AARHUS
UNIVERSITET
INSTITUT FOR MOLEKYLÆRBIOLOGI OG GENETIK

FAUPE – Forbedring af Afgrødernes Udbytte og Produktionsmæssige Egenskaber

Arbejdspakke 2 – NUE genekspressionsanalyser i byg

STØTTET AF
promilleafgiftsfonden
for landbrug

The main objectives of our activities have been to develop new, effective screening methods and to provide genomics tools to investigate Nitrogen Use Efficiency (NUE) in barley and drought tolerance in grasses. Regarding drought tolerance, we performed greenhouse experiments in perennial ryegrass and to study NUE we performed field experiments at the experimental farms in Taastrup, in collaboration with the Department of Plant and Environmental Sciences, Copenhagen University.

Transcriptomic response to nitrogen application in barley

A nitrogen gradient field plot experiment was performed using 6 barley cultivars in summer 2015. Plots for each cultivar were placed in 3 replicates in three blocks. Liquid nitrogen fertilizer was dispensed pre-emergently in gradient doses along plots as shown in Figure 1. Two medium to high N-efficient cultivars (Tocada and Invictus) and one low efficiency cultivar (Imperial) were selected for sampling. Two sampling replicates per plot were collected at six fixed nitrogen concentrations within a 6 weeks period starting 1st July 2015. Two developmental stages were chosen for sampling: Growth stage 5.9 and mature stage 9.2 (Zadocks scales GS59 and GS92 respectively). Flag leaf and spike samples were taken from the same plants, flash-frozen in liquid nitrogen and stored separately at -80ºC. In total, 648 samples were collected.

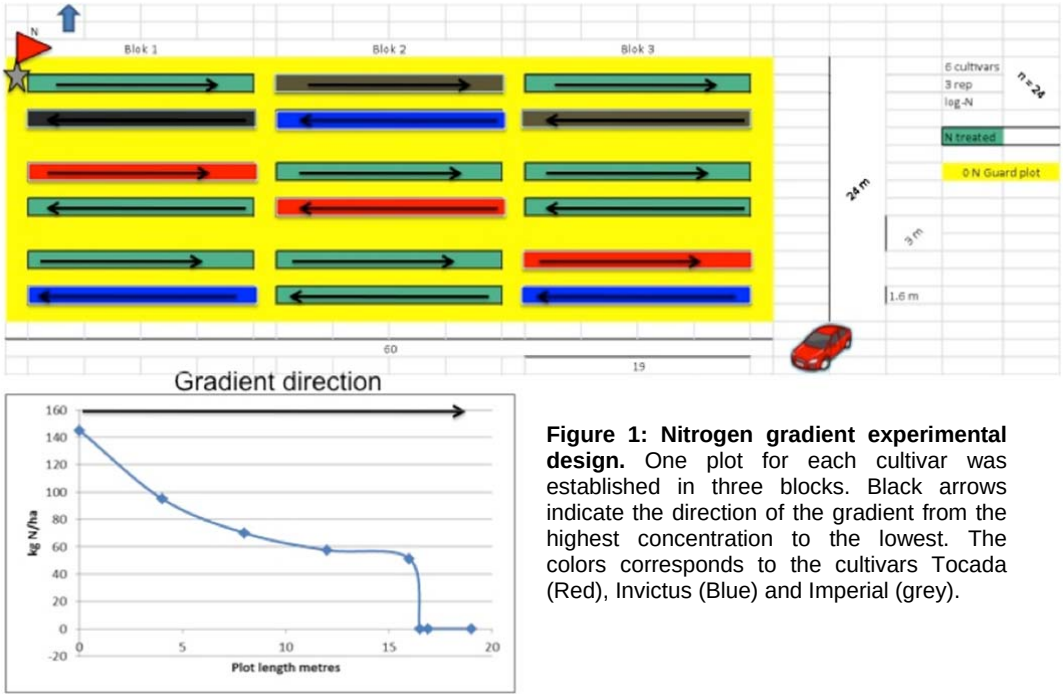


Figure 1: Nitrogen gradient experimental design. One plot for each cultivar was established in three blocks. Black arrows indicate the direction of the gradient from the highest concentration to the lowest. The colors corresponds to the cultivars Tocada (Red), Invictus (Blue) and Imperial (grey).

