

Doctoral (PhD) thesis

István Monostori

Gödöllő

2021



Hungarian University of Agriculture and Life Sciences

Association mapping-based approach to study nitrogen use efficiency, its genetic components in winter wheat

István Monostori

Gödöllő

2021

Doctoral School: Doctoral School of Plant Science

Science: Agricultural and Horticultural Sciences

Head of Doctoral School: Lajos Helyes DSc

Professor, Doctor of the Hungarian Academy of Sciences

Hungarian University of Agriculture and Life Sciences

Supervisor: Attila Vágújfalvi PhD

Senior research fellow

Eötvös Loránd Research Network, Centre for Agricultural Research



.....

Dr. Lajos Helyes

Head of Doctoral School

.....

Dr. Attila Vágújfalvi

Supervisor

1. BACKGROUND AND OBJECTIVES

Wheat (*Triticum aestivum* L.) became one of the most important food grain source for humans. Considering the predicted population growth, the increase in per capita consumption and the changes in diets, the global production of agricultural products needs to be highered by 60 percent by 2050.

To increase crop productivity the amount of applied nitrogen (N) fertiliser is expected to double by 2050, unless the nitrogen use efficiency of the plants increases significantly. Nevertheless, the utilisation of N fertilisers is rather inefficient nowadays since approximately 50-70% of the applied N vanishes from the plant - soil system polluting the environment. Therefore, a more efficient use of nitrogen fertilisers is necessary. The utilisation of N efficient cultivars that exploit more efficiently the soil nutrient resources can provide one way to increase nitrogen use efficiency (NUE). Traditional breeding practices and biotechnology methods can be used to breed lines with better nitrogen use efficiency that can achieve adequate yield even under poor N supply thus showing a good fertilizer reaction.

Since NUE is determined by multiple genetic factors and influenced heavily by the environment and by the genotype by environment interaction, its genetic dissection is quite challenging. Therefore, relatively little information is available on the genetic background of the nitrogen use efficiency of winter wheat, hence a better understanding of the genetic basis and its regulation is fundamental.

The aim of our study was to map the key genetic components involved in the regulation of nitrogen use efficiency, moreover to understand the interactions between NUE-related traits in winter wheat. To investigate NUE-related traits and their genetic and environmental regulation, field trials were evaluated on a Central European wheat collection of 93 cultivars at two N input levels across three seasons.

The following objectives have been set:

- Characterisation of the effect of N fertiliser treatment on nitrogen use efficiency and its components. Furthermore, investigation of N fertilizer effect on traits determining the development and productivity of winter wheat.
- Examination of the relationship between the agronomic traits in order to identify the key characteristics that determine nitrogen use efficiency
- Studying the impact of environmental factors and genetic polymorphisms on the investigated agronomic traits
- Characterisation of the genetic diversity, the population structure of the winter wheat collection used for the association analysis, moreover, the assessment of the linkage disequilibrium
- Studying the genetic background of nitrogen use efficiency in our winter wheat collection using genome-wide association analysis

2. Materials and Methods

2.1 Experimental design

Field trials were evaluated in a split plot design, in three replicates, at two nitrogen input levels across three consecutive seasons. N treatment was considered as main plots and varieties as sub plots. In the N0 treatment, only the naturally occurring nitrogen was available in the soil, while in the case of N120, N was top dressed at growth stage (120 kgN hectare⁻¹). Altogether 93 bread wheat varieties were phenotyped in sub plots.

2.2 Phenotypic evaluation

The entire collection was evaluated for 16 agronomically important or N input-related traits in six environments (3 year × 2 N levels). The investigated traits were: heading date (HD), plant height (PH), grain yield (GY), thousand grain weight (TGW), weight of straw (SY), spike number per meter (SN), grain number per spike (GN), harvest index (HI), N harvest index (NHI), grain protein content (GPC), NUE, nitrogen uptake efficiency (NUpE), nitrogen utilization efficiency (NUtE), amount of N taken up by the whole plant (NUp_{full}), amount of the N harvested in grain (NUp_{grain}), grain N accumulation efficiency (GNACE).

