Identification of QTL associated with nitrogen metabolism in a maize (*Zea mays L. ssp. mays*) testcross population

by

Ignacio Trucillo Silva

A dissertation submitted to the graduate faculty in

partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Plant Breeding

Program of Study Committee:

Michael Lee, Major Professor
Dianne Cook
Thomas Lübberstedt
Maria G. Salas-Fernandez
Erik Vollbrecht

Iowa State University
Ames, Iowa

2015

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

ABSTRACT

Maize is a widely cultivated crop in the world and its production relies heavily on nitrogen (N) fertilization. N is an essential mineral nutrient for plant growth and development. However, during the last decades excessive quantities of N have been applied by farmers, a surplus to what maize plants can uptake, and several problems have arisen, such as pollution of the ecosystem and an economic loss to farmers. Breeding maize hybrids that are more efficient in the use of N is a long term goal for plant breeders. Nonetheless, previous to breeding, the genetic basis of N-metabolism in maize would need to be elucidated. Herein, maize testcrosses (TC), derived from the IBMSyn10-DH crossed by an elite inbred, were: 1) Grown in hydroponic condition and several physiological traits related to N-metabolism were assessed on leaf and root tissues. After performing statistical analyses, quantitative trait loci (QTL) were identified; 2) Grown in field conditions under low and high N, several agronomic traits were determined, and statistical and QTL analyses were implemented.

A novel statistical approach was implemented to differentiate experimental errors from truthful phenotypic records in order to remove them for further genetic analysis. This automated method for outlier determination helped to focus the analysis on real data and obtain more reliable QTL mapping results.

Several QTL associated with N-metabolism were determined and numerous candidate genes underlying QTL genomic regions are proposed for further analysis. At least one rich QTL region, presenting three or more overlapping confidence intervals for QTL, were

determined at each of the ten chromosomes. These genomic regions may be valuable in the determination of N-metabolism in maize TC.

CHAPTER 1: GENERAL INTRODUCTION

Maize (*Zea mays L.*) is a widely cultivated crop and provides an abundant source of food, feedstock, biofuel and components included in several industrial products. Maize production is strongly dependent on Nitrogen (N) fertilization due to its importance as a mineral nutrient for plant growth and development. Indeed, the doubling of food production worldwide during the past 40 years has been associated with a 7-fold increase in the use of N-fertilizers (Hirel et al., 2007). From a productive perspective, sufficient N is required for amino acid metabolism, ear growth, and dry matter accumulation in maize kernels (Hirel et al., 2001). On the contrary, N deficiency adversely affects kernel number, dry matter accumulation and could result in a 14–80% decrease in grain yield (Uhart and Andrade, 1995).

Even though N-fertilization is a necessity for maize production, the intensification in the use of N-fertilizers generates several detrimental impacts, including extensive pollution of primary natural resources and numerous related economic issues. It is noteworthy that on average only 33-50% of the nitrate applied to the soil is accessed by cereal crops (Raun and Johnson, 1999) while the excess may be denitrified by soil bacteria (e.g. nitrate), volatilized (e.g. surface-applied urea-based fertilizers) (Nielsen, 2006), and, to greatest extent, lost by N-leaching. Indeed, N-leaching from the Mississippi River Basin is one of the main causes for the expanding hypoxic dead zone that develops each year on the Louisiana-Texas shelf of the Gulf of Mexico (Goolsby and Battaglin, 2000). Nitrate concentrations have increased several fold during the past 100 years in streams of the basin, and the annual delivery of nitrate from the Mississippi River to the Gulf has nearly tripled since the late 1950's. Furthermore, in the

state of Iowa it has been estimated that N is the second highest cost in maize production (after seed cost) and producers typically invest more than US$ 1B every year in N-fertilizers. Therefore, if 50-67% of the applied N is not utilized, that results in a total loss of US$ 568-761 M every year, just for Iowa (Iowa State University, Extension and Outreach, 2015).

Plant breeding programs were generally focused on selecting for high-yielding genotypes under high N-input systems. As a result, from 1961 to 2006 the amount of N fertilizers applied to agricultural crops increased by 7.4 fold, whereas the overall yield increase was only 2.4 fold. This implies that efficiency in the use of N has sharply declined (Hirel *et al.*, 2011). Thus, even though artificial selection could lead to cultivars showing high performance under high N- fertilization, those genotypes may be not the most efficient in the use of N.

Breeding maize with increase efficiency in the use of N may lead to a reduction of the annual inputs of N fertilizer, rendering a more sustainable agriculture while maintaining yields and concomitant profits. But, before breeding, a further understanding of some aspects of N metabolism and their genetic determination may be needed. Because of the major importance of the elucidation of the genetics underlying N-metabolism for the maintenance of a sustainable and profitable agriculture, maize breeders have been working on methods to determine the genotype-phenotype relationships and considerable genomic research on N-metabolism in maize has been and is currently conducted (*).

Much of today's commercial maize germplasm originates from seven progenitor lines, including B73 and Mo17 (Mikel and Dudley, 2006). Both inbreds differ in their response to N fertilization (Balko and Russell, 1980) and are the parents of the IBM

population (Lee et al., 2002). A total of 360 double haploid (DH) lines were generated from the IBMSyn10 population (Hussain et al., 2007), and 176 DH lines have been crossed to an elite inbred (property of DuPont Pioneer, closed pedigree). This high-resolution mapping population can be directly associated to the physical map established for B73 (www.maizesequence.org). Hence, this maize TC population may serve as an ideal resource for performing N-metabolism genetic studies.

Literature Review

Nitrogen metabolism

Plants N-metabolism is complex, influenced by the interplay of many physiological processes including signaling and regulatory pathways that integrate plant N-status and plant growth (Moose and Below, 2009). Several genomic regions are involved in the genetic control of N uptake by roots, translocation to leaves and remobilization from stalk and leaves to finally reach the grain. Furthermore, N-metabolism interacts directly, and is interconnected, with other biological pathways, such as carbon (C) and Phosphorus (P) metabolism. Photosynthesis occurs primarily in the source leaf and is the process were C is fixed. Simultaneously, N is incorporated into amino acids and proteins, while P mediates the synthesis of RNA and realization of energy (Schlüter et al., 2013).

In many plant species, the management of N, from a physiological perspective, can be divided in two main phases depending on the plant cycle. First, during a vegetative phase, young developing roots and leaves act as sink organs of N and amino acids. Those amino acids are further utilized in the synthesis of enzymes and proteins involved in plant architecture and all the components of the photosynthesis apparatus. Later on, during the

reproductive phase, generally starting after flowering, the pattern is characterized by the protein hydrolysis and remobilization on the N accumulated from roots and shoots to the storage organs (e.g. seeds) (Masclaux et al., 2001). Nevertheless, this arbitrary separation of the plant life cycle may not occur successively and the two phases may take place simultaneously. In maize, 45-65% of the N present in the grain is provided by remobilization from the N already accumulated before silking, while the remaining is obtained from post-silking N uptake (Gallais and Coque, 2005).

N is primarily absorbed from the soil by roots as nitrate and, to a lesser extent as ammonium. Several enzymes play a major role in the assimilation of these two inorganic N compounds in higher plants (Yemm and Folkes, 1958; Lea et al., 1990; Lea and Azevedo, 2007; Lea and Miflin, 2010) (Fig. 1.1). Nitrate is reduced to nitrite by nitrate reductase (NR) in the cytoplasm, followed by nitrite in the plastids by nitrite reductase (NiR), resulting in ammonium. Ammonium is predominantly assimilated by the action of two enzymes. First, glutamine synthetase (GS) assimilates ammonia into the amide position of glutamine, and later, glutamate synthase (GOGAT) transfers the amide group of glutamine to a C skeleton in the form 2-oxoglutarate, yielding two molecules of glutamate; thus, completing the assimilation of ammonia into amino acids (Lea and Miflin, 2010). The organic acid, 2-oxoglutarate, can be synthesized by isocitrate dehydrogenases and aspartate aminotransferases (AspAT), but the exact enzymatic origin is still unknown (Hodges et al., 2003). Moreover, alanine aminotransferase (AlaAT) catalyzes the transfer of the amino group from glutamate to pyruvate to yield 2-oxoglutarate and alanine. Asparagine synthetase (AS) can catalyze the synthesis of asparagine by amidation of aspartate using either glutamine or ammonium as an amino donor. Asparagine and glutamine are the major N-

transport and storage compounds from source to sink organs in non-leguminous plants (Lea and Ireland, 1999). In addition, phosphoenolpyruvate carboxylase (PEPC), an enzyme corresponding to the primary C-metabolism, is directly related with N-metabolism since assures the provision of C skeletons for amino acid synthesis, through catalyzing the addition of bicarbonate to phosphoenolpyruvate to form the four-C compound oxaloacetate.

Quantitative genetic approaches may facilitate the identification and characterization of genomic regions involved in the genetic variation of N-metabolism in maize. The estimation of metabolites content and enzyme activities coupled with the analysis of variation at the genome level may identify the genetic determinants of N-metabolism. Furthermore, a set of N-responsive agronomical traits, including plant height, ear height, grain yield, and flowering time, showing significant phenotypic and genotype variation have been reported in several investigations further described below (Agrama et al., 1999; Bertin and Gallais, 2000; Bertin and Gallais, 2001; Coque and Gallais, 2006). Hence, an integrated genetic analysis targeting physiology and agronomic traits may increase the knowledge of the number, location, effects, and identities of such genetic loci associated with N-metabolism, possible leading to new biological insights (Broman and Sen, 2009).

DNA markers

DNA markers can reveal sites of variation at the DNA level (Winter and Kahl, 1995; Jones et al., 1997), thus allowing the study of the variability between genotypes, determining reference points within chromosomes, genomic regions of agronomic importance, and may accelerate breeding programs by marked-assisted selection. DNA-based markers can be located in coding or non-coding genomic regions and their functions and sequences could be

known or unknown. Those markers are not affected by the environment, do not vary with the developmental stage of the individual, allow early detection of polymorphisms, are extremely abundant (Collard et al., 2005), and generally require low quantities of DNA for the analysis. Molecular markers used in this study are single-nucleotide polymorphisms (SNPs). SNPs are DNA sequence variations of one nucleotide present in the genome sequence, highly abundant, reproducible and accurate. The use of SNPs as biallelic genetic markers allows the rapid, highly automated genotyping (Wang et al., 1998).

QTL mapping

A Quantitative Trait Loci (QTL) is a region of any genome that is responsible for variation in the phenotypic variance of a quantitative trait (Doerge, 2002). Linkage mapping of QTL allows to experimentally estimating the mean and variance associated with a specific locus. The procedure relies on differences among the trait means of genotypes at a marker locus (Bernardo, 2010). Since molecular markers and specific statistical software became available QTL mapping allowed the routine detection of QTL in plants (Bernardo, 2008).

A general QTL mapping approach in plants involves four major steps. The first step is the crossing between two inbred parental lines, which may, or may not differ for the target traits. The resulting F_1 seed will be completely heterozygous at all markers and QTL. A number or line- cross populations derived from the F_1 can be used for QTL mapping, such an F_2 design, backcross, recombinant inbred lines (RIL), advanced intercross lines (AIL), and even hybrid combinations resulting from crossing a RIL population by a tester line. Each specific segregating population may have its advantages and disadvantages and have been characterized in several publications (Lynch and Walsh, 1998; Bernardo, 2010). The next

stage is the screening or genotyping of each individual of the population using molecular markers. Thirdly, the traits of interest are measured in the individuals of the population, a process called phenotyping. Finally, by statistical means, associations between marker loci and phenotypic variance are assessed. Those regions of the genome that show convincing evidence of association are defined as QTL (Broman and Sen, 2009).

The precision in the identification of a QTL is critical to the time, expense, and probability of success of further studies (e.g. positional cloning) (Remington et al., 2001). The precision of the estimation of the QTL position, referred as resolution, may vary substantially depending on several factors such as recombination frequency, marker density and population size (Yu et al., 2011). A mapping population presenting high recombination frequency, high marker density as well as high population size, may result in a higher mapping resolution. Concomitantly, resulting QTL confidence intervals (typically presented as 1-LOD interval) may be shorter, encompassing a lower amount of candidate genes compared to hundreds of candidate genes when dealing with a low-resolution population with significant linkage blocks across the genome.

Even though several QTL mapping methods are available, most implemented procedures are single-marker analysis (SMA), simple interval mapping (SIM), composite interval mapping (CIM), and multiple interval mapping (MIM). Under SMA, the simplest approach, an analysis of every marker with the trait is performed. Each marker-trait association test is performed independent of information from all other markers. A genetic map, with markers and genetic positions, is not required. However, the estimations of QTL position and effect may not be precise (Lynch and Walsh, 1998). In addition, the size of a QTL effect can be confounded with the distance of the QTL to the nearest marker. SIM

(Lander and Botstein, 1989) uses a genetic map for the location of QTL and the presence of a single QTL is performed in a separate analysis for each pair of adjacent markers (interval). Thus, the most likely position of a QTL and the size of the QTL effects are estimated more precisely compared to SMA. However, the construction of a genetic map is a necessity and results are biased when more than one QTL is present within a marker interval. With regard to CIM (Zeng, 1994), it considers the interval mapping test as in SIM, but incorporates markers significantly associated with the trait elsewhere in the genome in order to reduce background variation (Doerge, 2002; Bernardo, 2010). Those markers, called cofactors, are identified by forward or backward stepwise regression and the number can be selected by the researcher. A few limitations of CIM are that the method requires a genetic map, specialized software, and may require higher computational time. MIM (Kao et al., 1999) procedures builds a multiple-QTL model considering numerous marker intervals simultaneously. It uses a stepwise selection procedure by fitting individual QTL sequentially in the model, searching epistasis between significant QTL and refining of QTL effect and genetic position (Bernardo, 2010).

QTL mapping for nitrogen metabolism related traits in maize

During the past 20 years, maize N-metabolism has been the subject of numerous investigations. Many research efforts have been implemented towards elucidating the genetic basis behind the biological responses related to N-metabolism through QTL mapping analysis.

Agrama et al. (1999) studied a segregating population of 214 F_2 maize genotypes derived from the cross between B73 and G79, tolerant and intolerant under low-N stress

conditions, respectively. The population was genotyped with 185 restriction fragment length polymorphism (RFLP) probes and the traits analyzed were ear-leaf area, plant height, grain yield, number of ear and number of kernels per plant. Between two and six QTL per trait were identified under low and high N levels (i.e. LN and HN, respectively), with more QTL detected under LN.

Likewise, genomic regions associated with grain yield and its components were determined in a subsequent study focused on maize hybrids (i.e. TC) (Bertin and Gallais, 2000). The genotypes were originated from the cross between 99 RILs and a common tester, and those RILs were derived from crossing a French flint line (F_2) and an iodent late line (I_o). That same population was extensively studied in subsequent analysis. The TC materials were grown under LN and HN and genotyped with 152 marker loci. It was concluded that the genetic variability was expressed differentially under different N conditions and a total of 29 QTL were identified. Successively, Hirel et al. (2001) developed a quantitative genetics approach by associating metabolic functions and agronomic traits to DNA markers using information obtained from the previous investigation (Bertin and Gallais, 2000). Coincidences of QTL clustered mainly in chromosome 5 for yield and its components, as well as genes encoding cytosolic GS. The same research group reported that based on the coincidence between previously mapped QTL and genes encoding enzymes involved in N assimilation, NUE can be improved by marker-assisted selection and genetic engineering (Masclaux et al., 2001). Based on the same population, agronomic and physiological traits were used to detect QTL and determine their causal relationships in an integrated manner (Gallais and Hirel, 2004). Information from agronomic traits was gathered from maize hybrids; while physiological traits (nitrate content, NR and leaf GS activities) were studied

among 77 RILs from the same population). In agreement with a previous investigation (Bertin and Gallais, 2000), QTL coincidences with the GS locus served to point out the relevance of GS locus, positioned in chromosome 5, as a candidate gene responsible for phenotypic variation in the use of N. In addition, N-metabolism was studied during kernel germination in 140 F_6 RIL derived from the same population. In total, 152 RFLP markers were employed and nine QTL were detected. Similarly, coincidences were observed between QTL and genes encoding for GS (Limami et al., 2002). Moreover, Coque and Gallais (2008) studied, once again, a set of RIL derived from the population employed by Bertin and Gallais (2000) and related TC genotypes. Coincidences, as well as inconsistencies, in the detection of QTL were identified when comparing analysis on inbreds versus TC genotypes under different N conditions. QTL inconsistencies, especially under LN environment, demonstrate that only a few yield-QTL could be useful for marker-assisted selection (MAS) breeding (Agrama, 2005). Furthermore, the same population was investigated for QTL associated with metabolites, such as asparagine and glutamate, and activity of GDH and GS in the developing ear of maize (Canas et al., 2012). The population was genotyped with 203 genetic markers and co-location with QTL for grain yield determined in previous studies (Hirel et al., 2001; Coque et al., 2008) was identified. In addition, candidate genes associated with the determination of yield were identified, including Gln1.3 (GS locus), Gdh1 (GDH locus) and AS4 (AS locus).

In 2007, a QTL analysis based on a $F_{2:3}$ tropical population grown under LN and HN conditions was published. The traits included yield, plant height, kernel and ear number per plant, anthesis-silking interval, chlorophyll content and fresh weight of 100 kernels. In

general, inconsistencies in the detection of QTL for yield were determined under both N treatments (Ribaut et al., 2007).

Zhang et al. (2010) measured the activity of ten enzymes involved in the C- and N-metabolism of maize, using the IBMSyn4 population. The investigation focused on the analysis of leaf tissues, based on a segregating population with four rounds of intermating previous to the inbreeding process. That inbred population was genotyped with 2,200 genetic markers. The linkage analysis detected 73 QTL associated with enzyme activity and eight QTL associated with biomass. Most of the enzyme-activity QTL were located *in trans* (unlinked or even in a different chromosome) to the known genomic locations of the structural genes but, three *cis*-QTL were determined for NR, glutamate dehydrogenase and shikimate dehydrogenase.

In 2012, an investigation based on a set of 74 introgression lines (IL), derived from the cross of an elite Chinese inbred with diverse donors, including Mo17 and B73, grown under LN and HN was published (Liu et al., 2012). The population was genotyped with 189 simple sequence repeat markers and QTL mapping was performed for grain yield and yield component traits. More QTL were identified under LN compared to HN conditions (33 vs 23). In addition, QTL information (e.g. position, CI) for similar traits was collected from previous publications and integrated into a reference map. Thirty-seven consensus QTL regions were determined (18 under LN and 19 at HN, respectively) with an average CI of 22 cM. Thirteen candidate genes specifically expressed under LN were later identified, and those IL containing candidate genes in the introgressed segments were evaluated to determine the genetic effects of candidate genes.

Recently, Zhang (2015) determined several QTL and candidate genes associated with the accumulation of 12 metabolites directly related to C- and N-metabolism in the maize nested association mapping (NAM) population. An association mapping approach was implemented and 101 candidate genes were identified in the population derived from the cross between 25 genotypes, which represents a broad maize genetic variation, and B73 (as reference genotype).

Despite all the efforts in elucidating the genetics underlying N-metabolism in maize, very few investigations followed an integrative approach including agronomic and physiological traits. A research strategy of studying a representative hybrid-high-resolution population, while targeting physiological and agronomic traits measured at different stages from genotypes grown in hydroponics and field conditions under LN and HN treatment has not yet previously explored. Furthermore, a quantitative analysis for enzymes related to N-metabolism in both root and shoot tissues has not been reported in a population with these features and genotyped with a large amount of marker loci. Thus, a genetic mapping investigation for N-metabolism related traits based on a hybrid IBMSyn10-derived population may be particularly valuable for the scientific community.

Research Objectives

The objectives of this research were to:

- Determine and implement the best statistical method to deal with a raw unbalanced complex-dataset
- Identify genomic regions associated with N-metabolism enzyme activity and enzyme content from maize shoot tissues grown in hydroponics

- Identify genomic regions associated with N-metabolism enzyme activity and enzyme content from maize root tissues grown in hydroponics
- Identify genomic regions associated with N-metabolism related traits in maize grown in the field under low and high N conditions
- Identify candidate genes within QTL regions associated with N-metabolism

Dissertation Organization

This dissertation aims to identify QTL associated with N-metabolism related traits in a maize testcross (TC) population. The segregating population was generated from the cross between IBMSyn10DH lines and an elite inbred line provided by DuPont Pioneer. The IBMSyn10DH lines are derived from the cross between B73 and Mo17, and went through ten rounds of intermating before making doubled haploid lines. The resulting TC genotypes were planted in hydroponics and in field trials and the performance of several traits were measured. A general introduction addressing the importance of the theme, motivation, and the general methods used in this research are presented in the first chapter of the thesis dissertation.

The second chapter focuses on the statistical method used for the determination of outliers within the raw data. A novel approach, based on a jackknife resampling strategy, is described. Basically, a statistical model is fitted n times, systematically omitting one observation from the dataset, followed by the prediction of random effects each of the n times, with the aim of targeting “real outliers” based on the complete information gathered in

the experiment. In addition, the r code used in the analysis is provided within the supplementary information.

The third chapter focuses on the determination of QTL associated with N-metabolism related enzymes and metabolites from leaf tissues. The TC maize population was grown in hydroponics and leaf tissues were harvested at V4 vegetative stage (Abendroth et al., 2011). Leaf tissues from each genotype, in six replications, were subject to enzyme analysis and certain metabolites concentrations were determined. With the information gathered in the experiment, plus molecular markers information on 5,300 SNP, QTL mapping was implemented.

Chapter 4 addresses the identification of QTL associated with N-metabolism associated enzymes and metabolites from root tissues. As described in the previous chapter, a maize hybrid population was grown in hydroponics and root tissues were harvested from each of six replications of every genotype and QTL mapping was executed.

In the fifth chapter of the thesis, the analysis of QTL for N-metabolism related agronomic and physiological traits from a maize TC population grown in field experiments under LN and HN conditions is presented. A total of 176 hybrid genotypes were grown in field conditions and traits including plant height, ear height, grain yield, N leaf concentration at three stages, and flowering time were gathered and the information was used for QTL mapping analyses.

There are four manuscripts, included in chapters 2, 3, 4 and 5, to be submitted to different peer-reviewed journals, including Plant Physiology and Theoretical and Applied Genetics.

Finally, chapter 6 accounts for the overall conclusions of the investigation. A systemic investigation, targeting agronomic and physiological traits, measured in hydroponic and field conditions at vegetative and mature stages, respectively, was implemented. The maize population employed is derived from parents widely utilized throughout several commercial breeding programs in the U.S. and was subjected to ten rounds of intermating. The high resolution expected due to extensive recombination, coupled with the high number of molecular marker loci employed, resulted in an unprecedented accuracy for a QTL mapping study. In addition, several candidate genes were identified within QTL regions for further analysis and validation studies.

References

Abendroth LJ, Elmore RW, Boyer MJ, Marlay SK (2011) Corn growth and development. *In*. Extension Publication #PMR-1009, Iowa State University

Agrama HA (2005) Application of molecular markers in breeding for nitrogen use efficiency. Journal of Crop Improvement **15:** 175-211

Agrama HAS, Zakaria AG, Said FB, Tuinstra M (1999) Identification of quantitative trait loci for nitrogen use efficiency in maize. Molecular Breeding **5:** 187-195

Balko LG, Russell WA (1980) Response of maize inbred lines to N-fertilizer. Agronomy Journal **72:** 723-728

Bernardo R (2008) Molecular markers and selection for complex traits in plants: Learning from the last 20 years. Crop Science **48:** 1649-1664

Bernardo R (2010) Breeding for quantitative traits in plants, Ed 2nd. Stemma Press, Woodbury, MN

Bertin P, Gallais A (2000) Physiological and genetic basis of nitrogen use efficiency in maize. I. Agrophysiological results Maydica **45:** 53-66

Bertin P, Gallais A (2001) Genetic variation for nitrogen use efficiency in a set of recombinant inbred lines II - QTL detection and coincidences. Maydica **46:** 53-68

Broman KW, Sen S (2009) A Guide to QTL mapping with R/qtl. Springer, New York

Canas RA, Quillere I, Gallais A, Hirel B (2012) Can genetic variability for nitrogen metabolism in the developing ear of maize be exploited to improve yield? New Phytologist **194:** 440-452

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica **142:** 169-196

Coque M, Gallais A (2006) Genomic regions involved in response to grain yield selection at high and low nitrogen fertilization in maize. Theoretical and Applied Genetics **112:** 1205-1220

Coque M, Gallais A (2008) Genetic variation for N-remobilization and postsilking N-uptake in a set of maize recombinant inbred lines. 2. Line per se performance and comparison with testcross performance. Maydica **53:** 29-38

Coque M, Martin A, Veyrieras JB, Hirel B, Gallais A (2008) Genetic variation for N-remobilization and postsilking N-uptake in a set of maize recombinant inbred lines. 3. QTL detection and coincidences. Theoretical and Applied Genetics **117:** 729-747

Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. Nature Reviews Genetics **3:** 43-52

Gallais A, Coque M (2005) Genetic variation and selection for nitrogen use efficiency: a synthesis. Maydica **50:** 531-547

Gallais A, Hirel B (2004) An approach to the genetics of nitrogen use efficiency in maize. Journal of Experimental Botany **55:** 295-306

Goolsby DA, Battaglin WA (2000) Nitrogen in the Mississippi basin: Estimating sources and predicting flux to the Gulf of Mexico. *In* U.S. Geological Survey Fact Sheet, FS-135-00. U.S. Geological Survey, Reston, Virginia

Hirel B, Bertin P, Quillere I, Bourdoncle W, Attagnant C, Dellay C, Gouy A, Cadiou S, Retailliau C, Falque M, Gallais A (2001) Towards a better understanding of the genetic and physiological basis for nitrogen use efficiency in maize. Plant Physiology **125:** 1258-1270

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

Hodges M, Flesch V, Galvez S, Bismuth E (2003) Higher plant NADP(+)-dependent isocitrate dehydrogenases, ammonium assimilation and NADPH production. Plant Physiology and Biochemistry **41:** 577-585

Hussain T, Tausend P, Graham G, Ho J (2007) Registration of IBM2 SYN10 Doubled Haploid Mapping Population of Maize. Journal of Plant Registrations **1:** 81-81

Jones N, Ougham H, Thomas H (1997) Markers and mapping: we are all geneticists now. New Phytologist **137:** 165-177

Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics **152:** 1203-1216

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

Lea PJ, Azevedo RA (2007) Nitrogen use efficiency. 2. Amino acid metabolism. Annals of Applied Biology **151:** 269-275

Lea PJ, Blackwell RD, Chen FL, Hecht U (1990) Enzymes of ammonia assimilation. *In* PJ Lea, ed, Methods in Plant Biochemistry, Vol 3. Academic Press Limited, London, pp 257-276

Lea PJ, Ireland RJ (1999) Nitrogen metabolism in higher plants. *In* BK Singh, ed, Plant Amino Acids. Marcel Dekker, Inc, New York, pp 1-47

Lea PJ, Miflin BJ (2010) Nitrogen assimilation and its relevance to crop improvement. *In* CH Foyer, H Zhang, eds, Nitrogen metabolism in plants in the post-genomic era, Vol 42. Willey-Blackwell, Chichester, UK, pp 1-40

Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A (2002) Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant Molecular Biology **48:** 453-461

Limami AM, Rouillon C, Glevarec G, Gallais A, Hirel B (2002) Genetic and physiological analysis of germination efficiency in maize in relation to nitrogen metabolism reveals the importance of cytosolic glutamine synthetase. Plant Physiology **130:** 1860-1870

Liu R, Zhang H, Zhao P, Zhang Z, Liang W, Tian Z, Zheng Y (2012) Mining of candidate maize genes for nitrogen use efficiency by integrating gene expression and QTL data. Plant Molecular Biology Reporter **30:** 297-308

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits, Ed 1st. Sinauer Associates, Sunderland, MA

Masclaux C, Quillere I, Gallais A, Hirel B (2001) The challenge of remobilisation in plant nitrogen economy. A survey of physio-agronomic and molecular approaches. Annals of Applied Biology **138:** 69-81

Mikel MA, Dudley JW (2006) Evolution of north American dent corn from public to proprietary germplasm. Crop Science **46:** 1193-1205

Moose S, Below FE (2009) Biotechnology approaches to improving maize nitrogen use efficiency. Biotechnology in Agriculture and Forestry **63:** 65-77

Nielsen RL (2006) N loss mechanisms and nitrogen use efficiency. *In*, Ed Nitrogen management workshops. Purdue Agriculture Agronomy Extension, West Lafayette, IN

Raun WR, Johnson GV (1999) Improving nitrogen use efficiency for cereal production. Agronomy Journal **91:** 357-363

Remington DL, Ungerer MC, Purugganan MD (2001) Map-based cloning of quantitative trait loci: progress and prospects. Genetics Research **78:** 213-218

Ribaut J-M, Fracheboud Y, Monneveux P, Banziger M, Vargas M, Jiang C (2007) Quantitative trait loci for yield and correlated traits under high and low soil nitrogen conditions in tropical maize. Molecular Breeding **20:** 15-29

Schlüter U, Colmsee C, Scholz U, Braeutigam A, Weber APM, Zellerhoff N, Bucher M, Fahnenstich H, Sonnewald U (2013) Adaptation of maize source leaf metabolism to stress related disturbances in carbon, nitrogen and phosphorus balance. BMC genomics **14:** 442

Uhart SA, Andrade FH (1995) Nitrogen deficiency in maize: I. Effects on crop growth, development, dry matter partitioning, and kernel set. Crop Science **35:** 1376-1383

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen NP, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, Lander ES (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science **280:** 1077-1082

Winter P, Kahl G (1995) Molecular marker technologies for plant improvement. World Journal of Microbiology & Biotechnology **11:** 438-448

Yemm W, Folkes BF (1958) The metabolism of aminoacids and proteins in plants. Annual Review of Plant Physiology **9:** 245-280

Yu H, Xie W, Wang J, Xing Y, Xu C, Li X, Xiao J, Zhang Q (2011) Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. Plos One **6:** e17595

Zeng ZB (1994) A composite interval mapping method for locating multiple QTLs. *In* 5th World Congress on Genetics Applied to Livestock Production, Vol 21, Guelph, Ontario, Canada, pp 37-40

Zhang N, Gibon Y, Gur A, Chen C, Lepak N, Hoehne M, Zhang Z, Kroon D, Tschoep H, Stitt M, Buckler E (2010) Fine quantitative trait loci mapping of carbon and nitrogen metabolism enzyme activities and seedling biomass in the maize IBM mapping population. Plant Physiology **154:** 1753-1765

Zhang N, Gibon Y, Wallace JG, Kruger Lepak N, Li P, Dedow L (2015) Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population. Plant Physiology **168:** 575-583

(*) www.cimmyt.org/en/component/docman/doc_view/120-executive-summary-improved- maize-for-african-soils

www. nitrogenes.cropsci.illinois.edu/projects.cfm

Figures

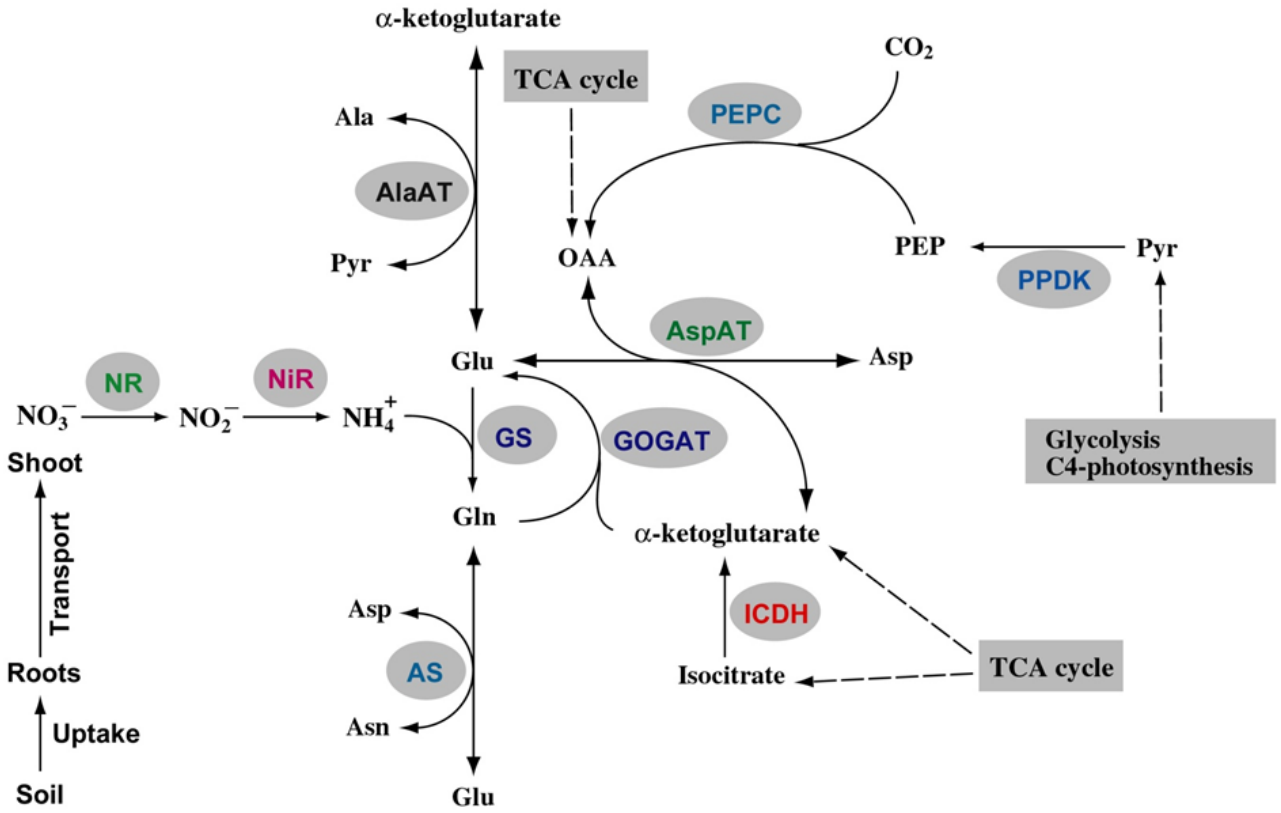

Figure 1.1. Main reactions involved in N-acquisition and assimilation in higher plants.

AlaAT, alanine aminotransferase; AS, asparagine synthase; AspAT, aspartate aminotransferase; GOGAT, glutamate synthase; GS, glutamine synthetase; ICDH, isocitrate dehydrogenase; NR, nitrate reductase; NiR, nitrite reductase; PEPC, phosphoenolpyruvate carboxylase; PPDK, pyruvate orthophosphate dikinase (Prepared by Kanwarpal S. Dhugga, 2015).

CHAPTER 2: A METHODICAL APPROACH FOR IDENTIFYING OUTLIERS IN COMPLEX DATA

A paper to be submitted to Agronomy Journal

Ignacio Trucillo-Silva[1], Dianne Cook[2], Michael Lee[1]

[1] Department of Agronomy, Iowa State University, Ames, IA 50010, USA.

[2] Department of Econometrics and Business Statistics, Monash University, Clayton, VIC 3800, Australia.

Abstract

Analysis of raw datasets can become tedious and laborious, leading researchers to launch directly into the statistical analysis with a routine analysis without carefully checking the quality of the data. This can result in the failure to find problems with the data or in the direct removal of valid data, which might overly influence the final results of the investigation. Statistical analysis of raw data has received considerably less emphasis than the subsequent genetic analysis, even though good data are well known to be essential and the foundation for any successful investigation. Here we describe an approach that can be used with readily available tools, to check the quality of a complex biological data set collected using a careful experimental design, and to determine which observations might be dismissed from the analysis. The approach entails five different steps using R, where observations not consistent with the rest of the dataset can be discarded, by an iterative jackknife process by targeting and removing those genotypes which generated outliers and re-fitting a statistical model. Improvements in the log-likelihood values, on the order of 200 units in magnitude, were achieved by removing just a few genotypes (three to eight).

Key words: statistics – jackknife – outliers – data mining – exploratory data analysis

Abbreviations: QTL, quantitative trait locus; N, nitrogen; TC, testcross; PEPC, phosphoenolpyruvate carboxylase; Ala AT, alanine aminotransferase; NiR, nitrite reductase; BLUP, best linear unbiased estimation; AIC, Akaike information criterion; BIC, Bayesian information criterion.

Introduction

Analysis of large datasets generated by intricate experimental designs is a great challenge and an emergent reality for biological researchers, especially in situations where the aim of the investigation is focused on a step ahead of the statistical analysis, such as mapping genes and quantitative trait locus (QTL). Generally, big datasets are complex and daunting, and researchers frequently launch directly into the statistical analysis with a routine analysis without carefully checking the quality of the data. This can result in failure to detect problems with the data that overly influence the final results of the investigation. Alternatively, problems with the data can also be missed if the statistical analysis is performed by a professional statistician, not involved with the data collection, or aware of the downstream objectives. Sometimes, the biologist might rely on specific software available for the analysis, which may not make it straightforward to also determine misleading data and exclude them from the analysis, or in other words proceed with the data cleaning with a statistical basis (i.e. statistical cleaning of data) (Hellerstein, 2008). The risk of stumbling into a serious pitfall increases dramatically with complex data sets, and no professional would want to spend a humongous amount of time in the analysis of raw data when the final objective is to obtain results and eventually write a manuscript describing the findings.

Statistical cleaning of raw data has received considerably less emphasis than the subsequent genetic analysis, even though good data are well known to be essential and the

foundation for any successful investigation. This paper describes an approach that can be used with readily available tools to check the quality of a large, complex biological data set collected using a careful experimental design. The principles follow those described in Tukey (1965) of exploratory data analysis, using modern technology.

A main goal of the approach is to optimize the use of the available data, by identifying overly influential values. Plots of the data play an important role in association with computationally intensive calculations. This approach should be applicable to many other types of biological data.

Materials and Methods

Data description

The data set used in this study comprises estimations of activity of eight Nitrogen (N) metabolism-related enzymes in 176 maize (*Zea mays* L.) testcross (TC) genotypes. Maize plants were grown in hydroponic conditions in an incomplete block design. Twenty genotypes were included in each hydroponic tank or set (incomplete block) with 12 replications. In a total of ten sets, 200 genotypes were initially planted. Each replication was arranged to take into account some variability in light. Furthermore, in each of the sets, two checks were included in each of the 12 replications (Fig. 2.1). These two checks were genotypes that were parental sources of the population in their TC genotype. They were designed to be used for calibration purposes. When maize plants reached V4 stage (Abendroth et al., 2011), both leaf and root tissues were harvested. Based on previous experience (K. Dhugga, personal communication, 2012), half of the replications were discarded due to lack of uniformity within plants of same genotype. Finally, plant tissues

from the six replications in each set were subdivided into two reaction plates to perform biochemical measurements. The enzymes studied, or response variables, included alanine aminotransferase (Ala AT), asparagine aminotransferase, asparagine synthetase, nitrate reductase, nitrite reductase (NiR), glutamine synthase, glutamate synthase, and phophoenol pyruvate carboxylase (PEPC).

Even though 200 genotypes were initially planted, 24 genotypes were excluded from the analysis a priori, based on poor genotypic information attributable to low DNA sampling quality due to DNA contamination. The resulting data set consisted of the activity of eight enzymes in 176 genotypes, replicated six times.

Statistical computing software

All analysis were performed in R (RCoreTeam, 2014) and several packages were used including plyr (Wickham, 2011), reshape (Wickham, 2007), ggplot2 (Wickham, 2009) , GGally (Emerson, Green, et al., 2013, Schloerke, Crowley, et al., 2014), ASReml (Butler, Cullis, et al., 2007), and asremlPlus (Brien, 2014).

Computing resource needed

All calculations were executed with a personal laptop with 2nd Gen Intel® CoreTM i5-2430M processor and 8GB DDR3 memory. In general, the system requirements would depend on the dimensions of the dataset and the calculations required for model fitting during the Jackknife step. The ASReml algorithm, based on the R package version, could be considered one of the fastest options.

Results and Discussion

Statistical cleaning

The approach is divided into five main steps consisting of visual inspection of the data, studying relationships between multiple response variables (enzyme activity), fitting statistical models, filtering genotypes (subsetting the data) and, finally, filtering influential measurements based on a Jackknife approach. These are described below.

Step 1: Visual inspection of the data

In this step, basic plots of the data are generated to examine the responses for different aspects of the experimental design. Based on the experimental design, we would expect the enzyme activity performance of checks to be similar across replicates of the same tank, even if they differ between replicates of different tanks. The main idea of this phase is to search for any data structure, in terms of statistical dependence or independence between measurements, or for problems that may affect model predictions in order to have a better grasp of the data set to fit an appropriate statistical model. Checks can be graphically displayed across hydroponic tanks, replications and biochemical plates for each different response variable.

In this particular dataset, non-uniform values for the response variables between replications of same tank were observed for the checks (Fig. 2.2). Furthermore, uniform values were observed between reaction plates but values between sets were substantially different. Originally, the strategy was to use the checks as covariates in the statistical model in order to remove measurement error in the tanks and plates, but check performance was too variable to achieve an appropriate calibration in each incomplete block or tank. In addition,

the variance observed suggests that it might be important to allow for non-constant variance across sets in the statistical model.

Step 2: Identification of relationships and patterns between enzymes

A scatterplot matrix of all variables was an effective means to examine the association between the response variables (e.g. ggpairs in the GGally R package). For a small number of variables, this is a good choice of plot, because it is possible to layout all pairs of variables in a reasonable space. Nevertheless, as in this experiment there are eight response variables, it pushes the limits of the scatterplot matrix because it would require 64 plots to be displayed on a page. Fig. 2.3 shows a selection of these plots. One of the key purposes of making these plots is to decide if a multivariate model or a univariate model would suffice. If the enzyme activities are correlated with each other, a multivariate response model may be the better choice in order to model jointly multiple response variables taking into account the dependency between those variables.

The enzymes PEPC to Ala AT activity (Fig.2.3A and Fig.2.3B) are plotted. Based on all genotypes, a slight pattern could be perceived (weak positive correlation). However, that pattern disappear when each genotype was considered separately, as is the case when analyzing only checks (Fig. 2.3A) or a few random selected genotypes (Fig 2.3B). Similar observations were noticed when analyzing other traits such as Ala AT versus nitrite reductase (NiR) (Fig. 2.3C and Fig. 2.3D). Therefore, as the association between enzyme activities was almost exclusively due to genotype, it is feasible to fit univariate models of enzyme activity on genotype, plate, and set.

Step 3: Fitting a statistical model

A linear mixed effect model was defined and fit using ASReml R package. The response variable is activity of enzyme (nM of substrate converted per gram of plant tissue). Tank (set), the light replicate and, plate were included as fixed effects in the model (where replicate and plate are nested in set), and check genotype effect was included as a continuous covariate. Finally, a random effect for the genotype was included in the linear model. The model can be represented as follows:

$$y = Xb + Zu + e$$

where y denotes an n x 1 vector of observed response values, b is a p x 1 vector of fixed effects (set, light, plate), X is an n x p design matrix, u is a q x 1 vector of random effects (genotypes), Z is a n x q design matrix, e is the error term, and E (u) = 0, E (e) = 0, Cov (u, e) = 0, Var (u) = G, and Var (e) = R. The G matrix had a compound symmetry structure on the genotype levels and R matrix is a diagonal matrix with different values for each set, allowing non-constant variance across sets.

From this model, best linear unbiased predictions (BLUP) for each genotype and best linear unbiased estimation (Henderson, 1975) for each of the checks were obtained and used in a posterior study for the identification of QTL.

Step 4: Filtering genotypes

The purpose of replicates in an experimental design is to estimate the variability among experimental units treated equally and assess the variability in each treatment. The variability in replicates is expected to be relatively small compared to the variability across treatments. This step is to evaluate the variability of replicates by genotype, and begin the model building with a small set of consistent genotypes. Fig. 2.4 shows the mean-centered

Ala AT activity of a sample of genotypes from different sets, sorted in each from least to most variable. Some genotypes presented highly variable enzyme activity, which may cause some problems for the model fitting. From the original 176 genotypes, 32 had very consistent enzyme responses, and these were used for the initial model fitting. Additional genotypes were added later in a stepwise fashion, in conjunction with the jackknife approach described in the next section.

To assess the effect of the data reduction with the full set, correlations between statistics calculated on the reduced set and the complete data were estimated. The Pearson correlation between BLUP values based on the complete data set and predictions based solely on the 32 initially selected genotypes was 0.72, which suggests fairly close agreement in the reduced set to the overall data. Furthermore, variances and correlations of BLUP values were calculated for the selected 32 genotypes across all sets and replications and, separately, for the reduced set of 32 genotypes based solely on most consistent replications and sets. The correlation was as high as 0.90. This suggests that the model fitting for these 32 genotypes is relatively robust and that some specific genotypes are mainly responsible for biasing the predictions.

Step 5: Filtering outliers using a Jackknife approach

In the Jackknife approach (Miller, 1974), the influence of observations is examined by fitting the model without the observation. Here, the model is re-fitted many times, with each genotype excluded once, and the BLUPS for the included data are examined. Fig. 2.5 illustrates the approach for three genotypes a, b, c. The red point corresponds to the BLUP for genotype c when genotype a is left out of the model, and the yellow point corresponds to the BLUP for genotype c when genotype b is left out of the model. The blue dashed line

indicates the leave-one-out BLUPS when excluding from the data set all other genotypes. Both the red and yellow points differ from the other estimates which suggests that genotypes a and b may be influential. An R function was created (provided in supplementary material), to make the calculations. Results of the analysis are shown in Fig. 2.6.

To identify outliers, the 1% trimmed mean (Tukey and McLaughlin, 1963) and standard deviations were calculated in the complete set of estimations for each genotype. Because in a normal distribution 99.7% of the data are within three standard deviations from the mean, estimations greater than three times the trimmed standard deviations were considered outliers.

The filtering process was continued iteratively by targeting and removing those genotypes that generated outliers in the 32 consistent benchmark genotypes and repeating the jackknife function. The objective was to determine an optimal situation where genotypes generating outliers in the BLUPs are discarded from the analysis while keeping as many possible genotypes. The iterations were continued until no more outliers were visible in the plots or in the event that members of the small set of consistent genotypes, the ones determined in the previous step, become the new target to be discarded. Attention was also given to the log-likelihood values when running the full model with the “clean” or simplified dataset versus raw data. In all cases the log-likelihood values improved several orders of magnitude with the refined data.

The jackknife process will essentially identify genotypes that have larger random effects. Recall that mixed effects models tend to estimate random effects that are contracted towards the fixed effects component. If genotypes with larger random effects are discarded, the overall mean and variance of the random effects might change, but they should be more

representative of all of the remaining genotypes because the influential ones, potentially biasing effects away from the majority, have been removed. The final purpose is to identify and remove only genotypes that generate unreliable estimates, and to avoid discarding a genotype that exhibits a real and consistent response. Herein, "unreliable" observations would include data severely affected by experimental error such as those resulting from technical faults during biochemical measurements.

