Exploration of the
Sonalum nigrum transcriptome
for the prediction of
extracellular peptides
involved in the ROS regulation.

PROJET PHAACE

new types of **PH**ytocytokines,
As an **A**lternative to **C**hemical h**E**rbicides
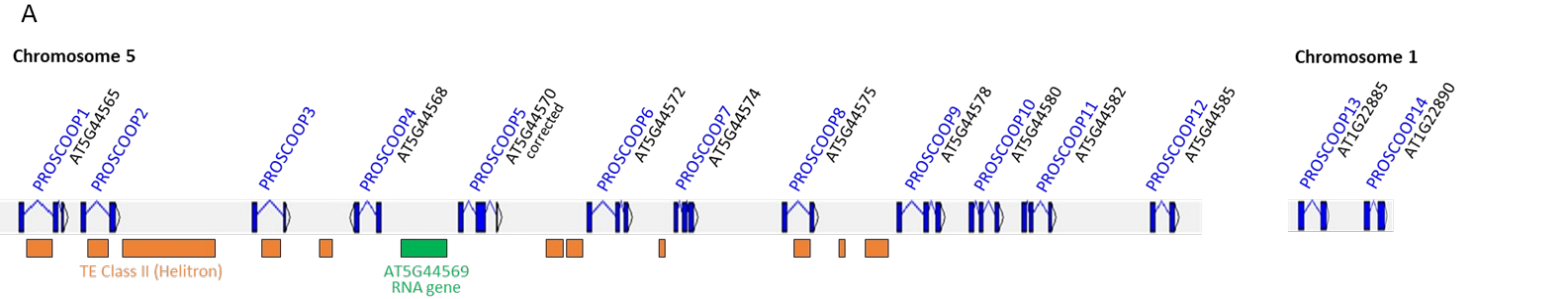SPE Innovation 2020-22

Coord: J.P. Renou, S. Aubourg
(IRHS Angers, INRAE/Univ Angers/IARA)

IRHS
Institut de Recherche en
Horticulture et Semences
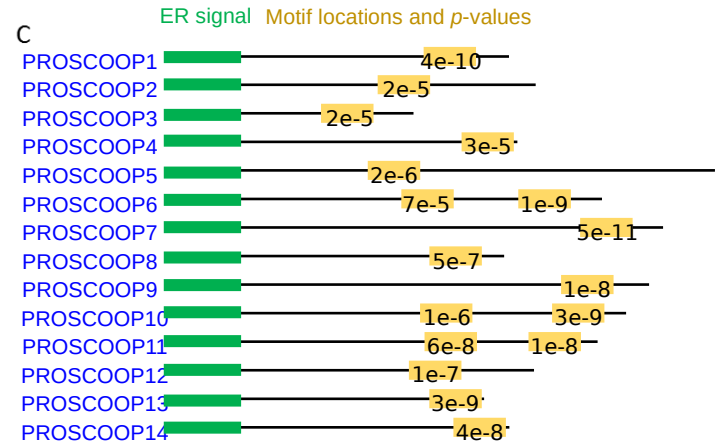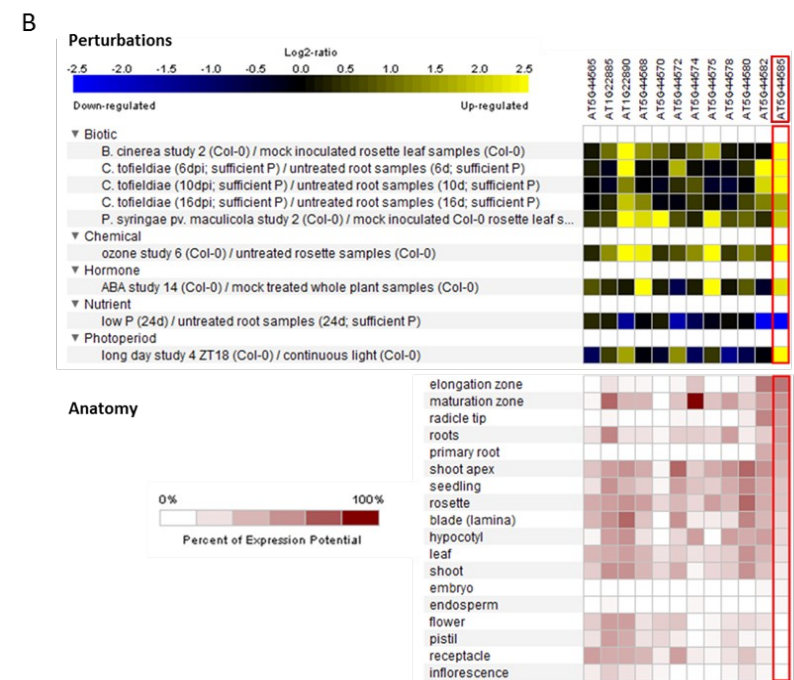INRAE L'INSTITUT agro Rennes Angers université angers

Discovery of the **SCOOP Family** through Orphan gene function inference by genomic data-mining: A new signal peptide family in Arabidopsis
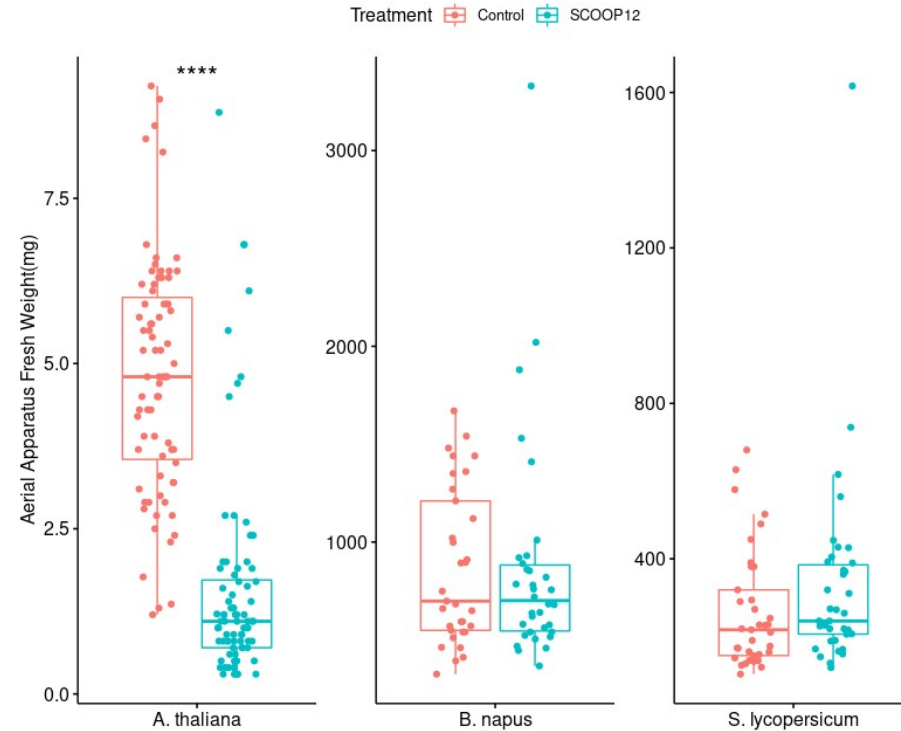

