Theor Appl Genet (2014) 127:2679–2693
DOI 10.1007/s00122-014-2407-7

ORIGINAL PAPER

A genome-wide identification of chromosomal regions determining nitrogen use efficiency components in wheat (*Triticum aestivum* L.)

Fabien Cormier · Jacques Le Gouis · Pierre Dubreuil · Stéphane Lafarge · Sébastien Praud

Received: 3 July 2014 / Accepted: 3 October 2014 / Published online: 19 October 2014

Abstract

Key message **This study identified 333 genomic regions associated to 28 traits related to nitrogen use efficiency in European winter wheat using genome-wide association in a 214-varieties panel experimented in eight environments.**

Abstract Improving nitrogen use efficiency is a key factor to sustainably ensure global production increase. However, while high-throughput screening methods remain at a developmental stage, genetic progress may be mainly driven by marker-assisted selection. The objective of this study was to identify chromosomal regions associated with nitrogen use efficiency-related traits in bread wheat (*Triticum aestivum* L.) using a genome-wide association approach. Two hundred and fourteen European elite varieties were characterised for 28 traits related to nitrogen use efficiency in eight environments in which two different nitrogen fertilisation levels were tested. The genome-wide association study was carried out using 23,603 SNP with a mixed model for taking into account parentage relationships among varieties. We identified 1,010 significantly associated SNP which defined 333 chromosomal regions associated with at least one trait and found colocalisations for 39 % of these chromosomal regions. A method based on linkage disequilibrium to define the associated region was suggested and discussed with reference to false positive rate. Through a network approach, colocalisations were analysed and highlighted the impact of genomic regions controlling nitrogen status at flowering, precocity, and nitrogen utilisation on global agronomic performance. We were able to explain 40 ± 10 % of the total genetic variation. Numerous colocalisations with previously published genomic regions were observed with such candidate genes as Ppd-D1, Rht-D1, NADH-Gogat, and GSe. We highlighted selection pressure on yield and nitrogen utilisation discussing allele frequencies in associated regions.

Communicated by Ian Mackay.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-014-2407-7) contains supplementary material, which is available to authorized users.

F. Cormier · P. Dubreuil · S. Lafarge · S. Praud (✉)
Centre de recherche de Chappes, Biogemma, Route d'Ennezat CS90126, 63720 Chappes, France
e-mail: sebastien.praud@biogemma.com

J. Le Gouis
Diversité et Ecophysiologie des Céréales, INRA, UMR 1095 Génétique, 5 chemin de Beaulieu, 63 039 Clermont-Ferrand, France

J. Le Gouis
Diversité et Ecophysiologie des Céréales, Université Blaise Pascal, UMR 1095 Génétique, 63 177 Aubière Cedex, France

Abbreviations

ADM_S	Straw dry matter at maturity
DArT	Diversity array technology
LD	Linkage disequilibrium
FLO	Flowering date
G	Genotype
G × E	Genotype × environment
GNY	Grain nitrogen yield
GPC	Grain protein content
GPD	Grain protein deviation
GY	Grain dry matter yield
HI	Harvest index
KS	Kernel per spike
N	Nitrogen
%N_S	Straw nitrogen content at maturity
NHI	Nitrogen harvest index

NSA	Straw nitrogen per area
NTA	Total nitrogen in plant at maturity
NUE	Nitrogen use efficiency
NUE_Prot	Nitrogen use to protein efficiency
NupE	Nitrogen uptake
NutE	Nitrogen utilisation efficiency
NutE_Prot	Nitrogen utilisation to protein efficiency
P	*P* value
PH	Plant height
QTL	Quantitative trait locus
QTN	Quantitative trait nucleotide
SA	Spike per area
SNP	Small nucleotide polymorphism
SSR	Single sequence repeat
TKW	Thousand kernel weight

Introduction

Global production of cereals has increased by around threefold since 1960 (FAO 2012) and is correlated with increased application of nitrogen (N) fertiliser. To date, the global growth in fertiliser demand is still positive as the demand for grain increases (FAO 2011). Thus, to sustainably enhance worldwide cereal production, it is necessary to increase production per N fertiliser unit.

Nitrogen use efficiency (NUE) is defined as grain yield divided by the available nitrogen. In bread wheat (*Triticum aestivum* L.), genetic progress on NUE-related traits has been assessed in various studies (Ortiz-Monasterio et al. 1997; Guarda et al. 2004; Muurinen et al. 2006; Cormier et al. 2013) and was mainly driven by selection on yield at a constant and high N level. This genetic progress should be at least maintained and preferably accelerated to deal with political, economic, and environmental concerns (Rothstein 2007; Pathak et al. 2011). Several promising ways to improve NUE have been proposed such as focusing on root architecture (Hirel et al. 2007; Foulkes et al. 2009; Kant et al. 2011) or on senescence and remobilisation (Gaju et al. 2011, Distelfeld et al. 2014). Although encouraging results have been obtained (Knyazikhin et al. 2013), phenotyping for NUE is still tedious as there are actually no high-throughput methods available (Manske et al. 2001; Tester and Langridge 2010). Moreover, G × N interactions have been observed on various agronomic traits (e.g. Le Gouis et al. 2000; Barraclough et al. 2010; Cormier et al. 2013) meaning that varieties may have to be tested in several N regimes. Thus, in a global context of fertiliser reduction, the ability to identify stable quantitative trait loci (QTL) controlling NUE-related traits and to implement this knowledge in breeding programs may condition a part of the future genetic gain. Various studies have already identified interesting quantitative trait loci (QTL) linked to N metabolism and response to N using biparental populations (e.g. An et al. 2006; Laperche et al. 2007; Habash et al. 2007; Guo et al. 2012; Xu et al. 2013). Originally developed in animal and human genetics, genome-wide association study (GWAS) is now used in numerous studies in crop species. Although GWAS has provided useful results in dissecting complex traits in wheat such as yield and its components (e.g. Crossa et al. 2007; Neumann et al. 2011), and yield response to nitrogen (Bordes et al. 2013), to our knowledge, this study is the first GWAS on NUE and NUE-related traits in small grain cereals.

GWAS overcomes the two main limitations suffered by biparental design of limited allelic diversity and poor mapping resolution due to limited recombination events during the creation of the population (Korte and Farlow 2013). However, the use of linkage disequilibrium (LD) to identify marker-trait association at the whole genome level has also some specific limitations. False positive association (Type I error) can easily arise from population structure. In addition, though the accumulation of recombination allows for a high-resolution mapping, it also decreases LD between causal mutation and markers, which in turn decreases the power of detection for a given number of markers. To deal with these major trade-offs, independent markers can be used to assess the relative kinship in the panel. This information is then used to control Type I error. The power issue can be solved by increasing the number of markers which is now possible with the use of wheat single nucleotide polymorphism (SNP) chips at relatively low cost (Wang et al. 2014).

In GWAS, results are mostly shown using simple Manhattan plots and there is no widespread method to well define associated chromosomal regions. Indeed, in a panel, the link between linkage disequilibrium and genetic or physical distance is much more complex than in a biparental population, where methods such as one LOD support interval or bootstrapping are commonly used to assess QTL confidence interval (e.g. Lander and Botstein 1989; Mangin et al. 1994; Visscher et al. 1996). Moreover, in strong LD regions, pairwise correlation between significant markers can approach genotyping accuracy rate. Thus, even with methods such as stepwise logistic regression to test whether a marker in a given set is necessary or sufficient to explain the association signals, finding the one likely to be closest to the causal mutation is nearly impossible (McCarthy and Hirschhorn 2008). Added to that, in high LD regions, the tested marker is correlated to many other SNPs that can contribute to the estimation of the kinship reducing the power of detection (Rincent et al. 2014). Thus, the most significant quantitative trait nucleotide (QTN) may not be the closest to the causal mutation. In low LD regions, it is possible that only one SNP is significant, and there is no simple way to define a region from the relationship of *P* value (*P*) with genetic/physical distance. In any

Table 1 Description of measured and calculated traits assessed in all environments for which adjusted means by varieties were calculated on a 214-lines wheat association panel