Quantification of the procedure

Improvements in the log-likelihood values, on the order of 200 units in magnitude, were achieved by removing just a few (three to eight) genotypes. Plotting the data before and after each cleaning step was very helpful to visualize the effect, but it is important to use numerical statistics like the log-likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) values, to help quantify the effect of excluding some genotypes on the model fitting. While analyzing Ala AT, eight genotypes were discarded and values of log-likelihood, AIC, and BIC changed from -4278.71, 8579.42, and 8634.46, to -4043.22, 8108.45, and 8162.95, respectively.

Conclusions

The correct application of statistical methods requires careful pre-processing of data in order to obtain valid conclusions. In the literature this has not received substantial attention but it is an extremely critical part of data analysis, and especially important with large and complex datasets. As the experiment gets larger, with more genotypes measured, and more treatment conditions applied, the complexity of the data increases, and may introduce additional problems or errors. To simply throwing everything into a model and

hope to get good results it is not a good idea. Understanding the data set to find problems and address them it is crucial in order to obtain reliable results. Steps similar to the ones used here to preview the enzyme activity data could be applied in several other situations to improve the analysis and related interpretations.

In this research, the results were used in a subsequent QTL analysis, in order to identify regions of the maize genome associated with N-metabolism related enzymes. A better understanding of the genetics underlying N-metabolism will provide insightful information for improving selection in maize.

Acknowledgments

We would like to thank Ignacio Alvarez-Castro for the assistance in setting the mixed model and Lendie Follett for the help in coding the Jackknife function.

References

Abendroth, L.J., R.W. Elmore, M.J. Boyer and S.K. Marlay. 2011. Corn growth and development. Extension Publication #PMR-1009, Iowa State University.

Brien, C. 2014. asremlPlus: collection of functions to augment the use of asreml in fitting mixed models. R package version 1.9-4.

Butler, D., B.R. Cullis, A.R. Gilmour and B.J. Gogel. 2007. Analysis of mixed models for S-language environments: ASReml-R reference Manual. Department of Primary Industries and Fisheries, Queensland, Brisbane, Australia.

Emerson, J.W., W.A. Green, B. Schloerke, J. Crowley, D. Cook, H. Hofmann, et al. 2013. The Generalized Pairs Plot. Journal of Computational and Graphical Statistics 22: 79-91. doi:10.1080/10618600.2012.694762.

Hellerstein, J. 2008. Quantitative data cleaning for large databases. United Nations Economic Commission for Europe, White paper.

Henderson, C.R. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31: 423-447. doi:10.2307/2529430.

Miller, R.G. 1974. Jackknife - A review. Biometrika 61: 1-15.

RCoreTeam. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Schloerke, B., J. Crowley, D. Cook, H. Hofmann, H. Wickham, F. Briatte, et al. 2014. GGally: Extension to ggplot2. R package version 0.4.8, Vienna, Austria.
Tukey, J.W. 1965. The technical tools of statistics. American Statistician 19: 23-28.
Tukey, J.W. and D.H. McLaughlin. 1963. Less vulnerable confidence and significance procedures for location based on a single sample (Trimming/Winsorization I.). Sankhyā, Ser. p. 331-352.
Wickham, H. 2007. Reshaping data with the reshape package. Journal of Statistical Software.
Wickham, H. 2009. ggplot2: elegant graphics for data analysis. Springer, New York.
Wickham, H. 2011. The Split-Apply-Combine Strategy for Data Analysis. Journal of Statistical Software 40: 1-29.

Figures

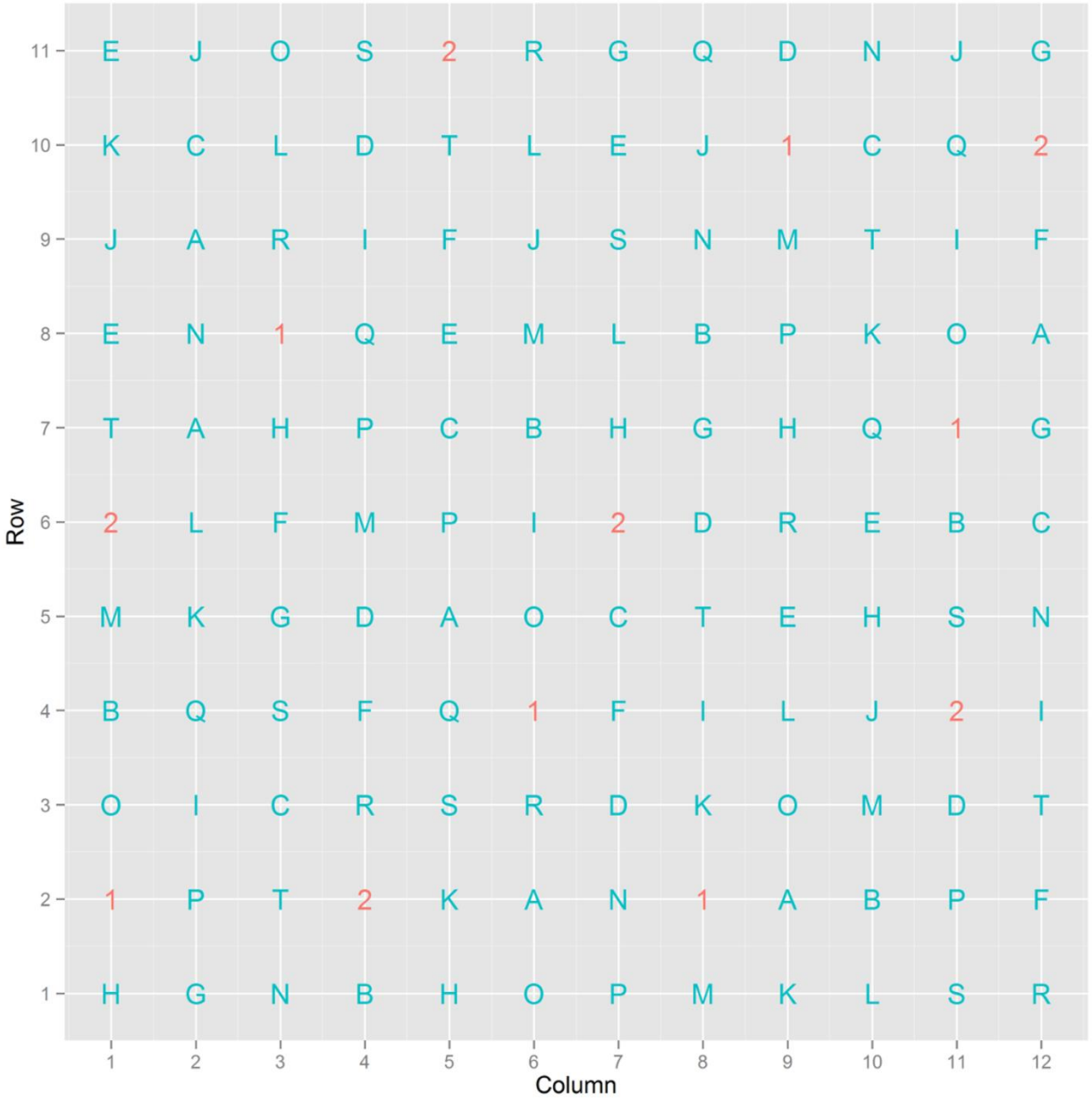

Figure 2.1. Experimental design of a set. Letters (blue) indicate genotypes, and 1, 2 (red) indicate checks.

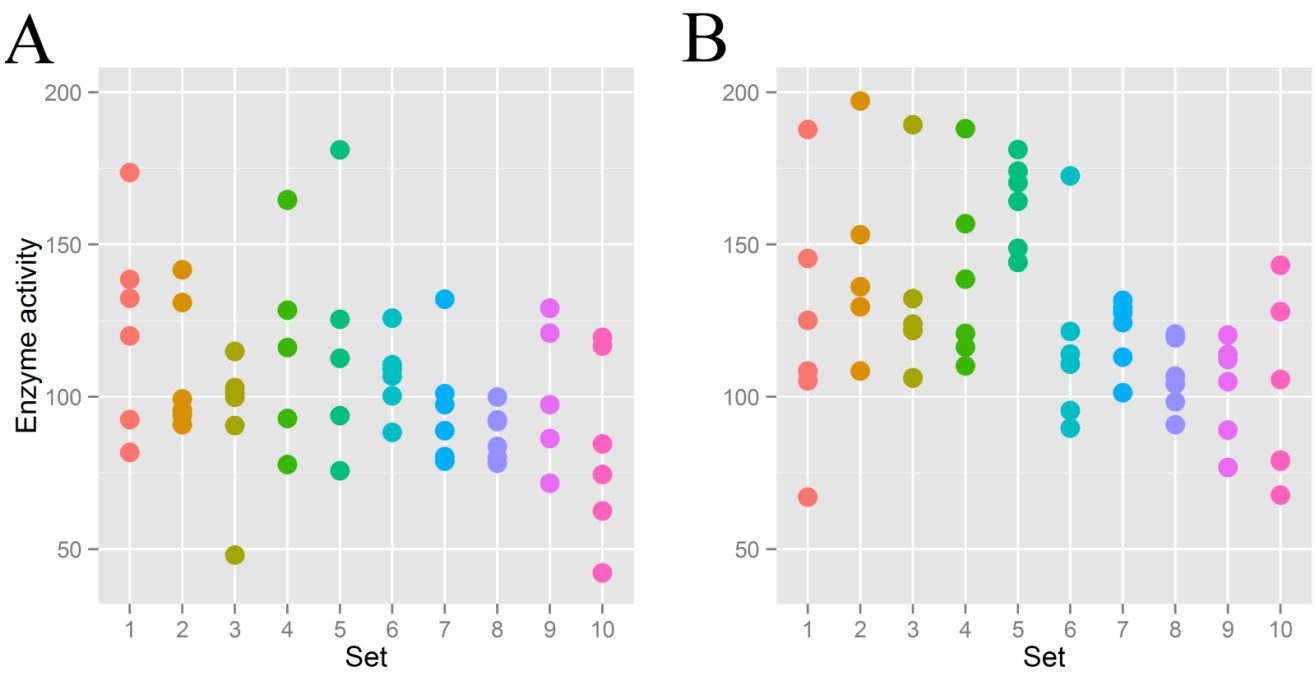

Figure 2.2. Enzyme activity of checks for each set, across incomplete blocks (hydroponic tanks).

(A) Check 1, B) Check 2. Color is redundantly representing set. Scale on the vertical axis is set to be the full range of activity values for all genotypes. The variability of the values in each set for both checks is much larger than it was expected.

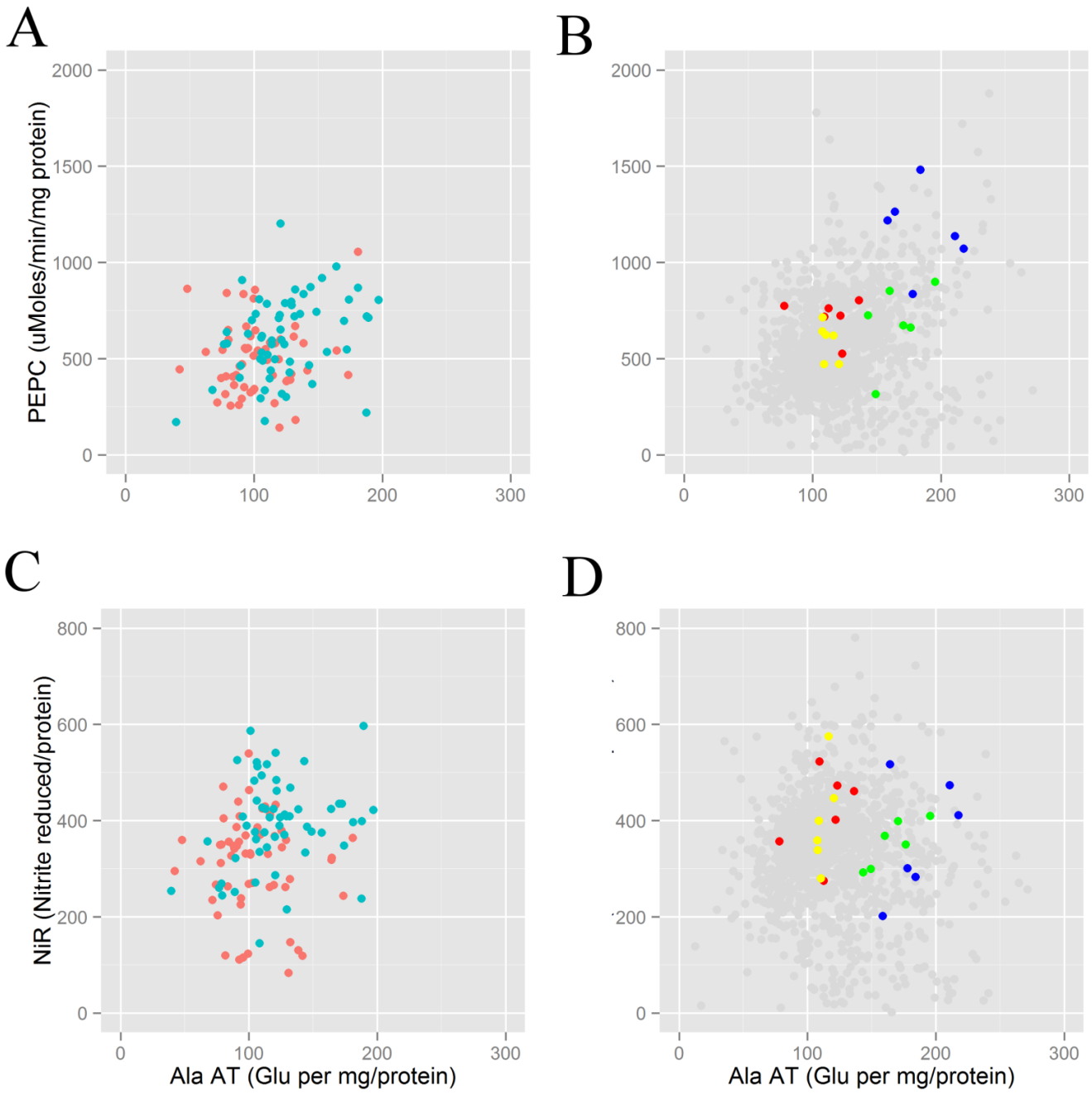

Figure 2.3. Bivariate analysis of enzyme performance (PEPC versus Ala AT (A and B), NiR versus Ala AT (C and D)).

(A), (C) red and green indicate checks 1and 2, respectively. (B), (D) Four random selected genotypes are depicted in red, green, yellow and blue; rest of genotypes in grey.

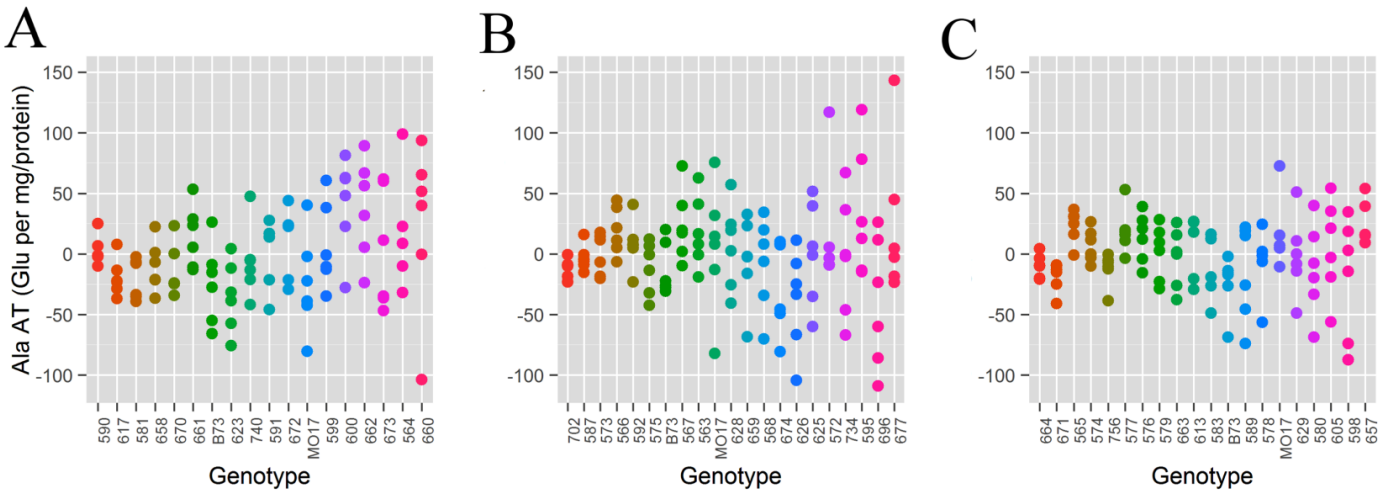

Figure 2.4. Identifying consistent genotypes across different sets: (A) Set 1, (B) Set 2, and (C) Set 3.

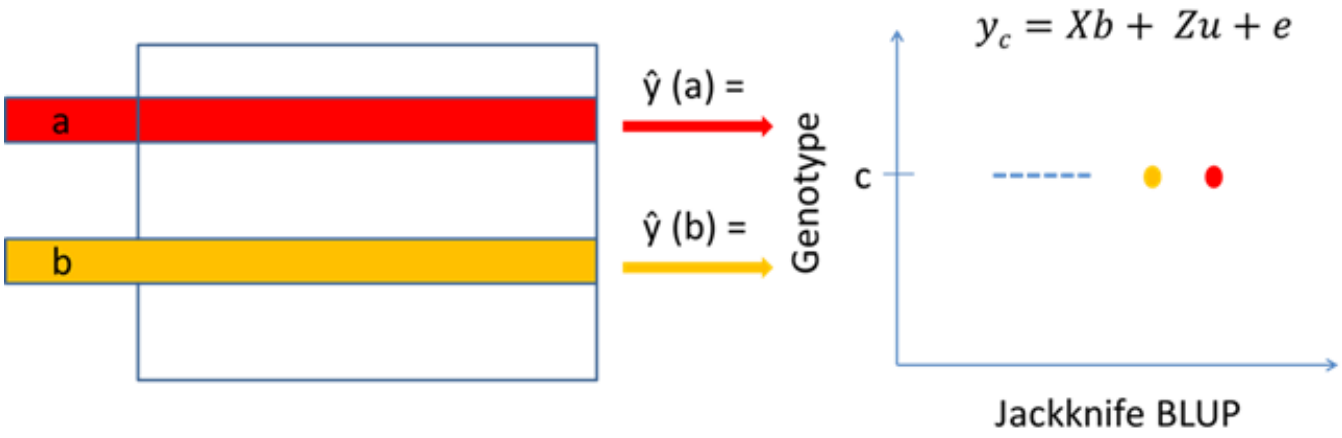

Figure 2.5. Scheme of random effects predictions based on a Jackknife approach.

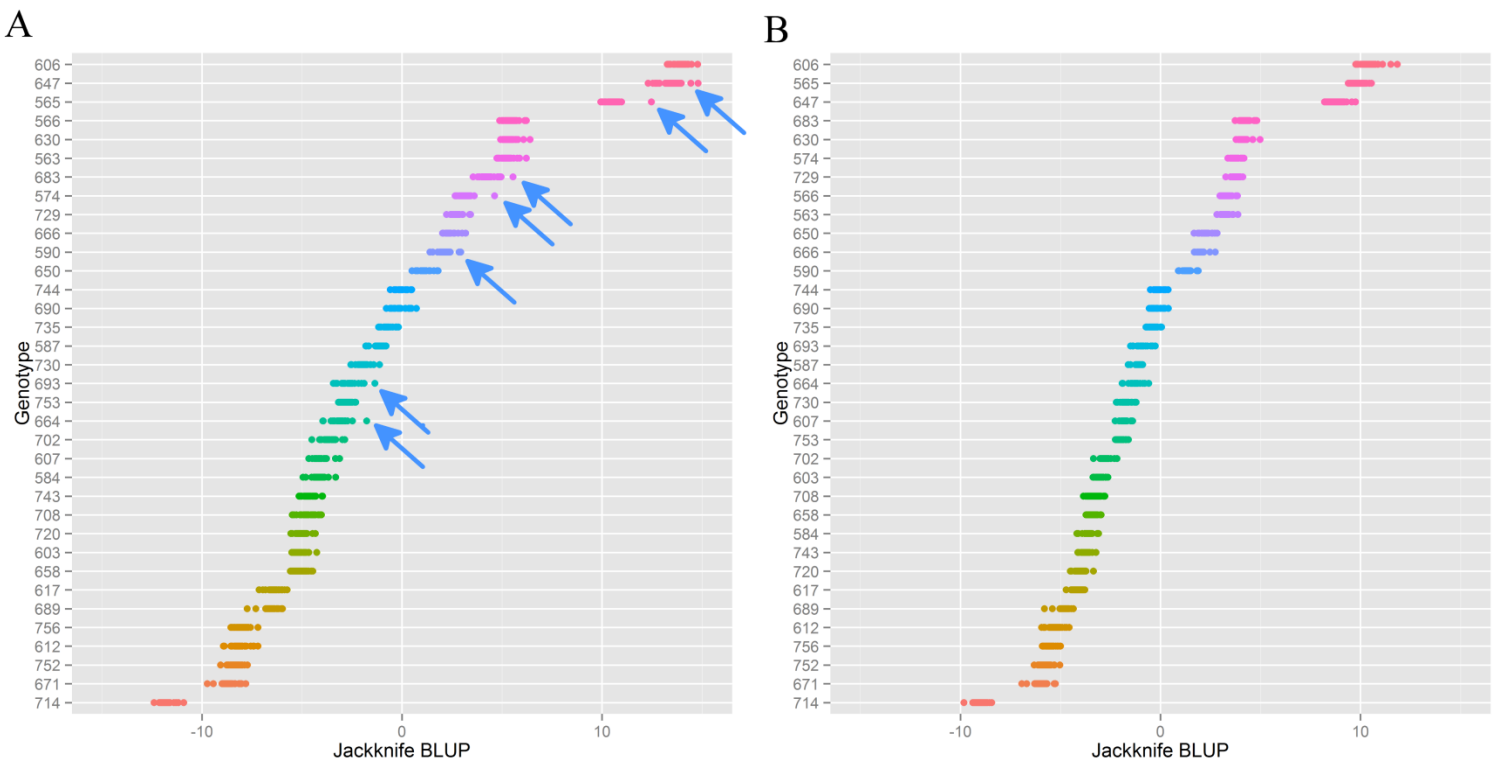

Figure 2.6. Results of cleaning approach for enzyme performance (Ala AT activity).

(A) Prediction of random effects based on raw data, (B) Prediction of random effects after cleaning approach, arrows pinpoint outliers.

APPENDIX A

R code for statistical data cleaning

```
#Loading packages
library(asreml)
library(ggplot2)
library(psych)
library(chemometrics)
library(reshape)
library(dae)
library(asremlPlus)

#Set working directory and open data file
setwd('add.path')
mydata <-  read.csv('filename.csv',header=T)

#Determine more consistent values (response variable: ala.at )
ggplot(data = subset(field,exp==2 & trt==1), aes(x = genotype, y=ala.at, color = genotype))+
geom_point(size=4) + ggtitle("")+ theme(aspect.ratio=1)
ala.at 1.1<-subset(field,exp==2 & trt==1)
sd1<- ddply(ala.at 1.1, .(genotype), summarise, ala.at _sd = sd (ala.at, na.rm=T))
all <- merge(ala.at 1.1,sd1,by="genotype")
all$genotype <-factor(all$genotype, levels=all[order(all$ ala.at _sd), "genotype"])
ggplot(data = all, aes(x = genotype, y= ala.at, color = geno))+
  geom_point(size=4) + ggtitle("")+ theme(aspect.ratio=1)


#Jackknife function
myjackknife <- function(data){
    mylist <- list()
    genos <- unique(data$genotype)
    for (i in genos){
    mydata <- subset(data, genotype != i)
      mymodel <- asreml(ala.at ~ model for your data, data = mydata)
     coefs <- coef(mymodel, pattern = 'test')
      name <- paste('genotype',i, sep = '')
      colnames(coefs) <- name
      mylist[[name]] <- coefs
  }
  return(mylist)
}

# Trim standard deviation functions
outfun <- function(x) {abs(x-mean(x,na.rm=TRUE, trim = .1)) > 3*sd.trim(x,
na.rm=TRUE)}
outfun2 <- function(x) {abs(x-mean(x,na.rm=TRUE, trim = .1)) > 3*winsor.sd(x, trim=0.05,
```

```
na.rm=TRUE)}
outfun3 <- function(x) {  abs(x-mean(x,na.rm=TRUE, trim = .1)) > 3*sd_trim(x, trim=0.05)}

#Run Jackknife function
jk_blups_ <- myjackknife()

#Save results as csv file
setwd("")
write.csv(jk_blups_, file="filename.csv")

#Read results
ala.graph<- read.csv("filename.csv", header=T, stringsAsFactors =  FALSE)

#Preparing the files to make graph
ala.graph[,1]<-gsub("test_","",ala.graph [,1])
colnames(ala.graph)[1] <- 'genotype'
m.ala.graph<-melt(ala.graph, id.vars=c("genotype"))
m.ala.graph$genotype<- as.factor(m.ala.graph$genotype)
m.ala.graph <- m.ala.graph[order(m.ala.graph$genotype, m.ala.graph$variable), ]

#Use the subset of more consistent genotypes (e.g. 35 genotypes)
m.ala.graph_35<-subset()

#Make the graph
jack<-m.ala.graph_35
jack$genotype <- factor(jack$genotype)
jack$variable <- factor(jack$variable)
jack[jack ==0] <- NA
jmedian<- ddply(jack, .(genotype), summarise, ala.at_median = median (value, na.rm=T))
jack <- merge(jack,jmedian,by="genotype")
jack$genotype <-factor(jack$genotype, levels=jack[order(jack$ala.at_median), "genotype"])

ggplot(data = jack, aes(x = value, y=genotype, color = genotype), na.rm=T)+
  geom_point(size=2) + ggtitle("Selected genotypes - Ala AT")+
  theme(plot.title = element_text(lineheight=.8, face="bold")) +
  theme(aspect.ratio=1) + theme(legend.position="") +
  scale_color_discrete(name="")

#Identifying outliers
results <- apply(ala.graph[,-1], 1, outfun)
results1 <- apply(results, 1, sum)
write.csv(results1, file="outliers.csv")
results1<- read.csv("outliers.csv", header=T)
results1[,1]<-gsub("genotype","", results1[,1])
colnames(results1)[1] <- 'genotype'
colnames(results1)[2] <- 'outliers'
```

```
results1$genotype <- as.factor(results1$genotype)
leave.out1<-subset(results1, outliers>10)
out.in.order1<-leave.out1[order(leave.out1$outliers, decreasing=TRUE), ]
out.in.order1

#subset minus genotype in leave.out1
mydata1<-
mydata1<-mydata
mydata1[mydata1$genotype== 709, 'ala.at'] <- NA

#Comparison of  loglikelihoods, AIC, BIC
ala.at_raw <- asreml(ala.at ~ set + set:rep + set:plate + check,  random = ~ test,
                  rcov = ~ at(set):units, data = mydata)

ala.at_clean1 <- asreml(ala.at ~ set + set:rep + set:plate + check, random = ~ test,
                 rcov = ~ at(set):units, data = mydata1)
info.crit.asreml(ala.at_clean1)
info.crit.asreml(ala.at_raw)

#Can run Jackknife function again, clean data and make new graph
```

CHAPTER 3: MAPPING OF QTL FOR N-METABOLISM RELATED ENZYMES AND METABOLITES IN A MAIZE TESTCROSS POPULATION GROWN IN HYDROPONICS: I. LEAF TISSUE ANALYSIS

A paper to be submitted to Plant Physiology

Ignacio Trucillo-Silva[1], Michael Lee[*1], Hari Kishan R. Abbaraju[2], Lynne Fallis[3], Hongjun Liu[4], and Kanwarpal S. Dhugga[5]

[1] Department of Agronomy, Iowa State University, Ames, IA, 50011, USA.
[2] Monsanto Company, St. Louis, MO, 63167, USA.
[3] Trait Discovery & Technology, DuPont Pioneer, Johnston, IA, 50131, USA.
[4] Sichuan Agricultural University, Chengdu, 611130, China.
[5] CIMMYT, Apartado Postal 6-641-06600, Mexico D.F, Mexico.
***** Corresponding author; mlee@iastate.edu.

Abstract

Nitrogen (N) availability is essential for plant growth and development. During last decades, several problems have arisen due to over-fertilization with N in rural areas. Breeding for maize with greater efficiency in the use of N may help to reduce contamination and increase profits. Nevertheless, previous to breeding, a better understanding of the genetics underlying N-metabolism will be needed. Herein, a linkage mapping analysis for N-metabolism related enzymes, metabolites, and proteins was performed based on leaf tissue, harvested from maize hybrids grown in hydroponics. A total of 44 quantitative trait loci (QTL) were identified, all of them located *in trans* compared to the genomic position of the correspondent structural genes. Epistasis between QTL was not significant for most of the

traits. Nevertheless, significant epistasis was determined in two QTL model explaining 2.5-5% of the genetic variance. The QTL models for different traits accounted from 7 to 31% of the genetic variance. Furthermore, 12 coding regions underlying 1-LOD QTL confidence intervals are proposed for further validation studies.

Introduction

Nitrogen (N) is one of the most important mineral nutrients for plant growth and development. N is required for the formation of enzymes and other proteins, for example, signaling and structural proteins. Whereas enzymes carry out metabolism, which produces precursors for plant growth, signaling molecules respond to environmental and other stimuli to keep metabolism optimally functional. N deficiency reduces dry matter accumulation, kernel number and could result in a substantial decrease in grain yield (Uhart and Andrade, 1995; DeBruin et al., 2013). On the other hand, oversupply of N is detrimental to the underground water, as being highly soluble, a substantial portion of it can leach into the water tables. Runoff resulting from heavy rains into streams and deltas leads to excessive algal growth, which adversely affects aquatic life by choking it off the oxygen supply. Over-fertilization in agricultural areas, aside from adversely affecting the ecosystem, causes economic loss to the farmers. The main causes of N loss are leaching, runoff, denitrification, and, volatilization (Nielsen, 2006). N from the Mississippi River Basin has been implicated as the main cause for the expanding hypoxic zone that develops each spring and summer on the Louisiana-Texas shelf of the Gulf of Mexico (Goolsby and Battaglin, 2000). Nitrate concentrations have increased several folds in streams of the Mississippi Basin during the past 100 years, and the annual delivery of nitrate from the Mississippi River to the Gulf has

nearly tripled since the late 1950's. Approximately one million mt (metric ton) of nitrate discharged annually from the Mississippi River Basin (Goolsby et al., 1999) could potentially produce more than 20 million mt of organic carbon annually in the Gulf of Mexico (Goolsby and Battaglin, 2000).

Improving N use efficiency (NUE) of maize would reduce N losses from the soil. NUE, which in cereals has been defined as the ratio of grain produced per unit of soil N, can be subdivided into two main components: N uptake efficiency (total plant N/soil N) and N utilization efficiency (total grain N/total plant N) (Moll et al., 1982; Dhugga and Waines, 1989). Since N uptake efficiency is derived from multiplying final biomass with N concentration, N uptake efficiency should in fact be referred to as N acquisition efficiency because it is difficult to separate the effect of feedback inhibition from a limitation in root uptake. Once absorbed by the roots, nitrate is transported to the leaves for reduction and incorporation into amino acids and other molecules, followed by incorporation into various macromolecules, including enzymes. A limitation at any point in the N metabolism pathway could limit N acquisition and utilization and, as a result, biomass production. This research was designed to evaluate the enzymes and proteins involved in N reduction and incorporation into organic molecules in order to determine associated QTL. QTL identified henceforth will help in selecting recombinants that combine desirable activities for improved NUE.

The pathway for N reduction and incorporation of reduced N into organic molecules is well understood (Fig. 3.1) (Yemm and Folkes, 1958; Lea, 1990; Lea and Azevedo, 2007; Lea and Miflin, 2010). Nitrate is reduced to nitrite by nitrate reductase (NR) in the cytoplasm, followed by reduction of nitrite in the plastids to ammonium by nitrite reductase (NiR). Ammonium thus generated is aminated into glutamine from glutamate by glutamine

synthetase (GS). Another enzyme, glutamine-2-oxoglutarate aminotransferase or glutamate synthase (GOGAT), then converts glutamine back to glutamate, producing an additional glutamate along the way from 2-oxoglutarate. Asparagine synthase (AS) produces asparagine and glutamate from glutamine and aspartate. Glutamate can serve as an amino donor for other amino acids, a reaction accomplished by different amino transferases. For instance, alanine aminotransferase (AlaAT) catalyzes the amino transfer to pyruvate resulting in 2-oxoglutarate and alanine (Miyashita et al., 2007), while aspartate aminotransferase (AspAT) forms 2-oxoglutarate and aspartate after transferring the amino group of glutamate to oxaloacetate. Following N assimilation, glutamate, asparagine, glutamine and other amino acids, constituents of proteins, are transported via vascular tissues to the growing organs or stored, as vegetative storage proteins, which can aid plant growth during periods of N deficiency.

N and C-metabolisms are highly interconnected (Nunes-Nesi et al., 2010). Certain metabolites and enzymes perform key roles in C metabolism and are regulated by the status of N in the cell (Sugiharto et al., 1990). Oxaloacetate, one of the carbon skeletons utilized in amino acids synthesis, is made from the addition of bicarbonate to phosphoenol pyruvate (PEP) by a reaction catalyzed by phosphoenol pyruvate carboxylase (PEPC). Pyruvate orthophosphate dikinase (PPDK) is responsible for catalyzing the regeneration of phosphoenol pyruvate. Ribulose-1,5-biphosphate carboxylase/oxygenase (Rubisco), considered as the most important enzyme on Earth, catalyzes the carboxylation of ribulose-1,5-biphosphate and produces triose phosphate, the building block of sugars (Farquharson, 2012). The information on the genetic basis underlying the regulation of plant C and N interactions is scarce.

A vast majority of today's commercial maize germplasm originated from seven progenitor lines, including B73 and Mo17 (Mikel and Dudley, 2006). Both these inbreds differ in their response to N fertilization (Balko and Russell, 1980) and are parents of the IBM (Intermated B73xMO17) mapping population (Lee et al., 2002). After ten rounds of random mating, 360 doubled haploid (DH) lines were generated from the IBMSyn10 population (Hussain et al., 2007) resulting in a higher-resolution mapping population that can be directly associated to the physical map established for the B73 inbred line (www.maizesequence.org). On the whole, the maize breeding community would greatly benefit from an understanding of the genetic basis of N-metabolism, especially at the testcross (TC) level, which is the type of cultivar usually planted in commercial fields.

Mapping of quantitative trait loci (QTL) is routinely implemented in plant breeding programs. Linkage mapping of QTL allows to experimentally estimate the mean and variance associated with a specific locus. The procedure relies on differences among the trait means of genotypes at a marker locus (Bernardo, 2010). The precision in the identification of a QTL can be critical to the time, expense, and probability of success of further studies (e.g., identification of candidate genes and positional cloning) (Remington et al., 2001). That precision in the estimation of the QTL position, referred as resolution, may vary substantially depending on several factors such as recombination frequency present in the mapping population, marker density and population size (Yu et al., 2011).

Several studies have shown association between QTL and N-metabolism related enzymes. For instance, Hirel et al. (2001) developed a quantitative genetics approach by associating metabolic functions and agronomic traits to DNA markers using information obtained from a previous investigation (Bertin and Gallais, 2000). Coincidences of QTL

clustered mainly in chromosome 5 for yield and its components, besides genes encoding cytosolic GS were identified under same genomic region. Contemporarily, the same research group published an article were it is claimed that, based on the coincidence between previously mapped QTL and genes encoding enzymes involved in N assimilation, NUE can be improved by marker-assisted selection and genetic engineering (Masclaux et al., 2001). Successively, after the analysis of 140 RIL genotyped with 152 marker loci, the identification of QTL for germination efficiency which co-located with genes encoding cytosolic GS has been reported (Limami et al., 2002). Furthermore, agronomic and physiological traits were used to detect QTL and determine their causal relationships in an integrated manner (Gallais and Hirel, 2004). For that investigation, agronomic traits were measured in a set of 99 hybrids by Bertin and Gallais (2001) while physiological traits were studied at the level of lines (77 RIL). After identifying several QTL coincidences with GS locus, it was concluded that GS locus in chromosome 5 was a candidate gene responsible for phenotypic variation in the use of N. Using the IBMSyn4 maize population, Zhang et al. (2010) measured activities of ten enzymes involved in carbon and N-metabolism. Seventy-three QTL associated with enzyme activities and eight QTL associated with biomass were identified. Most of the enzyme activities QTL were away from the known genomic locations of genes but three *cis*-QTL were identified for NR, glutamate dehydrogenase and shikimate dehydrogenase. Recently, a QTL analysis was performed for 12 metabolites directly related to C- and N-metabolism in the maize nested association mapping (NAM) population. An association mapping approach was implemented and 101 candidate genes were identified (Zhang et al., 2015). QTL associated with enzymes of N-metabolism in hybrid seedlings based on a high-resolution mapping population were not studied. We believe that the identification of QTL in

hybrid background is essential to mitigate the concern that the ones identified from an inbred population may not be relevant in hybrids.

In this investigation, a mapping population of TC genotypes, derived from the cross between IBMSyn10-DH lines and an elite inbred, was grown under hydroponics and leaf samples were analyzed in order to identify QTL associated with enzyme activity and metabolites involved in the N-metabolism pathway.

Materials and Methods

Plant material

A total of 176 TC genotypes derived from the cross between each IBMSyn10-DH line and an elite inbred were used. The IBMSyn10-DH population, developed by Hussain et al. (2007), is a set of DH lines derived from a population after ten generations of random mating from the cross between inbred lines B73 x Mo17. Each DH line was crossed by an elite inbred (PEI), property of DuPont Pioneer (closed pedigree), to generate the TC genotypes.

Experimental design

Kernels from each TC genotype were germinated in autoclaved paper rolls and sterilized water and, subsequently grown under hydroponic conditions. Ten tanks (i.e., sets) containing appropriate growing media were planted with a total of 264 seedlings in each set. In every set, 22 different genotypes were grown, and each genotype was replicated 12 times. Two genotypes (B73 and Mo17 each crossed to the PEI) served as controls, and were included in every set and replication.

The growing media consisted of $MgSO_4$.7H2O 0.5 mM, KH_2PO_4 0.5 mM, Fe-EDTA 0.1 mM, FeEDDHA 0.1 mM, $Ca(NO_3)2.4H_2O$ 1.25 mM, KNO_3 2.5 mM, Na(OH) 0.1 mM, and 0.4 L of trace elements (25 mM H_3BO_3, 2 mM $MnSO_4.H_2O$, 2 mM $ZnSO_4.7H_2O$, 0.5 mM $CuSO_4.5H_2O$, 0.5 mM $Na_2MoO_4.2H_2O$ and 50 mM KCl) in a total of 400 L solution per hydroponic tank. Two weeks after planting, the six most representative plants of each genotype, based on both their root and shoot development, were selected and transplanted into another hydroponic tank with same media.

When plants reached V4 stage (Abendroth et al., 2011), leaf samples were taken and stored at -80°C while the rest of the plant tissues were dried for 12 days at 48°C.

Biochemical assays

Activities of eight enzymes related to N-metabolism were determined. These enzymes were: NR, NiR, GS, GOGAT, AlaAT, AS, AspAT and PEPC, and specific protocols were adapted by K. Dhugga, R. Abbaraju and L. Fallis. GS, GOGAT, AspAT and PEPC assay protocols were adapted from Gibon (2004), while NR from Lea et al. (1990), NiR from Bourne and Miflin (1973), AS assay from Joy and Ireland (1990) and AlaAT protocol was modified from Ashton et al. (1990). Metabolites nitrate and glutamate were measured as byproducts of enzyme reactions. The concentration of the proteins of several enzymes were determined: PEPC (i.e., $PEPC_e$), PPDK, rubisco and Lox6. Each determination was based on ELISA (Engvall and Perlmann, 1971) protocols and DuPont Pioneer's proprietary antibodies. All measurements were determined by comparing absorbance of each specific biochemical reaction with known standards using a spectrophotometer (Spectramax Plus 384 Microplate Reader, Molecular Devices).

Plant tissues, leaf and root, were weighed and analyzed for N content by combustion analysis. Based on biomass weight and percentage of N measurements, total amount of N present in shoot (TN_s) and root (TN_r) tissues were calculated. N_{ratio} was calculated as the ratio between TN_s and TN_r.

Statistical analysis

All statistical analyses were implemented in R statistical program (RCoreTeam, 2014). Initial data analysis of the raw data was based on the ggplot2 package (Wickham, 2010) and GGally (Schloerke et al., 2014). As a first step, a univariate analysis, where a single variable is fitted in a model, followed by a multivariate approach, where multiple variables are analyzed simultaneously, was performed in order to comprehend the relationship among variables. The determination of outliers present in the dataset, based on a jackknife resampling strategy, was applied. As described in Trucillo-Silva (2015), a statistical model is fitted n times, systematically omitting one observation from the dataset, followed by the prediction of random effects for a subset of most consistent genotypes each of the n times. The aim of the process is to target "real outliers" based on the complete information gathered in the experiment and fine-tune the statistical model, quantified by improvements in log-likelihood, Akaike and Bayesian information criterion values after discarding misleading observations, while keeping informative and true observations for later analysis. The mixed model was fitted with ASReml R package (Butler et al., 2007) and correspondent mixed model equations were solved for prediction of random effects and estimation of fixed effects.

The statistical model can be represented as follows:

$$y = Xb + Zu + e$$

where y denotes a n x 1 vector of observed response values, b is a p x 1 vector of fixed effects, X is a n x p design matrix, u is a q x 1 vector of random effects, Z is a n x q design matrix, and e being the error term.

The following assumptions were used: E (u) = 0, E (e) = 0, Cov (u, e) = 0, and Var (u) = G and, Var (e) = R. The G matrix had a compound symmetry structure on the genotype levels and R matrix is a diagonal matrix with different values for each set, allowing non-constant variance across sets. The response variable was the activity of the enzyme, metabolite concentration, final ELISA determination and N content, respectively. Set, the light replicate and plate were included as fixed effects in the model (where replicate and plate are nested in set), and check genotype effect was included as a continuous covariate. Finally, a random effect for the genotype was included in the linear model. During the process described above, several genotypes were discarded separately for each trait. Extreme cases were traits as AlaAT with eight total genotypes discarded, and nitrate, PPDK, Lox6, $PEPC_e$ and N_{ratio} were no genotypes were taken out of the analysis and sample size totalized 176 genotypes.

Significance of genetic variance was calculated based on log-likelihood ratio test by comparing models with and without the TC random effect. Correlation was calculated, after Bonferroni correction for multiple comparisons, among BLUP values for each pair of traits, and repeatability was derived from the variance estimates from ASReml. As variance components were estimated for each of the different sets, a different value of repeatability was estimated for each set and then partial estimates were averaged correspondingly.

Genotypic information, genetic and physical maps

TC genotypes were analyzed with a total of 5,306 single nucleotide polymorphism (SNP) markers generated at Beijing Genomics Institute. Physical and genetic position of each SNP were determined and genetic maps were created using R/qtl (Broman et al., 2003). Based on the approach used for the determination of real outliers described previously (Trucillo-Silva et al., 2015), different genotypes were omitted from the analysis of each trait. As a result, a different genetic map was determined for each individual trait. Recombination fractions were estimated and Kosambi mapping function was implemented to calculate genetic map distances (Kosambi, 1944). Furthermore, as the recombination between linked loci increases every generation, leading to an expansion of the genetic map, mapping distances were adjusted with the purpose of comparison with previous investigations. The expansion factor was determined based on the following equation: $\propto = \frac{j}{2} + (2\mathrm{i} - 1)/2\mathrm{i}$, where j corresponds to the number of generations of intermating including the two generations for generating the F_2, and i is the number of inbred generations after intermating (Teuscher et al., 2005).

A total of 13 genetic maps were produced, depending on the specific genotypes included in the analysis of each trait, followed by adjustment of the genetic distances with the goal of comparing them with the previous QTL studies. Average spacing between markers was 2.13 cM (0.33 cM adjusted distance) while the maximal spacing between markers was nearly 45 cM (7 cM adjusted distance), located in chromosome 6. The average total map length was 11,275 cM. Real genetic map distances were reduced by a factor of 6.5 to estimate adjusted F_2 map distances (Fig. 3.3A) and final adjusted map was 1,734.65 cM in length (Fig.3.3B).

With regard to physical distance, the length of the total genome was 2,051.8 Mb and on average a marker was positioned every 400Kb. The widest gaps between markers, 69.8 and 67.4 Mb, were located in chromosomes 2 and 9, respectively.

QTL mapping and identification of candidate genes

Associations between phenotypes and genotypes were determined using QTL Cartographer (Basten et al., 2002). Single-marker analysis, followed by linear regression analysis and composite interval mapping (CIM) was performed. For CIM, Zmap (model 6) was implemented, using the ten most significant marker cofactors identified by forward and backward regression. In addition, QTL were scanned at intervals of 1 cM and at every marker while cofactors located within a window of 10 cM of the scanned position were excluded from the analysis. In order to determine LOD score thresholds of 5%, and to further identify significant QTL, 1,000 permutations were performed for every trait. Two nearby QTL were considered as different when LOD peaks were localized 20 cM or greater apart. Furthermore, a multiple interval mapping (MIM) analysis was performed by fitting all previously identified QTL from CIM analysis, and parameters were re-estimated and positions refined. In addition, all pairwise interactions between QTL in every model were studied for each trait. The significance was determined based on the information criterion: IC (k) = -2 (log (L) - kc (n)/2), where the penalty function corresponds to: c (n) = log (n) and a threshold of 0.0 was used (Basten et al., 2002). The proportion of the total phenotypic variance associated with each model was estimated.

In addition, physical genomic regions corresponding to 1-LOD QTL regions were examined, and putative genes related to N-metabolism were prioritized based on their

annotations at MaizeGDB (Lawrence et al., 2008) and NCBI (http://www.ncbi.nlm.nih.gov), and proposed as targets for further studies.

Results

Statistical analysis for N-metabolism related traits

Genetic variance was highly significant for all the traits studied (Table 1). In some hybrids, mean values exceed two standard deviations compared to the parental (Mo17 and B73) performance in their respective TC version. Repeatability ranged from 0.27 - 0.86. The lowest value of repeatability was for AS, while the highest value corresponded to the metabolite nitrate (Table 3.1).

