



Candidate genes and genome-wide association study of grain protein content and protein deviation in durum wheat

D. Nigro¹ · A. Gadaleta² · G. Mangini¹ · P. Colasuonno² · I. Marcotuli² · A. Giancaspro² · S. L. Giove² · R. Simeone¹ · A. Blanco¹

Received: 24 August 2018 / Accepted: 19 December 2018 / Published online: 2 January 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Main conclusion Stable QTL for grain protein content co-migrating with nitrogen-related genes have been identified by the candidate genes and genome-wide association mapping approaches useful for marker-assisted selection.

Grain protein content (GPC) is one of the most important quality traits in wheat, defining the nutritional and end-use properties and rheological characteristics. Over the years, a number of breeding programs have been developed aimed to improving GPC, most of them having been prevented by the negative correlation with grain yield. To overcome this issue, a collection of durum wheat germplasm was evaluated for both GPC and grain protein deviation (GPD) in seven field trials. Fourteen candidate genes involved in several processes related to nitrogen metabolism were precisely located on two high-density consensus maps of common and durum wheat, and six of them were found to be highly associated with both traits. The wheat collection was genotyped using the 90 K iSelect array, and 11 stable quantitative trait loci (QTL) for GPC were detected in at least three environments and the mean across environments by the genome-wide association mapping. Interestingly, seven QTL were co-migrating with N-related candidate genes. Four QTL were found to be significantly associated to increases of GPD, indicating that selecting for GPC could not affect final grain yield per spike. The combined approaches of candidate genes and genome-wide association mapping led to a better understanding of the genetic relationships between grain storage proteins and grain yield per spike, and provided useful information for marker-assisted selection programs.

Keywords Genome-wide association mapping · Grain protein content · Grain protein deviation · Grain yield · GPC and grain yield relationships · Nitrogen-related genes · QTL mapping · SNP markers

Abbreviations

| | |
|-------|--------------------------|
| AlaAT | Alanine aminotransferase |
| ASN | Asparagine synthetase |
| CG | Candidate gene |
| GDH | Glutamate dehydrogenase |
| GOGAT | Glutamate synthetase |
| GPC | Grain protein content |

| | |
|------|----------------------------------|
| GPD | Grain protein deviation |
| GS | Glutamine synthetase |
| GWAS | Genome-wide association study |
| GY | Grain yield |
| GYS | Grain yield per spike |
| NIR | Nitrite reductase |
| NR | Nitrate reductase |
| NRT2 | Nitrate transporter |
| NUE | Nitrogen use efficiency |
| PPDK | Pyruvate orthophosphate dikinase |
| QTL | Quantitative trait locus |

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00425-018-03075-1>) contains supplementary material, which is available to authorized users.

✉ A. Gadaleta
agata.gadaleta@uniba.it

¹ Department of Soil, Plant and Food Sciences, Genetics and Plant Breeding Section, University of Bari, Bari, Italy

² Department of Agricultural and Environmental Science, Research Unit of “Genetics and Plant Biotechnology”, University of Bari, Bari, Italy

Introduction

Grain protein content (GPC) is the most important quality trait in bread (*Triticum aestivum* L.) and durum wheat (*Triticum turgidum* L. ssp. *durum*), determining

the nutritional value and the baking quality as well as the pasta-making technology characteristics of flour and semolina. Mature wheat grain has a low amount of storage proteins compared to grain legumes, generally ranging from 10 to 20%. During the past 3 decades, the increase in GPC has mainly been achieved by intensifying nitrogen (N) fertilization. Recently, however, the high costs of N fertilizers and the detrimental environmental impacts of nitrates loss from the soil have stimulated interest in the possibility of reducing the amount of N applied to cereal cropping systems while maintaining the high productivity of modern cultivars. The development of new wheat varieties with high GPC has, therefore, become a breeding priority).

The possibility of improving GPC and yield should be considered in the light of the major recent advances in the identification of genes that might play a key role in the N uptake from the soil, and of those controlling the enzymes of amino acid metabolism, potentially involved in N transferring to the protein in the grain. Nitrogen metabolism is a quite complex pathway, involving several processes such as nitrate uptake, reduction, assimilation into amino acids, and translocation throughout the plant. A number of genes and transcription factors are involved in these steps (reviewed by Balyan et al. 2016). Nitrate transporters play a central role in nitrate uptake. NRT2 (nitrate transporter 2) family members act as high-affinity nitrate transporters in roots and their interplay enables efficient use of soil nitrate when poorly available. Some authors found them to be involved in nitrate-regulated root development in *Arabidopsis* (Garnett et al. 2009). The up-taken nitrate is then first reduced by cytosolic nitrate reductase (NR) to nitrite, which is subsequently imported into the chloroplast and further reduced by nitrite reductase (NIR) into ammonium. Ammonium is then incorporated into organic molecules by the glutamine synthetase (GS) and glutamate synthetase (GOGAT) pathways (Lea and Azevedo 2007). Several studies have been carried out on these genes, pointing out a correlation with GPC QTL in durum wheat (Gadaleta et al. 2011, 2014; Nigro et al. 2013, 2014, 2016).

In addition to these central enzymes, others have been found to be somehow involved in N utilization, such as glutamate dehydrogenase, alanine aminotransferase, and asparagine synthetase, all involved in amino acid metabolism. Glutamate dehydrogenase (GDH) catalyzes the interconversion between glutamate and 2-oxoglutarate (Habash et al. 2007), while alanine aminotransferase (AlaAT) is involved in the synthesis and degradation of alanine (Shrawat et al. 2008). Asparagine synthetase (ASN) catalyzes the synthesis of asparagine (Asn) and glutamate from glutamine (Gln) and aspartate. Several ASN genes have been identified in durum wheat, and *ASNI* is up-regulated in response to some mineral deprivation (Gao et al. 2016).

Besides genes and enzymes previously reported, which have a direct role in the nitrogen metabolism, there are others, working as regulatory and signaling factors when specific conditions occur (Xu et al. 2012). NIN-like protein (NLP7) is a transcription factor involved in regulation of nitrate assimilation in higher plants, as a key player in plant nitrate signaling (Castaings et al. 2009). Nitrogen remobilization related genes may also have an important role in N content and final yield. Over-expression of the cytosolic pyruvate orthophosphate dikinase (*PPDK*) gene was found to increase the seed weight in *Arabidopsis* (Taylor et al. 2010). Specifically, it accelerates N remobilization from leaves and thereby increases rosette growth rate and seed weight as well as N content. All these described genes and enzymes activities might then be a target for NUE crop improvement.

Grain yield and GPC are complex trait determined by multiple quantitative trait loci (QTL) that interact with each other and with the environment. Selecting for both traits simultaneously has been difficult in practical breeding programs, and little progress has been achieved in the last 30 years for developing new wheat lines with increased GPC and high or acceptable grain yield. GPC improvement has been hindered by the negative correlation with grain yield and yield-related traits in segregating populations and germplasm collections in all cereals grown under the same conditions of N availability (Simmonds 1995; Oury et al. 2003). Because of this inverse relationship, high yielding modern cultivars have generally lower GPC compared to older ones (Simmonds 1995). During the past decades, many QTL controlling GPC have been located on all wheat chromosomes in a number of studies conducted by linkage mapping based on biparental populations (recent reviews by Blanco et al. 2012; Giancaspro et al. 2016; Quraishi et al. 2017; Kumar et al. 2018) and by association mapping based on germplasm collections (Rapp et al. 2018). A major QTL on the short arm of chromosome 6B, accounting for up to 66% of the phenotypic variation, has been reported from the wild tetraploid wheat *T. turgidum* spp. *dicoccoides*, and the functional gene *Gpc-B1* has been cloned by Uauy et al. (2006). Uauy et al. (2006) found, indeed, that the ancestral wild wheat allele of this gene/QTL encodes an NAC transcription factor (NAM-B1) that accelerates senescence and increases nutrient remobilization from leaves to developing grains. The introgression of *Gpc-B1* has shown different effects on grain weight and yield in different genetic backgrounds, and a dedicated breeding effort was suggested to offset its potential negative effects (Brevis and Dubcovsky 2010). Most GPC QTL were detected in single mapping populations and/or in specific environmental conditions, and without considering the relationship with yield-related traits. However, some recent investigations have taken into account GPC and grain yield components simultaneously assessed on the same population

to identify GPC loci without pleiotropic effects and/or not closely linked to gene for low yield-related traits (see references in Table S1).

Monaghan et al. (2001) suggested the analysis of residuals from regression of grain protein concentration on grain yield (grain protein deviation, GPD) to identify genotypes with higher grain protein concentration than would be expected from their GY. The primary components of GY are the number of productive spikes per unit area, the average number of kernel per spike, and the average kernel weight. The product of kernel number per spike and kernel weight is grain yield per spike (GYS). These yield components are quantitatively inherited, and significant correlations and coincident QTL were most frequently observed with GY, while the number of spikes per unit area was the least coincident yield component (e.g., Börner et al. 2002; Habash et al. 2007; Cuthbert et al. 2008). Kernel number per spike, kernel weight, and GYS show higher heritability than GY, and are used as selection parameters in the early breeding generations. As QTL analysis and genome-wide association study (GWAS) involve phenotypic evaluations of hundreds of lines included in segregating populations and germplasm collections, these yield components are often used as proxies for GY. Despite these studies described the calculation and experimental design to assess GPD, a few information is still available about its genetic and physiological bases. Thus, the identification of genetic sources of elevated protein content without negative pleiotropic effects would be useful for improving GPC and GY simultaneously.

So far, the objectives of this study were to: (i) identify candidate genes (CGs) for GPC and provide their precise map position on high-density single-nucleotide polymorphism (SNP)-based consensus maps; (ii) identify stable QTL/genes for GPC and GPD by the CGs and GWAS approaches using a tetraploid wheat collection coupled with the 90 K iSelect SNP genotyping array (Wang et al. 2014); (iii) identify molecular markers associated with stable high GPC QTL without decreasing grain yield. The identification of loci controlling GPC will provide information on the genetic relationships between grain storage proteins and yield-related traits and on the genetic resources available to breeders to improve the nutritional and technological properties of wheat products, as well as the opportunity to develop closely associated markers to be used in marker-assisted wheat breeding.

Materials and methods

Plant materials and field experiments

A total of 240 accessions of tetraploid wheat (*Triticum turgidum* L., $2n = 4x = 28$; AABB genome) were grown

in southern Italy at the experimental fields at Valenzano (Bari) for 5 years (2009, 2010, 2012, 2013, and 2014, hereafter, reported as V09, V10, V12, V13, and V14), at Gaudio (Potenza) for 1 year (2013, hereafter reported as G13), and at Foggia for 1 year (2012, hereafter reported as F12). The collection included the accessions of seven *T. turgidum* subspecies: *durum* (130 accessions), *durum* var. *ethiopicum* (10), *turanicum* (20), *polonicum* (19), *turgidum* (16), *carthlicum* (14), *dicoccum* (19), and *dicocoides* (12). The genotypes (number/name, pedigree, year of release, and country) were described in more detail in Mangini et al. (2018). Details about population structure and genetic diversity of the wheat collection, as investigated with SSR and DArT markers, are provided by Laidò et al. (2014), and with SNP markers by Marcotuli et al. (2015). The collection has been used for the association mapping of loci controlling grain yield components (Mangini et al. 2018) and nutritional compounds (Marcotuli et al. 2015, 2018; Colasuonno et al. 2017; Nigro et al. 2017a).

A randomized complete block design with three replications and plots consisting of 1 m rows, 30 cm apart, with 80 germinating seeds per plot, was used in the seven field experiments. During the growing season, 10 g of nitrogen per m² and standard cultivation practices were adopted. Plots were hand harvested at maturity and GYS was determined dividing grain yield per row by the number of spikes per row (about 60–80 spikes). Grain protein content, expressed as a percentage of protein on a dry weight basis, was determined on a 2 g sample of whole-meal flour using near-infrared reflectance spectroscopy (Spectra Alyzer Premium, Zeutec Büchi, Rendsburg, Germany).

DNA extraction and SNP genotyping

Genomic DNA was isolated from fresh leaves using a method previously described by Sharp et al. (1988) and subsequently purified via phenol–chloroform extraction. DNA concentration, quality, and purity were checked by both agarose gel-electrophoresis and NanoDrop2000 (Thermo Scientific™). Genomic DNA of each accession was diluted to 50 ng/μL and then sent to TraitGenetics Laboratory, Gatersleben, Germany (<http://www.traitgenetics.de>) for the genotyping procedure. SNP genotyping was performed with the wheat 90 K Infinium iSelect array containing 81,587 gene-associated SNP markers (Wang et al. 2014) and following the manufacturer's recommendations as described in Akhunov et al. (2009). The genotyping assays used an Illumina iScan reader and analyzed using Genome Studio software version 2011.1.