Trait	Description	Formula	Units	H^2_G	Mean	SD
ABSN	Post-anthesis absorption	NTA–NFA	kg ha^{-1}	0.25	22.7	26.43
ADM_FLO	Above-ground dry matter at anthesis		kg ha^{-1}	0.69	10,618	2,222.50
ADM_S	Straw dry matter at maturity		kg ha^{-1}	0.84	7,288	1,861.32
DMGY	Dry matter grain yield		kg ha^{-1}	0.89	7,400	1,257.49
EFFG	Genetic efficiency	REMN/GNY	%	0.18	82.3	19.85
EFFREMN	Remobilization efficiency	REMN/NFA		0.27	77.3	7.56
FLO	Anthesis date		Days (after 1st January)	0.99	149.2	7.12
GNY	Grain N yield	GPC/5.7 × GY	kg ha^{-1}	0.50	127.9	35.44
GPC	Grain protein concentration		%	0.92	9.93	2.05
GPD	Grain protein deviation	GPC − a × GY − b	% of protein	0.80	0	0.78
HI	Harvest index	GY/(GY + ADM_S)	%	0.88	50.4	5.67
INN_FLO	N nutrition index	%N_FLO/(5.35 × ADM_FLO/1,000)$^{(-0.442)}$		0.63	0.69	0.19
NFA	N at anthesis	ADM_FLO × %N_FLO	kg ha^{-1}	0.16	138	48.82
NHI	N harvest index	GNY/NTA	%	0.63	81.1	5.71
NSA	Straw N per area	ADM_S × %N_S	kg ha^{-1}	0.50	30.4	14.17
NTA	total N per area	NSA + GNY	kg ha^{-1}	0.41	158	45.03
NUE	N use efficiency	GY/NTA_{max}*	kg DM kg^{-1} N	0.87	37.8	7.69
NUE_Prot	N use efficiency to protein	GPC/NTA_{max}*	% protein kg^{-1} N ha^{-1}	0.90	0.05	0.01
NupEFlo	N uptake at anthesis	NFA/NFA_{max}*	%	0.15	0.76	0.12
NupEMat	N uptake efficiency at maturity	NTA/NTA_{max}*	%	0.37	0.78	0.08
NutE	N utilisation efficiency	GY/NTA	kg DM kg^{-1} N	0.87	48.8	11.19
NutE_Prot	N utilisation efficiency to protein	GPC/NTA	% protein kg^{-1} N ha^{-1}	0.89	0.07	0.01
PH	Plant height		cm	0.95	76.6	8.43
REMN	N remobilization	NFA–NSA	kg ha^{-1}	0.25	109	39.21
SA	Spikes per area		nb spike m^{-2}	0.85	412	78.83
TKW	1,000-kernel weight		g	0.96	42.4	4.11
%N_FLO	N concentration at anthesis		%	0.80	1.29	0.34
%N_S	Straw N concentration at maturity		%	0.77	0.42	0.13

* NTA_{max} and NFA_{max} are defined as the respective 95th percentile of NTA and NFA (see Cormier et al. 2013)

case, *P* value depends on the QTL effect. This biases the *P* value support method of constructing "confidence interval" (Mangin et al. 1994). Thus, authors often fix a more or less arbitrary window around QTN peaks based on mean LD decay, for example, 1 Mb in maize for Tian et al. (2011), 200 kb in rice for Zhao et al. (2011), or 5 cM in wheat for Le Gouis et al. (2012). The method chosen to define an associated chromosomal region influences GWAS reliability and this issue remains under investigated.

Using 214 European elite varieties, 28 NUE-related traits, and 23,603 SNP, this study aimed to (1) estimate the power of such an elite panel to perform GWAS with respect to the method used to define associated chromosomal regions and false positive rate, (2) identify stable chromosomal regions involved in NUE-related traits and assess their transferability to the field, and (3) analyse colocalisations for NUE components and NUE-related traits to estimate pleiotropic effects associated with QTL-based selection.

Materials and methods

Phenotypic data

Phenotypic data are described in Cormier et al. (2013). Briefly, 225 European elite varieties were evaluated in eight environments defined as a combination of year, site, and nitrogen supply (two seasons, three sites, and two nitrogen supplies). The high N treatment corresponded to common agricultural practices. The low N treatment corresponded to a mean yield reduction of 20 % (Suppl data 1). Other crop inputs including weed, disease and pest control, potassium, phosphate and sulphur fertilisers, were applied at sufficient levels to prevent them from limiting yield. Plant growth regulators were applied to limit lodging in all environments. In each environment, 28 traits were measured or calculated (Table 1). From adjusted means by trial, overall adjusted means by varieties were

computed using a simple linear model with environment and genotype as fixed effects. These values were used in the GWAS. Generalised broad-sense heritabilities (H^2_G) were calculated using the formula proposed by Cullis et al. (2006) from the previous linear model with genotype as a random effect.

Genotyping and consensus map

Of the 225 varieties present in field trials, 214 were genotyped. SNP data consisted of a subset of SNP from an Illumina 90 K chip (Wang et al. 2014) together with SNP developed by Biogemma. Heterozygous loci were considered as missing data. Loci with a minor allele frequency inferior to 0.05 or loci which had available data for less than 150 varieties were not used. In total, we used 23,603 mapped SNP in this study.

We built a consensus map with the Biomercator software (Arcade et al. 2004). We used the map published by Le Gouis et al. (2012), based on Somers et al. (2004), as a reference. This map contains SSR and DArT markers, and the location of several major genes (Vrn, Ppd, Rht). SNP was projected on it, from non-published maps containing 535 markers in common with this reference map. The Strudel software was used to check map alignments (Bayer et al. 2011) and mapping errors were corrected.

Linkage disequilibrium

We used the r^2 estimator (Hill and Robertson 1968) to assess linkage disequilibrium (LD). LD was calculated for every pair of markers mapped on the same chromosome, and then r^2 was plotted against map distance. For every chromosome, LD decay (cM) is estimated at the point where a curvilinear function proposed by Hill and Weir (1998) intersects the threshold of the critical LD. Critical LD was the 95th percentile of the unlinked-r^2 assessed on 100,000 randomly chosen pairs of unlinked loci (mapped on different chromosomes) which were square root transformed to approximate a normally distributed random variable (Breseghello and Sorrells 2006).

Association mapping study

Following Patterson et al. (2006), we did not find any structure in this 214-varieties panel. Indeed, the largest eigenvalue was not significant ($P = 0.043$). Thus, we tested SNP-trait association using a mixed model K (Yu et al. 2006) written in R using the ASReml-R package (Butler et al. 2009) and expressed as:

$$y = 1\mu + S\alpha + Zu + \varepsilon$$

where y is a vector of estimated genetic values, 1 is a vector of 1's, μ is the intercept, α is the additive effect of the tested SNP, u is a vector of random polygenic effects assumed to be normally distributed $N(0, \sigma^2_y K)$ with K a matrix of relative kinship, S and Z are incidence matrices, ε is a vector of residual effects.

K was estimated as $1(n \times n) - R_{\text{dist}}$ where R_{dist} is the modified Rogers' distance (Rogers 1972) matrix based on 3,461 SNP spread over the genome and with less than 0.1 missing data and $1(n \times n)$ is a matrix of 1's of the same size as the R_{dist} matrix ($n = 214$).

To summarise, we tested 23,603 SNP on 28 traits using the adjusted means of 214 European elite varieties. There is no widespread method to define QTL boundaries from GWAS results. So, we proceeded as follows. First, for each trait, we computed LD between every significantly associated SNP (quantitative trait nucleotide—QTN). LD blocks were defined as a group of QTN belonging to the same LD cluster (clustering by average distance) using a cutoff of (1-"critical LD"). We define the initial QTL boundaries as the minimum and maximum map position of QTN belonging to the same LD block. Then, as previously described, we assessed LD between every mapped SNP within a window covering 10 % of the chromosome length and centred on each QTL. We used the LD decay to extend the previous boundaries. This second step aimed to take into account possible LD with the causal mutation at the first QTL boundaries (for detail Suppl data 2). We only defined QTL for LD blocks containing SNP mapped on the same chromosome. For each trait, QTL with overlapping boundaries were considered the same if the alleles increasing the trait value at each were themselves correlated positively.

Phenotype simulation and power

The statistical power provided by the panel was evaluated through simulation studies where $-\log_{10}(P)$ thresholds, narrow-sense heritability and variance explained by a SNP were the three modulated parameters. We set $-\log_{10}(P)$ threshold at 3, 4, 5, 6; narrow-sense heritability (h^2) at 0.3, 0.6, and 0.9; and variance explained by the SNP (π) at 0.010, 0.030, 0.050, 0.075, 0.100, 0.150, and 0.200.

Phenotypes were simulated as follows:

$$y_i = g_i + a_{ij} + \varepsilon_i \quad (1)$$

where y_i is the simulated phenotype of the variety i, g_i is the genetic additive background effect of variety i, a_{ij} the additive effect at the quantitative trait nucleotide (QTN) j of variety i allele, and ε_i a residual error term sampled from a normal distribution $N(0, \sigma\varepsilon^2)$.

First, $k = 100$ SNP were chosen to simulate the genetic background effect. This selection is made by forming k-means cluster based on the genotyping incidence matrix and selecting the SNP nearest the centroid of each cluster

(Lorenz et al. 2010). Thus, if g_i is the genetic background effect of variety i:

$$g_i = \sum_{k=1}^{k=100} a'_{ik}, \quad a'_{ik} = \begin{cases} 1 \\ 0 \end{cases} \tag{2}$$

with a'_{ik} the effect of the variety i allele at the locus k.