Correlation between N-metabolism related traits

Correlation analysis among all traits is presented in Fig. 3.2. From a total of 136 pairwise Pearson correlations, the percentage of correlation coefficients that were significantly different from zero at p-value<0.001, <0.01, or <0.05 was respectively 13, 5, and 4. Significant correlations between enzyme activity, metabolites, and protein concentrations were all positive. Negative significant correlations were found between TN_s and activities of AS (-0.31), AspAT (-0.31), GOGAT (-0.28), and among TN_r and N_{ratio} (-0.43). Strong, positive correlation were observed between TN_r and TN_s (0.89), AlaAT activity and glutamate (0.79), PPDK and $PEPC_e$ (0.69), as well as between AspAT and GS (0.56).

Identification of quantitative trait loci

A total of 44 QTL were identified across all traits spread across all the chromosomes. Chromosomes 6 and 8 possessed the largest and smallest numbers, eight and two, respectively of QTL (Table 3.2, Fig. 3.4).

The number of QTL detected per trait ranged from one (NR, PPDK and rubisco) to five (AS) (Fig. 3.5). Individual QTL explained on average of 8.8% of the variance with some explaining as much as of 16.2% (TN_s) or as low as 5.9% ($PEPC_e$). Most of the QTL, 77%, accounted for less than 10% of the variance.

Confidence intervals (CI 1-LOD) for QTL localization ranged from 2 – 28 cM (0.31-4.26 cM adjusted distance) in length, with an average of 8.36 cM. Those CI are equivalent to 0.2-12 Mb in physical distance, with an average CI length of 2.1 Mb.

A hotspot QTL region was localized on the short arm of chromosome 6, comprising QTL associated with five different phenotypes including enzyme activities (AS, GOGAT and PEPC) and metabolites (glutamate and nitrate). Furthermore, QTL for rubisco and $PEPC_e$ (Chr. 9), and for AS and GOGAT (Chr. 10) were respectively co-localized on the genetic map.

Multiple interval mapping – Epistasis

In most traits, epistatic interactions did not significantly improve the fit of the models, and epistatic effects were excluded from genetic models. Even though, epistatic effects were retained in the traits PEPC and nitrate, explaining 5 and 2.5% variation respectively. MIM models explained a significant portion of the total variance in AS and nitrate (over 30%) and, just over 7% for other traits (AlaAT, NR, PPDK) (Table 3.3).

Candidate genes

On average 61 candidate genes were identified within 1-LOD QTL regions across the maize genome; ranging from 1 to 278 genes, for Rubisco-1 and Glutamate-2 QTL, respectively. A subset of the putative genes could be associated to a N-metabolism pathway. Most promising candidate genes ID are GRMZM2G008714, GRMZM2G045171, GRMZM2G082780, GRMZM2G088235, GRMZM2G155974, GRMZM2G028574, GRMZM2G166366, GRMZM2G343519, GRMZM2G402582 and, GRMZM2G481529 (Table 3.4).

All of the QTL for enzyme activities, metabolites and proteins identified in this study are located away from the known genomic locations of their corresponding structural genes. For example QTL for GS were identified at chromosomes 1, 5, 6, and 8 at physical positions 80.15, 83.95, 150.20, and 2.55 Mb, respectively. The structural genes for GS1 and GS2 are located on chromosome 1 between 271.02 – 273.44 Mb and on chromosome 2 between 18.94-19.46 Mb, based on the following nearest loci on the IBM2 2008 Neighbors map, respectively. Thus, the putative N-metabolism related genes identified under the QTL regions might be involved in regulating the activity of the respective enzymes through alteration of the metabolite pools, as was previously reported (Zhang et al., 2010).

Discussion

In this investigation 44 QTL associated with N-metabolism were identified in a high-resolution maize TC mapping population. In addition, QTL models explaining even greater than 30% of the genetic variance were identified for certain phenotypes, such as AS and nitrate. These discoveries may lead to a better understanding of the genetics underlying N-

metabolism in maize and possibly, towards breeding genotypes with the ability to increase yield performance per N unit supplied.

Consistent with previous studies, all significant correlations between enzyme activities were positive, suggesting that the enzymes were co-regulated to varying extents (Zhang et al., 2010). Indeed, significant correlations between enzyme activities, metabolites and proteins were positive as well (Fig. 3.2). With an increase in nitrate concentration, an increase in the activities of NR, NiR GOGAT and GS activities would be expected. A positive correlation between nitrate and glutamate further support this viewpoint (Fig. 3.2). Although a significantly positive correlation (0.59) between the concentration of PEPC protein and activity was observed, its deviation from unity suggest either the enzyme was partially inactivated during extraction or the extract contained endogenous inhibitors of this enzyme. PEPC activity is known to be inhibited by aspartate, oxaloacetate, and malate (Huber and Edwards, 1975). In addition, Zhang et al. (2010) found negative correlations between enzyme activities and biomass. Likewise, TN_r, and TN_s, both estimated as the product of N concentration of root and shoot biomass, respectively, were negatively correlated with most enzyme activity, metabolite and ELISA determinations. In addition, plants showing high TN_r also presented high TN_s (R^2=0.89).

Compared to a previous investigation (Zhang et al., 2010), repeatability values were to a certain extent lower on average (mean value of 0.55 versus 0.65), however GS showed a higher value (0.35 versus 0.20). The differences in repeatabilities may be due to the fact that enzyme measurements were performed on a robot-based platform by Zhang et al. (2010), while manual procedures were employed in this investigation. Even though six replications were implemented, variation due to experimental error was not possible to be thoroughly

eliminated. Nevertheless, the significant genotypic effect for all traits strengthens the likelihood to identify responsible QTL.

QTL were associated with the activities of eight enzymes, two metabolites, four ELISA determinations, and three N-content phenotypes. In agreement with a previous investigation (Zhang et al., 2010), the same QTL were detected for NR, GS and AspAT localized in chromosomes 4, 5, and 9, respectively. Nevertheless, most of the QTL reported in other maize studies (Agrama et al., 1999; Hirel et al., 2001) failed to co-localize with the QTL identified in this analysis and were determined in different genetic location, outside a 20 cM window or were even unlinked (e.g., two QTL associated with NR were determined in chromosome 5 by Hirel et al., 2001, whereas a single QTL associated with NR was found in chromosome 4 in this investigation and by Zhang et al., 2010).

A few QTL detected in this study co-localized with QTL for different agronomic traits. For instance, GOGAT-1 QTL is exactly at the same location as a QTL previously associated with the determination of ear-per-plant, and PPDK-1 QTL co-localized with a grain yield QTL under high N and a QTL associated with number of kernels per year (Agrama et al., 1999). Hence, some genomic regions seem to be affecting more than one trait or the presence of pleiotropy.

Herein QTL associated with different phenotypes did not co-locate as regularly as it was determined in a previous study (Zhang et al., 2010). In that study, three genetic regions (1-LOD confidence interval around the LOD peak) on chromosomes 1, 6, and 7 containing QTL for several enzymes were identified; however no QTL signal was detected under those exact regions in this QTL analysis. That investigation was based on the IBM-Syn4 population and genotyped with 2,200 DNA marker loci. Thus, it is possible that the co-

location of some QTL was due to a lack of genetic resolution (Zhang et al., 2010). That concern is less important in this investigation because of the additional rounds of intermating for the creation of the segregating population and the higher marker density.

A lower number of QTL was identified per trait compared to a preceding investigation based on IRILs derived from IBMSyn4 population of maize (Zhang et al., 2010). The identification of QTL would depend essentially on the magnitude of the QTL effect and population size (Beavis, 1998). Because a large number of small-effect QTL were expected to be segregating in the genome, and based on the size of our segregating population (176 individuals), only a subset of the total number of real segregating QTL were expected to be identified. Moreover, the additional generations of intermating used to create this population could affect the number of QTL detected since QTL previously identified in large linkage blocks, might be separated into several smaller-effect QTL after recombination occurred. Hence, even though more recombination cycles are better for improving mapping accuracy and resolution, the power to detect a QTL, each with very small effect, would be expected to be less. Furthermore, inbred lines were analyzed by Zhang et al. (2010) whereas TC materials are used in this study. According to prior investigations (Beavis et al., 1994; Schon et al., 1994), little evidence of common QTL detection between inbred per se and TC progeny was found, suggesting that marker-assisted selection strategies based on QTL identified at the inbred level would not assure the selection of hybrids with superior performance. Genetic studies based on a TC mapping population might thus be preferred if the objective is to select superior hybrids based on these traits.

The extensive amount of unexplained genetic variance by the multiple QTL models across traits (92.75 – 68.25%) suggests that there might be several small-effect QTL still

undetected in this analysis. The sum of the effect of numerous QTL, each with small marginal effect, plus any type of epistasis they might be involved in, should account for all the unexplained genetic variance in the QTL models. It has been established that epistasis could make a large contribution to the genetic regulation of complex traits (Carlborg and Haley, 2004). However, significant epistasis was detected in two out of 17 traits (PEPC and nitrate), and epistatic effects were much smaller than additive effects (5 versus 12.89 for PEPC and 2.5 versus 18.61 for nitrate). Similarly, no significant epistatic effects between QTL was detected in a recent study based on the maize NAM population, which included both Mo17 and B73 (reference line) genotypes (Zhang et al., 2015).

Several annotated protein-coding genes were identified under QTL intervals determined in this investigation. Four of the maize genes detected and, annotated in B73 genome, were also identified in a previous meta-QTL investigation aiming to discover candidate genes for N-use efficiency in maize (Liu et al., 2012). Those genes were GRMZM2G046382 (heat shock protein), GRMZM2G116204 (auxin-binding protein), GRMZM2G360339 (B12D protein) and, GRMZM2G123633 (cell wall invertase 3). In addition, 11 candidate genes revealed important putative functions related to N-metabolism in Arabidopsis and rice. Three of them were proposed as candidate genes in a recent investigation on the maize NAM population (Zhang et al., 2015). Those genes were GRMZM2G008714, GRMZM2G045171, and GRMZM2G180625; coding for pyruvate kinase, sucrose synthase and glyceraldehyde-3-phosphate dehydrogenase, respectively. Pyruvate kinase is a key enzyme in the glycolytic pathway, that catalyzes the transphosphorylation from PEP and ADP to pyruvate and ATP (Valentin et al., 2000); sucrose synthase catalyzes a reversible reaction between sucrose and uridine diphosphate

glucose in order to mobilize sucrose into multiple pathways that utilize activated sugars (Subbaiah et al., 2007), while glyceraldehyde-3-phosphate dehydrogenase is an enzyme of the glycolytic pathway which catalyzes the conversion of glyceraldehyde-3-phosphate and NAD^+ to 1,3 diphosphoglycerate and NADH (Harris and Waters, 1976). Moreover, in a recent study (Simons et al., 2014), glyceraldehyde-3-phosphate dehydrogenase had shown to be a key gene related to the decrease in biomass yield on a gln1-3 mutant maize genotype with B73 background in modeled conditions based on proteomic and transcriptomic data. In addition, eight other genes were found: GRMZM2G082780, GRMZM2G028574, GRMZM2G088235, GRMZM2G155974, GRMZM2G166366, GRMZM2G343519, GRMZM2G402582, and GRMZM2G481529. GRMZM2G082780 and GRMZM2G028574 both had been described as PEPC putative genes. GRMZM2G088235 is an urease accessory protein involved in the N-recycling from ureide, purine, and arginine (Witte et al., 2005); and GRMZM2G155974 catalyzes the addition of glycine to γ-glutamyl-cysteine, generating glutathione. Glutathione is a key water-soluble antioxidant, the storage form and long-distance transport form of reduced sulfur (Zagorchev et al., 2013). It has been established that as proteins contain both N and sulfur, a deficiency of either would severely affect protein synthesis and plant growth (Bouranis et al., 2008). GRMZM2G166366 was annotated as an aspartate kinase which catalyzes the phosphorylation of aspartate to for β-aspartyl phosphate, and is responsible for the first step in the biosynthesis of the amino acids lysine, methionine, and threonine (Azevedo et al., 1992). GRMZM2G343519 was annotated as a glutaredoxin protein, which is involved in protective and regulatory mechanisms in maize (Yang et al., 2015). In accordance with the highly significant correlation between PPDK and $PEPC_e$ (0.69), the candidate gene GRMZM2G402582, was annotated as PPDK and it was located

under a $PEPC_e$ QTL ($PEPC_e$-3). Finally, the putative gene GRMZM2G481529, a cytosolic enolase or phosphopyruvate hydratase, is described as a metalloenzyme responsible for the catalysis of the conversion of 2-phosphoglycerate to PEP, having orthologs within sorghum and rice.

The eleven coding regions underlying QTL identified in this study constitute promising candidates for validation studies.

Conclusions

We identified 44 QTL associated with the physiological traits related to N-metabolism in a high-resolution maize TC mapping population. Furthermore, genetic QTL models accounting for 7 to 31 % of the genetic variance were derived, and 11 candidate genes within QTL genomic regions are identified. These QTL constitute candidates for integration into a breeding program to improve NUE in maize. The immediate next step would be to grow these TCs in the field and assay the enzymes at flowering time when the canopy is fully developed to determine whether and how much of the variation in grain yield they explain. The field grown TCs will also make it possible to determine whether the QTL identified from the seedlings in the growth chamber relate to ultimate trait: grain yield.

Acknowledgments

The authors would like to thank R.F. Baker for Plant Breeding – Department of Agronomy – Iowa State University and DuPont Pioneer for making this research possible and Iowa State University undergraduate students (especially Guan Yi Lai) who assisted with planting and

harvesting in the hydroponic experiments. Requests for testcross materials and IBMSyn10-DH lines may be directed to DuPont Pioneer and Dr. Michael Lee, respectively.

References

Abendroth LJ, Elmore RW, Boyer MJ, Marlay SK (2011) Corn growth and development. *In*. Extension Publication #PMR-1009, Iowa State University

Agrama HAS, Zakaria AG, Said FB, Tuinstra M (1999) Identification of quantitative trait loci for nitrogen use efficiency in maize. Molecular Breeding **5:** 187-195

Ashton ARB, J.N.Furbank, R.T.Jenkins, C.L.D.Hatch, M.D. (1990) Enzymes of C4 photosynthesis. In PJ Lea, ed, Methods in Plant Biochemistry. Academic Press, London, pp 39-72

Azevedo RA, Blackwell RD, Smith RJ, Lea PJ (1992) Three aspartate kinase isoenzymes from maize. Phytochemistry **31:** 3725-3730

Balko LG, Russell WA (1980) Response of maize inbred lines to N-fertilizer. Agronomy Journal **72:** 723-728

Basten CJ, Weir BS, Zeng ZB (2002) QTL Cartographer, Version 1.16. Department of Statistics, North Carolina State University, Raleigh, NC

Beavis WD (1998) QTL analyses: power, precision, and accuracy. In AH Paterson, ed, Molecular dissection of complex traits. CRC Press, Washington, D.C., pp 145-162

Beavis WD, Smith OS, Grant D, Fincher R (1994) Identification of quantitative trait loci using a small sample of topcrossed and F4 progeny from maize. Crop Science **34:** 882-896

Bernardo R (2010) Breeding for quantitative traits in plants, Ed 2nd. Stemma Press, Woodbury, MN

Bertin P, Gallais A (2001) Genetic variation for nitrogen use efficiency in a set of recombinant inbred lines II - QTL detection and coincidences. Maydica **46:** 53-68

Bouranis DL, Buchner P, Chorianopoulou SN, Hopkins L, Protonotarios VE, Siyiannis VF, Hawkesford MJ (2008) Responses to sulfur limitation in maize. In NA Khan, S Singh, U Shahid, eds, Sulfur assimilation and abiotic stress in plants. Springer-Verlag, Berlin, Heidelberg, pp 1-19

Bourne WF, Miflin BJ (1973) Studies on nitrite reductase in barley. Planta **111:** 47-56

Broman KW, Wu H, Sen Ś, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics **19**: 2990-2992

Butler D, Cullis BR, Gilmour AR, Gogel BJ (2007) Analysis of mixed models for S-language environments: ASReml-R reference Manual. Department of Primary Industries and Fisheries, Queensland, Brisbane, Australia

Carlborg O, Haley CS (2004) Epistasis: too often neglected in complex trait studies?. Nature Reviews Genetics **5**: 618-625

DeBruin J, Messina CD, Munaro E, Thompson K, Conlon-Beckner C, Fallis L, Sevenich DM, Gupta R, Dhugga KS (2013) N distribution in maize plant as a marker for grain yield and limits on its remobilization after flowering. Plant Breeding **132:** 500-505

Dhugga KS, Waines JG (1989) Analysis of nitrogen accumulation and use in bread and durum-wheat. Crop Science **29:** 1232-1239

Engvall E, Perlmann P (1971) Enzyme-linked immunosorbent assay (ELISA) quantitative assay of immunoglobulin-g. Immunochemistry **8:** 871-874

Farquharson KL (2012) Insight into Ribulose 1,5-Bis-Phosphate Carboxylase/Oxygenase Assembly in Maize. Plant Cell **24:** 3171-3171

Gallais A, Hirel B (2004) An approach to the genetics of nitrogen use efficiency in maize. Journal of Experimental Botany **55:** 295-306

Gibon Y, Blaesing OE, Hannemann J, Carillo P, Hohne M, Hendriks JHM, Palacios N, Cross J, Selbig J, Stitt M (2004) A robot-based platform to measure multiple enzyme activities in Arabidopsis using a set of cycling assays: Comparison of changes of enzyme activities and transcript levels during diurnal cycles and in prolonged darkness. Plant Cell **16:** 3304-3325

Goolsby DA, Battaglin WA (2000) Nitrogen in the Mississippi basin - Estimating sources and predicting flux to the Gulf of Mexico. *In*, U.S. Geological Survey Fact Sheet, FS-135-00. U.S. Geological Survey, Reston, Virginia

Goolsby DA, Battaglin WA, Aulenbach BT, Hooper RP (1999) Nitrogen flux and sources in the Mississippi river basin. *In* U.S. Geological survey toxic substances hydrology program - Proceedings of the technical meeting, Charleston, SC, Vol 2, pp 203 - 214

Harris JI, Waters M (1976) Glyceraldehyde-3-phosphate dehydrogenase. *In* PD Boyer, ed, The enzymes, Ed 3. Academic Press, New York, pp 1-49

Hirel B, Bertin P, Quillere I, Bourdoncle W, Attagnant C, Dellay C, Gouy A, Cadiou S, Retailliau C, Falque M, Gallais A (2001) Towards a better understanding of the genetic and physiological basis for nitrogen use efficiency in maize. Plant Physiology **125:** 1258-1270

Huber SC, Edwards, GE (1975) Inhibition of phosphoenolpyruvate from C4 plants by malate and aspartate. Canadian Journal of Botany **17:** 1925-1933

Hussain T, Tausend P, Graham G, Ho J (2007) Registration of IBM2 SYN10 doubled haploid mapping population of maize. Journal of Plant Registrations **1:** 81-81

Joy KW, Ireland RJ (1990) Enzymes of asparagine metabolism. *In*. Methods in plant biochemistry. Academic Press, London, pp 287-296

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugenics **12:** 172-175

Lawrence CJH, L.C., Schaeffer ML, Sen TZ, Seigfried TE, Campbell DA (2008) MaizeGDB: the maize model organism database for basic, translational, and applied research. International Journal of Plant Genomics **496957**: 1-10

Lea PJ (1990) Enzymes of ammonia assimilation. *In*. PJ Lea, ed, Methods in Plant Biochemistry, Vol3. Academic Press Limited, London, pp 257-276

Lea PJ, Azevedo RA (2007) Nitrogen use efficiency. 2. Amino acid metabolism. Annals of Applied Biology **151:** 269-275

Lea PJ, Miflin BJ (2010) Nitrogen assimilation and its relevance to crop improvement. *In*. CH Foyer, H Zhang, eds, Nitrogen metabolism in plants in the post-genomic era, Vol 42. Willey-Blackwell, Chichester, UK, pp 1-40

Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A (2002) Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant Molecular Biology **48:** 453-461

Limami AM, Rouillon C, Glevarec G, Gallais A, Hirel B (2002) Genetic and physiological analysis of germination efficiency in maize in relation to nitrogen metabolism reveals the importance of cytosolic glutamine synthetase. Plant Physiology **130:** 1860-1870

Liu R, Zhang H, Zhao P, Zhang Z, Liang W, Tian Z, Zheng Y (2012) Mining of candidate maize genes for nitrogen use efficiency by integrating gene expression and QTL data. Plant Molecular Biology Reporter **30:** 297-308

Masclaux C, Quillere I, Gallais A, Hirel B (2001) The challenge of remobilisation in plant nitrogen economy. A survey of physio-agronomic and molecular approaches. Annals of Applied Biology **138:** 69-81

Mikel MA, Dudley JW (2006) Evolution of north American dent corn from public to proprietary germplasm. Crop Science **46:** 1193-1205

Miyashita Y, Dolferus R, Ismond KP, Good AG (2007) Alanine aminotransferase catalyses the breakdown of alanine after hypoxia in Arabidopsis thaliana. Plant Journal **49:** 1108-1121

Moll RH, Kamprath EJ, Jackson WA (1982) Analysis and interpretation of factors which contribute to efficiency of nitrogen-utilization. Agronomy Journal **74:** 562-564

Nielsen RL (2006) N loss mechanisms and nitrogen use efficiency. *In* Ed Nitrogen management workshops. Purdue Agriculture Agronomy Extension, West Lafayette, IN

Nunes-Nesi A, Fernie AR, Stitt M (2010) Metabolic and Signaling Aspects Underpinning the Regulation of Plant Carbon Nitrogen Interactions. Molecular Plant **3:** 973-996

R Core Team (2014) R: A language and environment for statistical computing. *In*. R Foundation for Statistical Computing, Vienna, Austria

Remington DL, Ungerer MC, Purugganan MD (2001) Map-based cloning of quantitative trait loci: progress and prospects. Genetics Research **78:** 213-218

Schloerke B, Crowley J, Cook D, Hofmann H, Wickham H, Briatte F, Marbach M, Thoen E (2014) GGally: extension to ggplot2. *In*. R package version 0.4.8, Vienna, Austria

Schon CC, Melchinger AE, Boppenmaier J, Brunklausjung E, Herrmann RG, Seitzer JF (1994) RFLP Mapping in maize - Quantitative trait loci affecting testcross performance of elite European flint lines. Crop Science **34:** 378-389

Simons M, Saha R, Amiour N, Kumar A, Guillard L, Clement G, Miquel M, Li Z, Mouille G, Lea PJ, Hirel B, Maranas CD (2014) Assessing the metabolic impact of nitrogen availability using a compartmentalized maize leaf genome-scale model. Plant Physiology **166:** 1659-1674

Subbaiah CC, Huber SC, Sachs MM, Rhoads D (2007) Sucrose synthase: expanding protein function. Plant signaling & behavior **2:** 28-29

Sugiharto B, Miyata K, Nakamoto H, Sasakawa H, Sugiyama T (1990) Regulation of expression of carbon-assimilating enzymes by nitrogen in maize leaf. Plant Physiology **92:** 963-969

Teuscher F, Guiard V, Rudolph PE, Brockmann GA (2005) The map expansion obtained with recombinant inbred strains and intermated recombinant inbred populations for finite generation designs. Genetics **170:** 875-879

Trucillo-Silva I, Cook D, Lee M (2015) A methodical approach for identifying outliers in complex data. In preparation. Iowa State University

Uhart SA, Andrade FH (1995) Nitrogen deficiency in maize: I. Effects on crop growth, development, dry matter partitioning, and kernel set. Crop Science **35**: 1376-1383

Valentin G, Chiarelli L, Fortin R, Speranza ML, Galizzi A, Mattevi A (2000) The allosteric regulation of pyruvate kinase - A site-directed mutagenesis study. Journal of Biological Chemistry **275:** 18145-18152

Wickham H (2010) A layered grammar of graphics. Journal of Computational and Graphical Statistics **19:** 3-28

Witte CP, Rosso MG, Romeis T (2005) Identification of three urease accessory proteins that are required for urease activation in Arabidopsis. Plant Physiology **139:** 1155-1162

Yang F, Huyen Thanh B, Pautler M, Llaca V, Johnston R, Lee B-h, Kolbe A, Sakai H, Jackson D (2015) A maize glutaredoxin gene, Abphyl2, regulates shoot meristem size and phyllotaxy. Plant Cell **27:** 121-131

Yemm W, Folkes BF (1958) The metabolism of aminoacids and proteins in plants. Annual Review of Plant Physiology **9**: 245-280

Yu H, Xie W, Wang J, Xing Y, Xu C, Li X, Xiao J, Zhang Q (2011) Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. Plos One **6**

Zagorchev L, Seal CE, Kranner I, Odjakova M (2013) A Central Role for Thiols in Plant Tolerance to Abiotic Stress. International Journal of Molecular Sciences **14:** 7405-7432

Zhang N, Gibon Y, Gur A, Chen C, Lepak N, Hoehne M, Zhang Z, Kroon D, Tschoep H, Stitt M, Buckler E (2010) Fine quantitative trait loci mapping of carbon and nitrogen metabolism enzyme activities and seedling biomass in the maize IBM mapping population. Plant Physiology **154:** 1753-1765

Zhang N, Gibon Y, Wallace JG, Kruger Lepak N, Li P, Dedow L (2015) Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population. Plant Physiology **168:** 575-583

Figures

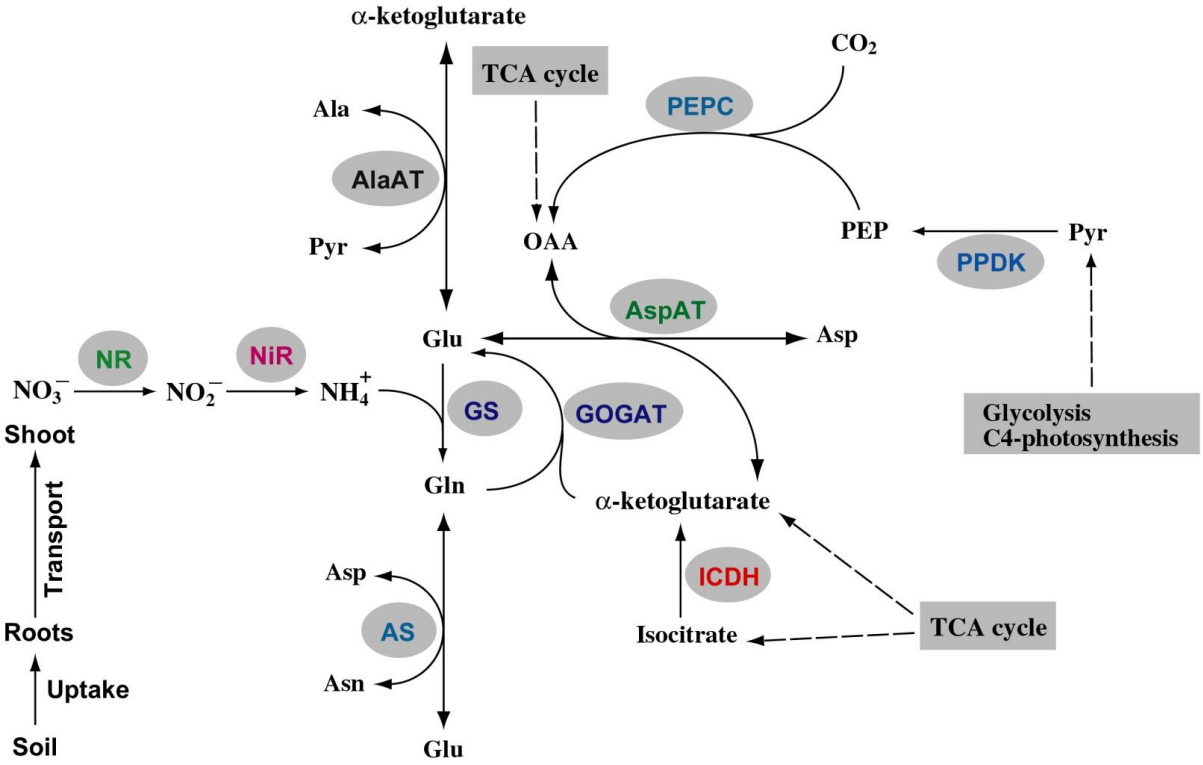

Figure 3.1. Enzymes and proteins involved in N-acquisition and assimilation in higher plants.

AlaAT, alanine aminotransferase; AS, asparagine synthase; AspAT, aspartate aminotransferase; GOGAT, glutamate synthase; GS, glutamine synthetase; ICDH, isocitrate dehydrogenase; NR, nitrate reductase; NiR, nitrite reductase; PEPC, phosphoenol pyruvate carboxylase; PPDK, pyruvate orthophosphate dikinase (Source: Kanwarpal S. Dhugga, 2015).

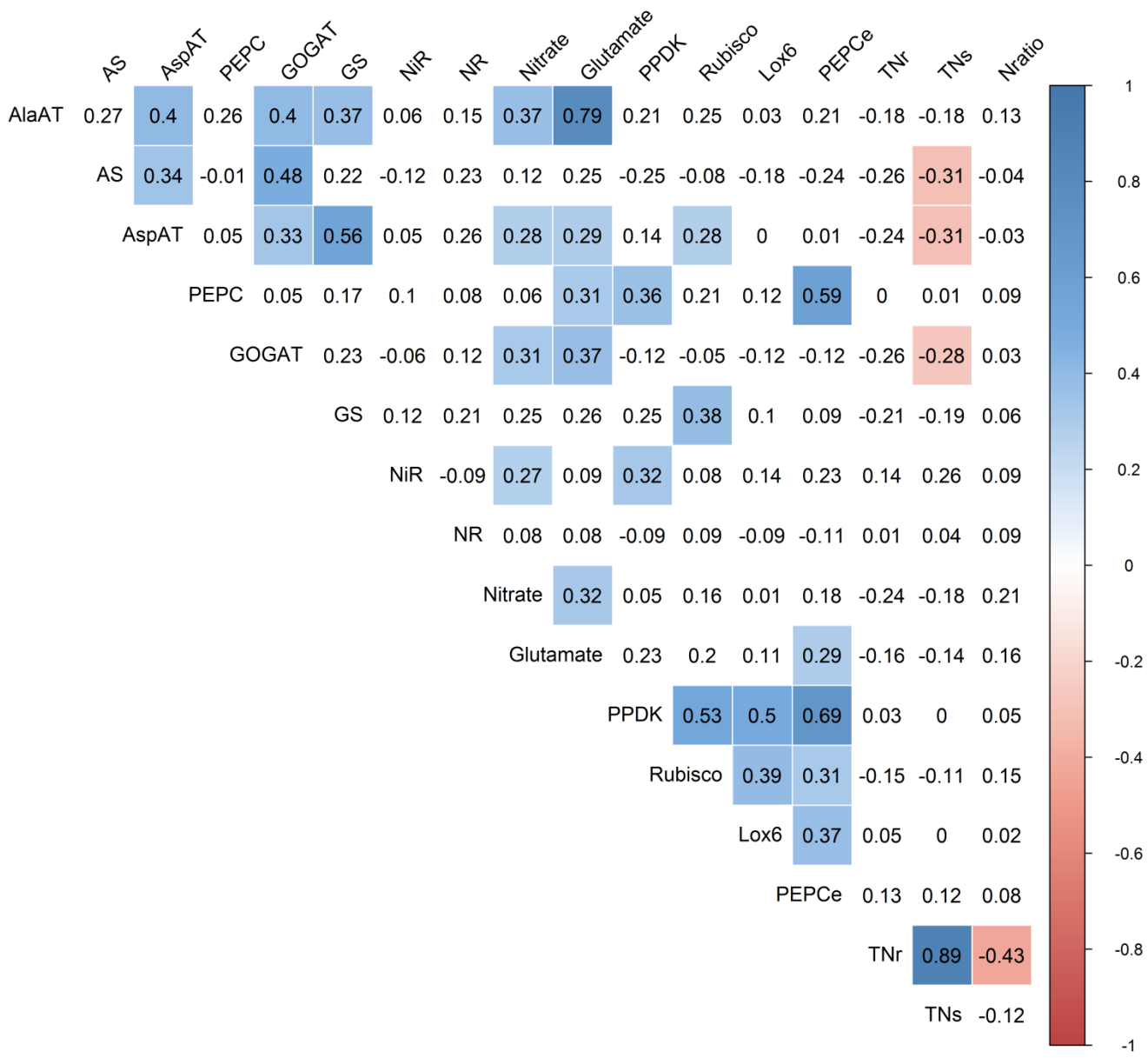

Figure 3.2. Correlation matrix-heatmap of N-metabolism related traits measured in leaf tissues in the IBMSyn10-DH TC population of maize.

Significant correlation (p-value<0.05) values are colored in blue (positive correlation) and red (negative correlation).

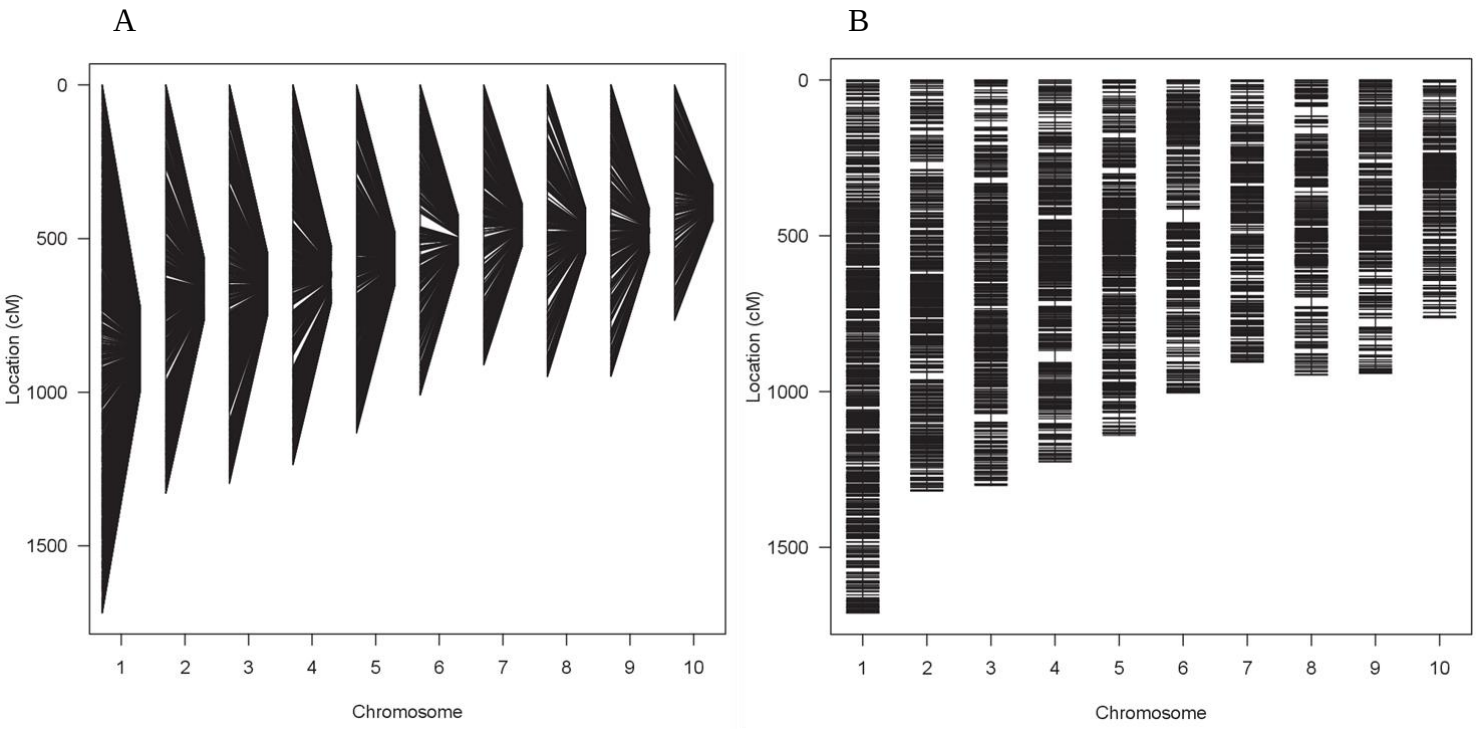

Figure 3.3. Adjustment of real genetic map and final adjusted F_2 genetic map.

(A) For each chromosome the line located on the left represents the real map estimated from the actual data while the line on the right corresponds to the adjusted genetic distance (cM)
(B) Final adjusted genetic map.

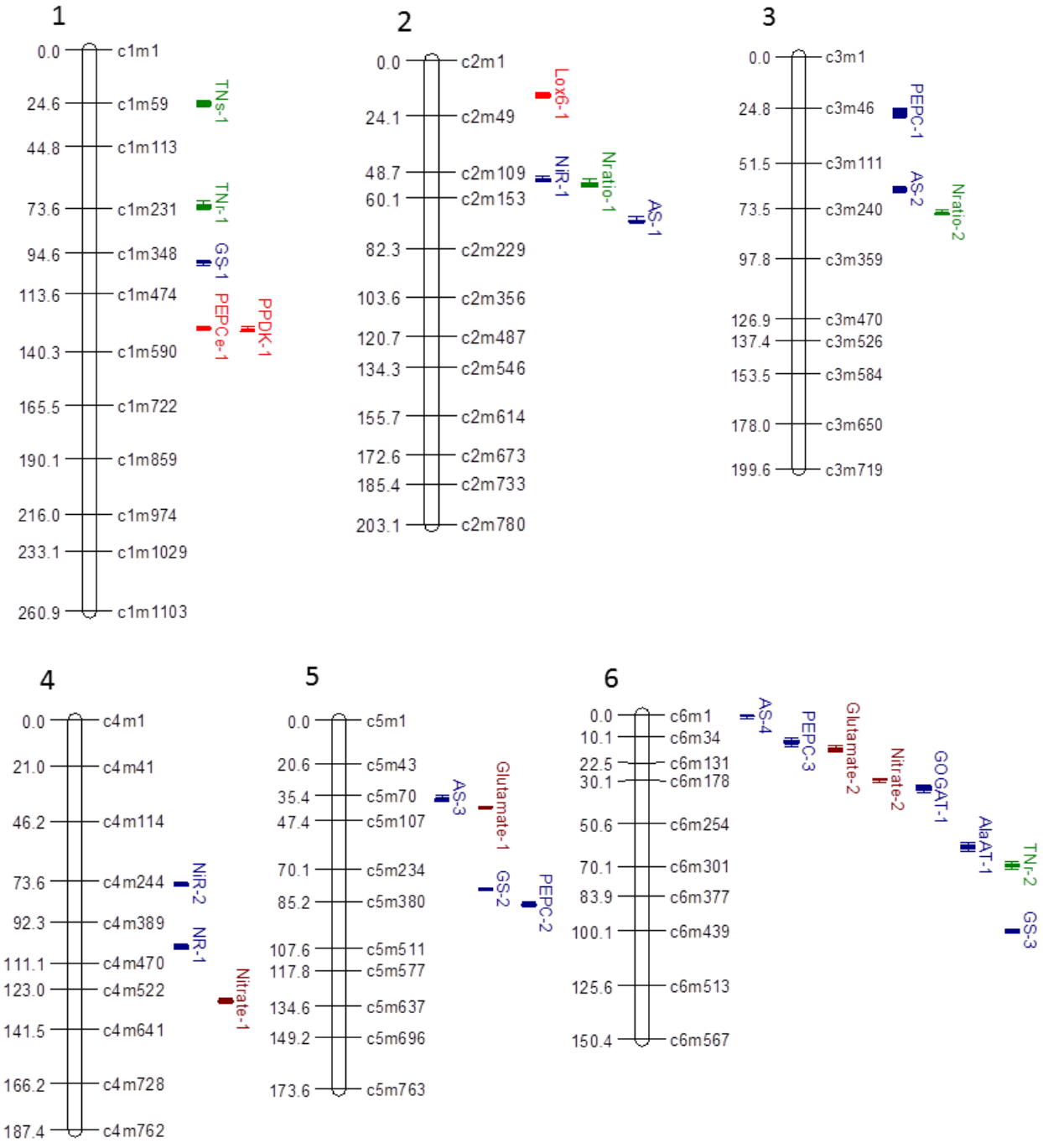

Figure 3.4. Genetic map and distribution of QTL associated with N-metabolism related traits identified in leaf tissues of the IBMSyn10-DH TC population of maize.

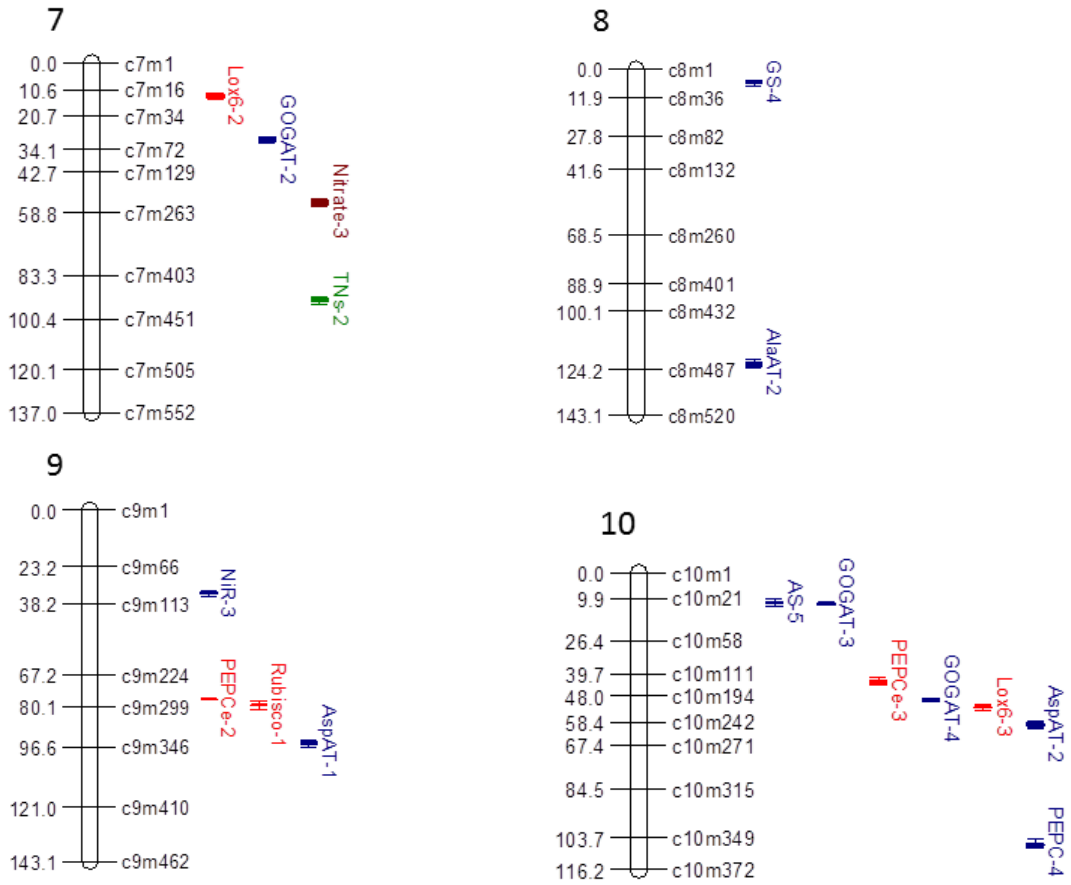

Figure 3.4 continued.

QTL positions shown at right of chromosomes (in cM) and lengths of bars are determined by 2-LOD confidence intervals. Only selected markers are displayed in the figure. QTL for enzyme activity are in blue, QTL for N content traits are in green, QTL for ELISA determination are in red, and QTL for metabolites are in brown. Figure created with MapChart 2.2 (Voorrips, 2002).

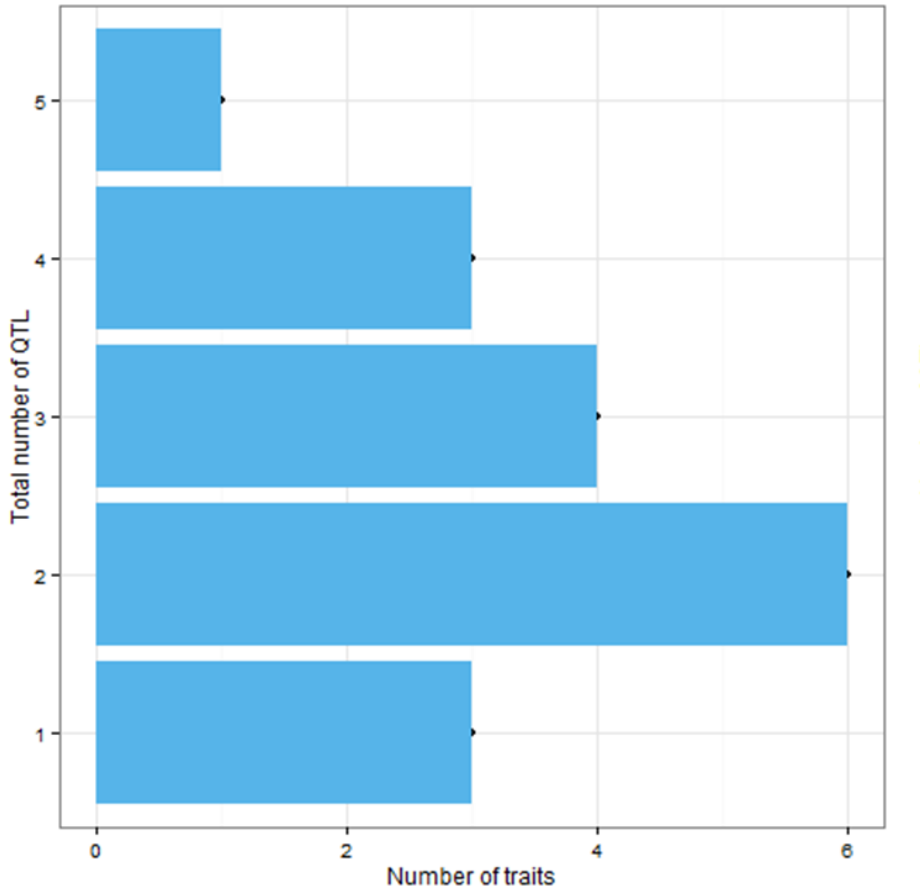

Figure 3.5. Total number of QTL associated with N-metabolism related traits in leaf tissues of the IBMSyn10-DH TC population of maize.

Tables

Table 3.1. Sample size, mean values for the population and checks, minimum values, maximum values, standard deviation, coefficient of variation, genetic effect p-value and repeatability of traits measured in leaf tissues of the IBMSyn10-DH TC population of maize.