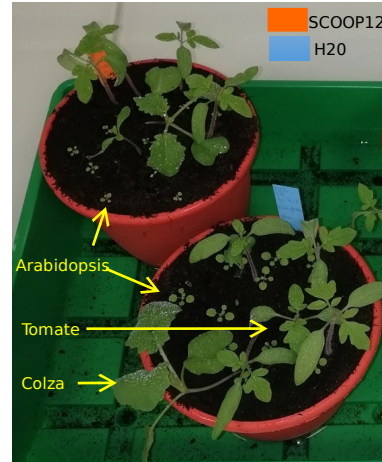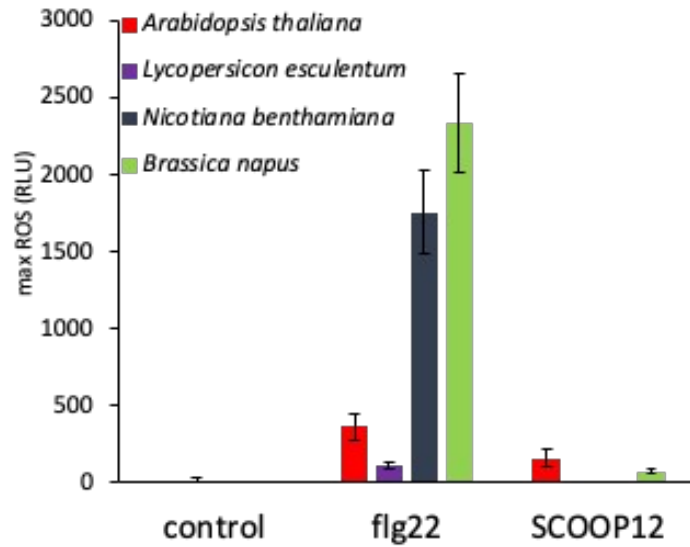From Pelletier, Gully *et al*, 2019

A synthetic SCOOP12 peptide treatment induces defense responses, ... and growth inhibition in a species specific fashion

From Pelletier, Gully *et al*, 2019

26 days seedlings after 14 days of treatment, 3 repetitions

Application : phytocytokine as selective bio-herbicide ?
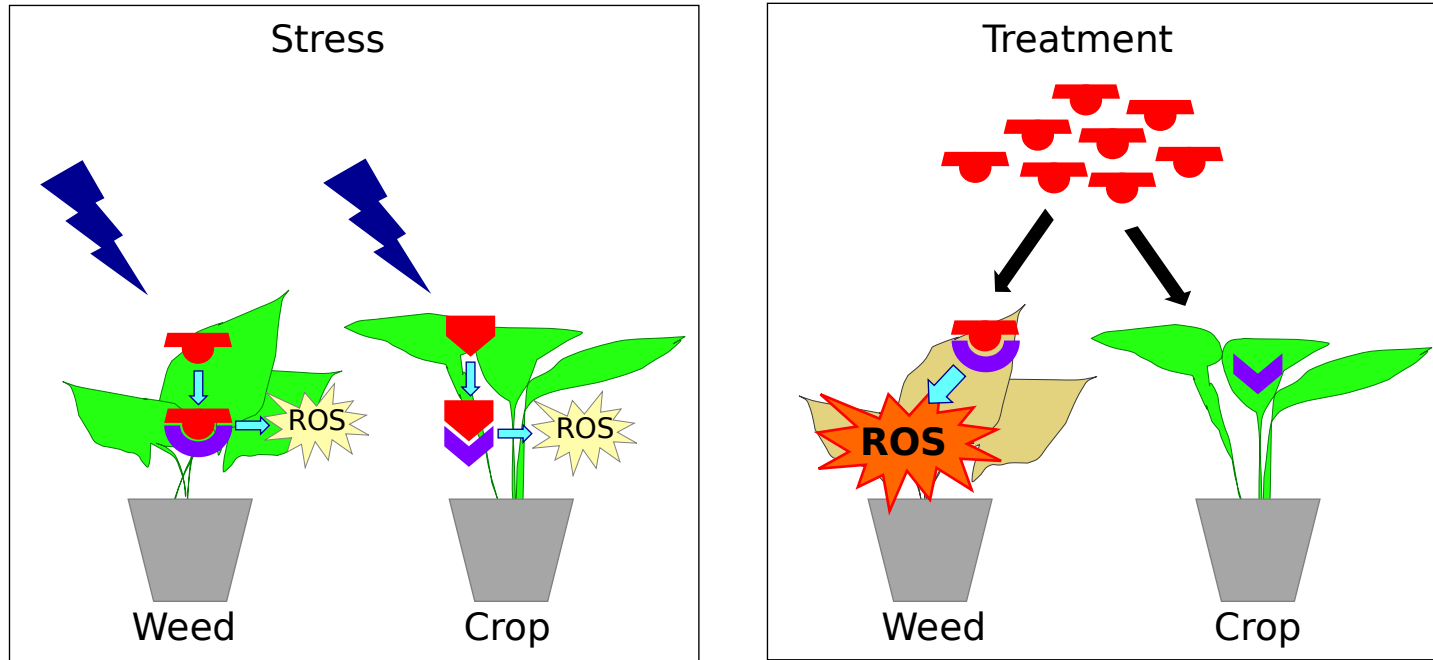
CONCEPT : specific **peptide-receptor** recognition. There are hundreds of putative candidates per species

HOW TO PROCEED ? Identify **SCOOP functional analogs** in other species (expressolog screening)

PUTATIVE OTHER APPLICATIONS : Growth regulators in horticulture, fruit thinning, activation of plant defenses....

Project objectives

Identification of genes encoding secreted peptides with function analogous to SCOOP12 (ROS burst, growth inhibition) in the weed *Solanum nigrum*
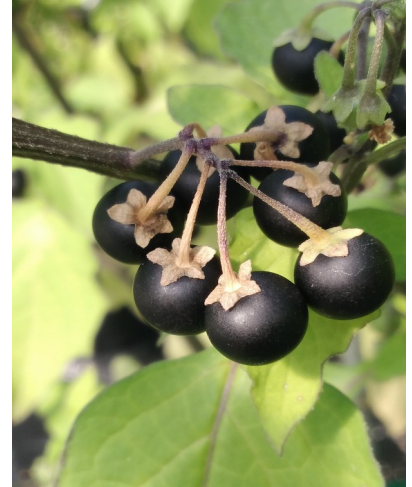

- **Assembly of *de novo* transcriptome reference**

- **Contig annotation**
  - Detection of ORF encoding short protein with signal peptide
  - Functional annotation (Interproscan, similarities in tomato and Arabidopsis)

- **Transcriptomic profiling**
  - RNA-seq differential analysis in stress conditions

- **Selection of candidates**

- **Test with synthetic predicted peptides**

De novo transcriptome assembly

1 complex sample
(mix of fruits, leaves, seedlings+zebularine, leaves+Xanthomonas, flowers, roots)

9 samples
(leaves, leaves +Atrazine, +H$_2$O$_2$, +SA, +Methyl viologen, +ABA, +Clavibacter, +Xanthomonas)

Sample preparation

De novo transcriptome assembly

RNAseq Long read
PacBio Sequel2
1 complex sample
(mix of fruits, leaves, seedlings+zebularine,
leaves+Xanthomonas, flowers, roots)

RNAseq Short read
Illumina Novaseq PE150
9 samples
(leaves, leaves +Atrazine, $+H_2O_2$, +SA, +Methyl
viologen, +ABA, +Clavibacter, +Xanthomonas)

