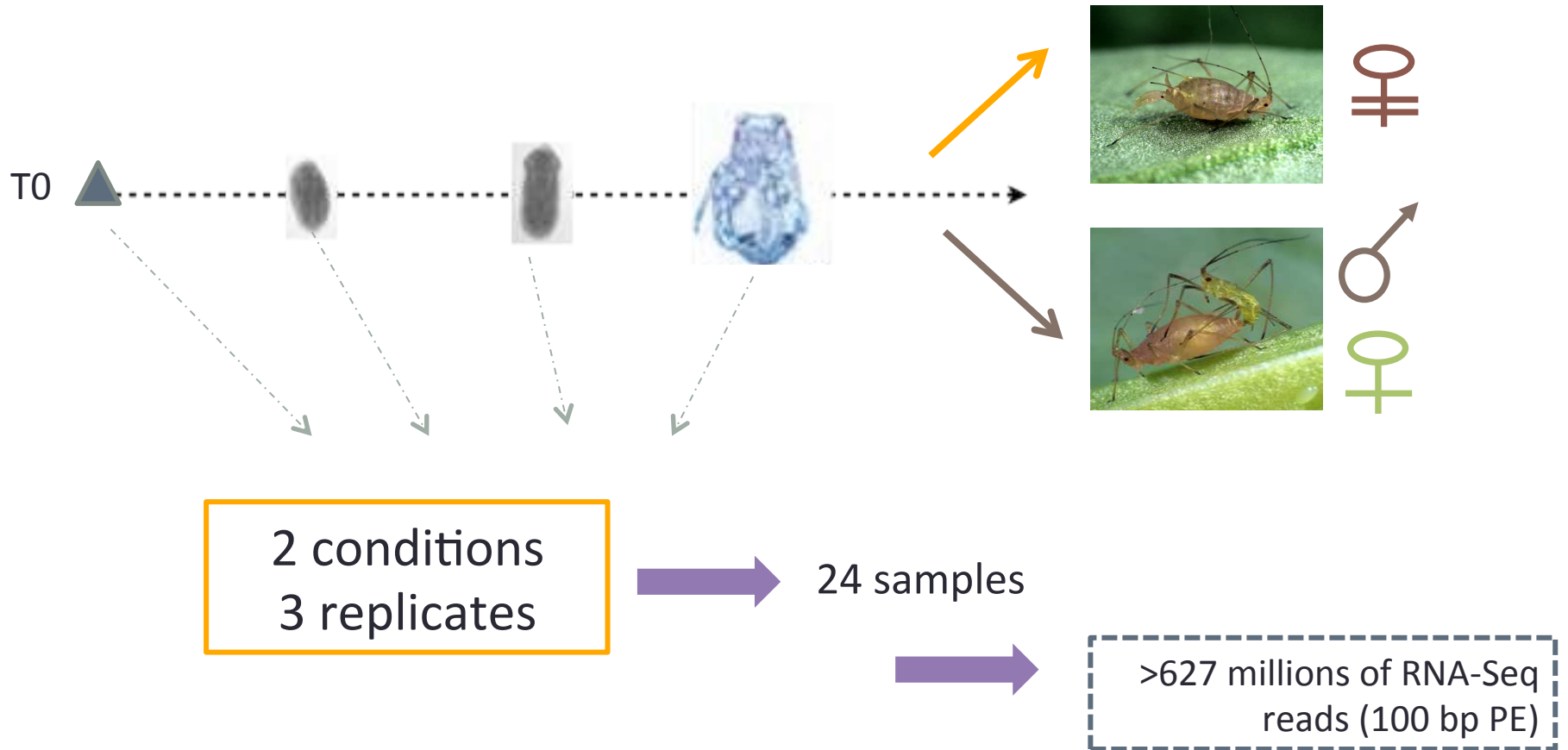
# Bioinformatics Platform for Agro-ecosystems Arthropods