SNP discovery in candidate genes for grain protein content

Genomic sequences of putative GPC related genes were retrieved from EnsemblePlant database (<http://plants.ensembl.org/>), and then blasted against the available data set of SNP marker sequences reported by Akhunov et al. (2009). Markers aligned with at least 80% identity were considered as markers within the coding sequences of the GPC candidate genes. Furthermore, the BLASTn analysis was extended to contigs assembled in the chromosome survey-sequencing project (<http://wheat-urgi.versailles.inra.fr/Seq-Repository>) to identify additional SNPs flanking the GPC candidate genes. SNPs markers within the coding sequences of the GPC candidate genes were investigated by screening the variants table section of genetic variation reported for each considered candidate gene in EnsemblePlant database (<http://plants.ensembl.org/>).

Statistical analysis and association mapping

Each year–location combination was considered as an environment. Analysis of variance of GPC for each environment and for the combined analysis of environments was carried out using the MSTAT-C software. Phenotypic correlations were calculated for GPC between the environments and between GPC and GYS. The estimation of GPD was calculated by covariance analysis of GPC and GYS. Variance component estimates were used to determine genetic variance (σ_G^2), environmental variance (σ_E^2), and broad-sense heritability ($h_B^2 = \sigma_G^2 / (\sigma_G^2 + \sigma_E^2 + \sigma_{G \times E}^2)$).

Mean values across replicates and mean values across replicates and environments of GPC were used in the GWAS for each environment and over environments, respectively. SNP markers with > 10% missing data points and with a minimum allele frequency (MAF) of less than 5%, and unmapped markers on the durum wheat consensus map (Maccaferri et al. 2014), were removed from the data matrix prior to marker–trait association analysis carried out by TASSEL v5.0 (<http://www.maizegenetics.net>). Associations between SNP markers and GPC were previously calculated using the most widely used statistical models to evaluate the effects of population structure (Q) and kinship (K): (i) the general linear model (GLM); (ii) the GLM including population structure as a covariate using the Q matrix derived from the principal component analysis as implemented in TASSEL (GLM + Q); (iii) the mixed linear model (MLM) based on the kinship matrix (MLM + K); (iv) the mixed linear model based on both Q matrix and K matrix (MLM + K + Q). The model MLM + K showed less deviations of the expected values from the observed values in the quantile–quantile (Q–Q) plots and was definitely used. For GWAS results, a threshold P value of 0.001 ($-\log_{10}(P) \geq 3.0$) was used to declare

significant marker–trait associations, while suggestive QTL were considered at the sub-threshold $2.5 < -\log_{10}(P) < 3.0$ when supported by co-migrating CGs. In agreement to the linkage disequilibrium estimates determined by Laidò et al. (2014), the value of 8 cM was used as the support interval to declare significant SNPs associated with a QTL. QTL were considered stables when detected at $-\log_{10}(P) \geq 3.0$ in at least three environments and in the mean across environments.

For the associations between CGs and GPC, the conservative Bonferroni correction for multiple testing was calculated by dividing $P < 0.01$ with the number of markers used in the analysis. Chromosome localization and map position of SNP markers were derived from the high-density linkage maps described by Maccaferri et al. (2014) for durum wheat and by Wang et al. (2014) for common wheat, used as reference maps.

Results

Phenotypic variation for grain protein content

The 240 accessions of the tetraploid wheat collection were evaluated for GPC in southern Italy in replicated field trials in seven environments. The analysis of variance revealed highly significant differences ($P < 0.01$) among genotypes in each of the seven environments (not shown); the combined analysis across environments, including 218 accessions evaluated in all seven environments, showed significant effects of environments, genotypes, and environment \times genotype interaction (Table 1). Mean, range, standard deviation, genetic variance, and heritability estimates (h_B^2) for GPC for each trial are reported in Table 2. The phenotypic variations of GPC greatly varied among the seven environments: from 11.5% at G13 to 16.2% at V09; the GPC mean value across environments was 14.5% with values ranging from

Table 1 Mean square from the combined analysis of variance for grain protein content in a tetraploid wheat collection evaluated in seven field environments

| Source of variation | df | Mean square | Expected mean square |
|-------------------------|------|-------------|--|
| Environment (L) | 6 | 1540.557*** | |
| Replication/environment | 14 | 18.154*** | |
| Genotype (G) | 218 | 41.349*** | $\sigma_c^2 + r\sigma_{gl}^2 + rL(\sigma_g^2)$ |
| $G \times L$ | 1308 | 2.703*** | $\sigma_c^2 + r\sigma_{gl}^2$ |
| Error | 3052 | 0.453 | σ_c^2 |

The error, genotypic, and genotype \times environments variances are indicated as σ_c^2 , σ_g^2 , and σ_{gl}^2 , respectively. r and L are used for replication and location

***Significant at the $P \leq 0.001$ level

Table 2 Means, ranges, standard deviation (SD), coefficient of variation (CV), genetic variance (σ_G^2), and heritability (h_B^2) of grain protein content in a tetraploid wheat collection evaluated in seven environments

| Trait | Environments | | | | | | | Across environments |
|-----------------|--------------|-----------|-----------|-----------|-----------|----------|-----------|---------------------|
| | V09 | V10 | F12 | V12 | V13 | G13 | V14 | |
| Mean collection | 16.2 | 15.4 | 15.3 | 14.6 | 14.6 | 11.5 | 13.9 | 14.5 |
| Range | 11.8–22.9 | 12.5–21.7 | 12.3–20.4 | 11.0–19.2 | 11.7–20.8 | 9.4–16.1 | 11.7–18.3 | 12.1–19.4 |
| SD | 2.01 | 2.14 | 2.09 | 2.76 | 2.65 | 1.27 | 1.22 | 1.40 |
| CV (%) | 5.04 | 4.31 | 3.84 | 5.80 | 4.32 | 5.81 | 4.27 | 4.64 |
| σ_G^2 | 3.798 | 4.443 | 1.979 | 2.520 | 2.513 | 1.430 | 1.370 | 1.840 |
| h_B^2 | 0.85 | 0.91 | 0.85 | 0.78 | 0.86 | 0.76 | 0.78 | 0.60 |

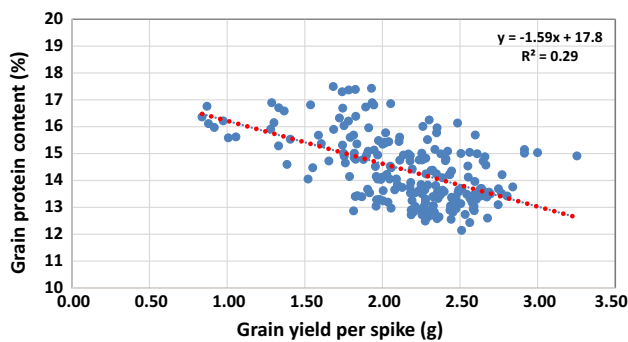


Fig. 1 Relationship between grain yield per spike and grain protein content in the tetraploid wheat collection observed in the mean across seven environments

12.1 to 19.4%. A normal frequency distribution of GPC was observed in each environment and in the mean values over the seven environments. Estimates of broad-sense heritability showed moderately high values in individual environments (from 0.76 at G13 to 0.91 at V10) and a relatively low value (0.60) over environments.

Correlations of GPC means between environments (Table S2) were all significant and ranged from 0.38 to 0.83, consistent with the strong environmental influence on phenotypic expression of GPC. Phenotypic correlation analysis between GPC and GYS, scored in the same collection and environments (Mangini et al. 2018) showed highly significant negative correlations at $P < 0.001$ in each environment (r value ranging from -0.23 at G13 to -0.62 at V12) and across environments ($r = -0.55$) (Table S3 and Fig. 1). The consistent negative correlations indicated the need to consider the relationship between GPC and GYS to identify favorable QTL/genes affecting only GPC.

Association of candidate genes with grain protein content

A total of 14 genes published as candidate genes for GPC and/or yield-related traits (Habash et al. 2007; Nigro et al.

2013, 2014; Gadaleta et al. 2011, 2014; Krapp 2015; Tian et al. 2015; Gao et al. 2016) were considered in the current study for the gene-based association mapping: *AlaAT*, *Fd-GOGAT*, *NADH-GOGAT*, *ASN1*, *GDH*, *NR*, *NIR*, *NRT2*, *NLP7*, *PPDK*, and the four glutamine synthetase genes (*GSI*, *GS2*, *GSe*, and *GSr*).

The BLASTn analysis between the CG sequences and the whole set of wheat SNPs of the 90 K iSelect allowed to identify 81 SNPs with at least 80% similarity. No SNP was identified within the *NLP7* sequence. Forty-three SNP markers were located on the consensus durum (Maccaferri et al. 2014) and/or bread wheat maps (Wang et al. 2014), respectively, while the remaining were unmapped and not further considered (Table 3). This enabled us to assign the CGs to specific wheat chromosome arm/groups: *PPDK* was mapped on 1BL, *AlaAT* on homoeologous chromosome group 4; *Fd-GOGAT* on homoeologous chromosome arms 2S; *NADH-GOGAT* on 3L; *GDH* on 5BL; *ASN1* on group 5; *NIR* was mapped on 6BL chromosome; *NR* on homoeologous chromosome arms 6L and 6BS, and *NRT2* on 6S. Different isoforms of GS family were mapped as follow: *GS2* on 2L, *GSI* on 6L, and *GSe* and *GSr* genes on homoeologous chromosome group 4 (Fig. 2). As indicated in Table 3, some markers were mapped in the common wheat consensus map but not in the durum one; in such cases, SNP markers co-migrating or mapping in the same contig of the CG sequence were used for the marker–trait association analysis. The minor allele frequency (MAF) ranged from 4.3% for *PPDK* to 45.5% for *Fd-GOGAT*.

Using the Bonferroni threshold to control for multiple testing, nine CGs were found significantly associated with GPC in the mean across environments and in at least three environments (*AlaAT*, *NADH-GOGAT*, *ASN1*, *NIR*, *NR*, *NRT2*, *GSI*, *GS2*, and *GSr*) (Table 4). While *Fd-GOGAT*-2B was significant only at one environment, *GDH*, *NADH-GOGAT*-3B, and *GSe1* were not significant in any environment or in the mean across environments. Notably, *AlaAT*-4A, *ASN1*-5A, *NR*-6A, and *GS2*-2B were significant in all seven environments, explaining up to 14.9, 14.6, 24.3, and 14.9% of phenotypic variance, respectively.

Table 3 Chromosome localization of candidate genes for grain protein content on the durum (Maccaferri et al. 2014) and bread wheat (Wang et al. 2014) consensus maps, and minor allele frequency (MAF) in the tetraploid wheat collection

