

Structural and Functional Annotations

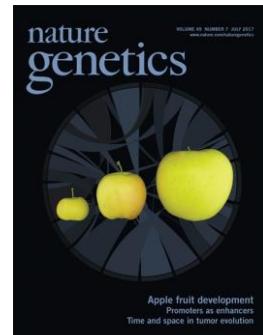
From prediction to validation

2 case stories

200x Illumina + 35x PacBio + 600x Bionano

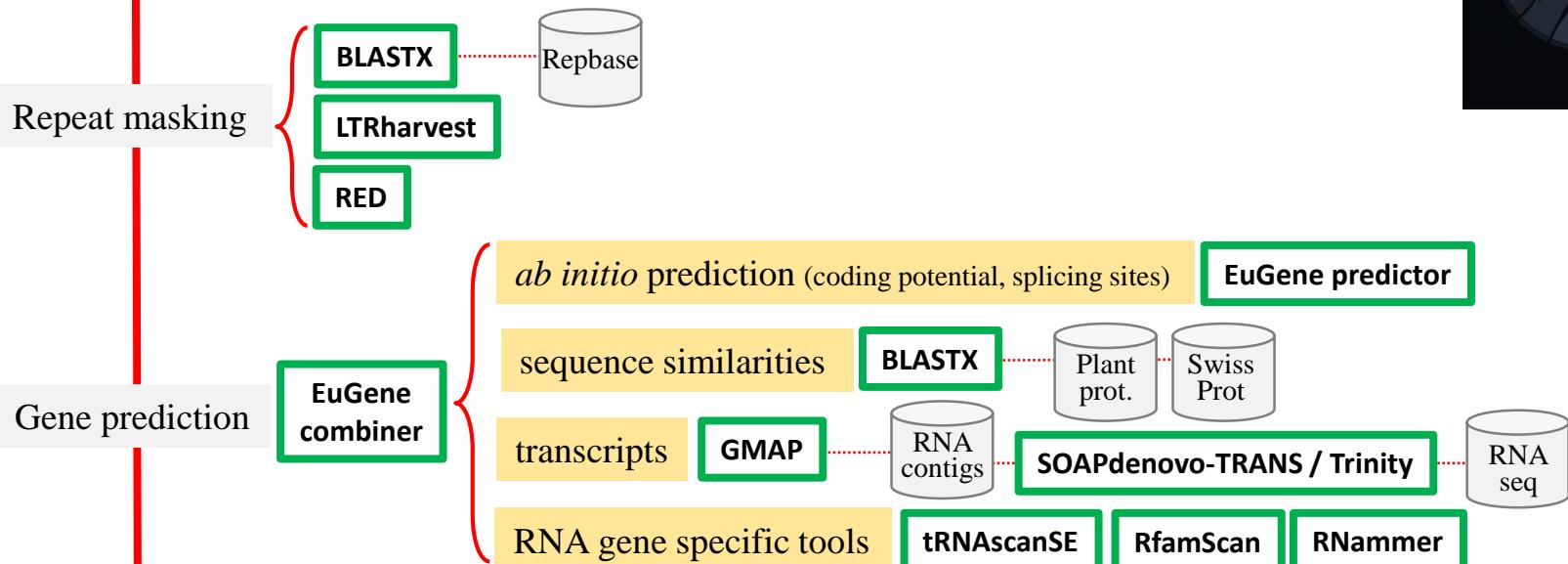
Malus domestica genome

Daccord et al., 2017



2150 contigs / 17 chr. (624,8 Mb)

Structural annotation



Proteins

motif
detection

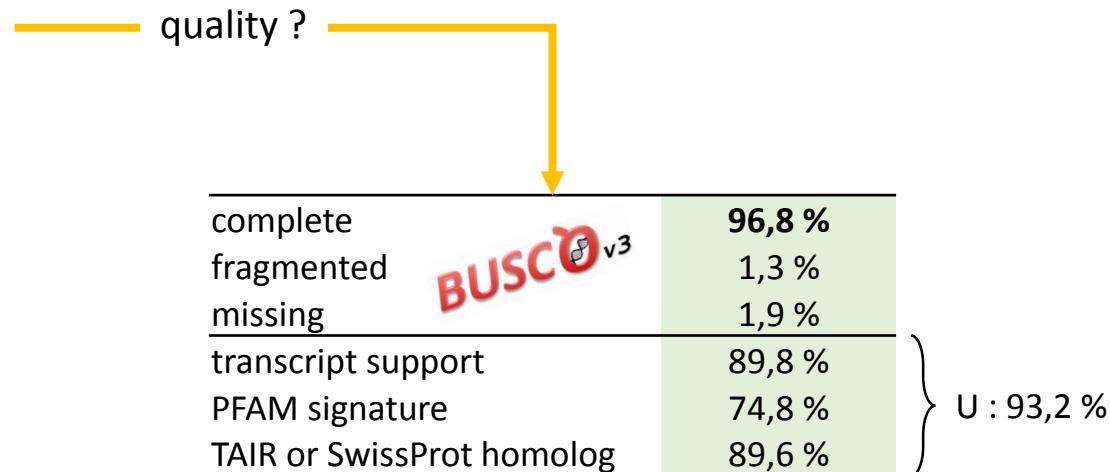
orthology

localization
signal

Inferred functions
Gene families
GO classification

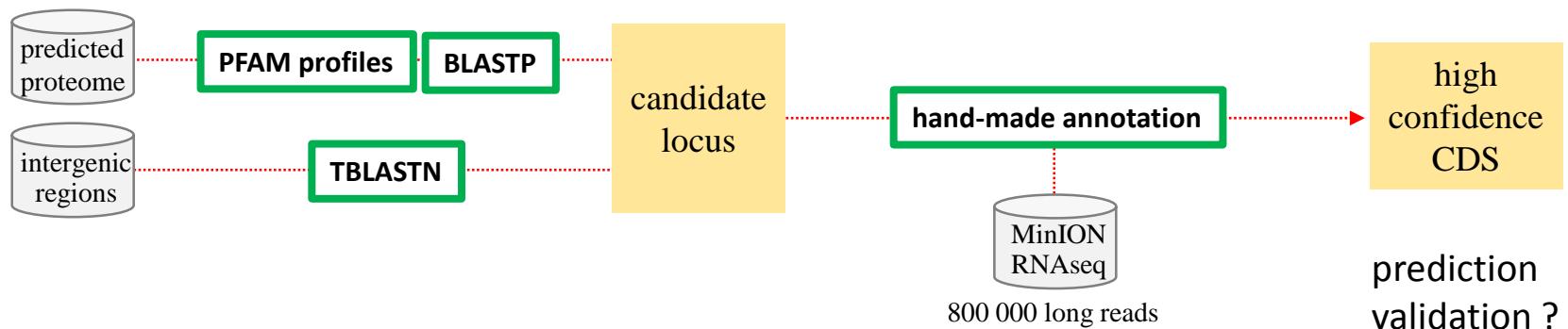
Functional annotation

coding genes	45116
snoRNA	410
rRNA	4369
tRNA	654
miRNA	141
snRNA	181
other ncRNA	54
undefined ncRNA	1816
total genes	52741



One step further...