980 M read pairs

Gentyane

ISOSEQ pipeline
cov1+
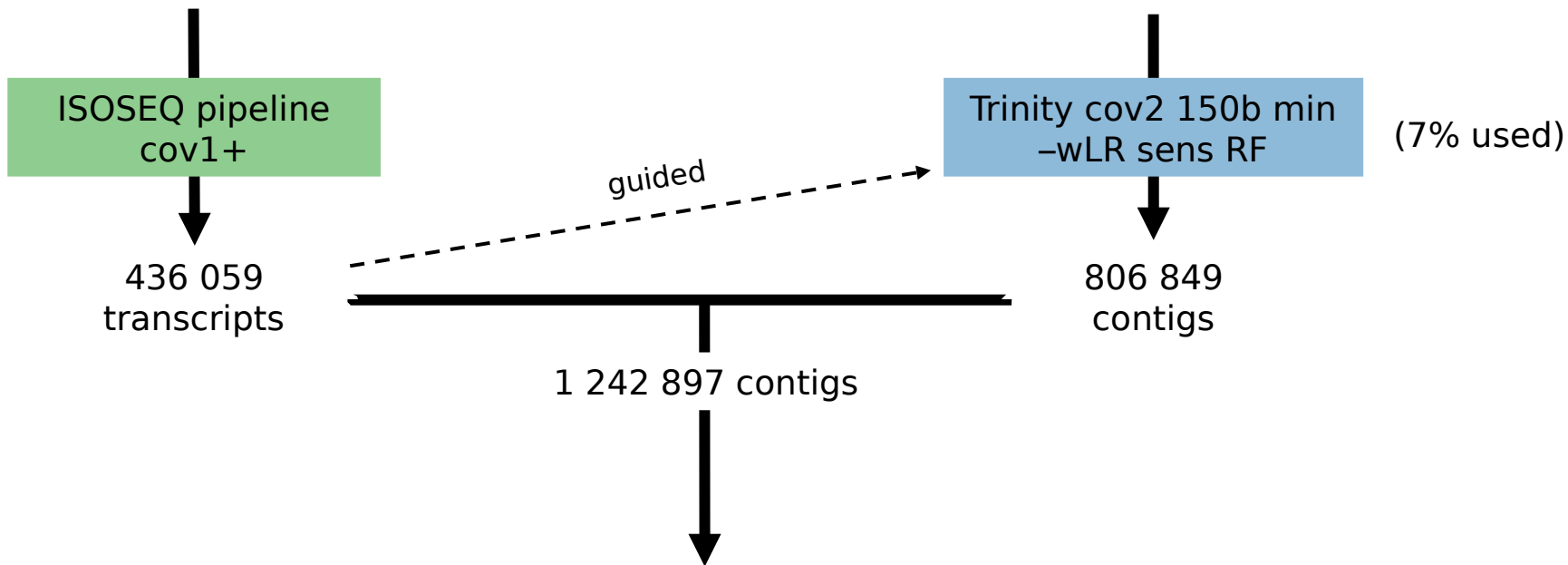
Trinity cov2 150b min
–wLR sens RF
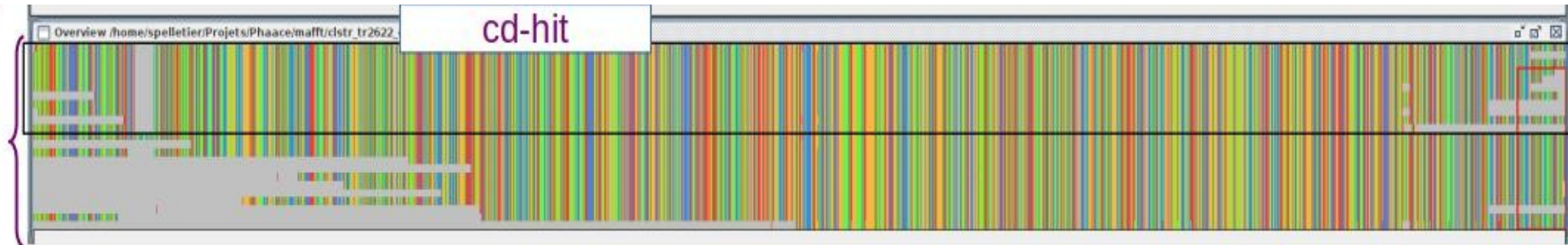
(7% used)

guided

436 059
transcripts

806 849
contigs

1 242 897 contigs

De novo transcriptome assembly

Warning :   partially spliced transcripts > no selection of longer contigs after Trinity
alleles and ohnologs (*Solanum nigrum* is hexaploïd, 2n=6x=?) ?
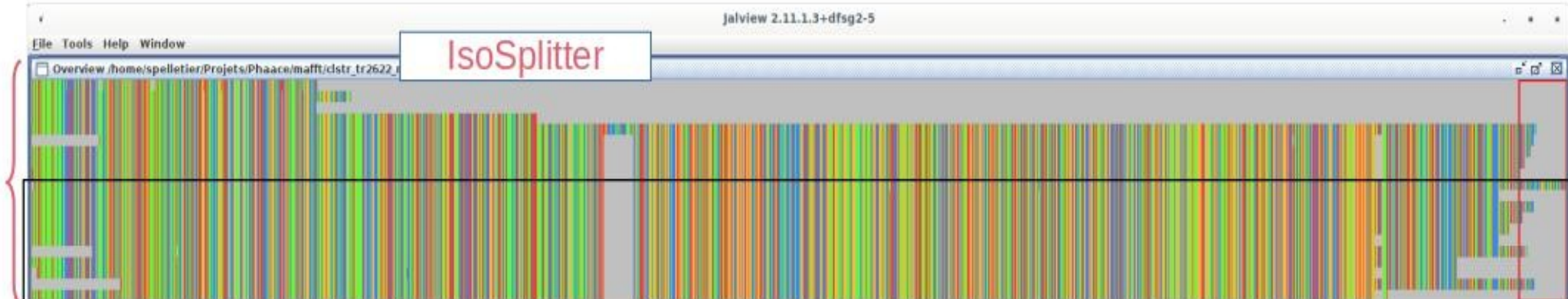
De novo transcriptome assembly

RNAseq Long read
PacBio Sequel2
1 complex sample
(mix of fruits, leaves, seedlings+zebularine, leaves+Xanthomonas, flowers, roots)

RNAseq Short read
Illumina Novaseq PE150
9 samples
(leaves, leaves +Atrazine, +H_2O_2, +SA, +Methyl viologen, +ABA, +Clavibacter, +Xanthomonas)