The GY, HD and PH were scored on a per plot basis, while GN, SN, TGW and SY were evaluated from above-ground biomass samples collected from a representative row of each plot. The straw and grain samples were milled and N content was determined by Dumas method using Rapid NIII nitrogen analyser. The amount of the N harvested in grain and in straw was calculated by multiplying the N concentration obtained from Rapid NIII analysis with the amount of GY and SY, respectively. NUE is defined as GY divided by the sum of naturally available N content in the soil and the amount of N allocated with fertiliser (kg hectare⁻¹). NutE is the GY divided by amount of N taken up by the whole aboveground plant. NUpE is the amount of N taken up by the whole aboveground plant divided by the sum of naturally available N content in the soil and the amount of N allocated with fertiliser. NHI is the ratio of N content harvested in the grain to amount of N taken up by the whole aboveground plant, which is a measure of N translocation efficiency. GNACE is the N content harvested in the grain divided by the sum of naturally available N content in the soil and the amount of N allocated with fertiliser, which serves as a measure of the overall efficiency.

2.5 Genotyping, analysis of population structure

Genotyping was performed with the wheat DArTseq® platform by Triticarte Pty. Ltd. This marker set was filtered on the basis of individual marker-related statistics. After filtering marker data, marker with known chromosomal position were selected and used for linkage disequilibrium (LD) and association analysis.

The population structure has been investigated using several different methods to verify the results. After the usage Bayesian statistical approach, the kinship relationships were also examined by phylogenetic and principal coordinate analysis; the results have been validated using publicly available pedigree databases.

2.6 Linkage disequilibrium analysis

Genome wide linkage disequilibrium analysis was performed using the LD function in the software TASSEL. LD decay was estimated separately in all chromosomes and also for the three genomes (A, B and D) in the entire population. To estimate the LD decay, significant r^2 values were plotted against the genetic distance between the marker pairs and a second degree smoothed Loess curve was fitted using SPSS 22 software. The pattern of linkage disequilibrium has been visualized on each chromosome by TASSEL software.

2.7 Genome wide association analysis

Genome wide association analysis was performed using TASSEL software for 11 investigated traits in all environments (three cropping seasons, two N fertilisation levels) and for the response to N fertilisation for 8 selected traits in order to identify genomic regions involved in the response to N fertilisation. Four different statistical models were adopted to calculate P-values for marker-trait associations to avoid spurious associations.

The principal component analysis and kinship-matrix used as covariate were calculated with TASSEL software. The Q-matrix was generated based on the results of the Structure software. The critical threshold for assessing the significance of marker-trait associations were calculated by false discovery rate separately for each trait in each year. An marker-trait association was defined significant if the calculated q-value passed the FDR threshold for a given trait in all of the four models.

3. Results and Discussion

3.1 Characterisation of the phenotypic traits

Large phenotypic differences were observed in nitrogen use efficiency, its main components and in other traits that determine the plant's adaptability. Genotype effect was highly significant for 14 traits, which reflects the possibility to further development of these traits.

The environmental effect was the main factor controlling most of the studied traits. For most of the traits greater differences have been found between seasons than between N treatments. However, N fertilisation had significant effect on most of the studied traits. The largest differences between two N treatments were detected for grain yield, $NU_{p_{\text{grain}}}$, and $NU_{p_{\text{full}}}$.

The large environmental effect experienced in our field trials was primarily due to the significant differences in the environmental conditions of the three seasons. "Environmental effect" in the actual experiment indicates not just the considerable effect of the weather conditions but it also interprets the differences in the soil N supply. Although, the soil type at each location was chernozem, the average N content significantly differed between the years. Based on these huge differences, the three experimental years were considered as three different environments, rather than replicates, therefore, a total of six experimental "settings" (3 environments x 2 N-levels) could be tested in our experiment.

A high variability in nitrogen use efficiency has been observed in the population in all environments, demonstrating for the breeders that the population has adequate potentials to develop NUE. In accordance with the results of previous studies, we recorded the highest NUE values in the season (2012/2013) with the lowest N availability, while in the season (2013/2014) with excellent N availability recorded the lowest NUE values.

Our results are consistent with previous studies which declared that contributions of NU_{pE} and NU_{tE} to the overall NUE depends mostly on the level of N supply. In our experiment, the differences in nitrogen use efficiency were determined a greater extent by NU_{pE} than by NU_{tE} in all seasons. When the available N was rare, the ability to absorb N is more important, which is mostly related to root characteristics. On the contrary, at high N levels, variation in NUE is mainly due to differences in NU_{tE} . The greatest importance of NU_{tE} was in the 2013/2014 season, which was characterised by very good N supply, but even then NU_{pE} was the main factor determining NUE.