Gene families of interest (defense, biocontrol...) tagged by at least one published plant member



50 gene families
1194 genes

Gene Families	gene nb	ok	site	ATG	exons	split	merg	underp	pseudo
Cysteine-rich secretory protein (PR-1)	20	18	0	1	1	0	0	0	5
Glycosyl hydrolase family 17 (PR-2)	76	58	0	7	5	4	2	0	15
Barwin domain protein (PR-4)	3	3	0	0	0	0	0	0	3
Thaumatin (PR-5)	35	25	1	2	1	1	5	0	6
Chitinase A/Glycosyl hydrolase family 18 (PR-8)	23	18	0	1	1	0	3	0	7
Protease inhibitor/seed storage/LTP family (PR-14)	27	26	0	0	0	0	0	1	4
Cupin/germin/oxalate oxidase (PR-15, PR-16)	33	29	1	1	2	0	0	0	16
Polyphenol oxidase (PPO)	12	11	0	0	1	0	0	0	6
Hydroxymethylglutaryl-CoA reductase (HMGR)	4	4	0	0	0	0	0	0	11
Farnesyl diphosphate synthase (FDS)	4	2	0	0	2	0	0	0	0
Terpene Synthase (TPS)	32	21	2	1	7	1	0	0	20
Pyridoxal phosphate (PLP)-dependent transferase/Alliinase/CSL	8	8	0	0	0	0	0	0	1
Ascorbate Peroxidase (APOX)	9	5	0	4	0	0	0	0	3
Glutathione S-transferase (GST)	50	42	0	1	2	2	3	0	12
Glucan synthase/Callose synthase	14	7	3	1	1	2	0	0	4
Pectinesterase	70	65	1	1	2	1	0	0	22
Cinnamoyl-CoA reductase/Epimerase (CAD)	55	43	4	2	3	2	1	0	21
Tubulin alpha	10	9	0	0	0	1	0	0	1
Actin	11	2	0	9	0	0	0	0	1

CDS ok	site	ATG	exons	split	merge	underpred	pseudo
753	18	55	54	23	20	32	239
78,8 %	1,9 %	5,8 %	5,7 %	2,4 %	2,1 %	3,4 %	20 %

955 functional genes

Pollen Ole e1/Extensin	31	28	0	1	2	0	0	0	8
Peptidase lyase C	25	24	0	0	1	0	0	0	5
NPR/NIM family	9	7	1	0	0	1	0	0	2
Salicylic acid carboxyl methyltransferase (SAM)	32	24	1	5	2	0	0	0	13
Wall-associated kinase (WAK)	14	11	0	1	1	1	0	0	8
Agglutinin	24	3	0	0	0	0	0	21	3
Translational Elongation Factor 1-alpha	7	4	0	2	1	0	0	0	0
Genes with transcription affected by <i>Venturia inaequalis</i>	10	8	0	1	1	0	0	0	nd
Genes with transcription affected by <i>Envinia amylobova</i>	6	6	0	0	0	0	0	0	nd
Aluminium activated malate transporter (ALMT)	25	22	1	1	0	1	0	0	1
Acetoacetyl-CoA thiolase (AACT)	3	2	0	1	0	0	0	0	0
HMG-CoA synthase (HMGS)	2	2	0	0	0	0	0	0	0
Mevalonate kinase (MVK)	2	2	0	0	0	0	0	0	0
Phosphomevalonate kinase (PMK)	2	2	0	0	0	0	0	0	0
Mevalonate diphosphate decarboxylase (MVD)	2	2	0	0	0	0	0	0	0
Isopentenyl diphosphate isomerase (IDI)	2	2	0	0	0	0	0	0	0
Geranyl diphosphate synthase (GDS)	6	5	0	0	1	0	0	0	2



Analysis of CATMA transcriptome data identifies hundreds of novel functional genes and improves gene models in the *Arabidopsis* genome

Sébastien Aubourg,^{✉1} Marie-Laure Martin-Magniette,^{1,2} Véronique Brunaud,¹ Ludivine Taconnat,¹ Frédérique Bitton,¹ Sandrine Balzergue,¹ Pauline E Jullien,³ Mathieu Ingouff,³ Vincent Thareau,⁴ Thomas Schiex,⁵ Alain Lecharny,^{1,4} and Jean-Pierre Renou^{✉1}

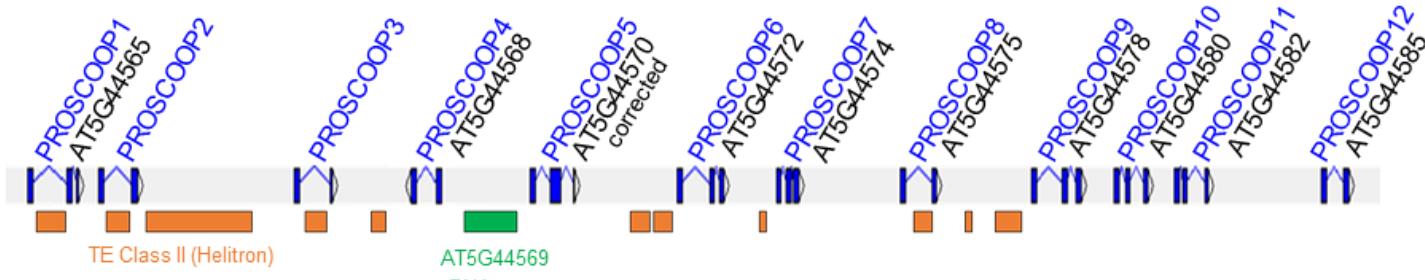
[BMC Genomics](#). 2007; 8: 401.

Focus on a small gene of **unknown function**, strongly induced by the necrogenic bacteria *Erwinia amylovora*

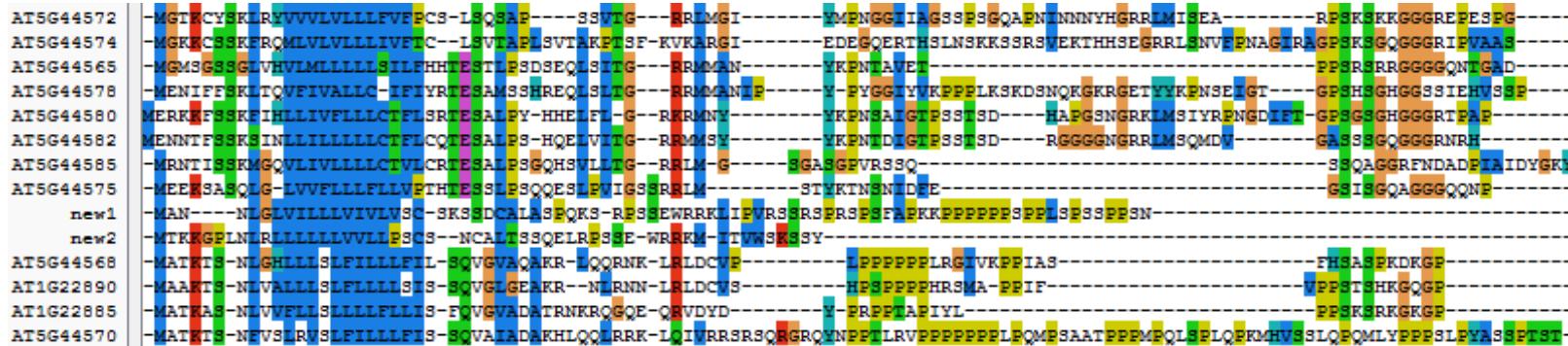
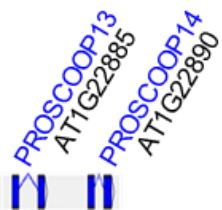