| Gene | Enzyme | SNP name | SNP id | Wheat map position | | | SNP | MAF (%) |
|----------------------|------------------------------|-------------------------|------------------|--------------------|--------------|--------------|------|---------|
| | | | | Chrom. | Durum map cM | Bread map cM | | |
| <i>AlaAT-4A</i> | Alanine aminotransferase | Kukri_c2963_272 | IWB43751* | 4AS | 11.7 | 26.5 | C/T | 45.7 |
| <i>AlaAT-4B</i> | | Excalibur_c17206_329 | IWB22932 | 4BL | – | 71.5 | – | – |
| <i>AlaAT-4B</i> | | wsnp_Ex_c22078_31257869 | IWA2683* | 4BL | 74.2 | 71.5 | G | 0.0 |
| <i>Fd-GOGAT-2A</i> | Fd-Glutamate synthase | wsnp_CAP12_c901_472535 | IWA991 | 2AS | 92.1 | 106.3 | G/A | 45.5 |
| <i>Fd-GOGAT-2A</i> | | RAC875_rep_c70093_1079 | IWB62645 | 2AS | 92.1 | 106.3 | G/A | 45.0 |
| <i>Fd-GOGAT-2A</i> | | wsnp_Ex_c15325_23565794 | IWA2006 | 2AS | 92.1 | 106.3 | C/T | 45.1 |
| <i>Fd-GOGAT-2A</i> | | wsnp_JD_c12088_12411845 | IWA5793 | 2AS | 92.1 | 106.3 | A/G | 45.2 |
| <i>Fd-GOGAT-2B</i> | | Kukri_rep_c104810_341 | IWB49035 | 2BS | 70.1 | 88.9 | G/A | 19.4 |
| <i>Fd-GOGAT-2B</i> | | BobWhite_rep_c51373_453 | IWB4951 | 2BS | 70.1 | 88.9 | T/C | 18.3 |
| <i>Fd-GOGAT-2B</i> | | IAAV7487 | IWB35350 | 2BS | 70.1 | 88.9 | A/G | 18.3 |
| <i>Fd-GOGAT-2D</i> | | BS00082604_51 | IWB11266 | 2Dx | – | 42.4 | – | – |
| <i>NADH-GOGAT-3A</i> | NADH-Glutamate synthase | Tdurum_contig42234_2052 | IWB71028 | 3AL | 72.2 | 88.0 | T/G | 19.8 |
| <i>NADH-GOGAT-3B</i> | | Excalibur_c33550_75 | IWB25353 | 3BL | 86.6 | 68.4 | T/G | 30.8 |
| <i>NADH-GOGAT-3B</i> | | Tdurum_contig42234_4440 | IWB71034 | 3BL | 86.4 | 68.4 | G/T | 33.5 |
| <i>NADH-GOGAT-3B</i> | | Tdurum_contig42234_6222 | IWB71036 | 3BL | 86.6 | 68.4 | A/G | 33.6 |
| <i>NADH-GOGAT-3B</i> | | GENE-1966_110 | IWB32748 | 3BL | 86.4 | 67.7 | C/T | 33.7 |
| <i>NADH-GOGAT-3B</i> | | Kukri_rep_c70198_1436 | IWB49991 | 3BL | 86.4 | 67.7 | A/G | 56.6 |
| <i>ASNI-5A</i> | | Asparagine synthetase | Kukri_c12738_882 | IWB40883 | 5AL | – | 42.9 | – |
| <i>ASNI-5A</i> | Ra_c24707_827 | | IWB51581* | 5AL | 43.0 | 42.8 | T/C | 23.3 |
| <i>ASNI-5B</i> | wsnp_RFL_Con-tig4574_5416228 | | IWA8604 | 5B | 42.0 | 41.4 | A/G | 47.3 |
| <i>ASNI-5D</i> | D_GBB4FNX-02I4ZJR_91 | | IWB18298 | 5D | – | 72.3 | – | – |
| <i>GDH-5B</i> | Glutamate dehydrogenase | RAC875_c91566_56 | IWB61037 | 5BL | 143.5 | – | G/A | 11.3 |
| <i>GDH-5B</i> | | Tdurum_contig60866_667 | IWB72758* | 5BL | 144.2 | 128.0 | G/A | 37.6 |
| <i>GDH-5B</i> | | D_contig23766_203 | IWB16150* | 5DL | – | 151.1 | – | – |
| <i>NIR-6B</i> | Nitrite Reductase | CAP7_c7415_267 | IWB14244 | 6BL | 102.1 | 73.4 | C/T | 22.8 |
| <i>NR-6A</i> | Nitrate Reductase | CAP12_c2701_221 | IWB13325 | 6AL | 82.1 | – | T/C | 13.1 |
| <i>NR-6A</i> | | BobWhite_c23839_119 | IWB1754* | 6AL | 82.1 | 86.4 | A/G | 20.6 |
| <i>NR-6B</i> | | wsnp_CAP12_c1388_706924 | IWA921 | 6BS | 13.9 | 18.2 | G/A | 11.4 |
| <i>NRT2-6A</i> | Nitrate transporter | BS00066615_51 | IWB9848 | 6AS | – | 33.9 | – | – |
| <i>NRT2-6A</i> | | RAC875_c6137_562 | IWB59593 | 6AS | 11.2 | – | G/T | 7.3 |
| <i>NRT2-6B</i> | | RAC875_rep_c119097_451 | IWB62494 | 6BS | 27.8 | 33.3 | G/T | 16.3 |

Table 3 (continued)

| Gene | Enzyme | SNP name | SNP id | Wheat map position | | | SNP | MAF (%) |
|----------------|-----------------------------|-----------------------------|----------|--------------------|--------------|--------------|------|---------|
| | | | | Chrom. | Durum map cM | Bread map cM | | |
| <i>PPDK-1B</i> | Pyruvate phosphate dikinase | wsnp_BE443930B_Ta_2_2 | IWA141 | 1BL | 54.8 | 70.9 | T/C | 4.3 |
| <i>GSI-6A</i> | Glutamine synthetase | BS00084250_51 | IWB11353 | 6AL | 67.5 | 82.4 | G/T | 3.9 |
| <i>GSI-6A</i> | | BS00084251_51 | IWB11354 | 6AL | 67.5 | 82.4 | A/C | 3.9 |
| <i>GSI-6A</i> | | GENE-4249_274 | IWB33870 | 6AL | 67.5 | – | A/C | 0.4 |
| <i>GSI-6B</i> | | CAP11_c991_160 | IWB13090 | 6BL | 90.1 | 71.8 | G/T | 39.7 |
| <i>GS2-2A</i> | | BS00057059_51 | IWB8706 | 2AL | 169.3 | 132.7 | G/A | 41.4 |
| <i>GS2-2A</i> | | BS00057060_51 | IWB8707 | 2AL | 169.3 | 132.7 | C/T | 41.1 |
| <i>GS2-2B</i> | | wsnp_Ra_c4660_8405634 | IWA7955* | 2BL | 148.0 | 115.8 | A/C | 26.0 |
| <i>GS2-2D</i> | | TA001249-1083 | IWB65367 | 2Dx | – | 76.4 | – | – |
| <i>GSr1-4A</i> | | wsnp_Ku_rep_c76865_75281903 | IWA7521* | 4AS | 42.7 | 47.0 | A/G | 24.5 |
| <i>GSr1-4B</i> | Ku_c39003_290 | IWB39342 | 4BL | 70.9 | 72.3 | T/G | 15.8 | |
| <i>GSe1-4A</i> | Tdurum_con-tig11169_401 | IWB66958* | 4AL | 68.7 | 49.0 | G/A | 48.8 | |

*SNP co-migrating or mapping in the same contig of the gene sequence

The regression analysis between the CGs and GPD detected significances in the mean across environments and in at least three environments for *AlaAT*, *NADH-GOGAT*, *NIR*, *NR*, *NRT2*, *GS2*, and *GSe1*, explaining 10.4% (*NADH-GOGAT*)–23.8% (*NIR*) of phenotypic variance in the mean over environments (Table 5). These results indicated the involvement of these CGs in the GPC without negative interaction with GYS. No statistical significances were found for the other CGs.

QTL detection by genome-wide association study (GWAS)

The wheat collection was genotyped using the 90 K iSelect array including 81,587 gene-associated SNPs (Wang et al. 2014). After excluding SNP markers with > 10% missing data points and with a minimum allele frequency (MAF) of less than 5%, and unmapped markers on the durum wheat consensus map (Maccaferri et al. 2014), 17,003 SNPs were used for the marker–trait association. The performance of the four majors statistical GWAS models (GLM, GLM + Q, MLM + K, and MLM + K + Q) were initially evaluated for controlling the relative kinship and population structure of the wheat collection, and then to minimize the marker–trait association rate of false-positives (type I error). Inspection of the deviation of the observed $-\log_{10}(P)$ values from the expected distribution in the Q–Q plots in each environment and across environments (Fig. S1) indicated that the models MLM + K and MLM + K + Q controlling relative kinship performed significantly better than GLM and the GLM + Q model controlling population structure for both

GPC and GPD. MLM + K model produced slightly better results than MLM + K + Q, and was definitively used in the GWAS analysis.

Many QTL for GPC consistent in one or two environments were detected on all 14 chromosomes; these were considered environment-specific QTL and were not reported in the current work, because we were interested to identify important stable QTL across environments, to be used in marker-assisted breeding programs. Eleven stable QTL for GPC were detected at $-\log_{10}(P) \geq 3.0$ in at least three environments and the mean across environments (Table 6). The identified QTL were located on the chromosome arms 2BS (two QTL), 3AL (two QTL), 3BL, 4AS, 5AS, 5BL, 6BL, 7AS, and 7BL. The QTL *QGpc.mgb-5B.2* on 5BL and *QGpc.mgb-3B* on 3BL were consistent in six and five environments, respectively, and in the mean across environments. The additive effect of each QTL ranged from 0.65 to 1.91 units and accounted 5.1–8.7% of phenotypic variation across environments. Two additional QTL (*QGpc.mgb-4B* and *QGpc.mgb-5B.1*) were detected at $-\log_{10}(P) \geq 3.0$ in 2–3 environments and were reported, because supported by co-migrating CGs (*GSr1-4A* and *GDH-5B*). Each QTL had additive effects in the same direction and was often significant at sub-threshold $2.5 < -\log_{10}(P) < 3.0$ in the other environments.

To take into account the negative correlation between GPC and GYS (Fig. 1 and Table S3), the marker–trait analysis was carried out on GPD data for each environment and for the mean value across environments (Table 6). Using the same criteria to detect a significant stable QTL, the GWAS analysis on GPD confirmed the QTL on chromosome arms

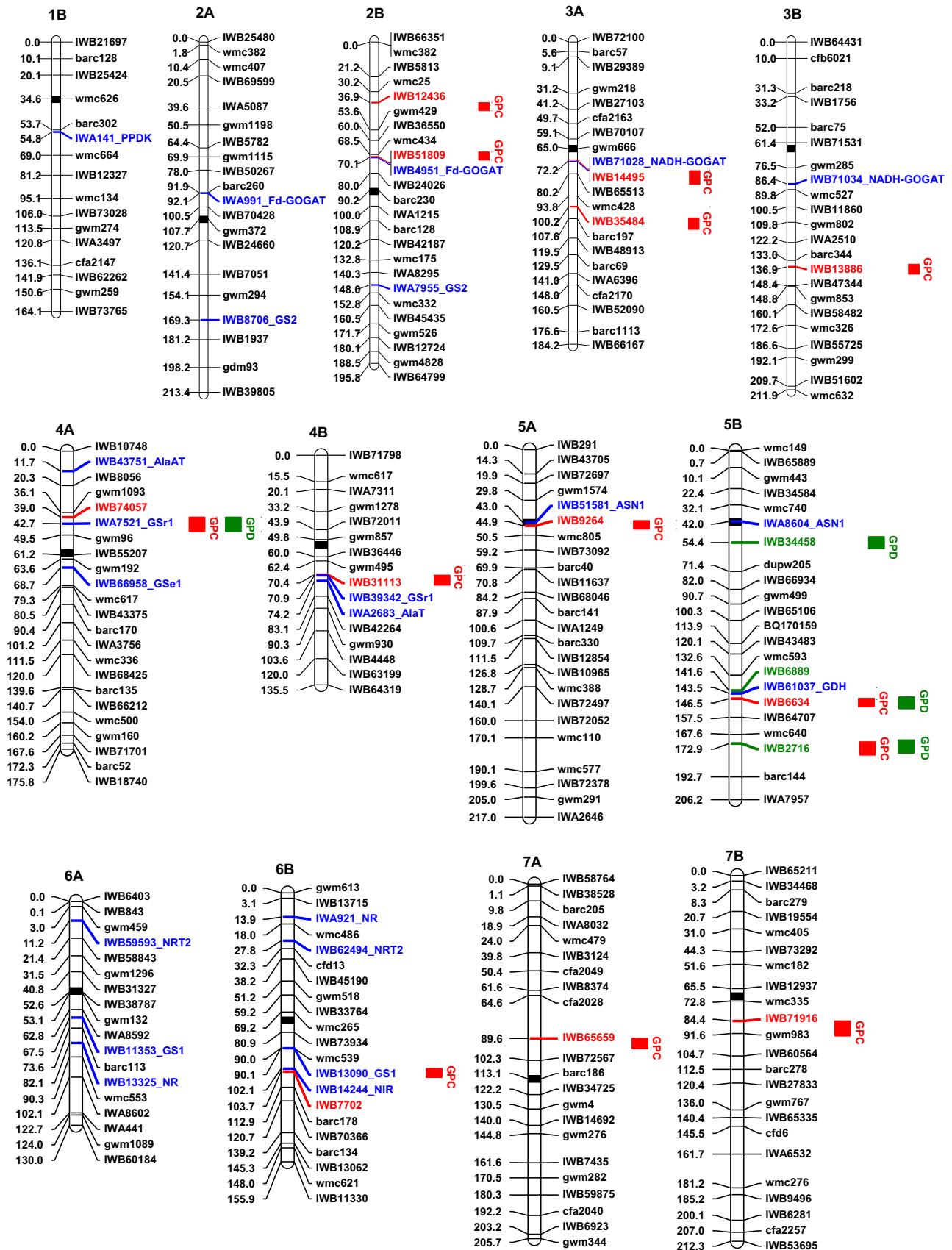


Fig. 2 Schematic representation of A and B genome chromosomes of the durum wheat consensus linkage map (Maccaferri et al. 2014) with map positions of candidate genes and QTL for grain protein content (GPC, red bars) and grain protein deviation (GPD, green bars). Each chromosome map is represented by the first and the last SNP markers, and by an SNP marker every about 20 cM. SSR markers have been also inserted every about 20 cM to compare the consensus SNP map with the published SSR-based maps. Markers are indicated on the right side and cM distances on the left side of the bar. Black solid regions of the chromosome bars indicate the centromeric region. Candidate genes for grain protein content are indicated after the corresponding SNP located in the gene sequence (in blue) or in the same map position of the co-migrating SNP marker located in the same contig

5BL at 172.9 cM (*QGpd.mgb-5B.3*) and detected one additional QTL on 5BL at 54.4 cM (*QGpd.mgb-5B.1*). Two additional QTL for GPD (*QGpd.mgb-4A* and *QGpd.mgb-5B.2*) were detected in two environments and were reported, because supported by the candidate genes *GSr1-4A* and *GDH-5B*. *QGpd.mgb-5B.2* at 141.6 cM on 5BL could coincide with *QGpc.mgb-5B.1* detected at 146.5 cM. These four GPD QTL are expressed independently from variation of GYS, while all the other 10 QTL for GPC failed to show significant stable effects indicating a genetic or physiological association between GPC and GYS.

