

BUILDING A SORGHUM GRAIN : A TRANSCRIPTOME ROADMAP TARGETING PROTEIN CONTENT AND DIGESTIBILITY

Séne M^{1,2}, Berger A^{1,2}, Calatayud C^{1,2}, Rios M^{1,2}, Bonicel J³, Morel MH³, Mameri H³, Pot D^{1,2}, Terrier N^{1,2}

¹ UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier - ² CIRAD, UMR AGAP Institut, F-34398 Montpellier - ³ UMR IATE, INRAE, Institut Agro Montpellier, University of Montpellier, Montpellier, France

1. Contexte

- Sorghum can contribute to increase the share of plant proteins
- A major problem limiting sorghum food and feed use is its low protein digestibility
 - intrinsic nature of the reserve proteins (prolamins, called kafirins) in protein bodies
 - appearance of additional disulfide bridges during grain development (and cooking)
- There is variability in sorghum :
 - for protein content in grains (Rami et al., 1998 ; Kimani et al., 2020)
 - for the composition in different kafirins (Cremer et al., 2014)
 - for protein digestibility (Cremer et al., 2014 ; Diatta, 2018 ; Duressa et al., 2020 ; Kardes et al., 2021)
- Allelic diversity for genes involved in the biosynthesis of different kafirins has been identified (Laidlaw et al., 2010)

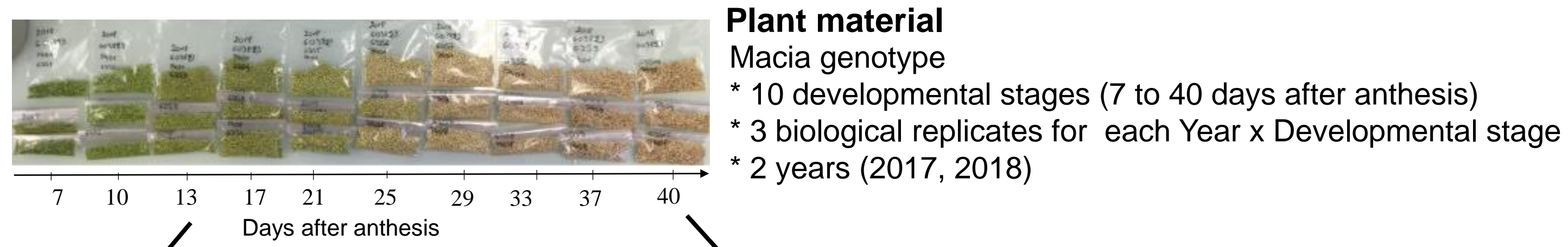
2. Objective

The aim of this study is to improve the understanding of the molecular and biochemical mechanisms leading to protein accumulation and involved in their low digestibility

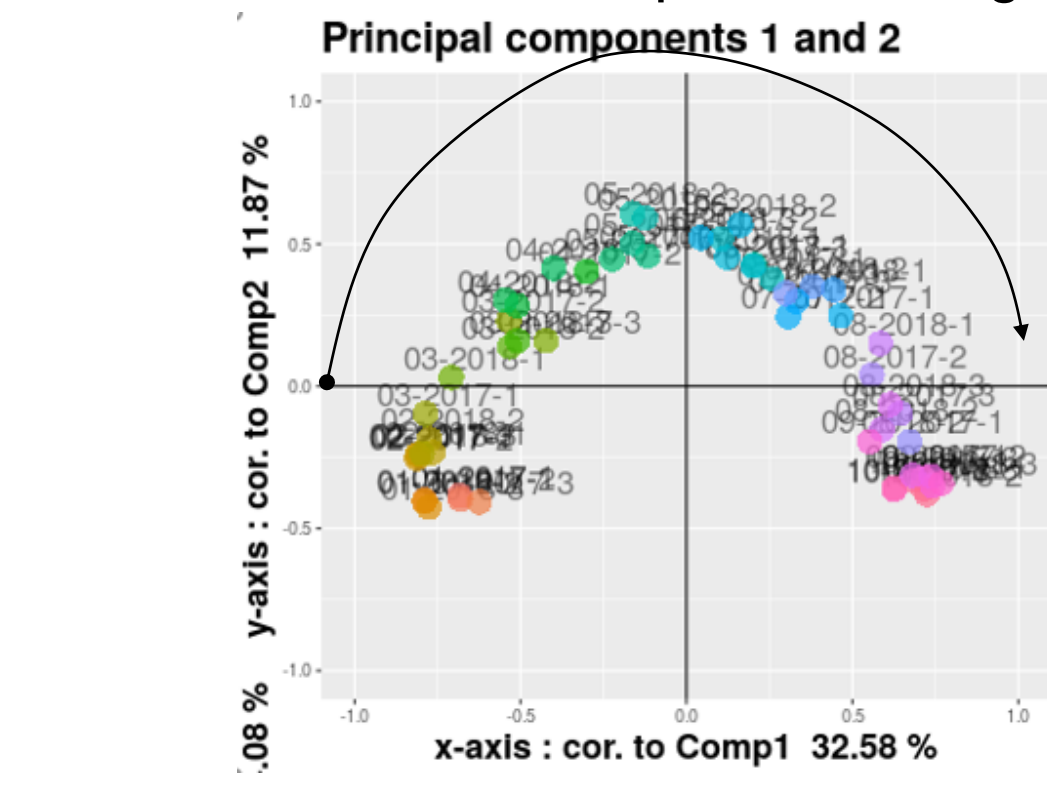
We addressed the two following questions :

