05/15/2019

Understanding sorghum cell wall deposition using comparative genomics and gene coexpression network

Lauriane Hennet, PhD candidate
12èmes journées du Réseau Français des Parois

Réseau Français des Parois

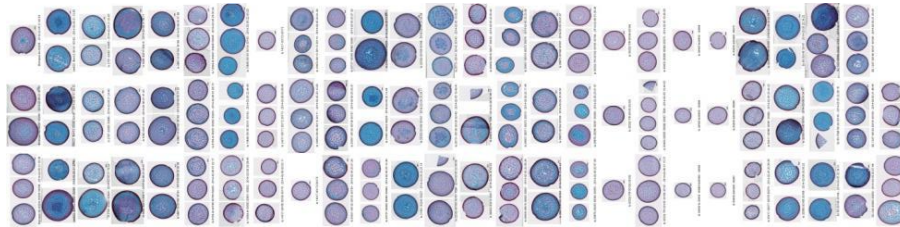Surpervisors: Nancy Terrier, David Pot, Delphine Luquet, Hervé Etienne
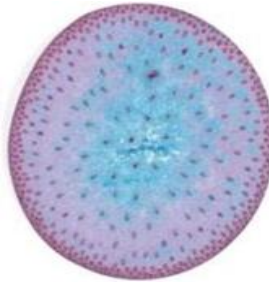
Secondary Cell Walls are target for breeding new sorghum lines

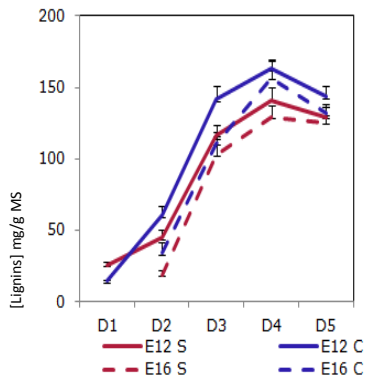## Sorghum stem

Secondary cell wall (SCW)

- Cellulose
- Xylan
- Lignin

**Biomass
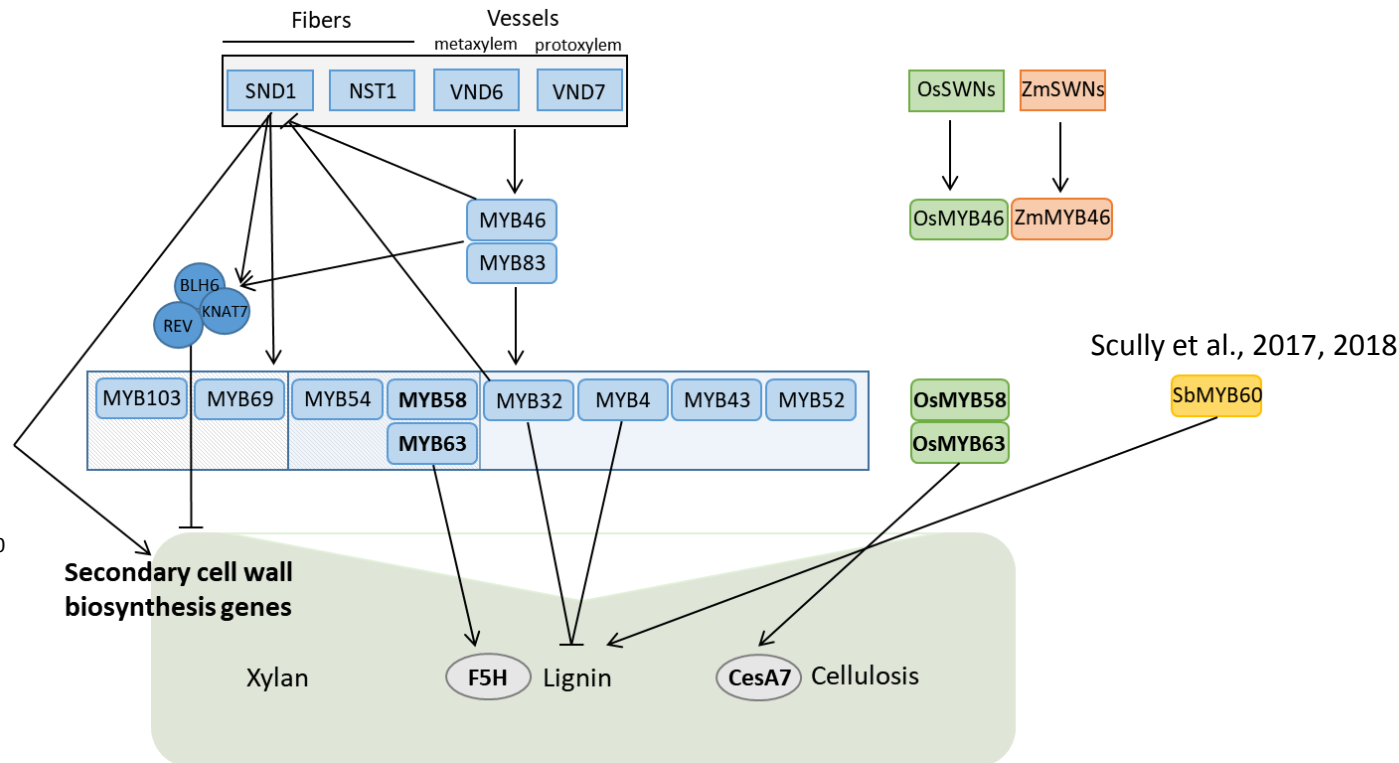Tolerance
Stiffness**

**Human and animal nutrition
Agromaterials**

McKinley 2017
Liu 2018
Perrier 2018

Secondary Cell Walls are target for breeding new sorghum lines

- Accumulation pattern in sorghum
- Biosynthesis enzymes well listed in sorghum   Rai et al., 2016
- Transcription factors mostly known in Arabidopsis   McKinley et al., 2016

Scully et al., 2017, 2018

D'après

McCarty et al. 2009
Zhong et al. 2007, 2008, 2010
Ko et al. 2009
Yamaguchi et al. 2011
Soyano et al. 2008

Fornale et at. 2006
Sonbol et al. 2009
Yoshida et al. 2013
Kubo et al. 2005
Mitsuda et al. 2005