Fireblight on Rosaceae

Chromosome 5

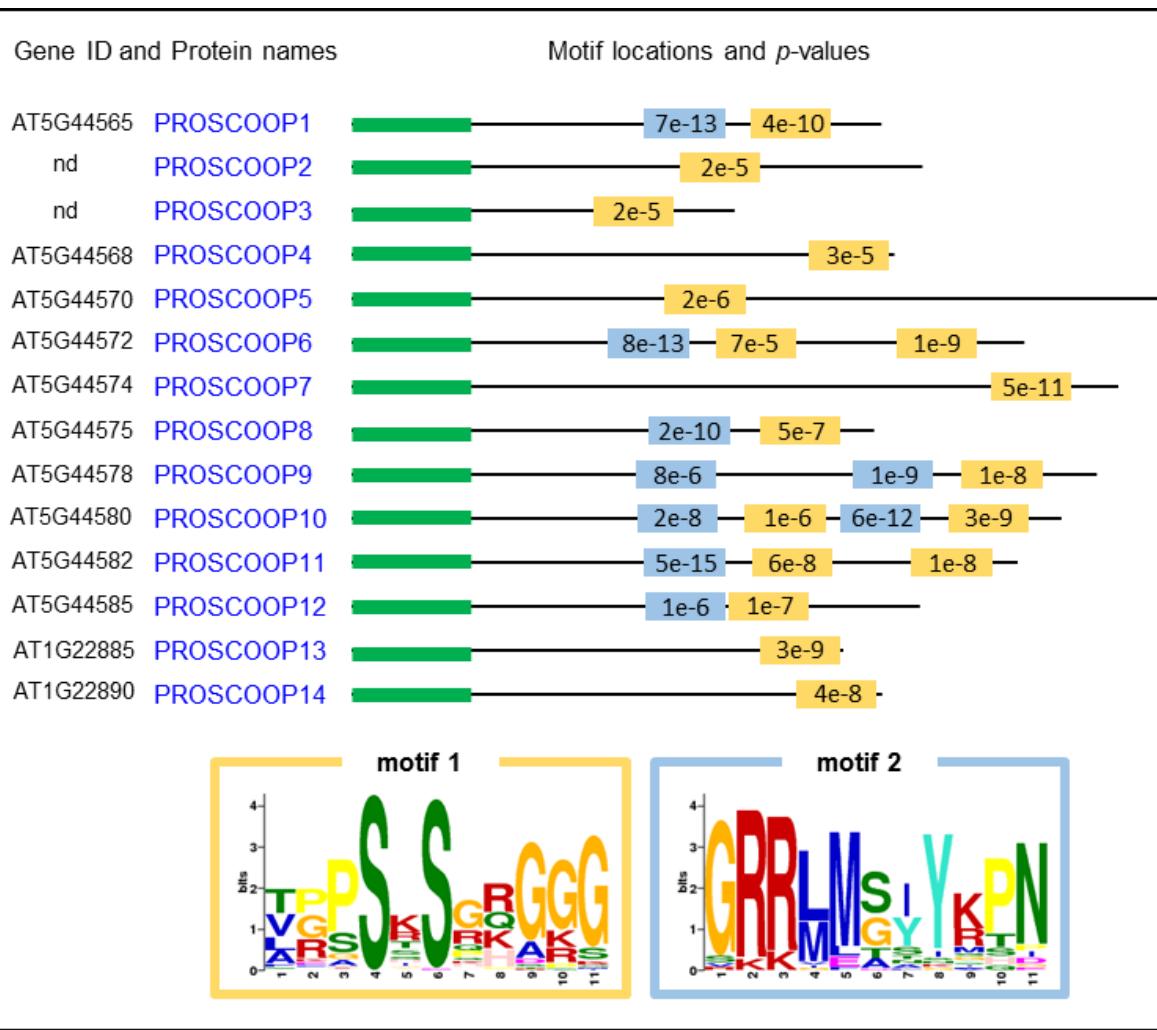


Chromosome 1



MEME

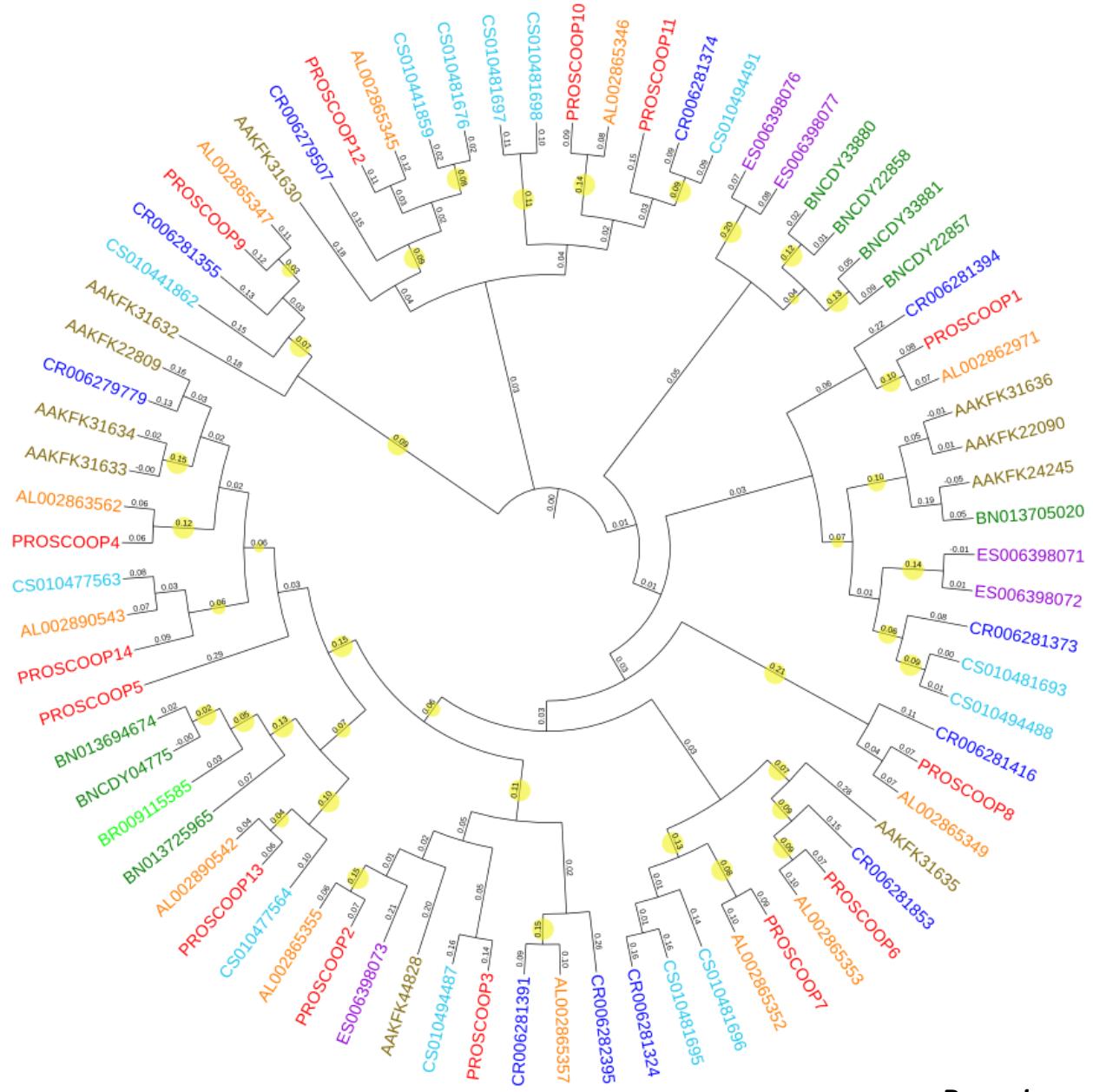
Motif searching



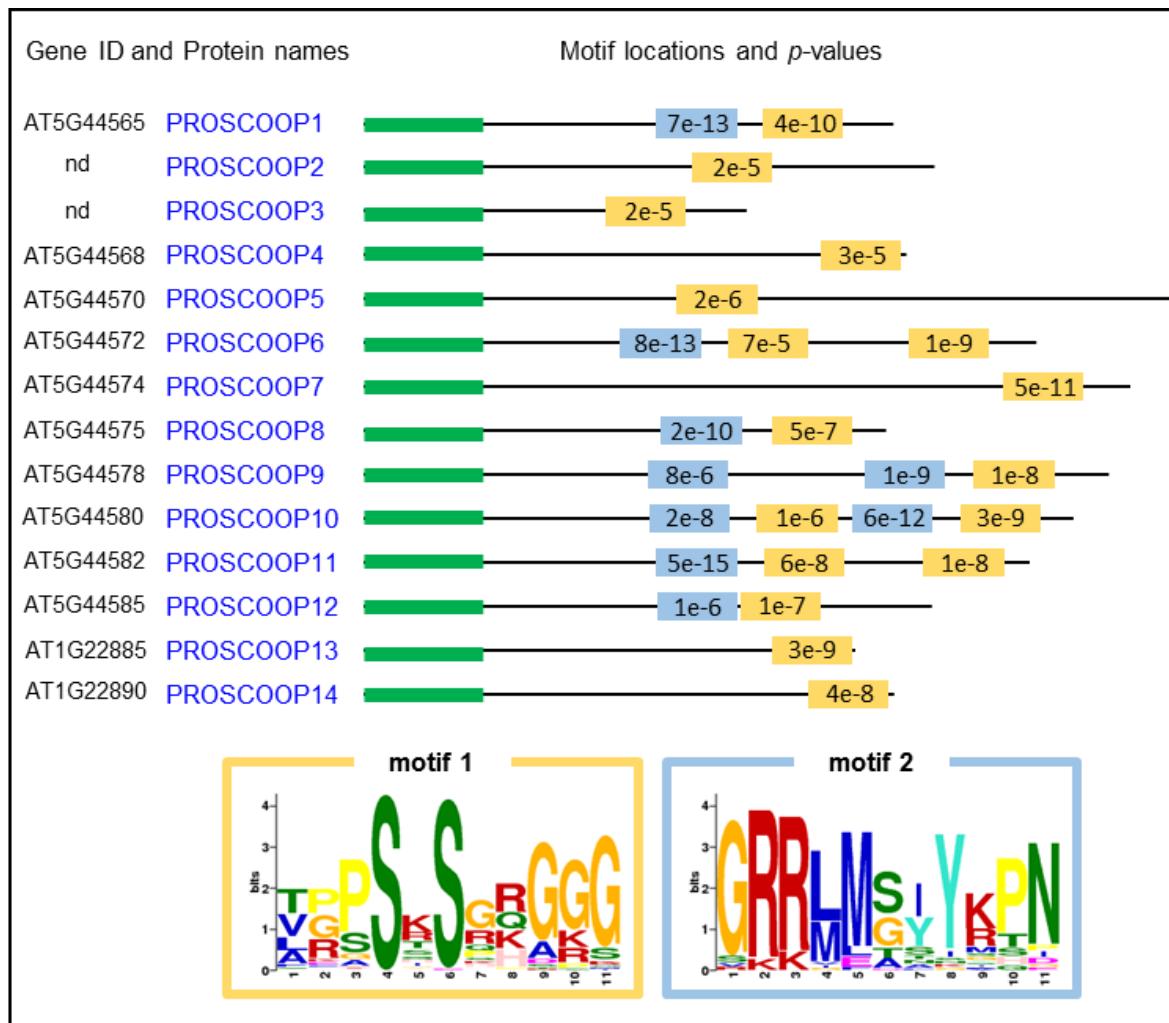
MAST

Uniprot screening

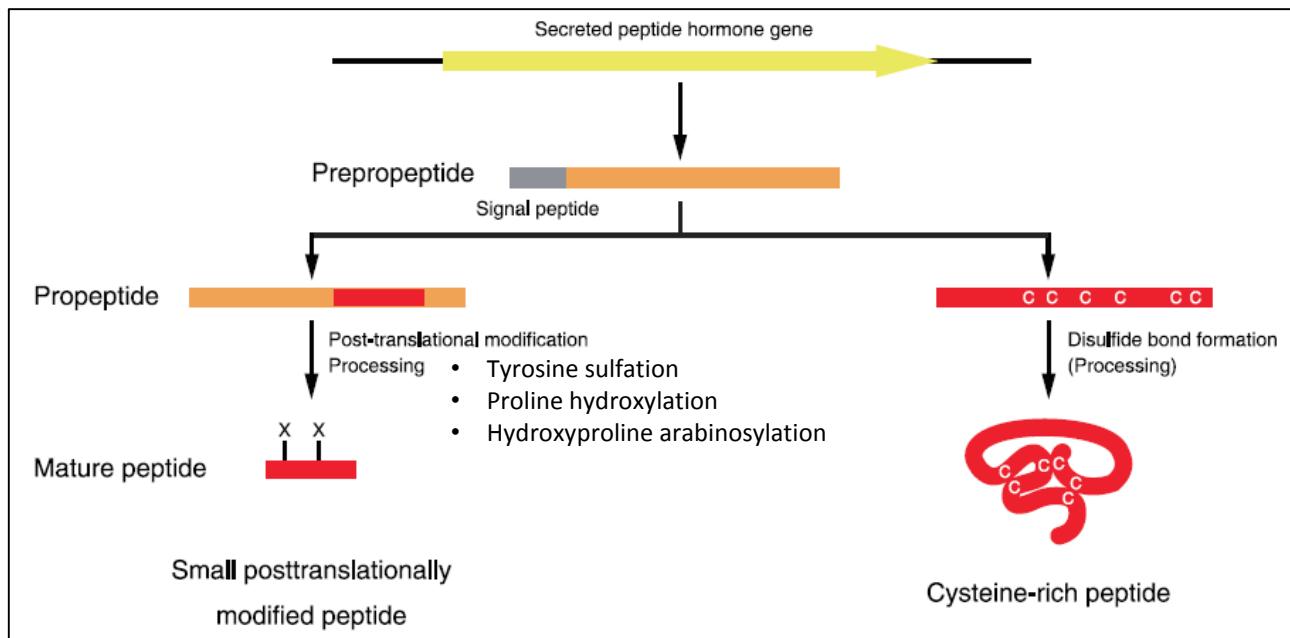
Arabidopsis thaliana
Arabidopsis lyrata
Arabis alpina
Capsella rubella
Camelina sativa
Eutrema salsugineum
Brassica napus
Brassica rapa



Brassicaceae specific



SCOOP motif Serine riCh endOgenOus Peptide



from Matsubayashi, 2011

Structural features of small post-translationally modified secreted peptides in plants :

- ✓ Multiple paralogous genes (from 3 to 30)
- ✓ Encode 70 to 120 aa pre-proptides
- ✓ Contain N-terminal secretion signal (RE)
- ✓ No or few Cysteine residues
- ✗ Sequence diversity excepted in the C-terminus corresponding to 5 to 15 aa mature peptide

Phytosulfokines

cell elongation, growth, development
pollen tube growth, vascularisation
attenuate pattern-triggered immunity