- Which molecular actors are involved in protein body establishment ?
- Which molecular factors are involved in the low digestibility of grain proteins ?

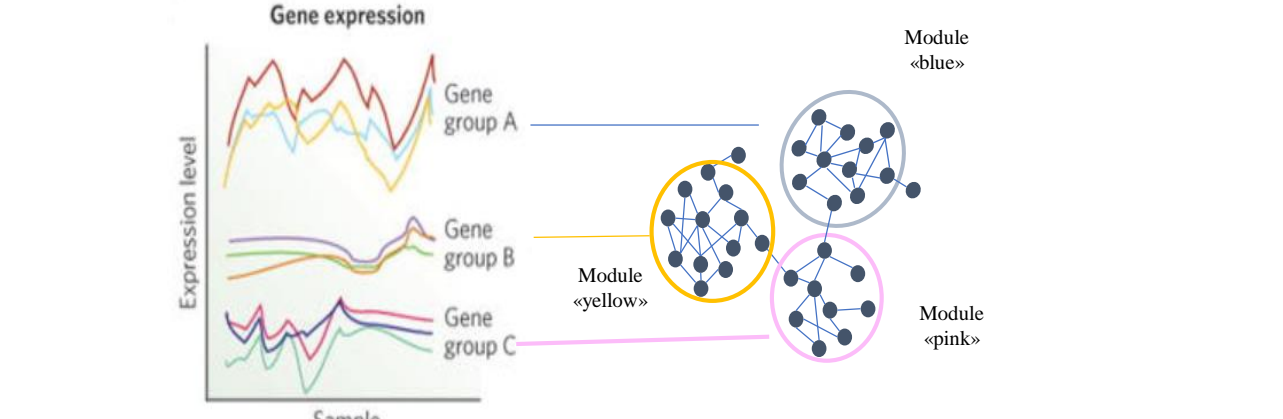
3. Plant material, gene expression data and biochemical traits