Trait	Unit	n[a]	Pop μ[b]	B73TC[c]	Mo17TC[d]	Min[e]	Max[f]	SD[g]	CV[h]	G effect P[i]	Rptblity[j]
AlaAT	Normalized Glu per mg/protein	168	158.18	143.47	161.94	147.54	168.69	4.77	3.02	7.44E-06	0.39
AS	Normalized Glu per mg/protein	173	804.71	755.67	819.72	772.93	844.35	11.34	1.41	1.88E-03	0.27
AspAT	Normalized Glu per mg/protein	172	560.26	510.58	547.22	540.96	582.23	7.54	1.35	4.74E-02	0.31
GOGAT	Normalized Glu per mg/protein	172	174.56	163.00	181.28	166.40	182.81	3.22	1.84	8.68E-06	0.39
GS	GHA mmoles/mg protein	170	267.16	254.55	267.65	255.80	278.23	4.67	1.75	1.49E-03	0.35
NiR	Nitrite reduced/mg protein	173	185.45	172.12	237.50	109.31	271.26	33.00	17.79	1.33E-03	0.70
NR	nMoles nitrite/mg protein	171	4.10	3.67	4.00	2.55	6.49	0.73	17.80	5.25E-08	0.62
PEPC	uMoles NADH/min/mg protein	171	274.61	264.73	322.94	182.52	404.83	44.34	16.15	1.64E-11	0.52
Nitrate	nMoles/mg protein	176	242.18	153.27	361.04	149.50	462.60	57.66	23.81	1.38E-10	0.86
Glutamate	Normalized Glu per mg/protein	173	135.13	125.16	141.21	122.82	146.57	4.99	3.69	1.88E-07	0.50
PPDK	ug/ml	176	242.17	226.18	226.62	164.63	339.81	31.82	13.14	4.54E-08	0.64
Rubisco	ug/ml	172	238.16	271.76	173.75	197.85	289.37	50.53	21.22	2.42E-08	0.49
Lox6	ug/ml	175	44.12	44.68	43.45	19.25	75.52	9.37	21.24	8.71E-10	0.74
$PEPC_e$	ug/ml	176	282.00	289.83	272.01	186.22	412.08	32.92	11.67	4.49E-08	0.61
TN_s	mg	172	30.59	44.24	32.92	19.00	44.18	5.50	17.98	4.98E-08	0.73
TN_r	mg	176	4.75	7.10	5.10	2.21	9.13	1.09	0.23	7.12E-07	0.70
N_{ratio}	ratio	176	6.70	6.23	6.45	5.91	7.37	0.26	3.88	3.92E-13	0.50

[a] Population size, [b] Population mean, [c,d] BLUP value for parental genotypes in testcross genotype, [e] Minimum value, [f] Maximum value, [g] Standard deviation, [h] Coefficient of variation (%), [i] p value of the genetic effect, [j] Repeatability, normalized values were multiplied by a factor of 1.131 for AlaAT, AS and AspAT, and by 1.151 for GOGAT.

Table 3.2. QTL associated with N-metabolism related traits measured in leaf tissues in the IBMSyn10-DH TC population of maize.

QTL name	Chr[a]	Marker[b]	G Pos (cM)[c]	G Interval (cM)[d]	Adj (cM)[e]	P Pos (Mb)[f]	P Interval (Mb)[g]	LOD	R^2 (%)	Add effect[h]	# Genes
AlaAT-1	6	217	400.21	390.85-402.76	61.57	107.70	107.45-108.25	4.57	6.79	-1.28	37
AlaAT-2	8	356	795.25	787.71-801.56	122.35	170.35	170.15-170.65	4.71	7.01	-1.31	30
AS-1	2	160	456.57	454.19-460.14	70.24	22.95	22.85-25.3	5.78	8.78	-4.83	130
AS-2	3	152	419.09	411.63-426.49	64.48	29.45	27.10-31.25	4.24	6.32	-4.02	125
AS-3	5	74	244.73	238.47-246.37	37.65	8.65	8.15-9.35	4.32	6.49	4.18	46
AS-4	6	5	3.91	2.95-5.91	0.60	1.40	1.20-1.65	6.47	9.89	5.08	24
AS-5	10	22	77.15	74-78.02	11.87	3.05	2.95-3.15	4.51	7.20	-4.33	8
AspAT-1	9	296	616.45	612.53-622.45	94.84	139.55	138.85-139.85	8.17	12.26	-2.82	50
AspAT-2	10	201	386.89	381.43-390.33	59.52	132.35	132.05-132.6	5.36	7.83	2.23	35
GOGAT-1	6	152	225.40	214.7-227.58	34.68	93.55	93.25-94.15	6.17	9.19	1.07	39
GOGAT-2	7	59	195.78	192.48-201.06	30.12	8.75	8.65-9.25	5.09	7.46	0.91	20
GOGAT-3	10	23	77.94	74.51-81.72	11.99	3.05	2.95-3.15	9.98	15.67	-1.41	8
GOGAT-4	10	166	320.94	319.77-323.77	49.38	115.35	114.15-116.15	8.29	12.70	-1.72	60
GS-1	1	322	643.24	641.24-645.21	98.96	80.15	77.15-80.65	4.92	7.34	1.36	104
GS-2	5	269	519.18	517.52-520.52	79.87	83.95	83.05-85.95	4.75	7.77	-1.38	63
GS-3	6	350	653.29	646.18-656.1	100.51	150.20	149.35-150.75	7.11	10.94	1.63	62
GS-4	8	18	35.30	31.62-40.53	5.43	2.55	2.35-2.75	4.93	7.36	1.32	5
NiR-1	2	108	337.51	334.5-342.38	51.92	15.05	14.95-15.65	5.45	9.27	10.79	31
NiR-2	4	203	490.66	485.86-491.81	75.49	50.00	44.00-53.1	5.54	8.61	-10.14	185
NiR-3	9	87	220.16	216.68-223.61	33.87	11.95	11.65-12.85	6.10	9.77	-11.65	40
NR-1	4	341	675.51	667.96-680.8	103.92	167.80	167.45-169.9	4.50	7.89	0.22	65

Table 3.2 continued.

QTL name	Chr[a]	Marker[b]	G Pos (cM)[c]	G Interval (cM)[d]	Adj (cM)[e]	P Pos (Mb)[f]	P Interval (Mb)[g]	LOD	R^2 (%)	Add effect[h]	# Genes[i]
PEPC-1	3	44	181.86	164.25-191.94	27.98	5.50	5.25-5.85	4.40	7.09	12.40	29
PEPC-2	5	303	562.79	561.1-568.02	86.58	156.25	151.80-157.05	5.53	9.04	-13.89	129
PEPC-3	6	41	82.21	78.15-87.06	12.65	6.75	6.35-6.95	4.24	6.81	12.18	21
PEPC-4	10	302	691.34	688.4-695.71	106.36	147.35	147.15-147.85	4.82	7.81	13.10	49
Nitrate-1	4	418	832.99	831.24-838.17	128.15	187.35	187.25-187.45	4.37	6.62	15.69	4
Nitrate-2	6	139	195.53	193.93-197.9	30.08	91.45	90.20-91.75	10.41	15.19	23.09	52
Nitrate-3	7	184	357.61	349.92-362.09	55.02	116.45	108.95-118.55	4.71	6.49	-17.06	205
Glutamate-1	5	85	266.85	265.2-269.13	41.05	10.15	10.05-10.25	5.22	7.74	1.46	11
Glutamate-2	6	57	105.01	99.94-112.77	16.16	13.95	11.80-23.60	4.48	6.58	1.31	278
PPDK-1	1	463	845.49	845.09-847.08	130.08	191.85	191.75-192.05	5.16	7.90	-9.57	9
Rubisco-1	9	243	516.74	515.54-517.53	79.50	122.65	122.55-122.75	6.10	9.60	5.83	1
Lox6-1	2	33	98.63	94.96-102.89	15.17	4.25	4.15-4.45	8.35	12.00	-3.30	32
Lox6-2	7	23	86.51	79.08-90.51	13.31	4.05	3.95-4.20	6.77	9.00	3.02	8
Lox6-3	10	185	346.43	340.37-344.59	53.30	127.35	127.15-127.65	7.04	9.00	3.03	24
$PEPC_e$-1	1	460	839.98	838.21-843.40	129.23	191.55	191.30-191.75	6.36	10.49	-11.34	15
$PEPC_e$-2	9	242	501.24	499.20-501.44	77.11	122.65	122.45-122.75	8.49	13.02	22.05	5
$PEPC_e$-3	10	117	275.85	273.07-280.88	42.44	70.70	68.85-76.95	4.14	5.88	-8.74	201
TN_s-1	1	51	163.97	155.61-166.82	25.23	8.35	7.95-8.45	10.47	16.17	-2.42	42
TN_s-2	7	311	603.06	599.58-607.5	92.78	160.65	160.55-160.90	4.22	5.98	1.59	14
TN_r-1	1	195	469.95	467.09-480.29	72.30	37.95	37.15-39.35	4.38	6.04	-0.28	79
TN_r-2	6	235	451.99	449.76-457.93	69.54	115.45	115.10-117.50	4.54	6.27	0.29	100
N_{ratio}-1	2	119	349.98	344.92-359.17	53.84	16.25	15.85-16.65	4.37	6.39	-0.07	40
N_{ratio}-2	3	212	493.48	490.50-497.62	75.92	133.50	129.25-136.95	7.03	10.68	-0.09	185

[a] Chromosome number, [b] Marker localized at LOD peak, [c] Genetic position of SNP in cM, [d] 1-LOD interval in cM, [e] Adjusted genetic position, [f] Physical position in Mb, [g] 1-LOD Physical interval, [h] Additive effect of respective QTL (a positive-signed effect represents an increasing allele from B73, while a negative-signed allele denotes an increasing allele from Mo17), [i] Total number of annotated genes underlying 1-LOD QTL region.

Table 3.3. Analysis of multiple QTL model for N-metabolism related traits measured in leaf tissues of the IBMSyn10-DH TC population of maize.

Phenotype	# QTL in model[a]	Model R^2 (%)[b]	QTL interacting[c]	R^2 epistasis (%)[d]
AlaAT	2	7.25		
AS	5	31.55		
AspAT	2	10.49		
GOGAT	4	27.09		
GS	4	21.10		
NiR	3	22.42		
NR	1	7.89		
PEPC	4	24.47	PEPC-1:PEPC-3	5.00
Nitrate	3	31.43	Nitrate-2:Nitrate-3	2.50
Glutamate	2	18.13		
PPDK	1	7.90		
Rubisco	1	9.60		
Lox6	3	26.73		
$PEPC_e$	3	17.78		
TN_s	2	12.77		
TN_r	2	8.12		
N_{ratio}	2	12.85		

[a] Number of significant QTL fitted in MIM model, [b] Total R^2 obtained by fitting significant QTL simultaneously in a MIM model, [c] Significant epistasis between QTL, [d] R^2 explained by epistasis solely.

Table 3.4. Candidate genes related to N-metabolism underlying QTL genomic regions in leaf tissue of the maize IBMSyn10-DH TC population.

Maize GDB ID	Corresponding gene annotation	Chr[a]	Start[b]	End[c]	QTL name
GRMZM2G008714	Piruvate kinase	10	147664124	147668582	PEPC-4
GRMZM2G045171	Sucrose synthase	4	168773364	168776492	NR-1
GRMZM2G082780	PEPC 4	3	29056230	29064249	Asn Syn-2
GRMZM2G088235	Urease protein	5	83898114	83902364	GS-2
GRMZM2G155974	Glutathione synthetase	3	133812995	133826187	N_{ratio}-2
GRMZM2G028574	PEPC 3	2	115914515	115915086	TN_r-2
GRMZM2G166366	Aspartate kinase 1	6	115555315	115557026	TN_r-2
GRMZM2G180625	Glyceraldehyde-3-phosphate dehydrogenase	6	6901483	6906034	PEPC-3
GRMZM2G343519	Glutaredoxin protein	10	73172286	73173446	PEPCe-3
GRMZM2G402582	PPDK	10	74699777	74700071	PEPCe-3
GRMZM2G481529	Cytosolic enolase, phosphopyruvate hydratase	1	38637579	38641262	TN_r-1

[a] Chromosome, [b,c] start and end location in bp.

CHAPTER 4: MAPPING OF QTL FOR N-METABOLISM RELATED ENZYMES AND METABOLITES IN A MAIZE TESTCROSS POPULATION GROWN IN HYDROPONICS: II. ROOT TISSUE ANALYSIS

A paper to be submitted to Plant Physiology

Ignacio Trucillo-Silva[1], Michael Lee[1]*, Hari Kishan R. Abbaraju[2], Lynne Fallis[3], Hongjun Liu[4], and Kanwarpal S. Dhugga[5]

[1] Present address: Department of Agronomy, Iowa State University, Ames, IA, 50011, USA.
[2] Present address: Monsanto Company, St. Louis, MO, 63167, USA.
[3] Present address: Trait Discovery & Technology, DuPont Pioneer, Johnston, IA, 50131, USA.
[4] Present address: Sichuan Agricultural University, Chengdu, 611130, China.
[5] Present address: CIMMYT, Apartado Postal 6-641-06600, Mexico D.F, Mexico.

***** Corresponding author; mlee@iastate.edu.

Abstract

Nitrogen (N) availability is essential for plant growth and development. During last decades, several problems have arisen due to over-fertilization with N in rural areas. Breeding for maize with greater efficiency in the use of N may help to reduce contamination and increase profits. Nevertheless, previous to breeding, a better understanding of the genetics underlying N-metabolism will be needed. Herein, a quantitative trait loci (QTL) mapping for N-metabolism related enzymes and metabolites was performed based on root tissue harvested from maize hybrids grown in hydroponic conditions. Twenty-six QTL were identified across all traits. QTL models explained 7-43% of the observed variance and no

significant epistasis was detected between QTL. A total of 14 candidate genes were proposed underlying 1-LOD QTL confidence interval regions. All the candidate genes are located *in trans*, unlinked or even in different chromosome, to the known genomic positions of each correspondent structural genes.

Introduction

Nitrogen (N) is one of the most important mineral nutrients for plant growth and development. In maize, sufficient N is required for amino acid metabolism, ear growth, and dry matter accumulation in kernels (Hirel et al., 2001). While N deficiency could result in a substantial decrease in grain yield (Uhart and Andrade, 1995), the oversupply of N causes a severe negative impact in the environment.

A significant proportion of the N added to soils is not uptake and utilize by plants and is lost to the environment. Important causes of N-loss are denitrification of the nitrate form by soil bacteria, volatilization of surface-applied urea-based fertilizers (Nielsen, 2006), and N-leaching and runoff. N contamination from the Mississippi River Basin has been implicated as one of the main causes for the overgrowth of algae that consumes the oxygen needed to support marine life, which develops each spring and summer on the Louisiana-Texas shelf of the Gulf of Mexico (Goolsby and Battaglin, 2000). Nitrate concentrations have increased several fold during the past 100 years in streams of the Basin, and the annual delivery of nitrate from the Mississippi River to the Gulf has nearly tripled since the late 1950's. According to the Louisiana Universities Marine Consortium, the hypoxic area, also called “dead zone”, is bigger than the states of Connecticut and Rhode Island combined, a 28% larger than last year (Schleifstein, 2015).

Improving N use efficiency (NUE) of maize would reduce N losses from the soil. NUE, which in cereals has been defined as the ratio of grain produced per unit of soil N, can be subdivided into two main components: N uptake efficiency (total plant N/soil N) and N utilization efficiency (total grain N/total plant N) (Moll et al., 1982; Dhugga and Waines, 1989). Breeding maize efficient in the use of N may render a more sustainable agriculture, leading to diminish in N fertilization while maintaining yields and an overall increase in profits. But, previous to breeding, the development of a comprehensive understanding of N-metabolism at the genetic level may be helpful or even necessary.

The pathway for N reduction and incorporation of reduced N into organic molecules has been described (Yemm and Folkes, 1958; Lea et al., 1990; Lea and Miflin, 2010) (Fig. 4.1). Nitrate is reduced to nitrite by nitrate reductase (NR) in the cytoplasm, followed by reduction of nitrite in the plastids to ammonium by nitrite reductase (NiR). Ammonium thus generated is aminated into glutamine from glutamate by glutamine synthetase (GS). Another enzyme, glutamine-2-oxoglutarate aminotransferase or glutamate synthase (GOGAT), then converts glutamine back to glutamate, producing an additional glutamate along the way from 2-oxoglutarate. Asparagine synthase (AS) produces asparagine and glutamate from glutamine and aspartate. Glutamate can serve as an amino donor for other amino acids, a reaction accomplished by different amino transferases. For instance, alanine aminotransferase (AlaAT) catalyzes the amino transfer to pyruvate resulting in 2-oxoglutarate and alanine (Miyashita et al., 2007), while aspartate aminotransferase (AspAT) forms 2-oxoglutarate and aspartate after transferring the amino group of glutamate to oxaloacetate. Following N assimilation, glutamate, asparagine, glutamine and other amino acids, constituents of

proteins, are transported via vascular tissues to the growing organs or stored, as vegetative storage proteins, which can aid plant growth during periods of N deficiency.

N and carbon (C) metabolisms are highly interconnected (Nunes-Nesi et al., 2010). Certain metabolites and enzymes perform key roles in C metabolism and are regulated by the status of N in the cell (Sugiharto et al., 1990). Oxaloacetate, one of the C skeletons utilized in amino acids synthesis, is made from the addition of bicarbonate to phosphoenol pyruvate (PEP) by a reaction catalyzed by phosphoenol pyruvate carboxylase (PEPC). Even though N and C metabolisms are essential for life, and several key enzymes and chemical reactions were determined, the genetic basis underlying the plant's ability to uptake and utilize N not completely understood and information is limited.

Mapping of quantitative trait loci (QTL) is routinely implemented in plant breeding programs. Linkage mapping allows the estimation of the mean and variance associated with a specific locus. The procedure relies on differences among the trait means of genotypes at a marker locus (Bernardo, 2010). The precision in the identification of a QTL can be critical to the time, expense, and probability of success of further studies (e.g., identification of candidate genes and positional cloning) (Remington et al., 2001). That precision in the estimation of the QTL position, referred as resolution, may vary substantially depending on several factors such as recombination frequency present in the mapping population, marker density and population size (Yu et al., 2011). The genomic region defined by a QTL could contain one or several genes. Thus, it is not a straightforward process to identify the genes underlying a QTL. Nevertheless, based on previous annotations and descriptions on model species, a few candidate genes could be proposed for further investigation.

Much of today's commercial maize germplasm originates from seven progenitor lines, including B73 and Mo17 (Mikel and Dudley, 2006). Both inbreds differ in their response to N fertilization (Balko and Russell, 1980) and are parents of the IBM (Intermated B73 x Mo17) mapping population (Lee et al., 2002). After ten rounds of random mating, 360 double haploid (DH) lines were generated from the IBMSyn10 population (Hussain et al., 2007) resulting in a higher-resolution mapping population that can be directly associated to the physical map established for B73 inbred (www.maizesequence.org). On the whole, this population serves as an outstanding resource for mapping studies in order to increase the understanding of the genetic basis of N-metabolism.

Several studies have shown association between QTL and N-metabolism related enzymes. Agronomic and physiological traits were used to detect QTL and determine their causal relationships in an integrated manner (Gallais and Hirel, 2004). Agronomic traits were measured in a set of hybrids by Bertin and Gallais (2001), however physiological traits were studied at the level of lines (77 RIL). Limami et al. (2002) studied 140 RIL and identify QTL of germination efficiency that co-localized with genes encoding cytosolic GS. Zhang et al. (2010) measured activity of ten enzymes involved in C and N-metabolism on leaf tissues, based on the IBMSyn4 maize population. Seventy-three QTL associated with enzyme activity and eight QTL associated with biomass were identified. Most of the enzyme activity QTL was *in trans* to the known genomic locations of genes but, three cis-QTL were located for NR, glutamate dehydrogenase and shikimate dehydrogenase. Recently, a QTL analysis based on leaf tissues was performed for 12 metabolites directly related to C- and N-metabolism in the maize nested association mapping (NAM) population. An association mapping approach was implemented and 101 candidate genes were identified (Zhang et al.,

2015). However, QTL associated with enzymes related with N-metabolism from root tissues, at hybrid level, and based on a high-resolution mapping population were not studied. Whilst the chemical reactions occurring within the root system are essential for N-acquisition, the vast majority of enzyme QTL studies were merely focused on leaf tissues. Hence, an investigation devoted to enzyme activity and metabolites related to N-metabolism on root tissues may provide additional insight into these aspects of N-metabolism.

In this investigation, root tissues from a mapping population of maize TC genotypes, derived from the cross between IBMSyn10-DH lines and an elite inbred, grown under hydroponics, were analyzed for enzyme activity and metabolites related to N-metabolism. Key genetic regions associated with enzyme activity and biochemical compounds were identified. Following QTL detection, confidence interval regions (1-LOD CI) were assessed for the identification of candidate genes associated with N-metabolism for further investigation. To our knowledge, this is the first report describing a QTL analysis for N-metabolism related enzymes and metabolites in maize root tissues.

Materials and Methods

Plant material

A total of 176 TC genotypes derived from the cross between each IBMSyn10-DH line and an elite inbred were used. The IBMSyn10-DH population, developed by Hussain et al.(2007), is a set of DH lines derived from a population after ten generations of random mating from the cross between B73 x Mo17. Each DH line was crossed by an elite inbred (PEI), property of DuPont Pioneer (closed pedigree), to generate the TC genotypes.

Experimental design

Kernels from each TC genotype were germinated in autoclaved paper rolls and sterilized water, and subsequently grown under hydroponic conditions. Ten tanks (i.e., sets) containing appropriate growth media were planted with a total of 264 seedlings in each tank. In every set, 22 genotypes were grown, and each genotype was replicated 12 times. Two genotypes (B73 and Mo17 each crossed to the PEI) served as controls, and were included in every set and replication.

The growth media consisted of $MgSO_4$.7H2O 0.5 mM, KH_2PO_4 0.5 mM, Fe-EDTA 0.1 mM, FeEDDHA 0.1 mM, $Ca(NO_3)2.4H_2O$ 1.25 mM, KNO_3 2.5 mM, Na(OH) 0.1 mM, and 0.4 L of trace elements (25 mM H_3BO_3, 2 mM $MnSO_4.H_2O$, 2 mM $ZnSO_4.7H_2O$, 0.5 mM $CuSO_4.5H_2O$, 0.5 mM $Na_2MoO_4.2H_2O$ and 50 mM KCl) in a total of 400 L solution per hydroponic tank. Two weeks after planting, the six most representative uniform plants of each genotype, based on both their root and shoot development, were selected and transplanted into another hydroponic tank with same media.

When plants reached V4 stage (Abendroth et al., 2011), root samples were taken and stored at -80°C while the rest of the plant tissues were dried for 12 days at 48°C.

Biochemical assays

Activity of eight enzymes related with N-metabolism pathway was determined in root samples of each genotype. The set of enzymes comprised NR, NiR, GS, GOGAT, AlaAT, AS, AspAT and PEPC, and specific protocols were adapted by K. Dhugga, R. Abbaraju and L. Fallis. GS, GOGAT, Asp AT and PEPC assay protocols were adapted from Gibon (2004), while NR from Lea et al. (1990), NiR from Bourne and Miflin (1973), AS from Joy and Ireland (1990), and AlaAT protocol was modified from Ashton et al. (1990). Metabolites

nitrate and glutamate were measured as byproducts of enzyme reactions. All measurements were determined by comparing absorbance of each specific biochemical reaction with known standards using a spectrophotometer (Spectramax Plus 384 Microplate Reader, Molecular Devices).

Plant tissues were weighed and analyzed for N content by combustion analysis. Based on biomass dry weight and percentage of N measurements, total amount of N present in root (TN_r) tissues was calculated. In addition, N_{ratio} was estimated as the ratio between TN_s and TN_r. The analysis of TN_s is presented in Trucillo-Silva et al. (2015).

Trait data analysis

Statistical analysis was implemented in R statistical program (RCoreTeam, 2014). Initial data analysis of the raw data was based on the ggplot2 package (Wickham, 2010) and GGally (Schloerke et al., 2014). As a first step, a univariate analysis, where a single variable is fitted in a model, followed by a multivariate approach, where multiple variables are analyzed simultaneously, was performed in order to comprehend the relationship among the variables. The determination of outliers present in the dataset, based on a jackknife resampling strategy, was applied. As described in Trucillo-Silva (2015), a statistical model is fitted n times, systematically omitting one observation from the dataset, followed by the prediction of random effects for a subset of most consistent genotypes each of the n times. The aim of process is to target "real outliers" based on the complete information gathered in the experiment and fine-tune the statistical model, quantified by improvements in log-likelihood, Akaike and Bayesian information criterion values after discarding misleading observations, while keeping informative and true observations for later analysis. The mixed

model was fitted with ASReml R package (Butler et al., 2007) and correspondent mixed model equations were solved for prediction of random effects and estimation of fixed effects. The statistical model can be represented as follows:

$$y = Xb + Zu + e$$

Where y denotes a n x 1 vector of observed response values, b is a p x 1 vector of fixed effects, X is a n x p design matrix, u is a q x 1 vector of random effects, Z is a n x q design matrix, and e being the error term.

The following assumptions were used: E (u) = 0, E (e) = 0, Cov (u, e) = 0, and Var (u) = G and, Var (e) = R. The G matrix had a compound symmetry structure on the genotype levels and R matrix is a diagonal matrix with different values for each set, allowing non-constant variance across sets. The response variable was the activity of the enzyme and the metabolite concentration, respectively. Set, the light replicate and plate were included as fixed effects in the model (where replicate and plate are nested in set), and check genotype effect was included as a continuous covariate. Finally, a random effect for the genotype was included in the linear model. During the process describe above, several genotypes were discarded separately for each trait. Extreme cases were AlaAT and NR were five genotypes were discarded, respectively; whereas no genotypes were taken out of the analysis and sample size totalized 176 genotypes for AspAT, AS, and GS. Furthermore, one and four complete sets of data were removed for glutamate and nitrate, respectively, due to contamination of samples and very low accuracy in the estimations.

Significance of genetic variance was calculated based on log-likelihood ratio test by comparing models with and without the TC random effect. Correlation was calculated among BLUP values for each pair of traits and significance was adjusted after Bonferroni correction

for multiple comparisons. Repeatability was derived from variance estimations from ASReml. As variance components were estimated for each different set, a different value of repeatability was estimated for each set and then partial estimates were averaged correspondingly.

Genotypic information and genetic maps

TC genotypes were analyzed with a total of 5,306 single nucleotide polymorphism (SNP) markers generated by the Beijing Genomics Institute. Physical and genetic position of each SNP was determined and genetic maps were created using R/qtl (Broman et al., 2003). Recombination fractions were estimated and Kosambi mapping function was implemented to calculate genetic map distances (Kosambi, 1944). Furthermore, as the recombination between linked loci increases every generation, leading to an expansion of the genetic map, mapping distances were adjusted with the purpose of comparison with previous investigations. The expansion factor was determined based on the following equation: $\propto = \frac{j}{2} + (2\mathrm{i} - 1)/2\mathrm{i}$, where j corresponds to the number of generations of intermating including the two generations for generating the F_2, and i is the number of inbred generations after intermating (Teuscher et al., 2005).

The real map was 11,265.25 cM length and map distances were reduced by a factor of 6.5 to estimate the adjusted F_2 map. The final adjusted map was 1,733.12 cM length with an average spacing between markers of 0.33 cM, while the maximal spacing between markers was nearly 7 cM, located in chromosome 6 (Fig. 4.3). With regard to physical distance, the length of the total genome was 2,051.75 Mb and on average there was a marker positioned every 400Kb. The biggest gaps between markers, 69.80 and 67.40 Mb, were located in chromosomes 2 and 9, respectively.

QTL mapping and identification of candidate genes

Associations between phenotypes and genotypes were determined using QTL Cartographer (Basten et al., 2002). Single marker analysis, followed by linear regression analysis and composite interval mapping (CIM) was performed. For CIM, Zmap (model 6) was implemented, using the ten most significant marker cofactors identified by forward and backward regression. In addition, QTL were scanned at intervals of 1 cM and at every marker while cofactors located within a window of 10 cM of the scanned position were excluded from the analysis. In order to determine LOD score thresholds of 5%, and to further identify significant QTL, 1,000 permutations were performed for every trait. Two nearby QTL were considered as different when LOD peaks were localized 20 cM or greater apart. Effects of QTL are expressed relative to the B73 allele. Therefore, a positive effect would imply an increase in the phenotypic value when the B73 allele is present, whereas a negative effect would indicate a reduction in the presence of B73 allele.

Furthermore, a multiple interval mapping (MIM) analysis was performed by fitting all previously identified QTL from CIM analysis, and parameters were re-estimated and positions refined. In addition, all pairwise interactions between QTL in every model were studied for each trait. The significance was determined based on the information criterion: IC (k) = -2 (log (L) - kc (n)/2), where the penalty function corresponds to: c (n) = log (n) and a threshold of 0.0 was used (Basten et al., 2002). The proportion of the total phenotypic variance associated with each model was estimated.

Candidate genes annotated on corresponding 1-LOD QTL confidence interval regions were examined from MaizeGDB (Lawrence et al., 2008) and Phytozome (Goodstein et al., 2012). Those candidate genes directly related to N-metabolism based on their descriptions on

model species, such as rice (*Oryza sativa*) and Arabidopsis (*Arabidopsis thaliana*), were proposed for further studies. Several other candidate genes may be promising candidates for further investigations, including transcription factors; however were not considered due to the difficulties to ascertain a direct relationship with N-metabolism in maize based on available descriptions.

Results

Statistical analysis for N-metabolism associated traits

Genetic variance was statistically significant for all traits and a wide range of values was obtained across traits. Repeatability values averaged 0.52, with values ranging from 0.38 - 0.70. The lowest value of repeatability was registered for AspAT, while the highest value corresponded to TN_r. Coefficient of variations ranged from 0.03 to 0.48 (for N_{ratio} and NR, respectively) (Table 4.1).

Correlation between N-metabolism related traits

Pearson correlation values were estimated between all traits (Fig 4.2). From a total of 66 pairwise correlations, 31% were extremely significant (p-value<0.001) and 4% were significant (p-value<0.05). Significant correlation values ranged from -0.43 to 0.55. The highest significant correlation was determined between AlaAT and AspAT and all significant correlations between enzyme activities, enzymes and metabolites, and between metabolites were positive. Close negative correlations were estimated between TN_r and N_{ratio} (-0.43) and N_{ratio} and GOGAT (-0.27).

Identification of quantitative trait loci

Twenty-six QTL were identified across all traits. Even though QTL were identified in all chromosomes, five QTL were detected in chromosome 7 while one QTL was identified in

chromosome 8 (Fig. 4.3). AlaAT-3 and AspAT-2 were the only QTL found to overlap their respective 1-LOD CI at chromosome 10. On average, 2.2 QTL were identified per trait, ranging from one QTL for some traits (GOGAT and NR) to four QTL for NiR. Most of the QTL (69%) explained less than 10% of the genetic variance, while 27% and 4 % of the identified QTL were associated with 10-25% and >25% of the variance, respectively. The QTL which accounted the highest amount of variance (31.5%) and presented the highest LOD score (23.4) was PEPC-1 QTL, located in chromosome 5. For that QTL, B73 allele showed a negative effect (-23.78 uMole NADH/min/mg protein). Furthermore, at 70 % of all QTL detected across traits, B73 showed a negative additive effect. For certain traits, B73 exhibited only a negative effect (e.g., AspAT, GOGAT, and GS) while for AS QTL only positive effects for B73 alleles was found.

Confidence intervals (CI 1-LOD) for QTL ranged from 1.04 - 24.46 cM (0.16 - 3.76 cM adjusted distance) length, with an average of 7.79 cM (1.2 cM adjusted distance). Those CI correspond to 0.2 to 21.1 Mb in physical distance, with a mean CI length of 2.46 Mb.

Multiple interval mapping

First order epistatic interactions between QTL identified previously by CIM were not significant for all traits, thus epistatic digenic effects were excluded from genetic models. Even though 43% of the total variance was explained in PEPC by fitting two QTL in a MIM model, other genetic models captured less than 10% of the phenotypic variance, such as AS, GOGAT, GS and NR (Table 4.3). On average, multiple QTL models explain 15.1 % of the phenotypic variance and two QTL were included in each of the models.

Candidate genes

On average of 63 genes were annotated underlying QTL 1-LOD regions, with CI regions having between six and 376 genes. Nevertheless, only a subset of the putative genes could be associated to N-metabolism pathway based on the description in model species. The most promising genes may be GRMZM2G028574, GRMZM2G111225, GRMZM2G136712, GRMZM2G155974, GRMZM2G166366, GRMZM2G374302, GRMZM2G409131, GRMZM2G466543, GRMZM2G473001, GRMZM2G481529, GRMZM2G493395, GRMZM5G817058, GRMZM2G575696 and GRMZM2G580894 (Table 4.4). Each of them had shown important putative functions, as PEPC, nitrilase, aspartate kinase, glutathione synthetase, aspartate kinase, arginine decarboxylase, phosphofructokinase, arogenate dehydratase, PEPC, phosphopyruvate hydratase, 1-deoxy-D-xylulose-5-phosphate synthase, phosphoribosyl transferase, and last two genes as S-adenosyl-methionine-dependent (SAM)-methyltransferase, respectively. In accordance with Trucillo-Silva (2015), all the QTL identified in this study are located on a different position to the known genomic location of each corresponding structural gene (e.g., GS QTL were identified at chromosomes 7 and 9 in this study, whereas GS1 and GS2 locus are located in chromosome 1 between 271.02-273.438 Mb and on chromosome 2 between 18.94-19.46 Mb, based on the following nearest loci on the IBM2 2008 Neighbors map, respectively). Therefore, the candidate genes identified under the QTL regions seem to affect in a *trans*-acting regulatory manner as previously described (Zhang et al., 2010). Proposed candidate genes are located in chromosomes 1, 2, 3, 4, 6, and 7. In addition, no candidate genes were proposed underlying QTL for AS, GS, NiR, PEPC, Nitrate and Glutamate.

Discussion

In this first genetic mapping investigation for N-metabolism related enzymes and metabolites in root tissues in plants, 22 QTL were identified in a TC mapping population. QTL models explaining more than 20% of the genetic variance were determined for certain phenotypes, such as NiR and PEPC. Since a higher-resolution mapping population and a high number of molecular markers were employed, the results are expected to be more precise and accurate compared to previous QTL studies on N-metabolism in leaf tissues. The findings would help to increase the knowledge of the genetics underlying N-metabolism in maize hybrids.

Even though numerous QTL associated with enzymes involved in N-metabolism were identified in previous studies (Agrama et al., 1999; Limami et al., 2002; Canas et al., 2012), only a few investigations were based on a representative and high-resolution mapping population, such as Zhang et al. (2010) and Zhang et al. (2015). Despite the fact that most traits in maize had shown low correlations between performances in the inbred and hybrid progeny (Hallauer et al., 2010), relatively few studies focused on hybrid populations (Bertin and Gallais, 2001; Gallais and Hirel, 2004). In order to fine map and account for the higher recombination rate, a high dense SNP marker platform was employed (5,300 SNP markers). Therefore, a higher-mapping resolution for a QTL analysis for root phenotypes was accomplished.

In accordance with previous studies, the activity of enzymes investigated, constituents of the N-metabolism pathway (except PEPC, member of the primary C-metabolism, albeit closely related to N-metabolism), seem to be co-regulated (Zhang et al., 2010; Trucillo-Silva et al., 2015). Hence, a positive correlation between enzyme activities, as well as within

metabolites concentration, was expected *a priori* and was confirmed (Fig. 4.2). In addition, significant correlations between enzyme activities and metabolites were as well positive. Furthermore, the correlation value between the two metabolites (nitrate and glutamate) was significant and showed almost exact value as in a previous investigation (Zhang et al., 2015). Even though all correlations between enzyme activities were positive, some of the correlation values determined in this investigation were not statistically significant in Trucillo-Silva et al. (2015) (e.g., between AlaAT and AspAT, AS and GS, and AS and both metabolites). The highest correlation value was estimated between AlaAT and AspAT (0.55), and the only co-location of QTL was as well between AlaAT-3 and AspAT-2 at chromosome 10. Even though no candidate genes associated with N-metabolism were found within that interval, it may be considered an important genomic region. Negative correlation was calculated between N_{ratio} and TN_r, values expected based on how N_{ratio} was estimated (TN_s/TN_r). In addition, GOGAT and N_{ratio} showed as well a negative correlation, which further supports the positive relationship between GOGAT activity and the accumulation of N in root tissues.

In comparison to previous studies (Zhang et al., 2010; Trucillo-Silva et al., 2015), in which leaf tissue was investigated, the analysis and determination of root enzyme activity is even more complex and laborious in order to obtain accurate and reliable data. Activity of root enzymes is more susceptible to fluctuation, due to several factors including the procedure employed for cleaning the roots previous to sampling, compared to enzyme activity in leaf tissues. Even though a protocol for sampling and cleaning roots was established and applied, six replications per genotype were included, and many other influencing factors were taken into account (e.g., uniform temperature and harvesting window period), repeatability estimations for many traits were relatively low (0.38-0.70)

compared to similar studies based on leaf tissues (e.g., repeatability for NR was 0.62 compared to 0.65 and 0.74 for leaf tissue analyses performed by Trucillo-Silva et. al. (2015), and Zhang et al., (2010), respectively).

QTL associated with eight enzyme activities, two metabolites and two N content traits were identified. A few QTL determined in this root study were likewise identified in analogous position in a previous QTL analysis on leaf tissues (Trucillo-Silva et al., 2015), such as a QTL associated with AS located on chromosome 5, and QTL for PEPC, nitrate and GOGAT (LOD peak values identified 2, 4 and 7 adjusted cM apart, respectively). In agreement with Zhang (2010), a QTL for AlaAT was detected in chromosomes 4, about 5 cM apart from the detected position in this study. Nonetheless, most of the QTL reported in other maize studies (Agrama et al., 1999; Hirel et al., 2001; Canas et al., 2012), fail to co-locate, were greater than 20 cM apart or even on different chromosomes, with the QTL identified in this investigation (e.g., QTL for GS activity were determined on chromosomes 7 and 9 in this study, whereas on chromosomes 4 and 5 in Canas et al., 2012).

A lower number of QTL was identified per trait compared to previous investigations (Zhang et al., 2010; Trucillo-Silva et al., 2015) based on leaf tissues. One main difference compared to those studies is that all phenotypes were measured solely on root tissues in this investigation, suggesting that similar traits are differentially regulated in roots and leaf tissues. Furthermore, the power to identify a QTL would depend essentially on the magnitude of the QTL effect and the size of the segregating population (Beavis, 1998). Because a large number of small-effect QTL segregating in the genome are expected, and due to the size of the segregating population (176 individuals), only a subset of the total number of real segregating QTL are expected to be identified. Moreover, in comparison to Zhang (2010), the

number of QTL detected might have been affected by the six additional generations of random mating during the creation of the population. Hence, QTL previously detected in large linkage blocks, might had been separated into several smaller-effects QTL after further recombination events occurred. Herein, the power to detect a QTL, each with a very small effect, would be expected to be lower. Furthermore, inbred lines were used in Zhang et al. (2010) whereas a TC mapping population is used in this study, and little evidence of common QTL detection between inbred per se and TC progeny has been determined in previous investigations (Beavis et al., 1994; Schon et al., 1994).

All the QTL identified in this study are located *in trans* to the actual position of structural genes. Even though the parents of the mapping population (B73 and Mo17) responded differentially to N-fertilization, those genotypes were not selected specifically to differ for the traits analyzed in this investigation. In accordance to the results, and because of the relevance of the enzymes in the N-pathway, most of the genetic variation present in the population was associated with variation of genes *in trans*; but significant variation in structural genes was not or slightly present. Hence, genomic regions associated with regulatory functions are of much importance in the determination of N-metabolism in this population.

Similar to Trucillo-Silva (2015), the MIM results across traits suggest that there might be several undetected small effect QTL responsible of the rest of the genetic variation (e.g., for PEPC and AS, two QTL explained 42.5% and 8.1% of the variance, respectively). The sum of the effect of numerous QTL, each with small marginal effect, plus any type of epistasis which they might be involved in, should account for all the unexplained genetic variance in the MIM QTL models. It has been established that epistasis can have a large

contribution to the genetic regulation of complex traits (Carlborg and Haley, 2004). However, statistically significant first order epistasis between identified QTL was not detected. Likewise, no significant epistasis between QTL was detected in a recent study based on the maize Nested Association Mapping (NAM) population, which included the parents of this population (B73 and Mo73) (Zhang et al., 2015).

From a total of 60,000 annotated genes across the maize genome, a limited amount was identified under 1-LOD QTL intervals. One of the genes was as well identified in a previous meta-QTL investigation aiming the discovery of candidate genes for N-use efficiency in maize (Liu et al., 2012). That gene is GRMZM2G368398 and is described as a transposon protein in maize, and as an oligopeptide transporter (Yellow stripe-like7) in Arabidopsis. An additional gene (GRMZM2G053958), which code for NAD(P)-binding Rossmann-fold superfamily protein was proposed as a candidate gene in a recent investigation based on C and N metabolism in the NAM population (Zhang et al., 2015). Herein, 14 candidate genes associated with N-metabolism are proposed for further studies. GRMZM2G028574 and GRMZM2G473001 are genes described to have PEPC activity. GRMZM2G111225 is annotated as a nitrilase enzyme, which catalyzes the hydrolysis of nitriles to carboxylic acids and ammonia, and is implicated in auxin biosynthesis in maize (Park et al., 2003). Furthermore, GRMZM2G136712, an aspartate kinase, catalyzes the phosphorylation of aspartate to generate, after a few more reactions, methionine, lysine and threonine. GRMZM2G155974 catalyzes the addition of glycine to γ-glutamyl-cysteine, generating glutathione. Glutathione is a key water-soluble antioxidant, the storage form and long-distance transport form of reduced sulfur (Zagorchev et al., 2013). GRMZM2G166366 was annotated as an aspartate kinase which catalyzes the phosphorylation of aspartate to for

β-aspartyl phosphate, and is responsible for the first step in the biosynthesis of the amino acids lysine, methionine, and threonine (Azevedo et al., 1992). GRMZM2G374302 codes for arginine decarboxylase, a key enzyme involved in the polyamine biosynthesis that decreases in concentration under N-deficiency conditions (Amiour et al., 2012). In addition, GRMZM2G409131 catalyzes the phosphorylation of D-fructose 6-phosphate to fructose 1,6-biphosphate, the entry point into glycolysis to lastly produce pyruvate (Plaxton and Podesta, 2006). GRMZM2G466543 codes for arogenate dehydratase, a gene that functions in the final steps of the aromatic amino acid pathway that produces two essential amino acids: tyrosine and phenylalanine (Holding et al., 2010). GRMZM2G481529 a cytosolic enolase or phosphopyruvate hydratase, is described as a metalloenzyme responsible for the catalysis of the conversion of 2-phosphoglycerate to PEP, necessary for sucrose synthesis from pyruvate in C4 plants (Karpilov et al., 1978), having orthologs within sorghum and rice. GRMZM2G493395 codes for 1-deoxy-D-xylulose-5-phosphate synthase, first step for thiamine and pyridoxol biosynthesis (Hans et al., 2004). GRMZM5G817058 is a phosphoribosyltransferase and acts in amino acid metabolism by catalyzing the first step in the biosynthesis of histidine (Morot-Gaudry et al., 2001). Finally, GRMZM2G575696 and GRMZM2G580894, both S-adenosyl-L-methionine (SAM)-dependent methyltransferases, are responsible of transferring methyl groups from a methyl donor SAM to N, oxygen, sulfur, and C atoms of several biomolecules, such as DNA, RNA, histones, and other proteins. These modifications may affect the expression of a wide variety of genes, signaling, nuclear division, and metabolisms (Bobenchik et al., 2011).

Further investigation is strongly recommended to confirm QTL regions associated with N-metabolism and validate candidate genes underlying those key genetic regions.

Fourteen coding regions are determined as promising candidates for future validation studies. The results accomplished in this investigation, in addition to all previous N-related studies, may help to improve the current understanding of N-metabolism in maize and to identify suitable targets for selection.

Conclusions

The genetics underlying N-metabolism in maize is complex. A promising approach to get insight into the genetic components and decipher the regulatory steps involved in N-metabolism, is by studying key enzymes, in a representative and high-resolution mapping population. In this study, 26 QTL associated with N-metabolism physiological traits were identified after analyzing root tissues from a high-resolution maize TC mapping population, derived from B73 and Mo17. Genetic QTL models accounting for 7 to 43 % of the genetic variance were determined and 14 candidate genes within QTL genomic regions were proposed for further investigation.

Acknowledgments

The authors thank R.F. Baker Center for Plant Breeding – Department of Agronomy – Iowa State University and DuPont Pioneer for making this research possible and Iowa State University undergraduate students (especially Guan Yi Lai) who assisted with planting and harvesting in the hydroponic experiments. Requests for testcross materials and IBMSyn10-DH lines may be directed to DuPont Pioneer and Dr. Michael Lee, respectively.