Narrow-sense heritability (h^2) is defined by:

$$h^2 = \frac{\sigma g^2 + \sigma j^2}{\sigma T^2} \tag{3}$$

where σj^2 the genetic variance related to QTN j different from k, σg^2 the variance related to the genetic background, and σT^2 the total variance.

The variance explained by QTN j (π) is defined by:

$$\pi = \frac{\sigma j^2}{\sigma T^2} \tag{4}$$

Total variance (σT^2) is deduced from Eqs. (3) and (4) as h^2 and π are fixed in each simulation study:

$$\sigma^2 = \frac{\sigma g^2}{h^2 - \pi} \tag{5}$$

Given the percentage of variance explained by QTN j (π), its additive effect (a_j) is calculated by Falconer and Mackay (1996) as:

$$a_j = \sqrt{\frac{\pi \times \sigma T^2}{p_j(1 - p_j)}} \tag{6}$$

with p_j the allele frequency of the reference allele at locus j. Thus, if variety i allele at QTN j was the reference allele, a_{ij} from Eq. (1) was equal to a_j, else a_{ij} was equal to $-a_j$.

Finally, the variance of the residual error term ($\sigma\varepsilon^2$) was computed as:

$$\sigma\varepsilon^2 = \left(1 - h^2\right) \times \sigma T^2 \tag{7}$$

In total, 400 SNP were randomly chosen to play in turn the role of the QTN j with $j \neq k$ (QTN $\neq$ genetic background effect) for each pair of h^2 and π parameter values. The statistical model used to detect associations between SNP and simulated phenotypes was the previously described model K. In the same way, QTL were defined following the two steps already described. Detection power was estimated by the ratio of the number of times a true QTN was located in the computed QTL to the total number of tests. The SNP selected as being the true QTN j was not tested per se.

Prediction

The percentage of total variance explained by each significant SNP was first assessed for each trait using a simple regression of overall adjusted mean on the SNP (r^2_{snp}). Then, for each trait, the predicted values of varieties were estimated by summing the allele effects assessed in GWAS at associated loci. To avoid redundancy, only one SNP per LD block was kept; that which explained the most variance.

This model was first used to predict overall adjusted means. It was then used to predict adjusted means in each of the eight individual environments. Consequentially, we computed two types of correlations (r^2): the correlation between predicted values and overall adjusted means (r^2_{adj}), and the correlation between predicted values and each of the eight individual environments (r^2_{env}).

To assess transferability of GWAS results to field trials, we calculated a prediction similarity [mean(r^2_{env})/r^2_{adj}] that we plotted as a function of trait heritability.

Colocalisation and network approach

To assess the impact of genetic correlation and pleiotropy, we analysed colocalisations through a network approach. QTL colocalisation between two traits were statistically tested using the probability of an hypergeometric law ("sampling without replacement"; Larsen and Marx 1985) with the total cumulative length of QTL for trait i and trait j and the total map length as parameters of the hypergeometric distribution. The cumulative length of QTL shared by trait i and j was the parameter of the probability. A fairly stringent threshold of $P = 0.001$ was set as the criteria of significance.

On the basis of significant colocalisations, inter-trait relationships were then studied through a network approach using traits as nodes and the percentage of one trait QTL overlapping another trait QTL as edges. Betweenness centrality was computed on each node following Opsahl et al. (2010) method with $\alpha = 0.5$ to equally take into account the number of edges and edges' weights in the calculation. To statistically test trait betweenness centralities values, this network was then permuted 500 times to assess the empirical distribution of betweenness centrality, and thus determine the statistical law underlying this distribution.

Results

Genetic map and linkage disequilibrium

The consensus genetic map obtained had a total length of 3,167 cM. To finely map QTL, LD has to decay rapidly and SNP density has to be high to ensure that at least one SNP is linked to the causal mutation. While diversity level is similar in the A and B genomes, it is greatly reduced in the D genome (Cadalen et al. 1997), contributing to its

higher levels of LD. Indeed, mean LD decay on genome A, B, and D was, respectively, 0.52, 0.70, and 2.14 cM. LD decay is the estimated distance from which two SNP are not genetically linked, meaning that their LD (r^2) is inferior to the critical LD. Critical LD was estimated from a sample of 100,000 pairs of unlinked SNP which revealed a mean unlinked-r^2 of 0.016 and a critical LD (95th percentile) of 0.23.

A rapid LD decay predicts a good mapping resolution in GWAS. Though as previously mentioned, it can decrease power if SNP density is not sufficient. SNP density ranged from 0.7 cM^{-1} for chromosome 4D to 14.6 cM^{-1} for chromosome 7A (Table 2). On genomes A and B, SNP density seemed sufficient with respect to LD decay. On genome D, the lower SNP density may be compensated for by the higher LD, but QTL will be less precisely defined.

Power assessment

Choosing a P value threshold has to balance the control of Type I error (false positive) with Type II error (false negative). Considering power simulation and the expectation of small-effect QTN, a $-\log_{10}(P)$ threshold of 3 was adopted as a criterion for significant marker-trait associations. Indeed, a more stringent threshold inflated Type II error and thus reduced extremely the power of detection, notably on QTN explaining less than 10 % of the variance (Fig. 1). At a QTN heritability of 5 % and a narrow-sense heritability of 0.6, power was dramatically reduced from 55 % to 7 % when $-\log_{10}(P)$ threshold increased from 3 to 6 (Fig. 1).

At a $-\log_{10}(P)$ score threshold of 3, when the genetic variance explained by the locus was greater than 10 %, trait heritability did not affect power, and Type II error was reduced. In general, the variance explained by the QTN was the main factor that influences the power of the study as compared to trait narrow-sense heritability. It should be noted that with a weakly stringent threshold of 3, the power to detect an association for a QTN, which explained 5 % of the total genetic variance, was 48, 55, and 60 %, for a trait narrow-sense heritability of 0.3, 0.6, and 0.9, respectively.

Table 2 SNP used in association: number of mapped SNP, coverage on the consensus map, SNP density and LD decay at a critical LD $r^2 = 0.23$. Critical LD was assessed as in Breseghello and Sorrells (2006)

Chr	SNP	Coverage (cM)	SNP density (cM^{-1})	LD decay (cM)
1A	1,246	110.4	11.3	0.49
1B	2,055	128.5	16	0.19
1D	430	121.7	3.5	2.71
2A	1,454	262.7	5.5	1.39
2B	2,362	205.8	11.5	0.70
2D	402	130.9	3.1	0.80
3A	1,151	155.1	7.4	0.68
3B	1,972	147.8	13.3	0.05
3D	253	104.7	2.4	1.07
4A	786	123.4	6.4	0.21
4B	849	143.3	5.9	0.70
4D	97	139.7	0.7	2.43
5A	1,604	186.1	8.6	0.32
5B	2,243	262.4	8.5	2.19
5D	327	115.6	2.8	0.94
6A	1,588	122.0	13	0.19
6B	1,603	115.0	13.9	0.05
6D	254	136.8	1.9	1.02
7A	1,782	122.2	14.6	0.38
7B	1,034	198.5	5.2	1.06
7D	246	134.9	1.8	6.00
Total	23,603	3,167.5	7.5	1.12

GWAS results

Overall, 1,010 SNP were significantly associated (QTN) to at least one of the 28 studied traits. Considering QTN, LD blocks and LD around associated regions, 333 QTL were mapped with a mean size of 3.2 cM. Ninety percent (between the 5th and 95th percentile) of QTL had a range within 0.1–14 cM indicating that the method used to define QTL is mostly efficient. In few cases, the assessments of LD decay in the chromosomal region containing QTN may

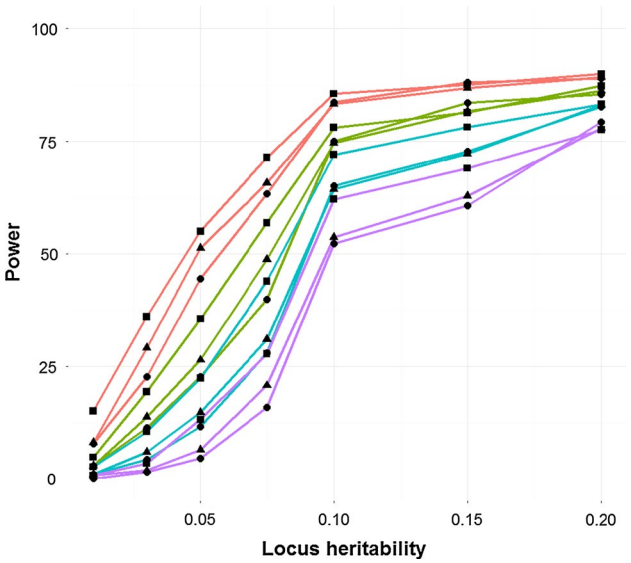

Fig. 1 Influence of trait heritability and $-\log10(P$ value) threshold on the relation between locus heritability and power of detection in a 214-lines wheat association panel. In *red*, *green*, *blue*, *violet*, respective LOD score thresholds are 3, 4, 5, and 6. *Square*, *triangle*, and *circle* represent a respective narrow-sense heritability of 0.9, 0.6, 0.3

not correctly fit and QTL boundaries must be used with caution.