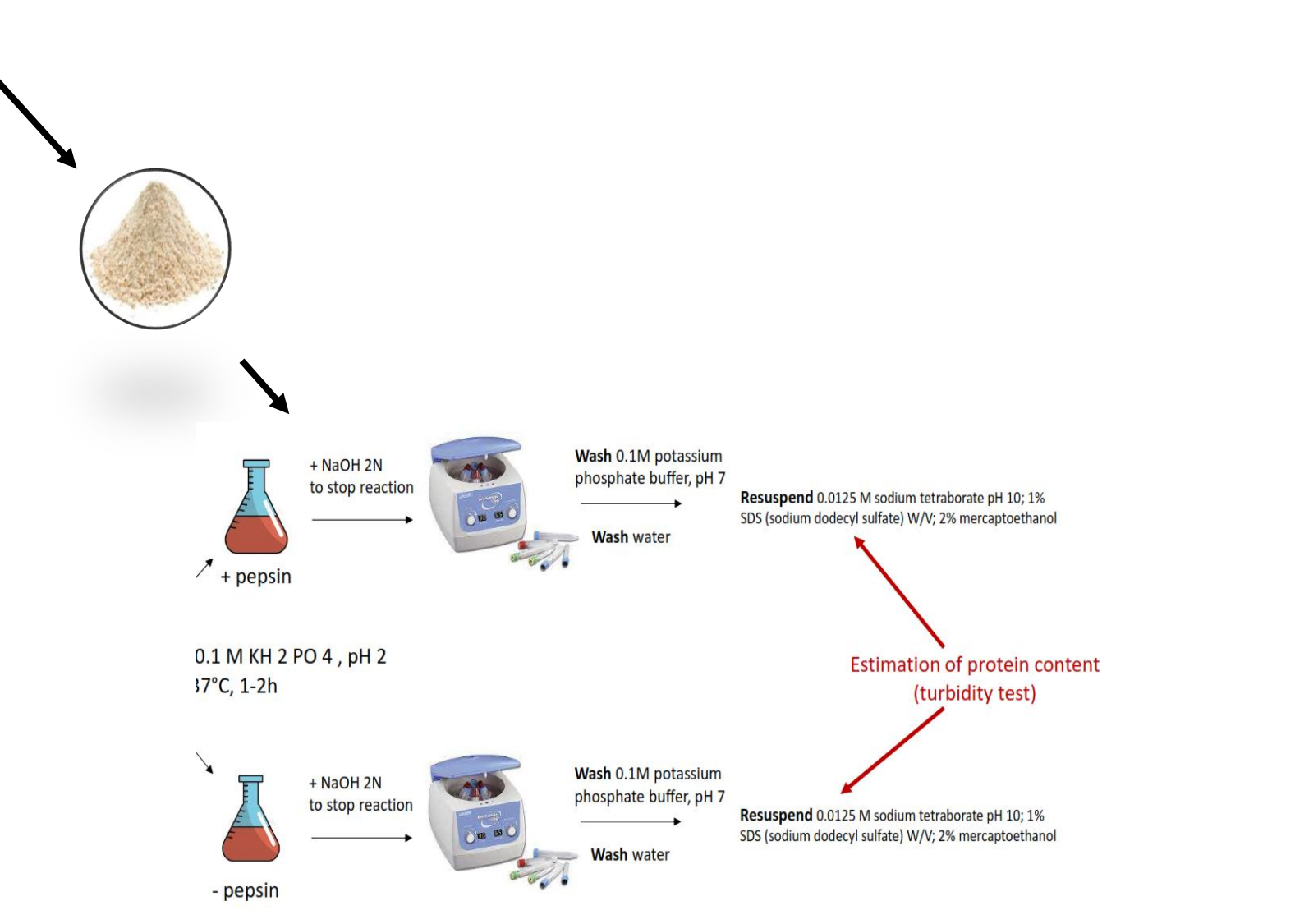
Gene expression data
RNAseq = 60 samples, 37325 genes



Gene network analysis (WGCNA, Cytoscape)



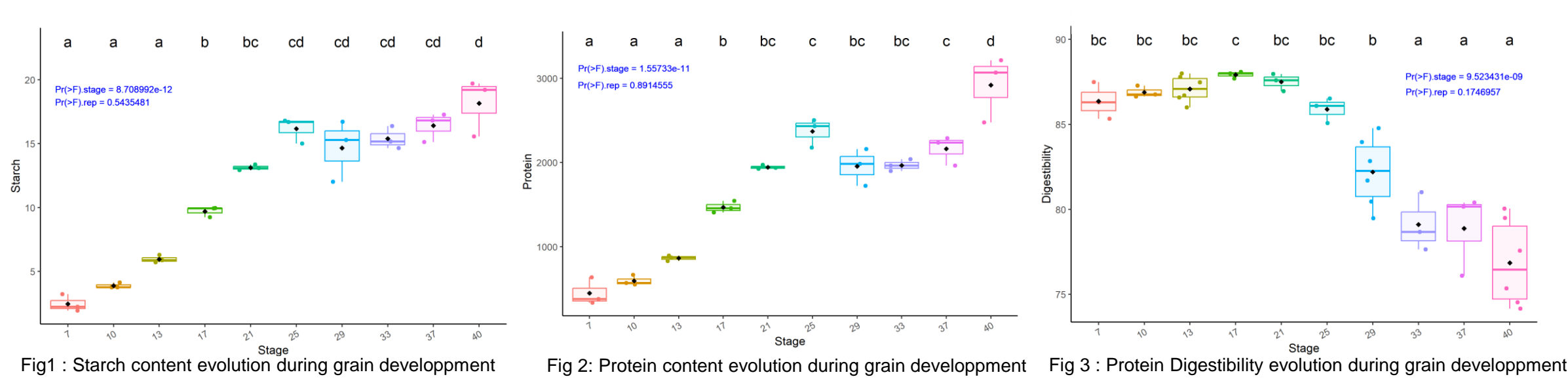
Plant material
Macia genotype
* 10 developmental stages (7 to 40 days after anthesis)
* 3 biological replicates for each Year x Developmental stage
* 2 years (2017, 2018)