## Application : the reproductive polyphenism



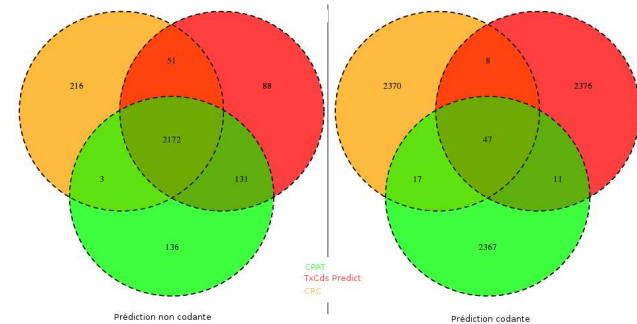
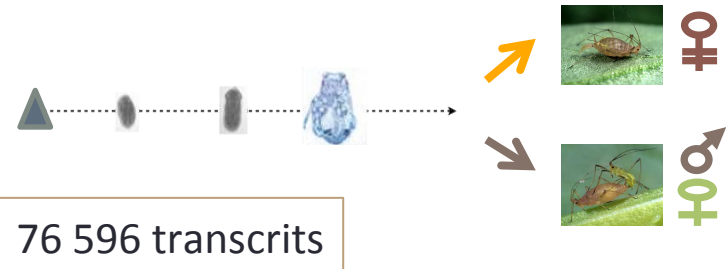
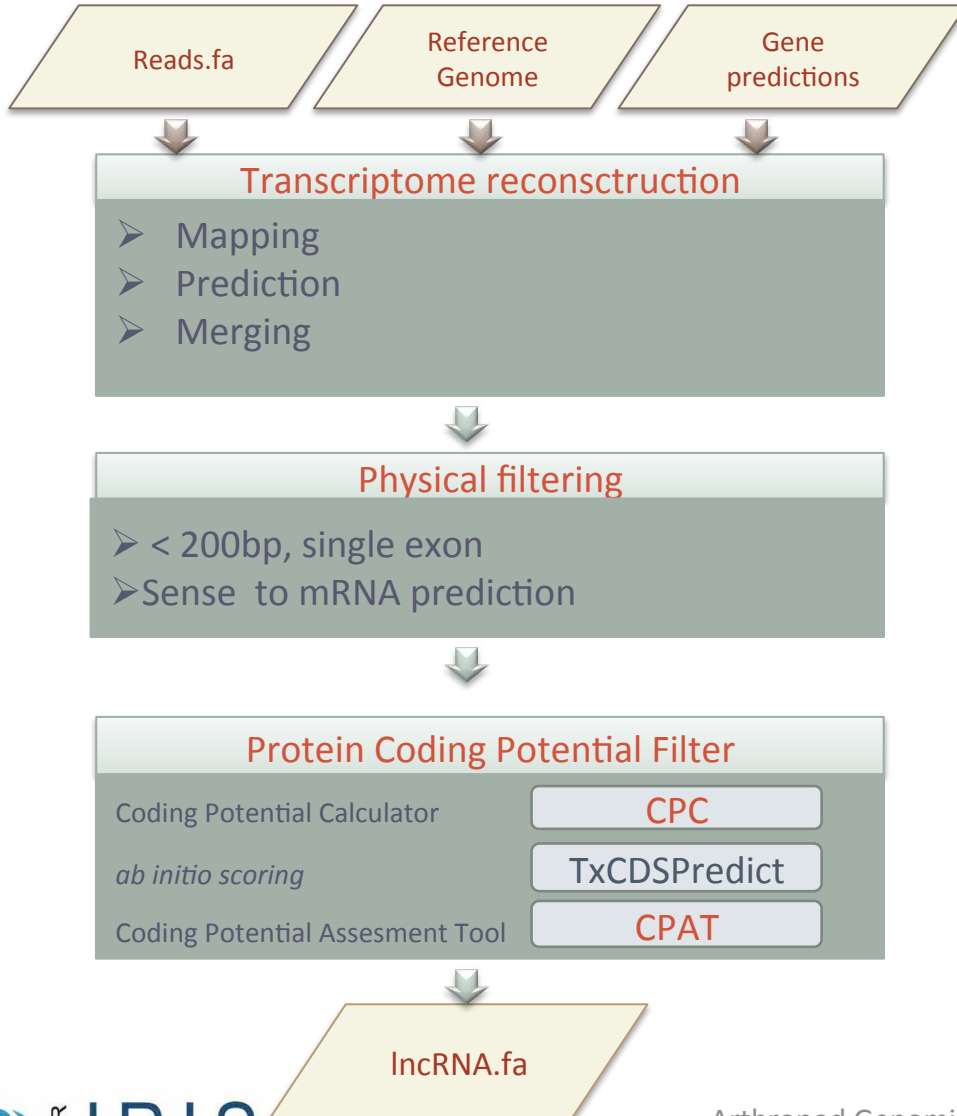
# Bioinformatics Platform for Agro-ecosystems Arthropods