Using gene co-expression network to identify candidate genes

Transcriptomics data sampled during development

Building a gene coexpression network

- Build a gene-gene similarity network
- Divide network into groups with similar expression
- Identify drivers genes and groups

1 Analysis of co expressed gene groups
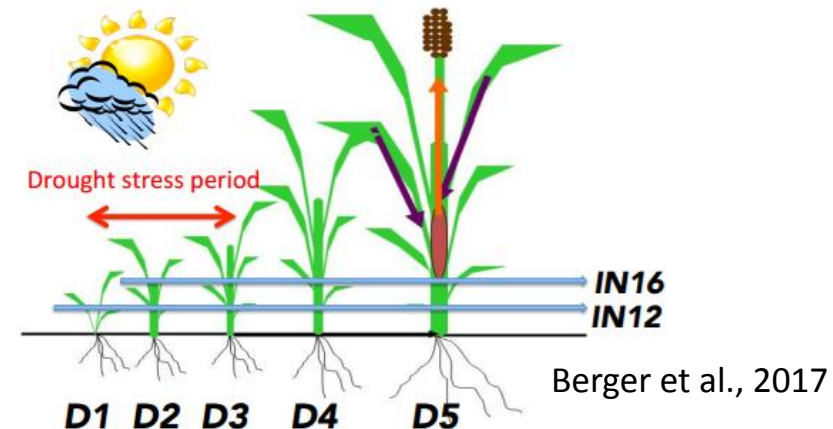- Enrichment
- Top genes

2 Use of bait genes known for their role in SCW

20.000 genes expressed throughout the development

- ## Sampled during 3 years of experiment in field
  - From the 1st month to maturity
  - 2, 3 and 5 dates
  - 2 to 4 internode ages
  - 200 samples x2 tech rep
  - Water stress and control conditions

2013 set-up

Berger et al., 2017

- 34000 genes with around 20000 genes expressed

  ➢ Different set-ups usable with co-expression network approach

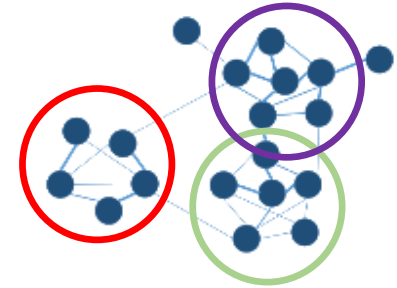# 12.000 genes gathered in 30 co-expressed genes groups

- WGCNA R package = Weighted Gene Coexpression Network Analysis

- Use of only the strongest links = 12.000 genes
- 30 modules or groups of co-expressed genes
    - Named by color
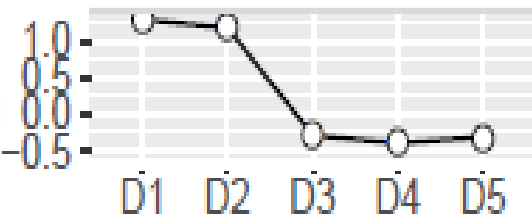   - Containing from a few genes to 1500+ genes
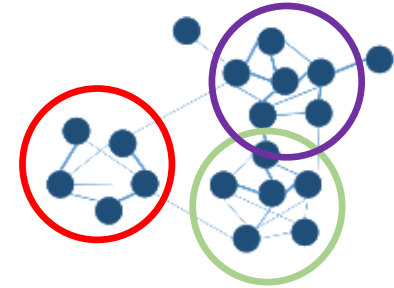
Langfelder P and Horvath S, 2008

12.000 genes gathered in 30 co-expressed genes groups

- WGCNA R package = Weighted Gene Coexpression Network Analysis

- Use of only the strongest links = 12.000 genes
- 30 modules or groups of co-expressed genes
  - Named by color
  - Containing from a few genes to 1500+ genes
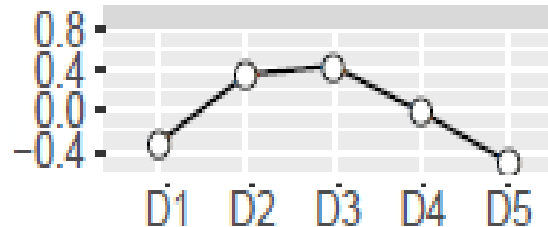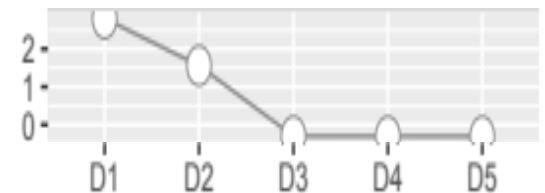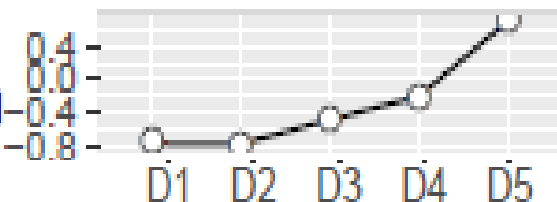
- Representative expression patterns

Langfelder P and Horvath S, 2008

Blue

Yellow

Turquoise

Brown

Black

Purple

12.000 genes gathered in 30 co-expressed genes groups

- WGCNA R package = Weighted Gene Coexpression Network Analysis

- Use of only the strongest links = 12.000 genes
- 30 modules or groups of co-expressed genes
  - Named by color
  - Containing from a few genes to 1500+ genes