Biochemical traits

- Protein Digestibility (%)
- Starch content (mg/grain)
- Protein content (µg/grain)

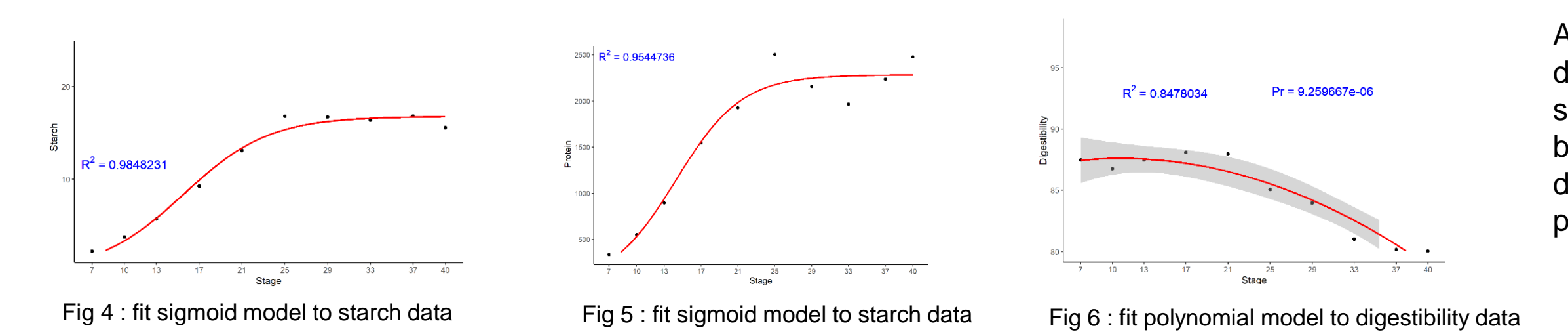
4. Evolution of biochemical traits during grain development



These traits are reshaped during grain development

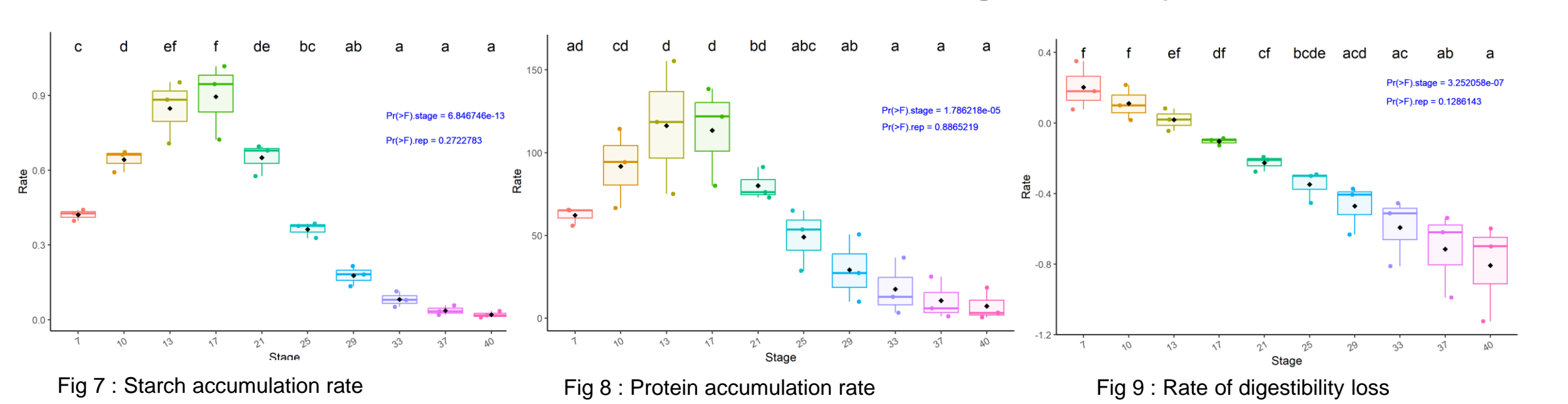
Need to calculate accumulation rates to study the correlation between gene expression and trait variation along grain development