## Application : the reproductive polyphenism



# Bioinformatics Platform for Agro-ecosystems Arthropods

## Application : the reproductive polyphenism



2 172 IncRNA

47 new mRNA



# Bioinformatics Platform for Agro-ecosystems Arthropods

## Application : the reproductive polyphenism

- Differentially expressed lncRNA (cuffdiff, p-value FDR < 0.05)

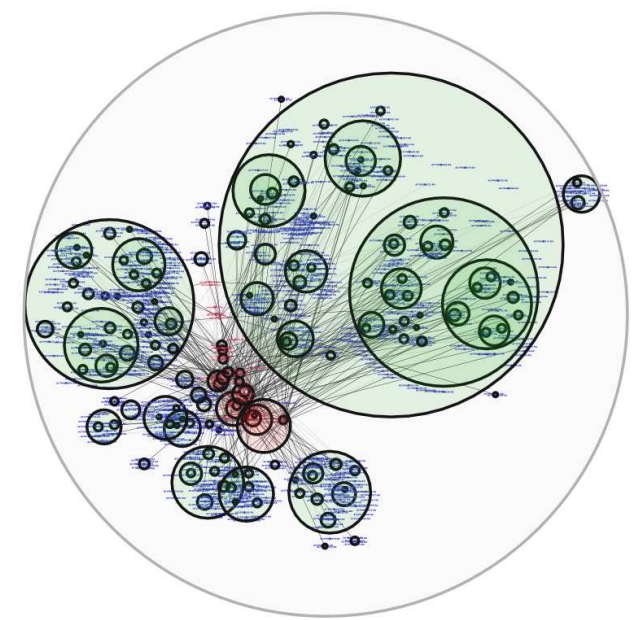
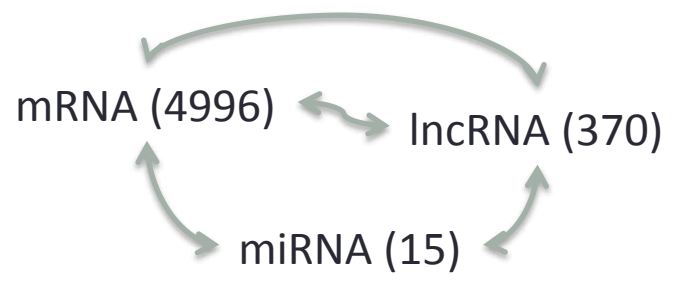
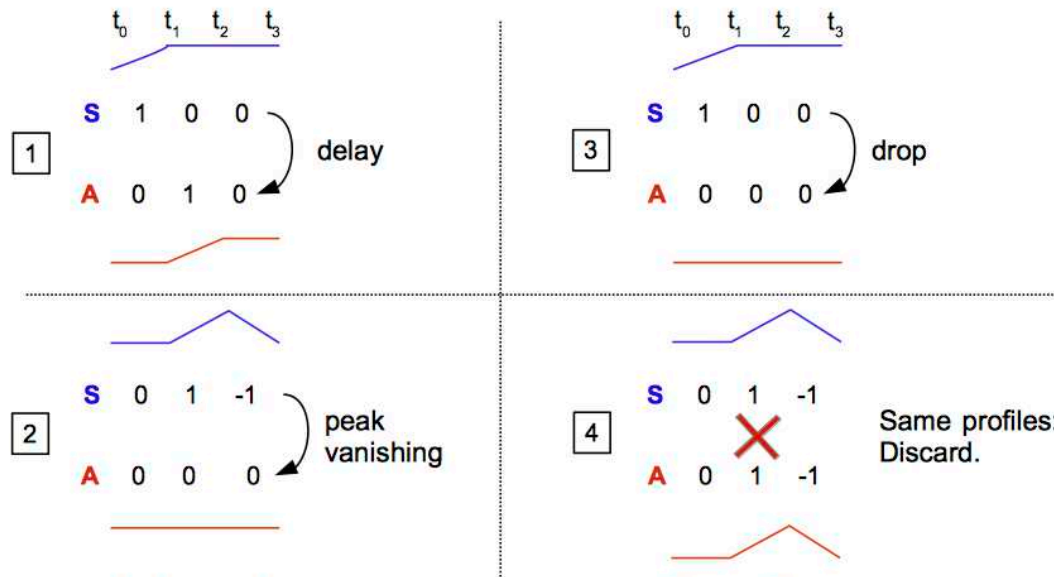
	T1	T2	T3
Sexual	2	0	4
Asexual	6	6	8