980 M read pairs

INRAE
Gentyane

ISOSEQ pipeline cov1+

Trinity cov2 150b min –wLR sens RF

(7% used)

guided

436 059 transcripts

806 849 contigs

1 242 897 contigs

EvidentialGene tr2aacds2

166 179 contigs

from Trinity: 115 959
from Sequel: 50 220

De novo transcriptome assembly
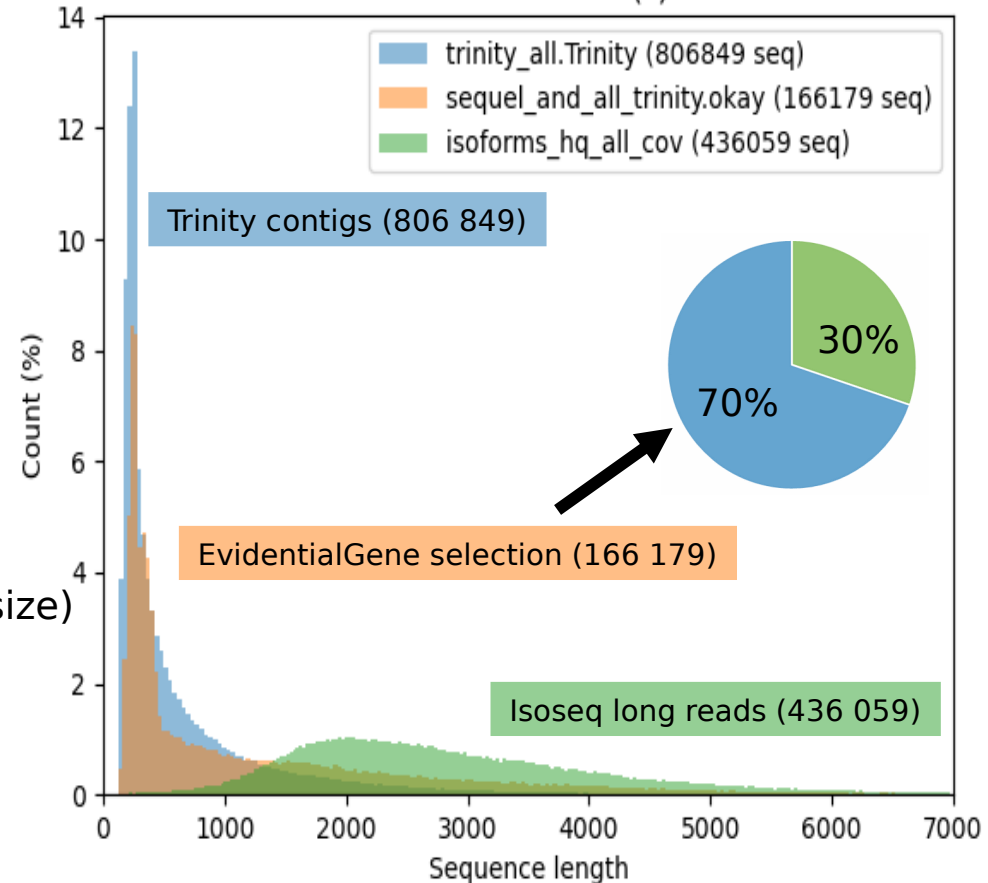
Quality controls (used for assembly pipeline optimization) of 166 179 contigs

- **Contig metrics** :
  - ➔ N50 = 2540 b
  - ➔ Average = 1285 b
  - ➔ Median = 633 b

- **Busco *Solanaceae*** (3 052 proteins) :
  - ➔ 95.9 % complete proteins
  - ➔ 1.9 % (59) missing proteins

- **BLASTX vs *S. lycopersicum* proteome**
  (ITAG4, 34 075 proteins) with cutoff 1e-5 :
  - ➔ 64 % of distinct proteins (min 30% of their size) tagged by 31 % of contigs

- **Short read mapping**
  (leaf bact. stress control sample) :
  - ➔ 83.43 % of mapping rate

fasta distribution(s)

Legend:
- trinity_all.Trinity (806849 seq)
- sequel_and_all_trinity.okay (166179 seq)
- isoforms_hq_all_cov (436059 seq)

Trinity contigs (806 849)

EvidentialGene selection (166 179)

Isoseq long reads (436 059)

70%   30%

Bioinformatics methods for prediction of gene families encoding extracellular peptides

166 179 contigs

ORFfinder

ORFparser

BLASTP

BLASTX

ITAG4

Basic:

- mainly short proteins harbouring an ER signal peptide in N terminal,

- a more or less conserved motif in C terminal (putative mature peptide)

ORF selection and annotation

Σ = 127 918 ORF

frameshift 3%

suspicious 7%

no ORF 23%

38261

66151

43880

ok 27%

unconserved 40%

ORFparser

Bioinformatics methods for prediction of gene families encoding extracellular peptides

Basic:

- mainly short proteins harbouring an ER signal peptide in N terminal,

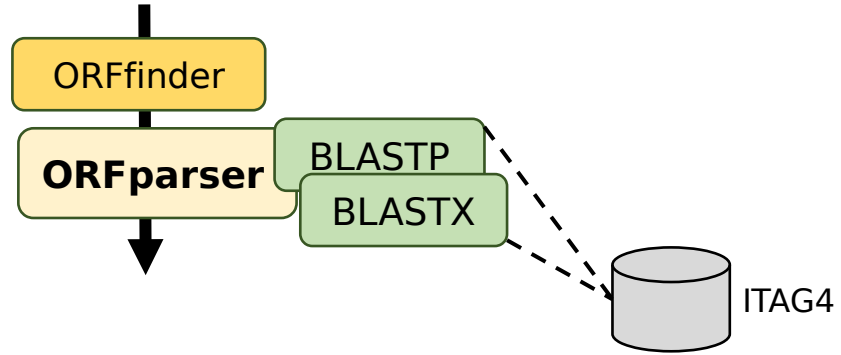- a more or less conserved motif in C terminal (putative mature peptide)

Bioinformatics methods for prediction of gene families encoding extracellular peptides
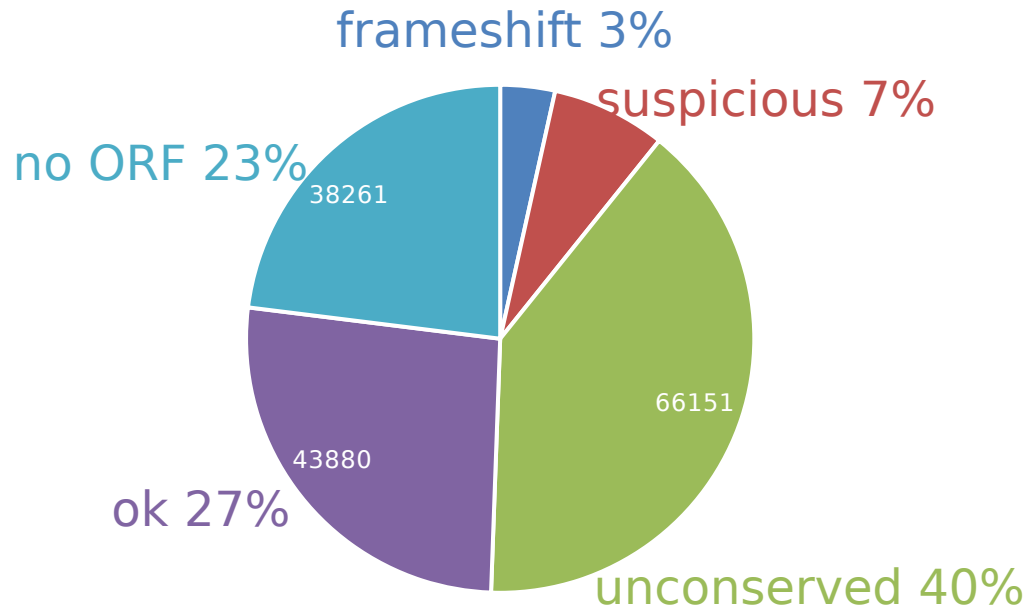
Basic:

- mainly short proteins harbouring an ER signal peptide in N terminal,

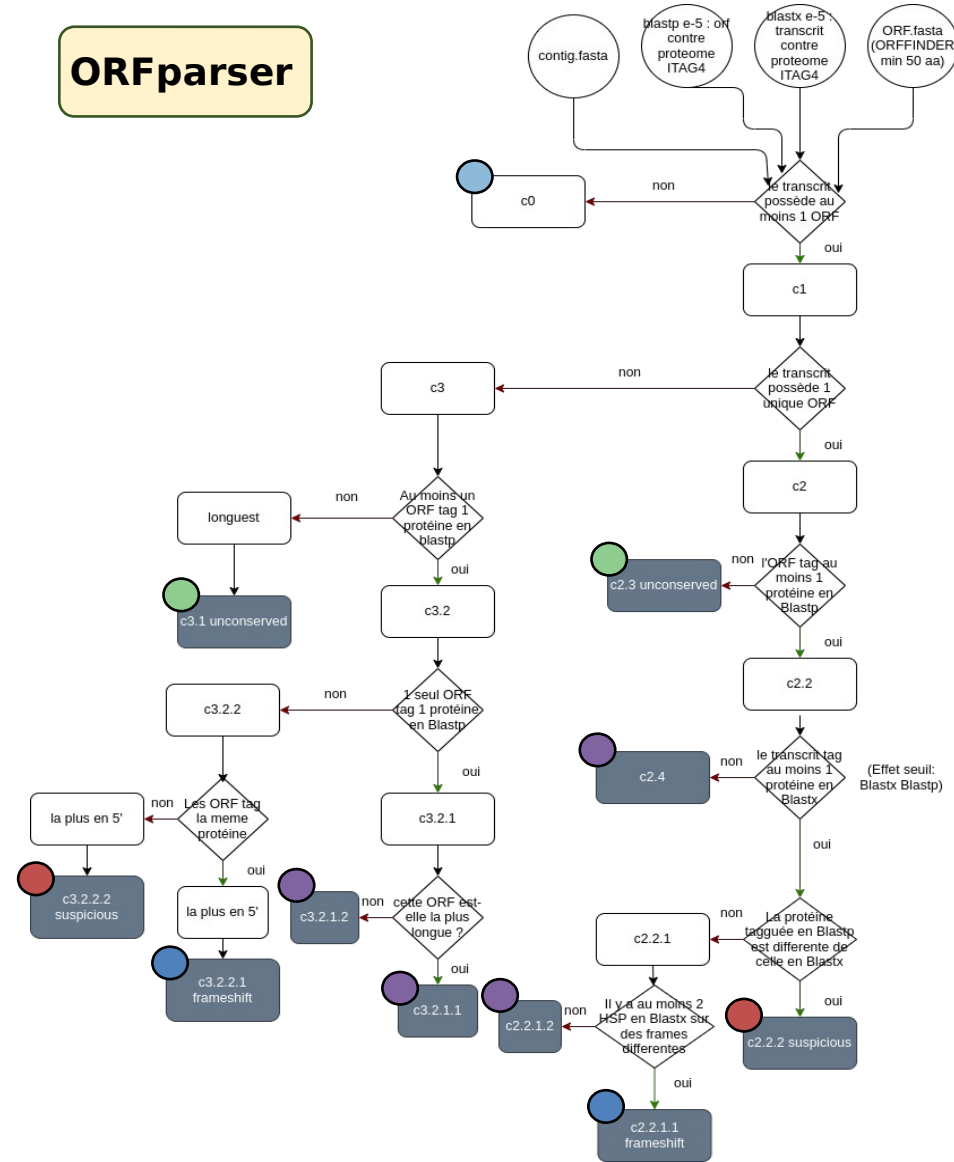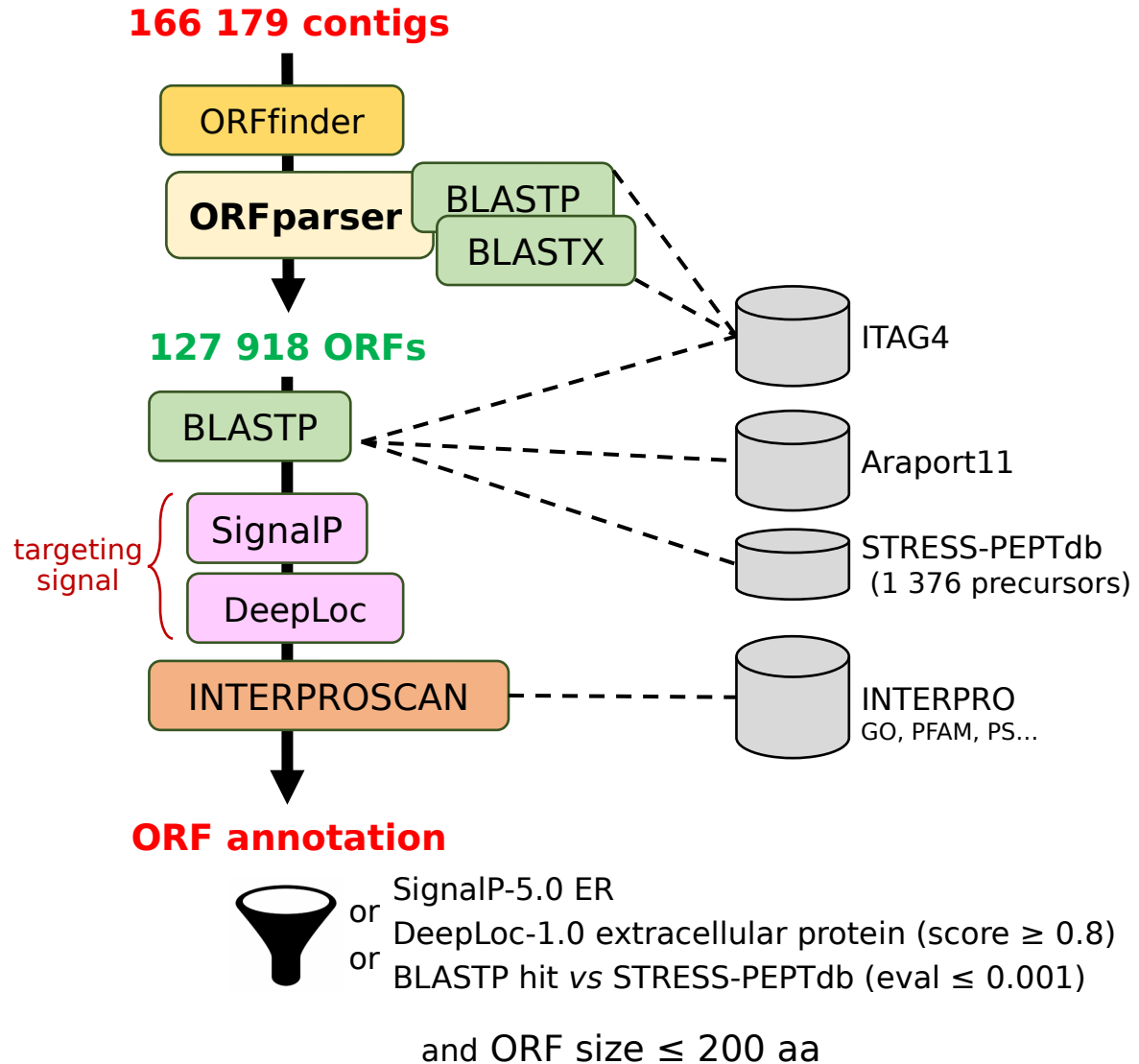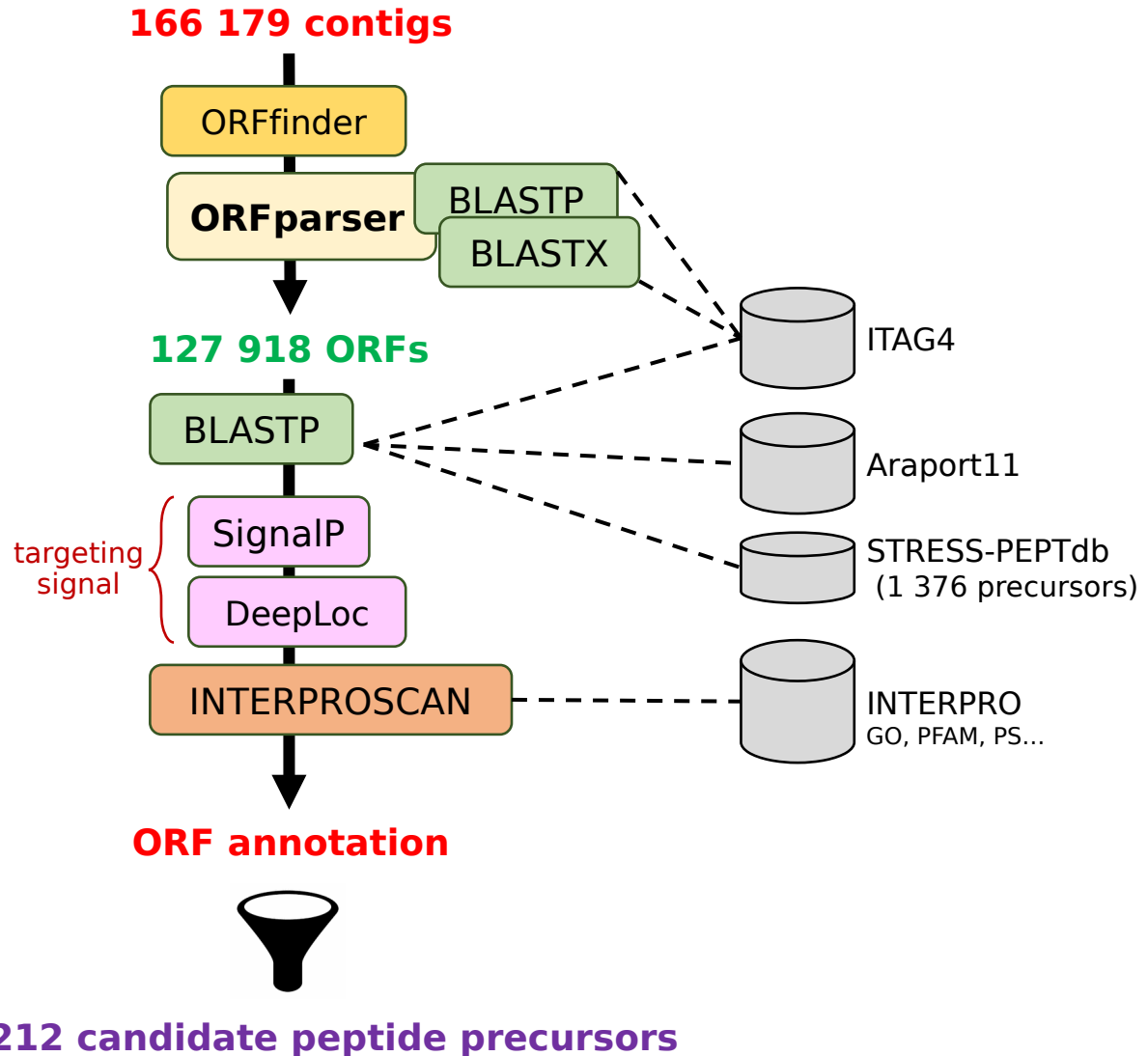- a more or less conserved motif in C terminal (putative mature peptide)