In agreement with SNP density and the genetic diversity, the number of QTL on genome D (42) was smaller than on genome A (142) and B (149). Homeologous group 2 maximised the number of QTL with 73 QTL. The number of QTL by trait ranged from 6 for NutE to 21 for %N_S (Table 3).

Predictions

First, we assessed the variance explained by each significant SNP (QTN). Then, we predicted overall adjusted means and each of the eight environments' adjusted means. On average, QTN explained 8.81 ± 4.79 % of the overall adjusted means (r^2_{snp}). On overall adjusted means, the best prediction (r^2_{adj}) was made on HI (Table 4). Using 20 SNP, we were able to explain 61.4 % of the genetic variation. Using 15 SNP on NUE, we were able to explain 55.7 % of the overall adjusted mean variation (Fig. 2) and 29.7 ± 4.9 % of the individual environment's variation (Table 4). On the environments' data (r^2_{env}), flowering date was the best predicted trait with 55.3 % of the variation explained on average.

Differences between predictions made on overall adjusted means (r^2_{adj}) and predictions on individual environment values (r^2_{env}) resulted from genotype × environment interactions. Thus, it was linked to trait broad-sense heritability. In fact, the transferability of our GWAS results to environmental values was exponentially proportional to trait broad-sense heritability (Fig. 3). This means that GWAS results became rapidly powerless to predict phenotypic values as broad-sense heritability decreased.

Colocalisation network

Altogether, the QTL covered 20 % (646/3,167) of the genetic map. There were colocalisations for 39 % of the QTL identified. Major regions of colocalisation were on chromosomes 1B, 2B, and 7A (Suppl data 3). Considering NUE and its two components, N uptake and N utilisation, there was no common QTL between NupEMat and NUE, but two NutE QTL (out of six) colocalised with NUE QTL and acted in the same way on both traits. NUE QTL (9/14) which colocalised with NutE_Prot QTL had opposite effect on these traits. By comparing QTL for the N uptake efficiency at flowering time (NupEFlo) and at maturity (NupEMat), we found that only one QTL was in common between these two traits.

Figure 4 provides a visual representation of the frequencies of QTL colocalisations. Using a bootstrap procedure with 500 permutations, it was assessed that the empiric betweenness centrality followed a gamma distribution (shape = 2.169, rate = 0.079; Suppl data 6). This distribution was used to test trait betweenness centrality. Four traits had a significant ($P < 0.05$) high betweenness centrality: INN_FLO, FLO, NutE, %N_Flo were ordered from the most significant to the less significant. We should notice that INN_FLO, %N_S, and FLO were not independent as we detected four chromosomal regions of colocalisations between these three traits. Two of them affected the three traits in the same ways. Two of them acted oppositely between FLO and the two other traits. All common QTL between %N_Flo and INN_FLO affected both traits in the same way.

Discussion

QTL definition and power

In most studies, authors fixed a window around QTN peaks often based on linkage disequilibrium to define associated chromosomal regions in GWAS. However, massive variation of LD exists along the chromosomes in wheat (Würschum et al. 2013). In this study, we suggested a method based on LD between QTN and LD within the chromosomal region of interest and assessed its power of detection. This method had the advantage of being based on LD decay in the chromosomal region of interest. Moreover, authors focus on *P* value methods (ad hoc and post hoc) to control false positive rate, although the way they design their associated region influences it. Indeed, linkage disequilibrium between causal mutations and associated SNP or mapping error can lead to the construction of a chromosomal region which does not contain the causal mutation even though the SNP-trait association was real.

Regarding power simulation and error type II, we chose a $-\log_{10}(P)$ threshold of 3 to validate SNP-trait associations. Our real false positive rate (error type I) was not only influenced by this $-\log_{10}(P)$ threshold. Indeed, in our real error Type I, we should consider all QTL which did not contain the causal mutation whether the SNP-trait association was real or not. Using the results of the power simulation studies, we estimated our real false positive rate at 7 % (for a QTN heritability between 5 and 10 %; Suppl data 2). If we had chosen a $-\log_{10}(P)$ threshold of 6, it would have been 3 %. Thus, increasing *P* value threshold reduced real error Type I for small-effect QTN yet drastically decreased power (Fig. 1). Moreover, for QTN with a heritability >10 %, a *P* value threshold superior to 3 slightly increased the real error Type I due to smaller QTL (Suppl data 2).

In GWAS, the real issue to control error Type I is not in the definition of a stringent *P* value threshold. It is in the development of a powerful method to define QTL boundaries, particularly in the case of GWAS oriented to gene

Table 3 QTL detected on a wheat association panel for 28 traits. QTL boundaries were defined as the minimum and maximum genetic position of QTN belonging to the same LD block (for LD blocks containing SNP mapped on the same chromosome) extended by the LD decay assessed on a window covering 10 % of the chromosome length centred on the mean genetic position. See Table 1 for trait abbreviations

TRAIT	QTL		Effect (%)[b]		MAF[c]		Size (cM)		Group 1			Group 2			Group 3			Group 4			Group 5		Group 6			Group 7		
	Total	Positif QTL (%)[a]	Mean	SD	Mean	SD	Mean	SD	A	B	D	A	B	D	A	B	D	A	B	D	A	B	A	B	D	A	B	D
ABSN	13	69	12.9	2.2	0.20	0.12	5.95	11.50		2		1	1			1		1	1		1	1	1	1			2	
ADM_FLO	12	50	2.1	0.5	0.23	0.12	2.14	2.71	1		1		2	1	1	1		1						1		2	1	
ADM_S	16	50	2.7	0.5	0.26	0.13	4.12	7.67		2			1	2	1	1			1	1	2	1	1			1	1	1
DMGY	10	70	2.3	0.7	0.19	0.13	0.79	0.82		1			2		2						1		2				1	1
EFFG	19	32	2.4	0.5	0.24	0.11	3.27	8.27	2	1		2	3			1		2			1	1	1	1		1	2	1
EFFREMN	12	67	1.2	0.3	0.18	0.12	2.33	3.00	1	2			1				2	1			1		1				3	
FLO	18	78	0.8	0.2	0.28	0.13	1.64	1.66	1			1	4	3	1	1					4		1		1	1		
GNY	11	36	1.5	0.3	0.28	0.12	6.96	9.07		3	1	1			1				1		1		1			1	1	
GPC	8	13	3.0	1.2	0.19	0.14	4.34	8.43					1		1	1		1			1	1			1		1	
GPD	8	38	0.16	0.04	0.31	0.15	1.63	1.16				2	2										2		1	1		
HI	18	72	1.6	0.6	0.25	0.16	1.73	2.20	1	1		2	2	1	1	3		1	1		1	1	1			1	1	
INN_FLO	7	14	2.1	0.2	0.27	0.09	4.86	6.40		1			2	1								2	1					
NFA	10	10	2.1	0.6	0.21	0.08	1.93	3.56					2		1	1		2			1	1		1				1
NHI	10	80	0.6	0.1	0.29	0.14	2.52	3.97	1	2		1							1		2	1			2			
NSA	14	43	3.4	1.4	0.24	0.16	1.62	2.81		1		1	3	1		2	1	1			1	1			1	1		
NTA	8	13	1.5	0.3	0.25	0.16	5.66	8.56		1	1	1	1								1		2		1			
NUE	14	57	2.3	0.6	0.22	0.12	3.00	6.61	1				1	1	2	2	1	1			1			1	2			1
NUE_Prot	11	18	2.9	1.1	0.22	0.15	4.30	7.13					2	1	1	1		1			1	2		1		1		
NupEFlo	7	0	2.2	0.4	0.22	0.11	2.36	4.23				1	1		1			1				1		1	1			
NupEMat	10	30	1.5	0.3	0.27	0.15	5.38	7.80		2	1	1						1			2	1	1			1		
NutE	6	67	2.6	0.6	0.22	0.13	1.22	1.73					1	1	1			1				1			1			
NutE_Prot	16	25	2.5	0.9	0.23	0.13	2.47	6.20	1				2		2	1	1	1			3		1		1	1	1	1
PH	14	14	2.9	0.5	0.20	0.11	7.67	14.97		1			2			1		1		1	2	1	2			2	1	
REMN	12	42	2.7	0.4	0.20	0.11	1.80	4.06		3			1					1				1	1			2	2	1
SA	11	45	4.3	1.2	0.15	0.12	1.44	1.61					1		2	2		1					2	1	1		1	
TKW	9	44	2.4	1.0	0.25	0.13	3.40	3.37		2				1		1		1				1	2				1	
%N_FLO	8	25	2.6	0.5	0.31	0.13	2.24	2.40					2	1								1	3	1				
%N_S	21	33	4.0	1.0	0.19	0.10	2.97	4.58		3		3	1	1	1	2		1	1	1		3	2				2	
Total									9	28	4	17	41	15	19	22	5	21	6	3	27	22	28	7	8	21	23	7