Discussion

Nitrogen metabolism is a very complex pathway involving several processes such as N uptake, reduction, assimilation, and translocation. The dissection and deciphering of these components is still a challenge for both scientists and breeders. As typical quantitative traits, NUE and GPC are influenced both by the actions of multiple genes and environmental factors. During the last decades, GPC improvement has been hampered by the tight negative correlation with grain yield in segregating populations and germplasm collections in all cereals grown under the same conditions of N availability (Simmonds 1995; Oury et al. 2003). Several physiological explanations for this general negative relationship have been proposed taking into consideration the dilution effect of grain nitrogen by carbohydrate compounds (Acreche and Slafer 2009) or the bio-energetic requirements for synthesis of carbohydrates and proteins (Bogard et al. 2010). Despite this general assessment, a number of bread (Sears 1998; Oury et al. 2003) and durum wheat lines (Clarke et al. 2005) showing increases in both GY and GPC have been selected. According to Sears (1998), it is possible to improve both GPC and GY simultaneously when an adequate source of genes increasing GPC is used in breeding programs. The use of GPD was proposed by Monaghan et al. (2001) to screen for increased grain protein concentration without a concurrent grain yield reduction. Several studies carried

out in wheat demonstrated the high heritability of GPD and provided a new selection tool to improve GY and GPC at the same time (Bogard et al. 2010; Thorwarth et al. 2018). The balance between N remobilization and N up-taken post-flowering strongly affects both GPC (Kichey et al. 2007) and GPD (Bogard et al. 2010).

Nitrogen-related candidate genes and relationships with GPC

Studying the genetic variability existing in CGs involved in the processes controlling the relationship between C and N metabolism could be an efficient way to overcome or reduce the negative correlation between GY and GPC, and to use allele-specific markers for the selection of high GPC genotypes in wheat breeding. As proposed by Garnett et al. (2009), nitrogen metabolism-related enzymes play important roles in N utilization efficiency. So far, one objective of our study has been to characterize different N-related CGs and investigate their relationships with GPC.

By genotyping the previously described tetraploid wheat collection with SNP marker array (Wang et al. 2014) and performing a BLASTn analysis of the entire SNP data set against the CG sequences, 1–8 SNPs within the coding sequences of 13 out of 14 examined CGs were identified. These genes were precisely located on genetic map, given the recent availability of the high-resolution consensus maps of both durum (Maccaferri et al. 2014) and common wheat (Wang et al. 2014). Regression analysis was performed between each considered CG and both GPC and GPD. Interestingly, six of them (*AlaAT*, *NADH-GOGAT*, *NR*, *NRT2*, *GS2*, and *NIR*) were found to be associated with both traits.

Alanine aminotransferase is a ubiquitous enzyme that may influence N assimilation and remobilization. The transamination reaction carried out by this enzyme results in the reduction of glutamate and pyruvate in a reversible manner, providing dual functions in carbon and nitrogen metabolism. Successful transgenic approach to increase NUE has been obtained in oil seed rape (Good et al. 2007) and in rice (Shrawat et al. 2008) by genetically modifying plants for *AlaAT* gene. The same situation was also reported for *NRT2* and *NR* genes. Both resulted to be associated with GPC and GPD by regression analysis in at least three environments and in the mean of all seven environments. *NRT2* genes are high-affinity nitrate transporters and most of them require a nitrate reductase protein to mediate nitrate transport (Kotur et al. 2012). Specifically, among the several *NRT2* transporters, the most active is *NRT2.1* (Li et al. 2007). A series of genetic and physiological studies have shown that high-affinity complex *NRT2.1*–*NR* protein plays a central part in efficient N uptake under low availability (Krapp et al. 2011; Gu et al. 2013). Taulemesse et al. (2015) showed a significant positive correlation between the expression patterns of

Table 4 Regression analysis between candidate genes and grain protein content in the tetraploid wheat collection evaluated in seven environments

| Gene | SNP id | Map position | | Environments | | | | | | | P | Mean effect | R ² (%) | Co-located QTL |
|----------------------|-----------|--------------|-------|--------------|---------|---------|---------|---------|---------|--------|---------|-------------|--------------------|-----------------------|
| | | Chrom. | cM | V09 | V10 | F12 | V12 | G13 | V13 | V14 | | | | |
| <i>AlaAT-4A</i> | IWB43751* | 4AS | 11.7 | 6.4*** | 7.1*** | 8.7*** | 7.2*** | 4.7** | 4.9** | 3.7* | 8.5*** | 1.17 | 14.9 | |
| <i>Fd-GOGAT-2A</i> | IWA991 | 2AS | 92.1 | - | - | - | - | - | - | - | - | - | - | |
| <i>Fd-GOGAT-2A</i> | IWB62645 | 2AS | 92.1 | - | - | - | - | - | - | - | - | - | - | |
| <i>Fd-GOGAT-2A</i> | IWA2006 | 2AS | 92.1 | - | - | - | - | - | - | - | - | - | - | |
| <i>Fd-GOGAT-2A</i> | IWA5793 | 2AS | 92.1 | - | - | - | - | - | - | - | - | - | - | |
| <i>Fd-GOGAT-2B</i> | IWB49035 | 2BS | 70.1 | - | - | - | - | - | - | 4.3* | - | - | - | <i>QGpc.mgb-2B.2</i> |
| <i>Fd-GOGAT-2B</i> | IWB4951 | 2BS | 70.1 | - | - | - | - | - | - | 4.3* | - | - | - | |
| <i>Fd-GOGAT-2B</i> | IWB35350 | 2BS | 70.1 | - | - | - | - | - | - | 4.3* | - | - | - | |
| <i>NADH-GOGAT-3A</i> | IWB71028 | 3AL | 72.2 | 4.1* | 6.3*** | 7.3*** | 5.9*** | - | 4.8** | 4.0* | 6.3*** | 1.27 | 11.0 | <i>QGpc.mgb-3A.1</i> |
| <i>NADH-GOGAT-3B</i> | IWB25353 | 3BL | 86.6 | - | - | - | - | - | - | - | - | - | - | |
| <i>NADH-GOGAT-3B</i> | IWB71034 | 3BL | 86.4 | - | - | - | - | - | - | - | - | - | - | |
| <i>NADH-GOGAT-3B</i> | IWB71036 | 3BL | 86.6 | - | - | - | - | - | - | - | - | - | - | |
| <i>NADH-GOGAT-3B</i> | IWB32748 | 3BL | 86.4 | - | - | - | - | - | - | - | - | - | - | |
| <i>NADH-GOGAT-3B</i> | IWB49991 | 3BL | 86.4 | - | - | - | - | - | - | - | - | - | - | |
| <i>ASNI-5A</i> | IWB51581* | 5AL | 43.0 | 7.1*** | 8.7*** | 9.2*** | 4.3* | 4.9** | 4.5** | 4.2* | 8.3*** | 1.42 | 14.6 | <i>QGpc.mgb-5A</i> |
| <i>ASNI-5B</i> | IWA8604 | 5B | 42.0 | 3.3 | 7.0*** | 5.7*** | 3.2 | 6.0*** | 3.4 | 2.2 | 5.8*** | 0.97 | 10.0 | |
| <i>GDH-5B</i> | IWB61037 | 5BL | 143.5 | - | - | - | - | - | - | - | - | - | - | |
| <i>GDH-5B</i> | IWB72758* | 5BL | 144.2 | - | - | - | - | - | - | - | - | - | - | <i>QGpc.mgb-5B.1°</i> |
| <i>NIR-6B</i> | IWB14244 | 6BL | 102.1 | 12.7*** | 8.5*** | 7.3*** | 6.0*** | 9.5*** | 7.6*** | - | 8.8*** | 1.41 | 15.6 | <i>QGpc.mgb-6B</i> |
| <i>NR-6A</i> | IWB13325 | 6AL | 82.1 | 11.1*** | 16.0*** | 10.1*** | 8.5*** | 8.7*** | 12.9*** | 5.6*** | 14.1*** | 2.24 | 24.3 | |
| <i>NR-6A</i> | IWB1754* | 6AL | 82.1 | 18.0*** | 21.3*** | 14.1*** | 14.7*** | 13.9*** | 16.6*** | 6.8*** | 21.4*** | 2.27 | 36.7 | |
| <i>NR-6B</i> | IWA921 | 6BS | 13.9 | - | - | - | - | - | - | - | - | - | - | |
| <i>NRT2-6A</i> | IWB59593 | 6AS | 11.2 | 4.7** | 5.3** | 4.3* | - | - | - | - | 4.2* | 1.51 | 7.2 | |
| <i>NRT2-6B</i> | IWB62494 | 6BS | 27.1 | - | - | - | - | - | - | - | - | - | - | |
| <i>GSI-6B</i> | IWB13090 | 6BL | 90.1 | 6.0*** | 7.0*** | 4.1* | 4.3* | 5.4*** | 6.1*** | - | 6.9*** | 1.10 | 12.1 | |
| <i>GSI-2A</i> | IWB8706 | 2AL | 169.3 | - | - | 3.7* | 6.5*** | - | - | - | 4.4** | 0.84 | 7.4 | |
| <i>GSI-2A</i> | IWB8707 | 2AL | 169.3 | - | - | 3.9* | 6.5*** | - | - | - | 4.4** | 0.85 | 7.6 | |
| <i>GSI-2B</i> | IWA7955* | 2BL | 148.0 | 8.4*** | 8.7*** | 6.9*** | 4.2* | 4.2* | 8.3*** | 4.1* | 8.3*** | 1.34 | 14.9 | |
| <i>GSI-1-4A</i> | IWA7521* | 4AS | 42.7 | - | - | - | - | - | - | - | - | - | - | <i>QGpc.mgb-4A</i> |
| <i>GSI-1-4B</i> | IWB39342 | 4BL | 70.9 | - | 7.1*** | 5.6*** | - | - | 5.0** | 5.1** | 5.6*** | 1.25 | 10.1 | <i>QGpc.mgb-4B°</i> |
| <i>GSI-1-4A</i> | IWB66958* | 4AL | 68.7 | - | - | - | - | - | - | - | - | - | - | |

The $-\log_{10}(P)$ values are reported for each candidate gene in each environment (V09, V10, F12, V12, V13, G13, and V14) and in the mean across environments. Phenotypic variation (R^2) and additive effect are reported only for candidate genes significant in the mean of all seven environments. Chromosome and map position from Maccaferri et al. (2014)

*, **, and ***Significant at $P \geq 0.05$, 0.01 and 0.001, respectively, using the Bonferroni threshold to control for multiple testing; - not significant; Phenotypic variation explained by the QTL (%)