- Representative expression patterns

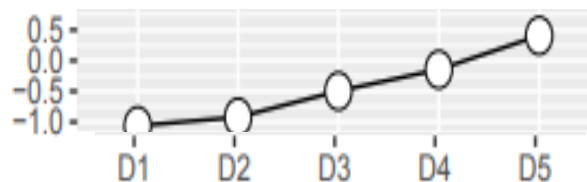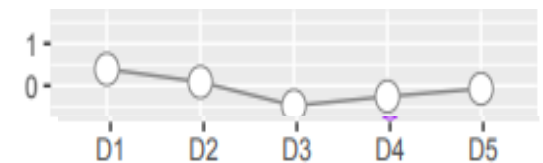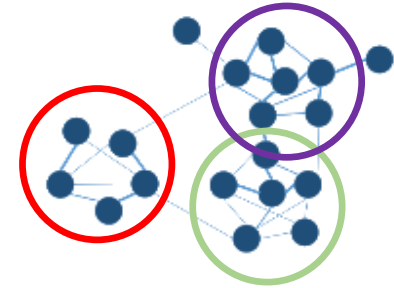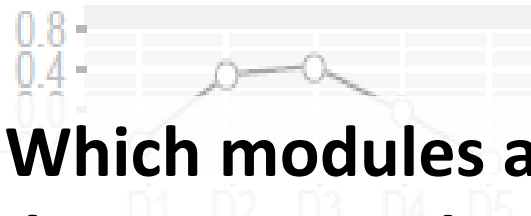Langfelder P and Horvath S, 2008
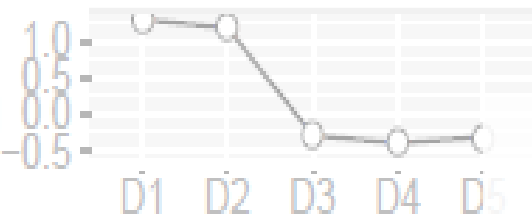
Blue

Yellow

Turquoise
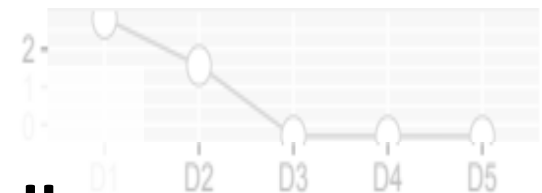
Which modules are gathering secondary cell wall genes ?

Brown

Purple

8
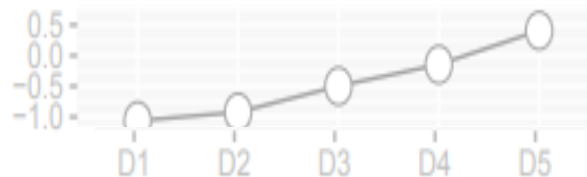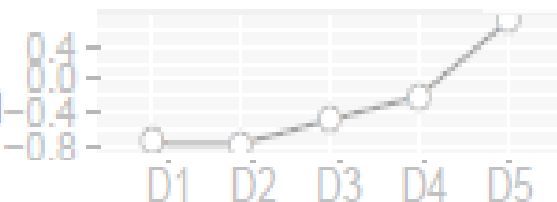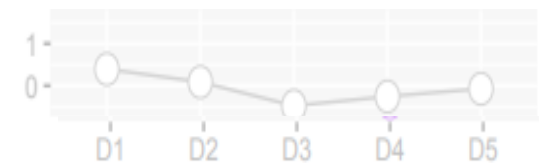
S/PCW GO terms are overrepresented in 5 modules
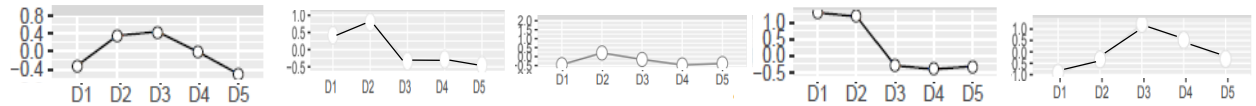
Nb of GO terms and strenght of the p-value

	Yellow	Lightsteelblue	Steelblue	Blue	Pink
SCW	🟥			🟩	🟩
Xylan	🟥		🟩	🟩	
Lignin	🟧				🟩
Suberin	🟩				
Shikimate	🟩				
Flavonoid	🟧			🟩	
Cellulose	🟧		🟩		
Phenylprop	🟩				🟩
CW		🟩	🟧	🟧	
Pectin	🟩	🟩			
Xylem					🟩
GO terms with p-value<5%	92	8	29	100+	75

S/PCW GO terms are overrepresented in 5 modules

	Yellow	Lightsteelblue	Steelblue	Blue	Pink
SCW					
Xylan					
Lignin					
Suberin					
Shikimate					
Flavonoid					
Cellulose					
Phenylprop					
CW					
Pectin					
Xylem					
GO terms with p-value<5%	92	8	29	100+	75
p-value for custom list enrichment	2.07*10-34	1.59*10-5	6.01*10-5	8.72*10-5	0.0093

Nb of GO terms and strenght of the p-value

Enrichment in custom list with enzymes involved in CW biosynthesis Rai et al.,2016
McKinley et al., 2016

S/PCW GO terms are overrepresented in 5 modules

|	Yellow	Lightsteelblue	Steelblue	Blue	Pink
SCW	🟥			🟩	🟩
Xylan	🟥		🟩	🟩	
Lignin	🟧				🟩
Suberin	🟩				
Shikimate	🟩				
Flavonoid	🟧			🟩	
Cellulose	🟧		🟩		
Phenylprop	🟩				🟩
CW		🟩	🟧	🟧	
Pectin	🟩				
Xylem					🟩
GO terms with p-value<5%	92	8	29	100+	75
p-value for custom list enrichment	2.07*10-34	1.59*10-5	6.01*10-5	8.72*10-5	0.0093

Nb of GO terms and strenght of the p-value

5 modules containing secondary and primary cell wall involved genes [1]
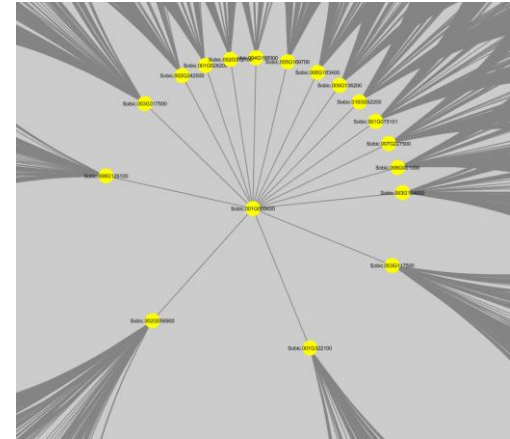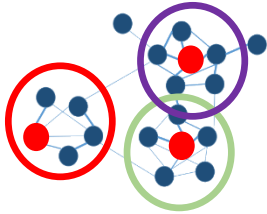
Hub genes of pink module include KNAT7