[a] Percentage of QTL for which the most frequent allele had a positive effect on trait

[b] Effect expressed in percentage of trait mean (except for GPD)

[c] *MAF* minor allele frequency

Table 4 Summary of GWAS results predictions made by SNP (r^2_{snp}) and using the sum of SNP effect on both overall adjusted mean (r^2_{adj}) and on eight individual environments (r^2_{env}). To avoid redundancy, for each LD block, the SNP which maximised the genetic variance explained was selected

Trait	SNP[a]	Prediction on adjusted means			Prediction on individual environments	
		r^2_{snp} (%)		r^2_{adj} (%)	r^2_{env} (%)	
		Mean	SD		Mean	SD
ABSN	14	6.0	0.9	37.7	6.9	3.3
ADM_FLO	13	6.9	4.4	40.9	18.5	13.1
ADM_S	17	6.5	3.8	52.8	27.7	4.2
DMGY	12	11.5	9.0	53.6	30.8	6.2
EFFG	20	6.1	1.0	42.3	7.0	3.6
EFFREMN	13	7.5	1.7	40.4	8.4	4.4
FLO	20	8.6	6.5	58.5	55.3	2.6
GNY	11	7.3	2.9	40.0	9.9	5.7
GPC	10	14.0	8.7	57.5	37.9	10.8
GPD	8	7.8	3.9	33.7	15.6	5.1
HI	20	8.6	6.4	61.4	32.4	4.3
INN_FLO	8	11.5	4.3	40.0	12.8	10.5
NFA	13	6.3	2.3	34.2	5.7	5.2
NHI	11	5.3	2.9	37.2	10.9	5.8
NSA	15	6.3	3.4	38.2	9.7	5.2
NTA	9	8.1	3.0	32.0	7.0	6.1
NUE	15	8.7	7.2	55.7	29.7	4.9
NUE_Prot	11	12.4	8.8	59.7	35.5	11.5
NupEFlo	9	7.4	3.0	27.7	5.2	5.6
NupEMat	11	6.4	2.9	31.4	6.9	4.3
NutE	6	8.7	6.4	38.3	23.2	9.1
NutE_Prot	18	10.1	8.7	59.8	34.4	7.4
PH	17	10.5	4.9	48.6	37.0	16.0
REMN	12	6.3	1.4	28.3	4.8	3.5
SA	12	7.4	3.8	41.0	22.1	8.1
TKW	10	8.1	2.9	39.0	32.3	3.6
%N_FLO	10	11.4	6.9	45.5	20.3	8.6
%N_S	21	8.3	4.4	57.8	25.8	13.5

[a] SNP number can differ from QTL number in Table 3 when LD blocks contained SNP mapped on different chromosomes (as no QTL was defined but one SNP was used in prediction)

discovery. This field has practically never been investigated and publications mainly focus on *P* value. We advocate balancing QTL coverage, real error Type I, and power altogether. An improvement of our methods could be to adapt the construction of the associated region to QTN heritability.

Power, locus heritability, and genetic determinism

The fraction of total genetic variance explained by a single significantly associated SNP (QTN) averaged 8.81 ± 4.79 % which is coherent regarding the simulation study. Indeed, the power started to be maximised from a locus heritability of 10 % (at a $-\log_{10}(P)$ threshold = 3, Fig. 1). Yet variability existed and fraction of total genetic variance ranged from GPC (14.0 ± 8.7 %) to NHI (5.3 ± 2.9 %).

When numerous QTN explained a small fraction of genetic variance, we can presume that the GWAS study was powerful and that the genetic determinism underlying this trait is highly polygenic. When QTN have larger locus heritability, the cause can be a less polygenic genetic determinism and/or a lack of power due to low narrow-sense heritability. Narrow-sense heritability estimates the proportion of additive variance on total variance (Falconer and Mackay 1996). Thus, narrow-sense heritability is also linked to the importance of epistasis in the trait genetic architecture. In this study, we have not searched for epistasis. However, several studies have highlighted its impact. For example, GPC is controlled by major protein concentration genes (Payne 1987; Uauy et al. 2006; Avni et al. 2013) and significant interactions between them (Dumur et al. 2004; Conti et al. 2011; Plessis et al. 2013). Another example is epistatic

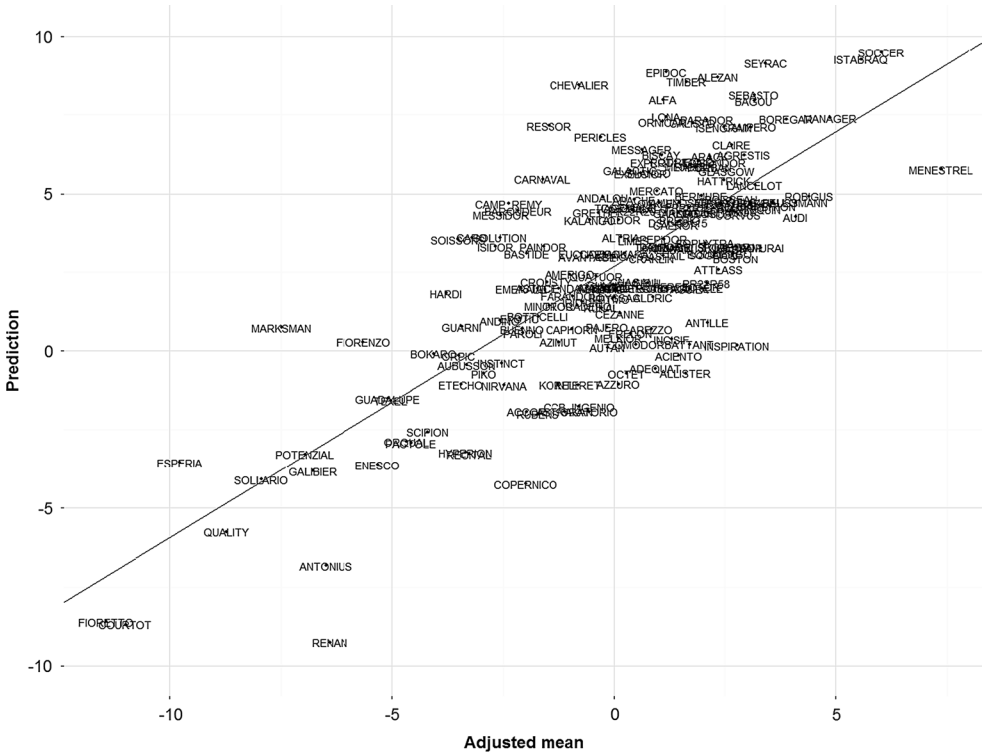

Fig. 2 Prediction of NUE values as a function of overall adjusted mean for 214 wheat lines. Predictions were made summing the effects of 15 significantly associated SNP. The following regression function is also plotted: $y = 0.86x + 2.66$ ($r^2 = 0.56$; $P < 0.001$)

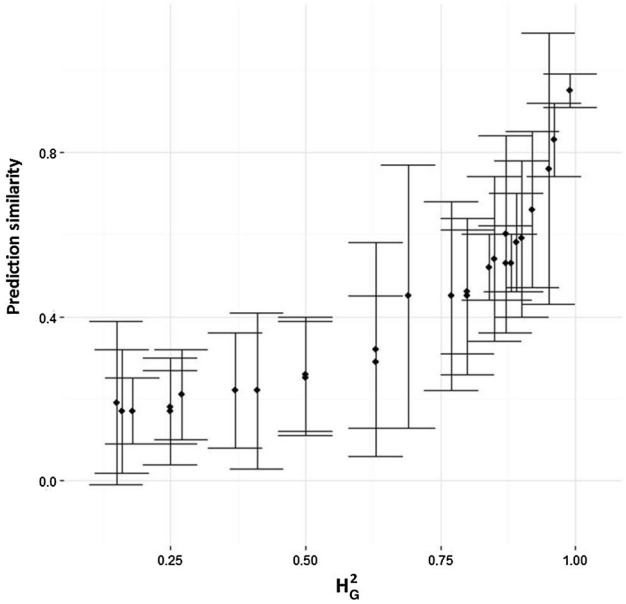

Fig. 3 Prediction similarity (r^2_{env}/r^2_{adj}) between predictions made on overall adjusted means (r^2_{adj}) and the ones made on individual environment's values (r^2_{env}) as a function of generalised heritability (H^2_G) of 28 traits. Means (*diamond*), standard deviations (*whisker*). Mean (r^2_{env}/r^2_{adj}) $= -0.39e^{H^2_G}$ ($r^2 = 0.88$; $P < 0.001$)