Table 5 Regression analysis between candidate genes and grain protein deviation in the tetraploid wheat collection evaluated in seven environments

| Gene | SNP id | Map position | | Environments | | | | | | | P | Mean effect | R ² (%) | Co-located QTL | | |
|----------------------|-----------|--------------|-------|--------------|---------|---------|--------|-----|-----|-----|--------|-------------|--------------------|----------------|------|------|
| | | Chrom. | cM | V09 | V10 | F12 | V12 | G13 | V13 | V14 | | | | | | |
| <i>AlaAT-4A</i> | IWB43751* | 4AS | 11.7 | 7.1*** | 3.7* | 6.8*** | – | – | – | – | – | 6.4*** | 0.69 | 11.6 | | |
| <i>Fd-GOGAT-2A</i> | IWA991 | 2AS | 92.1 | – | – | – | – | – | – | – | – | – | – | – | | |
| <i>Fd-GOGAT-2A</i> | IWB62645 | 2AS | 92.1 | – | – | – | – | – | – | – | – | – | – | – | | |
| <i>Fd-GOGAT-2A</i> | IWA2006 | 2AS | 92.1 | – | – | – | – | – | – | – | – | – | – | – | | |
| <i>Fd-GOGAT-2A</i> | IWA5793 | 2AS | 92.1 | – | – | – | – | – | – | – | – | – | – | – | | |
| <i>Fd-GOGAT-2B</i> | IWB49035 | 2BS | 70.1 | – | – | – | – | – | – | – | – | – | – | – | | |
| <i>Fd-GOGAT-2B</i> | IWB4951 | 2BS | 70.1 | – | – | – | – | – | – | – | – | – | – | – | | |
| <i>Fd-GOGAT-2B</i> | IWB35350 | 2BS | 70.1 | – | – | – | – | – | – | – | – | – | – | – | | |
| <i>NADH-GOGAT-3A</i> | IWB71028 | 3AL | 72.2 | – | 4.6** | 6.9*** | – | – | – | – | 5.4*** | 5.7*** | 0.81 | 10.4 | | |
| <i>NADH-GOGAT-3B</i> | IWB25353 | 3BL | 86.6 | – | – | – | – | – | – | – | – | – | – | – | | |
| <i>NADH-GOGAT-3B</i> | IWB71034 | 3BL | 86.4 | – | – | – | – | – | – | – | – | – | – | – | | |
| <i>NADH-GOGAT-3B</i> | IWB71036 | 3BL | 86.6 | – | – | – | – | – | – | – | – | – | – | – | | |
| <i>NADH-GOGAT-3B</i> | IWB32748 | 3BL | 86.4 | – | – | – | – | – | – | – | – | – | – | – | | |
| <i>NADH-GOGAT-3B</i> | IWB49991 | 3BL | 86.4 | – | – | – | – | – | – | – | – | – | – | – | | |
| <i>ASNI-5A</i> | IWB51581* | 5AL | 43.0 | 4.1* | 3.9* | 5.1** | – | – | – | – | – | – | – | 4.3* | | |
| <i>ASNI-5B</i> | IWA8604 | 5B | 42.0 | – | – | 3.9* | – | – | – | – | – | – | – | – | | |
| <i>GDH-5B</i> | IWB61037 | 5BL | 143.5 | – | – | – | – | – | – | – | – | – | – | – | | |
| <i>GDH-5B</i> | IWB72758* | 5BL | 144.2 | – | – | – | – | – | – | – | – | – | – | – | | |
| <i>NIR-6B</i> | IWB14244 | 6BL | 102.1 | 12.8*** | 10.0*** | 8.6*** | – | – | – | – | – | – | – | – | | |
| <i>NR-6A</i> | IWB13325 | 6AL | 82.1 | 6.5*** | 5.8*** | 6.5*** | – | – | – | – | – | – | – | – | 23.8 | |
| <i>NR-6A</i> | IWB1754* | 6AL | 82.1 | 14.1*** | 12.7*** | 11.9*** | 3.7* | – | – | – | – | – | – | – | – | 4.7 |
| <i>NR-6B</i> | IWA921 | 6BS | 13.9 | – | – | – | – | – | – | – | – | – | – | – | – | 22.9 |
| <i>NRT2-6A</i> | IWB59593 | 6AS | 11.2 | 5.6*** | – | – | – | – | – | – | – | – | – | – | – | 11.3 |
| <i>NRT2-6B</i> | IWB62494 | 6BS | 27.1 | – | – | – | – | – | – | – | – | – | – | – | – | |
| <i>GSI-6B</i> | IWB13090 | 6BL | 90.1 | – | – | – | – | – | – | – | – | – | – | – | – | |
| <i>GSI-2A</i> | IWB8706 | 2AL | 169.3 | – | 4.5** | 5.8*** | 5.2*** | – | – | – | – | – | – | – | – | 12.4 |
| <i>GSI-2A</i> | IWB8707 | 2AL | 169.3 | 3.8* | 4.5** | 5.9*** | 5.1** | – | – | – | – | – | – | – | – | 12.1 |
| <i>GSI-2B</i> | IWA7955* | 2BL | 148.0 | 3.8* | – | 3.8* | – | – | – | – | – | – | – | – | – | |
| <i>GSr1-4A</i> | IWA7521* | 4AS | 42.7 | – | – | – | – | – | – | – | – | – | – | – | – | |
| <i>GSr1-4B</i> | IWB39342 | 4BL | 70.9 | – | – | – | – | – | – | – | – | – | – | – | – | |
| <i>GSe1-4A</i> | IWB66958* | 4AL | 68.7 | – | 4.5** | 3.7* | – | – | – | – | – | – | – | – | – | |

The $-\log_{10}(P)$ values are reported for each candidate gene in each environment (V09, V10, F12, V13, G13, and V14) and in the mean across environments. Phenotypic variation (R^2) and additive effect are reported only for candidate genes significant in the mean of all seven environments. Chromosome and map position from Maccacferri et al. (2014)

*, **, and ***Significant at $P \geq 0.05$, 0.01, and 0.001, respectively, using the Bonferroni threshold to control for multiple testing; – not significant; Phenotypic variation explained by the QTL (%)

Table 6 Marker–trait associations for grain protein content (GPC) and protein content deviation (GPD) identified by GWAS (model MLM+K) in the tetraploid wheat collection

| QTL* | Closest marker | Chrom. | Arm | Position cM | Map Interval cM | SNP allele | SNP frequency | Environments | | | | | | | Mean | | Co-located candidate gene | |
|-----------------------------------|----------------|--------|-----|-------------|-----------------|------------|---------------|--------------|-----|-----|-----|-----|-----|-----|------|--------|---------------------------|----------------------|
| | | | | | | | | V09 | V10 | F12 | V12 | G13 | V13 | V14 | P | Effect | | R ² (%) |
| Grain protein content | | | | | | | | | | | | | | | | | | |
| <i>QGpc.mgb-2B.1</i> | IWB12436 | 2BS | | 36.9 | 36.9-41.9 | G/T | 21-184 | – | 3.7 | 3.3 | – | 2.9 | 4.4 | 3.5 | 3.0 | 1.39 | 5.2 | |
| <i>QGpc.mgb-2B.2</i> | IWB51809 | 2BS | | 70.1 | 69.2-70.1 | C/T | 18-199 | – | 4.0 | – | 2.6 | 3.8 | 3.1 | 2.7 | 4.7 | 1.77 | 8.7 | <i>Fd-GOGAT-2B</i> |
| <i>QGpc.mgb-3A.1</i> | IWB14495 | 3AL | | 72.2 | 69.6-78 | G/A | 19-198 | 3.0 | 5.0 | – | – | 2.5 | 3.3 | 3.5 | 3.4 | 1.37 | 6.0 | <i>NADH-GOGAT-3A</i> |
| <i>QGpc.mgb-3A.2</i> | IWB35484 | 3AL | | 100.2 | 100.2-107.3 | A/G | 11-210 | 3.3 | 3.3 | – | 4.0 | – | – | 3.0 | 3.4 | 1.51 | 5.9 | |
| <i>QGpc.mgb-3B</i> | IWB13886 | 3BL | | 136.9 | 132.8-138.9 | T/C | 41-177 | 4.5 | 3.0 | – | 4.5 | 3.1 | 3.4 | – | 4.0 | 1.91 | 7.4 | |
| <i>QGpc.mgb-4A</i> | IWB74057 | 4AS | | 39.0 | 39.0-47.5 | A/G | 92-126 | – | 3.7 | – | – | 3.6 | 2.8 | 3.1 | 3.0 | 0.65 | 5.1 | <i>GSt1-4A</i> |
| <i>QGpc.mgb-4B</i> ^o | IWB31113 | 4BL | | 70.4 | 65.5-70.9 | A/C | 38-171 | – | – | 2.5 | – | 3.4 | 3.1 | 3.5 | 2.8 | 0.78 | 4.7 | <i>GSt1-4B</i> |
| <i>QGpc.mgb-5A</i> | IWB9264 | 5AS | | 44.9 | 43.0-44.9 | C/T | 15-206 | 2.9 | 3.4 | – | 2.8 | 2.9 | 3.7 | 3.2 | 3.6 | 1.85 | 6.2 | <i>ASNI-5A</i> |
| <i>QGpc.mgb-5B.1</i> ^o | IWB6634 | 5BL | | 146.5 | 141.6-146.5 | G/A | 31-189 | 3.5 | 3.5 | – | – | 2.6 | 2.5 | – | 2.5 | 0.68 | 4.1 | <i>GDH-5B</i> |
| <i>QGpc.mgb-5B.2</i> | IWB2716 | 5BL | | 172.9 | 165.2-172.9 | A/G | 57-158 | 3.8 | 4.5 | 3.2 | 3.7 | 3.2 | 3.8 | – | 3.9 | 0.89 | 7.2 | |
| <i>QGpc.mgb-6B</i> | IWB7702 | 6BL | | 103.7 | 102.1-107 | A/G | 59-155 | – | – | 3.0 | – | 3.2 | 2.7 | 3.4 | 3.3 | 0.87 | 5.8 | <i>NIR-6B</i> |
| <i>QGpc.mgb-7A</i> | IWB65659 | 7AS | | 89.6 | 89.6-95.5 | C/T | 85-131 | – | – | – | 3.0 | 4.5 | – | 3.0 | 3.6 | 1.15 | 6.5 | |
| <i>QGpc.mgb-7B</i> | IWB71916 | 7BL | | 84.4 | 84.4-92.9 | C/T | 71-145 | 3.3 | – | – | – | 3.1 | 3.6 | – | 3.1 | 1.47 | 5.2 | |
| Grain protein deviation | | | | | | | | | | | | | | | | | | |
| <i>QGpd.mgb-4A</i> ^o | IWB74057 | 4AS | | 39.0 | 37.8-39.0 | A/G | 92-126 | 3.0 | – | – | – | 2.6 | 3.7 | – | 2.9 | 0.56 | 5.2 | <i>GSt1-4A</i> |
| <i>QGpd.mgb-5B.1</i> | IWB34458 | 5BL | | 54.4 | 53.1-54.4 | T/G | 47-170 | 3.2 | – | – | 3.1 | 3.1 | – | – | 3.0 | 0.80 | 5.4 | |
| <i>QGpd.mgb-5B.2</i> ^o | IWB6889 | 5BL | | 141.6 | 140.7-146.5 | G/A | 145-69 | 4.1 | 2.9 | – | – | – | – | – | 3.4 | 0.44 | 5.8 | <i>GDH-5B</i> |
| <i>QGpd.mgb-5B.3</i> | IWB2716 | 5BL | | 172.9 | 172.9- | A/G | 56-158 | 5.9 | 2.8 | – | 3.0 | 3.2 | 3.9 | – | 3.6 | 0.72 | 6.8 | – |

The $-\log_{10}(P)$ values are reported for each marker in each environment (V09, V10, F12, V12, G13, V13, and V14) and in the mean of the environments. Phenotypic variation (R^2) and additive effect are reported only for markers significant in the mean of all seven environments. Chromosome and map position from Maccacferri et al. (2014)

Suggestive QTL significant at $2.5 \leq -\log_{10}(P) \leq 3.0$ are reported in italic

– Not significant

*QTL significant at $-\log_{10}(P) \geq 3.0$ in at least three environments and in the mean across environments

^oQTL significant at $-\log_{10}(P) \geq 3.0$ in two or three environments co-located with candidate genes

the high-affinity NO_3^- transporter family member TaNRT2.1 and the nitrate reductase TaNR in wheat at different concentration of nitrate.

Moreover, Zhao et al. (2013) reported a transgenic approach to overexpress a tobacco *NR* gene in wheat, which resulted in an increased seed protein content, without increasing N fertilization. Our analysis showed also a significant and positive additive effect of both genes on both GPG and GPD, but not on GYS. These results support our finding in underlying the possibility of using NRT2.1–NR complex as a breeding target to improve wheat NUE, yield, and grain quality.