The harvest-index and nitrogen harvest-index traits explain a significant part of the variance of NUtE in the seasons 2013/2014 and 2014/2015. This indicates that high HI and NHI has great importance in development of NUtE.

In our current study, in line with the results of previous studies, GPC showed negative correlation with NUtE, which may make it difficult to improve nitrogen use efficiency. Because if grain yield is increased at fixed NUpE, it can cause a decline of GPC, unless NHI is increased.

3.2 Genetic diversity, population structure and linkage disequilibrium

In our current study, the B genome marker coverage was the highest, the A genome had a similar level of coverage, while D genome had significantly lower polymorphic marker coverage. The distribution of markers between three wheta genomes was consistent with previous studies results. An average LD decay distance for our total population was found to be 9 cM. Taking into consideration this marker density was sufficient to identify MTAs almost in the entire genome with the exception of some chromosomes of the D genome. Considering our population size, there was a relatively rapid decrease in LD, so in general terms a small LD was observed. This phenomenon can be explained by our Central European collection with a relatively high genetic diversity. The extent of LD decay showed a high degree of variability in different regions of the genome. Recombination deserts were identified in some regions of the genome: the largest LD blocks were indentified on 1B and 2D chromosomes. The long-distance LD blocks (recombination deserts) on the 1B and 2D chromosomes can be associated with the widespread use of 1BL.1RS translocation as a resistance source in hungarian breeding programmes and the use of the dwarfing allele of the *Rht8* gene on some lines.

Two subpopulations (Sp1 and Sp2) were identified with the Bayesian approach implemented in Structure software and confirmed by PCoA. The majority of the 79 cultivar belonged to Sp1, while Sp2 incorporated the remaining 14 cultivars. This structure is consistent with the breeding history of the cultivars.

3.3 Genome-wide mapping of agronomic and N-related traits

Altogether, 183 MTAs were identified that may play a role in the genetic regulation of winter wheat nitrogen use efficiency, its components and agronomically relevant traits. Seventeen chromosome region have been identified where LD regions of markers of different traits overlap, so these regions determine several traits. The distribution of identified MTAs between genomes followed the distribution of polymorphic markers between genomes, so B genome showed the

highest number of MTAs (93) and it was followed by the A genome (75), while only 15 MTAs were found on the D genome.

Since the weather, soil N content and precipitation distribution of all three seasons of our experiment differed significantly, different numbers of MTAs were detected between the cropping seasons in our experiment. In case of most investigated traits, it can be generally stated that we have identified more marker-trait associations under better nitrogen supply (more favourable conditions), than nitrogen deficiency.

We found that most of the identified MTAs were environmental specific and also that most of the MTAs were detected in one environment only. We can conclude that these environment specific MTAs reflect regions associated with adaptation to specific environmental conditions, their expression is influenced by environmental factors in a unique way, so these genes may have environmentally dependent genetic regulation. These environment specific MTAs reflect regions associated with adaptation to specific environmental conditions. Despite the environmental variability, 4 markers were significant for the same trait in two different environments. These genomic regions are the best candidates for more extensive studies because they may be involved in constitutive processes for agronomically important trait.

Three markers were associated in both treatments in the same year for the same trait indicating that these loci may not (or just partially) influenced by N fertilisation. Twenty-seven markers were associated with multiple traits, so these markers are not trait-specific. This situation occurs when high LD exists between loci for different traits or when the pleiotropy is the main cause of the genetic correlation between traits. Identification of genetic regions involved in determination of multiple traits can help us understand the complex regulatory mechanisms of these traits.

Markers associated with grain yield and its components may be suitable candidates to increase NUE by increasing winter wheat yield. Altogether, 12 loci were identified that are involved in the variation of grain yield, and several of these MTAs correspond to QTLs already reported in other studies, using different wheat mapping populations.

Based on the results of our association analysis, most marker associated with nitrogen use efficiency (and its components) were found to be environment specific. We identified 12 significant markers for nitrogen use efficiency across all treatments and cropping seasons. We observed that most markers associated with nitrogen use efficiency were also associated with nitrogen uptake efficiency and/or grain nitrogen accumulation efficiency. Either the marker is located inside a significant chromosome region, which also representative of the former traits. The identified environment specific MTAs can facilitate the exploration of the possible structural

or regulatory genes controlling NUE under different, specific environmental conditions. Eight significant chromosome regions associated to nitrogen uptake efficiency have been identified on chromosomes 1B, 2B, 3B, 5A, 5B and 7A.

