Could storage protein composition be modified by acting at the transcriptional level?

Julie Boudet, Adrien Mazuel, Annie Faye, Titouan Bonnot, Pierre Martre, Catherine Ravel

Could storage protein composition be modified by acting at the transcriptional level?

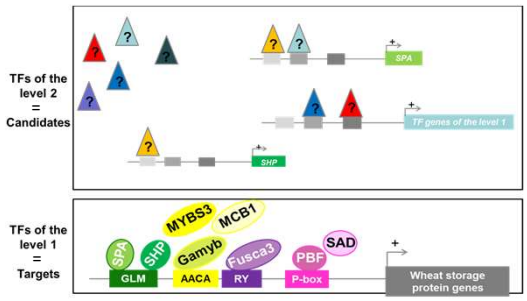
Julie Boudet, Adrien Mazuel, Annie Faye, Titouan Bonnot, Pierre Martre, Catherine Ravel

Université Clermont Auvergne, INRA, UMR 1095 GDEC (Génétique, Diversité et Écophysiologie des Céréales), Clermont-Ferrand, France.

CONTEXTE AND OBJECTIVES

Bread wheat (*Triticum aestivum*) is mainly used after transformations. All these transformations require a given end-use quality, which depends on seed storage protein (SSP) concentration and composition. Expression of SSP genes is mainly regulated by transcription factors (TFs), which specifically bind cis-motifs in the promoter region. This first level of the transcriptional regulation of SSP genes implies TFs able to bind cis-motifs included in their promoter. In barley, eight TFs are involved in this regulation. In wheat, all these TFs are found (SPA, SHP, PBF, SAD, GAMYB, MYBS3, MCB1 and FUSCA3). Five of them have been demonstrated to regulate SSP synthesis. These TFs could be also regulated by transcriptional proteins (second level of regulation).

Currently, knowledge on the TFs able to modulate the expression of genes coding for SPA, SHP, PBF, SAD, GAMYB, MCB1, MYBS3 and FUSCA3 is poor. Therefore, we aimed at identifying the TFs involved in the second level of regulation (called candidate TFs), studying their polymorphism using exome data provided by Whealbii project (https://www.whealbi.eu/fr/). Polymorphisms were then used to study their effects on the SSP composition content and composition by association mapping.

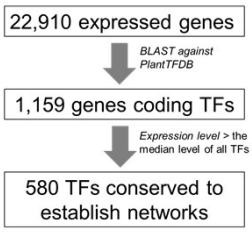


What are the candidate TFs (level 2) involved in the regulation of the target TFs of the first level ? Could they modulate SSP composition?

IDENTIFICATION OF CANDIDATE TRANSCRIPTION FACTORS BY A NETWORK-BASED STRATEGY

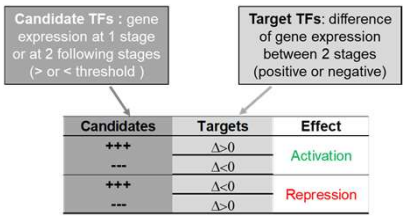
RNAseq data were obtained in collaboration with the BreedWheat program (https://breedwheat.fr/). mRNAs were extracted from grains of *Triticum monococcum*, a diploid wheat model species. Grains were harvested at different developmental stages during filling (300, 400, 500 and 600°Cday after anthesis) from plants cultivated under controlled conditions. Genes coding for TFs were extracted from the set of genes expressed in the grain by blast against PlantTFDB (http://planttfdb.cbi.pku.edu.cn/). These data were used to infer gene networks with the RulNet platform (http://rulnet.isima.fr). RulNet allows the connection between different -omic entities according to rules defined by the scientists with a biological meaning. Rulnet makes also possible to focus on rules involving attributes of special interested, which have to be declared as central attributes.

1-Putative candidate TFs selection



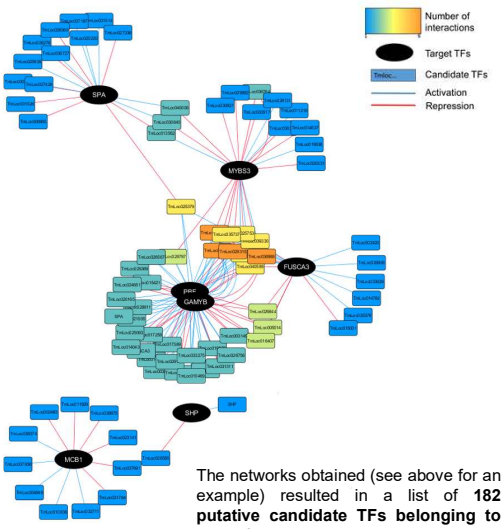
In the following, we focussed on the 580 TFs representing 53 TF families.

2-Rules for network construction between TFs



Three parameters reflect the quality of the rules: the support, the confidence and the lift. Two identities are connected if the support, the confidence, the lift are > at 0.2, 0.9 and 1.5 respectively.