B CLV3/CLE	
CLV3	----TDSKSLLLLFCFLDTHADL-----TQAHAHVQGLSNRKMMMKMSEVNGANCBAEAKTKG-LGLH-EE-RVPSCDPLHN-HNNPQRQPRNIFQLP
CLE48	MAAATKXGKSVFIV-L---SSS-AHSS-TKSFF-L---GETQ-DTK-A-MKK-E-KIDG-GTANEEVE-RVPVGSPLRNKRH-KPTP-----
CLE1	MAN---LKFLL-CIICVLSRSS-SASPV-----FPNADIGKR-BHM-IE-AEVLA-SM-EKLIMGRFNESM-----SPGGPPDPH
CLE2	MAKLSFTFDFLT-L---SSIAAGSPL-----EGARV-GVKVRGLPSI-----EATSPTEWDDQAGSH--GKS-P-----SPGGPPDPH
CLE3	MAS---LKLMLVCVLLLVLLTSLVH-ECPVLA-EERFSGSSRLKKIREL-FERLKE-WKGRSEGEETI-LGNTL-DSK-----SPGGPPDPH
CLE4	MAS---LKLMLVCVLLLV-LEF-SVH-QVLPV-AEESPDSGMRNKRMLR-LKRE-----LKVRSKDGQTVLG-TL-DSK-----SPGGPPDPH
CLE5	MATLILKOTE-I-IIIILISLTLSSQRLRSLRSYRAVSGNMDSQVLLHEGLD-LSK-FKGHNE---RRFLVSS-----D-----SPGGPPDPH
CLE6	MANLILKOSL-I-IIIILISLTFELSSQRLRSLRPTGMGDSQVLLRELGD-LSK-FKGQDE---RRFLVSD-----D-----SPGGPPDPH
CLE7	MAS---KALL-LFVMLT-F-LVVIEMEGRIVR-NSKTKDGSNDLKLRLGVY-VSE-LRIG---RELSONV-EVD-----SPGGPPDPH-HSYPLSSKPR
CLE8	-----MVKVLKRD5-MLU-ITLY-FL---TT5-M-AR-0-D---P---=L-VLGVEKD-VVPA
CLE9	-----MTM-THLNR-L-ILISLFLVSSLSSSTASVTTVDEGNRSTSRNFYRVRPFVPREHNHPHYHT-PHRSCDFI-RPYA-RS-MCIELQDHRSSRVA
CLE10	-----MKTNRNPINIL-VFF---LL-TTARA-T-R-NMT---N-R---THRTPVK-QHAY-YAHPRSCEFSV-PYRA-RS-MCIELQDHRSSRVA
CLE11	-----MTKPKPCPSFL-FHISLSSALF-V-F-L-LISP-AFTTS-YLKLGSI-NSS---L-LGK---I-LASN-FDFTPFL-KNKORTQRQD
CLE12	MLR15SSSSMAL-KFSQIL-FIVLNLSSLF-L---LHHNLYSLNFR-RLYLSNAV-EPSL-LKQ-HYSRVSLS-R-SK-VW-LSDR-FDFTPFLRSRSRHSN-HR
CLE13	-----MATTRVSHVVLGF-LWISLILF-V-SIG-LFGF-NFSSK-----PINFPPSVPTW-LA-YL-YR-G-RR-----LA-VLKFTFDPTFL-KDLRSR-N
CLE14	-----MKVWSQRLSLF1IWIFI1LAGHLHSSAGRKLPMSMTTIEFQRLSD
CLE16	M-EA-CSKHK-R-RIRRAYTTSTTG-Y-A-AVVF-C---GTF-V-FAQFGTSS5-A-LFA-PHOYPLSPRKAHGIF-----HEMASFPAK-KATVS-F-TGQ
CLE17	MTHVLRQGQGKRR-NHDVNMTCF-F-LFF---F-VFYVS-F-QIYLSS5-A---S-V-Y-S-R-V-H---AVSPP---P-PKPRRA-L-Y-STAP-FRGQ
CLE19	-----MJKGLMZLASSLTLIAFIHQSESASMRSLLNNGNSYEEEQVQLKYOSH
CLE20	-----MKHNKNHNPSPRPLCLLIVF1LWV1LSSKASRHIWERRRSPSKS-GENREFLPSQFPTV
CLE21	ML-1LSSR-Y-AMRKDVL-I-IVI-FTVLVLLIIISRSSSIQAGRFTMGRNRLS-A-ARSLYNN-HIKWV-Y---T-EMSENFKRVRSSR-S-L-R-R-K
CLE22	MGN-YSSR-R-KSRKHIT-TVALI-L-LL-LFF-----L-F-L-YAKAS-SSS-PN1HHSHTH-GSL-KKSGNLDPKLHLDLSNA-ASSRGSS---Y-TN---YEG
CLE25	-----MGCGNTRAL---YVGTAI-S-GT-V---I-VG1-LANS-APS---S-S---PE-NV4-TL-R-F-G-X
CLE26	-----MRNNR-SLQLWTF---FTWVG---LW-TOAD-----VW-NON-KEDOTKETTIAVWNMNSD4-----QZQLEQDLSR
CLE45	MLGSSTRSMFFLLVCGILLADNRVNNSMRHREFFI-KETQAEGV-QTEETKUS5-TGVO-QHITEDQEMLKJ-KNRVLEEVN-KD-KIKE-EQ-ERKNN
TDIF/CLE41	MATSNDQTNIKSS---SRTL---L-FEFSLL-BFSSL-TIPMTR---D---STSM-VAPFKRVE-LESVPMASSTMDLRS---KASTRS-R-TS---R-KR-EF-GND
CLE42	-----MRSFT---TISUWF-E-FNL---IITGHTRDIQD---DIEGSNVH---D-MAVT-E-GEKRR-E-FRV-----R-FRT-M-TLIGKKG-I-BAN
TDIF/CLE44	MATTBOTSIS-CFHVQVIR---ITIPLF---E-ICP-TSSNN---D---HESKNTKMPASKRFI-LTOPSTSSTMKRN---TAH-RPSSGTS5SARBRERFAE

CLE family

shoot/floral meristem development
root apical meristem development
vascular cell differentiation

C	RGF	<pre> MVSIRVICYLLVFSVLQVHAKVSANFNQAPQMKNSEGLGASNGTQIAKKHAEDVIENRKLTKHVNVKVEANEKNGLEIESKEMVKKRKNKRLTKTESLTADYDNPQHHIPRE MTNITFLSCLLFLNLLFLCFSPLSFGHDEKDLVSDVGSNAKYMKHLGDNDAAKQVWRGR--SQEFSKETTMMH--KTTKKTETM--VEEDD--LVAY-TADYPRKPHPPXK MTL-SKILCVLIIQLCSFRYSLDHGQNQSRDFVTAIKAHY--GD-VNKMH-TRGRKLMMASEKEAEATMGRKRNEERTNSK--SEVEDD--LVAY-TADYPRKPHPPXK MRFTIIIVIAFLIIQSLLEEQILVYARKGREACHKSLSDYQGQDQSSTLHPKELYDAVPKRVRGRATRAEKFQVATMNMDSWSKISGASKHLIVERKLGHFRKSSSFKWPKKKKSSGSPFVAFYD-YDRGPRHPPR MSSHVASHASILLFL-LFLH-HSDSRHLNHVII-TATRSFLSKDQNVSSSTSKEPVKVSRFPV-PGPKLHH-RRHPPLI-FADYPKPSTRPRB MCSLRISGLVITYVCFK--LSSSVNGCASAFLRSRHKHVVAS-BDFVNGGERRALGVTGETEVWMV-PDOPHRKPTTNEKS-B MEMKKMSYANLTILLFLFLLAFQGGSRDDQHMMVTRKDTDSMGRKLKSLKPINTPKHNGKFEYDQGSHDQEYREVELYD-GORKYKPVW -MKLIRVTLLCALAILLVLVTPS-SQLQ-KHPYSSPSQQLSKVKTVM-ATRKLMIZIESEPTMSHESGGSEQLRVTRSSGSKDEEAKLSELSEEEAKALAKYLSM-DTRTRRRRNNV-KALPLDP -MARVSHKSFLVALLLFISSPTQARSRL-EWVRNRNTLVEWKSQRKTRHGGESSDVGMLDN-ANSANKRDI-NR-7 </pre>	13
	RGF1		13
	RGF2		13
	RGF3		13
	RGF4		13
	RGF5		13
	RGF6		13
	RGF7		13
	RGF8		13
	RGF9		13

Root meristem Growth Factor

root development interact with auxin

D	IDA	MAPCRTMVLFVFLAASSSVAARICAT-H---EM-----KKNIKR-LTFKNNSHTI---YLPKGVPDPAASPKRHNFSVLSPH-----	77
IDL1		-----MNLSHKTMFMTL-YIVFLIFGYSNATARIQPIKLSETIEVQTRSQRQEIGGFTFK-GRV---HSFSKRVLV-----PSPSIRHNSVNNLKH-----	86
IDL2		MSSRNQRSRITSSSSVFFRTTRILLLILLLFGCNGARTNVRNSKPKHHDAVSSS-----TK---QI---LPRLPHFVP-----ASPSPKRHNIDIGLSSWHRSSP-----	95
IDL3		MSSRSRHSR-----KYQLTR-TIPILVLLVLLSCCNGART-TNWFNTSPKQKDVSPPHDHVHHQVQDKHPSLGPRLPQFVP-----ASPSPKRHNIEGLSSTTKT-----	99
IDL4		-----HYTPRFLWPRRLRRLSPLFLLCPIIHCNDRACTSSSSVYFRNPNNHESNSNTVRGHLRPLRHLPSAPASPSPKRHNIDIGLQKALLSP-----	93
IDL5		-----MGNKRIKAMLLVMVMMVESWRCIEADSLRFLYSSSPLFOPFVKWRPNPNRHIIQNOGNCNDPYYTPESSGF-LPLTLRPLBASPSPKRHNLYGOSTSNHCRP-----	103

Inflorescence Deficient in Abscission

floral abscission
cell wall, lateral roots

E	PSY1	MIFVW-[D]-L-VCI-[T]ITTSRARNPVSVSGFENGSQFQRSLVNVFDY[DPSANPKHDPV] MIFGTT-[D]-L-ICIDLTPLVTSPPNTLQSQ-IVKTFQTSRFLMTIEDY[DPSANTRHDPV] MFYSSSSS[IGC]L[FIT]FALLS-SARLSLSEN-EMTVGPERSLVSTNDY[DPTANGRHDPV]
	At3g47295	
	At2g29955	