- Finding hub genes by calculating centrality scores
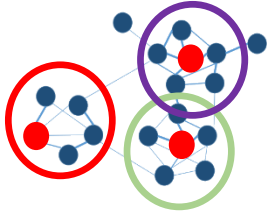
Hub genes of pink module include KNAT7

- Finding hub genes by calculating centrality scores

Sobic.001G075101	KNAT1KNOTTED-like from Arabidopsis thaliaa
Sobic.001G131400	
Sobic.001G322100	
Sobic.001G526200	KNAT7KNOTTED-like homeobox of Arabidopsis thalina
Sobic.002G056900	
Sobic.002G242500	WRKY2WRKY DNA-binding protin
Sobic.002G296600	
Sobic.002G312700	
Sobic.003G104600	NmrA-like negative transcriptional regulator family pro
Sobic.003G117500	RALFL9 - Rapid ALkalinization Factor RALF family protein precursor
Sobic.003G317500	
Sobic.005G136200	Chalcone and stilbene synthase family protese
Sobic.005G169700	
Sobic.006G001000	Cytochrome p450
Sobic.006G128100	
Sobic.006G183400	RBL1 RHOMBOIDlike
Sobic.007G227500	
Sobic.010G052200	

Hub genes in favorites modules

- 5 modules containing hub genes important for further analysis
- List in the *to be published* paper

Using gene co-expression network to identify candidate genes

**Transcriptomics data sampled
during development**

Building a gene coexpression network

- Build a gene-gene similarity network
- Divide network into groups with similar expression
- Identify drivers genes and groups

1 Analysis of co expressed gene groups
- Enrichment
- Top genes

2 Use of bait genes known for their role in SCW

Using gene co-expression network to identify candidate genes

**Transcriptomics data sampled
during development**

Building a gene coexpression network

- Build a gene-gene similarity network
- Divide network into groups with similar expression
- Identify drivers genes and groups

1 Analysis of co expressed gene groups
- Enrichment
- Top genes

2 Use of bait genes known for their role in SCW

TRANSCRIPTION FACTORS
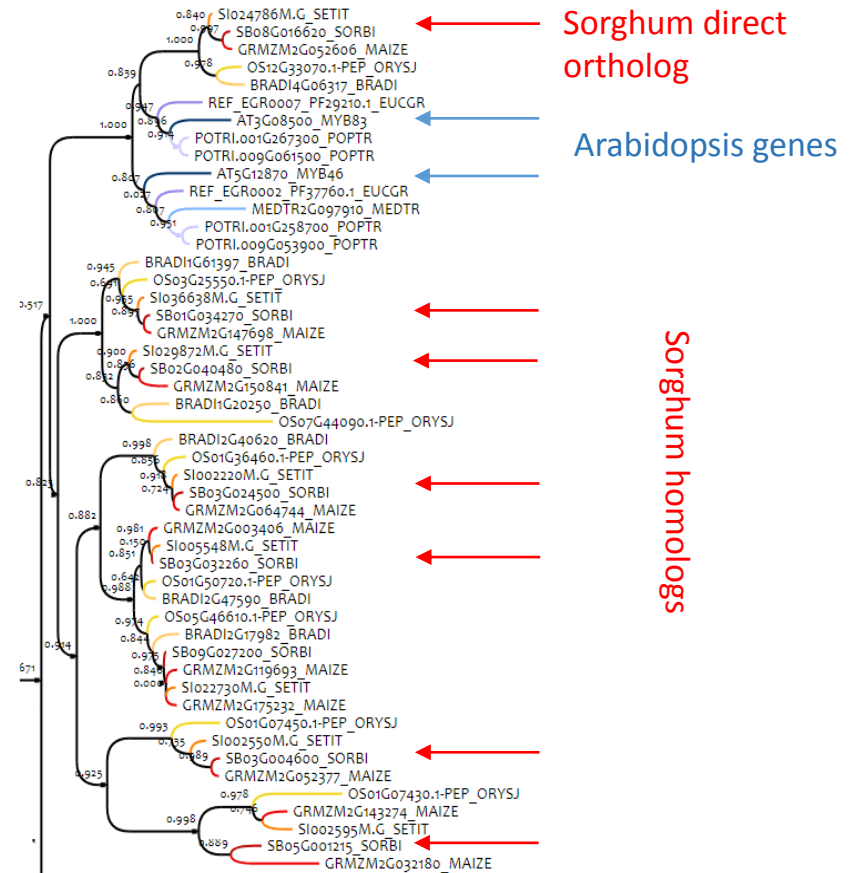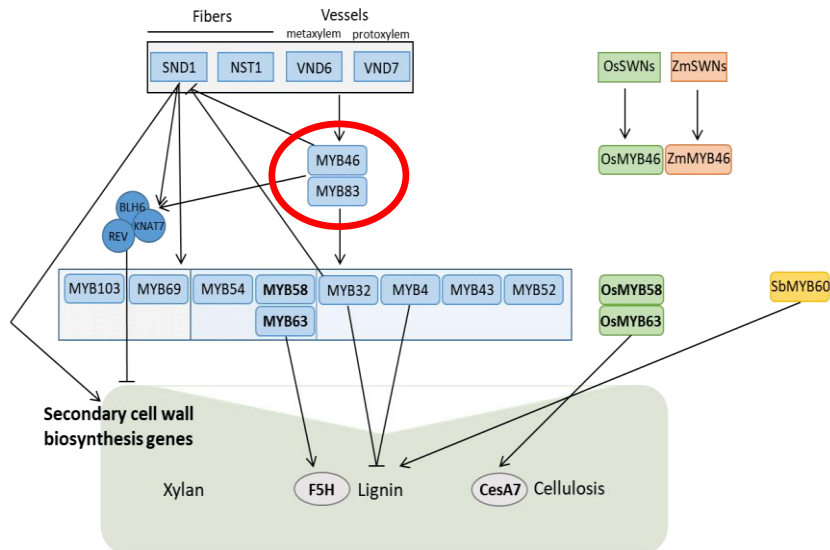
Synthesising MYB and NAC in sorghum

- Phylogenetic trees including 9 species
  - Arabidopsis, eucalyptus, poplar, medicago, brachypodium, setaria, maize, rice, sorghum

- MYB = 1700+ proteins and 150 in sorghum
- NAC = 900+ proteins and 90 in sorghum
Sanjari et al., 2019

