Theor Appl Genet (2006) 112: 1205–1220
DOI 10.1007/s00122-006-0222-5

ORIGINAL PAPER

Marie Coque · André Gallais

Genomic regions involved in response to grain yield selection at high and low nitrogen fertilization in maize

Received: 15 April 2005 / Accepted: 15 January 2006 / Published online: 22 March 2006
© Springer-Verlag 2006

Abstract In order to validate the role of genomic regions involved in nitrogen use efficiency and detected in a population of recombinant inbred lines (RIL), we have applied from the same population a recurrent selection for adaptation to low N-input (N0) and to high N-input (N1). Variation of allele frequency at neutral marker during the two cycles of recurrent selection may provide information about markers linked to QTLs. Significant temporal variation of allele frequency was investigated using the test of Waples, which tests the hypothesis of genetic drift versus selection. Most genomic regions (12/19) responding to selection were detected for selection at high N-input and only two were common to selection at high and low N-inputs. This was consistent with the greater grain yield response to selection observed for the population selected under high N-input compared with the population selected under low N-input, when they were evaluated at high N-fertilization. In contrast, when they were evaluated at low N-input both types of selection gave similar yield. As was expected, in the first cycle we observed selection of markers linked to grain yield QTLs. In the course of the second cycle three situations were observed: the confirmation of most regions already selected in C1 including all C1 regions overlapping with grain yield QTLs; the non-confirmation of some C1 regions (2/9); and the identification of new genomic zones (10/17). The detected marker–QTL associations revealed the consistency of the involvement of some traits, such as root architecture and glutamine synthetase activity, which would be of major importance for grain yield setting whatever the nitrogen fertilization.

Communicated by H. H. Geiger

M. Coque · A. Gallais (✉)
Station de Génétique Végétale du Moulon, INRA/UPS/CNRS/INAPG, 91190 Gif sur, Yvette, France
E-mail: gallais@moulon.inra.fr
Tel.: +33-1-69332369
Fax: +33-1-69332340

Introduction

Increase in nitrogen (N) fertilizer use has greatly contributed to the increase in crop yields. However, the success of N-fertilization has been associated with environmental hazards. Nowadays, farmers must optimize the use of N-fertilizer not only to limit ground water pollution by nitrates, but also to preserve their net income. Therefore, the challenge for breeders is the development of varieties with better nitrogen use efficiency (NUE) leading to high yields at lower N-input.

For maize, NUE has already been defined by Moll et al. (1987) as the grain yield per unit of nitrogen from soil and nitrogen fertilizer. Genetic variability for NUE has been shown in different studies and appears to be differently expressed according to the N-level (Beauchamp et al. 1976; Pollmer et al. 1979; Balko and Russel 1980; Reed et al. 1980; Russel 1984; Moll et al. 1987; Landbeck 1995; Bertin and Gallais 2000; Presterl et al. 2002). At high N-input, variation in N-uptake contributes more to variation in NUE than N-utilization efficiency (grain yield/N-uptake), whereas the reverse is observed at low N-input (Bertin and Gallais 2000). These results suggest that limiting steps in N-metabolism may be different under high and low N-levels. Detection of QTLs partly corroborates this conclusion (Agrama et al. 1999; Bertin and Gallais 2001). At low N-input, detected QTLs are more related to N-content and N-remobilization, whereas at high N-input they are more related to post-anthesis N-uptake. However, it appears that QTLs for yield and yield components detected at low N-input could be a subset of QTLs detected at high N-input (Bertin and Gallais 2001).