166 179 contigs

ORFfinder

ORFparser — BLASTP / BLASTX

127 918 ORFs

BLASTP

SignalP } targeting signal

DeepLoc }

INTERPROSCAN

ORF annotation

6 212 candidate peptide precursors

ITAG4

Araport11

STRESS-PEPTdb (1 376 precursors)

INTERPRO GO, PFAM, PS...

Transcriptomic profiling

9 biological samples, 3 repetitions

Biotic stresses : *Xanthomonas Campestris, Clavibacter michiganensis* (infiltrated) } *vs* control (vacuum)

Abiotic stresses : Atrazine, H202, ABA, Methylviologen (paraquat), (spray) } *vs* control (H_2O spray)

Other : SA

RNA-seq (Illumina Novaseq PE150) @Novogene

Read mapping on 166 179 *S. nigrum* contigs (Salmon)

Average mapping rate of **82.55 %** [79.4 – 84.3]

Differential analysis (anaDiff)

Average common dispersion : 0.62 (biotic), 0.36 (abiotic)

Transcriptomic profiling

|logFC| > 0.5 ; BH ≤ 10%

3 366 transcripts (contigs)
diff expressed at least in one sample

	up	**down**
ABA	2	2
Atrazine	622	419
H_2O_2	189	6
MEV	899	10
SA	8	0
Clavibacter	335	26
Xanthomonas	1 440	252
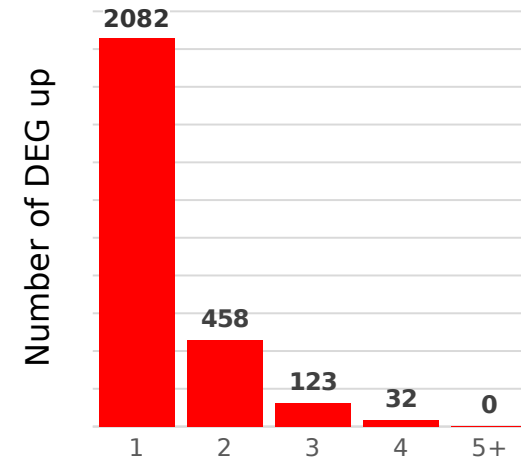
2 695 up at least in **one** sample