At the end of the sampling period, the cultivar Imperial was completely lodged. Therefore we decided to extract RNA from the cultivars Tocada and Invictus only. Five nitrogen levels were selected: 0, 50, 70, 95 and 145 kg/ha. Samples simultaneously collected from 3 replicate plots were treated as source materials for 3 biological replicates for RNAseq experiments. Two time points (designated as T1 and T2) were chosen for further analyses. T1 represented a rapid expansion phase GS59, in which barley plants have the highest N requirements, while T2 represented the end of the senescence stage GS92, in which N is re-mobilized from the foliage leaves to maturing grains. Total RNA samples were prepared from flag leaf samples and from a pool of 8 grains from the middle part of the spikes in each individual plant.

Paired-end Illumina HiSeq sequencing data were obtained for 120 total RNA samples. Raw sequence reads (2x100 nt) were subjected to adapter trimming and quality filtering, resulting in 693,547,158 cleaned paired reads for the variety Invictus and 689,645,450 cleaned paired reads for

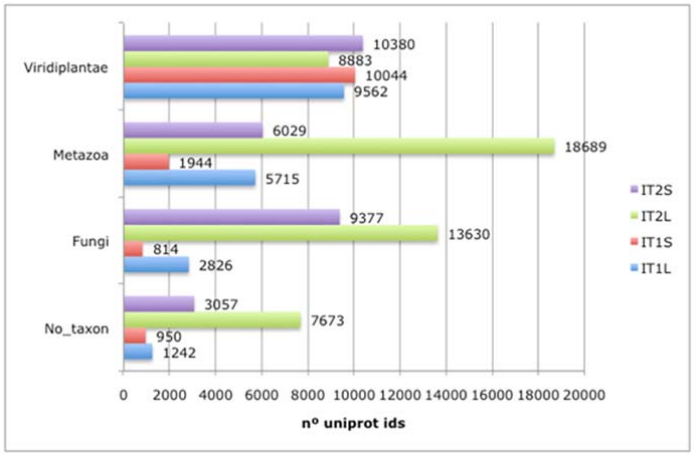
the variety Tocada.

For data analysis, three different strategies were applied:

1. DE analysis using de novo assemblies for each cultivar as a reference

Only 3 differentially expressed (DE) transcripts were found at T1, while several hundred at T2. This was most probably due to the fact that biotic contamination (pathogen infection) of the senescent tissue samples collected at time point T2 was extremely high (Figure 2).

Figure 2. Taxonomic groups identified in the cultivar Invictus. *De novo* assembled transcripts were translated to protein sequences, and the longest ORF was used to query the Uniprot database. The graph represents the number of proteins belonging to plants and other taxonomic groups in the flag leaf at T1 (IT1L) and T2 (IT2L), and in the spike at T1 and T2 (IT1S and IT2S).



2. DE analysis using transcript reference sequences from the barley genome

For this approach 64.095 transcript reference sequences from the barley genome (*Hordeum_vulgare.ASM32608v1.31*) was as used as a common reference for both cultivars. This strategy provided two main advantages:

- Reads originating from environmental contamination could be excluded, and
- Using common references allows more robust and reliable assessment of expression level differences.

The software STAR Aligner was used to map reads to the reference and the Trinity pipeline was used to detect differentially expressed genes (DEGs) at different nitrogen levels compared to the control (0 kg/ha added nitrogen). Using this approach only one single DEG could be found. Among replicates a high variability could be detected. This variability could be greatly reduced by excluding outlying samples from the analysis (1 sample out of 15 samples per tissue). Figure 3 shows the number of DEGs genes detected before and after removing outlying samples.

Genotype	sampling date	Tissue	DE genes before removing outlayers	DE genes after removing outlayers
Invictus	T1	Leaf	1	1 (MLOC37763) upregulated in [N]=50kg/ha and N=95kg/ha
Invictus	T1	Spike	0	1 (MLOC_62754) downregulated in [N]=50kg/ha
Invictus	T2	Leaf	0	5
Invictus	T2	Spike	0	293
Tocada	T1	Leaf	1	1 (MLOC37763) upregulated in [N]=70kg/ha
Tocada	T1	Spike	0	0
Tocada	T2	Leaf	0	9
Tocada	T2	Spike	0	1

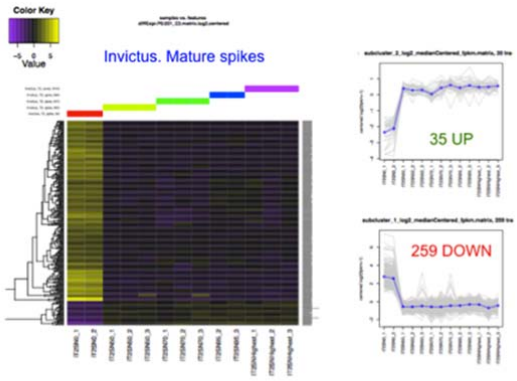


Figure 3. DE genes in response to increasing nitrogen concentration. Table: number of DE genes identified before and after removing outliers. Graph: DE genes in Invictus mature spike.