NADH-GOGAT gene has been long considered one of the major candidate genes for cereal NUE, as supported by a cross-genome ortho-meta QTL study carried out by Qurashi et al. (2011). Transgenic approaches have been used in both wheat and rice (Yamaya et al. 2002; Lu et al. 2011), finding that its over-expression led an increase in grain weight (80% maximum), and on the contrary, its suppression reduced yield per plant and thousand kernel weight, phenotypic indications of nitrogen starvation, supporting that *NADH-GOGAT* is, indeed, a key step in nitrogen utilization and grain filling. Association of *NADH-GOGAT* and GPC was also reported in durum wheat (Nigro et al. 2017b). A different situation was shown for *GS2* genes. We found the homoeologous plastidic *GS2-2B* and *GS2-2A* being associated with GPC and GPD, respectively. Among all N metabolism-related genes, glutamine synthetase, which catalyzes the first step in assimilating inorganic N into organic compounds, has been one of the most extensively studied and has been used to improve NUE in wheat (Thomsen et al. 2014). Plastidic *GS2* has been shown to be involved in re-assimilation of N during photorespiration (Tabuchi et al. 2007). Correlation studies from Kichey et al. (2007) provided evidence of the role of the *GS* enzyme in the control of NUE. In addition, a number of QTL related to GY and GPC co-localizing with structural genes encoding plastidic *GS2* (Gadaleta et al. 2011; Bordes et al. 2013) were identified. As reviewed by Thomsen et al. (2014), the effects of overexpressing *GS* on NUE-related physiological traits, biomass, and yield have generally been inconsistent. However, a very recent paper of Hu et al. (2018) showed that transgenic expression of a specific *TaGS2* allele increased flag leaf *GS* activity, leaf functional duration, N uptake, grain N concentration, nitrogen harvest index, harvest index, grain yield, and yield components under both low N and high N conditions, suggesting that the control of GPC- and NUE-related processes may be specific, depending both on the species examined but most of all on the genetic variability within the species, and then, on single genotype (Simons et al. 2014; Nigro et al. 2017a). Besides the evaluation of CGs associations with both GPC and GPD, regression analysis was also carried out between candidate genes and grain

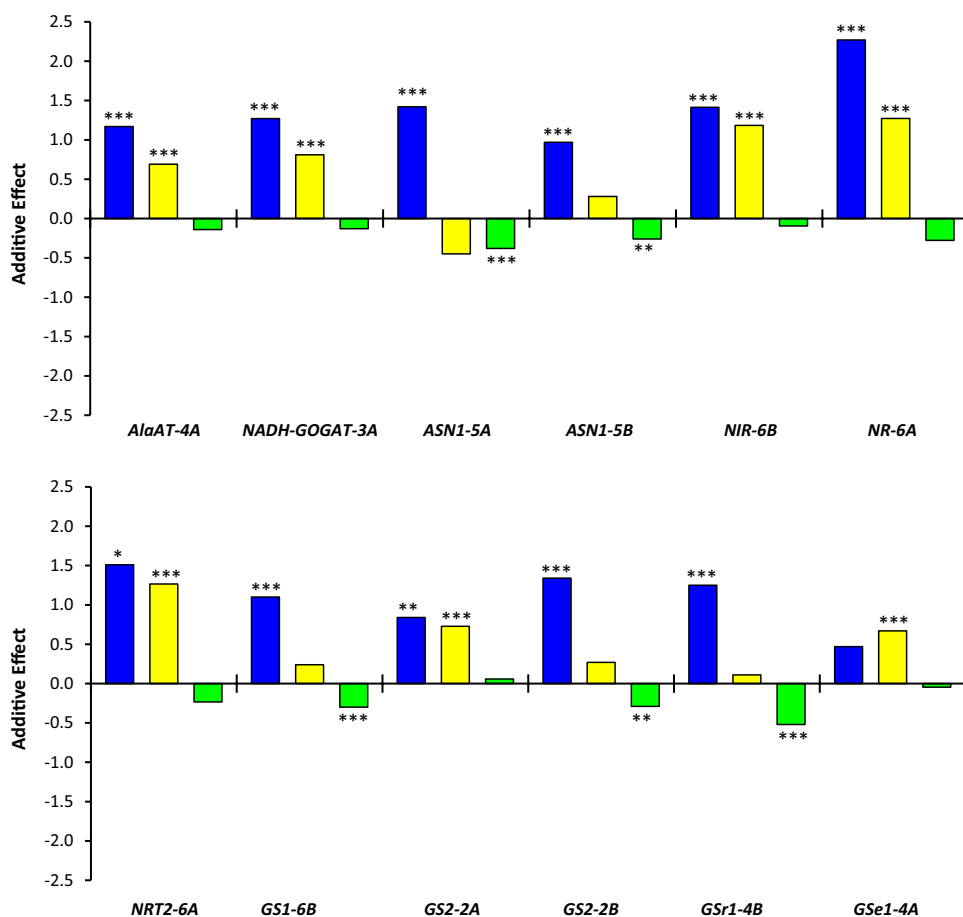
yield per spike (data from Mangini et al. 2018) in the tetraploid collection (Fig. 3). The most interesting genes were, indeed, those having a significant and positive effect both in GPD and GYS, or at least significant and positive for GPD and not significant in GYS, meaning that an increasing in protein content is not negatively correlated with yield. Interestingly, the six CGs previously discussed, being associated with both GPC and GPD, reflected this situation. Specifically, *AlaAt-4A*, *NADH-GOGAT-3A*, *NIR-6B*, *NR-6A*, *NRT2-6A*, and *GS2-2A* showed a positive and significant additive effect on GPD (and GPC), but had no significant effect on GYS. *GSe1-4A* showed a positive additive effect only on GPC but not on GPD, but anyway, had no effect on GYS. A different situation was observed for the other five genes, *ASN1* (both 5A and 5B homoeologous), *GS1-6B*, *GS2-2B*, and *GSr1-4B*. These genes, belonging to complex gene families, showed inverse significant additive effects for GPD and GYS, according to the known negative correlation existing between the two traits.

Stable QTL for grain protein content detected by GWAS

Objectives of the current study were the identification of stable QTL for GPC, and to analyze the genetic basis of the negative relationship between GPC and GYS in a tetraploid wheat collection, evaluated in seven replicated field trials in southern Italy. The different order of magnitude of the mean square of the environments in comparison to that of the genotypes (Table 1) indicated the high influence of annual environmental variations on the GPC. The GPD has been also considered for identifying QTL for high GPC not associated with low grain yield. The wide phenotypic variation for GPC observed in each of the seven environments (Table 2) can be attributed to the composition of the *T. turgidum* collection including wild and semi-domesticated accessions, landraces, and modern durum cultivars. The genome-wide association study with 17,003 SNP-derived genes, conducted by the GLM and the three widely adopted MLM models (MLM + Q, MLM + K, and MLM + Q + K), indicated the MLM + K as the most suitable model for the association mapping analysis for both GPC and GPD, thus, confirming the previous results of GWAS on grain yield components and grain quality traits carried out on the same wheat collection (Colasuonno et al. 2017; Nigro et al. 2017b; Mangini et al. 2018), and supporting several previous studies on the suitability of the MLM + K model for reducing potential spurious marker–trait associations (see review by Gupta et al. 2014).

Eleven stable GPC QTL, distributed on nine chromosomes of the A and B genomes, were detected at $-\log_{10}(P) \geq 3.0$ in at least three environments and the mean across environments (Table 6 and Fig. 2); two additional

Fig. 3 Additive effect of candidate genes for grain protein content (blue bars), grain protein deviation (yellow bars), and grain yield per spike (green bars) identified by regression analysis in the mean values across seven environments in a tetraploid wheat collection. The effect refers to the allele with a positive effect on grain protein content. *, **, and *** Significant at $P \geq 0.05$, $P \geq 0.01$, and $P \geq 0.001$, respectively, using the Bonferroni threshold to control for multiple testing



QTL, co-locating with two candidate genes, were detected in two environments.

The high-resolution consensus linkage map of durum wheat (Maccaferri et al. 2014), used in the current study as reference map for SNP markers associated with GPC QTL, includes both SNP and SSR markers, and that made possible the comparison of the detected QTL with previously reported QTL in the literature. Several QTL for GPC have been mapped on all bread and durum wheat chromosomes (see reviews by Blanco et al. 2012; Quraishi et al. 2017; Kumar et al. 2018; Rapp et al. 2018). A summary of 22 peer-reviewed papers on GPC QTL that considered yield-related traits in the same experiment is reported in Table S1. Many QTL were detected in single mapping population and may be attributed to the different genotypes of the parental lines, the high number of QTL/genes controlling GPC, epistatic gene actions, marker density of the linkage maps used in QTL analyses, and the statistical models used for detecting marker–trait associations. Many other QTL were detected in individual environments; these can be considered “adaptive” QTL as expressed under specific environmental conditions or in accordance with the level of an environmental factor (Collins et al. 2008). Most QTL for GPC identified in the current study validated QTL previously detected in

different genetic background (Table S1). Interestingly, the detected QTL on chromosome arms 3BL (*QGpc.mgb-3B*) and 5BL (*QGpc.mgb-5B.2*) were detected in 5 and 6 out of 7 environments, respectively, and in the mean across environments. *QGpc.mgb-5B.2* was also expressed as QTL for GPD (*QGpd.mgb-5B.3*), meaning that it can increase GPC without negative effect on grain yield per spike. *QGpc.mgb-5B.2* maps in the same 5B region of a GPC QTL previously described by Zanetti et al. (2001), Habash et al. (2007) and Wang et al. (2012).

The GWAS analysis on GPD detected one QTL on 4AS and three QTL on 5BS and 5BL (two QTL) chromosome arms. Three of these QTL, which were independent from grain yield per spike, were coincident with three GPC QTL (*QGpd.mgb-4A*, *QGpd.mgb-5B.2*, and *QGpd.mgb-5B.3*). The QTL *QGpd.mgb-4A* was previously identified by Groos et al. (2003) and Habash et al. (2007) by GLM and MLM analyses, and by Blanco et al. (2012) and Wang et al. (2012) by conditional QTL mapping. The QTL *QGpd.mgb-5B.1* at 54.4 cM on 5B chromosome was also detected by the other authors by (Groos et al. 2003; Habash et al. 2007; Wang et al. 2012; Mahjourimajid et al. 2016) in biparental mapping populations, and in a Central European durum wheat collection (Rapp et al. 2018).

We positioned *QGpd.mgb-5B.2* and *QGpd.mgb-5B.3* on the long arm of chromosome 5B at 31.3 cM from each other (141.6 and 172.9 cM, respectively). *QGpd.mgb-5B.2* was found to be co-localized with the candidate gene *GDH*. This region has been reported to carry at least one GPC QTL not associated with yield-related traits by Zanetti et al. (2001), Habash et al. (2007), Suprayogi et al. (2009), Wang et al. (2012), Bordes et al. (2013) and Mahjourimajid et al. (2016). However, the different linkage maps with different types of molecular markers and coverage do not enable a reliable QTL comparison, and it is hard to ascertain if the above studies detected one or two QTL on 5BL chromosome arm.

The lower number of putative QTL for GPD compared to the number of QTL for GPC can be partly attributed to the negative correlations observed between GPC and GYS in each of the seven environments and across environments (Fig. 1 and Table S3). A higher number of putative QTL for the primary traits than for the mathematically derived traits were previously found by Wang et al. (2012) in maize and by Rapp et al. (2018) in wheat. Wang et al. (2012) specifically studied the impact of mathematically deriving traits from primary traits, and concluded that derived traits had a lower detection power and a higher false discovery rate than for their primary traits.

Co-location of candidate genes with QTL

Among the detected GPC QTL, some were quite interesting as they co-localized with N metabolism-related genes. *QGpc.mgb-2B.2* localized at 70.1 cM on 2BS chromosome arm of the durum consensus map (Maccaferri et al. 2014) in the same position of *Fd-GOGAT-2B* (Table 4). Despite the regression analysis on GPC and GPD detected no significant effect of this gene on these traits (it was found to be significant only for GPC in one environment), an SNP marker co-migrating with the gene was found to be significantly associated with GPC by GWAS at $-\log_{10}(P)=4.7$, with an additive effect of 1.77 and explaining 8.7% of the phenotypic variance, the higher value reported in this study. In a recent study, Zeng et al. (2017) showed that an amino acid substitution in rice *Fd-GOGAT* genes caused by a single-nucleotide polymorphism led to an increased GPC, confirming its important role as a potential candidate in NUE improvement.

Another SNP marker, IWB9264, at 44.9 cM on 5AS chromosome arm, was found to be associated with a QTL for GPC, *QGpc.mgb-5A*. This SNP was found to be located 1.9 cM downstream to SNP IWB51581, which happened to be located in the coding region of one *asparagine synthetase* gene isoform. Asparagine is one of the main amino acids involved in N recycling and remobilization. Asparagine synthetase (*ASN*) was reported to have combined effect with GS in transportation of nitrogen-containing molecules, during N remobilization phase. Authors have reported its involvement

in nitrogen remobilization in flag leaves during grain filling in wheat (Curci et al. 2018), suggesting the role of *ASN* genes in response to N status in crops.