GO enrichement
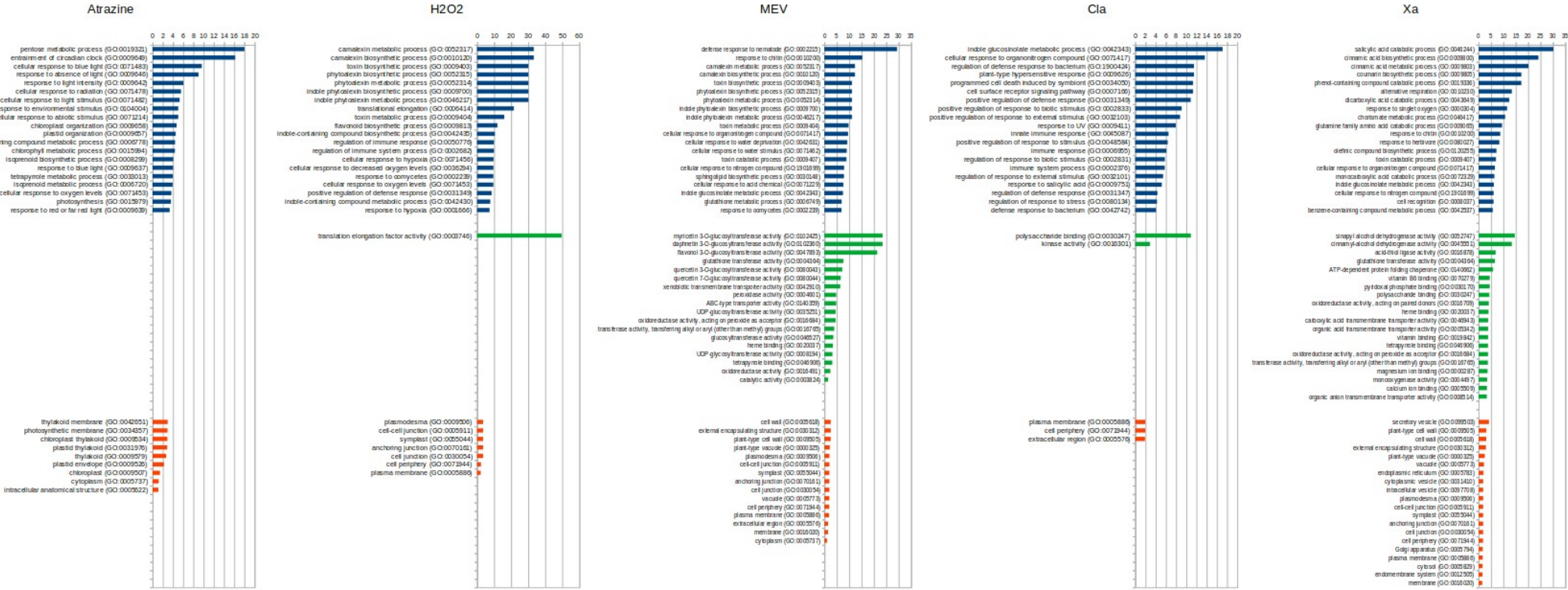
GO biological process complete
GO molecular function complete
GO cellular component complete

GO enrichement

GO biological process complete

GO enrichement

GO cellular component complete

H2O2

plasmodesma (GO:0009506)
cell-cell junction (GO:0005911)
symplast (GO:0055044)
anchoring junction (GO:0070161)
cell junction (GO:0030054)
cell periphery (GO:0071944)
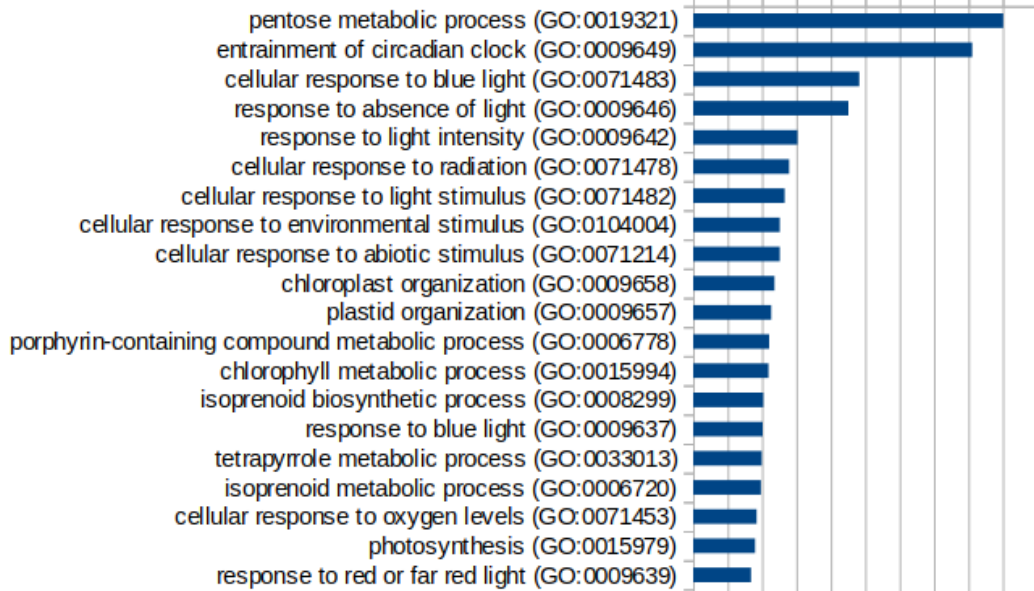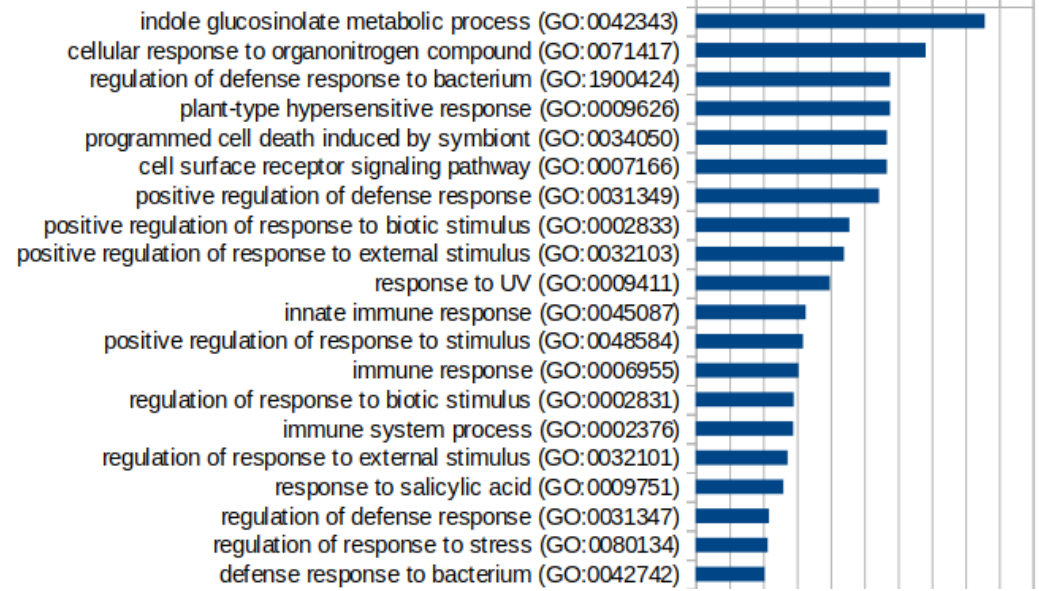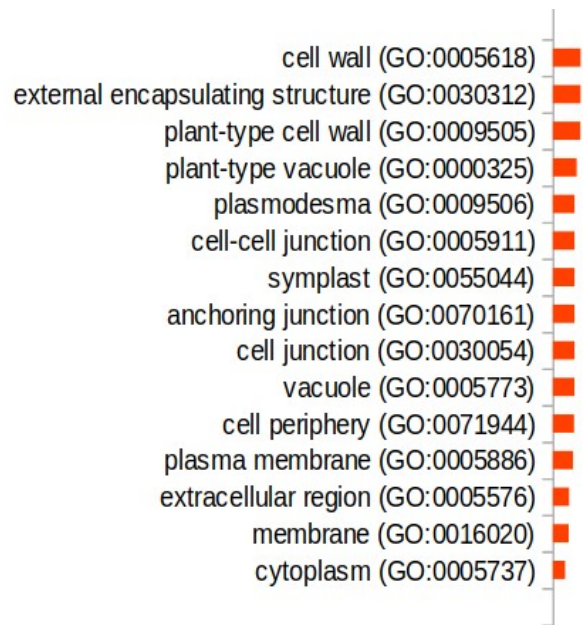plasma membrane (GO:0005886)

Cla

plasma membrane (GO:0005886)
cell periphery (GO:0071944)
extracellular region (GO:0005576)