References

Abendroth LJ, Elmore RW, Boyer MJ, Marlay SK (2011) Corn growth and development. *In*. Extension Publication #PMR-1009, Iowa State University

Agrama HAS, Zakaria AG, Said FB, Tuinstra M (1999) Identification of quantitative trait loci for nitrogen use efficiency in maize. Molecular Breeding **5:** 187-195

Amiour N, Imbaud S, Clement G, Agier N, Zivy M, Valot B, Balliau T, Armengaud P, Quillere I, Canas R, Tercet-Laforgue T, Hirel B (2012) The use of metabolomics integrated with transcriptomic and proteomic studies for identifying key steps involved in the control of nitrogen metabolism in crops such as maize. Journal of Experimental Botany **63:** 5017-5033

Ashton AR, Burnell JN, Furbank RT, Jenkins CLD, Hatch MD (1990) Enzymes of C_4 photosynthesis. Academic Press, London

Azevedo RA, Blackwell RD, Smith RJ, Lea PJ (1992) Three aspartate kinase isoenzymes from maize. Phytochemistry **31:** 3725-3730

Balko LG, Russell WA (1980) Response of maize inbred lines to N-fertilizer. Agronomy Journal **72:** 723-728

Basten CJ, Weir BS, Zeng ZB (2002) QTL Cartographer, Version 1.16. Department of Statistics, North Carolina State University, Raleigh, NC

Beavis WD (1998) QTL analyses: power, precision, and accuracy. *In* Molecular dissection of complex traits, pp 145-162

Beavis WD, Smith OS, Grant D, Fincher R (1994) Identification of quantitative trait loci using a small sample of topcrossed and F4 progeny from maize. Crop Science **34:** 882-896

Bernardo R (2010) Breeding for quantitative traits in plants. *In*, Ed 2nd. Stemma Press, Woodbury, MN

Bertin P, Gallais A (2001) Genetic variation for nitrogen use efficiency in a set of recombinant inbred lines II - QTL detection and coincidences. Maydica **46:** 53-68

Bobenchik AM, Augagneur Y, Hao B, Hoch JC, Ben Mamoun C (2011) Phosphoethanolamine methyltransferases in phosphocholine biosynthesis: functions and potential for antiparasite therapy. Fems Microbiology Reviews **35:** 609-619

Bourne WF, Miflin BJ (1973) Studies on nitrite reductase in barley. Planta **111:** 47-56

Broman KW, Wu H, Sen Ś, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics **19:** 2990-2992

Butler D, Cullis BR, Gilmour AR, Gogel BJ (2007) Analysis of mixed models for S-language environments: ASReml-R reference Manual. Department of Primary Industries and Fisheries, Queensland, Brisbane, Australia

Canas RA, Quillere I, Gallais A, Hirel B (2012) Can genetic variability for nitrogen metabolism in the developing ear of maize be exploited to improve yield? New Phytologist **194:** 440-452

Carlborg O, Haley CS (2004) Epistasis: too often neglected in complex trait studies? Nature Reviews Genetics **5:** 618-625

Dhugga KS, Waines JG (1989) Analysis of nitrogen accumulation and use in bread and durum-wheat. Crop Science **29:** 1232-1239

Gallais A, Hirel B (2004) An approach to the genetics of nitrogen use efficiency in maize. Journal of Experimental Botany **55:** 295-306

Gibon Y, Blaesing OE, Hannemann J, Carillo P, Hohne M, Hendriks JHM, Palacios N, Cross J, Selbig J, Stitt M (2004) A robot-based platform to measure multiple enzyme activities in Arabidopsis using a set of cycling assays: comparison of changes of enzyme activities and transcript levels during diurnal cycles and in prolonged darkness. Plant Cell **16:** 3304-3325

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research **40:** D1178-D1186

Goolsby DA, Battaglin WA (2000) Nitrogen in the Mississippi basin: Estimating sources and predicting flux to the Gulf of Mexico. *In* U.S. Geological Survey Fact Sheet, FS-135-00. U.S. Geological Survey, Reston, Virginia

Hallauer AR, Carena MJ, Filho JBM (2010) Quantitative genetics in maize breeding, Vol 6. Iowa State University Press, Ames, IA

Hans J, Hause B, Strack D, Walter MH (2004) Cloning, characterization, and immunolocalization of a mycorrhiza-inducible 1-deoxy-D-xylulose 5-phosphate reductoisomerase in arbuscule-containing cells of maize. Plant Physiology **134:** 614-624

Hirel B, Bertin P, Quillere I, Bourdoncle W, Attagnant C, Dellay C, Gouy A, Cadiou S, Retailliau C, Falque M, Gallais A (2001) Towards a better understanding of the genetic and physiological basis for nitrogen use efficiency in maize. Plant Physiology **125:** 1258-1270

Holding DR, Meeley RB, Hazebroek J, Selinger D, Gruis F, Jung R, Larkins BA (2010) Identification and characterization of the maize arogenate dehydrogenase gene family. Journal of Experimental Botany **61:** 3663-3673

Hussain T, Tausend P, Graham G, Ho J (2007) Registration of IBM2 SYN10 doubled haploid mapping population of maize. Journal of Plant Registrations **1:** 81-81

Joy KW, Ireland RJ (1990) Enzymes of asparagine metabolism. *In* Methods in plant biochemistry. Academic Press, London, pp 287-296

Karpilov LS, Novitskaia IL, Kuz'min AN, Maslov AI, Popova EI (1978) Reversibility of glycolysis in leaves of C4-plants. Biokhimiia **42:** 1810-1816

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugenics **12:** 172-175

Lawrence CJH, L.C., Schaeffer ML, Sen TZ, Seigfried TE, Campbell DA (2008) MaizeGDB: the maize model organism database for basic, translational, and applied research. International Journal of Plant Genomics **496957:** 1-10

Lea PJ, Blackwell RD, Chen FL, Hecht U (1990) Enzymes of ammonia assimilation. *In* PJ Lea, ed, Methods in Plant Biochemistry, Vol 3. Academic Press Limited, London, pp 257-276

Lea PJ, Miflin BJ (2010) Nitrogen assimilation and its relevance to crop improvement. *In* CH Foyer, H Zhang, eds, Nitrogen metabolism in plants in the post-genomic era, Vol 42. Willey-Blackwell, Chichester, UK, pp 1-40

Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A (2002) Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant Molecular Biology **48:** 453-461

Limami AM, Rouillon C, Glevarec G, Gallais A, Hirel B (2002) Genetic and physiological analysis of germination efficiency in maize in relation to nitrogen metabolism reveals the importance of cytosolic glutamine synthetase. Plant Physiology **130:** 1860-1870

Liu R, Zhang H, Zhao P, Zhang Z, Liang W, Tian Z, Zheng Y (2012) Mining of candidate maize genes for nitrogen use efficiency by integrating gene expression and QTL data. Plant Molecular Biology Reporter **30:** 297-308

Mikel MA, Dudley JW (2006) Evolution of north American dent corn from public to proprietary germplasm. Crop Science **46:** 1193-1205

Miyashita Y, Dolferus R, Ismond KP, Good AG (2007) Alanine aminotransferase catalyses the breakdown of alanine after hypoxia in Arabidopsis thaliana. Plant Journal **49:** 1108-1121

Moll RH, Kamprath EJ, Jackson WA (1982) Analysis and interpretation of factors which contribute to efficiency of nitrogen-utilization. Agronomy Journal **74:** 562-564

Morot-Gaudry JF, Job D, Lea PJ (2001) Amino acid metabolism. *In* PJ Lea, JF Morot-Gaundry, eds, Plant Nitrogen. Springer Verlag, Berlin, pp 167-211

Nielsen RL (2006) N loss mechanisms and nitrogen use efficiency. *In*, Ed Nitrogen management workshops. Purdue Agriculture Agronomy Extension, West Lafayette, IN

Nunes-Nesi A, Fernie AR, Stitt M (2010) Metabolic and signaling aspects underpinning the regulation of plant carbon nitrogen interactions. Molecular Plant **3:** 973-996

Park WJ, Kriechbaumer V, Muller A, Piotrowski M, Meeley RB, Gierl A, Glawischnig E (2003) The nitrilase ZmNIT2 converts indole-3-acetonitrile to indole-3-acetic acid. Plant Physiology **133:** 794-802

Plaxton WC, Podesta FE (2006) The functional organization and control of plant respiration. Critical Reviews in Plant Sciences **25:** 159-198

R Core Team (2014) R: a language and environment for statistical computing. *In* R Foundation for Statistical Computing, Vienna, Austria

Remington DL, Ungerer MC, Purugganan MD (2001) Map-based cloning of quantitative trait loci: progress and prospects. Genetics Research **78:** 213-218

Schleifstein M (2015) 2015 Gulf "dead zone" larger than Connecticut, Rhode Island combined. *In* NOLA.com | The Times-Picayune, New Orleans, LA

Schloerke B, Crowley J, Cook D, Hofmann H, Wickham H, Briatte F, Marbach M, Thoen E (2014) GGally: extension to ggplot2. *In* R package, Ed 0.4.8, Vienna, Austria

Schon CC, Melchinger AE, Boppenmaier J, Brunklausjung E, Herrmann RG, Seitzer JF (1994) RFLP Mapping in maize - Quantitative trait loci affecting testcross performance of elite European flint lines. Crop Science **34:** 378-389

Sugiharto B, Miyata K, Nakamoto H, Sasakawa H, Sugiyama T (1990) Regulation of expression of carbon-assimilating enzymes by nitrogen in maize leaf. Plant Physiology **92:** 963-969

Teuscher F, Guiard V, Rudolph PE, Brockmann GA (2005) The map expansion obtained with recombinant inbred strains and intermated recombinant inbred populations for finite generation designs. Genetics **170:** 875-879

Trucillo-Silva I, Cook D, Lee M (2015) A methodical approach for identifying outliers in complex data. *In* preparation. Iowa State University

Trucillo-Silva I, Lee M, Abbaraju HKR, Fallis L, Liu H, Dhugga KS (2015) Mapping of QTL for N-metabolism related enzymes and metabolites in a maize testcross population grown in hydroponics: I. Leaf tissue analysis. *In* preparation. Iowa State University

Uhart SA, Andrade FH (1995) Nitrogen deficiency in maize: I. Effects on crop growth, development, dry matter partitioning, and kernel set. Crop Science **35:** 1376-1383

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. Journal of Heredity **93:** 77-78

Wickham H (2010) A layered grammar of graphics. Journal of Computational and Graphical Statistics **19:** 3-28

Yemm W, Folkes BF (1958) The metabolism of aminoacids and proteins in plants. Annual Review of Plant Physiology **9:** 245-280

Yu H, Xie W, Wang J, Xing Y, Xu C, Li X, Xiao J, Zhang Q (2011) Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. Plos One **6:** e17595

Zagorchev L, Seal CE, Kranner I, Odjakova M (2013) A central role for thiols in plant tolerance to abiotic stress. International Journal of Molecular Sciences **14:** 7405-7432

Zhang N, Gibon Y, Gur A, Chen C, Lepak N, Hoehne M, Zhang Z, Kroon D, Tschoep H, Stitt M, Buckler E (2010) Fine quantitative trait loci mapping of carbon and nitrogen metabolism enzyme activities and seedling biomass in the maize IBM mapping population. Plant Physiology **154:** 1753-1765

Zhang N, Gibon Y, Wallace JG, Kruger Lepak N, Li P, Dedow L (2015) Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population. Plant Physiology **168:** 575-583

Figures

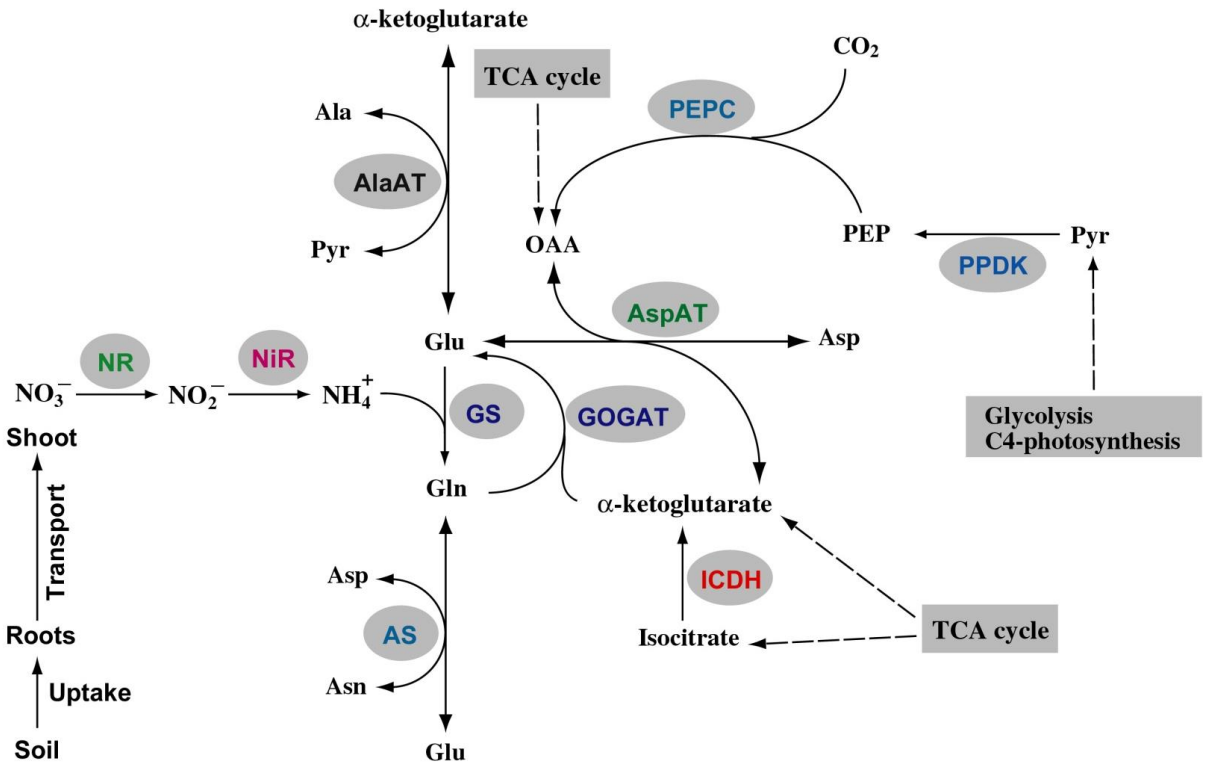

Figure 4.1. Enzymes and proteins involved in N-acquisition and assimilation in higher plants.

AlaAT, alanine aminotransferase; AS, asparagine synthase; AspAT, aspartate aminotransferase; GOGAT, glutamate synthase; GS, glutamine synthetase; ICDH, isocitrate dehydrogenase; NR, nitrate reductase; NiR, nitrite reductase; PEPC, phosphoenolpyruvate carboxylase; PPDK, pyruvate orthophosphate dikinase (Source: Kanwarpal S. Dhugga, 2015).

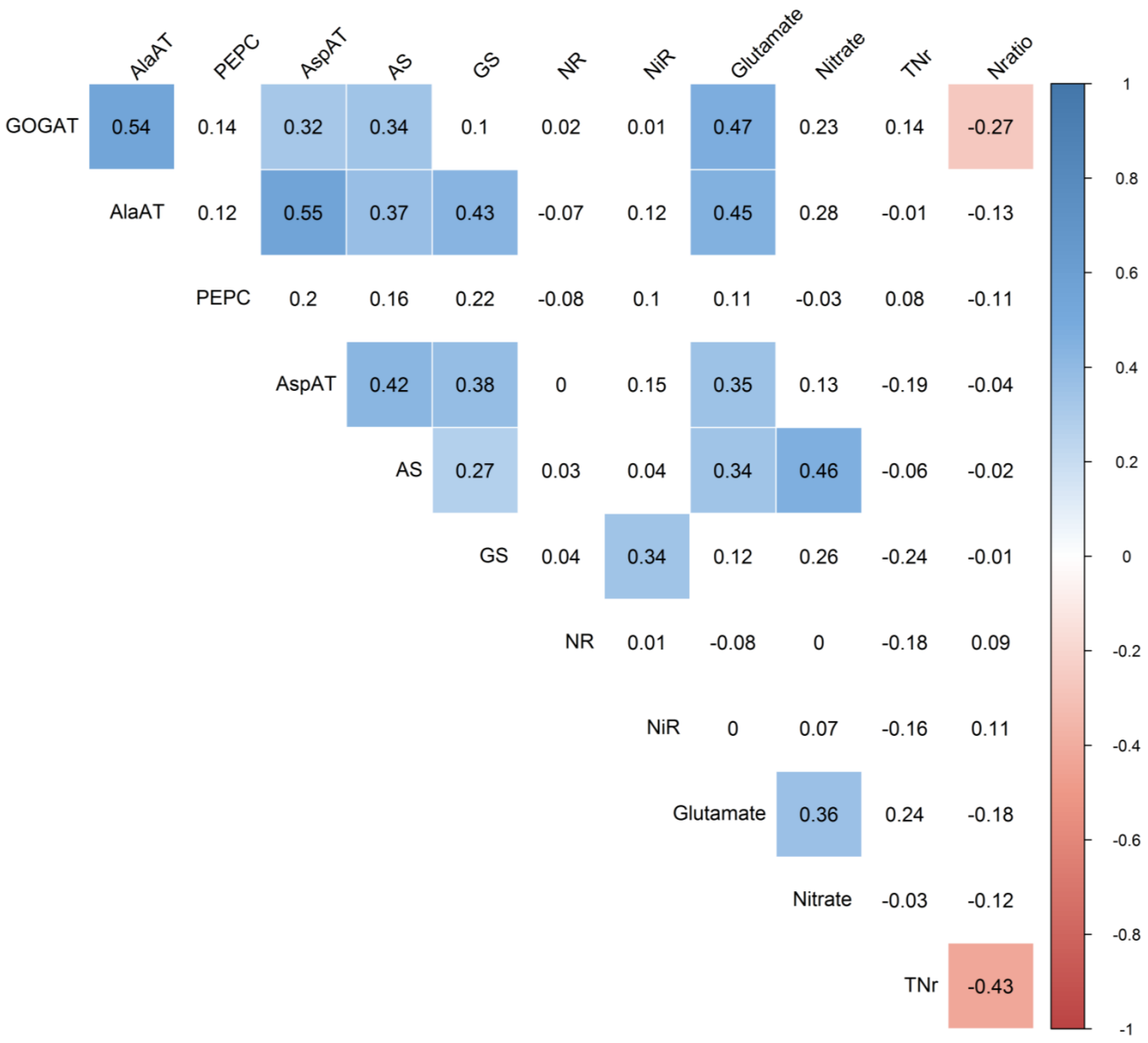

Figure 4.2. Correlation matrix-heatmap of the N-metabolism related enzymes and metabolites measured on root tissues in the maize IBMSyn10-DH TC population.

Significant correlation values are colored in blue (positive correlation) and red (negative correlation)

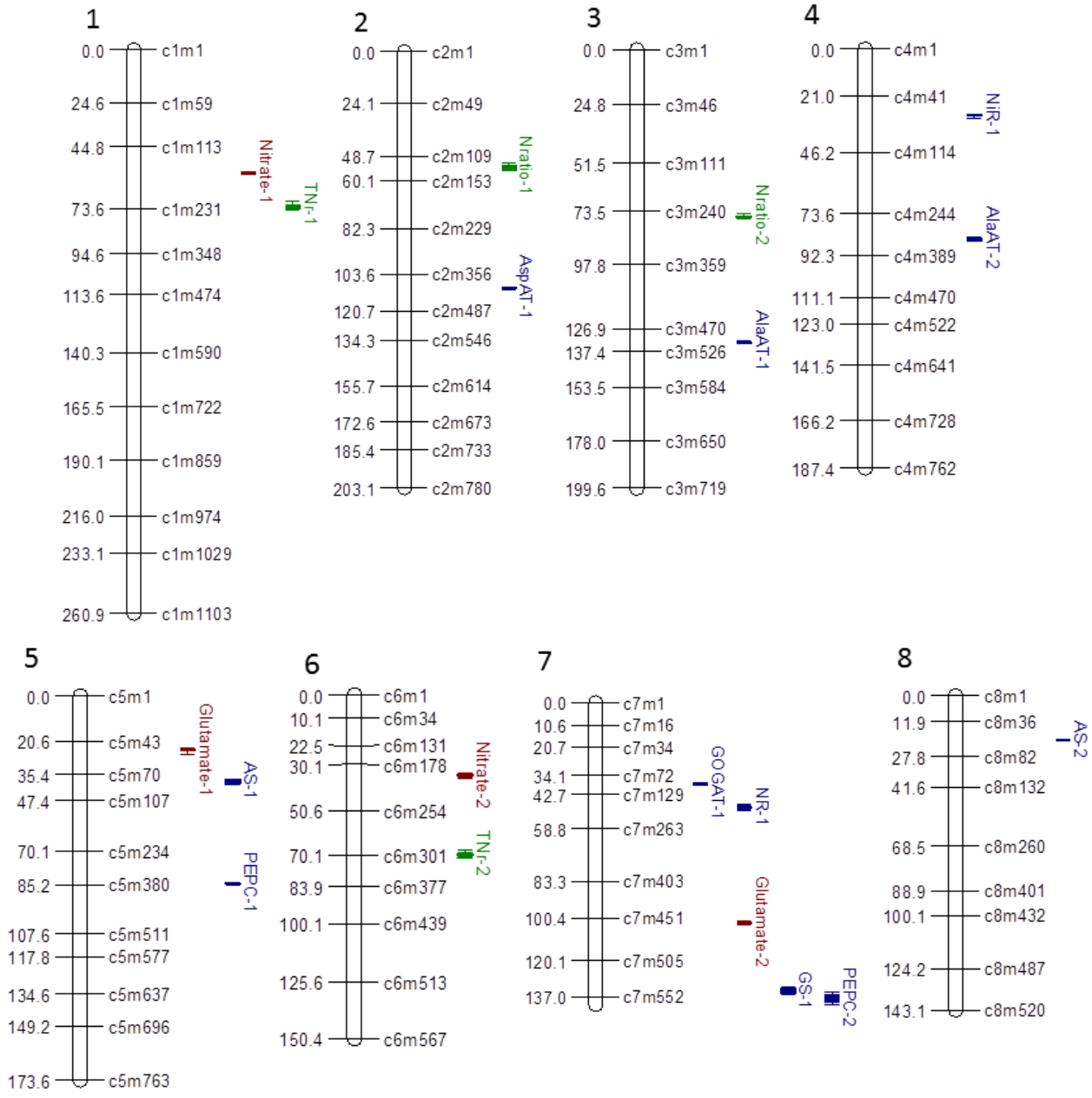

Figure 4.3. Genetic map and distribution of QTL associated with N-metabolism related enzymes and metabolites measured on root tissues in the maize IBMSyn10-DH TC population.

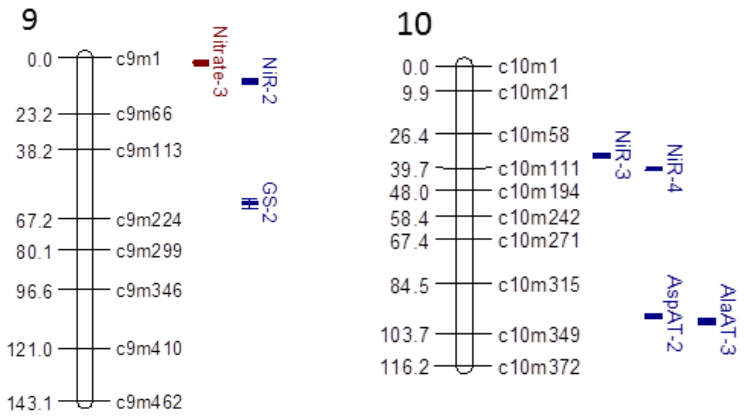

Figure 4.3 continued.

QTL positions shown at right of chromosomes and lengths of bars are determined by 2-LOD confidence intervals. QTL associated with enzyme activity are in blue, while QTL associated with metabolites are in red. Only selected markers are displayed in the figure. Figure created with MapChart 2.2 (Voorrips, 2002).

Tables

Table 4.1: Sample size, mean values for the population and checks, minimum values, maximum values, standard deviation, coefficient of variation, genetic effect p-value and repeatability of traits measured root tissues from the IBMSyn10-DH TC population of maize.

Trait	Unit	n[a]	Pop μ[b]	B73TC[c]	Mo17TC[d]	Min[e]	Max[f]	SD[g]	CV[h]	G effect P[i]	Rptblity[j]
AlaAT	nMole norm Glu/mg protein/0.5 h	171	241.11	293.38	220.65	189.83	318.37	24.78	0.10	1.75E-12	0.51
AS	nMole norm Glu/mg protein/0.5 h	176	472.65	474.28	476.30	412.90	538.98	20.43	0.04	1.55E-08	0.48
AspAT	nMole norm Glu/mg protein/0.5 h	176	930.63	963.72	929.91	835.90	1070.29	38.97	0.04	6.75E-07	0.39
GOGAT	nMole norm Glu/mg protein/0.5 h	175	182.80	190.14	192.76	146.11	220.28	12.81	0.07	1.98E-08	0.45
GS	nMole GHA /mg protein/0.5 h	176	407.11	453.90	348.95	353.72	471.38	22.08	0.05	2.66E-06	0.38
NiR	nMole nitrite reduced/mg protein	172	699.90	648.80	589.15	623.88	779.67	33.14	0.05	1.12E-08	0.48
NR	nMole nitrite/mg protein	171	2.82	2.78	1.40	0.03	6.81	1.36	0.48	4.40E-11	0.65
PEPC	uMole NADH/min/mg protein	172	394.85	357.26	423.80	320.09	506.06	40.87	0.10	<1.00E-12	0.62
Nitrate	nMole/mg protein	157	199.16	245.32	189.28	160.51	236.25	17.26	0.09	2.99E-09	0.52
Glutamate	nMole Glu/mg protein/0.5 h	106	194.38	223.96	192.06	166.71	231.09	12.06	0.06	4.33E-09	0.57
TN_r	mg	176	4.75	7.10	5.10	2.21	9.13	1.09	0.23	7.12E-07	0.70
N_{ratio}	ratio	176	6.70	6.23	6.45	5.91	7.37	0.26	3.88	3.92E-13	0.50

[a] Population size, [b] Population mean, [c,d] BLUP value for parental genotypes in test cross genotype, [e] Minimum value, [f] Maximum value, [g] Standard deviation, [h] Coefficient of variation (%), [i] p value of the genetic effect, [j] Repeatability, normalized values were multiplied by a factor of 1.131 for AlaAT, AS and AspAT, and by 1.151 for GOGAT

Table 4.2. QTL associated with N-metabolism related enzymes and metabolites from root tissue analysis in the IBMSyn10-DH TC maize population.

QTL name	Chr[a]	Marker[b]	G Pos (cM)[c]	G Interval (cM)[d]	Adj (cM)[e]	P Pos (Mb)[f]	P Interval (Mb)[g]	LOD	r^2 (%)	Add effect[h]	# Genes[i]
AlaAT-1	3	429	867.44	865.09-868.13	133.45	180.25	192.00-194.20	4.33	7.16	-10.27	84
AlaAT-2	4	264	555.33	552.56-557.69	85.44	58.45	143.75-146.05	4.93	8.14	7.76	63
AlaAT-3	10	292	643.8	641.20-648.66	99.05	139.85	146.05-146.25	4.89	8.06	-7.41	17
AS-1	5	79	249.44	246.16-258.51	38.38	9	9.00-9.85	5.52	8.55	6.27	41
AS-2	8	45	131.64	131.16-132.20	20.25	6.2	6.85-7.15	4.66	7.16	6.13	6
AspAT-1	2	339	715	714.43-718.51	110	137.15	165.65-169.35	7.32	11.18	-17.1	115
AspAT-2	10	288	627.89	625.33-631.09	96.6	139.25	145.25-145.45	6.37	9.21	-12.41	6
GOGAT-1	7	84	245.75	241.98-248.84	37.81	11.9	13.85-14.10	6.22	9.77	-4.53	11
GS-1	7	409	873.64	865.44-880.88	134.41	158.15	173.65-174.25	4.25	6.85	-6.17	37
GS-2	9	164	393.47	388.70-398.86	60.53	23.85	26.65-28.30	4.35	7.02	-6.28	63
NiR-1	4	55	191.89	189.18-194.26	29.52	6.85	6.75-7.25	6.76	10.61	11.86	24
NiR-2	9	29	61.2	57.89-70.10	9.42	3.65	3.65-4.65	6.28	9.81	11.1	29
NiR-3	10	77	230.71	221.50-233.11	35.49	10.05	10.45-10.85	5.69	10.57	-13.59	18
NiR-4	10	101	262.07	260.48-263.82	40.32	14.35	15.95-19.05	7.67	12.21	16.93	96
NR-1	7	136	315.63	313.05-321.94	48.56	25.6	78.95-100.05	5.49	9.11	-0.43	376
PEPC-1	5	298	551.15	548.52-552.56	84.79	75.25	127.35-139.25	23.4	31.54	-23.78	218
PEPC-2	7	409	887.11	882.06-906.52	136.48	158.15	174.55-175.55	4.36	4.48	-8.95	60
Glutamate-1	5	50	159.48	157.65-162.73	24.54	5.55	5.45-5.65	8.76	15	-6.84	17
Glutamate-2	7	327	666.28	663.82-668.75	102.5	137.55	163.50-164.05	5.75	9.09	-5.43	33
Nitrate-1	1	127	370.69	369.37-372.41	57.03	18.1	23.45-23.80	6.99	14.01	-5.37	12
Nitrate-2	6	155	227.31	223.79-231.96	34.97	86.35	93.55-94.85	5.04	9.48	-4.1	48
Nitrate-3	9	11	12.29	6.99-16.15	1.89	1.75	1.25-1.60	4.79	9.13	4.12	12
TN_r-1	1	195	469.95	467.09-480.29	72.3	37.95	37.15-39.35	4.38	6.04	-0.28	79
TN_r-2	6	235	451.99	449.76-457.93	69.54	115.45	115.10-117.50	4.54	6.27	0.29	100
N_{ratio}-1	2	119	349.98	344.92-359.17	53.84	16.25	15.85-16.65	4.37	6.39	-0.07	40

Table 4.2 continued.

QTL name	Chr[a]	Marker[b]	G Pos (cM)[c]	G Interval (cM)[d]	Adj (cM)[e]	P Pos (Mb)[f]	P Interval (Mb)[g]	LOD	r^2 (%)	Add effect[h]	# Genes[i]
Nratio-2	**3**	**212**	**493.48**	**490.50-497.62**	**75.92**	**133.5**	**129.25-136.95**	**7.03**	**10.68**	**-0.09**	**185**

[a] Chromosome number, [b] Marker localized at LOD peak, [c] Genetic position of molecular marker in cM, [d] 1-LOD interval in cM, [e] Adjusted genetic position, [f] Physical position in Mb, [g] 1-LOD Physical interval, [h] Additive effect of QTL (a positive-signed effect represents an increasing allele from B73, while a negative-signed allele denotes an increasing allele from Mo17), [i] Number of annotated genes underlying 1-LOD QTL confidence interval

Table 4.3. Analysis of multiple QTL model for N-metabolism related enzymes and metabolites measured on root tissue from the maize IBMSyn10-DH TC population.

Phenotype	# QTL in model[a]	Model R^2 (%)[b]
AlaAT	2	11.65
AS	2	8.07
AspAT	2	12.06
GOGAT	1	9.77
GS	2	6.59
NiR	4	26.42
NR	1	9.11
PEPC	2	42.53
Nitrate	3	15.77
Glutamate	2	18.86
TN_r	2	8.12
N_{ratio}	2	12.85

[a] Number of QTL fitted in MIM model, [b] Total R^2 obtained by fitting significant QTL simultaneously in a MIM model

Table 4.4. Candidate genes underlying 1-LOD QTL regions associated with N-metabolism related enzymes and metabolites measured on root tissue from the maize IBMSyn10-DH TC population.

Maize GDB ID	Corresponding gene annotation	Chr[a]	Start[b]	End[c]	QTL name
GRMZM2G028574	PEPC 3	6	115914515	115915086	TN_r-2
GRMZM2G111225	Nitrilase 2	4	145590144	145596571	AlaAT-2
GRMZM2G136712	Aspartate kinase	7	80189428	80201455	NR-1
GRMZM2G155974	Glutathione synthetase	3	133812995	133826187	N_{ratio}-2
GRMZM2G166366	Aspartate kinase	6	115555315	115557026	TN_r-2
GRMZM2G374302	Arginine decarboxylase	4	144862958	144868207	AlaAT-2
GRMZM2G409131	Phosphofructokinase	7	82344751	82349620	NR-1
GRMZM2G466543	Arogenate dehydratase 6	2	166506882	166509171	AspAT-1
GRMZM2G473001	PEPC 1	7	86459173	86464913	NR-1
GRMZM2G481529	Phosphopyruvate hydratase	1	38637579	38641262	TN_r-1
GRMZM2G493395	Deoxy-D-xylulose-5-phosphate synthase	7	14086686	14089909	GOGAT-1
GRMZM5G817058	Phosphoribosyl transferase	7	80946776	80947644	NR-1
GRMZM2G575696	SAM-methyltransferase	7	85199074	85200388	NR-1
GRMZM2G580894	SAM-methyltransferase	7	83464904	8347015	NR-1

[a] Chromosome, [b,c] start and end location in bp.

CHAPTER 5: MAPPING OF QTL FOR N-METABOLISM RELATED TRAITS IN A MAIZE TESTCROSS POPULATION GROWN IN THE FIELD UNDER LOW AND HIGH NITROGEN CONDITIONS

A paper to be submitted to Theoretical and Applied Genetics

Ignacio Trucillo-Silva[1], Michael Lee[1]*, Thomas Lübberstedt[1], Pedro J. Gonzalez-Portilla[1], Hongjun Liu[2], Jason De Bruin [3], Jeff Schussler[3], Juan Pablo San Martin[3], and Kanwarpal S. Dhugga[4]

[1] Department of Agronomy, Iowa State University, Ames, IA, 50011, USA.
[2] Sichuan Agricultural University, Chengdu, 611130, China.
[3] DuPont Pioneer, Johnston, IA, 50131, USA.
[4] International Maize and Wheat Improvement Center, Mexico D.F, 6-641-06600, Mexico.
***** Corresponding author; mlee@iastate.edu.

Abstract

Nitrogen (N) availability is essential for plant growth and development. During last decades, several problems have arisen due to over-fertilization with N in rural areas. Breeding for maize with greater efficiency in the use of N may help to reduce contamination and increase profits. Nevertheless, previous to breeding, a better understanding of the genetics underlying N-metabolism will be needed. Herein, a quantitative trait loci (QTL) mapping for N-metabolism related agronomic and physiological traits was performed based on a maize hybrid-high-resolution population grown in the field under low (L) and high (H) N conditions. A total of 45 QTL were detected in a combined analysis (across three experiments at each N level) while 117 QTL were identified in the split analysis (at each

experiment by N treatment combination). In regard to the combined analysis, multiple QTL model explained 5.7-33.4% of the phenotypic variance and epistasis was significant for only one trait. Furthermore, 22 candidate genes underlying QTL regions were proposed for further analysis. With regard to the split analysis, QTL models explained from 2 - 43% of the variance, and 50 candidate genes associated with N-metabolism, underlying 1-LOD QTL regions, were targeted for further investigation. In addition, 23 candidate genes described as phosphate transporters and cellulose synthases were identified within QTL regions.

Introduction

Nitrogen (N) is one of the most important mineral nutrients for plant growth and development. In maize, sufficient N is required for amino acid metabolism, ear growth, and dry matter accumulation in maize kernels (Hirel et al., 2001). On the contrary, N deficiency reduces kernel number, dry matter accumulation and could result in a 14–80% decrease in grain yield (Uhart and Andrade, 1995).

Even though N fertilization is a necessity for maize production, the oversupply of N generates several problems, including pollution of primary natural resources and numerous related economic issues. On average only 33-50% of the nitrate applied to the soil is accessed by cereal crops (Raun and Johnson, 1999). Important causes of N loss are denitrification of the nitrate form by soil bacteria and volatilization of surface-applied urea-based fertilizers (Nielsen, 2006), but N-leaching is the principal cause of N-loss. In fact, N leaching from the Mississippi River Basin is considered one of the main causes for the expanding hypoxic zone, or oxygen depletion area which can no longer support aquatic organisms, that develops each year on the Louisiana-Texas shelf of the Gulf of Mexico (Goolsby and Battaglin, 2000).

Nitrate concentrations have increased several fold during the past 100 years in streams of the basin, and the annual delivery of nitrate from the Mississippi River to the Gulf has nearly tripled since the late 1950's.

Breeding maize for traits associated with N-metabolism could render a more sustainable agriculture, leading to a reduction in the use of N fertilizer while maintaining yields and an overall increase in profits. Nevertheless, prior to breeding, the development of a comprehensive understanding of N-metabolism and its relationship with yield and developmental traits may be necessary.

Much of today's commercial maize germplasm originates from seven progenitor lines, including B73 and Mo17 (Mikel and Dudley, 2006). Both inbreds differ in their response to N fertilization (Balko and Russell, 1980) and are parents of the IBM population (Lee et al., 2002). After ten generations of intermating, 360 doubled haploid (DH) lines have been generated from the IBMSyn10 population (Hussain et al., 2007) resulting in a high-resolution mapping population that allows the identification of a limited number of positional candidate genes using the physical map established for inbred B73 (www.maizesequence.org). Especially, an understanding N-metabolism at the testcross (TC) level is relevant from an applied perspective.

Mapping of quantitative trait loci (QTL) is routinely implemented in plant breeding programs. Linkage mapping of QTL allows the mean and variance associated with a specific locus to be estimated. The procedure relies on differences among the trait means of genotypes at a marker locus (Bernardo, 2010). Furthermore, QTL mapping studies conducted in different environments is a plausible strategy to identify genomic regions and genes which respond to stress conditions and are responsible for genotype by environment

interaction (Veldboom and Lee, 1996). However, a QTL region may contain few or numerous genes, depending on the level of recombination and genomic size of the QTL region, and the gene density therein. Hence, identification of specific candidate genes underlying QTL regions might not be straightforward. Nevertheless, based on the description of previously annotated genes, a few candidate genes for N-metabolism have already been identified for further investigation (Masclaux et al., 2001; Gallais and Hirel, 2004; Liu et al., 2012; Jansen et al., 2015).

Grain yield is the final outcome of the interaction between numerous complex biological pathways during plant ontogeny, determined by several genes, and affected by environmental conditions including temperature, water, and nutrient availability. Some of the key genes are associated with mineral nutrient uptake, assimilation and remobilization, while other genes are responsible for traits such as flowering time, plant and ear height. A wide variety of nutrients are crucial for plant development but N is considered the most limiting mineral nutrient for maize, and the understanding of the genetics and regulation of N-metabolism related genes is still limited.

Plant growth relies on the activity of the primary metabolism in the source leaf, where carbon (C) fixation takes place during photosynthesis but higher amounts of N integrated into amino acids and proteins are required, and phosphorus (P) for the creation of RNA and realization of energy. When plants suffer deficiencies from any of those mineral macro nutrients, plant growth and development would primarily rely on the ability of re-adjustment of the cellular C-N-P homeostasis (Schlüter et al., 2013). C-, N-, and P– metabolisms are coordinated by metabolite cross-talk, availability of substrates, phytohormones signals and provision of final products, and any environmental stress or physiological alteration would

generate a complex cascade of reactions in order to readjust plant homeostasis. Thus, it is expected that genes having a major role in primary C or P metabolism may be potentially identified as candidate genes underlying N-metabolism related traits (Liu et al., 2012).

Several studies have shown associations between QTL and N-metabolism-related agronomic traits. In a previous investigation (Agrama et al., 1999), 214 F_2 maize genotypes grown at LN and HN conditions, were genotyped with 185 restriction fragment length polymorphism (RFLP) markers, and QTL associated with N-use efficiency were identified. Traits analyzed were ear-leaf-area, plant height, yield, ears-per-plant, kernels-number-per-ear, and kernel weight, and 5-11 QTL were detected, correspondingly. Likewise, genomic regions associated with grain yield and its components were determined in a subsequent study (Bertin and Gallais, 2000). That investigation focused on 99 maize hybrids grown under LN and HN, genotyped with 152 marker loci. The genetic variability was expressed differentially under LN and HN conditions (distinct QTL were detected) and a total of 29 QTL were identified. Co-location between those QTL and QTL for physiological traits related to N-assimilation, such as nitrate content and GS activity, has been reported and GS was proposed as a candidate gene (Hirel et al., 2001). In addition, N-metabolism was studied during kernel germination in a population of 140 F_6 recombinant inbred lines, derived from the cross of a French flint line (F2) by an iodent line. The population was genotyped with 152 RFLP and nine QTL were detected. In accordance with previous studies, coincidences were determined between QTL and genes encoding for GS (Limami et al., 2002).

Since genetic variation has been reported to be expressed differentially at LN and HN conditions (Agrama et al., 1999; Bertin and Gallais, 2000; Gallais and Coque, 2005) and N-remobilization and post-silking N-uptake appears to be distinctively determined in lines

compared to hybrid genotypes (Coque and Gallais, 2008), a QTL analysis conducted on a maize hybrid population, derived from wide-spread parental genotypes (e.g., B73 and Mo17) from different heterotic groups, which have undergone a high number of recombinant events, may provide even more reliable and accurate QTL associations.

The objectives of this study were to (i) investigate the genetic variance present in a high-resolution maize TC population, derived from the cross between IBMSyn10-DH lines by an elite inbred, grown in the field under LN and HN conditions (ii) identify QTL associated with N-metabolism related traits, and (iii) proposed N-metabolism candidate genes underlying QTL for further studies.

Materials and Methods

Plant materials

A total of 176 TC genotypes, derived from the cross between each IBMSyn10-DH line and an elite inbred were used in this study. The IBMSyn10-DH population, developed by Hussain et al. (2007), is a set of DH lines derived from a population after ten generations of random mating from the cross between B73 x Mo17. Each DH line was crossed by NSSZ3 (i.e., PEI), an elite non-stiff stalk inbred, property of DuPont Pioneer, to generate the TC offspring. Initially, 200 genotypes were planted in the field experiments. However 24 genotypes were omitted from successive analysis due to DNA contamination during the extraction process resulting in misleading genotypic information.

Experimental design

Three experiments were planted at Johnston, IA and Marion, IA during 2011 and 2012. Plots were arranged in a split-plot design with N fertilization as treatment (Fig. 5.1A).

Two N levels were applied: high N (HN) and low N (LN), each fertilized with 269 kg ha^{-1} and with 78 kg ha^{-1} of N, respectively (Fig. 5.1B-C). Two replications were planted at each N level, in a two-row plot of 5.3 m length, with a density of 89,000 plants ha^{-1}. At Johnston location, field trials were planted in 2011 and 2012 (Experiments 1 and 3, respectively) under irrigation. Plots at Marion were grown in 2011 and 2012 (Experiment 2) under rain-fed conditions. However, the experiment planted at Marion 2011 was completely discarded due to a severe storm that damaged most plants. All fields were kept free of weeds throughout the growing seasons.

The 2011 growing season was characterized by high temperatures during crop establishment, followed by excessive rainfall. In addition, there were extremely high temperatures around flowering time and widespread high winds. The summer of 2012 was dry and relatively hot (Tables A5.1; A5.2).

Phenotypic measurements

Nine traits were measured on a plot basis, including plant (PHT) and ear height (EHT), flowering time (GDD), yield, N leaf content at 20, 45 and 60 days after flowering (N20DAF, N45DAF and N60DAF, respectively), and N remobilization at two stages. Height was computed, as the distance (cm) from the soil surface to the ear node (EHT) and to tassel tip at male flowering time (PHT). Flowering time was calculated as the growing degree days (°C) accumulated from planting to 50% of plants in the plot exerting 50% anthers. N leaf content (%) was determined by combustion (Dumas, 1826), from samples taken at different reproductive stages from the leaf immediately above the uppermost ear from four plants per plot. Plants sampled were selected based on phenotypic uniformity). In addition, plants and leaves were identified for replicating successive samplings on the same observation unit. N

remobilization (%) was estimated as N content at first leaf sampling (N20DAF) minus N content at following sampling date (N45DAF or N60DAF), divided by N content at first sampling. Depending on the experiment, one or two determinations of N remobilization were estimated: between N20DAF and N45DAF (R2045) and/or between N20DAF and N60DAF (R2060). Some traits were not computed, such as N45DAF in Experiment 1 and height measurements at Experiment 3. In addition, data were not recorded for N60DAF in Experiment 2, neither for seven TC genotypes from Experiment 1. Plots were machine-harvested at physiological maturity and grain yield was estimated on a plot basis and values adjusted to 14% moisture content.

Statistical analysis

All statistical analysis were performed with R statistical program (RCoreTeam, 2014). An analysis of variance (ANOVA) was conducted with a full model were each trait was fitted at a time in order to estimate variance components. The sources of variation included experiment, treatment nested into experiment, replication nested into treatment, genotype, genotype by experiment, and genotype by treatment interactions. As genotype by environment interaction for most traits was statistically significant (both genotype by experiment and at times genotype by treatment interactions), and the ranking of genotypes and checks varied substantially from one experiment to another, statistical (and concomitantly QTL analysis) were initially conducted across locations (i.e., combined analysis) and successively at each experiment and treatment combination separately (i.e., split analysis) (Table A5.3).

Initial data analysis of raw data was based on the ggplot2 package (Wickham, 2009) and GGally (Schloerke et al., 2014). As a first step, a univariate analysis, where a single

variable is fitted in a model, followed by a multivariate approach, where multiple variables are analyzed simultaneously, was performed in order to comprehend the relationship among variables. The determination of outliers present in the dataset, based on a jackknife resampling strategy, was applied. As described in Trucillo-Silva (2015), a statistical model is fitted n times, systematically omitting one observation from the dataset, followed by the prediction of random effects for a subset of most consistent genotypes each of the n times. The aim of the process is to target "real outliers" based on the complete information gathered in the experiment and fine-tune the statistical model, quantified by improvements in log-likelihood, Akaike and Bayesian information criterion values after discarding misleading observations, while keeping informative and true observations for later analysis. The mixed model was fitted with ASReml R package (Butler et al., 2007) and correspondent mixed model equations were solved for prediction of random effects and estimation of fixed effects. The statistical model can be represented as follows:

$$y = Xb + Zu + e$$

where y denotes a n x 1 vector of observed response values, b is a p x 1 vector of fixed effects, X is a n x p design matrix, u is a q x 1 vector of random effects, Z is a n x q design matrix, and e being the error term.

The following assumptions were used: E (u) = 0, E (e) = 0, Cov (u, e) = 0, and Var (u) = G and, Var (e) = R. The G matrix had a compound symmetry structure on the genotype levels and R matrix is the direct product of two autoregressive correlation matrices in order to take into account spatial adjustment based on rows and columns arrangement on each field experiment. The response variables were yield, GDD, PHT, EHT, N20DAF, N45DAF, N60DAF, R2045, and R2060. In the combined analysis, each N treatment was analyzed

separately and experiment and replication nested into experiment were included as fixed effect in the model, and check genotype effect was included as a continuous covariate. A random effect for the TC genotype was included in the linear model and spatial adjusted BLUP values were predicted for each genotype. Likewise, for the split analysis, replication was included as fixed effect in the model, and check genotype effect was included as a continuous covariate. Finally, a random effect for the TC genotype was included in the linear model and spatial adjusted BLUP values were predicted for each genotype.

After conducting the approach described in Trucillo-Silva (2015) with the raw dataset, different numbers of genotypes were omitted from the analysis of each trait. In the combined analysis, the mean sample size was n= 175, with no genotypes omitted for EHT, N20DAF, N60DAF, R2045, and R2060 under both LN and HN conditions, while a maximum of four genotypes were discarded in Yield and GDD, at LN as well as at HN. However, in the split analysis the average sample size for the different analysis was 170 genotypes, with a minimum number for PHT at Experiment 1 under HN (156), and a complete population size with all 176 genotypes (no genotype was discard) for GDD, N20DAF and R2060, at Experiment 3 at LN, HN and HN condition, respectively.

Significance of genetic variance was calculated based on log-likelihood ratio test by comparing a full model considering TC random effect versus a reduced model without including the term. Correlation analysis was determined among spatial adjusted BLUP values for each pair of traits, and significance was adjusted based on Bonferroni correction for multiple comparisons. Repeatability for each trait was derived from the variance estimations from ASReml as $\mathrm{Vg}/((\mathrm{lr}\ \mathrm{Vg} + rVgxe + Ve)/lr)$, where V_g is the genetic variance due to TC genotypes, V_{gxe} is the variance attributable to the interaction between genotype and

environment (G x E), V_e is the residual variance, l is the number of environments, and r denotes the number of replications.