5. Modeling starch protein contents and protein digestibility evolution



According to their developmental course, starch and protein were fit by a sigmoid model, while digestibility was fit by a polynomial model

6. Kinetics of starch, protein and protein digestibility



Protein and starch accumulation rates increase progressively until days 13 and 17, before gradually decreasing and stabilizing from day 33. The rate of loss of digestibility gradually decreases during grain development.

References

Cremer et al., 2014. Grain Sorghum Proteomics : Integrated Approach toward Characterization of Endosperm Storage Proteins in Kafirin Allelic Variants. *J. Agric. Food Chem.* 2014, 62, 40, 9819.
Diatta-Hoigate., 2022. Natural variability for protein digestibility and grain quality traits in a West African Sorghum Association Panel. *Journal of Cereal Science* 107, 103504.
Duressa et al., 2020. Identification of variant α-kafirin alleles associated with protein digestibility in grain sorghum. 60 : 2467-2478
Kimani et al., 2020. Genome-wide association study reveals that different pathways contribute to grain quality variation in sorghum (*Sorghum bicolor*). *BMC Genomics*. 21 (112)
Laidlaw et al., 2010. Allelic variation of the β-, γ- and δ-kafirin genes in diverse Sorghum genotypes. *Theoretical and Applied Genetics* 121(7):1227-37
Moretti and Laurindo., 2017. Protein disulfide isomerases: Redox connections in and out of the endoplasmic reticulum". *The Chemistry of Redox Signaling*. 617: 106–119.
Rami et al., 1998. Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (*Sorghum bicolor* L. Moench). *Theoretical and Applied Genetics* 97 (605–616)