Study of coincidences between QTLs for agronomic and physiological traits and key genes of N-metabolism have allowed the characterization of QTLs and the identification of candidate genomic regions involved in the NUE variation (Hirel et al. 2001; Gallais and Hirel 2004). However, considering the high risk of false QTL detection, especially when a low population size is used

(Beavis 1994; Melchinger et al. 1998; Bernardo 2004; Melchinger et al. 2004), before investing in the validation of candidate genes and developing marker-assisted selection (MAS), it is necessary to be confident in the QTL detection. A possible way to validate QTLs is to develop near isogenic lines (NIL) (Paterson et al. 1990; Lin et al. 2000) for candidate regions. Another way consists in developing a recurrent phenotypic selection from a population in linkage disequilibrium, due only to physical linkage, and studying the change in neutral marker frequency, which may provide insight into the genomic regions submitted to the selection. Indeed, the phenomenon of selective sweep allows the assumption that selection should affect patterns of linked neutral variation around selected loci, to an extent determined by the local recombination rate and the strength of selection (Smith and Haigh 1974). From a segregating population, the first cycle of selection is equivalent to an experiment of unidirectional selective genotyping which has been used to detect or validate QTLs (Ayoub and Mather 2002; Zhang et al. 2003; Wingbermuehle et al. 2004). The difference with selective genotyping as proposed by Darvasi and Soller (1992) is that the study is not based on differences in the mean values of marker genotypes, but on the change in marker allele frequencies, as proposed by Stuber et al. (1980, 1982). However, the main problem to solve when studying more than one cycle of selection is to test whether the allelic variations are due to selection or genetic drift.

In the present paper, we reported results of a divergent recurrent selection for adaptation to low and high N-fertilization, based on a recombinant inbred line (RIL) population where QTLs for NUE have already been detected (Bertin and Gallais 2001). Our objectives were to identify chromosome regions responding to selection for adaptation to low and high N-levels and to study their possible coincidences with QTLs already detected for agronomic and physiological traits as with candidate genes.

Material and methods

Theoretical aspects

For a locus directly involved in the determination of genotypic value of the considered trait, assuming a sufficiently small effect of the locus, the change in frequency of A (Δp) due to one cycle of recurrent selection in a biallelic population can be written by adapting the formula given by Griffing (1961) or Falconer and Mackay (1996) to take into account the inbreeding of the population:

$$\Delta p \sim i(1+F)pq\frac{\alpha}{\sigma} = i(1+F)pqr_{\mathrm{P}}, \quad (1)$$

where p is the frequency of A before selection ($q=1-p$), σ the phenotypic variance, α the substitution effect, i the standardized selection intensity, F the coefficient of inbreeding ($F=1$ with RILs) and r_{P} the square root of the percentage of phenotypic variance explained by the locus. The formula is also applicable for a marker linked to a gene. In this case, and considering only the first cycle of selection, α has to be replaced by the substitution effect of the marker $\alpha^{*}=\alpha(1-2c)$, with c being the recombination rate between the marker and the QTL and r_{P} becoming the square root of the phenotypic variance explained by the marker. As we considered testcross value, which is an additive trait (Gallais 1991), in our study $\alpha=a_{\mathrm{T}}$ is half the difference between the two homozygotes. We have verified that assumptions about genetic effects are not too strong; even with values of r_{P}^2 around 20% the predictive value of formula 1 remains quite reliable (data not shown). Then formula 1 can be written at the first selection cycle as:

$$\Delta p = 0.5ir_{\mathrm{P}} = 0.5i\lambda hr_{\mathrm{G}}, \quad (2)$$

where $\lambda=1-2c$ is the linkage parameter introduced by Schnell (1961), h the square root of the heritability of the considered trait and r_{G} the square root of the percentage of genetic variance explained by the QTL. The advantage of this formula is to clearly show parameters affecting the change in marker frequency: selection intensity, distance between the marker and the QTL, genetic effect of the linked QTL and heritability of the trait.

Knowing the change in marker frequency, it is then possible to estimate the effect of the QTL linked to the marker: $r_{\mathrm{P}}^2=(\Delta p/0.5i)^2$. The sign of the effect will be given by the sign of Δp. Such an approach should lead to approximately the same results as those of QTL detection.