Genotypic information and genetic maps

TC genotypes were analyzed with a total of 5,306 SNPs markers generated by Beijing Genomics Institute (Liu et al., 2015). Physical and genetic position of each SNP were determined and genetic maps were created using R/qtl (Broman et al., 2003). Recombination fractions were estimated and Kosambi mapping function was employed to calculate genetic map distances (Kosambi, 1944). Furthermore, as the recombination between linked loci increases every generation, leading to an expansion of the genetic map, mapping distances were adjusted to an F_2 map (Teuscher et al., 2005) in order to compare the outcomes with previous investigations. The expansion factor was determined based on the following equation: $\propto = \frac{j}{2} + (2\mathrm{i} - 1)/2\mathrm{i}$, where j corresponds to the number of generations of intermating including the two generations for creating the F_2, and i is the number of inbred generations after intermating.

The genetic map was 11,228.24 cM length or 1,727.42 cM in F_2 adjusted distance, with an average interval between markers of 2.12 cM. The 5,306 SNPs markers were spread across all chromosomes, with a maximum of 919 markers present in chromosome 1, and a minimum of 320 marker loci for chromosome 10 (Fig. 5.2).

QTL mapping and identification of candidate genes within QTL regions

Associations between phenotypes and genotypes were determined using QTL Cartographer (Basten et al., 2002). Single marker analysis, followed by linear regression analysis and composite interval mapping (CIM) was performed. For CIM, Zmap (model 6) was implemented, using the ten most significant marker cofactors identified by forward and

backward regression. In addition, QTL were scanned at intervals of 1 cM and at every marker, while cofactors located within a window of 10 cM of the scanned position were excluded from the analysis. In order to determine 5% LOD scores thresholds to define the significance of QTL, 1,000 permutations were performed. Two nearby QTL were considered as different when LOD peaks were localized 20 cM or greater apart. Effects of QTL are expressed relative to the B73 allele. As a result, a positive effect would imply an increase in the phenotypic value when the B73 allele is present, whereas a negative effect would indicate a reduction in the presence of B73 allele.

As stated above, even though QTL analysis was initially performed for each N treatment across locations, due to the presence of extensive G x E interactions, the analysis was additionally performed separately for each experiment and treatment combination. Furthermore, a multiple interval mapping (MIM) analysis was implemented by fitting all previously identified QTL from CIM analysis. In addition, all pairwise interactions between QTL in every model were studied for each trait. The significance was determined based on the information criterion: IC (k) = -2 (log (L) - kc (n)/2), where the penalty function corresponds to: c (n) = log (n) and a threshold of 0.0 was used (Basten et al., 2002). The proportion of the total phenotypic variance associated with each model was estimated.
In addition, physical genomic regions corresponding to 1-LOD confidence interval (CI) QTL regions were examined for the presence of annotated genes at MaizeGDB (Lawrence et al., 2008) and Phytozome (Goodstein et al., 2012). Candidate genes related to N-metabolism were prioritized based on their descriptions on model species, such as rice (*Oryza sativa*) and *Arabidopsis thaliana*, and proposed as targets for further studies.

Results

Statistical analysis

In the combined analysis, genetic variance was statistically highly significant (p-value<0.001) for nearly all traits across experiments at both LN and HN treatments (Table 5.1). For a few traits, namely N60DAF and R2060, genetic variance was statistically significant (p-value<0.05). In addition, all traits showed a wide distribution of values. Mean values for yield, EHT, PHT, N20DAF, N45DAF, and N60DAF were greater under HN than LN; the opposite pattern was observed for GDD, R2045 and R2060.

In general, repeatability values were higher under HN compared to LN, with mean values of 0.30 and 0.40 for LN and HN, respectively; and an overall mean of 0.35 across treatments. The highest value for repeatability was found for N45DAF under HN (0.69), the lowest value was for yield under LN conditions (0.15). The estimated coefficient of variation values, or relative standard deviation, ranged from 0.7-13.7 % for GDD and R2045, respectively. ANOVA results confirmed that G x E is a highly significant (p-value<0.001) source of variation for all of the traits (Table A5.3).

In addition to the combined analysis, the statistical analysis was performed for each experiment and treatment combination (Table A5.4). Repeatability values were higher in the split analysis compared to the combined analysis due to the extensive of Vgxe. Nevertheless, CV values were higher in the split compared to the combined analysis (e.g., EHT at LN was 2.3% in the combined analysis compared to 9.3 and 5.1% for EHT at LN at Experiments 1 and 2, respectively).

Correlations between traits

Analysis of correlations was performed across Experiments at each N level. From a total of 72 comparisons, 22% were highly significant (p-value<0.001), 4% showed intermediate significance (p-value<0.01), and 7% were statistically significant (p-value<0.05; Fig. 5.3). Correlation values between Yield and PHT, Yield and GDD, EHT and PHT, EHT and GDD, GDD and PHT, N20DAF and N45DAF, N45DAF and R2045, and N60DAF and R2060 were all statistically significant at both N treatments. Some correlation estimates were significant only at LN (e.g., between N20DAF and N60DAF, and N45 with EHT), while other correlations were significant only at HN (e.g., between Yield and EHT, and R2045 and R2060). Similar numbers of significant correlations were found at LN and HN (11 and 12, respectively).

Furthermore, analysis of correlations was conducted between traits at each experiment and treatment combination (Tables A5.5-A5.7). From a total of 126 comparisons, 24% were highly significant, 4% showed intermediate significance, and 3% where statistically significant. Close correlations were found between PHT and EHT, under both HN and LN conditions, and both traits were closely correlated to GDD. In general, close correlations were computed between N leaf content and N remobilization. Furthermore, Yield showed closed correlations with PHT and EHT in Experiment 2 at HN level.

Identification of quantitative trait loci

Analysis across experiments

Composite interval mapping

In the combined analysis, a total of 45 QTL were identified (Fig. 5.2; Table 5.2). QTL were identified in all chromosomes, ranging from 11 (chromosome 3) to one QTL per

chromosome (chromosomes 9 and 10). Twenty-three QTL were determined under LN and 22 under HN. On average, 2.5 QTL were identified per trait. Even though four QTL were identified for some traits such as EHT at LN and PHT at HN, no QTL were identified for N20DAF at HN. The percentage of explained variance by an individual QTL varied from 16%, for N60DAF-HN-2 to 5%, for R2045-HN-1. On average, each QTL explained 9% of the variance and most QTLs (67%) were responsible of less than 10% of the variance. Confidence intervals (1-LOD score) for QTL localization ranged from 2.04 to 21.89 cM (0.31-3.37 cM F_2 adjusted distance), with an average CI length of 8.68 cM (1.34 cM F_2 adjusted distance). Furthermore, those CI are equivalent to a physical map distance of 0.15 – 6.05 Mb, with an average of 1.08 Mb. All identified QTL CI covered in total nearly 3 % of the genome, or about 344.67 cM (53.03 cM F_2 adjusted distance).

A few QTL co-locate or were identified in close proximity on the genetic map. Based on how N remobilization was estimated and results from correlation analysis, most of QTL sharing genetic positions are associated with N leaf content and N remobilization (e.g., N60DAF-LN-2 and R2060-LN-1, N45DAF-HN-1 and R2045-HN-2). In addition, a few QTL associated with the same trait under both LN and HN conditions were found to co-locate (e.g., N45DAF-LN-3 and N45DAF-HN-1 at chromosome 3, GDD-LN-1 and GDD-HN-2 at chromosome 5) (Fig. 5.2; Table 5.2). Furthermore, some QTL associated with different traits were identified in extremely close positions, such as Yield-LN-1 and GDD-LN-1, and EHT-LN-3 and PHT-1-1 (peak LOD identified 14 and 5.7 real map cM apart, respectively).

Multiple interval mapping and epistasis analysis

A few multiple QTL models explained even greater than 31% of the variance (e.g., N45DAF under both LN and HN conditions and R2045 at HN) in the combined analysis

while other QTL model, R2060 at LN, accounted 5.7% of the phenotypic variance. For EHT at HN a single QTL was identified, explaining 8% of the variance. Furthermore, epistasis between QTL was not statistically significant in most of the models. However, digenic epistasis was significant for R2045 at HN and accounted 3.5% of the variance (Table 5.3).

Candidate genes within QTL

On average of 48 genes are annotated underlying QTL 1-LOD CI, ranging from five to 177 genes, for N60DAF-HN-3 and GDD-LN-1 QTL, respectively. Twenty-three of the candidate genes may have important roles associated with N-metabolism based on the descriptions for model species (Table 5.4). Most of the genes are related to the translocation of proteins and metabolites within the plant (e.g., GRMZM2G076593, an amino acid transporter and GRMZM2G143190, a major facilitator superfamily protein).

QTL analysis at each experiment by nitrogen level combination

Composite interval mapping

In the split analysis, 117 QTL were identified in total (Fig. A5.1; Table A5.8). Twenty-seven percent (or 12) of the QTL identified previously in the combined analysis were likewise identified in this individual experiment analysis (Table 5.2, shared QTL are marked with a rectangle). Most of the shared QTL between analyses (75%) were identified under same N treatment and QTL were consistent in the parental contribution of the allele and the magnitude of their effects (r=0.93). QTL were determined in all chromosomes, and chromosome 3 presented the highest number of QTL (22) while chromosome 8 showed the lowest quantity (7). Similar amount of QTL were identified under HN and LN conditions, 56 and 61 correspondingly. Some chromosomes presented more associations under LN conditions (e.g., chromosome 8) while other chromosomes had shown more QTL under HN

conditions (e.g., chromosome 5). On average, 13 QTL were identified per trait, and seven and six QTL were determined under LN and HN, respectively. Individual QTL effect explained on average 8.5% of the total variance per trait, varying from 17.5% (for QTL R20601LN-1) to 4.5% (for QTL GDD3LN-1) (Table A5.8). Nearly 80% of the detected QTL explained individually less than 10% of the variance.

Confidence intervals (1-LOD score) for QTL localization ranged from 1.00 to 38.7 cM (0.15-5.95 cM F_2 adjusted distance), with an average CI length of 9.08 cM (1.39 cM F_2 adjusted distance). Furthermore, those CI are equivalent to a physical distance of 0.10 – 34.65 Mb, with an average of 2.07 Mb. All identified QTL CI covered in total nearly 8 % of the maize genome, or about 894.28 cM (137.58 cM F_2 adjusted distance).

Several QTL were identified in close proximity on the genetic map or overlapped 1-LOD CI. Numerous QTL associated with N leaf content at different growth stages co-locate at three noticeable "QTL hotspots". Those QTL are located in chromosome 1(N20DAF3LN-2, N60DAF3LN-1, N20DAF2LN-1 and N20DAF2HN-1), chromosome 4 (R20601LN-5, R20453HN-3, R20603HN-2 and N60DAF3HN-1), and chromosome 6 (R20603HN-3, N45DAF2LN-4 and R20452LN-3). As expected, based on how N remobilization was calculated and results from correlation analysis, some QTL for N remobilization were detected on essentially same location to their correspondent N leaf content traits. That is the case for N60DAF1LN-2 and R20601LN-1 on chromosome 1, R20603HN-2 and N60DAF3HN-1 on chromosome 4, N45DAF2LN-5 and R20452LN-4 located on chromosome 7. Further highly-dense QTL regions are localized in chromosome 1 around 45 cM, at chromosome 2 position 356 cM, chromosome 3 at 50 cM, 250 cM and 460 cM, and at chromosome 5 at position 234 cM. There are 21 overlaps between QTL physical 1-LOD

intervals. Furthermore, numerous QTL may be considered stable due to the identification in extremely close position in different experiments (e.g., R20601LN-3 and R20603LN-1 at chromosome 3, Yield2HN-4 and Yield3HN-3 at chromosome 7), across different N conditions (e.g., N20DAF2LN-1 and N20DAF2HN-1 at chromosome 1, EHT2LN-2 and EHT2HN-4 at chromosome 3) and across both experiments and N conditions (e.g., N20DAF2LN-2 and N20DAF3HN-2 at chromosome 3) (Fig. A5.1; Table A5.8).

Multiple interval mapping and epistasis analysis

A few multiple QTL models explained even greater than 30% of the variance (e.g., R2060 at Experiment 1-LN and Experiment 3-HN, yield at Experiment 2-HN and Experiment 3-HN, N45DAF at Experiment 2), while the QTL model for GDD at exp2trt1 accounted 9.4% of the phenotypic variance. For some phenotypes (EHT, N20DAF, R2045, and N60DAF), single QTL were fitted in the MIM model at certain experiment by treatment combination, and 2-9% of the variance was explained. In addition, epistasis between QTL was statistically significant in four of the MIM models. Digenic epistasis accounted from 5% of the variance, for N60DAF model at Experiment 3 under LN, to 0.6% for EHT at Experiment 2 under HN (Table A5.9).

Candidate genes within QTL

On average 60 genes were identified underlying QTL CI, ranging from seven (under EHTHN-1 and EHTHN-5) to 597 (for R20452HN-1) genes. Nevertheless, a subset of 50 candidate genes was considered related to N-metabolism (Table 5.10). Most of those genes are associated with the translocation of metabolites within the plant, including transporters for nitrate, ammonium, amino acids, and sucrose; while others are structural genes of enzymes involved in the N-pathway, such as nitrite reductase (NiR), nitrate reductase (NR),

alanine aminotransferase and asparagine synthase; and C-primary metabolism, as phosphoenolpyruvate carboxylase. Six of the candidate genes are localized within the QTL "hotspots" on chromosome 1 (GRMZM2G050481, GRMZM2G085210, GRMZM2G119511, and GRMZM2G359559) and on chromosome 4 (GRMZM2G079381 and GRMZM2G428027). Furthermore, 23 genes associated with phosphate transporters and cellulose synthases were identified within QTL genomic regions (Table 5.11). One of those candidate genes is situated on the QTL "hotspot" region localized on chromosome 4 (GRMZM2G060630).

Discussion

The elucidation of the genetics underlying N-metabolism in maize TC provides a basis for breeding genotypes which can produce grain yield in a more efficient manner enhancing both productivity and sustainability. In the present investigation, 45 QTL associated with N-metabolism related traits were identified in the analysis across three experiments at two different N levels, and 117 QTL were found when studying each experiment by N treatment combination separately. Twelve of the QTL identified in the combined analysis were detected as well in the split analyses. Besides, 23 and 73 candidate genes were identified within QTL regions in the combined and the split analysis, respectively; and are proposed for further N-metabolism studies.

Strengthens of using an IBMSyn10-derived population

A TC segregating population derived from the cross between two founder lines (B73 and Mo17) of several current U.S. commercial germplasm was utilized. Hence, as the alleles are present in several commercial breeding germplasms, the results may be representative of elite germplasm and can be associated hybrid cultivars, the type of cultivar extensively

planted by farmers. Furthermore, fine mapping may be doable due to the ten generations of random mating during the creation of the population and a densely coverage of the genome with 5,306 polymorphic SNPs markers. Compared to previous studies (Agrama et al., 1999; Gallais and Hirel, 2004), were different populations and fewer molecular marker loci were employed, smaller QTL intervals were obtained. Thus, the determination of smaller CI QTL allowed the identification of a limited number of candidate genes associated with N-metabolism related traits.

Quality of the study

Herein, a novel approach for the determination of real outliers was implemented for the analysis of raw data. Even though high quality data are essential and the foundation for a successful investigation, statistical analysis of raw data has received considerably less emphasis than the subsequent genetic analysis. Many researchers generally launch directly into the statistical analysis with a routine analysis without carefully checking the quality of the data (Trucillo-Silva et al., 2015). Consequently, the presence of incorrect or inconsistent data may significantly distort the results of an investigation (Hellerstein, 2008), and may produce spurious QTL mapping results. Herein, the usefulness of the gathered information was optimized, followed by the prediction of spatial adjusted BLUP for each genotype. After the calculation of BLUPs, QTL mapping analysis was conducted using the real genetic map in order to obtain the maximum mapping accuracy possible based on the algorithm implemented during the genetic mapping (W. Beavis, personal communication, 2014).

In this investigation, maize hybrids were planted in two locations in Iowa under HN and LN conditions. N treatment implemented was effective and statistically significant, and causal of significant genotype by environment interaction. Accordingly, statistical and

concomitant genetic analysis were conducted for each N level across experiments and, in consistency with previous studies (Tuberosa et al., 2002; Fernandez et al., 2008), separately for each environment and treatment combination.

A significant amount of genotypic variance was identified for all traits, allowing the identification of promising genomic regions associated with the observed variation. Repeatability estimations showed moderate-low values in the combined analysis, as well as moderate-high values for the split analysis, varying for each trait and treatment combination. Thus, the range of repeatability values is comparable to previous QTL investigations for similar traits (Messmer et al., 2009; Semagn et al., 2013).

The analysis of correlation between traits, measured either at LN and HN, resulted in entirely anticipated outcomes based on published literature. In agreement with previous studies (Hallauer et al., 2010; Yin et al., 2011), Yield showed a significant positive correlation with PHT, EHT, and GDD, at both LN and HN conditions. This outcome was expected since, generally, plants with longer life cycle usually showed higher grain yield potential than shorter cycle plants. Furthermore, consistent with Veldboom (1996) and Austin (2001), higher plants showed as well higher EHT. In addition, hybrids at LN had uptake less amount of N from the soil, probably due to poor N availability and root development. Consequently those plants had remobilized N in higher proportions, than hybrids grown under HN level, in order to achieve ear development and further grain production. That is in agreement with former studies (Gallais and Hirel, 2004; Coque and Gallais, 2008), which stated that post-anthesis N uptake was negatively correlated to N remobilization. Unsurprisingly, correlation values between N remobilization and N leaf

content were statistically significant due to the fact that N leaf content was a direct component of N remobilization calculation.

Comparison with previous QTL investigations

Several QTL identified in the combined analysis are coincident with QTL detected in preceding maize studies. For instance, on chromosome 7, three QTL for fresh weight of 100-kernels showed an overlap with the 1-LOD CI for a Yield QTL herein identified. Furthermore, a QTL associated with male flowering and a QTL related to PHT were determined at same QTL 1-LOD CI at chromosome 8 and 4, respectively, on an investigation focused on 236 recombinant inbred lines planted in Mexico and Zimbabwe (Messmer et al., 2009). In addition, a Yield QTL previously localized on chromosome 5 under HN conditions (Coque and Gallais, 2006), was likewise detected in this analysis, however it was detected at LN. Similarly, the Yield QTL found in chromosome 5 under HN was also detected in a previous investigation based on 256 $F_{2:3}$ families evaluated in five tropical environments (Lima et al., 2006). Similarly, numerous QTL detected in the split analysis, even though undetected in the combined analysis, are in agreement with previous investigations. In consistency with Agrama (1999), a QTL for PHT under LN condition (PHT2LN-1) was determined at a very close proximity (chromosome 3, 62.2 cM). Other PHT QTL (PHT1HN-1) was likewise detected in extremely proximity position (chromosome 1, position 48.76 cM). Furthermore, two QTL associated with grain yield (Yield1LN-1 QTL and Yield1LN-2 QTL) did co-locate at chromosome 1, position 44.64 cM and chromosome 10, position 67.93 cM, respectively. Additionally, the QTL associated with yield in chromosome 1 was as well associated with PHT and ear leaf area under LN conditions (Agrama et al., 1999). In agreement with Semagn (2013), in which a meta-QTL analysis was performed for yield and

flowering traits based on 18 maize populations, four Yield QTL and three QTL associated with flowering traits were mapped in the corresponding physical confidence interval as in our investigation (Yield2LN-1, Yield2HN-1, Yield2HN-2 and Yield2HN-3; GDD1LN-1, GGD1HN-1 and GDD1LN-3). Likewise, a Yield QTL determined in here (Yield1LN-2), was identified on an extremely proximate genetic position by Tuberosa (2002) under two water regimes.

Interestingly, co-location between QTL identified in this study and QTL associated with N-metabolism related enzymes (Trucillo-Silva et al., 2015) was also determined. At chromosome 2, there were 1-LOD QTL confidence intervals overlaps between QTL associated with EHT2HN-3 and Lox6-1, N20DAF3HN-1 and Nratio-1, and R20601LN-2 with Asp AT-1 from root tissues. Likewise, in chromosome 6, there were overlaps between QTL for N60DAF3LN-3 and Glutamate-2 from leaf tissues, N20DAF2LN-3 and Nitrate-2 from leafs, GDD2HN-2 with Totalnr-2 from leafs, R20603HN-3 with GOGAT-1 from leafs, and between both N45DAF2LN-4 and R20452LN-3with Nitrate-2 from roots. Similarly, on chromosome 7, QTL for PHT2HN-2 with GOGAT-2 from leafs, and Yield3LN-2 with PEPC-2 from roots shared genetic locations.

In accordance with previous investigations (Gallais and Hirel, 2004; Liu et al., 2007; Cai et al., 2012; Liu et al., 2012), different QTL were identified under LN and HN conditions, reflecting a different genetic basis underlying the target traits depending on specific environmental conditions. Furthermore, several QTL (~25%) detected in the combined analysis under certain N condition were not identified under the same N level when analyzing each experiment separately (e.g., EHT-LN-1 and EHT2-HN-2 at chromosome 1, adjusted position 55.01 cM, detected on the combined and split analysis,

respectively), and numerous QTL where detected in one single location (e.g., Yield1-LN-2 detected exclusively at Experiment 1). That inconsistency across experiments suggests the presence of important QTL by experiment interaction that may play a major role as a contributor of the genotype by experiment variance. The lack of coincidence in the detection of QTL at different environments may be a consequence of the specific environmental features that characterized each of the three experiments. Environmental conditions (e.g., precipitation and temperature) varied substantially from one experiment to another and might have had a direct impact on plant responses. Furthermore, even though N treatment was determined to be statistically significant at each experiment, the total amount of N available for maize plants could not be precisely controlled under field conditions. Several factors may have affected N availability per plant, including root architecture, water content, presence of other macro and micro nutrients, and specific soil characteristics.

Inconsistencies in the detection of QTL with preceding investigations could be due to numerous causes including the usage of different segregating populations, environmental features, sampling variation, approach implemented for the analysis of raw data, and further aspects related to methodologies. The comparison of QTL results might be biased due to different segregating populations, thus different segregating alleles, and probably different genetic control mechanisms occurring on each genetic background. In addition, each population may have experienced a different amount of recombination, affecting the mapping resolution. Some previous investigations were based on populations subjected to a few generations of random mating (Beavis et al., 1991; Agrama et al., 1999). Hence, the resulting QTL associations may correspond to clusters of linked QTL. However, the real number of QTL underlying complex traits is expected to be considerably larger. In addition, each QTL

is expected to have an effect substantially smaller compared to the results obtained from studies based on conventional populations were QTL effects may be overestimated (Huang et al., 2010). Furthermore, due to differential gene expression under specific environmental circumstances, the comparison of results from QTL mapping studies based on different environments may be challenging. In addition, incongruences between results from the analysis of different samples from the same segregating population might arise as a merely artifact of random sampling. Furthermore, the incorrect management of the raw phenotypic data, such as the removal of outliers based purely on visual interpretations without a statistical basis or criteria, might become another causal of discrepancies across studies. In addition, QTL mapping results may vary based on the methodologies used, including the implementation of different QTL mapping models or methods, number of cofactors fit in the mapping model, determination of significance thresholds, and number of genetic markers utilized. Lastly, differences in the phenotyping precision and protocols employed for measuring specific traits might cause non-QTL-co-location across studies.

Importance of candidate genes and consistency with previous investigations

A total of 12 QTL detected in the combined analysis were identified successively in the split analysis, and four candidate genes related to N-metabolism were identified under those QTL regions. Those candidate genes code for a urease accessory protein (GRMZM2G063452), a major facilitator superfamily protein – peptide transporter (GRMZM2G085411), a citrate transporter (GRMZM2G086258), and an adenine nucleotide transporter (GRMZM5G886294). The first gene is responsible of the activation of urease, an enzyme involved in the recycling of N from ureide, purine, and arginine catabolism in plants (Witte et al., 2005). Peptide transporters mobilize di- and tripeptides, playing an important

role in the recycling of organic N (Ouyang et al., 2010) and a peptide transporter ortholog protein was found to be associated with LN tolerance in rice (Nischal et al., 2012). Moreover, citrate transportation is of main importance because its conversion provides C skeletons for N assimilation and reducing equivalents for several biosynthetic reactions (Popova and de Carvalho, 1998). Adenine nucleotides play a vital role in plant physiology, representing the major energy source of the cells, and adenine nucleotide transporters are the responsible of the transport of nucleotides across intracellular membranes (Haferkamp et al., 2011). Numerous of the 19 candidate genes identified under 1-LOD CI QTL regions detected solely in the combined analysis were targeted in previous investigations. A candidate gene identified under a QTL for PHT at HN, codes for ammonium transporter 2 (GRMZM2G043193) and was used as a microarray probe in a previous study, showing consistent expression in adult to post-flowering stage (V5-R31) (Liseron-Monfils et al., 2013). In addition, a phosphoglucomutase gene (GRMZM2G109383) identified under R2060-LN-2 QTL, was as well identified in a recent investigation on the maize Nested Association Mapping (NAM) population under a QTL associated with glucose (Zhang et al., 2015). That enzyme facilitates the conversion of glucose-1-phosphate to glucose-6-phosphate, playing a major role in glycolysis. In addition, the gene GRMZM2G088253, described as an urease accessory protein and detected under GDD QTL under both N conditions, was as well proposed as a candidate gene in a recent analysis of enzymes associated with N-metabolism from leaf tissue (Trucillo-Silva et al., 2015). Moreover, GRMZM2G066413 and GRMZM2G021606, both transcripts involved in primary C metabolism and described as PEP/P translocator and phosphoglycerate mutase family,

respectively; showed significant expression changes at LN versus HN conditions in maize (Schlüter et al., 2012).

Similarly, several candidate genes proposed to be associated with N-metabolism in the split analysis were as well identified in previous investigations. Two of the genes, coding for NiR and NR (GRMZM2G079381 and GRMZM2428027), were as well emphasized as priori candidates in a recent association study based on the NAM population (Zhang et al., 2015). In addition, 14 of the suggested candidate genes were determined to be differentially expressed under LN versus HN conditions (Schlüter et al., 2012) . Five of those genes code for transcripts involved in primary N metabolism (GRMZM2G079381, GRMZM2G088064, GRMZM2G101125, GRMZM2G104546, and GRMZM2G428027), while three genes are involved in primary C metabolism (GRMZM2G035599, GRMZM2G050481, GRMZM2G088064, and GRMZM2G), and six genes are involved in phosphate homeostasis (GRMZM2G009779, GRMZM2G035579, GRMZM2G045473, GRMZM2G083655, GRMZM2G086430, and GRMZM2G155123).

Usefulness for Plant Breeding

The findings of this investigation may contribute to the understanding of N-metabolism at the maize TC level and provide knowledge for future genetic studies. However, additional experimentation will be needed in order to completely elucidate the genetics underlying N-metabolism in maize. The localization of candidate genes that may be functionally related to the traits under investigation does not guarantee complete evidence in order to assure that the annotated genes are the ultimate responsible for the variation in the trait phenotypes. Further studies would be required for the validation of those candidate genes that may comprise the evaluation across multiple genetic backgrounds, re-sequencing

of candidate genes followed by association studies, fine mapping, and functional studies to manipulate the expression of the target genes by gene knock-out (e.g., mutation based), knock-down (e.g., VIGS approach), and/or overexpression. Based on the current knowledge on maize N-metabolism related traits, several small-effect QTL underlie the observed phenotypic variation. Thus, superior genotypes, in terms of N-utilization, may be challenging to be designed on a strictly traditional Mendelian genetic basis (e.g., identifying the exact combination of parents in order to create offspring carrying all desired arrange of alleles after certain number of crossing or backcrossing events), and marker-assisted or genomic selection (whole genome prediction) strategies may be more promising breeding approaches.

Conclusions

In summary, 45 and 117 QTL associated with N-metabolism agronomic and physiological traits were detected in a maize TC mapping population grown in the field at LN and HN in a combined analysis (across experiments, but separately at HN and LN) and a split analysis (at each experiment and N-treatment combination), respectively. Multiple QTL models explained 6-33% of the phenotypic variance and epistasis was significant for only one trait in the combined analysis. Furthermore, 23 candidate genes underlying QTL regions were proposed for further analysis. Whereas in the split analysis, QTL models explained 2-43% of the variance, and 50 candidate genes associated with N-metabolism were targeted for further investigation. In addition, 23 candidate genes associated with phosphate transporters and cellulose synthases were as well detected under the 1-LOD QTL CI regions. Further investigation on the genetics underlying N-metabolism in maize would be necessary with the

aim of developing ideotypes having the ability to maintain or even increase yields with a reduction on N fertilizer inputs leading to a more sustainable agriculture.

Acknowledgments

This work was supported by R.F. Baker Center of Plant Breeding – Department of Agronomy - Iowa State University, USDA_s National Institute of Food and Agriculture (project number: IOW05180) , and DuPont Pioneer. The authors would like to thank Sharon Cerwick, Ryan Pape, Angela Bryant, Beverly Krejsa, and especially to Guan Yi Lai for the help in collecting phenotypic data. Requests for TC materials and IBMSyn10-DH lines may be directed to DuPont Pioneer and Dr. Michael Lee, respectively.

References

Agrama HAS, Zakaria AG, Said FB, Tuinstra M (1999) Identification of quantitative trait loci for nitrogen use efficiency in maize. Molecular Breeding **5:** 187-195

Austin DF, Lee M, Veldboom LR (2001) Genetic mapping in maize with hybrid progeny across testers and generations: plant height and flowering. Theoretical and Applied Genetics **102:** 163-176

Balko LG, Russell WA (1980) Response of maize inbred lines to N-fertilizer. Agronomy Journal **72:** 723-728

Basten CJ, Weir BS, Zeng ZB (2002) QTL Cartographer, Version 1.16. Department of Statistics, North Carolina State University, Raleigh, NC

Beavis WD, Grant D, Albertsen M, Fincher R (1991) Quantitative trait loci for plant height in 4 maize populations and their associations with qualitative genetic-loci. Theoretical and Applied Genetics **83:** 141-145

Bernardo R (2010) Breeding for quantitative traits in plants, Ed 2nd. Stemma Press, Woodbury, MN

Bertin P, Gallais A (2000) Physiological and genetic basis of nitrogen use efficiency in maize. I. Agrophysiological results Maydica **45:** 53-66

Broman KW, Wu H, Sen Ś, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics **19:** 2990-2992

Butler D, Cullis BR, Gilmour AR, Gogel BJ (2007) Analysis of mixed models for S-language environments: ASReml-R reference manual. Department of Primary Industries and Fisheries, Queensland, Brisbane, Australia

Cai H, Chu Q, Gu R, Yuan L, Liu J, Zhang X, Chen F, Mi G, Zhang F (2012) Identification of QTLs for plant height, ear height and grain yield in maize (*Zea mays* L.) in response to nitrogen and phosphorus supply. Plant Breeding **134:** 502-510

Coque M, Gallais A (2006) Genomic regions involved in response to grain yield selection at high and low nitrogen fertilization in maize. Theoretical and Applied Genetics **112:** 1205-1220

Coque M, Gallais A (2008) Genetic variation for N-remobilization and postsilking N-uptake in a set of maize recombinant inbred lines. 2. Line per se performance and comparison with testcross performance. Maydica **53:** 29-38

Dumas JBA (1826) Annales de chimie, Vol 33

Fernandez MGS, Hamblin MT, Li L, Rooney WL, Tuinstra MP, Kresovich S (2008) Quantitative trait loci analysis of endosperm color and carotenoid content in sorghum grain. Crop Science **48:** 1732-1743

Gallais A, Coque M (2005) Genetic variation and selection for nitrogen use efficiency: a synthesis. Maydica **50:** 531-547

Gallais A, Hirel B (2004) An approach to the genetics of nitrogen use efficiency in maize. Journal of Experimental Botany **55:** 295-306

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research **40:** D1178-D1186

Goolsby DA, Battaglin WA (2000) Nitrogen in the Mississippi basin: Estimating sources and predicting flux to the Gulf of Mexico. *In* U.S. Geological Survey Fact Sheet, FS-135-00. U.S. Geological Survey, Reston, Virginia

Haferkamp I, Fernie AR, Neuhaus HE (2011) Adenine nucleotide transport in plants: much more than a mitochondrial issue. Trends in Plant Science **16:** 507-515

Hallauer AR, Carena MJ, Filho JBM (2010) Quantitative genetics in maize breeding, Vol 6. Iowa State University Press, Ames, IA

Hellerstein J (2008) Quantitative data cleaning for large databases. *In*. Technical report, United Nations Economic Commission for Europe

Hirel B, Bertin P, Quillere I, Bourdoncle W, Attagnant C, Dellay C, Gouy A, Cadiou S, Retailliau C, Falque M, Gallais A (2001) Towards a better understanding of the genetic and physiological basis for nitrogen use efficiency in maize. Plant Physiology **125:** 1258-1270

Huang Y-F, Madur D, Combes V, Ky CL, Coubriche D, Jamin P, Jouanne S, Dumas F, Bouty E, Bertin P, Charcosset A, Moreau L (2010) The genetic architecture of grain yield and related traits in Zea mays L. revealed by comparing intermated and conventional populations. Genetics **186:** 395-404

Hussain T, Tausend P, Graham G, Ho J (2007) Registration of IBM2 SYN10 doubled haploid mapping population of maize. Journal of Plant Registrations **1:** 81-81

Jansen C, Zhang Y, Liu H, Gonzalez-Portilla PJ, Lauter N, Kumar B, Trucillo-Silva I, Martin JPS, Lee M, Simcox K, Schussler J, Dhugga K, Luebberstedt T (2015) Genetic and agronomic assessment of cob traits in corn under low and normal nitrogen management conditions. Theoretical and Applied Genetics **128:** 1231-1242

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugenics **12:** 172-175

Lawrence CJH, L.C., Schaeffer ML, Sen TZ, Seigfried TE, Campbell DA (2008) MaizeGDB: the maize model organism database for basic, translational, and applied research. International Journal of Plant Genomics **496957:** 1-10

Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A (2002) Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant Molecular Biology **48:** 453-461

Lima MDA, de Souza CL, Bento DAV, de Souza AP, Carlini-Garcia LA (2006) Mapping QTL for grain yield and plant traits in a tropical maize population. Molecular Breeding **17:** 227-239

Limami AM, Rouillon C, Glevarec G, Gallais A, Hirel B (2002) Genetic and physiological analysis of germination efficiency in maize in relation to nitrogen metabolism reveals the importance of cytosolic glutamine synthetase. Plant Physiology **130:** 1860-1870

Liseron-Monfils C, Bi YM, Downs GS, Wu W, Signorelli T, Lu G, Chen X, Bondo E, Zhu T, Lukens LN, Colasanti J, Rothstein SJ, Raizada MN, Wu WQ, Lu GW, Chen X, Zhu T (2013) Nitrogen transporter and assimilation genes exhibit developmental stage-selective expression in maize (Zea mays L.) associated with distinct cis-acting promoter motifs. Plant Signaling and Behavior **8:** 1-14

Liu H, Niu Y, Gonzalez-Portilla PJ, Zhou H, Yang X, Zuo T, Wang L, Zhang S, Li X, Qin C, Tai S, Jansen C, Shen Y, Lin H, Lee M, Ware D, Zhang Z, Lübberstedt T, Pan G (2015) Genome wide genetic marker discovery and genotyping using next-generation sequencing in maize IBM Syn10 mapping population. BioMed Central**:** (in press)

Liu R, Zhang H, Zhao P, Zhang Z, Liang W, Tian Z, Zheng Y (2012) Mining of candidate maize genes for nitrogen use efficiency by integrating gene expression and QTL data. Plant Molecular Biology Reporter **30:** 297-308

Liu Z, Tang J, Wang C, Tian G, Wei X, Hu Y, Cui D, Liu ZH, Tang JH, Wang CL, Tian GW, Wei XY, Hu YM, Cui DQ (2007) QTL analysis of plant height under N-stress and N-input at different stages in maize. Acta Agronomica Sinica **33:** 782-789

Masclaux C, Quillere I, Gallais A, Hirel B (2001) The challenge of remobilisation in plant nitrogen economy. A survey of physio-agronomic and molecular approaches. Annals of Applied Biology **138:** 69-81

Messmer R, Fracheboud Y, Baenziger M, Vargas M, Stamp P, Ribaut J-M (2009) Drought stress and tropical maize: QTL-by-environment interactions and stability of QTLs across environments for yield components and secondary traits. Theoretical and Applied Genetics **119:** 913-930

Mikel MA, Dudley JW (2006) Evolution of north American dent corn from public to proprietary germplasm. Crop Science **46:** 1193-1205

Nielsen RL (2006) N loss mechanisms and nitrogen use efficiency. *In*, Ed Nitrogen management workshops. Purdue Agriculture Agronomy Extension, West Lafayette, IN

Nischal L, Mohsin M, Khan I, Kardam H, Wadhwa A, Abrol YP, Iqbal M, Ahmad A (2012) Identification and comparative analysis of microRNAs associated with low-N tolerance in rice genotypes. Plos One **7:** e50261

Ouyang J, Cai Z, Xia K, Wang Y, Duan J, Zhang M (2010) Identification and analysis of eight peptide transporter homologs in rice. Plant Science **179:** 374-382

Popova TN, de Carvalho M (1998) Citrate and isocitrate in plant metabolism. Biochimica et Biophysica Acta **1364:** 307-325

Raun WR, Johnson GV (1999) Improving nitrogen use efficiency for cereal production. Agronomy Journal **91:** 357-363

RCoreTeam (2014) R: a language and environment for statistical computing. *In*. R Foundation for Statistical Computing, Vienna, Austria

Schloerke B, Crowley J, Cook D, Hofmann H, Wickham H, Briatte F, Marbach M, Thoen E (2014) GGally: extension to ggplot2. *In*. R package version 0.4.8, Vienna, Austria

Schlüter U, Colmsee C, Scholz U, Braeutigam A, Weber APM, Zellerhoff N, Bucher M, Fahnenstich H, Sonnewald U (2013) Adaptation of maize source leaf metabolism to stress related disturbances in carbon, nitrogen and phosphorus balance. BMC genomics **14:** 442

Schlüter U, Mascher M, Colmsee C, Scholz U, Braeutigam A, Fahnenstich H, Sonnewald U (2012) Maize source leaf adaptation to nitrogen deficiency affects not only nitrogen and carbon metabolism but also control of phosphate homeostasis. Plant Physiology **160:** 1384-1406

Semagn K, Beyene Y, Warburton ML, Tarekegne A, Mugo S, Meisel B, Sehabiague P, Prasanna BM (2013) Meta-analyses of QTL for grain yield and anthesis silking interval in 18 maize populations evaluated under water-stressed and well-watered environments. BMC genomics **14:** 313

Teuscher F, Guiard V, Rudolph PE, Brockmann GA (2005) The map expansion obtained with recombinant inbred strains and intermated recombinant inbred populations for finite generation designs. Genetics **170:** 875-879

Trucillo-Silva I, Cook D, Lee M (2015) A methodical approach for identifying outliers in complex data. *In* preparation. Iowa State University

Trucillo-Silva I, Lee M, Abbaraju HKR, Fallis L, Liu H, Dhugga KS (2015) Mapping of QTL for N-metabolism related enzymes and metabolites in a maize testcross population grown in hydroponics: I. Leaf tissue analysis. *In* preparation. Iowa State University

Tuberosa R, Sanguineti MC, Landi P, Michela Giuliani M, Salvi S, Conti S (2002) Identification of QTLs for root characteristics in maize grown in hydroponics and analysis of their overlap with QTLs for grain yield in the field at two water regimes. Plant Molecular Biology **48:** 697-712

Uhart SA, Andrade FH (1995) Nitrogen deficiency in maize: I. Effects on crop growth, development, dry matter partitioning, and kernel set. Crop Science **35:** 1376-1383

Veldboom LR, Lee M (1996) Genetic mapping of quantitative trait loci in maize in stress and nonstress environments .2. Plant height and flowering. Crop Science **36:** 1320-1327

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. Journal of Heredity **93:** 77-78

Wickham H (2009) ggplot2: elegant graphics for data analysis. *In*. Springer, New York

Witte CP, Rosso MG, Romeis T (2005) Identification of three urease accessory proteins that are required for urease activation in Arabidopsis. Plant Physiology **139:** 1155-1162

Yin X, McClure MA, Jaja N, Tyler DD, Hayes RM (2011) In-season prediction of corn yield using plant height under major production systems. Agronomy Journal **103:** 923-929

Zhang N, Gibon Y, Wallace JG, Kruger Lepak N, Li P, Dedow L (2015) Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population. Plant Physiology **168:** 575-583

Figures

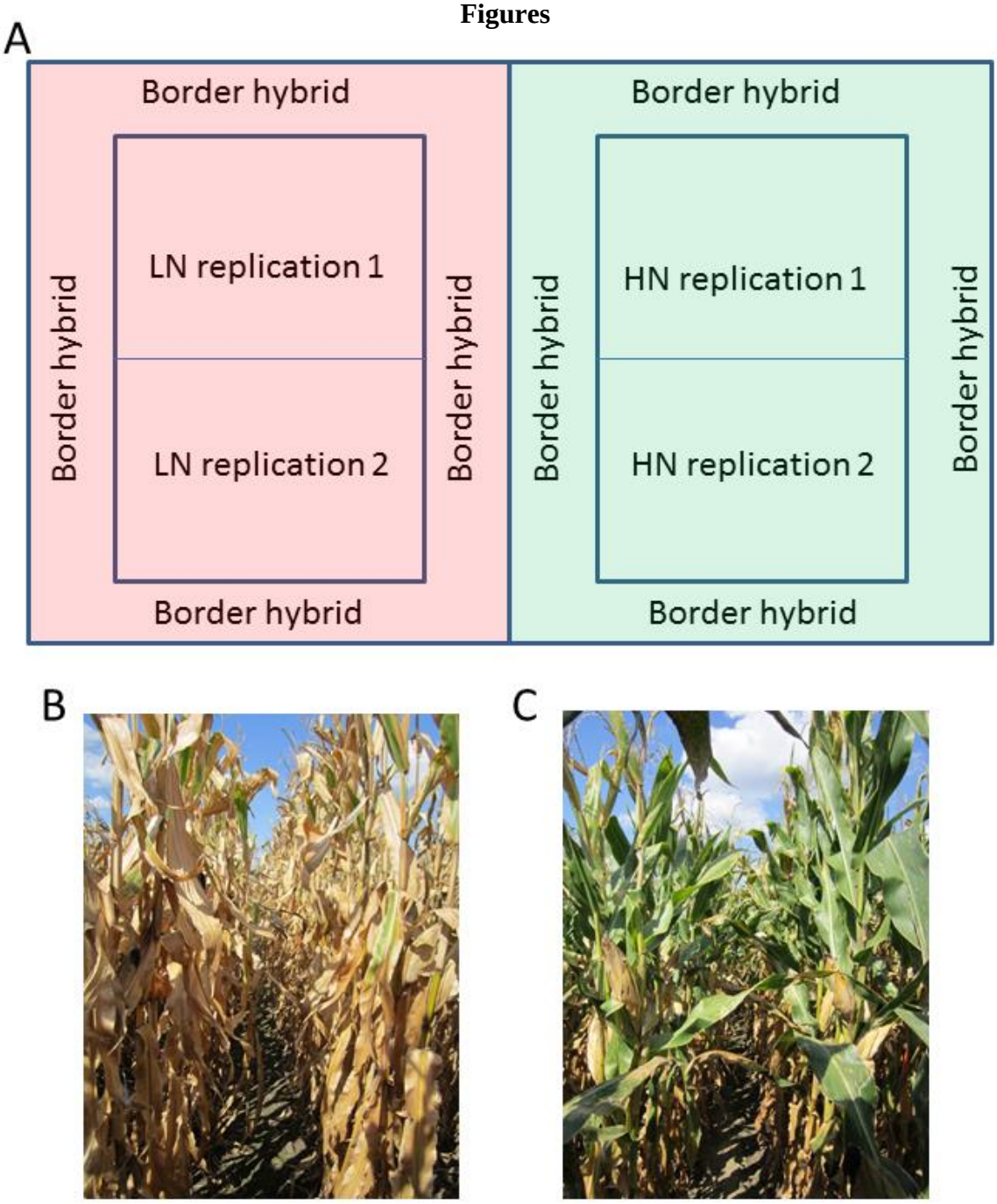

Figure 5.1. Experimental design and N-treatment effect on the maize TC IBMSyn-10 DH population.

(A) Layout of field experiments in a single location. LN on the left (red) and HN on the right (green); (B) LN effect on a random plot at Johnston, 2012; (C) HN effect on a random plot grown at Johnston, 2012.