Interestingly, two of the QTL detected with GWAS analysis confirmed the involvement of two candidate genes in GPC control, according to the result obtained with regression analysis. Specifically, peak marker IWB14495 identifying *QGpc.mgb-3A.1* is localized on 3AL chromosome, at 72.2 cM of the consensus map, in the same position of SNP marker IWB71028, localized in the coding region of *NADH-GOGAT* gene (Table 4). *NADH-GOGAT* has been reported co-localizing with an important QTL for NUE in a META analysis study (Quraishi et al. 2011). In addition, Salse et al. (2013), proposed that this gene could be used to improve grain filling either using genetic manipulation or by selecting the best alleles, both in wheat and in other cereals. In the previous studies on durum wheat, we also found a strong correlation between *NADH-GOGAT* gene expression and grain protein content (Nigro et al. 2013, 2017a). However, the most interesting result was reported for the SNP marker IWB7702, localized 1.6 cM apart from SNP IWB14244 on chromosome arm 6BL, which was found in the coding region of *NIR-6B* gene. This locus was, indeed, found to be significantly associated with QTL for GPC both by regression analysis and GWAS. Regression analysis between candidate genes and GPD resulted in the *NIR-6B* gene being highly significantly associated with the trait, $(-\log_{10}(P)=13.1)$, with an effect of 1.18 and explaining 23.8% of the phenotypic variance, Table 5). *Nitrite reductase* gene was also found to have positive and significant additive effect on both GPC and GPD, but none on GYS, making it a suitable candidate gene to be considered for improving GPD in wheat.

As previously reported, GWAS analysis carried out on GPD identified four QTL, three of which coincident with QTL for GPC. Two of them co-located with N-related CGs. Peak marker IWB74057 identifying *QGpd.mgb-4A* (coincident with *QGpc.mgb-4A*) is localized on 4AS chromosome, at 39.0 cM of the consensus map, in the coding region of *GSr1-4A* gene. The homoeologous gene *GSr1-4B* was found to co-localize with a detected QTL for GPC, *QGpc.mgb-4B*, but not with GPD QTL. Peak marker IWB6889, on 5BL, identified *QGpd.mgb-5B.2*, 1.9 cM upstream *GDH-5B* gene. 4.9 cM upstream, another GPC QTL was identified, *QGpc.mgb-5B.1*, co-locating with the same CG.

Comparison between candidate genes and GWAS analyses for QTL detection

The CG and GWAS approaches have been widely utilized for association mapping studies in a number of higher plants including the major cereal crops, such as rice, wheat, and maize. The CG strategy is often utilized in the identification

of structural genes/QTL or regulatory genes involved in a specific metabolic pathway, while the GWAS has been generally used to identify the genomic regions controlling a quantitative trait. In the current study, we used both approaches to identify genes/QTL involved in the control of the grain protein content in wheat. The SNPs located in the sequences of the candidate genes *NADH-GOGAT*, *ASNI*, *NIR*, and *GSr1* were significantly associated with GPC by both GWAS and CGs. However, the SNP markers of the genes *AlaAT*, *NR*, *NRT2*, *GSI*, and *GS2* were found to be significantly associated with GPC QTL only by GLM, while SNP markers of *Fd-GOGAT* and *GDH* genes were consistent only in the MLM analysis. *PPDK* was excluded from the regression analysis because of its allele frequency lower than 0.05. Similar results were obtained in some recent investigations on different crop species (see review Gupta et al. 2014). In rice, Zhao et al. (2013) consistently detected the semi-dwarf gene, *SD1*, by both the GLM and MLM models, while several SNPs near other height-controlling genes, such as *OsBAK1* and *DGL1*, were only detected by the regression analysis; they suggested that the MLM model might overcompensate for population structure and relatedness, leading to false negatives. The regression analysis generally has high false-positive rates because of the confounding effect of population structure and relatedness among individuals (Brescghello and Sorrells 2006). On the contrary, GWAS could effectively eliminate the excess of false-positives for most traits, but it also likely eliminated true positives (Gupta et al. 2014). In some situation, GWAS seems to be ignoring most of the important findings on the genetics and physiology of the traits of interest (Donnelly 2008). To overcome the limitation of both association mapping approaches and to increase the power and precision of QTL detection, the CG approach could be used following GWAS and select the CGs corresponding to some QTL for further study (Weng et al. 2011). Alternatively, the CG approach could be simultaneously used with GWAS as in a recent investigation in maize (Lipka et al. 2013) and in the current study. The combination of molecular biological knowledge and population genetic models has been recently proposed to increase statistical power of association mapping and to include GWAS results in a biological context (Marjoram et al. 2014).

Conclusions

The identification of SNP markers within the coding sequences of QTL/genes associated with GPC can contribute to elucidating the mechanism of synthesis and accumulation of storage proteins in the wheat kernels, to understand the relationship between GPC and GYS, and to exploit the genetic variability of wheat collection for the technological and nutritional improvement of end-finished

products. The CG approach carried out in parallel with GWAS can overcome some limitations of both association mapping methods, thus reducing type I (false-positive) and type II (false-negative) error rates, and, therefore, increasing the power and precision of QTL detection. The detection of useful QTL/genes to be transferred in commercial cultivars by marker-assisted and/or genomic selection should be carried out taking into account the general negative relationship between GPC and important yield-related traits to identify favorable alleles for the improvement of GPC without compromising grain yield. Moreover, marker–trait association analysis should be carried out on phenotypic data measured in more environments for detecting stable QTL that can play a relevant role in breeding programs. In the current study, we detected 11 stable GPC QTL consistent in at least three environments, and in the mean across environments; these QTL had always additive effects in the same direction and were consistent at suggestive sub-threshold $2.5 < -\log_{10}(P) < 3.0$ in the other environments. Interestingly, seven QTL were co-migrating with N-related candidate genes. Four QTL were found to be significantly associated to increases of grain protein deviation, indicating that selecting for GPC could not affect final grain yield per spike. These conclusions suggest that the identified SNP markers could be efficiently used in wheat breeding programs for developing new genotypes possessing a better nutritional quality and improved technological value of wheat grain.

Author contribution statement DN, AG, and AB conceived and designed research, analyzed data and wrote, and reviewed and edited the manuscript. RS reviewed and edited the manuscript. GM worked on investigation, data curation, and visualization. PC, IM, AG, and SLG contributed on resources and data. All authors approved the manuscript.

Acknowledgements The research project was supported by grants from Ministero dell’Istruzione, dell’Università e della Ricerca, project ‘ISCOCEM’, and from Regione Puglia “Programma regionale a sostegno della specializzazione intelligente e della sostenibilità sociale ed ambientale-FutureInResearch”.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

- Acreche MM, Slafer GA (2009) Grain weight, radiation interception and use efficiency as affected by sink-strength in Mediterranean wheats released from 1940 to 2005. *Field Crops Res* 110(2):98–105

- Akhunov E, Nicolet C, Dvorak J (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theor Appl Genet* 119(3):507–517
- Balyan HS, Gahlaut V, Kumar A, Jaiswal V, Dhariwal R, Tyagi S, Agarwal P, Kumari S, Gupta PK (2016) Nitrogen and phosphorus use efficiencies in wheat: physiology, phenotyping, genetics, and breeding. In: Janick J (ed) *Plant breeding reviews*, vol 40. Wiley, Hoboken, pp 167–234
- Blanco A, Mangini G, Giancaspro A, Giove S, Colasuonno P, Simeone R, Gadaleta A (2012) Relationships between grain protein content and grain yield components through quantitative trait locus analyses in a recombinant inbred line population derived from two elite durum wheat cultivars. *Mol Breed* 30(1):79–92
- Bogard M, Allard V, Brancourt-Hulmel M, Heumez E, Machet JM, Jeuffroy MH et al (2010) Deviation from the grain protein concentration-grain yield negative relationship is highly correlated to post-anthesis N uptake in winter wheat. *J Exp Bot* 61:4303–4312
- Bordes J, Ravel C, Jaubertie JP, Duperrier B, Gardet O, Heumez E, Balfourier F (2013) Genomic regions associated with the nitrogen limitation response revealed in a global wheat core collection. *Theor Appl Genet* 126(3):805–822
- Börner A, Schumann E, Fürste A, Cöster H, Leithold B, Röder M, Weber W (2002) Mapping of quantitative trait loci determining agronomic important characters in hexaploid wheat (*Triticum aestivum* L.). *Theor Appl Genet* 105(6–7):921–936
- Breseghele F, Sorrells ME (2006) Association analysis as a strategy for improvement of quantitative traits in plants. *Crop Sci* 46(3):1323–1330
- Brevis JC, Dubcovsky J (2010) Effects of the chromosome region including the grain protein content locus *Gpc-B1* on wheat grain and protein yield. *Crop Sci* 50:93–104
- Castaings L, Camargo A, Pocholle D, Gaudon V, Texier Y et al (2009) The nodule inception-like protein 7 modulates nitrate sensing and metabolism in Arabidopsis. *Plant J* 57:426–435
- Clarke JM, McCaig TN, DePauw RM, Knox RE, Clarke FR, Fernandez MR, Ames NP (2005) Strongfield durum wheat. *Can J Plant Sci* 85:651–654
- Colasuonno P, Lozito ML, Marcotuli I, Nigro D, Giancaspro A, Mangini G et al (2017) The carotenoid biosynthetic and catabolic genes in wheat and their association with yellow pigments. *BMC Genomics* 18:122. <https://doi.org/10.1186/s12864-016-3395-6>
- Collins NC, Tardieu F, Tuberosa R (2008) Quantitative trait loci and crop performance under abiotic stress: where do we stand? *Plant Physiol* 147(2):469–486
- Curci PL, Bergès H, Marande W, Maccaferri M, Tuberosa R, Sonante G (2018) Asparagine synthetase genes (*AsnS1* and *AsnS2*) in durum wheat: structural analysis and expression under nitrogen stress. *Euphytica* 214(2):36. <https://doi.org/10.1007/s10681-017-2105-z>
- Cuthbert JL, Somers DJ, Brûlé-Babel AL, Brown PD, Crow GH (2008) Molecular mapping of quantitative trait loci for yield and yield components in spring wheat (*Triticum aestivum* L.). *Theor Appl Genet* 117:595–608
- Donnelly P (2008) Progress and challenges in genome-wide association studies in humans. *Nature* 456:728–731
- Gadaleta A, Nigro D, Giancaspro A, Blanco A (2011) The glutamine synthetase (*GS2*) genes in relation to grain protein content of durum wheat. *Funct Integr Genomics* 11:665–670
- Gadaleta A, Nigro D, Marcotuli I, Giancaspro A, Giove SL, Blanco A (2014) Isolation and characterization of cytosolic glutamine synthetase (*GSe*) genes and association with grain protein content in durum wheat. *Crop Pasture Sci* 65:38–45
- Gao R, Curtis TY, Powers SJ, Xu H, Huang J, Halford NG (2016) Food safety: structure and expression of the asparagine synthetase gene family of wheat. *J Cereal Sci* 68:122–131
- Garnett T, Conn V, Kaiser BN (2009) Root based approaches to improving nitrogen use efficiency in plants. *Plant Cell Environ* 32:1272–1283
- Giancaspro A, Giove SL, Zito D, Blanco A, Gadaleta A (2016) Mapping QTLs for *Fusarium* head blight resistance in an interspecific wheat population. *Front Plant Sci* 7:1381. <https://doi.org/10.3389/fpls.2016.01381>
- Good AG, Johnson SJ, De Pauw M, Carroll RT, Savidov N, Vidmar J, Stroehrer V (2007) Engineering nitrogen use efficiency with alanine aminotransferase. *Botany* 85(3):252–262
- Groos C, Robert N, Bervas E, Charmet G (2003) Genetic analysis of grain protein content, grain yield and thousand-kernel weight in bread wheat. *Theor Appl Genet* 106:1032–1040
- Gu R, Duan F, An X, Zhang F, von Wiren N, Yuan L (2013) Characterization of AMT-mediated high-affinity ammonium uptake in roots of maize (*Zea mays* L.). *Plant Cell Physiol* 54:1515–1524
- Gupta PK, Kulwal PL, Jaiswal V (2014) Association mapping in crop plants: opportunities and challenges. *Adv Genet* 85:109–147
- Habash DZ, Bernard S, Schondelmaier J, Weyen J, Quarrie SA (2007) The genetics of nitrogen use in hexaploid wheat: N utilisation, development and yield. *Theor Appl Genet* 114:403–419
- Hu M, Zhao X, Liu Q, Hong X, Zhang W, Zhang Y, Tong Y (2018) Transgenic expression of plastidic glutamine synthetase increases nitrogen uptake and yield in wheat. *Plant Biotechnol J* 16(11):1858–1867
- Kichey T, Hirel B, Heumez E, Dubois F, Le Gouis J (2007) In winter wheat (*Triticum aestivum* L.), post-anthesis nitrogen uptake and remobilisation to the grain correlates with agronomic traits and nitrogen physiological markers. *Field Crops Res* 102:22–32
- Kotur Z, Mackenzie N, Ramesh S, Tyerman SD, Kaiser BN, Glass AD (2012) Nitrate transport capacity of the *Arabidopsis thaliana* NRT2 family members and their interactions with AtNAR2.1. *New Phytol* 194(3):724–731
- Krapp A (2015) Plant nitrogen assimilation and its regulation: a complex puzzle with missing pieces. *Curr Opin Plant Biol* 25:115–122
- Krapp A, Berthomé R, Orsel M, Mercey-Boutet S, Yu A, Castaings L, Daniel-Vedele F (2011) Arabidopsis roots and shoots show distinct temporal adaptation pattern towards N starvation. *Plant Physiol* 157(3):1255–1282
- Kumar A, Jain S, Elias EM, Ibrahim M, Sharma LK (2018) An overview of QTL identification and marker-assisted selection for grain protein content in wheat. In: Sengar RS, Singh A (eds) *Eco-friendly agro-biological techniques for enhancing crop productivity*. Springer, Singapore, pp 245–274
- Laidò G, Marone D, Russo MA, Colechia SA, Mastrangelo AM, De Vita P et al (2014) Linkage disequilibrium and genome-wide association mapping in tetraploid wheat (*Triticum turgidum* L.). *PLoS One* 9(4):e95211. <https://doi.org/10.1371/journal.pone.0095211>
- Lea PJ, Azevedo RA (2007) Nitrogen use efficiency. 2. Amino acid metabolism. *Ann Appl Biol* 15(3):269–275
- Li W, Wang Y, Okamoto M, Crawford NM, Siddiqi MY, Glass AD (2007) Dissection of the *AtNRT2. 1: AtNRT2. 2* inducible high-affinity nitrate transporter gene cluster. *Plant Physiol* 143(1):425–433
- Lipka AE, Gore MA, Magallanes-Lundback M, Mesberg A, Lin H, Tiede T, DellaPenna D (2013) Genome-wide association study and pathway level analysis of tocopherol levels in maize grain. *G3-Genes Genom Genet* 3(8):1287–1299
- Lu Y, Luo F, Yang M, Li X, Lian X (2011) Suppression of glutamate synthase genes significantly affects carbon and nitrogen metabolism in rice (*Oryza sativa* L.). *Sci China Life Sci* 54:651–663
- Maccaferri M, Cane MA, Sanguineti MC, Salvi S, Colalongo MC, Massi A, Fahima T (2014) A consensus framework map of durum wheat (*Triticum durum* Desf.) suitable for linkage disequilibrium analysis and genome-wide association mapping. *BMC Genomics* 15(1):873. <https://doi.org/10.1186/1471-2164-15-873>