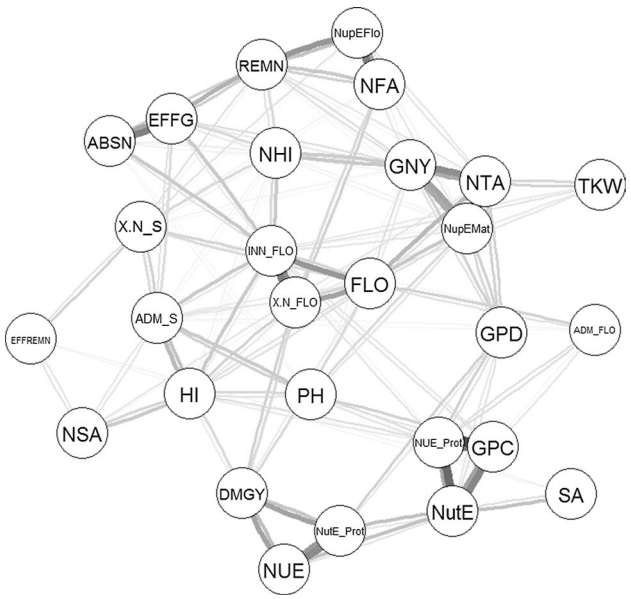

Fig. 4 Network of QTL colocalisations for 28 traits measured on a 214-lines wheat association panel. This network is based on the percentage of common QTL between traits after correction using a hypergeometric law to determine significant colocalisations ($P < 0.001$). Link thickness is function of the percentage of common QTL, from 5 % for the thinnest to 100 % for the thickest (values in Suppl data 5)

contribution in the genetic control of PH is important as revealed by Novoselovic et al. (2004), Zhang et al. (2008), and Wu et al. (2010). Using a doubled haploid wheat population, Zhang et al. (2008) estimated first-order epistatic contribution up to 19.9 % of the PH phenotypic variation.

Authors have often focused on epistatic interactions between SNP having a significant additive effect. However, epistatic interactions between SNP without additive effect can also explain genetic variability (Huang et al. 2014) as detected for heading date (Le Gouis et al. 2012). Nonetheless, whole genome scan for epistasis is a real computational and analytic challenge, which will surely help pathways mining (Philipps 2008; Mackay 2014).

Candidate genes and comparison with previously published QTL

Altogether, we detected 333 QTL on 28 traits. Significant colocalisations (QTL boundaries overlapping) between some of them and candidate genes or previously published QTL deserve to be pointed out. Regarding major genes for precocity, only the photoperiod sensitivity gene Ppd-D1 on chromosome 2D colocalised with QTL of FLO, HI, INN_FLO, %N_FLO, %N_S, affecting all these traits in the same way (late genotype have higher HI, INN_FLO, %N_FLO, and %N_S). Ppd-D1 also colocalised with an ADM_S QTL, with an opposite effect. Two factors can explain that Vrn genes were not associated to precocity: (1) this panel contains only winter wheat varieties and (2) only autumn trials were sown with vernalization requirements fulfilled.

On chromosome 4D, the dwarfing gene Rht-D1 (Rht2) was tested and had an expected significant effect on PH and ADM_S.

Similarly, the three closely mapped genes coding the glutenins and gliadins (Glu3A, Glu3B, and Gli) not surprisingly colocalised with QTL of NUE and NutE_Prot located on chromosome 1A. Moreover, the structural gene for high molecular weight glutenins GluD1 located on chromosome 1D lay within the boundaries of QTL affecting GNY, NTA, and NupEMat.

Several genes from the N assimilation pathway have already been associated to NUE QTL including the genes coding for glutamate synthase (NADH-Gogat) located in QTL on chromosome 3A, and 3B (Quraishi et al. 2011). On chromosome 3A, this colocalised with QTL of NFA, NupEFlo, and %N_S. On chromosome 3B, the NADH-Gogat gene colocalised with QTL of NUE_Prot, GPC, and ABSN. The gene for glutamine synthetase GS1 on 6A (Habash et al. 2007) colocalised with a cluster of QTL for EFFREMN, GPD, NutE_Prot, DMGY, and %N_S. Several publications already mentioned this region as affecting grain number per ear (Habash et al. 2007; Quarrie et al. 2005), NupEMat (An et al. 2006; Xu et al. 2013), root dry weight (An et al. 2006), %N_S and DMGY (Xu et al. 2013).

On chromosome 4B, a QTL of %N_S colocalised with numerous previously published QTL of nitrogen efficiency-related trait (An et al. 2006; Guo et al. 2012), glutamate dehydrogenase and glutamine synthase activity (Fontaine et al. 2009), harvest index (Xu et al. 2013), ears, spike, and grain-related trait (Quarrie et al. 2005; Habash et al. 2007; Laperche et al. 2007; Fontaine et al. 2009), and root morphology (Laperche et al. 2006). Previously published results were in part due to the presence of Rht-B1 (Rht1) in this chromosomal region. In our case, a diagnostic marker for Rht-B1 was tested and no significant effect was detected for any trait most probably because of the unbalanced allele frequencies of the combination of Rht-B1 and Rht-D1 (0.05, 0.65, 0.18, and 0.12 for the four allelic classes *Rht-B1b/Rht-D1b*, *Rht-B1b/Rht-D1a*, *Rht-B1a/Rht-D1b*, and *Rht-B1a/Rht-D1a*). The glutamine synthetase gene GSe (Habash et al. 2007) mapped using the SSR gpw7026 (Sourdille et al. 2004; Fontaine et al. 2009) was also within this QTL confidence interval and may be a good candidate gene to investigate.

On chromosome 2A, the Rbcs (Xpsr109) gene for the small subunit of the chloroplast photosynthetic enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) was located in a %N_S QTL, and has already been shown to colocalise with a QTL for N grain concentration (Laperche et al. 2006), and from a meta-QTL analysis on yield and yield-related traits (Zhang et al. 2010). Considering the small size of this QTL in this study (1.6 cM), and the link between N remobilisation and Rubisco subunit expression and degradation (Hörtensteiner and Feller 2002; Gregersen et al. 2008), Rbcs has to be considered as a good candidate gene.

Further investigations are needed on two promising regions where no obvious candidate genes were found within QTL boundaries. On chromosome 5B (gwm67-BCD351), a region linked to the INN_FLO colocalised with QTL previously published by Fontaine et al. (2009) on carbon percentage in flag leaf, and Habash et al. (2007) on nitrogen percentage in peduncle. As the nitrogen nutrition index (INN) refers to the minimum N concentration enabling maximum biomass growth (Justes et al. 1994), this confirms the effect of this region on nitrogen/carbon balance before remobilisation. On chromosome 7B (wPt-3530-wPt-7113), Laperche et al. (2007) published a QTL of %N_S which colocalised with one of this study affecting the same trait. This region also appeared in Laperche et al. (2006) as being linked to the lateral root number and the primary root length, and in Habash et al. (2007) for GNC.

Breeding strategies

As we worked on a panel composed of commercial varieties mostly registered between 1985 and 2010, results of this study have to be discussed in light of selection pressures. Although QTL have been detected, if favourable alleles are already fixed in the more recent varieties, those QTL are not so useful in future breeding.

As expected, favourable alleles are more frequent in recent varieties for QTL affecting traits under a high selection pressure than on QTL affecting untargeted traits. We estimated a positive correlation ($P < 0.001$; $r^2 = 0.48$) between the frequencies of alleles having a positive effect (in varieties released from 2005) and genetic progresses assessed by Cormier et al. (2013). Cormier et al. (2013) showed that in this panel of European elite varieties, NUE was increased by improving N utilisation (NutE: +0.20 % year^{-1}) and remobilisation (NHI: +0.12 % year^{-1}; %N_S: −0.52 % year^{-1}) through a major positive selection pressure on grain yield (DMGY: +0.45 % year^{-1}), while maintaining constant N uptake. In agreement, we found that for DMGY QTL, NutE QTL, and %N_S QTL, the median frequency of favourable alleles (in varieties released from 2005) was, respectively, 88, 68, and 79 % (Suppl data 7). Moreover, for a given trait, the frequency of alleles having a positive effect in recent varieties is directly linked to the genetic correlation between this trait and DMGY ($P < 0.001$; $r^2 = 0.49$; Suppl data 7). Thus, favourable alleles are already well represented in new varieties at QTL associated to traits directly (e.g. DMGY) or indirectly (e.g. NutE) targeted by breeding. This study has provided information to facilitate their monitoring.