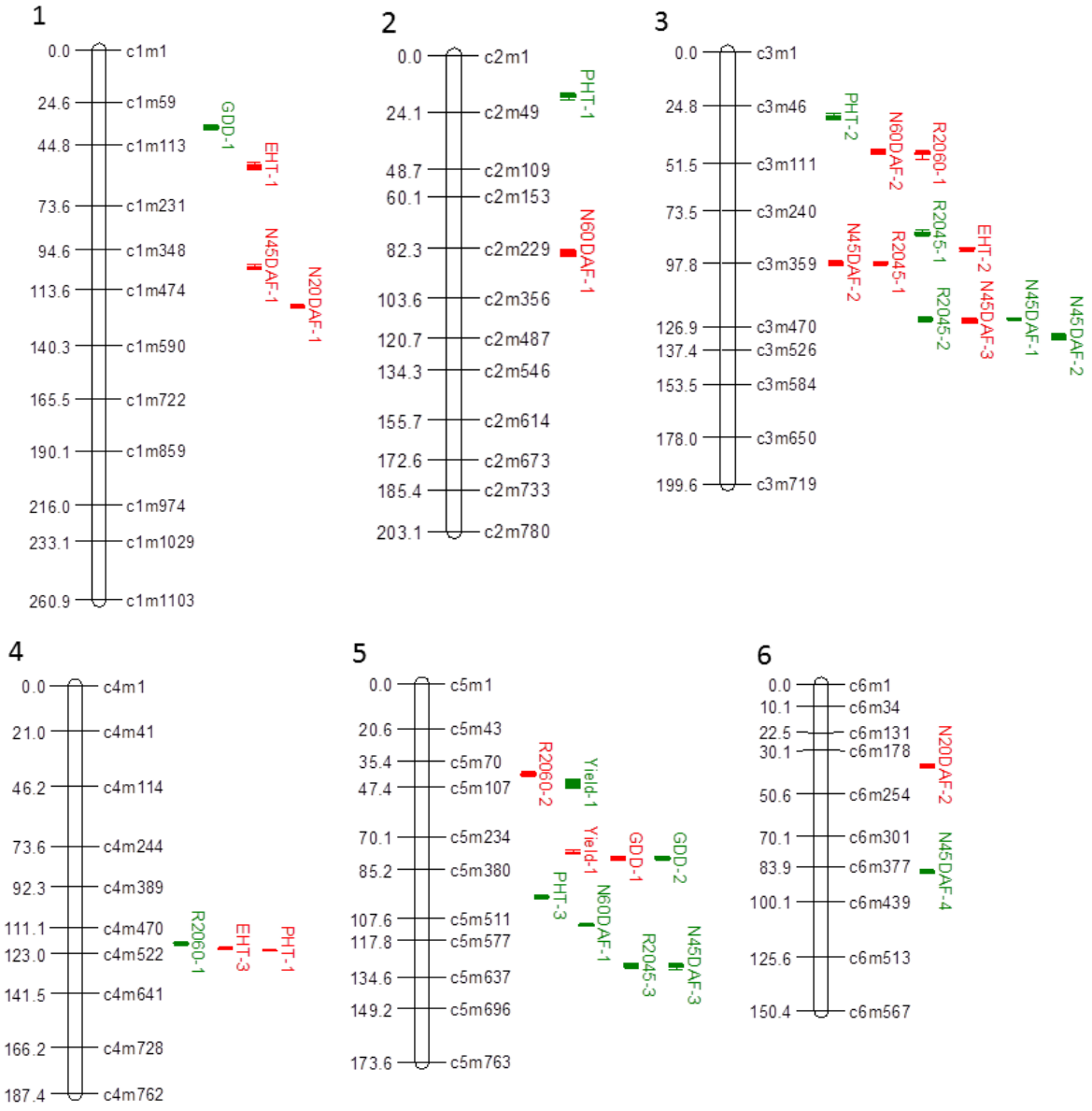

Figure 5.2. Genetic map and distribution of QTL identified across experiments at LN and HN in the IBMSyn10-DH population of maize.

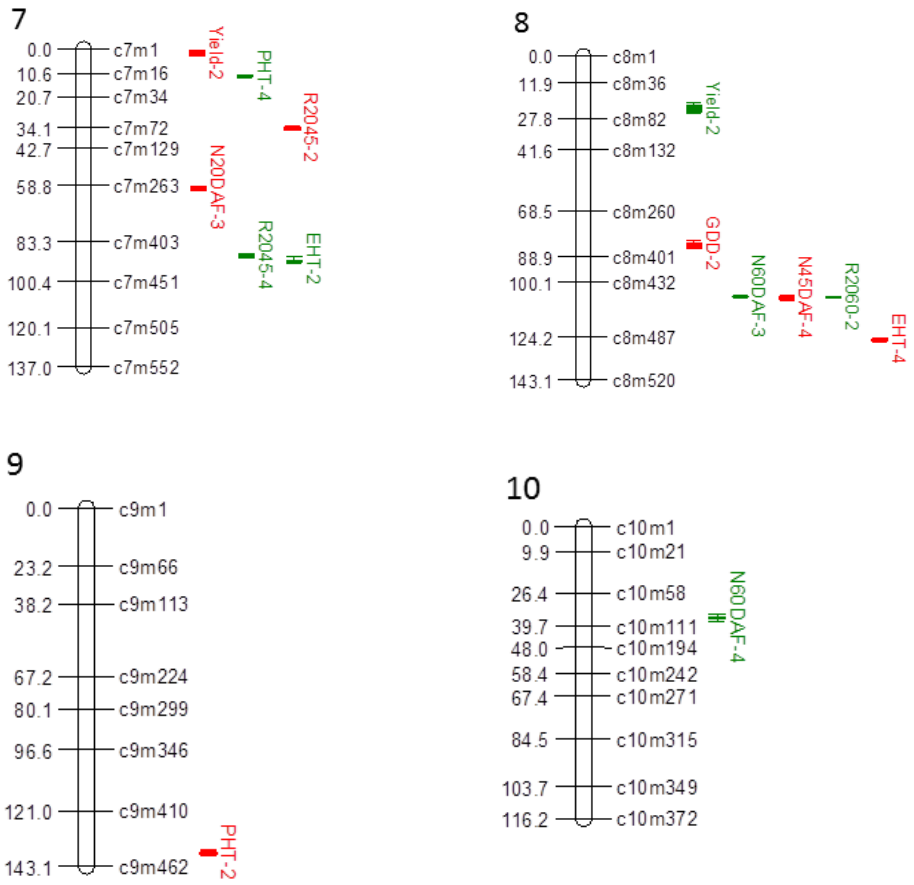

Figure 5.2 continued.

QTL depicted in red were identified under LN and in green under HN conditions. QTL positions shown at right of chromosomes (in cM) and lengths of bars are determined by 2-LOD confidence intervals. Only selected markers displayed in the figure. QTL names correspond to name of the trait followed by QTL number. Figure created with MapChart 2.2 (Voorrips, 2002).

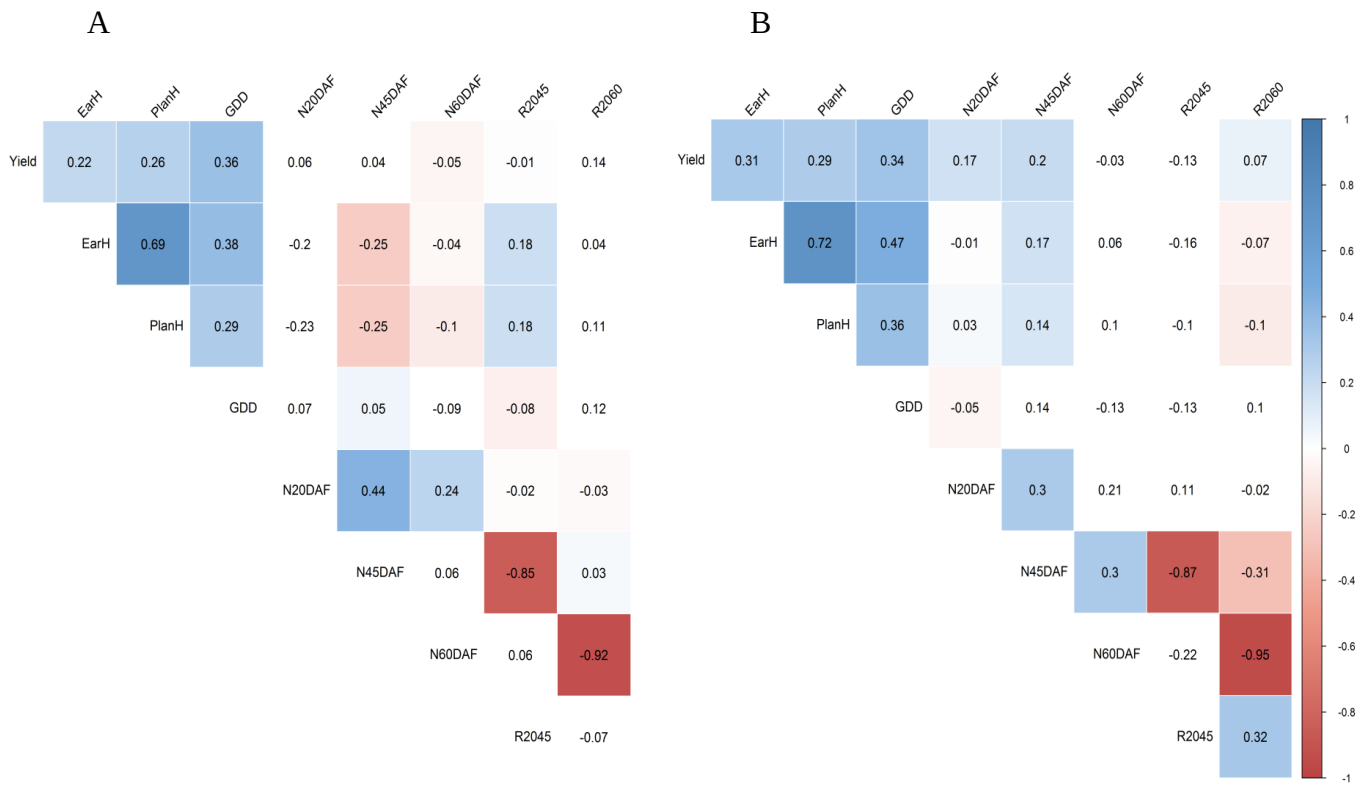

Figure 5.3. Correlation matrix-heatmap of N-metabolism related traits in the IBMSyn10-DH TC population of maize across locations at each LN (panel A) and HN treatment (panel B).

Significant correlation values (p-value<0.05) are colored in blue (positive correlation) and red (negative correlation).

Tables

Table 5.1. Statistical analysis of field traits measured on the IBMSyn10-DH TC population across experiments.

Trait	Unit	Trt	n[a]	Pop μ[b]	Min[e]	Max[f]	SD[g]	B73TC[c]	Mo17TC[d]	CV %	G effect P[i]	Rptblity[j]
Yield	MT Ha^{-1}	LN	172	6.00	5.44	6.44	0.18	7.81	4.11	3.06	<1.00E-17	0.145
		HN	172	11.24	10.31	11.96	0.29	11.19	9.11	2.58	<1.00E-17	0.299
EHT	cm	LN	176	112.59	106.00	118.48	2.58	118.37	114.14	2.29	0.0005003	0.402
		HN	176	113.52	105.19	122.79	3.57	123.07	103.38	3.14	9.72E-09	0.486
PHT	cm	LN	175	263.41	253.65	271.26	2.86	265.87	258.25	1.09	0.0009248	0.386
		HN	175	268.16	262.42	276.53	2.61	270.91	256.31	0.97	0.000651	0.327
GDD	GDD[c]	LN	172	795.29	779.75	812.31	5.70	799.04	812.52	0.72	<1.00E-17	0.448
		HN	172	792.64	779.34	805.03	5.41	804.47	798.79	0.68	<1.00E-17	0.391
N20DAF	%	LN	176	2.93	2.85	3.03	0.02	3.04	2.96	0.82	<1.00E-17	0.170
		HN	176	3.65	3.55	3.74	0.04	3.65	3.59	1.03	<1.00E-17	0.256
N45DAF	%	LN	173	1.87	1.67	2.01	0.06	1.88	1.94	3.38	<1.00E-17	0.346
		HN	174	2.58	2.09	3.01	0.15	2.69	2.24	5.77	<1.00E-17	0.691
N60DAF	%	LN	176	1.13	1.07	1.22	0.03	1.04	1.11	2.30	0.0374378	0.191
		HN	176	1.93	1.79	2.05	0.05	1.86	1.80	2.51	0.0122585	0.230
R2045	%	LN	176	35.99	31.55	41.88	2.05	37.43	32.96	5.69	6.15E-11	0.368
		HN	176	27.19	17.63	39.74	3.72	23.25	36.32	13.68	<1.00E-17	0.631
R2060	%	LN	176	61.17	58.59	63.16	0.85	64.90	63.15	1.39	0.0387483	0.191
		HN	176	49.89	47.12	53.14	1.16	50.25	46.75	2.33	0.0126479	0.230

[a] Population size, [b] Population mean, [c] Minimum value, [d] Maximum value, [e,f] BLUP value for parental genotypes in testcross genotype, [g] Standard deviation, [h] Coefficient of variation (%), [i] p value of the genetic effect, [j] Repeatability.

Table 5.2. QTL identified by CIM across experiments under LN and HN conditions associated with N-metabolism related traits in the IBMSyn10-DH TC population of maize ordered by trait.

QTL name	Chr[a]	Marker #[b]	Pos (cM)[c]	G Interval[d]	Adj (cM)[e]	P pos (Mb)[f]	P interval (Mb)[g]	LOD	R2 (%)	Add[h]	# Genes[i]
Yield-LN-1	5	254	503.62	501.69-505.66	77.48	80.35	77.45-80.45	5.7	8.13	-0.05	117
Yield-LN-2	7	4	14.14	8.53-19.02	2.18	2.05	1.95-2.15	7.55	11.44	0.07	16
Yield-HN-1	5	98	302.18	288.83-310.72	46.49	12.4	11.95-13.15	5.55	7.99	0.09	78
Yield-HN-2	8	54	146.98	140.55-160.27	22.61	8.35	8-8.45	4.14	5.86	0.08	34
EHT-LN-1	1	125	357.57	352.04-367.33	55.01	21.85	21.75-23.15	4.23	5.96	-0.66	62
EHT-LN-2	3	270	589.54	587.17-594.3	90.7	159.35	159.25-159.85	5.72	8.16	0.81	13
EHT-LN-3	4	392	785.54	782.61-785.69	120.85	181.25	180.95-181.35	8.5	12.58	0.99	12
EHT-LN-4	8	361	816.69	812.14-820.26	125.64	171.15	171-171.15	4.49	6.35	0.7	20
EHT-HN-1	7	310	593.06	590.94-596.79	91.24	160.55	160.45-160.65	5.3	8.02	1.19	15
PHT-LN-1	4	395	791.24	789.73-791.77	121.73	181.55	181.45-181.7	6.9	10.74	1.5	11
PHT-LN-2	9	369	896.52	892.53-899.7	137.93	153.85	153.75-154.05	6.28	10.31	1.01	31
PHT-HN-1	2	36	109.27	105.71-116.91	16.81	4.75	4.65-4.95	4.65	5.93	0.65	34
PHT-HN-2	3	50	193.03	191.22-203.42	29.7	6.15	6.05-6.4	8	10.55	0.93	16
PHT-HN-3	5	360	636.83	635.04-638.07	97.97	171.9	171.5-172.25	4.35	5.52	-0.64	34
PHT-HN-4	7	21	74.9	74.35-77.28	11.52	3.85	3.75-3.95	8.91	12.04	-1.25	8
GDD-LN-1	5	266	517.59	514.9-527.11	79.63	82.75	81.95-88	5.65	8.16	-1.7	177
GDD-LN-2	8	255	541.56	539.36-551.56	83.32	123.45	123.25-124.25	5.74	8.48	1.71	57
GDD-HN-1	1	77	239.22	233.96-245.16	36.8	12.4	12.25-12.6	4.64	7.43	1.74	41
GDD-HN-2	5	269	521.28	517.95-524.07	80.2	85.95	83.05-86.85	9.13	14.06	-2.11	101
N20DAF-LN-1	1	428	789.74	785.69-794.81	121.5	180.45	179.95-182.15	7.75	9.78	-0.01	79
N20DAF-LN-2	6	161	248.75	240.08-251.28	38.27	95.45	95.35-96.55	4.61	5.56	0.01	50
N20DAF-LN-3	7	202	388.01	386.06-390.95	59.69	124.25	123.75-124.85	8	10.12	-0.01	31
N45DAF-LN-1	1	350	671.53	666.56-674.73	103.31	91.05	88.55-91.15	5.79	7.3	-0.02	59
N45DAF-LN-2	3	298	631.41	626.87-633.99	97.14	167.2	166.45-167.65	6.1	7.72	-0.02	48

Table 5.2 continued.

QTL name	Chr[a]	Marker #[b]	Pos (cM)[c]	G Interval[d]	Adj (cM)[e]	P pos (Mb)[f]	P interval (Mb)[g]	LOD	R2 (%)	Add[h]	# Genes[i]
N45DAF-LN-3	3	389	807.07	797-813.14	124.16	184.05	183.15-184.25	8.64	11.32	-0.02	53
N45DAF-LN-4	8	336	693.92	691.02-700.18	106.76	166.55	166.05-166.75	7.47	9.55	-0.02	69
N45DAF-HN-1	3	386	800.47	798.9-802.98	123.15	183.65	183.45-183.8	6.54	8.37	-0.05	17
N45DAF-HN-2	3	421	856.31	845.71-861	131.74	190.35	188.65-190.65	4.78	5.97	0.04	88
N45DAF-HN-3	5	481	840.5	837.59-844.72	129.31	196.55	196.55-196.75	7.88	10.56	0.05	11
N45DAF-HN-4	6	292	563.49	557.68-566.8	86.69	131.45	130.2-131.9	5.82	7.38	0.04	68
N60DAF-LN-1	2	208	548.52	543.4-555.64	84.39	41.05	40.45-41.85	4.17	6.5	-0.01	60
N60DAF-LN-2	3	92	302.3	297.11-306.22	46.51	11.95	11.5-12.05	4.24	6.42	0.01	20
N60DAF-HN-1	5	414	719.78	718.51-722.55	110.74	183.45	183.15-183.85	6.88	10.34	-0.02	65
N60DAF-HN-2	8	337	692.82	691.02-695.1	106.59	166.65	166.55-166.75	10.18	16	-0.02	20
N60DAF-HN-3	10	81	236.21	235-239.13	36.34	11.15	11.05-11.25	5.24	7.71	0.01	5
R2045-LN-1	3	298	632.77	627.41-636.04	97.35	167.2	166.3-167.65	8.43	11.56	0.73	53
R2045-LN-2	7	69	223.94	220.4-228.16	34.45	10.05	9.85-10.35	4.51	5.9	-0.56	23
R2045-HN-1	3	248	542.45	541.4-551.56	83.45	152.35	152.25-155.75	4.2	5.21	-0.93	137
R2045-HN-2	3	386	801.38	796.85-806.02	123.29	183.65	183.15-183.95	4.77	5.95	0.92	42
R2045-HN-3	5	483	843.54	836.55-845.51	129.78	196.75	196.65-197.25	7.98	10.39	-1.33	19
R2045-HN-4	7	303	579.52	572.28-583.08	89.16	159.85	158.95-159.95	4.79	6.39	-1.08	87
R2060-LN-1	3	93	304.01	300.14-309.31	46.77	12.05	11.85-12.35	5	7.63	-0.25	32
R2060-LN-2	5	87	270.84	266.53-274.69	41.67	10.35	10.15-11.55	5.7	8.8	0.26	77
R2060-HN-1	4	382	769.66	768.36-772.45	118.41	179.75	179.65-179.95	4.15	6.95	-0.31	28
R2060-HN-2	8	337	692.82	691.01-694.05	106.59	166.65	166.55-166.75	7.92	13.02	0.47	20

[a] Chromosome number, [b] Marker localized at LOD peak, [c] Genetic position of SNP in cM, [d] 1-LOD interval in cM, [e] Adjusted genetic position, [f] Physical position in Mb, [g] 1-LOD Physical interval, [h] Additive effect of respective QTL (a positive-signed effect represents an increasing allele from B73, while a negative-signed allele denotes an increasing allele from Mo17), [i] Number of genes annotated underlying 1-LOD QTL CI. QTL names correspond to trait name followed by experiment number, N treatment and last number being QTL number for the respective trait. QTL names with a rectangle were also identified when analyzing each experiment and treatment combination separately.

Table 5.3. Multiple QTL models per trait analyzed across experiments in the IBMSyn10-DH population of maize.

Trait	Treatment	# QTL in model[a]	Model R^2 (%)[b]	R^2 Epistasis (%)[c]
Yield	LN	2	19.38	
	HN	2	15.32	
EHT	LN	4	25.00	
	HN	1	8.02	
PHT	LN	2	8.94	
	HN	4	22.62	
GDD	LN	2	14.21	
	HN	2	9.39	
N20DAF	LN	3	23.15	
	HN	0	0.00	
N45DAF	LN	4	32.18	
	HN	4	31.19	
N60DAF	LN	2	8.15	
	HN	3	18.54	
R2045	LN	2	12.69	
	HN	4	33.45	3.50
R2060	LN	2	5.72	
	HN	2	13.35	

[a] Number of significant QTL fitted in MIM model, [b] Total R^2 obtained by fitting significant QTL simultaneously in a MIM model, [c] R^2 explained by epistasis solely.

Table 5.4. Candidate genes associated with N-metabolism within identified QTL genomic regions across experiments in the IBMSyn10-DH population of maize.

Maize GDB ID	Corresponding gene annotation	Chr[a]	Start[b]	End[c]	QTL name
GRMZM2G007909	Nucleotide-sugar transporter family protein	7	159902823	159930978	R2045-HN-4
GRMZM2G021605	Phosphoglycerate mutase family protein	4	181259909	181262496	EHT-LN-3
GRMZM2G043193	Ammonium transporter 2	5	171561353	171564218	PHT-HN-3
GRMZM2G047119	ABC-2 and Plant PDR ABC-type transporter family protein	10	11047269	11048312	N60DAF-HN-3
GRMZM2G055216	Nucleotide-sugar transporter family protein	7	160597155	160600964	EHT-HN-1
GRMZM2G059124	Urease accessory protein D	5	85905051	85906027	GDD-HN-2
GRMZM2G063452	Urease accessory protein D	5	85896081	85896580	GDD-LN-1
GRMZM2G066413	Glucose-6-phosphate/PEP/P translocator-related protein	3	12295538	12298243	R2060-LN-1
GRMZM2G076593	Amino acid transporter	10	11087001	11089279	N60DAF-HN-3
GRMZM2G085411	Major facilitator superfamily protein, peptide transporter PTR2	1	180424719	180427842	N20DAF-LN-1
GRMZM2G086258	Dicarboxylate transport 2.1, citrate transporter	1	181895436	181907817	N20DAF-LN-1
GRMZM2G109383	Phosphoglucomutase/phosphomannomutase family protein	5	10865997	10872126	R2060-LN-2
GRMZM2G138698	Acid phosphatase 27, nucleotide	8	171124026	171128225	EHT-LN-4
GRMZM2G138756	Acid phosphatase 24, nucleotide pyrophosphatase/phosphodiesteras	8	171131158	171136471	EHT-LN-4
GRMZM2G143190	Major facilitator superfamily protein	1	90912230	90914709	N45DAF-LN-1
GRMZM2G154211	Sulfate transporter 3;1	1	12350236	12355974	GDD-HN-1
GRMZM2G170326	Magnesium transporter 2	6	95771854	95778367	N20DAF-LN-2
GRMZM2G326259	Potassium transporter	6	130519264	130520551	N45DAF-HN-4
GRMZM2G345226	Potassium uptake permease 6, K transporter	3	154110005	154112525	R2045-HN-1
GRMZM2G396550	Potassium uptake transporter 3, K transporter	3	154125169	154128809	R2045-HN-1
GRMZM2G433162	Amino acid permease 2, amino acid transporter	10	11200703	11203045	N60DAF-HN-3
GRMZM5G843192	ABC transporter family protein	5	172066957	172071242	PHT-HN-3
GRMZM5G886294	Adenine nucleotide transporter 1	5	83208803	83213993	GDD-LN-1

[a] Chromosome, [b,c] start and end location in bp

APPENDIX B

Figures

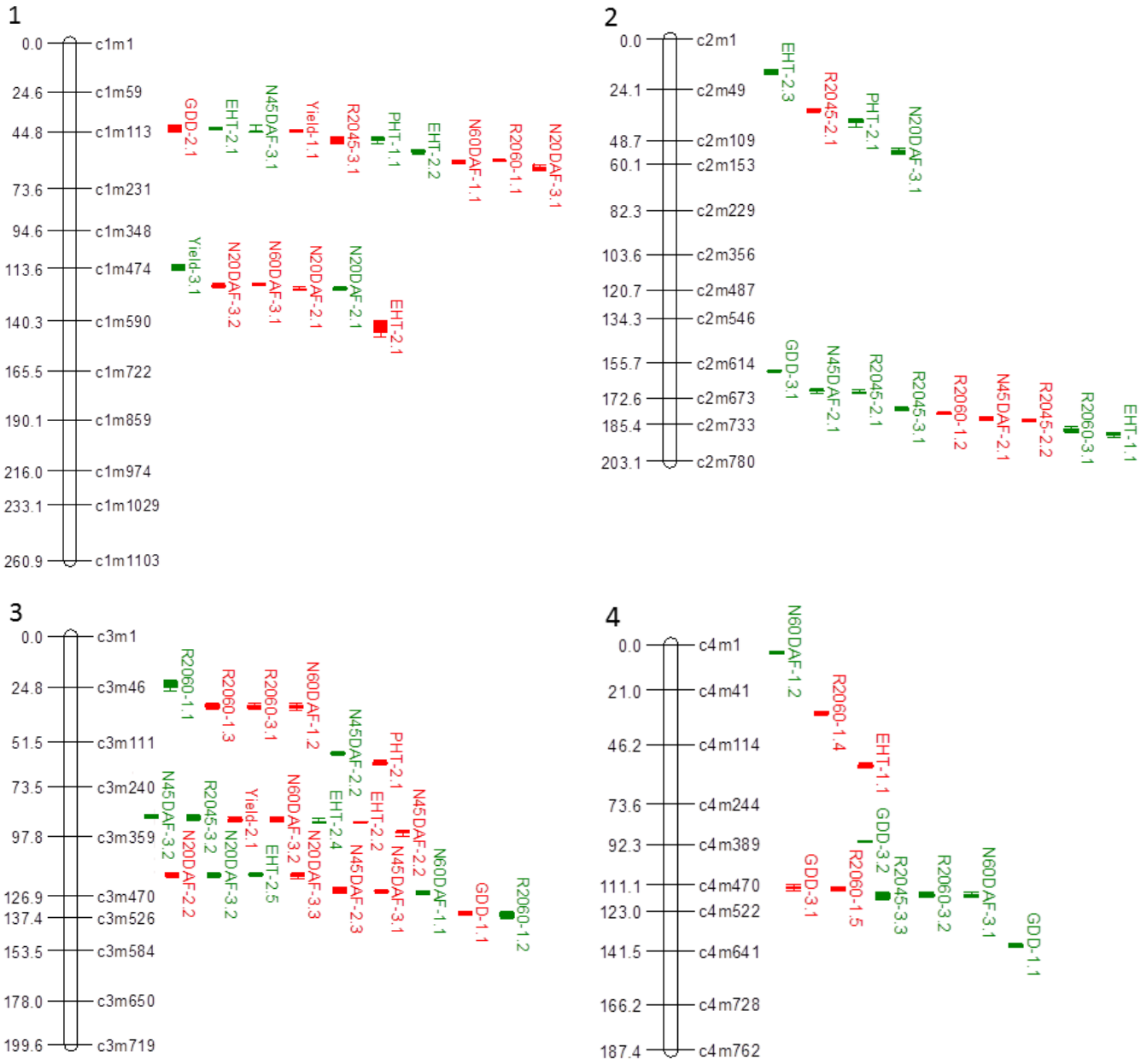

Figure A5.1. Genetic map and distribution of QTL identified at each experiment at low N and high N conditions in the IBMSyn10-DH population of maize.

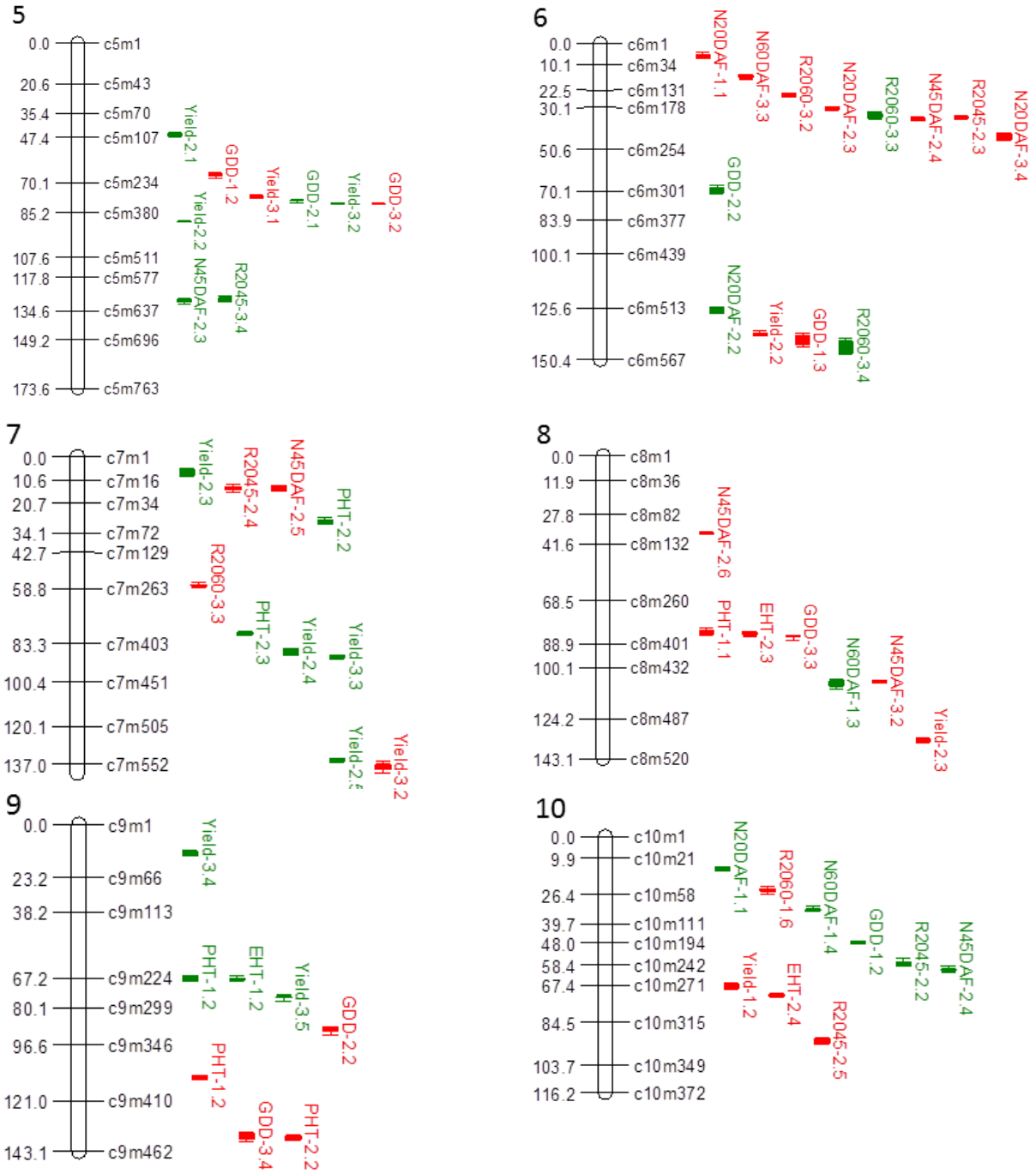

Figure A5.1 continued.

QTL depicted in red were identified under LN and in green under HN conditions. QTL positions shown at right of chromosomes (in cM) and lengths of bars are determined by 2-LOD confidence intervals. Only selected markers displayed in the figure. QTL names correspond to key trait name followed by experiment and QTL number correspondingly. Figure created with MapChart 2.2 (Voorrips, 2002).

Tables

Table A5.1. Average monthly temperatures (°C) by experiment.

	Month											
Experiment	1	2	3	4	5	6	7	8	9	10	11	12
1	-8.92	-3.97	2.78	8.83	16.06	21.50	26.39	23.22	16.53	13.06	5.33	-0.28
2	-2.06	-0.53	12.33	11.97	19.69	24.58	24.56	22.50	17.47	8.72	4.53	-0.78
3	-2.89	-1.08	12.11	12.47	19.14	22.89	27.58	22.64	17.64	10.39	5.31	-2.14

Data extracted from http://www.usclimatedata.com/climate

Table A5.2. Monthly precipitation (mm) by experiment.

	Month												
Experiment	1	2	3	4	5	6	7	8	9	10	11	12	Total
1	22.86	17.53	36.83	108.97	153.16	190.25	48.26	69.34	21.59	11.18	66.55	57.15	803.66
2	32.00	16.26	66.80	91.69	34.54	72.39	34.04	75.44	53.34	87.63	44.70	29.21	638.05
3	10.16	42.42	50.29	130.05	91.69	68.33	55.63	81.53	42.16	51.31	32.00	44.96	700.53

Data extracted from http://www.usclimatedata.com/climate

Table A5.3. Analysis of variance of field traits measured on the IBMSyn10-DH TC population.

Trait	Source[a]	DF	F value	p-value	
Yield	E	2	349.06	< 2.2e-16	***
	T (E)	3	569.57	< 2.2e-16	***
	R (T)	2	40.05	< 2.2e-16	***
	G	177	2.11	1.63E-13	***
	GxE	345	1.23	0.006	**
	GxT	176	0.81	0.964	
EHT	E	1	3628.08	< 2.2e-16	***
	T (E)	2	76.39	< 2.2e-16	***
	R (T)	2	9.35	9.93E-05	***
	G	176	4.56	< 2.2e-16	***
	GxE	169	4.44	< 2.2e-16	***
	GxT	176	1.05	0.346	
PHT	E	1	10161.39	< 2.2e-16	***
	T (E)	2	31.26	1.07E-13	***
	R (T)	2	36.58	8.65E-16	***
	G	176	4.06	< 2.2e-16	***
	GxE	169	4.22	< 2.2e-16	***
	GxT	176	1.09	0.23	
GDD	E	2	5535.09	< 2.2e-16	***
	T (E)	3	62.06	< 2.2e-16	***
	R (T)	2	18.74	9.29E-09	***
	G	177	8.13	< 2.2e-16	***
	GxE	345	4.57	< 2.2e-16	***
	GxT	177	1.03	0.37	
N20DAF	E	2	55.84	< 2.2e-16	***
	T (E)	3	1762.46	< 2.2e-16	***
	R (T)	2	19.40	4.90E-09	***
	G	177	2.04	2.16E-12	***
	GxE	346	1.77	6.69E-13	***
	GxT	177	1.11	0.15	
N45DAF	E	1	499.59	< 2.2e-16	***
	T (E)	2	1955.93	< 2.2e-16	***
	R (T)	2	23.71	9.26E-11	***
	G	177	4.48	< 2.2e-16	***
	GxE	177	1.53	6.09E-05	***
	GxT	177	1.53	6.17E-05	***
N60DAF	E	1	22.60	2.34E-06	***
	T (E)	2	1555.64	< 2.2e-16	***

Table A5.3 continued.

Trait	Source[a]	DF	F value	p-value	
	R (T)	2	11.26	1.49E-05	***
	G	177	1.61	8.98E-06	***
	GxE	169	1.49	0.000233	***
	GxT	177	1.11	0.17	
R2045	E	1	529.10	< 2.2e-16	***
	T (E)	2	490.79	< 2.2e-16	***
	R (T)	2	30.93	1.03E-13	***
	G	177	4.16	< 2.2e-16	***
	GxE	177	1.39	0.001569	**
	GxT	177	1.29	0.01	*
R2060	E	1	3.72	0.05	.
	T (E)	2	773.70	< 2.2e-16	***
	R (T)	2	16.05	1.43E-07	***
	G	177	1.67	1.71E-06	***
	GxE	169	1.58	2.65E-05	***
	GxT	177	1.10	0.19	

[a] Sources of variation: E, environment, T, treatment, R, replication, G, genotype, GxE, genotype by environment interaction, GxT, genotype by treatment interaction; * p-value <0.05, ** p-value<0.01 and *** p-value<0.001.

Table A5.4. Statistical analysis of field traits measured on the IBMSyn10-DH TC population at each experiment by treatment combination.

Trait	Unit	Exp	Trt	n[a]	Pop μ[b]	Min[c]	Max[d]	B73TC[e]	Mo17TC[f]	SD[g]	CV%[h]	G effect P[i]	Rptblity[j]
Yield	MT Ha^{-1}	1	LN	163	8.07	6.47	9.30	8.60	7.57	0.55	6.77	8.33E-15	0.4
		1	HN	167	10.01	7.36	12.03	8.07	7.32	0.97	9.64	1.17E-15	0.45
		2	LN	169	6.44	4.42	8.44	9.26	4.49	0.73	11.34	4.44E-16	0.45
		2	HN	173	11.5	9.25	13.62	13.52	8.85	0.84	7.32	<1.00E-17	0.73
		3	LN	174	10.17	8.05	11.41	12.21	9.45	0.55	5.37	<1.00E-17	0.61
		3	HN	171	12.14	8.58	14.18	14.09	10.38	0.97	8.02	<1.00E-17	0.68
EHT	cm	1	LN	164	112.01	86.45	136	131.04	112.55	10.4	9.28	3.30E-05	0.83
		1	HN	160	113.8	99.98	132.31	123.92	102.33	6.49	5.7	1.92E-10	0.61
		2	LN	175	145.8	126.18	165.46	148.4	144.59	7.46	5.11	<1.00E-17	0.74
		2	HN	175	138.06	122.26	160.52	147.09	129.31	7.26	5.26	<1.00E-17	0.73
PHT	cm	1	LN	162	259.13	231.01	279.67	256.91	245.05	9.91	3.82	2.96E-08	0.5
		1	HN	156	268.22	255.84	278.2	253.04	255.56	4.52	1.69	3.22E-05	0.42
		2	LN	175	323.09	297.64	346.34	330.29	316.11	9.00	2.79	<1.00E-17	0.75
		2	HN	175	326.07	301.31	351.26	348.03	317.55	9.12	2.80	5.55E-17	0.7
GDD	°C	1	LN	168	804.88	767.16	850.83	823.38	771.41	15.64	1.94	<1.00E-17	0.72
		1	HN	164	792.51	744.22	841.48	813.47	793.6	17.89	2.26	<1.00E-17	0.82
		2	LN	171	754.59	728.43	784.27	751.82	785.63	10.6	1.4	1.18E-12	0.61
		2	HN	172	744.95	718.57	769.42	753.59	742.8	10.33	1.39	5.20E-12	0.46
		3	LN	176	829.19	800.73	865.95	825.3	851.36	14.61	1.76	<1.00E-17	0.85
		3	HN	169	834.53	795.44	884.63	837.55	858.95	16.19	1.94	<1.00E-17	0.69
N20DAF	%	1	LN	168	2.93	2.66	3.23	2.92	3.04	0.11	3.74	6.41E-11	0.35
		1	HN	166	3.83	3.57	4.11	3.73	3.64	0.10	2.50	4.56E-08	0.37
		2	LN	172	3.04	2.77	3.37	3.54	3.49	0.10	3.18	4.19E-13	0.49
		2	HN	175	3.37	3.05	3.66	3.54	3.49	0.11	3.39	4.19E-13	0.49

Table A5.4 continued.

Trait	Unit	Exp	Trt	n[a]	Pop μ[b]	Min[c]	Max[d]	B73TC[e]	Mo17TC[f]	SD[g]	CV%[h]	G effect P[i]	Rptblity[j]
N20DAF	%	3	LN	174	2.86	2.69	3.08	3.14	2.86	0.07	2.57	2.86E-06	0.40
		3	HN	176	3.63	3.39	3.91	3.55	3.56	0.09	2.53	2.39E-15	0.58
N45DAF	%	2	LN	175	1.81	1.21	2.22	1.82	1.69	0.18	10.06	1.47E-13	0.61
		2	HN	170	2.34	1.79	2.73	2.47	2.10	0.17	7.24	1.15E-10	0.57
		3	LN	169	1.96	1.66	2.26	2.04	2.14	0.11	5.45	2.45E-12	0.58
		3	HN	174	2.82	2.40	3.22	2.91	2.48	0.14	4.89	5.55E-16	0.65
N60DAF	%	1	LN	166	1.19	0.95	1.46	1.18	1.11	0.09	7.59	8.14E-06	0.40
		1	HN	162	1.90	1.55	2.33	1.61	1.56	0.15	7.96	5.52E-08	0.36
		3	LN	174	1.63	1.38	2.01	1.61	1.69	0.12	7.10	5.67E-11	0.58
		3	HN	173	1.93	1.45	2.34	2.02	1.59	0.14	7.00	3.97E-08	0.52
R2045	%	2	LN	171	40.31	26.76	59.81	41.3	44.16	6.20	15.38	3.57E-14	0.63
		2	HN	170	31.63	22.21	45.66	32.08	40.92	4.70	14.85	4.34E-11	0.59
		3	LN	172	31.84	26.63	37.16	31.92	25.57	2.05	6.43	1.00E-05	0.39
		3	HN	170	21.94	14.08	30.13	17.52	30.85	2.89	13.18	3.27E-10	0.51
R2060	%	1	LN	164	59.6	47.69	67.06	62.85	65.17	3.2	5.36	8.64E-06	0.42
		1	HN	162	50.67	38.66	59.87	56.32	53.07	4.04	7.97	5.81E-08	0.44
		3	LN	174	66.96	59.1	72.96	71.25	63.52	2.81	4.20	6.66E-07	0.49
		3	HN	176	46.75	37.46	57.17	43.85	54.51	3.35	7.17	5.57E-07	0.49

[a] Population size, [b] Population mean, [c] Minimum value, [d] Maximum value, [e,f] BLUP value for parental genotypes in testcross genotype, [g] Standard deviation, [h] Coefficient of variation (%), [i] p-value of the genetic effect, [j] Repeatability

Table A5.5. Pearson pairwise correlation analysis between traits at Experiment 1.

	PHT	EHT	R2060	N60DAF	GDD	Yield	N20DAF
PHT		**0.55**	0.08	-0.11	**0.35**	0.21	-0.08
EHT	**0.60**		0.00	-0.03	**0.44**	0.16	-0.16
R2060	0.01	-0.01		**-0.90**	0.06	-0.02	0.06
N60DAF	-0.02	-0.07	**-0.91**		-0.10	0.05	0.28
GDD	**0.33**	**0.41**	0.09	-0.20		0.21	-0.14
Yield	0.15	0.03	0.08	-0.04	0.07		0.08
N20DAF	0.02	-0.12	0.01	**0.30**	**-0.38**	0.15	

*LN above diagonal and HN below diagonal, coefficients in bold are significant after Bonferroni correction, underlined means p-value<0.05, italics means p-value<0.01 and plain bold means p-value<0.001.

Table A5.6. Pearson pairwise correlation analysis between traits at Experiment 2.

	PHT	EHT	R2045	N45DAF	GDD	Yield	N20DAF
PHT		**0.70**	0.09	-0.09	**0.30**	*0.25*	-0.15
EHT	**0.75**		0.13	-0.13	**0.33**	0.21	-0.20
R2045	-0.02	-0.02		**-0.94**	-0.12	-0.09	0.00
N45DAF	0.04	0.08	**-0.92**		0.08	0.08	*0.28*
GDD	0.25	**0.32**	-0.08	0.09		*0.27*	-0.15
Yield	**0.35**	**0.34**	-0.21	0.21	0.26		-0.06
N20DAF	0.09	0.13	0.01	*0.30*	0.02	0.13	

*LN above diagonal and HN below diagonal, coefficients in bold are significant after Bonferroni correction, underlined means p-value<0.05, italics means p-value<0.01 and plain bold means p-value<0.001.

Table A5.7. Pearson pairwise correlation analysis between traits at Experiment 3.

	R2045	R2060	N60DAF	N45DAF	GDD	Yield	N20DAF
R2045		0.03	-0.02	**-0.75**	0.02	-0.08	0.03
R2060	**0.52**		**-0.94**	-0.11	0.19	-0.09	-0.13
N60DAF	**-0.45**	**-0.94**		0.24	-0.17	0.08	**0.35**
N45DAF	**-0.81**	**-0.41**	**0.48**		-0.08	0.24	**0.53**
GDD	-0.09	0.00	0.05	0.17		0.17	-0.03
Yield	-0.06	0.16	-0.16	0.17	*0.27*		0.13
N20DAF	0.21	0.06	0.18	**0.32**	0.18	0.07	

*LN above diagonal and HN below diagonal, coefficients in bold are significant after Bonferroni correction, underlined means p-value<0.05, italics means p-value<0.01 and plain bold means p-value<0.001.

Table A5.8. QTL identified by CIM at each experiment by treatment combination associated with N-metabolism related traits in the IBMSyn10-DH TC population of maize ordered by trait, experiment and N treatment.

QTL name	Chr[a]	Marker[b]	G Pos (cM)[c]	G Interval(cM)[d]	Adj (cM)[e]	P pos (Mb)[f]	P Interval (Mb)[g]	LOD	R^2 (%)	Add[h]	#Genes[i]
Yield1LN-1	1	103	290.18	287.79-291.51	44.64	16.45	15.85-16.75	6.19	10.40	-0.19	26
Yield1LN-2	10	226	441.54	433.26-444.64	67.93	136.95	136.65-137.05	4.51	7.24	0.16	18
Yield2LN-1	3	267	584.50	580.06-588.97	89.92	158.75	158.65-159.25	5.12	8.08	0.23	9
Yield2LN-2	6	430	900.77	893.86-901.29	138.58	163.95	163.85-164.55	5.75	8.49	-0.24	44
Yield2LN-3	8	377	873.64	870.85-877.56	134.41	172.65	172.45-172.75	5.56	8.10	0.21	22
Yield2HN-1	5	98	302.18	295.94-307.08	46.49	12.40	11.95-12.70	7.75	11.13	0.30	45
Yield2HN-2	5	326	582.59	581.54-583.77	89.63	162.65	162.35-163.35	6.32	8.60	-0.33	35
Yield2HN-3	7	12	46.53	37.81-53.54	7.16	2.85	2.75-2.95	5.66	8.32	0.25	16
Yield2HN-4	7	301	563.50	558.57-574.71	86.69	158.75	157.55-159.85	5.03	7.18	0.26	108
Yield2HN-5	7	407	882.81	875.03-883.60	135.82	174.55	174.35-174.70	6.03	8.33	0.26	24
Yield3LN-1	5	254	501.87	498.49-508.11	77.21	80.35	77.25-81.75	5.81	8.85	-0.17	116
Yield3LN-2	7	412	895.52	890.76-905.76	137.77	175.45	175.05-175.55	4.80	7.20	0.15	19
Yield3HN-1	1	397	732.65	721.76-737.34	112.72	160.85	148.25-160.95	5.06	5.77	-0.24	261
Yield3HN-2	5	271	523.96	522.46-524.45	80.61	86.85	85.95-87.80	10.45	13.09	-0.66	47
Yield3HN-3	7	304	580.42	575.57-583.17	89.30	159.85	158.95-160.05	5.54	6.36	0.27	58
Yield3HN-4	9	37	79.59	78.15-88.03	12.24	5.25	4.95-5.35	4.52	5.26	-0.23	8
Yield3HN-5	9	226	489.96	488.85-491.81	75.38	113.75	110.05-113.85	4.15	4.67	0.22	131
GDD1LN-1	3	440	881.83	877.77-884.7	135.67	196.45	196.05-196.75	7.95	17.04	-11.36	36
GDD1LN-2	5	183	432.65	428.48-435.40	66.56	37.00	35.95-37.80	4.47	7.45	-4.68	55
GDD1LN-3	6	433	913.29	903.49-929.20	140.51	164.75	164.55-164.85	4.80	9.41	5.40	22
GDD1HN-1	4	438	905.10	899.70-907.82	139.25	211.30	202.35-211.90	6.24	11.61	-7.49	267
GDD1HN-2	10	158	313.47	313.31-314.95	48.23	111.05	110.55-111.55	8.28	15.07	10.42	24
GDD2LN-1	1	94	280.51	273.65-290.98	43.16	15.15	15.05-16.05	4.65	7.33	-2.96	42
GDD2LN-2	9	282	586.23	580.05-587.17	90.19	137.15	136.75-137.45	4.46	6.95	3.16	49

Table A5.8 continued.