In our study 10 MTAs were spread along on 7 chromosomes for NUtE. Most markers were trait-specific, while the remaining markers were significant with GPC. We identified several previously reported MTAs, which supports the results of our research. Several previously not published QTLs and chromosome regions have been identified. Some of these QTLs and chromosome regions could help breeders to increase nitrogen use efficiency in the future.

4. Conclusions and Recommendations

To satisfy future demands, the increase of wheat yield and yield stability is needed. Further significant increases in nitrogen fertiliser doses cannot be reasonable from an environmental and economic point of view due to increasing losses in the soil-plant system. One possibility to increase nitrogen use efficiency can be provided by N efficient genotypes, which are bred using traditional breeding and biotechnology methods. These genotypes are capable of achieving relatively high yields even under low nitrogen availability, but they also have a good fertilizer reaction, so they efficiently use the applied fertilizer.

Although it has huge importance, the development of nitrogen use efficiency is challenging. Therefore our aim was to identify the key characteristics involved in the development of nitrogen use efficiency in winter wheat and also to identify the regulatory genetic components involved in this complex phenomenon.

We confirmed that nitrogen use efficiency was mainly determined by nitrogen uptake efficiency. Meanwhile, the efficient development of nitrogen uptake efficiency requires coordinated improvement of both nitrogen uptake and utilization efficiency. Therefore, in order to improve nitrogen use efficiency, it is necessary to put emphasis on high root density and post-flowering nitrogen uptake. In addition, to develop NUE one of the preferred ways is to improve nitrogen remobilisation.

During our field trials we observed a high degree of phenotypic variability for most of the traits, demonstrating that the population has adequate potential to develop NUE. The genetic dissection of nitrogen use efficiency and its components is quite challenging because they are influenced heavily by environmental factors by genotype by environment interactions as well. In addition, the genetic regulation of the N uptake and N transport is also complex as such, they are regulated not only transcriptionally but also at post-transcription and post-translational level, thus making the testing and evaluation very complicated.

We have chosen a genotyping by sequencing method to ensure the appropriate marker density from the gene rich regions of wheat genom. In our study, 183 markers were identified that may play a role in the genetic regulation of winter wheat nitrogen use efficiency, its components and agronomically relevant traits. Many of the regions have been previously reported, which supports the results of our research.

We found that most of the identified MTAs were environmental specific due to the different N availability and weather conditions between seasons leading to a high level of genotype \times environment interaction for most of the studied traits. Twenty-seven MTAs were found to be associated with multiple traits, which means that these loci can be considered as multi-trait loci. The other possible explanation is that these markers represent regulatory regions, or these markers are associated with genes with pleiotropic effects. Marker-trait associations observed in multiple environments are the best candidates for marker-based selection.

Our current results should be confirmed by further studies using collections with (much) larger number of genotypes representing a higher genetic variability. In this case, the genetic map resolution could be improved and the statistical force **could** be increased. Due to the resolution of our genetic map, the marker - associated with a given trait - may be located several cM distances from the functional gene(s). In order to better determine the position of the functional genes, it would be practical to further narrow the associated genome regions with new markers by fine mapping methods. The discovery of molecular markers, closely located to the locus that regulate NUE would allow the selection of NUE determining genes, without any expensive and time-consuming phenotyping processes.

Previous studies have shown that the modification of the expression of certain genes involved in the N metabolism is not a really effective way to improve NUE. A better approach is the simultaneous regulation of the network of genes responsible for nitrogen uptake, transport, assimilation and remobilisation. This aim can be achieved by the modulation of transcription factors, thus simultaneously regulating many genes of N metabolism, which can be **a** key method in the improvement of nitrogen use efficiency.

5. New scientific results

1. In our Central European winter wheat collection the nitrogen fertilizer treatment increased the grain yield, the nitrogen uptake and the amount of N remobilized into grain yield the best. It was observed that N treatment has no significant effect on flowering time, while it influenced the thousand grain weight, nitrogen utilization efficiency and nitrogen harvest-index in **a** year-dependent manner.