3-Identification of candidate TFs

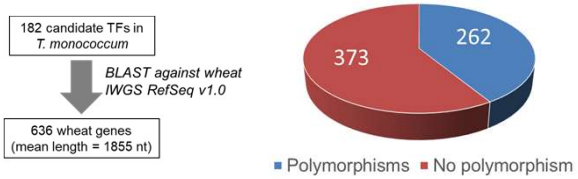


The networks obtained (see above for an example) resulted in a list of **182 putative candidate TFs belonging to 36 TF families.**

ALLELE MINING AND ASSOCIATION TO VALIDATE CANDIDATE TRANSCRIPTION FACTORS

The wheat orthologs/paralogs of *T. monococcum* candidates were searched by a blast analysis using sequences of the genes coding these TFs against the wheat pseudomolecule (The International Wheat Genome Sequencing Consortium, 2018). The blast results were analysed to find the coordinates of the orthologs and paralogs of candidate TFs on the wheat pseudomolecule. These positions were used to extract polymorphisms within the wheat sequences of all candidate TFs from the variant file produced by Whealbi.

1-Wheat gene candidates search and polymorphisms identification



262 genes (40%) contained 1,386 polymorphisms (SNPs) and 120 insertion-deletions (size >1 nucleotide) i.e. 1 polymorphism for about 313 nucleotides. 60% of genes contained no polymorphism because they were not captured or their sequence is conserved.

2-Association analysis

Genetic association was performed in a part of the wheat collection studied in Whealbi (105 lines), which was phenotyped for SSP content and composition by Plessis et al (2013) in three environments (at Clermont-Ferrand, Le Moulon with a high level of nitrogen fertilisation and Le Moulon with a low level of nitrogen fertilisation). Briefly, the traits analysed concerned total SSP, and all SSP fraction i.e. total gliadins and glutenins, high and low molecular weight glutenins, α/β-, δ-, w1-2 and w5 gliadins.

Eighteen candidate TFs were found to be associated at pvalue=0.001 with at least one trait concerning SSP composition.

Nine candidate TFs are associated with gliadins (indicate in blue in the table)

The TF indicated in bold characters is significantly associated with traits related to N content, gliadins and glutenins synthesis, in particular the gliadin to glutenin ratio or to the high molecular weight to low molecular weight glutenin ratio, which were important for the technological quality of wheat.

T. monococcum candidate TFs	Family	Chromosomal location of its wheat homolog
TmLoc039820	C2H2 family protein	chr1A
TmLoc023107	C2H2 family protein	chr1B
TmLoc023394	C2H2 family protein	chr4A
TmLoc033275	C3H family protein	chr3B
TmLoc040006 *	C3H family protein	chr7B
TmLoc012338	C3H family protein	chr7B
TmLoc014043	bZIP family protein	chr1B
TmLoc027274 *	bZIP family protein	chr7A
TmLoc030640	HB-PHD family protein	chr1A
TmLoc013010	HB-PHD family protein	chr3A
TmLoc026756	Trihelix family protein	chr1B
TmLoc041873	NAC family protein	chr3A
TmLoc041017	WRKY family protein	chr3B
TmLoc010838	GRAS family protein	chr3B
TmLoc021005	TALE family protein	chr5A
TmLoc038798 *	ARR-B family protein	chr6B
TmLoc020220 *	SBP family protein	chr7B
TmLoc013461	MYB_related family protein	chr7B

** indicates associations found in several locations*

CONCLUSIONS AND PERSPECTIVES

The network based-strategy resulted in a list of 182 TFs putatively involved in the regulation of TFs regulated SSP synthesis. Eighteen candidate TFs were associated with at least one trait concerning SSP composition. These associations reported have to be considered with caution because only 105 accessions were used. Nethertheless, some of them may be considered robust as they were found in several locations and/or for several related traits. Finally, storage protein composition could be modified by acting at the transcriptional level. We have to confirm the results with larger collections. Findings markers in genes without polymorphism could allow their statistical validation. In addition, genes validated by association mapping must be functionally validated. Thus, our strategy has reduced the number of candidate genes which was necessary before biological approaches.

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/ 2007-2013) under the grant agreement n°FP7- 613556, Whealbi project (http://www.whealbi.eu/project/).

REFERENCES

Vincent J., Martre P., Gouriou B., Ravel C., Dai Z., Petit J.-M. and Pailloux M. (2015) RulNet: A web-oriented platform for regulatory network inference, application to wheat –omics data. PLoS One, 10, e0127127

Plessis A., Ravel C., Bordes J., Balfourier F. and Martre, P. (2013) Association study of wheat grain protein composition reveals that gliadin and glutenin composition are trans-regulated by different chromosome regions. J. Exp. Bot. 64, 3627–3644

The International Wheat Genome Sequencing Consortium (IWGSC) (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 36: Issue 6403, eaar7191. DOI: 10.1126/science.aar7191