Selection procedure

To apply the recurrent selection, the same RIL population as in Bertin and Gallais (2001) study was used. It results from the cross between an early flint line (F2) from France and an iodent line (Io). A random sample of 99 out of 200 derived lines was used for QTL detection. This gives the population C0, from which two cycles (C1 and C2) of recurrent selection for combining ability with a tester were developed, using an unrelated inbred line tester (F252). For the evaluation of the C0 population, testcross progenies were studied at two N-levels: 175 kg N/ha applied at the time of sowing (N1) and no nitrogen fertilization (N0). After soil analysis, the nitrogen being supplied by the soil was estimated to be at least 50–60 kg N/ha. The experiment was conducted over two consecutive years, 1994 and 1995, in the same location (Plant Breeding Station of "Le Moulon" Gif/Yvette, to the south of Paris). The experimental design was a split-plot, with N-fertilization as first level and genotypes as second level, with three replications and two-row plots, 5 m long and 80 cm between rows. Grain yield was estimated as the mean of the 2 years of

experiment. We selected the best 18 lines in each N-condition of evaluation on the basis of their yield performance corrected for earliness. This leads to the C1N1 and the C1N0 populations. For each N-condition of selection (N0 and N1-selections), the selected lines were intercrossed for two generations. To favour random intercrossing, the first intercrossing was hand-developed by complete diallel crossing, whereas the second intercrossing was developed in isolation, with each F1 cross from the first intercrossing taken as female and the mixture of all F1 crosses as males. Each female was harvested separately and, after shelling, an equal quantity of their seeds was taken to form the next generation. Each new population was then selfed, with 200 selfed ears per population. These families were crossed with the tester line F252 and testcross progenies were evaluated in 2001 and 2002, in two locations per year. In 2001 the experiment was at "Le Moulon" and Mons en Chaussée (northern France), and in 2002 it was conducted at "Le Moulon" and Rennes (western France). The experimental design was the same as for the C0 experiment. The population developed for N1-selection was evaluated at N1-level: about 180 kg/ha; and the population developed for N0-selection was evaluated without N-fertilizer. A second cycle of recurrent selection was then initiated on the basis of the average yield for the 2 years of experiments. For the selection, grain yield (GY) was corrected for grain moisture (MOI) which is directly related to earliness, according to the equation $GY_{cor} = GY_{obs} - (b\ MOI_{obs})$ (b being the regression coefficient of grain yield on grain moisture, level $b = 1.7$ for N1 and $b = 1.2$ for N0). The best 20 lines were selected for N0 and N1-selections. This led to the populations C2N1 and C2N0. In order to save time, and partly because this was not required for the study of marker frequency, selected lines were not intercrossed and evaluated for yield production. Furthermore, the expected testcross value of the population from intercrossing is equal to the expected value of selected plants (Gallais 1991).

Agronomic evaluation of populations selected in C1

After crossing with the tester F252, populations resulting in the first cycle of selection, i.e. C1N0 and C1N1 populations, were evaluated for grain yield and grain moisture. Two N-levels (with and without N-fertilizer) were studied, in 2001 at two locations and in 2002 and 2003 at four locations, leading to 10 environmental conditions with three replications always with two-row plots (5 m long, 80 cm between rows).

Genetic map and markers

We used the genetic map based on RFLP (restriction fragment length polymorphism) markers of the RIL population resulting from the cross between F2 and Io published by Causse et al. (1996). In addition, in order to easily and rapidly study the change of allelic frequencies at neutral markers, a set of 143 microsatellite markers (SSR) was genotyped and assigned on this published map using the software package ActionMap (Albini et al. 2003). We then used Mapmaker/exp (version 3.0b) to map the 143 SSR according to their assignment information. This new map, combining RFLP and SSR loci, contained 385 markers with an average distance of 7 cM (Fig. 1).

Marker analysis

The allele frequencies of the 143 SSR markers, previously mapped, were studied for each population. This implied the study of marker allele frequencies for the 99 RILs in C0 population, the 19 selected plants of C1N1 and C1N0 populations and the 20 selected plants of C2N1 and C2N0 populations. All marker frequencies were expressed in reference to the parent line *F2*.