- Phylogenetic trees including 9 species
  - Arabidopsis, eucalyptus, poplar, medicago, brachypodium, setaria, maize, rice, sorghum

- MYB = 1700+ proteins and 150 in sorghum
- NAC = 900+ proteins and 90 in sorghum
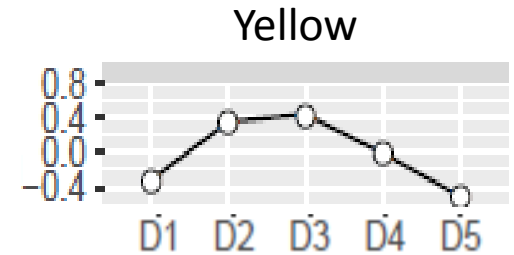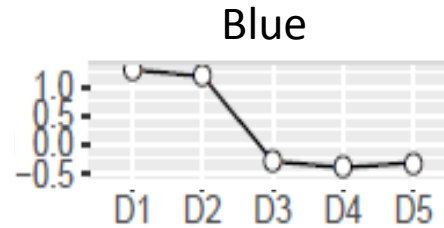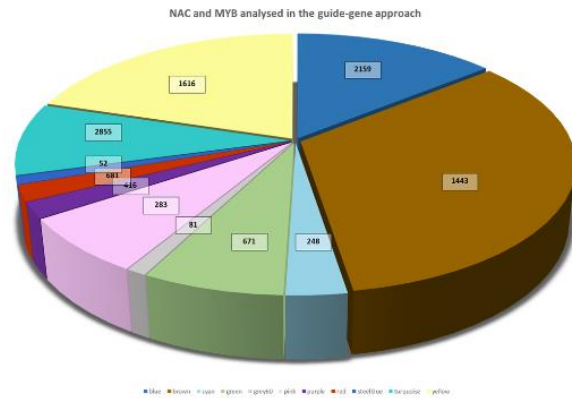 Sanjari et al., 2019

Half of NAC and MYB tested are in favorite modules
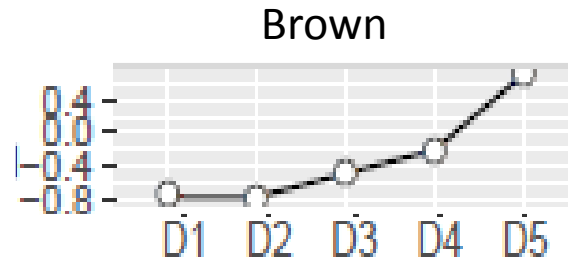
- 100 NAC and MYB bait genes used

Blue

Yellow

Brown

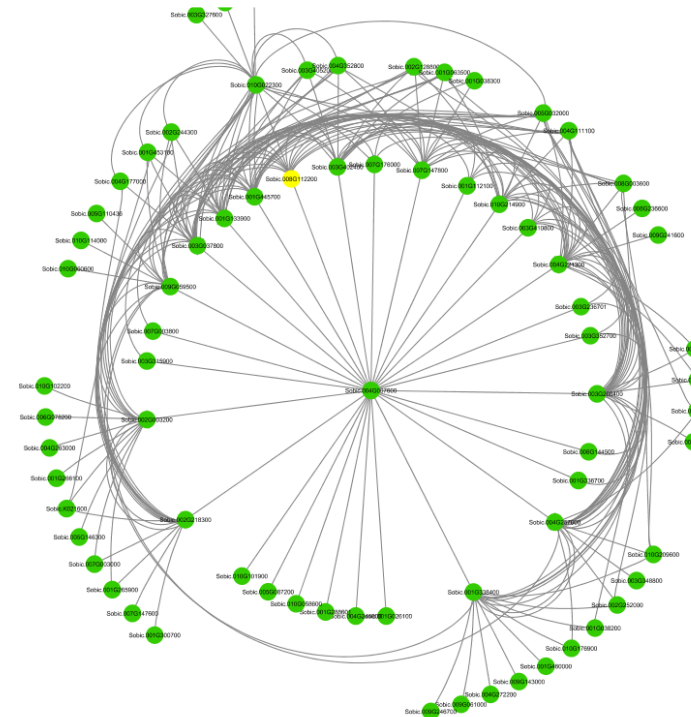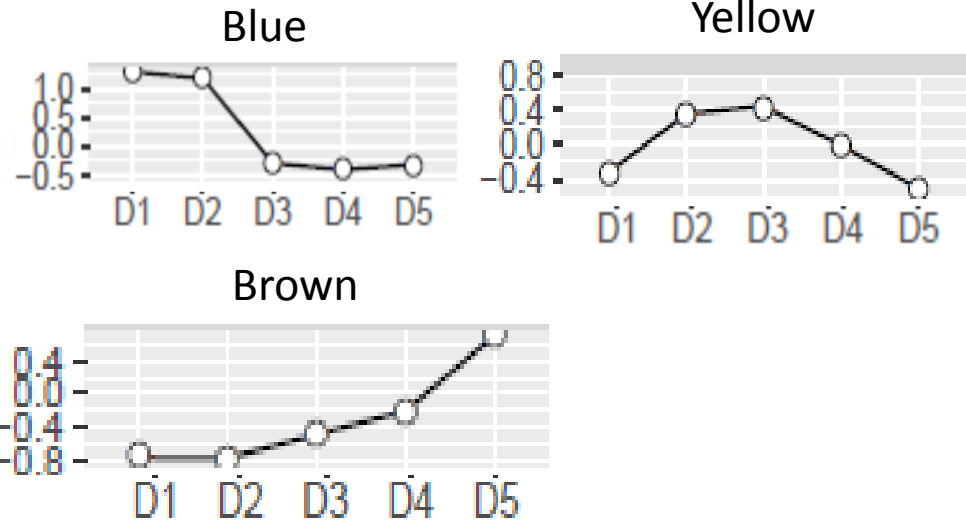NAC and MYB analysed in the guide-gene approach