- 6 091 interactions | 2 172 lncRNA

Intergenic : 3 671	« Same strand » : 1887 Divergent : 1014 Convergent : 767	
Genic : 2 420	Intronic : 674 Exonic : 1731 overlapping: 15	« Nested » : 743

# Bioinformatics Platform for Agro-ecosystems Arthropods

## Integration



# Bioinformatics Platform for Agro-ecosystems Arthropods Availability

Analyze Data Workflow Shared Data Visualization Help User Using 56.7 MB

Welcome to the BIPAA Galaxy server

For more information or help, visit the [BIPAA website](#) or write us to [bipaa@rennes.inra.fr](mailto:bipaa@rennes.inra.fr)

INRA SCIENCE & IMPACT

BIPAA Projects Publications About

Quick access Your choice... Search

Identify yourself

Shared Data Visualization

History

Unnamed history 56.5 MB

- 6: T1K\_R2\_F0x100\_GL349623.fq
- 5: T1A\_R2\_F0x100\_GL349623.fq
- 4: myhexatab.cpat
- 3: logitmodel.RData
- 2: GL349623.qtf
- 1: GL349623.fa

Data Libraries

Published Histories

Published Workflows

Published Visualizations

Published Pages

# Bioinformatics Platform for Agro-ecosystems Arthropods

## Availability

The screenshot shows the BIPAA website interface. At the top, there is a navigation bar with links for 'BIPAA', 'Projects', 'Publications', 'About', and 'Identify yourself'. Below this is a large green banner with the text 'Bioinformatics Platform for Agro-ecosystems Arthropods'. The main content area features a heading 'Bioinformatics Platform for Agro-ecosystems Arthropods' followed by several paragraphs of text describing the platform's mission, its association with INRA and INRIA, and its collaborative nature. To the right, there is a 'BIPAA Resources' sidebar with links to 'Aphidbase', 'LepidoDB', 'Galaxy server', and 'Create an account'. Below the text, there are two circular icons for 'Aphidbase' and 'LepidoDB', and a 'Galaxy' logo. At the bottom, there is a note about creating an account using the BIPAA registration form.

BIPAA Projects Publications About Identify yourself

## Bioinformatics Platform for Agro-ecosystems Arthropods

BIPAA

### Bioinformatics Platform for Agro-ecosystems Arthropods

BIPAA is a central platform to assist genomics and post-genomics programs developed on insects associated to agroecosystem, for the Plant Health and Environment Division of **INRA**. It shares resources and standards critical to guarantee data integration and interoperability, and improves the coherence of the protocols set up by its various partners. Apart from giving access to the data, the main utility of BIPAA is to settle an environment allowing a larger community to elaborate complex genomics analyses, browsing, mixing or crossing heterogeneous data.

BIPAA works with different INRA labs, and is mainly associated with national networks settled on aphids, lepidopteran and their associated parasitoids and predators. It is also involved in international projects on insect genome sequencing such as **ISK**.

The platform aims at developing particular treatments on a large panel of "omics" (e.g. genomics, transcriptomics, epigenomics, methylomics or proteogenomics) in strong collaboration with the INRIA **GenScale** team headed by Dominique Lavenier at Rennes.

The team benefits from the housing of the **Genouest** platform, an **IBISA** platform, member of **RENABI** (French network of bioinformatics platforms) headed by Olivier Collin at INRIA Rennes.

Collaborative programs with BIPAA are driven by the two French networks **BAPOA** and **ADALEP**.

BIPAA created and manages two dedicated website on genomic data:



A dedicated **Galaxy server** is available with specific resources and tools for the analysis of insects genomic data:



To create an account, use the **BIPAA registration form**.

**BIPAA Resources**

- Aphidbase
- LepidoDB
- Galaxy server
- Create an account

*Bioinformatics Platform for Agro-ecosystems Arthropods*  
**Acknowledgments**

- Gaël Le Trionnaire, Aurore Gallot
- Valentin Wucher
- Julie Jaquiéry
- Audrey David, Thomas Derrien
- Denis Tagu