Test of global differentiation between populations

To evaluate the global differentiation between all pairwise populations, we have used the multilocus estimation of Wright's parameter F_{ST} which provides an indication of the distance between two or more populations (Weir and Cockerham 1984; Michalakis and Excoffier 1996). To test the significance, the distribution of pairwise F_{ST} values under the hypothesis of no difference between the populations is obtained by permuting haplotypes between populations using the software package Arlequin (http://www.anthropologie.unige.ch/arlequin/). Then, the P value of the test is the proportion of permutations leading to an F_{ST} value larger or equal to the observed one. An exact test of population differentiation was also used by testing the hypothesis of a random distribution of k different genotypes among r populations as described by Raymond and Rousset (1995). This test is analogous to Fisher exact test on a 2×2 contingency table extended to an $r \times k$ contingency table performed thanks to Arlequin software. All potential states of the contingency table are explored with a Markov chain algorithm. The P value of the test is then taken as the proportion of the explored tables having a probability smaller or equal to the observed contingency table.

Test of selection effect versus genetic drift

In what follows, for he sake of simplicity, we have used the expression "selected markers" to designate markers which were detected as associated with selected loci, although the selection is not at the marker level but on the QTL. We have also used the expression "*Io* allele" or "*F2* allele" to express that the allele under selection at the