QTL name	Chr[a]	Marker[b]	G Pos (cM)[c]	G Interval(cM)[d]	Adj (cM)[e]	P pos (Mb)[f]	P Interval (Mb)[g]	LOD	R^2 (%)	Add[h]	#Genes[i]
GDD2HN-1	5	266	514.59	513.91-514.91	79.17	82.75	82.40-83.05	8.97	13.24	-3.98	16
GDD2HN-2	6	224	460.28	444.68-463.01	70.81	112.45	112.25-112.80	5.15	6.68	2.74	16
GDD3LN-1	4	363	731.02	727.49-733.71	112.46	175.55	174.05-175.95	4.35	4.54	-3.37	57
GDD3LN-2	5	273	525.23	524.91-526.00	80.80	88.00	87.80-88.75	11.73	13.53	-5.64	26
GDD3LN-3	8	268	555.81	552.26-556.64	85.51	131.1	129.00-131.30	9.98	11.23	4.98	80
GDD3LN-4	9	367	884.77	878.29-895.61	136.12	153.65	153.55-153.85	4.86	5.34	3.89	21
GDD3HN-1	2	221	577.01	573.97-579.05	88.77	44.35	44.25-45.95	6.93	9.74	6.10	63
GDD3HN-2	4	291	592.17	591.26-592.60	91.10	153.55	148.65-149.75	10.84	16.10	6.88	31
PHT1LN-1	8	253	544.59	537.32-547.48	83.78	123.05	120.15-123.45	4.74	7.80	-2.79	103
PHT1LN-2	9	331	719.91	717.46-721.55	110.76	145.35	145.25-145.7	5.65	9.46	3.27	21
PHT1HN-1	1	110	316.91	310.31-320.51	48.76	17.95	17.85-18.75	6.48	10.88	-1.58	28
PHT1HN-2	9	190	437.11	433.47-443.68	67.25	91.25	77.45-95.65	4.56	7.43	1.27	359
PHT2LN-1	3	140	404.27	398.86-407.02	62.20	23.00	22.75-23.50	4.98	7.05	2.71	19
PHT2LN-2	9	369	895.52	886.45-896.61	137.77	153.85	153.65-154.05	6.18	8.89	2.92	30
PHT2HN-1	2	86	256.48	253.33-261.45	39.46	11.35	11.25-11.45	5.44	8.06	2.69	15
PHT2HN-2	7	58	193.71	182.17-194.93	29.80	8.65	7.95-8.75	5.46	8.38	-2.76	30
PHT2HN-3	7	270	511.96	508.59-513.48	78.76	147.15	146.30-147.25	5.00	7.36	2.62	42
EHT1LN-1	4	128	362.25	359.16-367.33	55.73	20.05	19.75-20.60	5.05	8.82	-3.42	20
EHT1HN-1	2	399	775.84	771.40-777.53	119.36	188.35	187.55-188.5	5.47	7.44	-1.82	35
EHT1HN-2	9	190	438.86	438.55-440.60	67.52	91.25	90.55-91.35	11.51	17.13	2.73	20
EHT2LN-1	1	498	908.29	907.82-946.52	139.74	199.95	199.80-206.05	4.17	5.24	1.77	217
EHT2LN-2	3	272	591.78	589.22-593.01	91.04	159.70	159.35-159.85	11.13	15.33	3.29	16
EHT2LN-3	8	258	549.58	543.40-550.52	84.55	124.25	123.55-124.75	5.29	6.76	2.07	55
EHT2LN-4	10	238	465.97	461.38-469.64	71.69	138.85	138.75-139.15	4.45	5.63	-2.02	26
EHT2HN-1	1	94	281.90	278.77-284.85	43.37	15.15	15.15-15.25	6.87	9.25	-3.54	7
EHT2HN-2	1	125	358.52	354.08-360.21	55.16	21.85	21.75-22.70	5.26	6.94	-2.05	37
EHT2HN-3	2	35	106.52	99.63-110.79	16.39	4.65	4.45-4.85	6.02	8.41	2.23	27

Table A5.8 continued.

QTL name	Chr[a]	Marker[b]	G Pos (cM)[c]	G Interval(cM)[d]	Adj (cM)[e]	P pos (Mb)[f]	P Interval (Mb)[g]	LOD	R^2 (%)	Add[h]	#Genes[i]
EHT2HN-4	3	270	590.07	588.22-593.30	90.78	159.35	159.25-159.7	6.49	8.70	2.44	10
EHT2HN-5	3	363	757.95	754.12-761.24	116.61	179.35	179.25-179.45	6.49	8.70	2.27	7
N20DAF1LN-1	6	14	40.11	35.48-45.64	6.17	2.65	2.35-3.05	4.17	7.69	-0.03	20
N20DAF1HN-1	10	27	92.84	91.89-96.80	14.28	3.50	3.35-4.05	5.46	9.22	0.03	34
N20DAF2LN-1	1	439	802.93	802.98-806.02	123.53	186.00	183.90-187.55	8.23	11.77	-0.03	127
N20DAF2LN-2	3	360	755.32	753.12-762.24	116.20	178.70	177.45-179.25	4.06	5.79	-0.03	61
N20DAF2LN-3	6	141	199.82	197.34-203.42	30.74	91.75	91.45-92.15	4.10	5.53	0.02	28
N20DAF2HN-1	1	439	800.58	797.90-805.02	123.17	186.00	182.95-187.70	4.88	8.05	-0.03	176
N20DAF2HN-2	6	415	822.41	819.26-830.47	126.52	161.85	161.85-162.05	5.62	8.87	-0.04	15
N20DAF3LN-1	1	149	405.85	402.94-415.15	62.44	26.70	26.15-27.15	4.50	6.55	0.02	36
N20DAF3LN-2	1	427	791.47	789.63-796.85	121.76	180.20	179.95-182.15	6.11	9.07	-0.02	76
N20DAF3LN-3	3	364	758.61	753.12-763.28	116.71	179.45	179.25-180.45	4.05	6.13	-0.02	52
N20DAF3LN-4	6	178	288.52	278.77-298.1	44.39	102.65	98.05-104.75	4.45	6.47	0.02	265
N20DAF3HN-1	2	120	353.16	349.00-359.16	54.33	16.40	16.15-16.55	5.70	7.99	-0.03	14
N20DAF3HN-2	3	362	757.61	753.12-762.24	116.56	179.25	178.70-179.45	4.44	6.18	-0.02	20
N45DAF 2LN-1	2	345	723.76	721.55-728.67	111.35	170.55	170.35-174.15	5.70	6.62	-0.05	131
N45DAF 2LN-2	3	292	622.95	618.75-626.87	95.84	165.85	164.90-166.05	4.47	5.47	-0.04	38
N45DAF 2LN-3	3	384	801.36	796.85-811.1	123.29	183.15	183.00-184.15	5.46	6.63	-0.05	46
N45DAF 2LN-4	6	156	232.33	229.92-237.04	35.74	94.75	94.60-95.05	4.80	5.51	0.05	59
N45DAF 2LN-5	7	25	93.36	88.07-98.86	14.36	4.40	4.20-4.55	5.33	6.17	0.05	12
N45DAF 2LN-6	8	78	238.91	237.04-240.08	36.76	14.05	14.05-14.25	5.14	5.93	-0.07	8
N45DAF 2HN-1	2	269	639.23	636.03-641.11	98.34	97.00	72.15-106.65	6.41	8.92	-0.05	596
N45DAF 2HN-2	3	124	373.04	368.33-375.45	57.39	17.05	16.65-17.05	5.68	7.77	-0.05	16
N45DAF 2HN-3	5	482	841.15	837.59-845.71	129.41	196.65	196.55-197.25	4.45	5.97	0.05	21
N45DAF 2HN-4	10	207	393.58	389.82-396.42	60.55	133.2	132.95-133.35	5.63	7.69	-0.05	18
N45DAF 3LN-1	3	389	809.04	807.06-812.14	124.47	184.05	183.95-184.25	6.53	9.73	-0.03	10
N45DAF 3LN-2	8	335	692.83	692.01-696.10	106.59	166.35	166.15-166.65	8.06	12.27	-0.04	46

Table A5.8 continued.

QTL name	Chr[a]	Marker[b]	G Pos (cM)[c]	G Interval(cM)[d]	Adj (cM)[e]	P pos (Mb)[f]	P Interval (Mb)[g]	LOD	R^2 (%)	Add[h]	#Genes[i]
N45DAF 3HN-1	1	102	292.09	289.94-293	44.94	16.25	15.95-16.60	4.51	6.18	-0.04	21
N45DAF 3HN-2	3	265	570.87	566.8-573.97	87.83	158.55	157.50-158.65	7.96	11.73	0.05	32
N60DAF 1LN-1	1	140	387.75	385.66-390.74	59.65	24.75	24.75-25.10	8.65	15.02	-0.04	10
N60DAF 1LN-2	3	63	223.91	221.75-227.87	34.45	8.15	7.80-8.25	5.85	9.75	0.03	19
N60DAF 1HN-1	3	391	814.04	810.1-819.26	125.24	184.25	184.25-184.65	8.33	13.79	0.07	23
N60DAF 1HN-2	4	7	24.48	22.23-27.36	3.77	1.35	1.35-1.95	7.57	11.81	-0.05	21
N60DAF 1HN-3	8	334	700.49	691.01-707.3	107.77	166.15	165.65-166.65	4.02	5.83	-0.04	69
N60DAF 1HN-4	10	74	215.00	213.88-218.79	33.08	10.05	9.95-10.35	4.38	6.39	0.04	15
N60DAF 3LN-1	1	428	788.61	783.65-792.81	121.32	180.45	179.75-182.15	5.25	6.67	-0.03	84
N60DAF 3LN-2	3	268	585.75	580.05-588.22	90.12	158.9	158.75-159.25	4.24	5.31	-0.03	9
N60DAF 3LN-3	6	56	102.67	99.35-106.75	15.80	13.80	12.95-19.85	6.31	8.13	0.03	193
N60DAF 3HN-1	4	377	755.43	750.04-757.16	116.22	178.65	177.95-178.65	5.69	9.09	0.04	33
R20452LN-1	2	72	223.51	218.67-226.61	34.39	9.45	9.35-9.55	6.92	8.64	-2.01	22
R20452LN-2	2	356	732.53	729.32-734.26	112.7	178.00	177.85-179.05	9.36	12.09	2.37	47
R20452LN-3	6	156	230.16	226.61-233.53	35.41	94.75	94.15-95.05	4.94	6.00	-1.65	38
R20452LN-4	7	24	92.63	86.68-96.24	14.25	4.20	4.05-4.55	4.56	5.50	-1.54	18
R20452LN-5	10	282	604.48	593.71-608.11	93.00	145.05	144.85-145.05	4.64	5.61	-1.70	14
R20452HN-1	2	269	645.02	642.16-647.24	99.23	97.00	72.00-106.65	8.69	12.7	1.77	597
R20452HN-2	10	199	373.56	368.49-380	57.47	132.05	130.95-132.35	4.86	6.68	1.24	39
R20453LN-1	1	109	317.54	309.21-325.59	48.85	17.85	17.65-19.00	4.44	7.57	-0.60	58
R20453HN-1	2	315	697.04	692.01-699.14	107.24	147.25	145.05-147.95	4.16	5.78	0.73	82
R20453HN-2	3	266	577.38	570.89-582.09	88.83	158.65	157.50-158.75	4.76	6.66	-0.81	32
R20453HN-3	4	375	756.83	749.04-762.24	116.44	177.95	177.60-178.65	4.62	6.44	-0.78	49
R20453HN-4	5	483	835.58	831.47-843.67	128.55	196.75	196.65-197.25	5.28	7.94	-0.91	18
R20601LN-1	1	140	386.35	382.96-388.87	59.44	24.75	24.65-24.85	11.22	17.47	1.42	12
R20601LN-2	2	333	708.88	706.56-712.47	109.06	161.25	160.60-165.75	4.70	6.65	0.90	124
R20601LN-3	3	57	217.81	214.7-227.58	33.51	7.15	6.85-7.25	8.34	13.06	-1.23	54

Table A5.8 continued.

QTL name	Chr[a]	Marker[b]	G Pos (cM)[c]	G Interval(cM)[d]	Adj (cM)[e]	P pos (Mb)[f]	P Interval (Mb)[g]	LOD	R^2 (%)	Add[h]	#Genes[i]
R20601LN-4	4	68	204.85	201.86-209.76	31.52	9.75	9.65-9.95	6.09	8.79	-1.06	23
R20601LN-5	4	371	735.31	732.27-738.22	113.12	177.25	176.70-177.35	5.95	8.58	1.03	18
R20601LN-6	10	48	157.76	154.2-161.43	24.27	6.05	5.85-6.15	5.22	7.45	-0.91	10
R20601HN-1	3	38	152.11	144.44-161.29	23.40	4.85	4.75-5.25	5.02	7.90	1.20	15
R20601HN-2	3	440	880.83	875.79-892.59	135.51	196.45	195.95-198.05	4.32	6.73	-1.09	96
R20603LN-1	3	61	225.72	218.71-229.92	34.73	7.95	7.25-8.15	4.15	5.40	0.67	30
R20603LN-2	6	111	157.69	155.61-163.73	24.26	84.55	84.45-85.15	6.97	9.52	-0.89	21
R20603LN-3	7	190	371.60	370.37-374.3	57.17	120.90	120.50-121.65	4.14	5.77	0.75	35
R20603HN-1	2	384	761.77	758.2-770.41	117.20	185.95	185.85-186.15	4.43	6.27	0.91	18
R20603HN-2	4	376	753.50	747.99-757.16	115.92	178.35	177.95-178.65	6.13	9.04	-1.03	33
R20603HN-3	6	152	224.24	215.67-230.91	34.50	93.55	93.40-94.60	4.75	6.75	-0.92	48
R20603HN-4	6	434	935.29	920.02-955.68	143.89	165.55	164.95-166.05	5.04	7.54	-0.98	71

[a] Chromosome number, [b] Marker localized at LOD peak, [c] Genetic position of SNP in cM, [d] 1-LOD interval in cM, [e] Adjusted genetic position, [f] Physical position in Mb, [g] 1-LOD Physical interval, [h] Additive effect of respective QTL (a positive-signed effect represents an increasing allele from B73, while a negative-signed allele denotes an increasing allele from Mo17), [i] Number of genes annotated underlying 1-LOD QTL CI. QTL names correspond to trait key name followed by experiment, N treatment and last number constitutes the QTL number for each respective trait.

Table A5.9. Multiple QTL models per trait in each experiment by treatment combination in the IBMSyn10-DH population of maize.

Experiment	Treatment	Trait	# QTL in model[a]	Model R^2 (%)[b]	R^2 epistasis (%)[c]
1	LN	Yield	2	13.22	
1	LN	EHT	1	4.93	
1	LN	PHT	2	15.17	
1	LN	GDD	3	9.80	
1	LN	N20DAF	1	5.85	
1	LN	N60DAF	2	12.71	
1	LN	R2060	6	34.69	
1	HN	EHT	2	15.08	
1	HN	PHT	2	13.08	
1	HN	GDD	2	6.59	
1	HN	N20DAF	1	8.78	
1	HN	N60DAF	4	22.39	
1	HN	R2060	2	11.42	
2	LN	Yield	3	19.28	
2	LN	EHT	4	21.93	
2	LN	PHT	2	8.97	
2	LN	GDD	2	9.40	
2	LN	N20DAF	3	22.46	
2	LN	N45DAF	6	42.57	
2	LN	R2045	5	35.35	3.10
2	HN	Yield	5	36.96	
2	HN	EHT	5	32.32	0.60
2	HN	PHT	3	17.23	
2	HN	GDD	2	16.98	
2	HN	N20DAF	2	9.60	
2	HN	N45DAF	4	33.41	3.30
2	HN	R2045	2	20.88	
3	LN	Yield	2	12.01	
3	LN	GDD	4	33.89	
3	LN	N20DAF	4	24.23	
3	LN	N45DAF	2	16.76	
3	LN	N60DAF	3	23.38	5.20
3	LN	R2060	3	17.92	
3	LN	R2045	1	2.10	
3	HN	Yield	5	39.33	
3	HN	GDD	2	16.73	
3	HN	N20DAF	2	9.81	
3	HN	N45DAF	2	11.08	

Table A5.9 continued.

Experiment	Treatment	Trait	# QTL in model[a]	Model R^2 (%)[b]	R^2 epistasis (%)[c]
3	HN	N60DAF	1	7.10	
3	HN	R2060	4	31.14	
3	HN	R2045	4	28.42	

[a] Number of significant QTL fitted in MIM model, [b] Total R^2 obtained by fitting significant QTL simultaneously in a MIM model, [c] R^2 explained by epistasis solely.

Table A5.10. Candidate genes involved with N-metabolism underlying identified QTL genomic regions in each location by treatment combination in the IBMSyn10-DH population of maize.

Maize GDB ID	Corresponding gene annotation	Chr[a]	Start[b]	End[c]	QTL name
GRMZM2G004079	PEPC-related kinase 2	1	201226717	201229957	EHT2LN1-1
GRMZM2G010920	MYB-like HTH transcriptional regulator family protein	5	77781452	77784526	Yield3LN-1
GRMZM2G017170	Transmembrane amino acid transporter family protein	1	103766057	103773502	Yield3HN-1
GRMZM2G019742	Senescence-associated gene 12	2	103783794	103784578	R20452HN-1
GRMZM2G025078	Nitrogen regulatory PII-like, alpha/beta	1	103740306	103741070	PHT1HN-1, R20453LN-1
GRMZM2G034302	Sucrose transporter 2	1	15069084	15074473	GDD2LN-1
GRMZM2G046002	PEPC family protein	8	165965364	165968685	N60DAF1HN-3
GRMZM2G047404	Glucose-6-phosphate/phosphate translocator-related	2	173925593	173929511	N45DAF2LN-1
GRMZM2G050481	Alanine:glyoxylate aminotransferase	1	185331193	185332687	N20DAF2LN-1
GRMZM2G057724	Aspartic proteinase A1	7	157659513	157665812	Yield2HN-4
GRMZM2G065757	Aspartic proteinase A1	6	165214021	165220621	R20603HN-4
GRMZM2G073219	Phosphoenolpyruvate carboxykinase 1	4	209044351	209046566	GDD1HN-1
GRMZM2G076526	ABC-2 type transporter family protein	9	92255434	92263891	PHT1HN-2
GRMZM2G077069	Phloem protein 2-A13	10	136725465	136728946	Yield1LN-2
GRMZM2G079381	Nitrite reductase 1	4	178613063	178616618	R20603HN-2, R20453HN-3, N60DAF3HN-1
GRMZM2G085210	Major facilitator superfamily protein	1	177685320	177689170	N20DAF2HN-1, N20DAF2LN-1
GRMZM2G088018	Alanine aminotransferase 2	5	6455410	6456081	GDD1LN-2
GRMZM2G088028	Alanine aminotransferase 2	5	7956058	7958178	GDD1LN-2
GRMZM2G088064	Alanine aminotransferase 2	5	178613063	178616618	GDD1LN-2
GRMZM2G104546	Aspartate kinase-homoserine dehydrogenase ii	2	173487079	173506976	N45DAF2LN-1
GRMZM2G106213	ADP glucose pyrophosphorylase 1	2	174024306	174034507	N45DAF2LN-1
GRMZM2G101125	Transmembrane amino acid transporter family protein	4	175243397	175248435	GDD3LN-1
GRMZM2G116478	Transmembrane amino acid transporter family protein	9	111116518	111117272	Yield3HN-5
GRMZM2G119248	Asparagine synthase family protein	1	102171297	102171994	Yield3HN-1
GRMZM2G119249	Shikimate kinase like 1	1	165655084	165656780	Yield3HN-1
GRMZM2G119300	Glucose-1-phosphate adenylyltransferase family protein	3	179779657	179782313	N20DAF3LN-3

Table A5.10 continued.

Maize GDB ID	Corresponding gene annotation	Chr[a]	Start[b]	End[c]	QTL name
GRMZM2G119511	Alanine:glyoxylate aminotransferase	1	185769464	185771702	N20DAF2HN-1
GRMZM2G124353	Alanine:glyoxylate aminotransferase 2	1	15352527	15355919	GDD2LN-1
GRMZM2G137421	Nitrate transporter 1:2	6	156233140	156235981	N20DAF3LN-4
GRMZM2G137868	Alanine:glyoxylate aminotransferase	1	184941358	184944047	N20DAF2HN-1, N20DAF2LN-1
GRMZM2G141480	Phloem protein 2-A11	3	178325842	178335080	N20DAF2LN-2
GRMZM2G156486	Nitrilase/cyanide hydratase family protein	1	15989647	15994433	GDD2LN-1, N45DAF3HN-1, YIELD1LN-1
GRMZM2G164714	Phosphoenolpyruvate carboxylase family protein	10	132029047	132032448	R20452HN-2
GRMZM2G164743	Major facilitator superfamily protein	10	18190325	18199212	R20452HN-2
GRMZM2G173016	Nucleotide-sugar transporter family protein	2	103657680	103658845	N45DAF2HN-1, R20452HN-1
GRMZM2G327050	Nitrate transporter 1:2	6	103783794	103800000	N20DAF3LN-4
GRMZM2G335218	Ammonium transporter 2	8	165655084	165656780	N60DAF1HN-3, N45DAF3LN-2
GRMZM2G347457	Nitrate transporter 1:2	6	103740306	103741070	N20DAF3LN-4
GRMZM2G355906	Glutamate decarboxylase 2	2	45349199	45380559	GDD3HN-1
GRMZM2G359559	Alanine:glyoxylate aminotransferase	1	187035638	187037628	N20DAF2HN-1, N20DAF2LN-1
GRMZM2G385263	Nitrate reductase 2	9	111116518	111117272	Yield3HN-5
GRMZM2G392988	Sucrose synthase 3	8	124361368	124364917	EHT2LN-3
GRMZM2G403620	MYB-like HTH transcriptional regulator family protein	1	154968134	154972600	Yield3HN-1
GRMZM2G410704	Sucrose synthase 6	1	17721176	17724401	R20453LN-1
GRMZM2G428027	Nitrite reductase 1	4	177685320	177689170	R20453HN-3
GRMZM2G439542	PEP/phosphate translocator 2	2	187826182	187826819	EHT1HN-1
GRMZM2G701289	Ammonium transporter 1;1	3	6455410	5456081	R20601LN-3
GRMZM5G803404	ABC transporter family protein	2	101898444	101899537	N45DAF2HN-1, R20452HN-1
GRMZM5G821252	Nitrate transporter 1:2	9	83733269	83733894	PHT1HN-2
GRMZM5G869453	Pyruvate kinase family protein	3	15989647	15994433	R20603LN-1

[a] Chromosome, [b,c] start and end location in bp

Table A5.11. Candidate genes associated with phosphate transporters and cellulose synthase underlying QTL genomic regions identified in the analysis of each experiment by treatment combination in the IBMSyn10-DH population of maize.

Maize GDB ID	Corresponding gene annotation	Chr[a]	Start[b]	End[c]	QTL name
GRMZM2G009779	Phosphate transporter 1;7	2	99885266	99886925	R20452HN-1, N45DAF2HN-1
GRMZM2G009800	Phosphate transporter 1;7	2	99926376	99930029	R20452HN-1, N45DAF2HN-1
GRMZM2G018241	Cellulose synthase family protein	2	161757546	161763704	R20601LN-2
GRMZM2G024182	Cellulose synthase 1	9	87864926	87867458	PHT1HN-2
GRMZM2G027794	Cellulose-synthase-like C12	8	172525179	172529685	Yield2LN-3
GRMZM2G028353	Cellulose synthase 6	2	170393027	170398878	N45DAF2LN-1
GRMZM2G045473	Phosphate transporter 1;5	2	99381968	99383837	N45DAF2HN-1, R20452HN-1
GRMZM2G060630	Phosphate transporter 3;1	4	178520983	178525630	R20603HN-2, R20453HN-3, N60DAF3HN-1
GRMZM2G082580	Cellulose synthase 6	2	171408315	171412367	N45DAF2LN-1
GRMZM2G090126	Phosphate transporter 3;1	1	203917963	203920413	EHT2LN-1
GRMZM2G092186	Cellulose synthase 1	1	151430784	151432436	Yield3HN-1
GRMZM2G112377	Phosphate transporter 1;7	1	202585823	202587997	EHT2LN-1
GRMZM2G124089	Cellulose-synthase-like C12	8	172598290	172599189	Yield2LN-3
GRMZM2G132169	Laccase 12	3	183701021	183703536	N45DAF2LN-3
GRMZM2G150404	Cellulose synthase family protein	2	161768961	161771897	R20601LN-2
GRMZM2G170208	Phosphate transporter 1;5	2	99805638	99807544	N45DAF2HN-1, R20452HN-1
GRMZM2G173710	HPT phosphotransmitter 4	8	124015413	124020226	EHT2LN-3
GRMZM2G349834	Cellulose synthase 1	6	102842181	102845471	N20DAF3LN-4
GRMZM2G389588	Cellulose synthase 1	5	36617909	36620570	GDD1LN-2
GRMZM2G410085	Phosphate transporter traffic facilitator1	6	18041267	18046263	N60DAF3LN-3
GRMZM2G451646	Cellulose synthase 1	6	98887115	98889430	N20DAF3LN-4
GRMZM2G701031	Phosphate transporter 4;5	6	99037001	99037266	N20DAF3LN-4
GRMZM5G856598	Phosphate transporter 4;3	3	178916302	178919735	N20DAF2LN-2, N20DAF3HN-2

[a] Chromosome, [b,c] start and end location in bp

CHAPTER 6: GENERAL CONCLUSIONS

A linkage mapping analysis was conducted in an integrated manner aiming to identify genomic regions associated with N-metabolism in a maize TC population derived from B73 and Mo17. First, enzymes and metabolites were analyzed from root and leaf tissues at a vegetative stage (V4) from plants grown in hydroponic conditions. Subsequently, agronomic traits were measured in same TC genotypes, grown in the field under LN and HN conditions. This investigation provided insightful and valuable information in order to partially elucidate the genetic control of N-metabolism in a maize TC population.

A methodical approach for the determination of real outliers in the different datasets was implemented. Even though it is well established that high quality data are essential for the success of quantitative trait loci (QTL) mapping experiments (Bernardo, 2010), the management of raw data, including the determination of outliers with a statistical basis, has received considerably less emphasis than the subsequent genetic analysis. Hence, several spurious associations between genetic regions and variation in phenotype performance may have arisen as a result of the misinterpretation in the identification of overly influential values. On the other hand, real associations may have been missed due to the omission of valid observations based merely on totally subjective rationale. In order to optimize the use of the available information generated in the experiments, an approach for determination of outliers with a statistical basis was implemented. The approach is divided into five main steps consisting of visual inspection of the data, studying relationships between multiple response variables (e.g. enzyme activity), fitting statistical models, filtering genotypes (subsetting the data) and, finally, filtering influential measurements based on a Jackknife approach. As a result of implementing the approach described above, improvements in the

log-likelihood values, on the order of 200 units in magnitude, were achieved by removing just a few (three to eight) genotypes. It is important, while following the steps of the described approach, to find any problems and address them in order to obtain the most reliable results. Steps similar to the ones used here to survey the enzyme activity data may be applied in many other circumstances, to improve the accuracy of results. The R code used in this analysis is provided in the supplementary information and can be readily adapted for any similar initiatives.

From the linkage mapping analysis based on leaf tissue, harvested from maize hybrids grown in hydroponics, a total of 44 QTL were identified. Epistasis between QTL was not significant for most of the traits. Nevertheless, significant epistasis was determined for two QTL model explaining 2.5-5% of the genetic variance. The QTL models for different traits accounted from 7 to 31% of the genetic variance. Furthermore, 12 coding regions underlying 1-LOD QTL confidence intervals (CI) were identified as promising gene candidates associated with N metabolism for further validation studies. Moreover, all QTL identified were in *trans* compared to the genomic position of the correspondent structural genes.

In the similar analysis of enzymes and metabolites on root tissues, 22 QTL were identified. QTL models for explained 8-43% of the genetic variance and no significant epistasis was detected between QTL. A total of ten candidate genes were proposed underlying 1-LOD QTL CI regions. Similar to the findings with leaf tissue analysis, all candidate genes were located in *trans*, unlinked or even in different chromosome, to the known genomic positions of the correspondent structural genes.

In the analysis of agronomic and physiological traits from TC maize grown in the field under LN and HN conditions, 45 QTL were detected in a combined analysis (across three

experiments) while 117 QTL were identified in the split analysis (at each experiment by N treatment combination). In the case of the combined analysis, multiple QTL model explained 5.7-33.4% of the phenotypic variance and epistasis was significant for only one trait. Furthermore, 22 candidate genes underlying QTL regions were proposed for further analysis. With regard to the split analysis, QTL models explained from 2 to nearly 43% of the variance, and 50 candidate genes associated with N metabolism, underlying 1-LOD QTL regions, were targeted for further analysis. In addition, 23 candidate genes within identified QTL regions were also pinpointed for future investigation; all of them were associated with phosphate transporter and cellulose synthase.

Numerous hotspot QTL regions were identified in the maize genome across the hydroponics and field experiments. Several QTL did co-locate and various were determined in close proximity to each other (Fig 6.1). At least one rich QTL region, presenting three or more overlapping CI for QTL, was determined at each of the ten chromosomes.

The results of this integrated investigation provide an insight in order to achieve a more holistic comprehension of N metabolism in maize TC. Several genomic regions responsible of the variance in the performance of certain N-metabolism related traits and candidate genes within QTL regions have been targeted for further investigations. A better comprehension of the genetics underlying N-metabolism in maize would be necessary in order to: develop ideotype maize hybrids with an optimal performance of certain key enzymes and transporters related to N-metabolism, promote a more sustainable agriculture with a decrease in N fertilizer inputs while maintaining yields, leading to an overall increase in profits while reducing environmental contamination.

References

Bernardo, R. 2010. Breeding for quantitative traits in plants. Stemma Press, Woodbury, MN.
Voorrips, R.E. 2002. MapChart: Software for the graphical presentation of linkage maps and QTLs. Journal of Heredity 93: 77-78

Figures

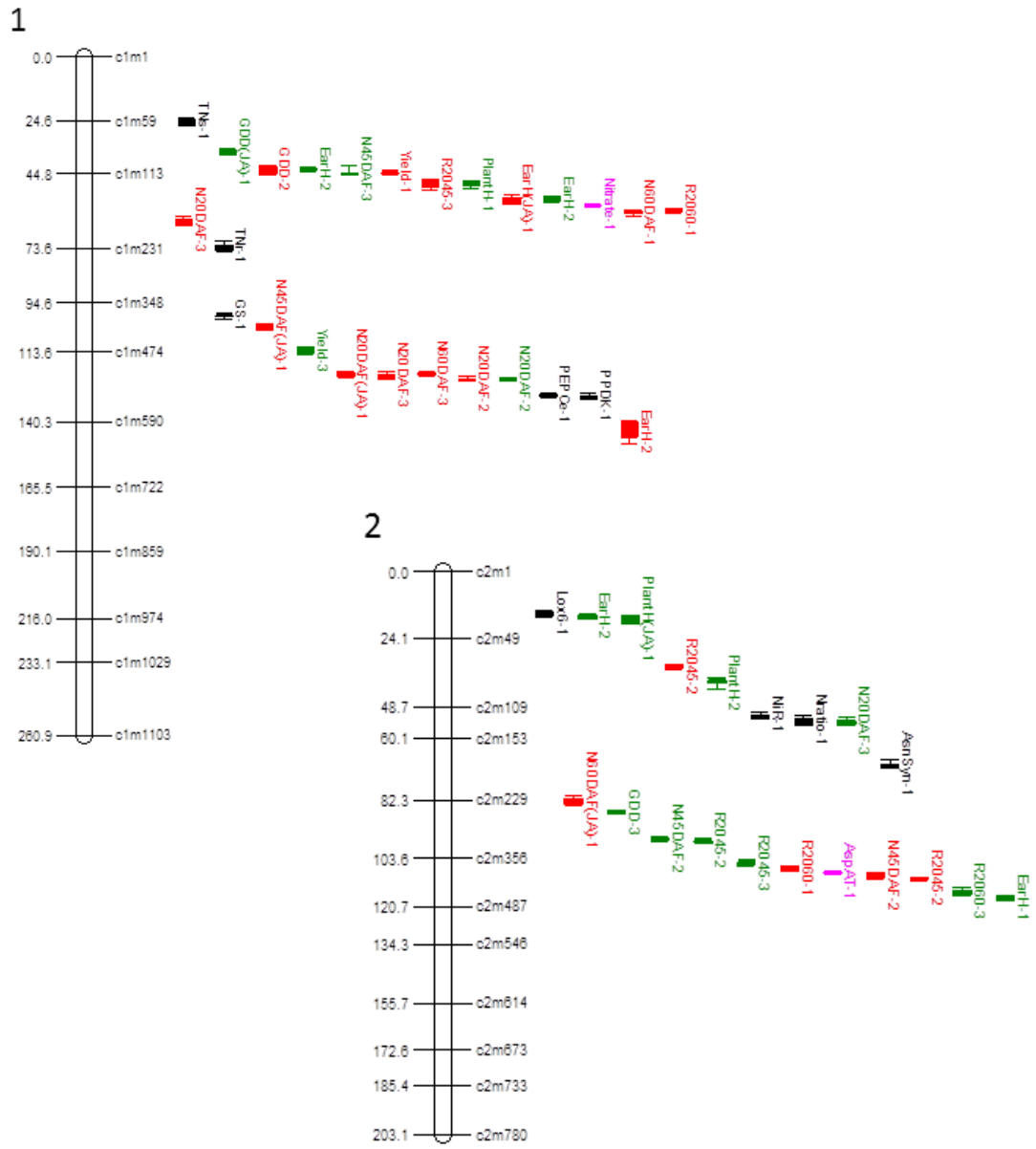

Figure 6.1. QTL identified across experiments in the maize IBMSyn10-DH TC population.

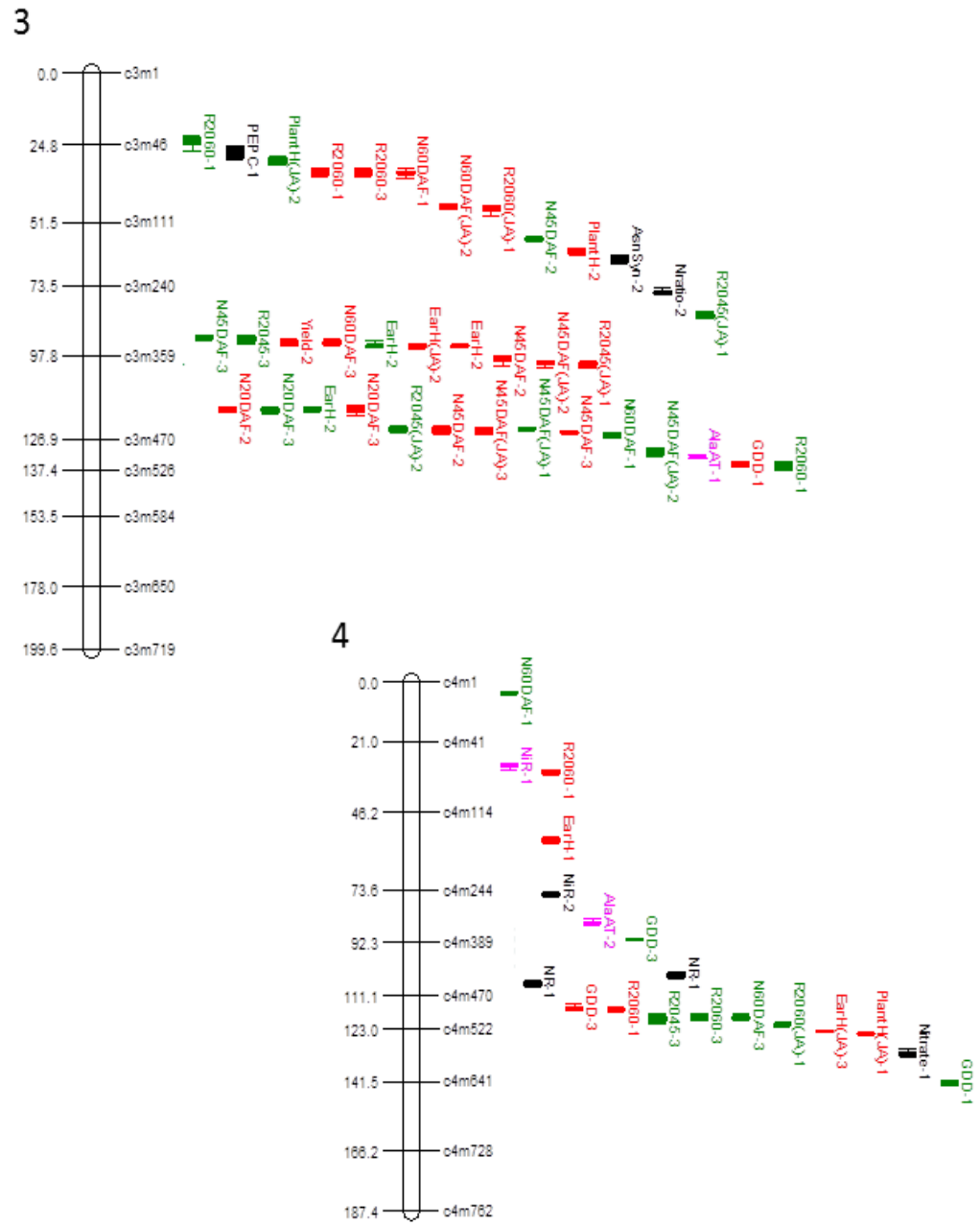

Figure 6.1 (continued).

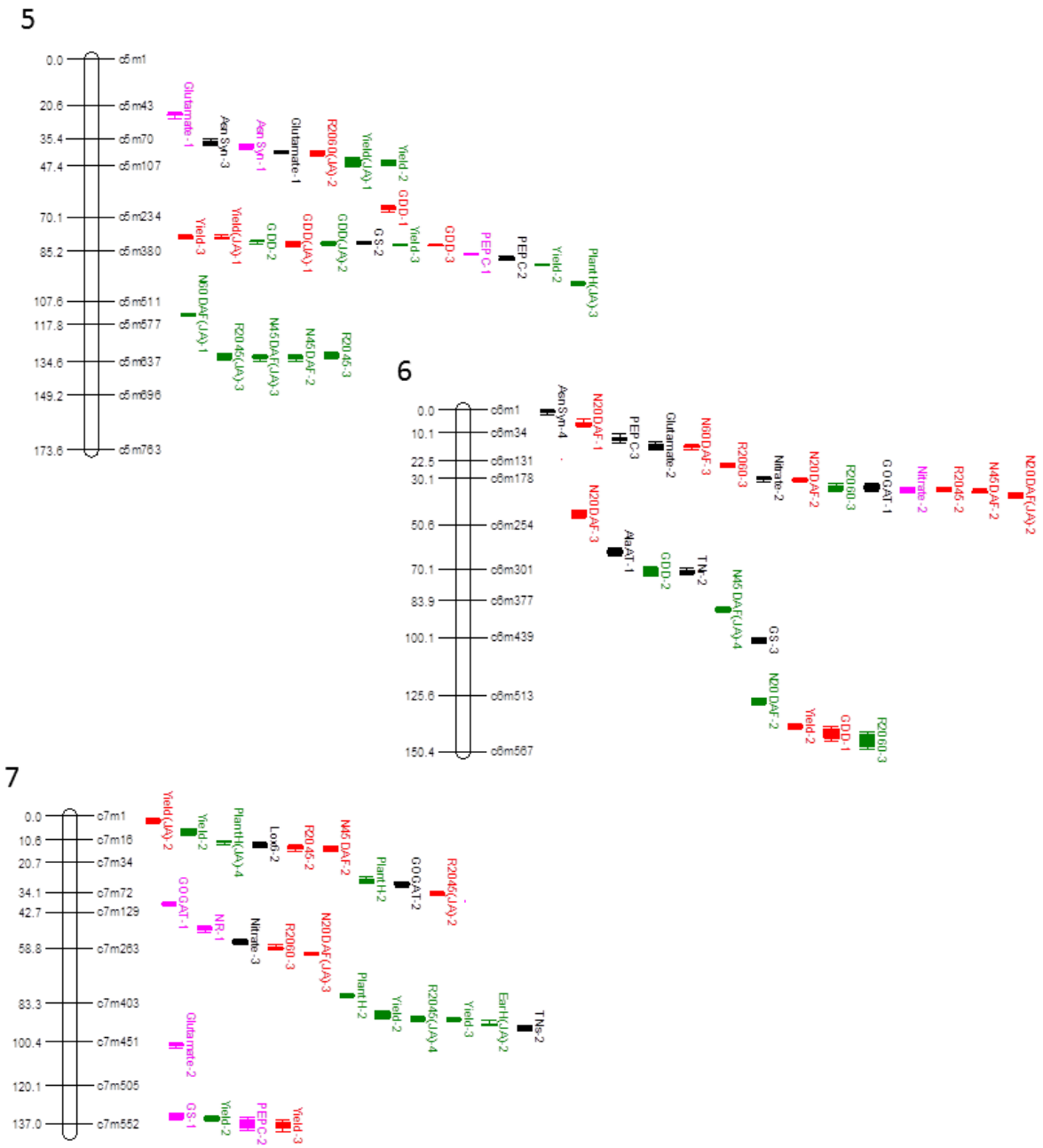

Figure 6.1 (continued).

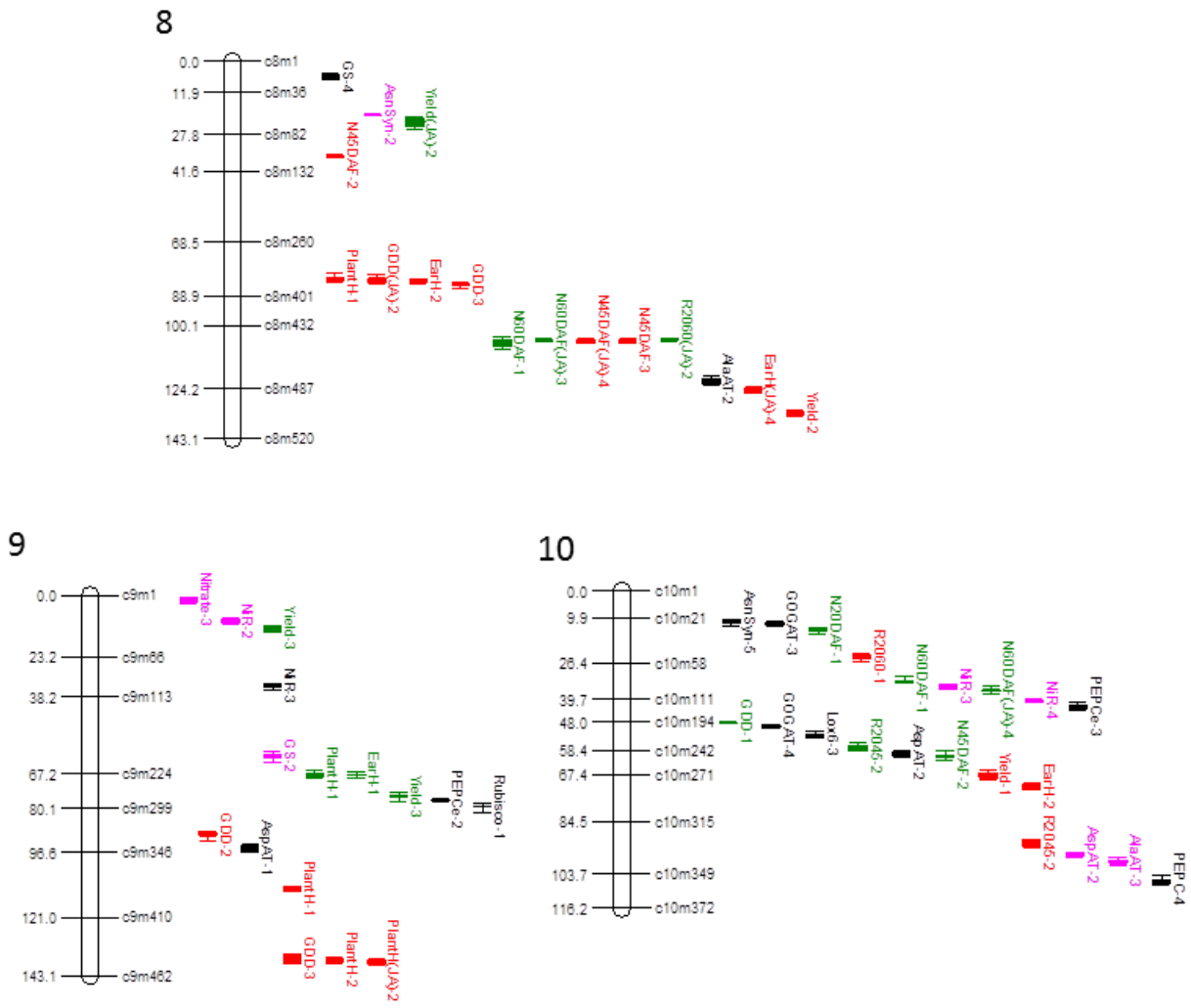

Figure 6.1 (continued).
QTL identified from hydroponic experiments are depicted in black (leaf tissue analysis) and pink (root tissue analysis), while QTL detected in field experiments are shown in red (low N) and green (high N). QTL name followed by (JA) meaning QTL identified in the combined experiments analysis (joint analysis). Figure created with MapChart 2.2 (Voorrips, 2002)

GENERAL ACKNOWLEDGMENTS

I would like to thank Dr. Michael Lee for giving the opportunity to study and learn plant breeding at Iowa State University. He was not only my major advisor but also someone I can trust and was always willing to provide any kind of help. I am grateful to all the committee members of my program of study and Dr. Kanwarpal Dhugga for their active participation throughout this research.

I want to thank to all the exceptional friends I made in Ames, including Joe and Adelaida Cortes, Andres Fuentes-Ramirez, Franco Matias-Ferreyra, Diego Ortiz, Pablo Pineyro, Pablo Barbieri and their respective families, for their support. I am also grateful to all the plant breeding and genetics graduate students, especially to Pedro Gonzalez and Vikas Belamkar. In addition, I want to acknowledge Guan Yi Lai for his remarkable work in the field and lab experiments. Furthermore, I must give thanks to my friend and former advisor Carlos Sala for the continuous encouragement into pursuing a higher education degree.

I am thankful to my parents Vicente and Sonia, and my sister Maria for their constant support. Finally, this achievement would not have been possible without the love of my wife Veronica and our adorable daughter Agustina. Thank you so much!