MEV

cell wall (GO:0005618)
external encapsulating structure (GO:0030312)
plant-type cell wall (GO:0009505)
plant-type vacuole (GO:0000325)
plasmodesma (GO:0009506)
cell-cell junction (GO:0005911)
symplast (GO:0055044)
anchoring junction (GO:0070161)
cell junction (GO:0030054)
vacuole (GO:0005773)
cell periphery (GO:0071944)
plasma membrane (GO:0005886)
extracellular region (GO:0005576)
membrane (GO:0016020)
cytoplasm (GO:0005737)

Xa

secretory vesicle (GO:0099503)
plant-type cell wall (GO:0009505)
cell wall (GO:0005618)
external encapsulating structure (GO:0030312)
plant-type vacuole (GO:0000325)
vacuole (GO:0005773)
endoplasmic reticulum (GO:0005783)
cytoplasmic vesicle (GO:0031410)
intracellular vesicle (GO:0097708)
plasmodesma (GO:0009506)
cell-cell junction (GO:0005911)
symplast (GO:0055044)
anchoring junction (GO:0070161)
cell junction (GO:0030054)
cell periphery (GO:0071944)
Golgi apparatus (GO:0005794)
plasma membrane (GO:0005886)
cytosol (GO:0005829)
endomembrane system (GO:0012505)
membrane (GO:0016020)

Transcriptomic profiling

2 695 up at least in **one** sample

6 212 candidates   **158**   **2 695 up**

curation, selection
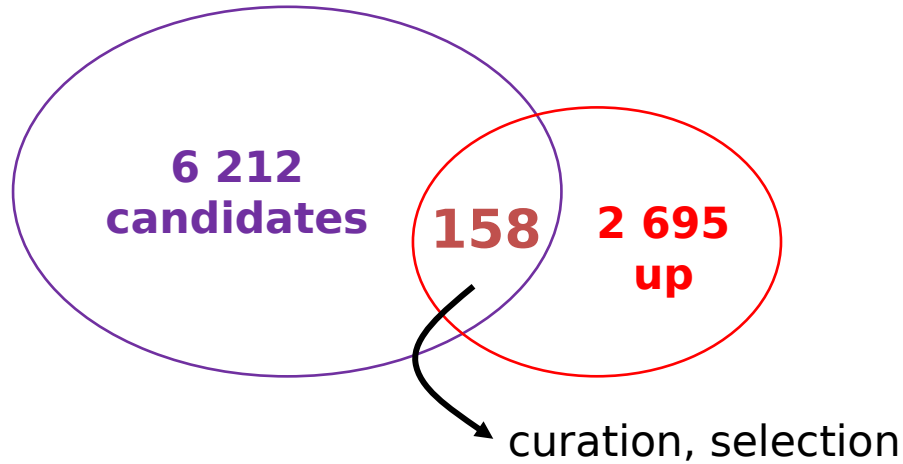
Selection criteria : up in bacteria
or H_2O_2 or MEV or Atrazine or SA

Candidate overview

PTMP	C-terminally Encoded Peptide	CEP	2
PTMP	Small phytoCytokines REgulating defense and Water loss	SCREW/CTNIP	1
PTMP	Inflorescence Deficient in Abscission (-like)	IDA	1
PTMP	Pamp-Induced secreted Peptide	PIP	5
PTMP	Phytosulfokine	PSK	1
HRGP	ArabinoGalactan Protein	AGP	7
HRGP	Phytocyanin/Early NODulin-Like	ENODL	2
HRGP	Extensin	EXT	2
CRP	Kunitz-type Trypsin Inhibitor	KTI	3
CRP	Non-specific Lipid Transfer Protein	ns-LTP	13
CRP	Plant Defensin Family (-Like)	PDF	6
CRP	Pollen allergen Ole e 1		1
CRP	Snakin/Gibberellic Acid-Stimulated Arabidopsis	GASA	1

nd	(Cysteine-rich, Antigen5, Pathogenesis-related)-derived	CAPE	4
nd	Glycine Rich Protein	GRP	4
nd	Phospholipase A1	PLA	1
nd	Auxine-responsive protein	(PF04526-DUF)	2
nd	Disease-resistance responsive protein	(PF03018)	3
nd	HR-like lesion-inducing protein	(PF05514)	2
nd	unknown	(PF11938-DUF)	1
nd	unknown	(PF11820-DUF)	1
nd	unknown		2
nd	unknown	(PF06521-PAR1)	4
nd	Ethylene-responsive proteinase inhibitor (PF00280)		1
nd	Peptidoglycan-binding LysM domain-containing protein (PF01476)		2
nd	Metallocarboxypeptidase inhibitor (PF02977) not in Ath		2
nd	proteinase inhibitor, enhance tolerance to oxydative stress (PF02428)		5
nd	PR-4 homolog, hevein/chitinase-like (PF00967-Barwin)		2
nd	chitin-binding protein (PF00187) Solanaceae		1
nd	SAR protein (PF03058) Solanaceae + bacteria		1

nd	unknown A, Solanaceae		3
nd	unknown B, Solanaceae		2
nd	unknown C, Solanum commersonii		2
nd	unknown D, Solanaceae		3
nd	unknown E, Solanaceae		2
nd	unknown F, Solanaceae		1
nd	unknown G, not in Ath		1
nd	unknown H, mainly Solanaceae not in Ath		1
nd	unknown I, Solanaceae		1
nd	unknown J, Solanaceae (size ?)		1

Long or not secreted proteins   42
False positive ORF   16

Legend:
- PTMP
- CRP
- HRGP
- other
- unknown Sol.
- not targeted
- false positive
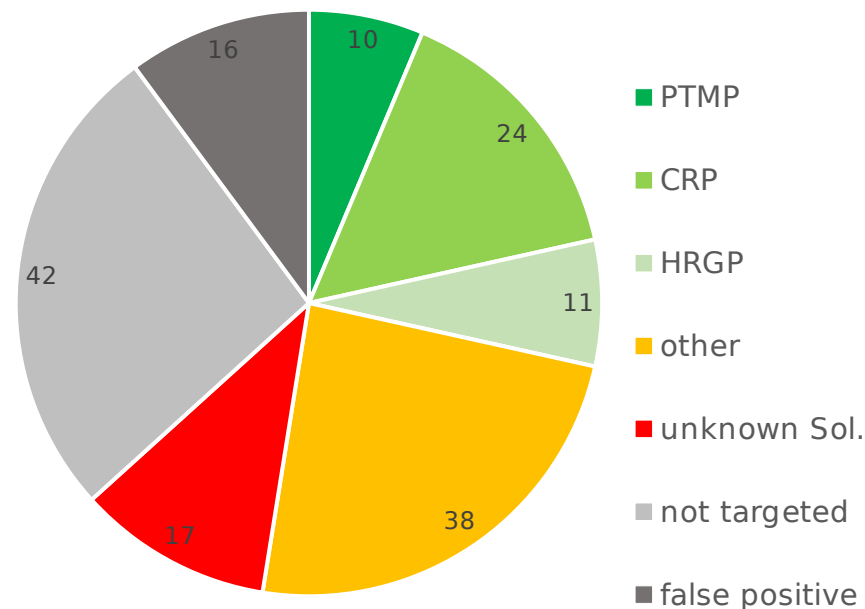
In progress :

- Tests with synthetic predicted peptides (phD, Loup Tran Van Canh)

- Valorization of the ROS transcriptomic analysis of *Solanum nigrum*

Acknowledgements

**Thanks
for your attention**