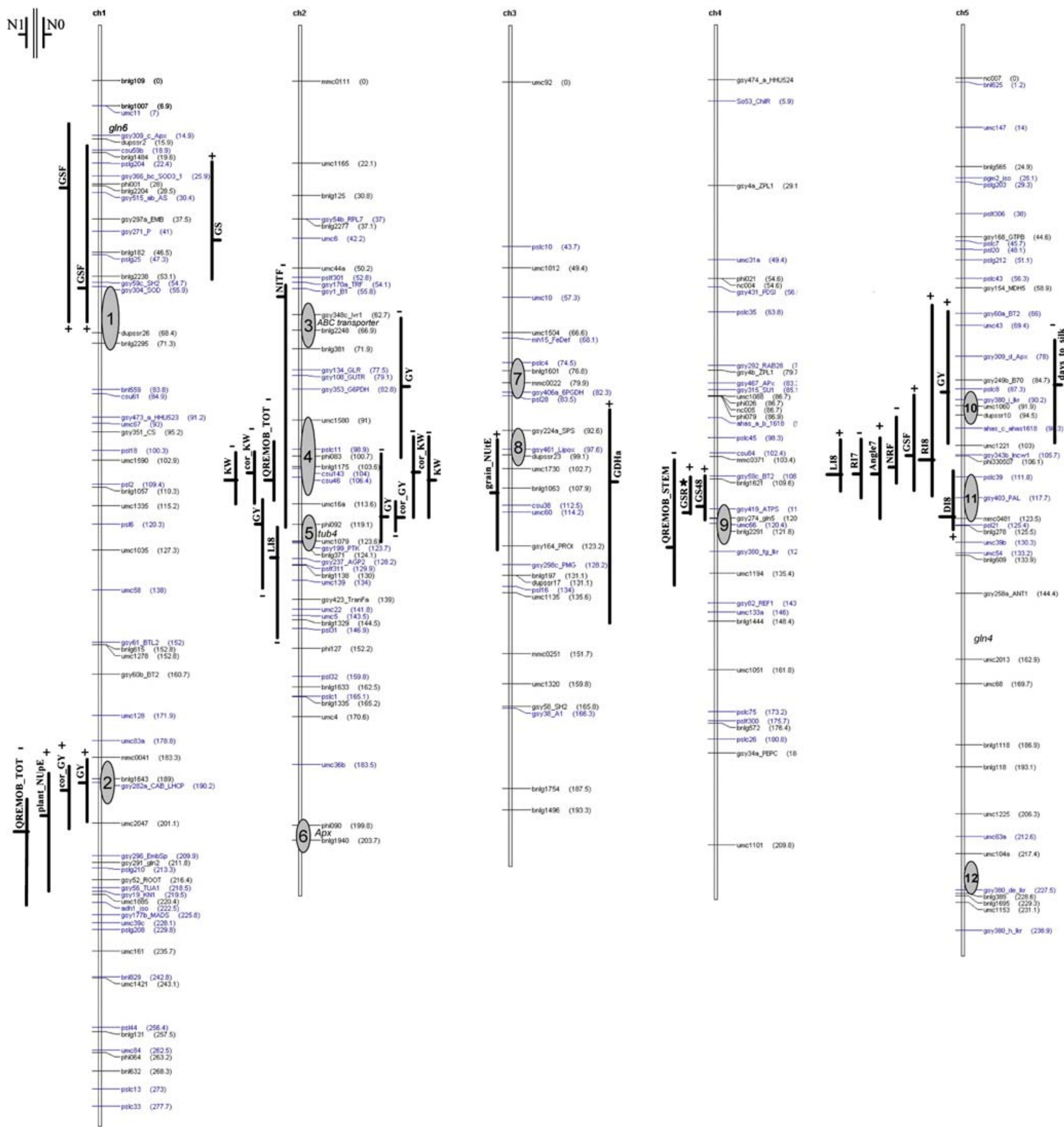

Fig. 1 Genetic map of population *F2×Io* of maize. Chromosomes 1 to 10. In *brackets near locus name on the right of the chromosome*: absolute distance in cM. QTLs are represented on the *left or on the right* of the chromosome according to the N-condition of detection: the *vertical arrow* defines the confidence interval; the *horizontal arrow* is proportional to the phenotypic variance explained by the QTL which is on the most probable position of the QTL. The *plus and minus signs below or above QTL arrow* indicate which parent brings the favourable allele: *plus* for Io and *minus* for F2. Each region targeted by selection is identified by *ellipses*. All the QTLs have been detected based on the means of 2 years of experiment except the QTL denoted with a *star*. Interesting genes not mapped in our population but projected on our map using the software Biomercator were noted in *italic* near the chromosome. Traits are defined and referenced in Table 1

''selected marker'' comes from either the parental line Io or F2. To test whether a marker was selected, we have used a single locus test of temporal variation in allele frequency derived from the test proposed by Waples (1989). It tests whether the observed variation in allele frequency between two populations can be explained by