- Mahjourimajd S, Taylor J, Rengel Z, Khabaz-Saberi H, Kuchel H, Okamoto M, Langridge P (2016) The genetic control of grain protein content under variable nitrogen supply in an Australian wheat mapping population. *PLoS One* 11(7):e0159371. <https://doi.org/10.1371/journal.pone.0159371>
- Mangini G, Gadaleta A, Colasuonno P, Marcotuli I, Signorile AM, Simeone R et al (2018) Genetic dissection of the relationships between grain yield components by genome-wide association mapping in a collection of tetraploid wheats. *PLoS One* 13(1):e0190162. <https://doi.org/10.1371/journal.pone.0190162>
- Marcotuli I, Houston K, Waugh R, Fincher GB, Burton RA, Blanco A, Gadaleta A (2015) Genome wide association mapping for arabinoxylan content in a collection of tetraploid wheats. *PLoS One* 10(7):e0132787. <https://doi.org/10.1371/journal.pone.0132787>
- Marcotuli I, Colasuonno P, Blanco A, Gadaleta A (2018) Expression analysis of cellulose synthase-like genes in durum wheat. *Sci Rep* 8:15675. <https://doi.org/10.1038/s41598-018-34013-6>
- Marjoram P, Zubair A, Nuzhdin SV (2014) Post-GWAS: where next? More samples, more SNPs or more biology? *Heredity* 112(1):79–88
- Monaghan JM, Snape JW, Chojecki AJS, Kettlewell PS (2001) The use of grain protein deviation for identifying wheat cultivars with high grain protein concentration and yield. *Euphytica* 122(2):309–317
- Nigro D, Gu YQ, Huo N, Marcotuli I, Blanco A, Gadaleta A, Anderson OD (2013) Structural analysis of the wheat genes encoding NADH-dependent glutamine-2-oxoglutarate amidotransferases genes and correlation with grain protein content. *PLoS One* 8:e73751. <https://doi.org/10.1371/journal.pone.0073751>
- Nigro D, Blanco A, Anderson OD, Gadaleta A (2014) Characterization of *ferredoxin-dependent glutamine-oxoglutarate amidotransferase (Fd-GOGAT)* genes and their relationship with grain protein content QTL in wheat. *PLoS One* 9:e103869. <https://doi.org/10.1371/journal.pone.0103869>
- Nigro D, Fortunato S, Giove SL, Paradiso A, Gu YQ, Blanco A, de Pinto MC, Gadaleta A (2016) *Glutamine synthetase* in durum wheat: genotypic variation and relationship with grain protein content. *Front Plant Sci* 7:971. <https://doi.org/10.3389/fpls.2016.00971>
- Nigro D, Fortunato S, Giove SL, Mangini G, Yacoubi I, Simeone R, Blanco A, Gadaleta A (2017a) Allelic variants of glutamine synthetase and glutamate synthase genes in a collection of durum wheat and association with grain protein content. *Diversity* 9(4):52. <https://doi.org/10.3390/d9040052>
- Nigro D, Laddomada B, Mita G, Blanco E, Colasuonno P, Simeone R, Gadaleta A, Pasqualone A, Blanco A (2017b) Genome-wide association mapping of phenolic acids in tetraploid wheats. *J Cereal Sci* 75:25–34
- Oury FX, Berard P, Brancourt-Hulmel M, Depatureaux C, Doussignault G, Galic N, Giraud A, Heumez E, Lecompte C, Pluchard P, Rolland B, Rousset M, Trottet M (2003) Yield and grain protein concentration in bread wheat: a review and a study of multi-annual data from a French breeding program. *J Genet Breed* 57:59–68
- Quraishi UM, Abrouk M, Murat F, Pont C, Foucrier S et al (2011) Cross-genome map based dissection of a nitrogen use efficiency ortho-metaQTL in bread wheat unravels concerted cereal genome evolution. *Plant J* 65:745–756
- Quraishi UM, Pont C, Ain QU, Flores R, Burlot L, Alaux M, Salse J (2017) Combined genomic and genetic data integration of major agronomical traits in bread wheat (*Triticum aestivum* L.). *Front Plant Sci* 8:1843. <https://doi.org/10.3389/fpls.2017.01843>
- Rapp M, Lein V, Lacoudre F, Lafferty J, Müller E, Vida G, Leiser WL (2018) Simultaneous improvement of grain yield and protein content in durum wheat by different phenotypic indices and genomic selection. *Theor Appl Genet* 131(6):1315–1329
- Salse J, Quraishi UM, Pont C, Murat F, Le Gouis J, Lafarge S (2013) Grain filling of a plant through the modulation of NADH-glutamate synthase. Patent Application No. 13/576,610
- Sears RG (1998) Improving grain protein concentration and grain yield in USA hard winter wheat. In: Fowler DB, Geddes WE, Johnston AM, Preston KR (eds) *Wheat protein production and marketing*. In: Proceedings of Wheat Symposium, Saskatoon, Printcrafters Inc., Winnipeg, Canada, pp 63–67
- Sharp PJ, Kreis M, Shewry PR, Gale MD (1988) Location of β -amylase sequences in wheat and its relatives. *Theor Appl Genet* 75:286–290
- Shrawat AK, Carroll RT, DePauw M, Taylor GJ, Good AG (2008) Genetic engineering of improved nitrogen use efficiency in rice by the tissue-specific expression of alanine aminotransferase. *Plant Biotechnol J* 6:722–732
- Simmonds NW (1995) The relation between yield and protein in cereal grain. *J Sci Food Agric* 67:309–315
- Simons M, Saha R, Guillard L, Clément G, Armengaud P, Cañas R, Maranas CD, Lea PJ, Hirel B (2014) Nitrogen use efficiency in maize (*Zea mays* L.): from omics studies to metabolic modeling. *J Exp Bot* 65:5657–5671
- Suprayogi Y, Pozniak CJ, Clarke FR, Clarke JM, Knox RE, Singh AK (2009) Identification and validation of quantitative trait loci for grain protein concentration in adapted Canadian durum wheat populations. *Theor Appl Genet* 119(3):437–448
- Tabuchi M, Abiko T, Yamaya T (2007) Assimilation of ammonium ions and reutilization of nitrogen in rice. *J Exp Bot* 58:2319–2327
- Taulemesse F, Le Gouis J, Gouache D, Gibon Y, Allard V (2015) Post-flowering nitrate uptake in wheat is controlled by N status at flowering, with a putative major role of root nitrate transporter NRT2.1. *PLoS One* 10(3):e0120291. <https://doi.org/10.1371/journal.pone.0120291>
- Taylor L, Nunes-Nesi A, Parsley K, Leiss A, Leach G, Coates S, Hibberd JM (2010) Cytosolic pyruvate, orthophosphate dikinase functions in nitrogen remobilization during leaf senescence and limits individual seed growth and nitrogen content. *Plant J* 62(4):641–652
- Thomsen HC, Eriksson D, Møller IS, Schjoerring JK (2014) Cytosolic glutamine synthetase: a target for improvement of crop nitrogen use efficiency? *Trends Plant Sci* 19:656–663
- Thorwarth P, Piepho HP, Zhao Y, Ebmeyer E, Schacht J, Schachtschneider R, Kazman E, Reif JC, Wurschum T, Friedrich C, Longin CFH (2018) Higher grain yield and higher grain protein deviation underline the potential of hybrid wheat for a sustainable agriculture. *Plant Breed* 137(3):326–337
- Tian H, Fu J, Drijber RA, Gao Y (2015) Expression patterns of five genes involved in nitrogen metabolism in two winter wheat (*Triticum aestivum* L.) genotypes with high and low nitrogen utilization efficiencies. *J Cereal Sci* 61:48–54
- Uauy C, Brevis JC, Dubcovsky J (2006) The high grain protein content gene *Gpc-B1* accelerates senescence and has pleiotropic effects on protein content in wheat. *J Exp Bot* 57:2785–2794
- Wang LIN, Cui FA, Wang J, Jun LI, Ding A, Zhao C, Wang H (2012) Conditional QTL mapping of protein content in wheat with respect to grain yield and its components. *Thai J Genet* 91(3):303–312
- Wang SW, Forrest D, Allen K, Chao A, Huang SBE, Maccaferri M et al (2014) Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotechnol J* 12(6):787–796
- Weng J, Xie C, Hao Z, Wang J, Liu C, Li M, Li X (2011) Genome-wide association study identifies candidate genes that affect plant height in Chinese elite maize (*Zea mays* L.) inbred lines. *PLoS One* 6(12):e29229. <https://doi.org/10.1371/journal.pone.0029229>
- Xu G, Fan X, Miller AJ (2012) Plant nitrogen assimilation and use efficiency. *Annu Rev Plant Biol* 63:153–182

- Yamaya T, Obara M, Nakajima H, Sasaki S, Hayakawa T et al (2002) Genetic manipulation and quantitative-trait loci mapping for nitrogen recycling in rice. *J Exp Bot* 53:917–925
- Zanetti S, Winzeler M, Feuillet C, Keller B, Messmer M (2001) Genetic analysis of bread-making quality in wheat and spelt. *Plant Breed* 120:13–19
- Zeng DD, Qin R, Li M et al (2017) The ferredoxin-dependent glutamate synthase (OsFd-GOGAT) participates in leaf senescence and the nitrogen remobilization in rice. *Mol Genet Genomics* 292(2):385–395
- Zhao XQ, Nie XL, Xiao XG (2013) Over-expression of a tobacco nitrate reductase gene in wheat (*Triticum aestivum* L.) increases seed protein content and weight without augmenting nitrogen supplying. *PloS One* 8(9):74678. <https://doi.org/10.1371/journal.pone.0074678>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.