7. Gene co-expression modules and their relationships with biochemical traits

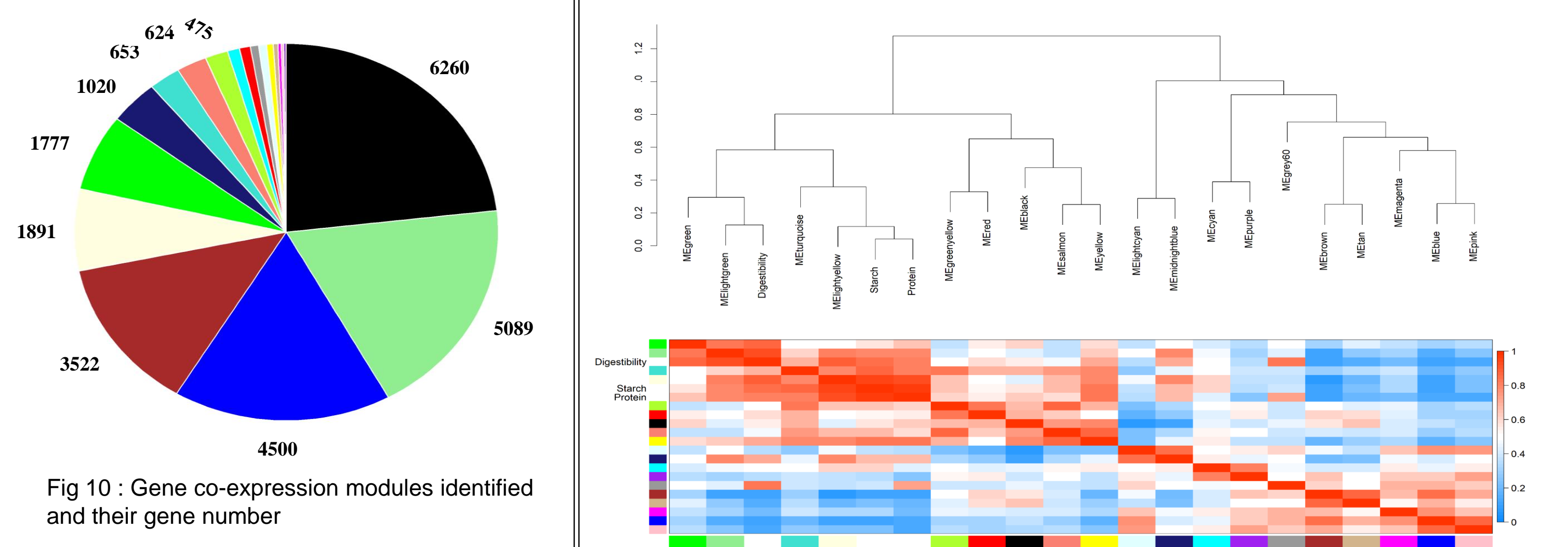


Fig 10 : Gene co-expression modules identified and their gene number

19 modules were identified, of which the lightgreen, black and blue ones contain the largest number of genes

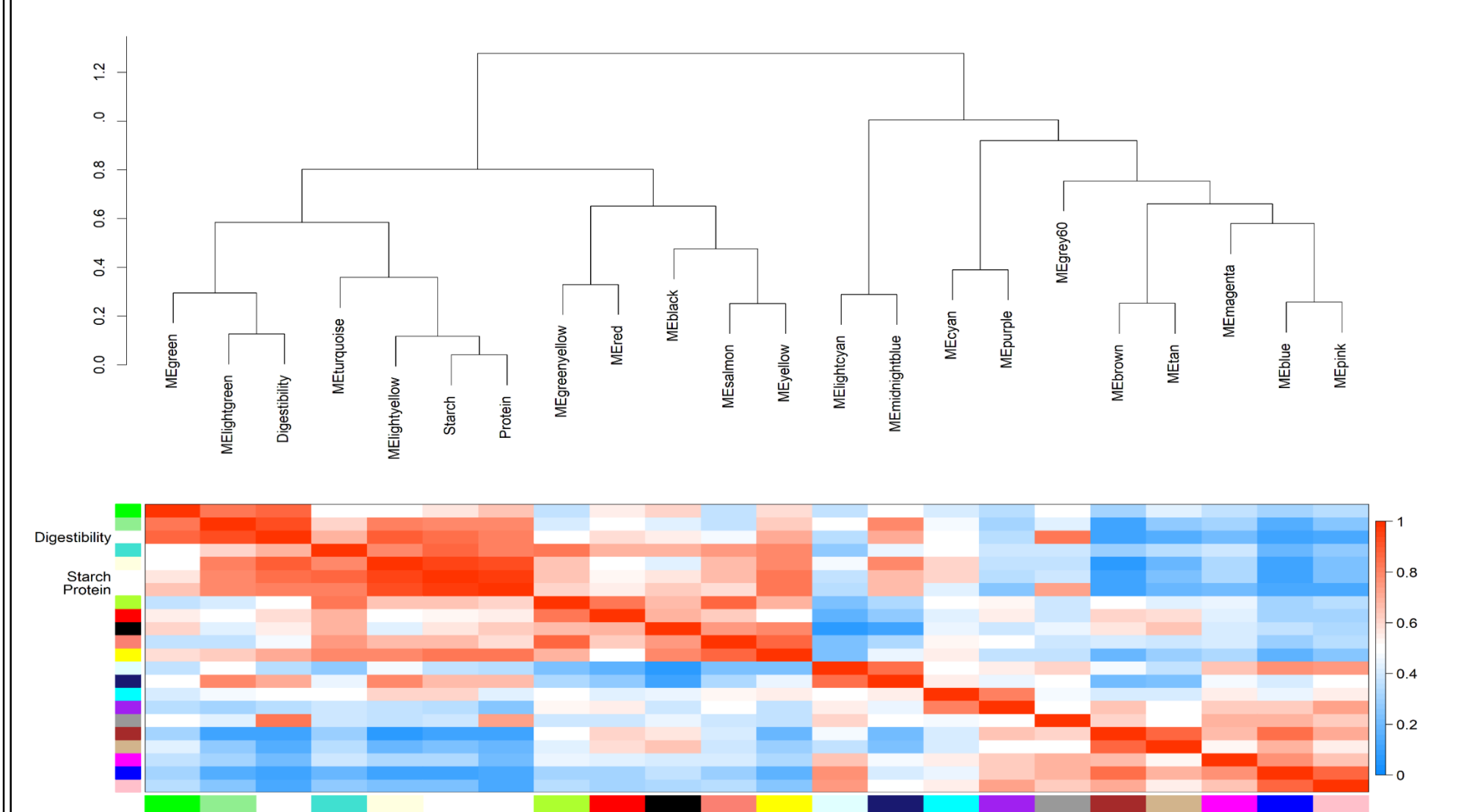


Fig 11 : Relationships between gene co-expression modules and kinetics of biochemical traits