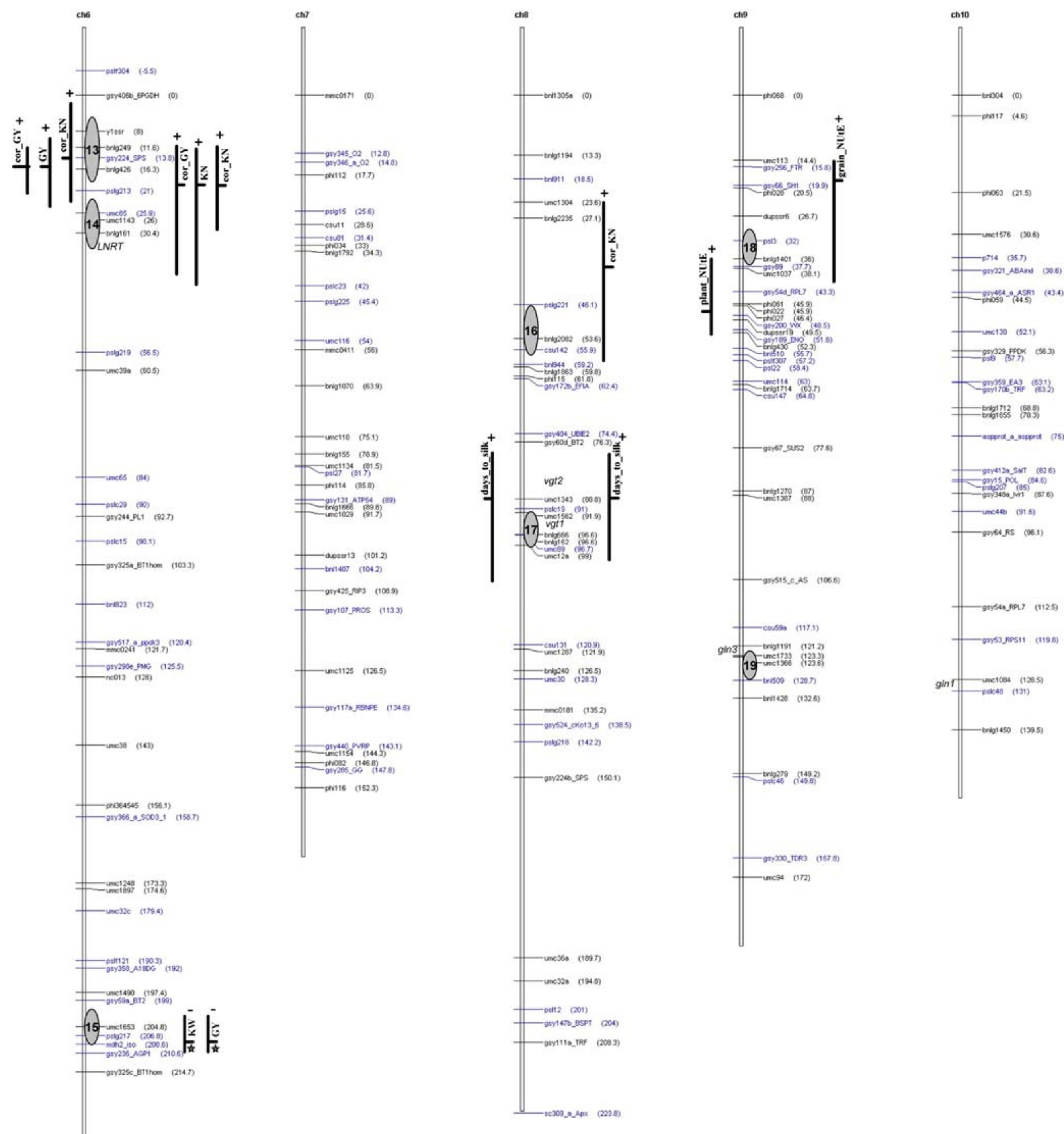

Fig. 1 (Contd.)

genetic drift (null hypothesis) or selection, taking into account population histories. Denoting by p_x and p_y the frequency of the *F2* allele in the two populations to be compared, this consists in a *t* test such that: $t = (p_x - p_y)/(\sqrt{V(p_x - p_y)})$, where $V(p_x{-}p_y)$ is the variance of the difference between the frequencies. The main problem is in the derivation of $V(p_x{-}p_y)$ taking into account the genetic drift at each generation (Appendix). This test was performed, for temporal variation, at each N-level between C0 and C1, C0 and C2 and C1 and C2 populations. But it was also performed between N-level populations: C1N0 versus C1N1 and C2N0 versus C2N1 in order to identify loci selected in a specific N-condition. To have an acceptable global α-risk, taking into account the great number of tests (equal to the number of markers), it was necessary to have a more restrictive α-threshold than for an individual test. This is why we have considered four α-levels: 5, 3, 1 and 0.5%.

QTL data and study of QTL colocalization

In order to characterize genomic regions responding to selection, we studied the coincidences between these regions and QTLs detected in the RIL population F2×Io. The QTLs published by Bertin and Gallais (2001) for grain yield, kernel weight, kernel number and flowering time traits were detected on the RFLP genetic map (Causse et al. 1996). Therefore, to have more direct correspondence with selection results, we redetected QTLs with the new genetic map for these sets of traits and for grain yield, kernel weight and kernel number corrected for earliness (including both silking date and grain moisture). Data were based on the averages of the experiments of 2 years in N0 and N1-conditions (Tables 3, 4). This detection was done by simple interval mapping method using the software Plab-QTL (Utz and Melchinger 1995). In addition, on population C0, in order to evaluate the effect of markers responding to selection, an analysis of variance (ANOVA) was done for each marker from the same agronomic data (Table 6). Indeed, as shown by formula 1, change in marker frequency due to selection depends directly on the percentage of phenotypic variance explained by the marker. Most of the other QTLs, detected in our population, reported in this study have already been published. QTLs for "field" traits related to NUE were identified by Bertin and Gallais (2001) and Gallais and Hirel (2004). QTLs for physiological traits at young stages were detected by Hirel et al. (2001) and at post-silking stage by Dubois et al. (2003). We also considered QTLs for root architecture detected by Guingo et al. (1998) and QTLs for germination detected by Limami et al. (2002). As these QTL detections were based on the RFLP genetic map (Causse et al. 1996) to study the coincidence between QTLs and SSR markers responding to selection, these sets of QTLs were projected on the newly built genetic map using the software package BioMercator (Arcade et al. 2004) (Fig. 1). Traits abbreviations are defined and referenced in Table 1.