PSY family

cellular proliferation
regulate plant defense responses

F CEP1

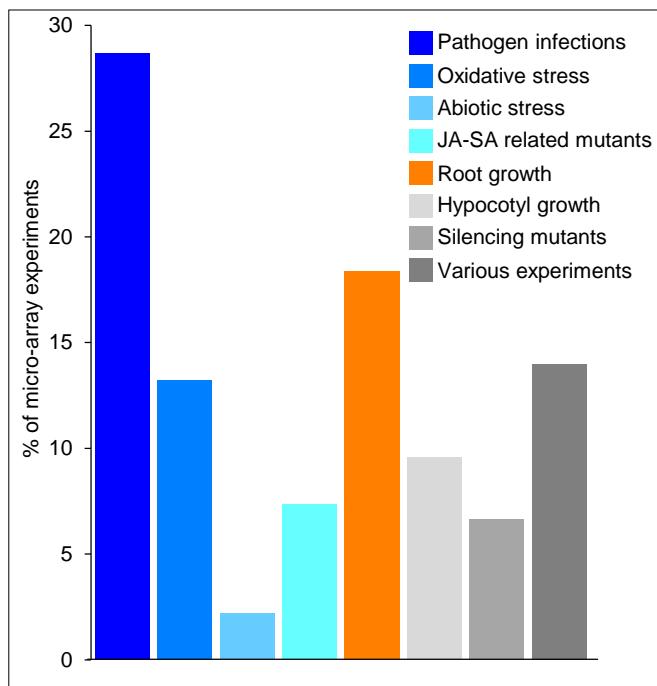
CEP1	-----MGMSNRSVSTSTF-----	-----ALAVLVLH-IDQTEER-----	-----KTTSLIELEGIKY-KTEAE-HPS-----	-----IVVTYTRRGV-----	-----QKEVIAHPT-----	-----FP-----	-----P-----	-----G-----	-----S-----	-----P-----	-----V-----	-----H-----	-----N-----	-----GRH-----
At1g59835	MKLFIITVLTISRVFQDTPATTA	KSKMVGHEHNFNLDDPF	GHTFGCFFKVLKKLGGDE	NLNKRNIFAPTFEDS	GCHGPVRNLKNTFD	P-----P-----	-----P-----	-----C-----	-----G-----	-----C-----	-----G-----	-----P-----	-----C-----	-----P-----
At2g23446	-----MATIN-VYY-FAF-----	-----TTS-V-G-S-----	-----IEGRKLT-----	-----K-----	-----TV-T-TSE-----	-----ERAGGSVLSSPPPEP-----	-----LESPP-----	-----S-----	-----G-----	-----D-----	-----P-----	-----P-----	-----C-----	-----G-----
At2g35612	-----MSVRSCLVTF-----	-----T-----	-----F-----	-----FLF-----	-----L-----	-----V-----	-----V-----	-----I-----	-----V-----	-----H-----	-----V-----	-----V-----	-----G-----	-----H-----
At5g6815	-----MESFMGOKTLYIACY-FMLM-----	-----V-----	-----E-----	-----G-----	-----F-----	-----G-----	-----V-----	-----G-----	-----R-----	-----I-----	-----R-----	-----I-----	-----M-----	-----K-----