Studying correlations between traits using QTL colocalisations rather than genetic correlations has the advantage of taking into account trait genetic architecture and the power with which we can dissect them. Moreover, it gives a better estimation of the pleiotropic effect of QTL-based selection on a trait. Indeed, the genetic correlation is symmetric ($r_{a/b} = r_{b/a}$), contrary to the percentage of QTL colocalising between two traits. For example, based on our detection, selection on GPC QTL will surely affect NUE_Prot as all GPC QTL are also NUE_Prot QTL. However, only 73 % of QTL for GPC would be affected by selection on NUE_Prot QTL.

Results of colocalisation analyses revealed that we should select on INN_FLO, FLO, NutE, and %N_Flo QTL to maximise the number of affected traits. As 57 % (4/7) of INN_FLO QTL and 50 % (4/8) of % N_Flo QTL were also FLO QTL, effect of phenology and pre-anthesis uptake are mixed. Thus, QTL controlling flowering time should be our first concern. Anthesis corresponds to a physiological transition and consequently, the date of this transition has a major impact on genotype × environment (G × E) interaction (Kamran et al. 2014). In this study, we observed an average genotypic flowering time standard deviation of 7 days. As varieties were tested in a small range of slightly contrasted environments, anthesis date directly affected G × E interaction and above all varieties' genetic values, favouring genotypes adapted to these environments. This created a confounding effect of major phenology genes (Reynolds et al. 2009) which are more likely to be associated to agronomic traits.

None of the central traits (INN_FLO, FLO, NutE, and %N_Flo; Fig. 4) was linked to final N uptake. As mentioned before, recent breeding efforts improved N remobilisation and N utilisation, and not N uptake (Cormier et al. 2013). Thus, selection pressure enhanced N utilisation centrality in our network (Fig. 4). In this panel, the low genetic variance of the N uptake was not sufficient to reveal meaningful correlations with other agronomic traits and thus significant QTL colocalisations. Nevertheless, as a component of NUE, N uptake is a promising lever of action (Hirel et al. 2007; Foulkes et al. 2009). This study has provided tools to start selecting for N uptake in elite varieties without fastidious phenotyping or can be used as an entry point in investigating genes and pathways controlling this trait (Korte and Farlow 2013) with further investigations in a more diverse panel.

Results on QTL colocalisations highlighted the importance of focusing on pre-anthesis nitrogen status, especially on INN_FLO which had a good heritability (0.63) and for which QTL have also the same effect on TKW and NUE_Prot.

Conclusions

Identification of chromosomal regions associated with nitrogen use efficiency-related traits at both high N levels and moderate N will help breeding for better adapted varieties. To our knowledge, this work is the first published study that reports GWAS results on N use efficiency in small grain cereals using a high marker density for precise mapping of genomic regions. Using an LD-based method to define QTL boundaries, 333 QTL were identified on 28 traits. Several colocalisations between our QTL and previously published QTL were pointed out. Using a network approach on colocalisation frequencies between traits, this study highlighted the interest of working on N status at flowering, and underscores the effect of recent breeding on N utilisation efficiency.

Author contributions Statistical analyses and manuscript were conducted by FC during his PhD thesis co-directed by SP and JLG. PD provided useful help on power assessment methods and interpretations. JLG, SL, and SP were implicated in methods, interpretations, and in reviewing the manuscript.

Acknowledgments These analyses were a part of a PhD thesis (129/2012) supported by the ANRT (Association Nationale de la Recherche et de la Technologie). Data were obtained thanks to the support of the ANR ProtNBle project (06 GPLA016). The authors are also grateful to experimental units staff at Estrées-Mons experimental unit (INRA) and at Villiers-le-Bâclea and Vraux (Arvalis), and to

Quang Hien Le for English style proofing. Sincere thanks to Ian Mackay for his involvement in the reviewing process.

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standards The authors declare that the experiments comply with the current laws of the country in which they were performed.

References

An D, Su J, Liu Q, Zhu Y, Tong Y, Li J, Jing R, Li B, Li Z (2006) Mapping QTLs for nitrogen uptake in relation to the early growth of wheat (*Triticum aestivum* L.). Plant Soil 284:73–84

Arcade A, Labourdette A, Falque M, Mangin B, Chardon F, Charcosset A, Joets J (2004) BioMercator: integrating genetic maps and QTN towards discovery of candidate genes. Bioinform 20:2324–2326

Avni R, Zhao R, Pearce S, Jun Y, Uauy C, Tabbita F, Fahima T, Slade A, Dubcovsky J, Distelfeld A (2013) Functional characterization of GPC-1 genes in hexaploid wheat. Planta 239:313–324

Barraclough PB, Howarth JR, Jones J, Lopez-Bellido R, Parmar S, Shepherd CE, Hawkesford MJ (2010) Nitrogen efficiency of wheat: genotypic and environmental variation and prospects for improvement. Eur J Agron 33:1–11

Bayer M, Milne I, Stephen G, Shaw P, Cardle L, Wright F, Marshall D (2011) Comparative visualization of genetic and physical maps with Strudel. Bioinformatics 27:1307–1308

Bordes J, Ravel C, Jaubertie JP, Duperrier B, Gardet O, Heumez E, Pissavy AL, Charmet G, Le Gouis J, Balfourrier F (2013) Genomic regions associated with the nitrogen limitation response revealed in a global wheat core collection. Theor Appl Genet 126:805–822

Breseghello F, Sorrells M (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177

Butler DG, Cullis BR, Gilmour AR, Gogel BJ (2009) ASReml-R reference manual. Technical report, Queensland Department of Primary Industries. http://www.vsni.co.uk/software/asreml/

Cadalen T, Boeuf C, Bernard S, Bernard M (1997) An intervarietal molecular marker map in *Triticum aestivum* L. Em. Thell. and comparison with a map from a wide cross. Theor Appl Genet 94:367–377

Conti V, Roncallo PF, Beaufort V, Cervigni GL, Miranda R, Jensen CA, Echenique VC (2011) Mapping of main and epistatic effect QTL associated to grain protein and gluten strength using a RIL population of durum wheat. J Appl Genet 52:287–298

Cormier F, Faure S, Dubreuil P, Heumez E, Beauchêne K, Lafarge S, Praud S, Le Gouis J (2013) A multi-environmental study of recent breeding progress on nitrogen use efficiency in wheat (*Triticum aestivum* L.). Theor Appl Genet 126:3035–3048

Crossa J, Burgueno J, Dreisigacker S, Vargas M, Herrera-Foessel SA, Lillemo M, Singh RP, Trethowan R, Warburton M, Franco J, Reynolds M, Crouch JH, Ortiz R (2007) Association analysis of historical bread wheat germplasm using additive covariance of relatives and population structure. Genetics 177:1889–1913

Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. J Agric Biol Envir S 11:381–393

Distelfeld A, Avni R, Fischer AM (2014) Senescence, nutrient remobilization, and yield in wheat and barley. J Exp Bot 65:3783–3798

Dumur J, Jahier J, Bancel E, Laurière M, Bernard M, Branlard G (2004) Proteomic analysis of aneuploidy lines in the homeologous group 1 of the hexaploid wheat cultivar Courtot. Proteomics 4:2685–2695

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longmans Green, Harlow

FAO (2011) Current world fertilizer trends and outlook to 2015. ftp://ftp.fao.org/ag/agp/docs/cwfto15.pdf

FAO (2012) World agriculture towards 203/2050, the 2012 revision. http://www.fao.org/docrep/016/ap106e/ap106e.pdf

Fontaine JX, Ravel C, Pageau K, Heumez E, Dubois F, Hirel B, Le Gouis J (2009) A quantitative genetic study for elucidating the contribution of glutamine synthetase, glutamate dehydrogenase and other nitrogen-related physiological traits to the agronomic performance of common wheat. Theor Appl Genet 119:645–662

Foulkes M, Hawkesford M, Barraclough P, Holdsworth M, Kerr S, Kightley S, Shewry P (2009) Identifying traits to improve the nitrogen economy of wheat: recent advances and future prospects. Field Crop Res 114:329–342

Gaju O, Allard V, Martre P, Snape JW, Heumez E, Le Gouis J, Moreau D, Bogard M, Griffiths S, Orford S, Hubbart S, Foulkes MJ (2011) Identification of traits to improve the nitrogen-use efficiency of wheat genotypes. Field Crop Res 123:139–152

Gregersen PL, Holm PB, Krupinska K (2008) Leaf senescence and nutrient remobilisation in barley and wheat. Plant Biol 10:37–49

Guarda G, Padovan S, Delogu G (2004) Grain yield, nitrogen-use efficiency and baking quality of old and modern Italian bread wheat cultivars grown at different nitrogen levels. Eur J Agron 21:181–192

Guo Y, Kong FM, Xu YF, Zhao Y, Liang X, Wang YY, An DG, Li SS (2012) QTL mapping for seedling traits in wheat grown under varying concentrations of N, P and K nutrients. Theor Appl Genet 124:851–865