Half of NAC and MYB tested are in favorite modules

- 100 NAC and MYB bait genes used

Blue

Yellow

Brown
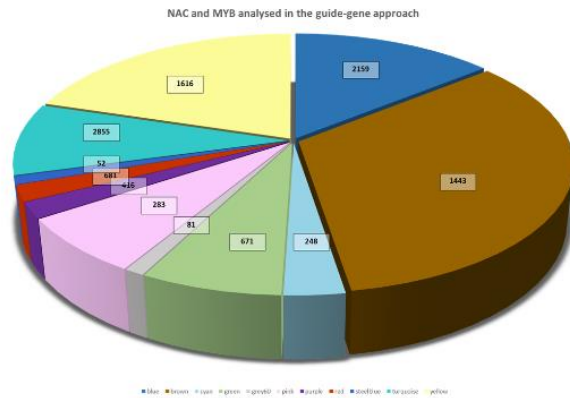
- MYB46 *top15'stop15* co-expressed genes
 - Capture the strongest links
  - 15x15 genes

- NAC and MYB favorite candidates:

Ortholog	NST1b	NST1a	MYB46	SND3	SND2	VND7	VND4/5/6	VND4/5/6
SCW	red	red	orange	red	red			
Lignin	red	orange	green	red	orange			
Xylan	red	red	orange	red	red			
Pectin	orange	green		orange	green	green		green
CW	orange	green		orange	green			green
Cellulose	orange	orange	orange	orange	orange			
Phloem loading		orange	orange		orange	orange		
Wax						orange		
Stem vascular tissue pattern formation			green					
Arabinan		green	green					
Polyphenols						orange		
Nb signific GO terms	32	34	31	34	32	15	13	56

Unknown MYB identified as a good candidate

	Sobic.002G141300
Ortholog	MYB not annotated
SCW	
Lignin	
Xylan	
Pectin	
CW	
Cellulose	
Phloem loading	
Nb signific GO terms	35
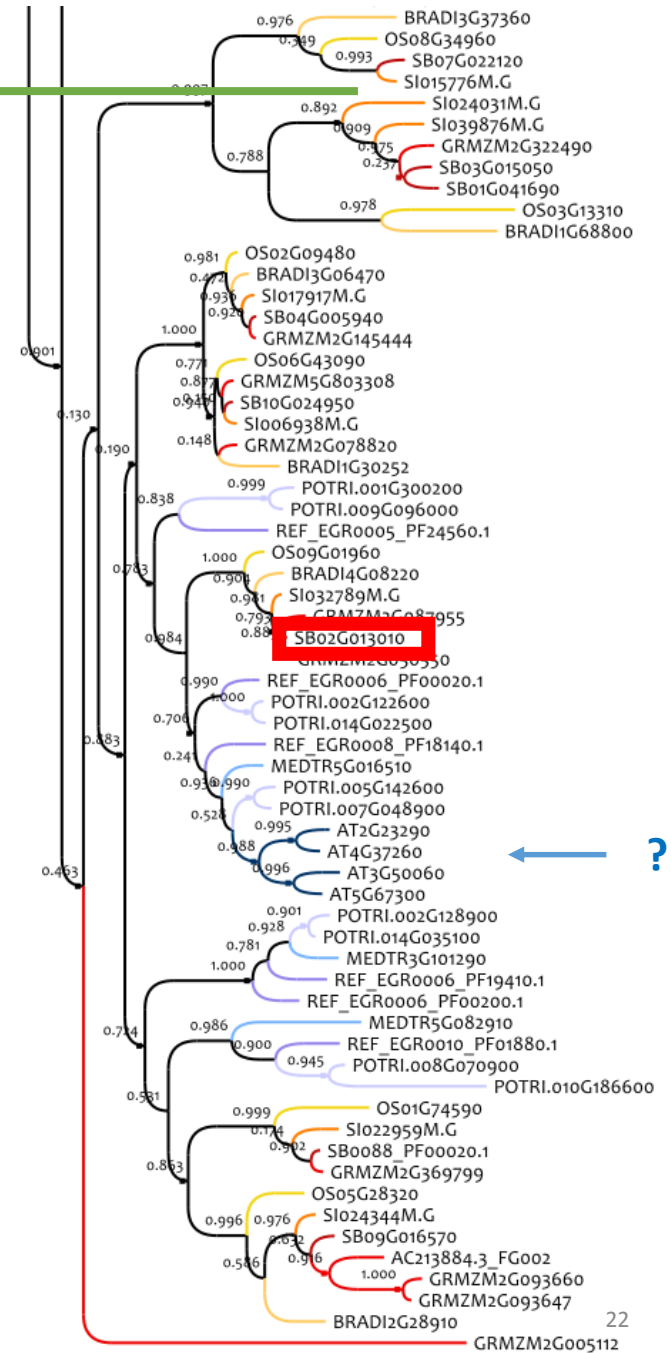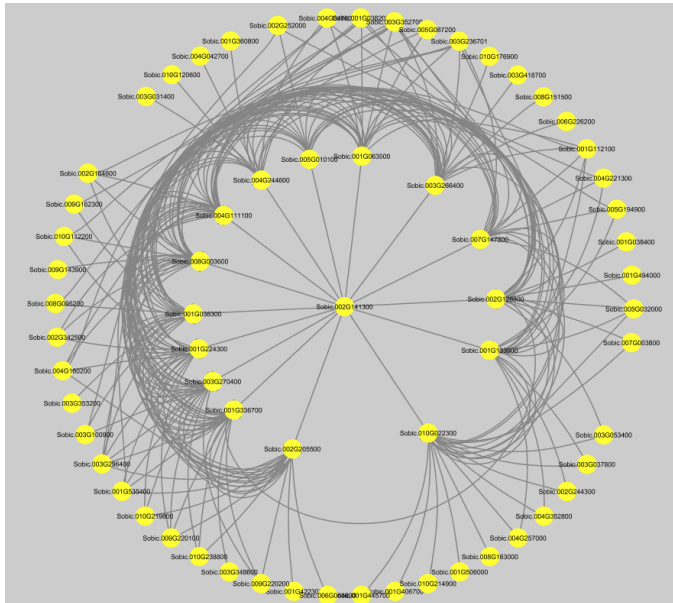
Unknown MYB identified as a good candidate

	Sobic.002G141300
Ortholog	MYB not annotated
SCW	
Lignin	
Xylan	
Pectin	
CW	
Cellulose	
Phloem loading	
Nb signific GO terms	35