Cter-Encoded Peptide family

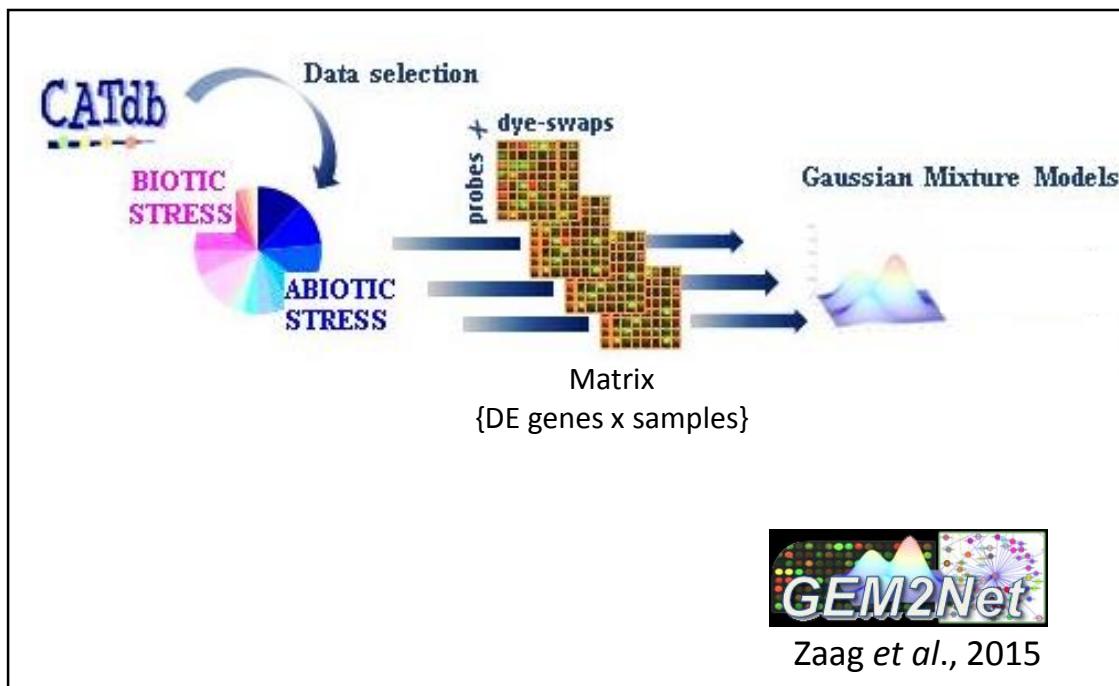
root development

Transcriptome mining

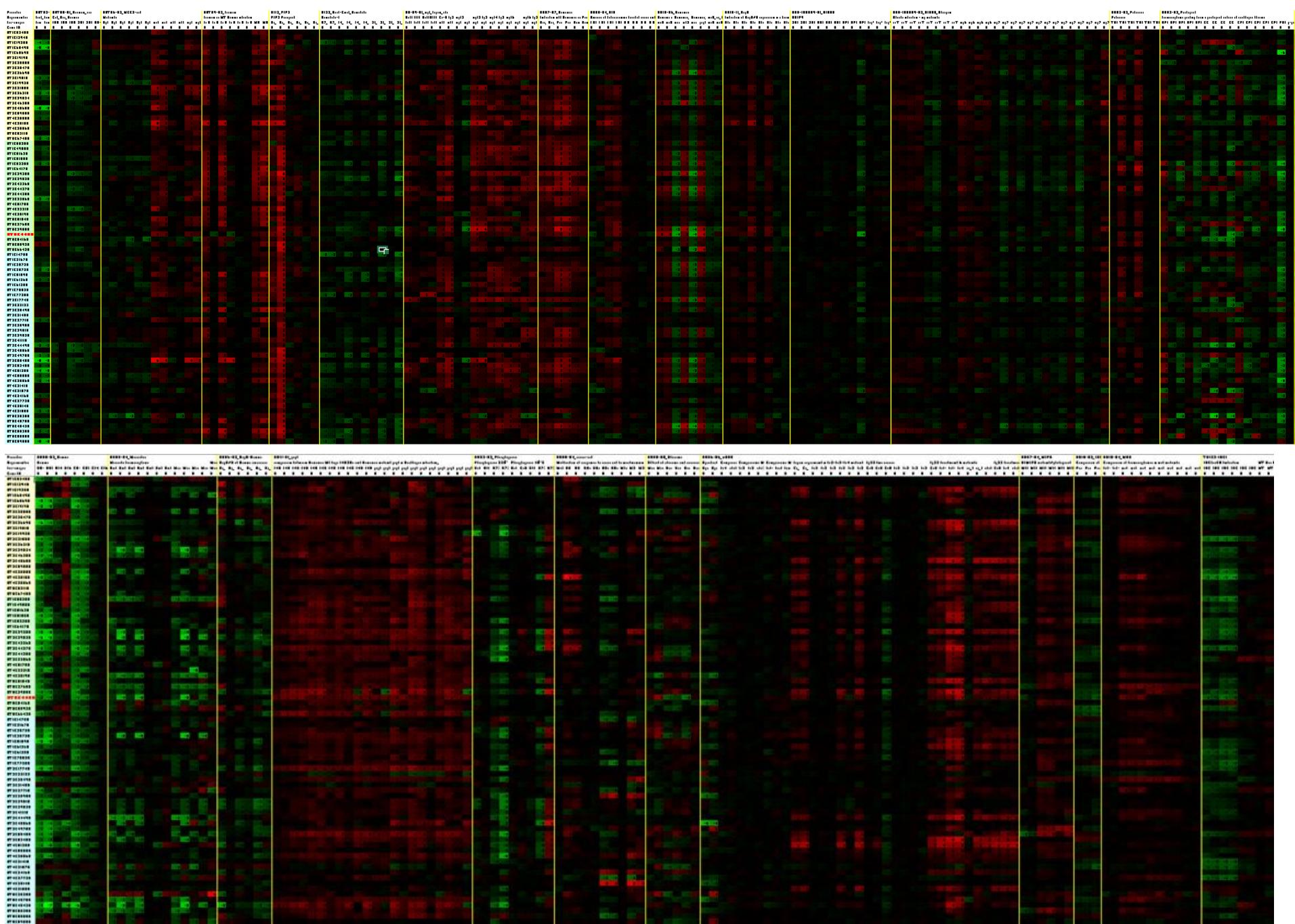
Differential expression - Metadata

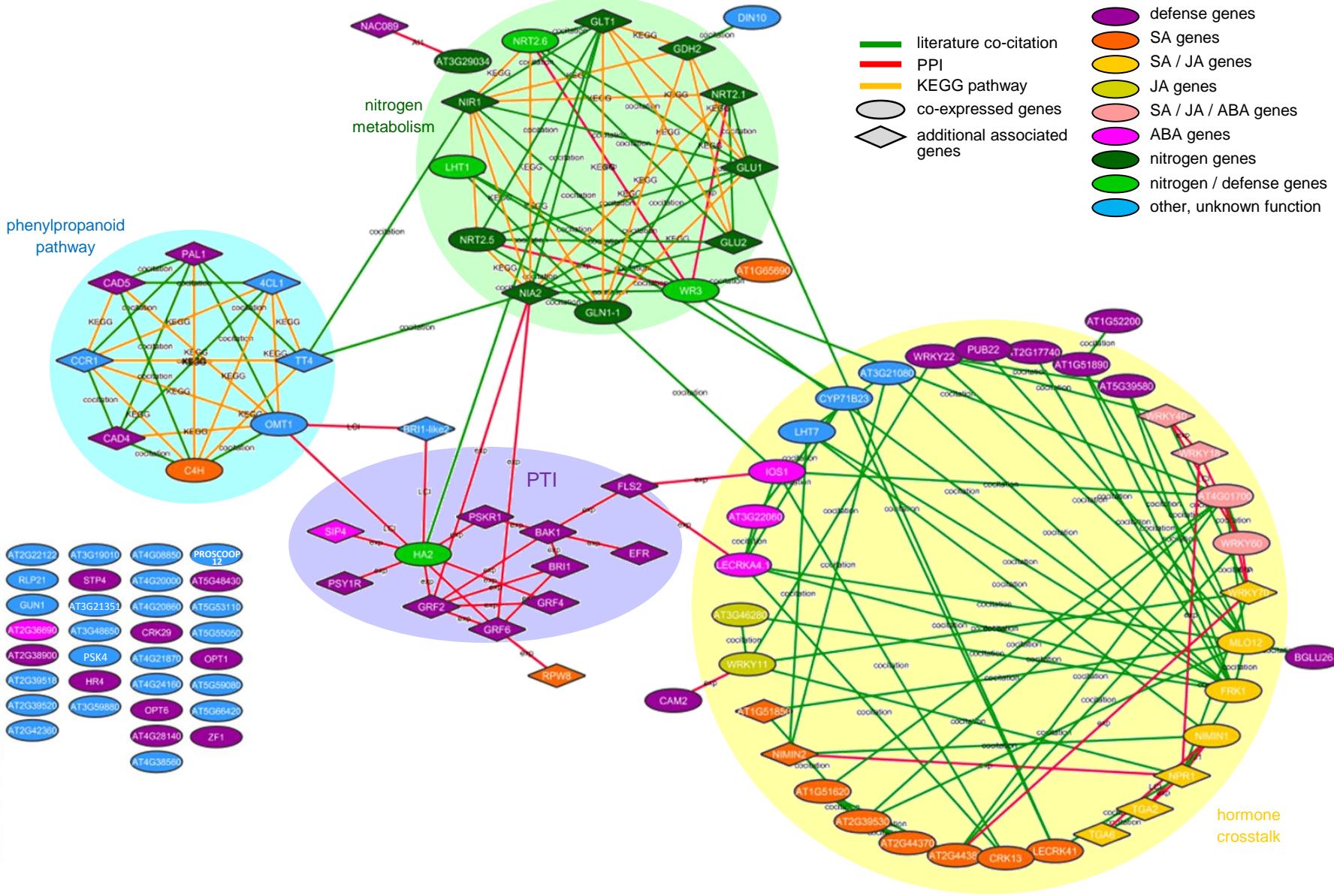


Clustering



83 genes co-regulated with *PROSCOOP12* (response to necrotrophic bacteria and fungi)

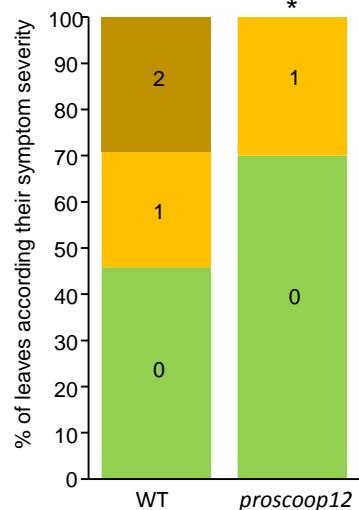




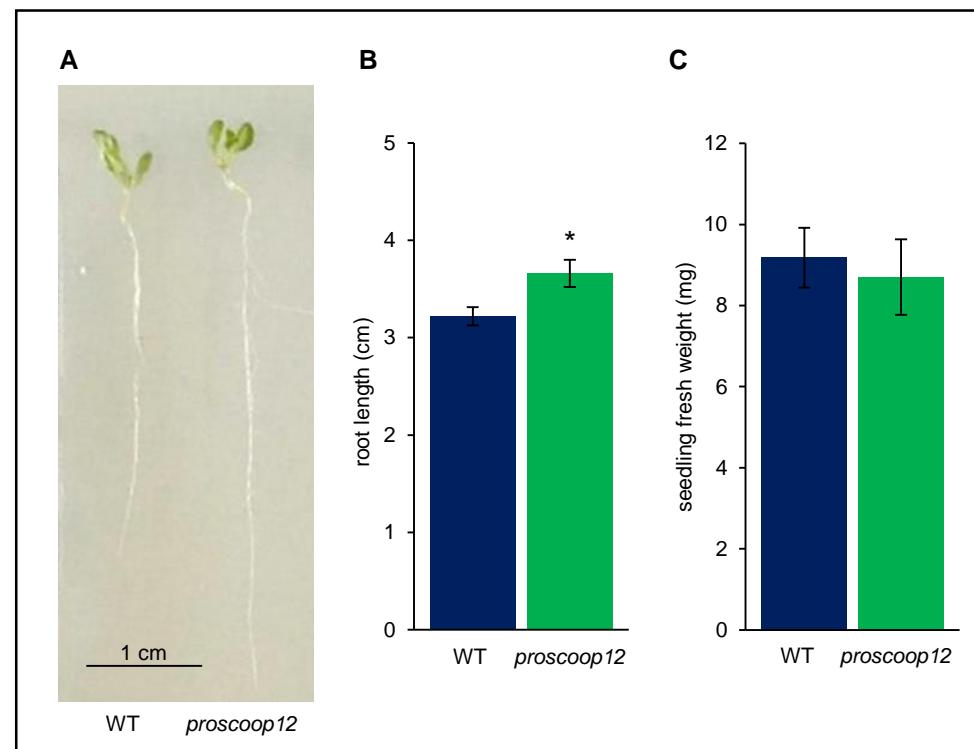
Relational annotation of *PROSCOOP12*
117 genes