Table 1 Definition of trait abbreviations and their published references

Traits	Definition	Reference
GS	Glutamine synthetase activity of leaf below the ear 15 days after flowering in field	Dubois et al. 2003
GSF	Glutamine synthetase activity of young leaves in hydroponical conditions	Hirel et al. 2001
NRF	Nitrate reductase activity of young leaves in hydroponical conditions	Hirel et al. 2001
GSR	Roots glutamine synthetase activity at young stage in hydroponical conditions (QTL detected by means of 1 year experiment)	Unpublished
GDHa	Aminating glutamate dehydrogenase activity of leaf below the ear 15 days after flowering in field conditions	Dubois et al. 2003
Plant_NUtE	Plant N-utilization efficiency (plant biomass/total N-uptake = 1/N-content) in field conditions	Bertin and Gallais 2001
Plant_NUpE	Whole plant N-uptake	Bertin and Gallais 2001
NITF	Nitrate content of young leaves in hydroponical condition	Hirel et al. 2001
KW	Thousand kernel weight in field conditions	Bertin and Gallais 2001
cor_KW	Thousand kernel weight in field conditions corrected by earliness	Tables 3 and 4
KN	Number of grains per plant in field conditions	Bertin and Gallais 2001
cor_KN	Number of grains per plant in field conditions corrected by earliness	Tables 3 and 4
GY	Grain yield in field conditions	Bertin and Gallais 2001
cor_GY	Grain yield in field conditions corrected by earliness	Tables 3 and 4
Days_to_silk	Flowering date in day since 1 June in field	Bertin and Gallais 2001
QREMOB_STEM	Nitrogen remobilization from the stem in field conditions	Gallais and Hirel 2004
QREMOB_TOT	Total nitrogen remobilization in field	Gallais and Hirel 2004
T50	Grain germination efficiency	Limami et al. 2002
GS48	Glutamine synthetase activity at early stage of germination (48 h)	Limami et al. 2002
LI8	Length of internode 8	Guingo et al. 1998
DI8	Diameter of internode 8	Guingo et al. 1998
Angle7	Direction of roots at internode 7	Guingo et al. 1998
RI7	Number of roots on internode 7	Guingo et al. 1998
RI8	Number of roots on internode 8	Guingo et al. 1998

Results and discussion

Agronomic evaluation of C1 populations

The yield performances of populations C1N1 and C1N0 were only significantly different for evaluation at high N-input, with the expected superiority of C1N1 (Table 2). This means that for adaptation to high N-input, direct selection (i.e. under N1-condition) was more efficient than indirect selection (i.e. under N0-condition), whereas for adaptation to low N-input, both types of selection, direct and indirect, gave the same result. The superiority of direct selection versus indirect selection at high N-input is consistent with the observed grain yield heritabilities and genetic correlation between grain yield under N0 and N1-conditions ($\rho = 0.75$) (Bertin and Gallais 2000). Indeed, to have indirect selection become more efficient in environment 1 than direct selection in environment 2, it is required to have: $\rho^2 h_1^2 > h_2^2$ (Gallais 1983), ρ being the genetic correlation between environment 1 and environment 2 and h_i^2 ($i = 1$, 2) the heritability for environment i. Heritability in C0 at high N-input ($h^2_{N1} = 0.69$) being higher than at low N-input ($h^2_{N0} = 0.53$), for the adaptation to high N-input indirect selection is necessarily less efficient than direct selection ($\rho^2 h^2_{N0} = 0.30 < 0.69$). Similarly, for the adaptation to low N-input, heritability at high N-input was not sufficiently high to allow indirect selection to be more efficient than direct selection ($\rho^2 h^2_{N1} = 0.39 < 0.53$). In addition, heritability in N0 was not sufficiently high to lead to a significant advantage in direct selection. The re-estimation of heritabilities at the level of C1

Table 2 Grain yield (0.1 t/ha) response to selection of populations C1N1 and C1N0 evaluated in N0 and N1-conditions