Sobic.001G224300 CESA7 - cellulose synthase
Sobic.001G422300 laccase 17
Sobic.001G445700 O-Glycosyl hydrolases family 17
Sobic.001G535400 4-GALACTOSYLTRANSFERASE
Sobic.002G205500 CESA9 - cellulose synthase
Sobic.002G342500 alpha 14-glycosyltransferase
Sobic.003G296400 CESA4 - cellulose synthase
Sobic.003G352700 laccase 17
Sobic.003G353200 laccase 2
Sobic.004G042700 endoglucanase
Sobic.004G047600 Polysaccharide biosynthesis
Sobic.004G244600 endoglucanase
Sobic.005G032000 anthocyanin 5-O-glucosyltransferase
Sobic.005G032000 UDP-Glycosyltransferase super
Sobic.009G162300 laccase 17
Sobic.009G220100 Exostosin
Sobic.010G112200 Xylanase inhibitor N-terminal
Sobic.010G120600 UDP-glucosyl transferase domain
Sobic.010G214900 4-coumarate:CoA ligase 2
Sobic.010G238800 glycosyltransferase family 43
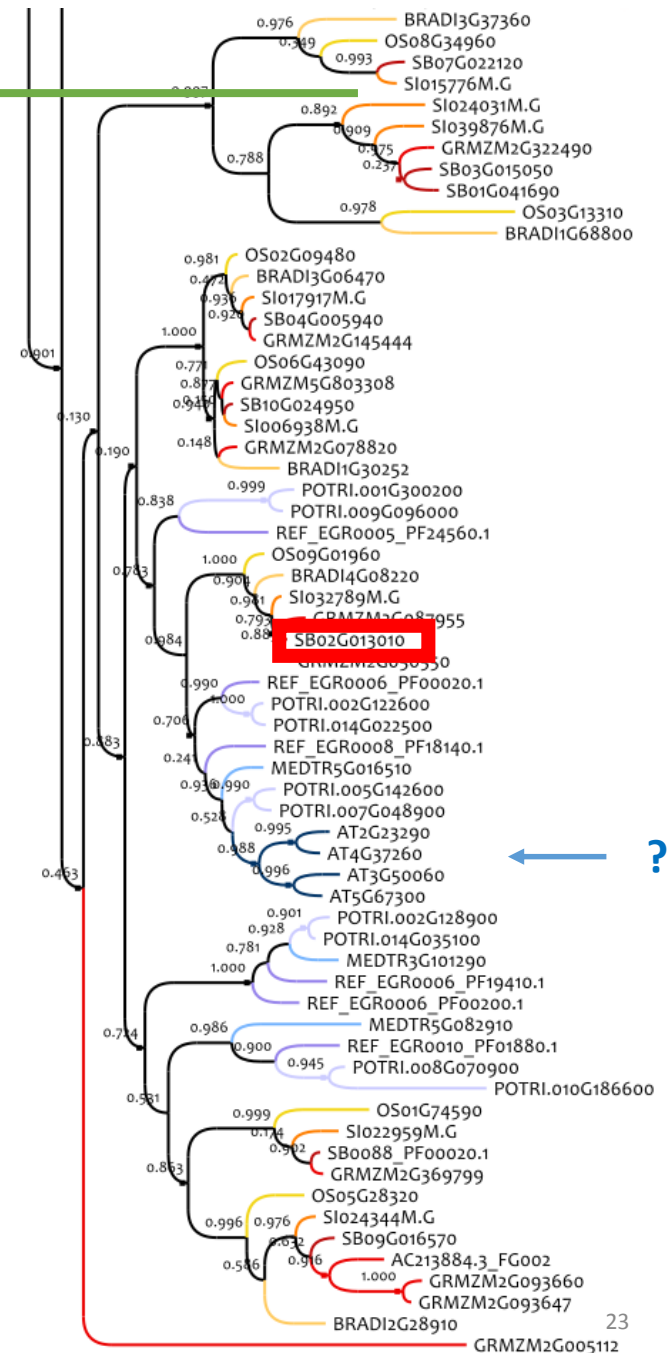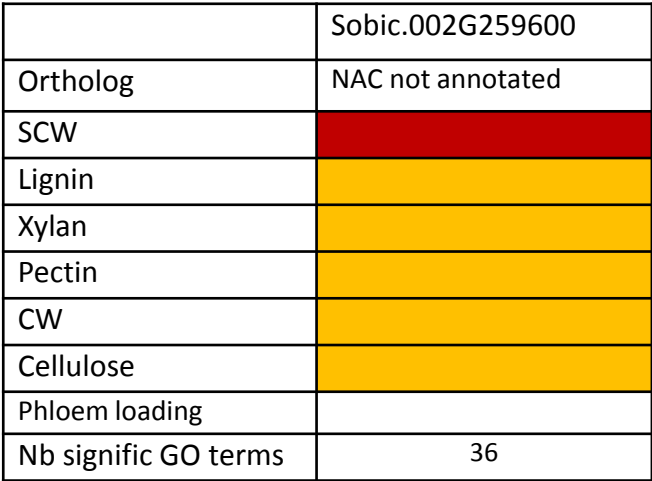
23