Validation 1 : mutant phenotyping (T-DNA and CRISPR-Cas9 lines)

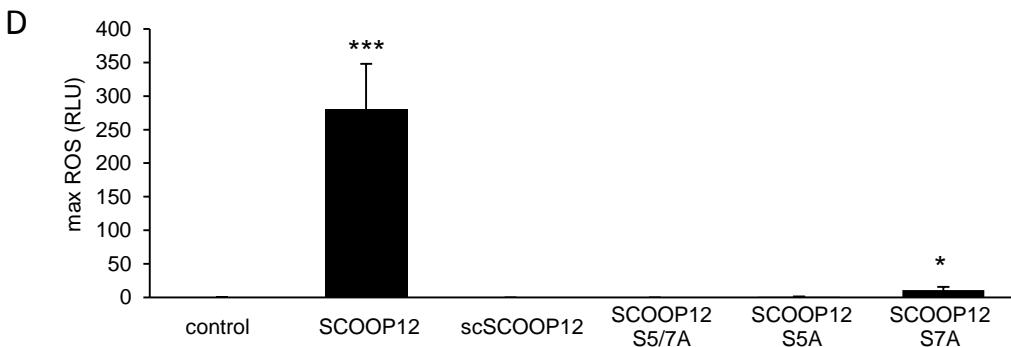
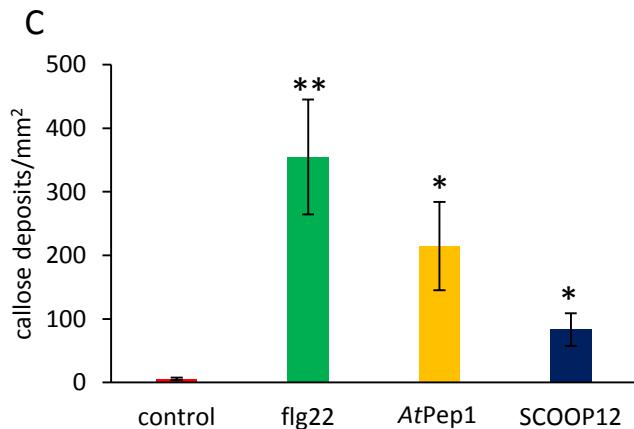
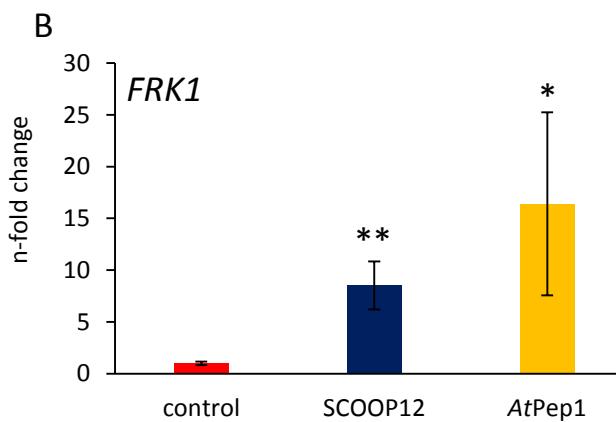
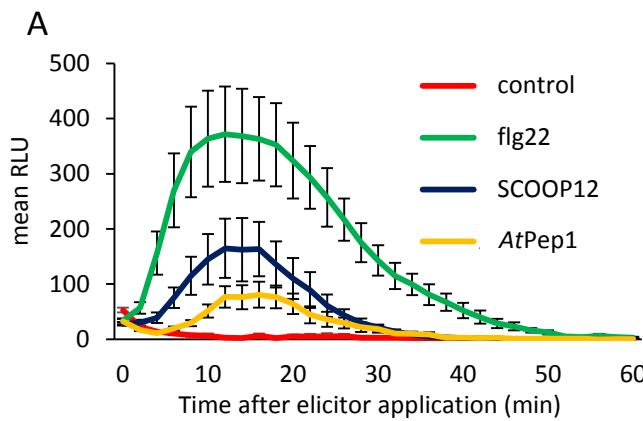
Erwinia amylovora infection



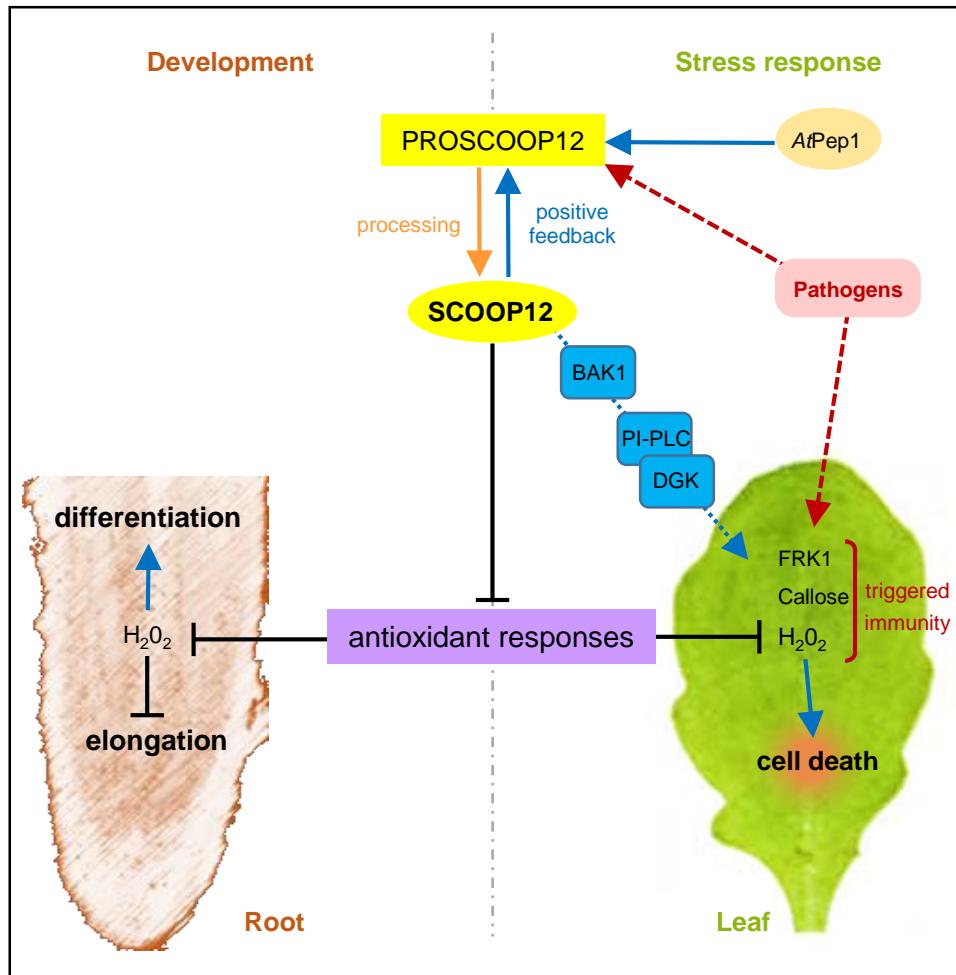
root development



Validation 2 : synthetic peptide



Functional model



The SCOOP12 peptide regulates defense response and root elongation in *Arabidopsis thaliana*.

K. Gully, S. Pelletier, M.-C. Guillou, M. Ferrand, S. Aligon, I. Pokotylo, A. Perrin, E. Vergne, M. Fagard, E. Ruelland, P. Grappin, E. Bucher, J.-P. Renou, S. Aubourg. J. Exp. Bot. (in press)

Acknowledgements



Nicolas Daccord (PhD student)
Florian Barbazange (M2 student)
Sylvain Gaillard
Charles-Eric Durel
Etienne Bucher

Jérôme Gouzy (LIPM, Toulouse)



Kay Gully (PhD student)
Sandra Pelletier
Marie-Charlotte Guillou (M2 student)
Philippe Grappin
Jean-Pierre Renou

Eric Ruelland (iEES, Paris)

Mathilde Fagard (IJPB, Versailles)