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

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theo Appl Genet 38:226–231

Hill WG, Weir BS (1998) Variances and covariances of squared linkage disequilibria in finite populations. Theor Popul Biol 33:54–78

Hirel B, Le Gouis J, Ney B, Gallais A (2007) The challenge of improving nitrogen use efficiency in crop plants: toward a more central role for genetic variability and quantitative genetics within integrated approaches. J Exp Bot 58:2369–2387

Hörtensteiner S, Feller U (2002) Nitrogen metabolism and remobilisation during senescence. J Exp Bot 53:927–937

Huang A, Xu S, Cai X (2014) Whole-genome quantitative trait locus mapping reveals major role of epistasis on yield of rice. PLoS One 9(1):e87330. doi:10.1371/journal.pone.008733

Justes E, Mary B, Meynard JM, Machet JM, Thelier-Huche L (1994) Determination of a critical nitrogen dilution curve for winter crops. Ann Bot Lond 74:397–407

Kamran A, Iqbal M, Spaner D (2014) Flowering time wheat (*Triticum aestivum* L.): a key factor for global adaptability. Euphytica 197:1–26

Kant S, Bi YM, Rothstein S (2011) Understanding plant response to nitrogen limitation for the improvement of crop nitrogen use efficiency. J Exp Bot 62:1499–1509

Knyazikhin Y, Schull M, Stenberg P, Mõttus M, Rautiainen M, Yang Y, Marshak A, Latorre Carmona P, Kaufmann R, Lewis P, Disney M, Vanderbilt V, Davis A, Baret F, Jacquemoud S, Lyapustin A, Myneni R (2013) Hyperspectral remote sensing of foliar nitrogen content. PNAS 110:185–192

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:29

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Laperche A, Devienne-Barret F, Maury O, Le Gouis J, Ney B (2006) A simplified conceptual model of carbon/nitrogen functioning for QTL analysis of wheat adaptation to nitrogen deficiency. Theor Appl Genet 113:1131–1146

Laperche A, Brancourt-Hulmel M, Heumez E, Gardet O, Hanocq E, Devienne-Barret F, Le Gouis J (2007) Using genotype × nitrogen interaction variables to evaluate the QTL involved in wheat tolerance to nitrogen constraints. Theor Appl Genet 115:399–415

Larsen R, Marx M (1985) An introduction to probability and its applications. Prentice-Hall Inc, Englewood Cliffs

Le Gouis J, Beghin B, Heumez E, Pluchard P (2000) Genetic differences for nitrogen uptake and nitrogen utilisation efficiencies in winter wheat. Eur J Agron 12:163–173

Le Gouis J, Bordes J, Ravel C, Heumez E, Faure S, Praud S, Galic N, Remoué C, Balfourier F, Allard V, Rousset M (2012) Genome-wide association analysis to identify chromosomal regions determining components of earliness in wheat. Theor Appl Genet 124:597–611

Lorenz AJ, Hamblin MT, Jannink JL (2010) Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. PLoS One 5(11):e14079. doi:10.1371/journal.pone.0014079

Mackay T (2014) Epistasis and quantitative traits: using model organisms to study gene–gene interactions. Nature Rev Genet 15:22–33

Mangin B, Goffinet B, Rebai A (1994) Constructing confidence intervals for QTL location. Genetics 138:1301–1308

Manske GGB, Ortiz-Monasterio IJ, Vlek PLG (2001) Techniques for measuring genetic diversity in roots. In: Application of physiology in wheat breeding. CIMMYT, Mexico, pp 208–218

McCarthy MI, Hirschhorn JN (2008) Genome-wide association studies: potential next steps on a genetic journey. Hum Mol Genet 17:156–165

Muurinen S, Slafer GA, Peltonen-Sainio P (2006) Breeding effects on nitrogen use efficiency of spring cereals under northern conditions. Crop Sci 46:561–568

Neumann K, Kobiljski B, Dencic S, Varshney RK, Börner A (2011) Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). Mol Breed 27:37–58

Novoselovic D, Baric M, Drezner G, Gunjaca J, Lalic A (2004) Quantitative inheritance of some wheat plant traits. Genet Mol Biol 27:92–98

Opsahl T, Agneessens F, Skvoretz J (2010) Node centrality in weighted networks: generalizing degree and shortest paths. Soc Netw 32:245–251

Ortiz-Monasterio I, Sayre KD, Rajaram S, McMahon M (1997) Genetic progress in wheat yield and nitrogen use efficiency under four N rates. Crop Sci 37:898–904

Pathak RR, Lochab S, Raghuram N (2011) Plant systems | improving plant nitrogen-use efficiency. In: Comprehensive biotechnology, 2nd edn. Elsevier, Amsterdam, pp 209–218

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. Plos Genet 2:e190

Payne P (1987) Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. Am Rev Plant Physiol 38:141–153

Philipps PC (2008) Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet 9:855–867

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

Quarrie SA, Steed A, Calestani C, Semikhodskii A, Lebreton C, Chinoy C, Steele N, Pljevljakusic D, Waterman E, Weyen J, Schondelmaier J, Habash DZ, Farmer P, Saker L, Clarkson DT, Abugalieva A, Yessimbekova M, Turuspekov Y, Abugalieva S, Tuberosa R, Sanguineti M-C, Hollington PA, Aragues R, Royo A, Dodig D (2005) A high density genetic map of hexaploid wheat (*Triticum aestivum* L.) from the cross Chinese Spring × SQ1 and its use to compare QTL for grain yield across a range of environments. Theor Appl Genet 110:865–880

Quraishi UM, Abrouk M, Murat F (2011) Cross-genome map based dissection of a nitrogen use efficiency ortho-meta QTL in bread wheat unravels concerted cereal genome evolution. Plant J 65:745–756

Reynolds M, Manes Y, Izanloo A, Langridge P (2009) Phenotyping approaches for physiological breeding and gene discovery in wheat. An Appl Biol 155:309–320

Rincent R, Moreau L, Monod H, Kuhn E, Melchinger AE, Malvar RA, Moreno-Gonzalez J, Nicolas S, Madur D, Combes V, Dumas F, Altmann T, Brunel D, Ouzunova M, Flament P, Dubreuil P, Charcosset A, Mary-Huard T (2014) Recovering power in association mapping panels with variable levels of linkage disequilibrium. Genet 197:375–387

Rogers JS (1972) Measures of genetic similarity and genetic distances. Studies in Genetics. University of Texas Publication 7213, pp 145–153

Rothstein S (2007) Returning to our roots: making plant biology research relevant to future challenges in agriculture. Plant Cell 19:2695–2699

Somers D, Isaac P, Edwards K (2004) A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Theor Appl Genet 109:1105–1114

Sourdille P, Gandon B, Chiquet V, Nicot N, Somers D, Murigneux A, Bernard M (2004) Wheat génoplante SSR mapping data release: a new set of markers and comprehensive genetic and physical mapping data. Funct Integr Genomics 4:12–25

Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. Sci 327:818–822

Tian F, Bradbury PJ, Briwn PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, Mcmullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of maize identifies genes affecting leaf architecture. Nat Genet 43:159–162

Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J (2006) A NAC gene regulating senescence improves grain protein, zinc and irons content in wheat. Science 314:1298–1300

Visscher PM, Thompson R, Haley CS (1996) Confidence intervals in QTL mapping by bootstrapping. Genetics 143:1013–1020

Wang S, Debbie W, Forrest K, Allen A et al (2014) Characterization of polyploidy wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. Plant Biotech J. doi:10.1111/pbi.12183

Wu X, Wang Z, Chang X, Jing R (2010) Genetic dissection of the developmental behaviours of plant height in wheat under diverse water regimes. J Exp Bot 61:2923–2937

Würschum T, Langer S, Longin CFH, Korzun V, Akhunov E, Ebmeyer E, Schachschneider R, Kazman E, Schacht J, Reif JC (2013) Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. Theor Appl Genet 126:1477–1486

Xu Y, Wang R, Tong Y, Zhao H, Xie Q, Liu D, Zhang A, Li B, Xu H, An D (2013) Mapping QTL for yield and nitrogen-related traits in wheat: influence of nitrogen and phosphorus fertilization on QTL expression. Theor Appl Genet 127:59–72

Yu JM, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zhang K, Tian J, Zhao L, Wang S (2008) Mapping QTL with epistatic effects and QTL × environment interactions for plant height

using a doubled haploid population in cultivated wheat. J Genet Genomics 35:119–127

Zhang LY, Liu DC, Guo XL, Yang WL, Sun JZ, Wang DW, Zhang A (2010) Genomic distribution of quantitative trait loci for yield and yield-related traits in common wheat. J Integr Plant Biol 52:996–1007

Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML et al (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat Comm 2:467–477