Starch, protein and digestibility kinetics are correlated with each other
Strong positive correlations are observed between :
✓ starch, protein, lightyellow and turquoise modules
✓ digestibility and the lightgreen module

8. The Lightyellow gene co-expression module contains genes and Transcription Factors (TF) involved in protein and starch synthesis, accumulation and storage

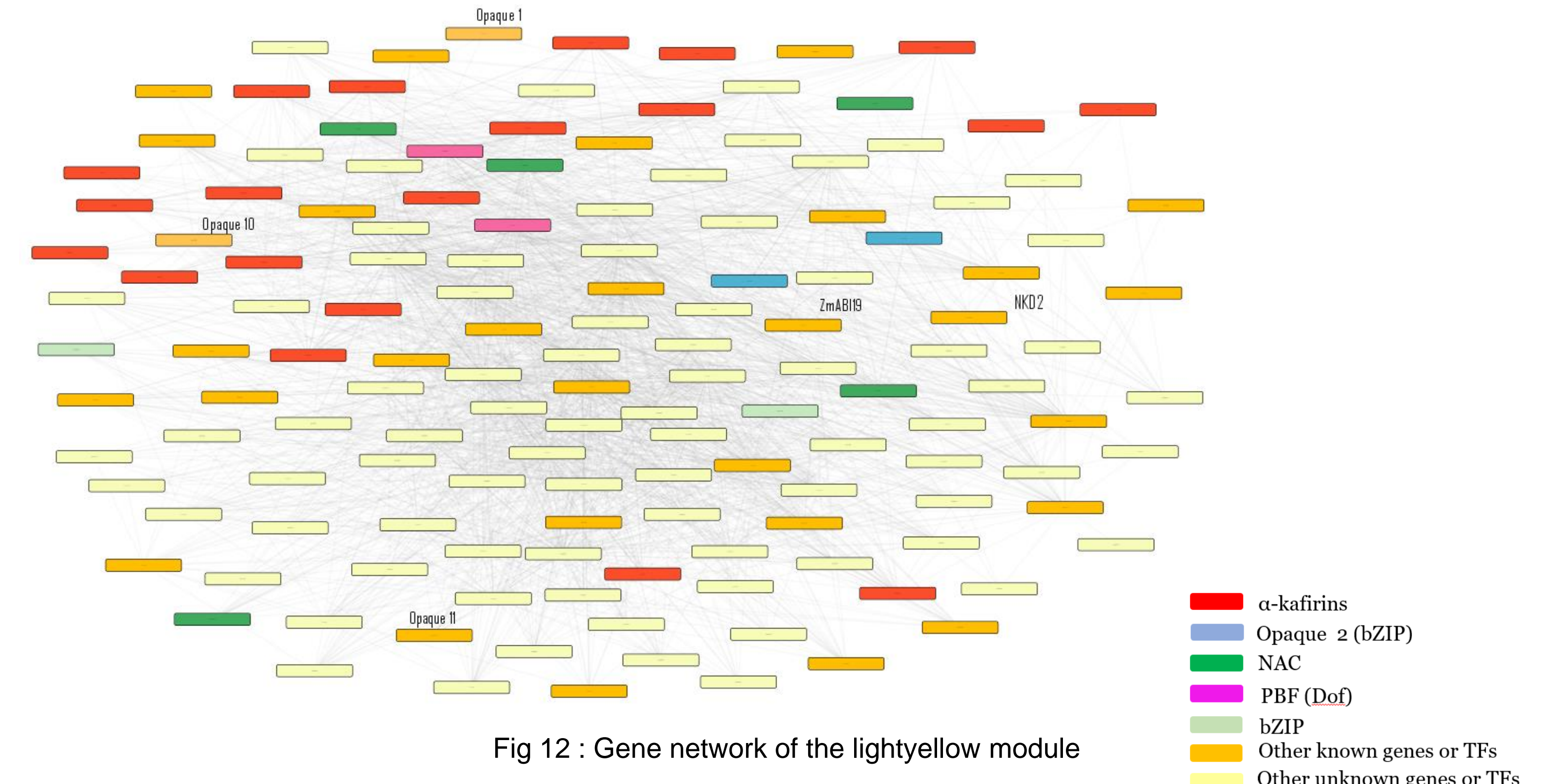


Fig 12 : Gene network of the lightyellow module

This gene co-expression network contains :