In flag leaves at T1 we identified a gene (MLOC_37763) that was up-regulated in both cultivars. It represents a Delta1-pyrroline-5-carboxylate synthase1 (P5CS1) gene which catalyzes the first two steps of the proline biosynthetic pathway and plays a central role in the regulation of this process in plants. Available information about the function this gene suggests a role in drought stress response. However, it was also found that this gene might be induced by nitrogen starvation. In addition, it has been shown that exogenous nitrogen supply enhances proline production in transgenic tobacco plants over-expressing this gene. Proline acts as one of the main reserve sources for carbon, nitrogen and energy during stress recovery (Kishor et al., 2005).

In spikes at T1 a down-regulated gene in Invictus (MLOC_62754) was found. It is a Type2 endoglucanase. To our best knowledge, this enzyme is involved in nitrogen uptake (NU), but it also plays an important role in plant development and reproduction. In the cultivar Tocada no DEGs could be found at this time point.

Despite the fact the nitrogen fertilzer will be applied in the early stage of plant development, we found more transcriptomic changes at the maturing stage. In senescent flag leaves at T2 we found 5 genes that showed increased expression responding to increasing nitrogen concentrations. Among these genes there is a Yellow Leaf Specific gene9 (MLOC_70921), which is similar to the known stress response gene YLS9. Other interesting genes from this group are nitrogen starvation/assimilation related genes such as Nitrate transporter NRT2.1 (Chakraborty et al., 2015). We could also identify a gene for a Reticuline oxidase-like protein (MLOC_77132), which is a representative of flavoproteins induced after by nitrogen addition in *Arabidopsis* seedlings (Scheible et al., 2004). Interestingly, we could identify a Basic 7S globulin gene (MLOC_26558) which was never found as expressed in leaves before. 7S globulin subunits are involved in the distribution of nitrogen sources in soybean seeds (Yamada et al., 2014). Further interesting genes are a putative CPRD2 (MLOC_67318), which is highly similar to the rice inactive Tetrahydrocannabinolic acid synthase (also considered to be a stress protein). Finally, MLOC_17627 is a NEN1-like protein. NEN proteins are NAC-dependent exonucleases required for enucleation process during phloem sieve element differentiation and are required for long distance allocation of nutrients. *Arabidopsis* mutants deficient for this gene develop shorter roots (Furuta et al., 2014, Heo et al., 2017).

Interestingly, the highest number of DE genes where found in maturing spikes. Altogether at different nitrogen levels we found 35 up-regulated and 259 down-regulated DEGs. We obtained functional annotation information for 80. 5% of these genes using the Blast2GO software and GO terms could be assigned to 67% of these genes. Twenty percent of the DEGs were associated with cellular nitrogen compound metabolic processes. The most represented molecular functions were protein binding and oxidoreductase activity. However, under down-regulated genes the GO category DNA binding proved to be over-represented. Only a few DEGs could be found in the cultivar Tocada and the majority of those genes showed variable expression pattern among different

nitrogen concentrations. To improve the robustness of DE analysis we applied another approach as described below:

3. DE analysis by parallel evaluation of replicates

In order to reduce variability among RNA sample replicates that represent replicate field plots and are prone to inherit high level of random variations,we used during this strategical approach in which we considered the three biological RNASeq data replicates as data of three independent parallel experiments. Expression profiling were carried out separately usingn the three experimental datasets. Quantitative gene expression data at different nitrogen concentration treatments were compered to the corresponfing cultivar-specific control (0 kg/ha added nitrogen). DEGs that consequently showed the same type of expression differences in all of the three parallel evaluations were recorded. The results of this comparative analysis is shown in Table 1. These results again reflect a relatively high variability in number of DEGs among replicates, especially in senescent leaves (IT2L and TT2L). Typically less than 10% of the up-regulated genes are common in the three replicates at a given nitrogen concentration.