2. It was observed that nitrogen use efficiency was mostly determined by nitrogen uptake efficiency. It was demonstrated that under good nitrogen supply, the importance of nitrogen utilization efficiency has increased, which was most strongly determined by the harvest-index and the nitrogen harvest-index.
3. The phenotypic variability of most traits was mainly determined by environmental factors. However, the treatment influenced the phenotypic variance to the greatest extent for nitrogen use efficiency, nitrogen uptake efficiency and grain nitrogen accumulation efficiency.
4. Our genetic map, containing 3290 markers, covered a genetic distance of 5880 cM. The average distance between markers in the entire genome was 1.79 cM, while on genomes A, B and D there were 1.75 cM, 1.22 cM and 5.77 cM, respectively. A large variability of linkage disequilibrium has been observed through the genome. The largest recombination deserts have been identified on chromosomes 1B and 2D. The population structure and kinship analysis showed that our collection most likely consists of two subpopulations.
5. The genome wide association analysis proved that 130 markers are significantly associated with at least one trait in one environment. Altogether, 183 MTAs were identified that may play a role in the genetic regulation of winter wheat nitrogen use efficiency, its components and agronomically relevant traits. Seventeen significant chromosome regions have been identified within the distance of the mean linkage disequilibrium. Moreover, we also identified genetic components that determine nitrogen response in 8 agronomic traits.

6. Publications

Publications related to the dissertation

1. **Monostori I**; Szira F; Tondelli A; Árendás T; Gierczik K; Cattivelli L; et al. (2017) Genome-wide association study and genetic diversity analysis on nitrogen use efficiency in a Central European winter wheat (*Triticum aestivum* L.) collection. PLoS ONE 12(12): e0189265. **IF: 2.806**
2. **Monostori I**; Árendás T; Hoffman B; Galiba G; Gierczik K; Szira F; Vágújfalvi A: Relationship between SPAD value and grain yield can be affected by cultivar, environment and soil nitrogen content in wheat. Euphytica (2016) 211: 103. **IF: 1.626**

Other publications in international scientific journals with impact factor:

1. Boldizsár Á; Soltész A; Tanino; Kalapos B; Marozsán-Tóth Zs; **Monostori I**, Dobrev P; Vankova R; Galiba G. (2021) Elucidation of molecular and

hormonal background of early growth cessation and endodormancy induction in two contrasting *Populus* hybrid cultivars. *BMC Plant Biol* 21, 111.

2. Ivanizs L; **Monostori I**; Farkas A; Megyeri M; Mikó P; Türkösi E; Gaál E; Lenykó-Thegze A; Szőke-Pázi K; Szakács É; Darkó É; Kiss T; Kilian A; Molnár I (2019) Unlocking the Genetic Diversity and Population Structure of a Wild Gene Source of Wheat, *Aegilops biuncialis* Vis., and Its Relationship With the Heading Time. *Front. Plant Sci.* 10:1531. **IF: 4.298**
3. Tarawneh R A; Szira F; **Monostori I**; Behrens A; Alqudah A M; Thumm S; Lohwasser U; Röder M S.; Börner A; Nagel M (2019) Genetic analysis of drought response of wheat following either chemical desiccation or the use of a rain-out shelter. *Journal of Applied Genetics* 60:2 pp. 137-146. **IF: 1.725**
4. **Monostori I**; Heilmann M; Kocsy G; Rakszegi M; Ahres M; Altenbach SB; Szalai G; Pál M; Toldi D; Simon-Sarkadi L; Harnos N; Galiba G; Darko É (2018) LED Lighting – Modification of Growth, Metabolism, Yield and Flour Composition in Wheat by Spectral Quality and Intensity. *Front. Plant Sci.* 9:605. **IF: 4.298**
5. Gierczik K; Novák N; Ahres M; Székely A; Soltész A; Boldizsár Á; Gulyás Zs, Kalapos B; **Monostori I**; Kozma-Bognár L; Galiba G; Vágújfalvi A: Circadian and Light Regulated Expression of CBFs and their Upstream Signalling Genes in Barley. *International Journal of Molecular Sciences* (2017) 18:(8) 1828; **IF: 3.226**
6. Judit B; Maccaferri M; Cané M.A; **Monostori I**; Spitkó T; Kuti Cs; Mészáros K; Láng L; Pál M; Karsai I: Phenotypical and physiological study of near-isogenic durum wheat lines under contrasting water regimes. *South African Journal of Botany* (2017) 108: 248-255 **IF:1.427**
7. **Monostori I**; Szira F, Galiba G, Rakszegi M, Bálint AF: Micronutrient contents and nutritional values of commercial wheat flours and flours of field-grown wheat varieties-A survey in Hungary. *Cereal Research Communications* (2014) 42(2) 293-302. **IF₂₀₁₄: 0.607**