- The majority (20/23) of genes coding for α-kafirin (main sorghum protein, Laidlaw et al., 2010)
- Orthologs of maize and rice genes (Op1, Op10, Op11, etc), involved in protein body structure, protein and starch synthesis, endosperm development, etc (Fenf et al., 2018; Gontarek et al., 2016; Zhang et al., 2015)
- Known TFs (Op2, NAC, ZmABI19, NKD, etc), whose maize orthologs regulate grain protein and starch accumulation, development and filling (Yang et al., 2021; Zhang et al., 2019; Zhang et al., 2015)
- New TFs not yet identified as being involved in the molecular determinism of our target traits : CO-like, bHLH, C2H2, etc.

9. Two of the most correlated genes with digestibility in lightgreen module are involved in protein condensation.

The putative effect of a gene on protein digestibility in the lightgreen module was estimated by calculating the level of correlation (Gene Significance, Tab 1) between the gene expression and the rate of loss of protein digestibility during grain development.

Tab 1: The ten genes most correlated with loss of digestibility rate in the lightgreen module

Sorghum gene name	Ortholog species	Gene Symbol	Annotation	TF_name	GO	References	Gene Significance	Modul Membership
1. Sobc...	S. bicolor	-	granule-bound starch synthase 1b, chloroplastic/amyloplastic	-	GO:0016757	-	0.91	0.92
2. Sobc...	S. bicolor	-	non-specific lipid-transfer protein-like protein At.	-	GO:008289	-	0.90	0.78
3. Sobc...	O. sativa	OsMADS29	MADS-box protein	MIKC_MADS	GO:0003677	Li et al., 2023 ; Nayyar et al., 2013	0.88	0.91
4. Sobc...	O. sativa	OsMADS7/OsMADS45	MADS-box protein	MIKC_MADS	GO:0003677	Li et al., 2023 ; Zhang et al., 2018	0.88	0.93
5. Sobc...	Z. mays	Dsk33	Pyrimidine reductase, riboflavin biosynthesis	-	GO:0008703	Dai et al., 2021 ; Dai et al., 2019	0.87	0.76
6. Sobc...	A. thaliana	DMP	Myo-inositol-phosphate	-	GO:0046854	Risai et al., 2015	0.86	0.94
7. Sobc...	Z. mays	PDI	Protein disulfide isomerase	-	-	Dai et al., 2021	0.86	0.83
8. Sobc...	O. sativa	OsPK2	Plastic pyruvate kinase	-	GO:0000287	Li et al., 2023 ; Cai et al., 2018	0.85	0.91
9. Sobc...	Z. mays	PDI	Protein disulfide isomerase	-	GO:0003756	Dai et al., 2021	0.85	0.95
10. Sobc...	O. sativa	OsMADS1-ql.GY3	MADS-domain transcription factor	MIKC_MADS	GO:0003677	Li et al., 2023 ; Liu et al., 2018	0.85	0.83

The lightgreen module contains two PDI genes (Protein disulfide isomerase), two of which are among the ten genes most significantly linked to loss of digestibility rate. PDI have oxidoreductase and isomerase properties and are specifically responsible for protein folding in the endoplasmic reticulum (Moretti and Lauridon, 2017), which probably make them major players of the variation in grain protein digestibility.

10. Conclusion and perspectives

In addition to the structural genes and key transcription factors identified in the lightyellow and lightgreen modules, exploration of the other modules also enabled us to detect other genes linked to sorghum secondary proteins: β-kafirin and δ-kafirin (turquoise module) and γ-kafirin (salmon module).

Considering the recalcitrant nature of sorghum to stable transformation (Poster 19), in order to evaluate the roles of these newly identified TFs, a simplified cellular system of overexpression in sorghum protoplasts will be used.

Acknowledgements

- This work was supported by the SOKAFI project funded by INRAE and the NitroSorg project funded by CASDAR semences.
- Mamadou Sene receives support for his PhD grant from CIRAD and INRAE CLIMAE Metaprogramme