Table 1. Number of UP-regulated genes compared to [N]=0 in every tissue and nitrogen concentration.
I=Invictus, T=Tocada, T1=GS55, T2=GS92, L=leaf, S=spike.

			number of UP-regulated genes compared to N=0kg/ha										
			N=50kg/ha		N=70kg/ha		N=95kg/ha		N=145kg/ha		pregulated at every (N		
Tissue	plot	Total non-redundant DE genes	per plot	common in 3 plots	per plot	common in 3 plots	per plot	common in 3 plots	per plot	common in 3 plots	per plot	common in 3 plots	Gene IDs
IT1L	R1	2615	538	31	477	72	599	64	407	23	79	4	MLOC_4831, **MLOC_37763**
	R2	2895	521		652		564		657		110		MLOC_2259, MLOC_70149
	R3	3030	735		893		724		961		186		
IT1S	R1	3094	514	116	473	41	726	26	637	12	71	1	MLOC_66445
	R2	3546	1019		891		608		1209		267		
	R3	3179	629		601		388		870		56		
IT2L	R1	8647	1432	1	1237	4	2105	6	657	0	41	0	
	R2	7924	10		837		774		1377		0		
	R3	2003	313		398		264		184		15		
IT2S	R1	4284	698	19	686	20	619	8	712	15	72	1	MLOC_68931
	R2	3962	724		720		684		714		195		
	R3	2561	429		597		578		430		67		
TT1L	R1	3149	825	17	771	72	514	33	761	90	101	4	MLOC_4831, **MLOC_37763**
	R2	2877	715		604		476		808		125		MLOC_12678, MLOC_53443
	R3	3436	593		785		669		670		151		
TT1S	R1	3149	718	5	602	16	716	7	523	23	78	0	
	R2	3418	1154		1138		1141		1126		545		
	R3	2676	543		804		519		600		84		
TT2L	R1	8425	1411	2	2026	0	1061	0	1671	0	56	0	
	R2	1835	620		325		1185		75		2		
	R3	391	141		42		17		34		0		
TT2S	R1	2932	493	4	1079	8	464	3	410	4	50	0	
	R2	2984	677		495		915		592		79		
	R3	2914	508		860		556		681		76		

In green leaves (time point T1) only 4 genes could be identified that consequently showed up-regulation. However, two of these genes are common in both cultivars and one (MLOC_37763) was the already found in the previous analysis. Filtering out DEGs that only occurred in one of the three parallel evaluations and keeping those that show the same direction of expression differences in at least replicaes of any given treatment can deliver a considerably higher number of potentially interesting DEGs.

These data illustrate the challenges and pitfalls of quantitative transcript profile analyses related to complex traits under field conditions and the experiences obtained during the evaluation of present data provide valuable information for designing similar future experiments.

References

Chakraborty, N., Sharma, P., Kanyuka, K., Pathak, R.R., Choudhury, D., Hooley, R., and Raghuram, N. (2015). G-protein α-subunit (GPA1) regulates stress, nitrate and phosphate response, flavonoid biosynthesis, fruit/seed development and substantially shares GCR1 regulation in A. thaliana. Plant molecular biology 89, 559-576.

Furuta, K.M., Yadav, S.R., Lehesranta, S., Belevich, I., Miyashima, S., Heo, J.-O., Vatén, A., Lindgren, O., De Rybel, B., and Van Isterdael, G. (2014). Arabidopsis NAC45/86 direct sieve

element morphogenesis culminating in enucleation. Science 345, 933-937.

Heo, J.-O., Blob, B., and Helariutta, Y. (2017). Differentiation of conductive cells: a matter of life and death. Current Opinion in Plant Biology 35, 23-29.

Kishor, P.K., Sangam, S., Amrutha, R., Laxmi, P.S., Naidu, K., Rao, K., Rao, S., Reddy, K., Theriappan, P., and Sreenivasulu, N. (2005). Regulation of proline biosynthesis, degradation, uptake and transport in higher plants: its implications in plant growth and abiotic stress tolerance. Curr Sci 88, 424-438.

Scheible, W.-R., Morcuende, R., Czechowski, T., Fritz, C., Osuna, D., Palacios-Rojas, N., Schindelasch, D., Thimm, O., Udvardi, M.K., and Stitt, M. (2004). Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes, and the regulatory infrastructure of Arabidopsis in response to nitrogen. Plant physiology 136, 2483-2499.

Yamada, T., Mori, Y., Yasue, K., Maruyama, N., Kitamura, K., and Abe, J. (2014). Knockdown of the 7S globulin subunits shifts distribution of nitrogen sources to the residual protein fraction in transgenic soybean seeds. Plant cell reports 33, 1963-1976.