Unknown NAC identified as a good candidate
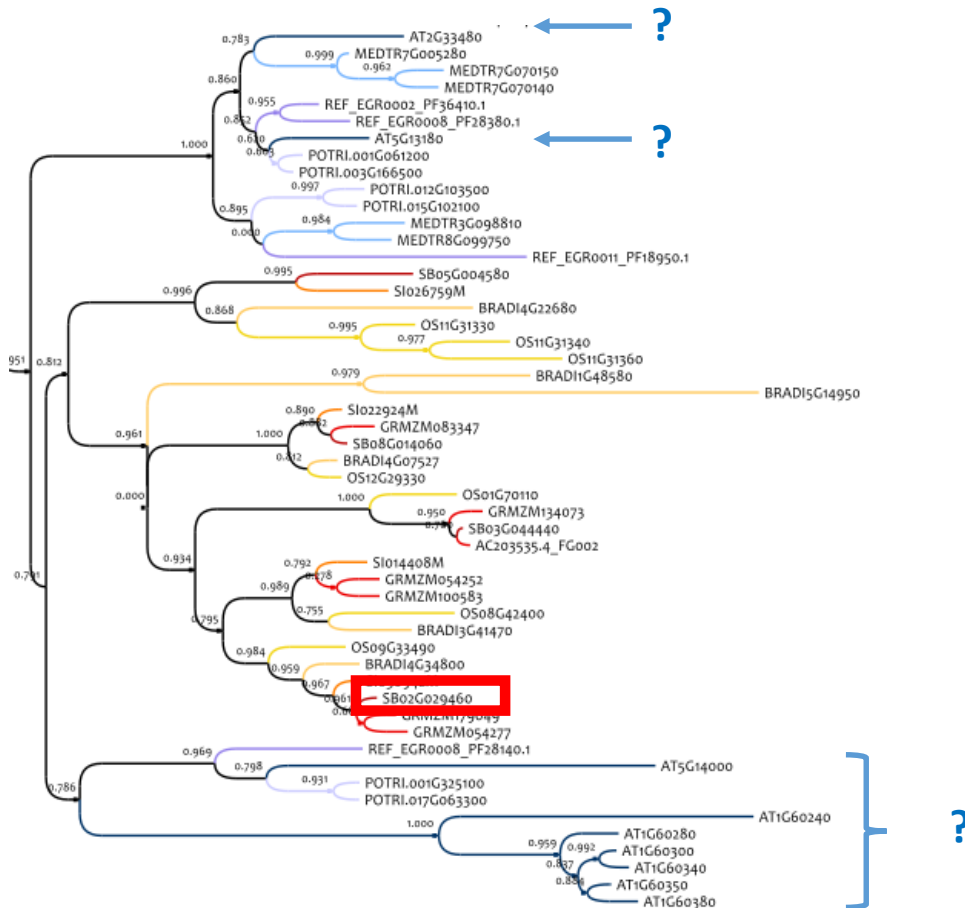
	Sobic.002G259600
Ortholog	NAC not annotated
SCW	
Lignin	
Xylan	
Pectin	
CW	
Cellulose	
Phloem loading	
Nb signific GO terms	36

Unknown NAC identified as a good candidate

	Sobic.002G259600
Ortholog	NAC not annotated
SCW	
Lignin	
Xylan	
Pectin	
CW	
Cellulose	
Phloem loading	
Nb signific GO terms	36

Sobic.001G224300 CESA7 - cellulose synthase
Sobic.001G422300 laccase 17
Sobic.002G205500 CESA9 - cellulose synthase
Sobic.002G342500 alpha 1,4-glycosyltransferase
Sobic.003G296400 CESA4 - cellulose synthase
Sobic.003G352700 laccase 17
Sobic.003G353200 laccase 2
Sobic.004G047600 Polysaccharide biosynthesis
Sobic.007G039100 myb domain protein 103
Sobic.009G026101 glycosyltransferase family 43
Sobic.009G162300 laccase 17
Sobic.009G220100 Exostosin family protein
Sobic.009G220200 Exostosin family protein
Sobic.010G112200 Xylanase inhibitor N-terminal
Sobic.010G134000 Phosphofructokinase family protein
Sobic.010G238800 glycosyltransferase family 43

Main NAC and MYB validated and new TF identified

- 5 modules containing hub genes important for further analysis
- List in the *to be published* paper

- Major NAC and MYB sorghum ortholog validated
- Most undirect orthologs seem also involved in SCW and CW
- Not annotated NAC and MYB identified

Conclusion

- 2 NAC and 2 MYB are being tested in transient and stable transformation AND transcriptomic
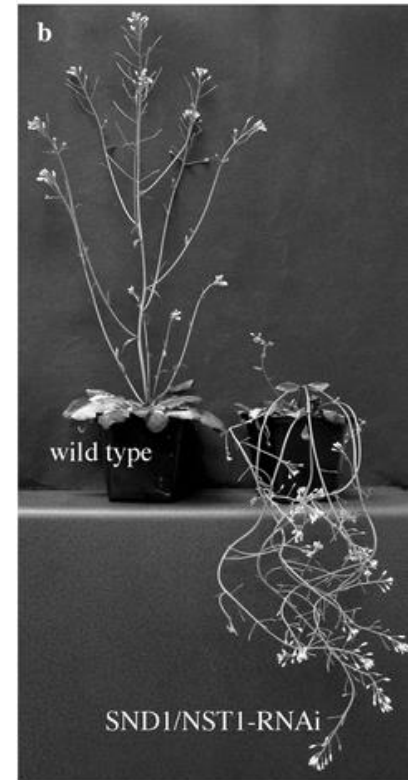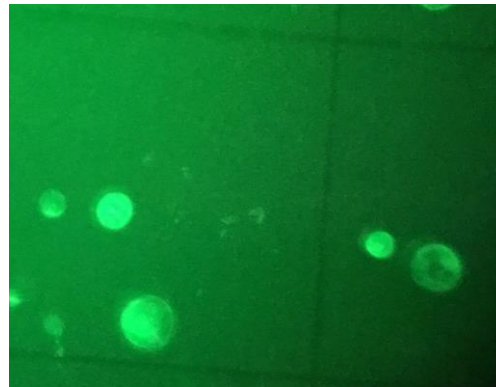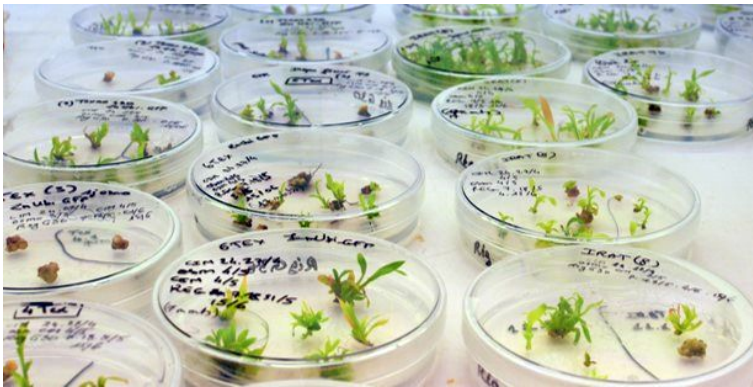
- Not annotated NAC and MYB to be tested

- Overlap bewteen these lists and GWAS SNP

- Selection signature, allele diversity

Zong et al. 2007

THANK YOU ☺