	N-level for grain yield evaluation	
	N0	N1
N0-selection	57.9[a]	70.2[b]
N1-selection	58.8[a]	77.5[c]

Results of grain yield are based on the means of 10 environments (i.e. location × year combinations)
Two means with a different letters (a, b, c) are significantly different at 0.05

Table 3 Results of QTLs detection based on means of Bertin and Gallais (2001)

Traits	Chr	N-condition	LOD	Left_marker	Abs_pos	IC	r^2_P
GY	ch1	N1	2.32	bnlg1643	190	184–200	10.3
	ch2	N0	2.02	bnlg381	82	64–100	9
		N1	2.34	phi092	120	114–136	10.4
		N0	2.02	umc16a	118	100–124	9
		N0	2.07	phi127	156	140–170	9.3
	ch4	N1	2.59	umc1051	166	150–192	11.6
		N0	3.22	gys34a_PEPC	198	188–208	16.2
	ch5	N1	2.19	gsy249b_B70	88	66–102	9.8
		N1	2.25	umc2013	166	150–176	10.1
		N0	2.28	gsy258a_ANTI	154	138–172	10.1
	ch6	N1	2.87	bnlg249	16	10–24	12.8
KW	ch1	N1	3	bnlg2295	94	82–102	13
		N1	3.67	umc1335	118	110–126	15.7
		N0	5.69	umc1335	118	112–122	23.3
		N0	3.24	umc1035	140	128–152	14
	ch2	N1	3	bnlg1175	108	100–114	13
		N0	2.33	bnlg1175	108	96–118	10.3
	ch4	N1	4.14	phi079	98	88–110	17.5
		N0	3.68	phi079	94	72–110	15.7
	ch5	N1	2.75	gsy258a_ANTI	158	144–170	12
		N0	2.29	gsy258a_ANTI	158	144–170	10.1
KN	ch1	N1	2.34	bnlg2204	30	22–36	10.5
		N1	4.61	bnlg1057	112	106–122	19.5
		N1	4.61	umc1035	146	136–160	19.5
		N1	2.35	umc1421	248	238–258	10.7
	ch3	N1	2.12	umc1135	148	136–158	9.5
	ch4	N0	2.66	gsy34a_PEPC	204	192–208	13.6
	ch6	N0	2.46	bnlg426	20	12–42	10.9
	ch8	N1	2.16	umc36a	190	140–196	10.8
Days_to_silk	ch5	N0	2.57	gsy249b_B0	86	74–102	11.3
	ch8	N1	2.36	umc1343	90	80–108	10.4
		N0	2.31	umc1343	90	80–104	10.2
	ch10	N1	2.36	bnlg1655	84	70–96	10.2
		N1	2.79	gsy64_RS	110	100–122	12.3
		N0	3.24	gsy64_RS	112	102–124	14.1

The table displays results for grain yield (GY), kernel number (KN) and kernel weight (KW)
Chr chromosome, *LOD* value of the LOD score test, *Left_marker* name of the flanking left marker, *Abs_pos* absolute position of the QTL, *IC* confidence interval of the position of the QTL, r^2_P percentage of phenotypic variance explained by the QTL

populations gives $h^2_{N1}=0.71$ and $h^2_{N0}=0.50$, taking into account differences in the experimental design. This confirms the lower grain yield heritability at low N-input than at high N-input.

Predicted versus observed change in marker allele frequencies and re-estimation of QTL effects

Application of formula 2 with $i=1.46$ and $r^2_P=0.12$ leads to $\Delta p=0.25$; this is qualitatively in accordance with the observed significant allele frequency variations which varied between 0.25 and 0.39 (Table 6). However, due to the selection of Δp values greater than the statistical threshold (0.23 for a 0.05 threshold), on average such values are necessarily overestimated.

From the observed change in gene frequency, knowing the selection intensity, it is possible to estimate r^2_P (Table 6). This leads to consistent values in comparison with those estimated for grain yield and its components in the QTL detection experiment (Tables 3, 4), although a higher estimation of r^2_P was found for regions 10 and 13 on chromosomes 5 and 6, respectively. This could be the result of random variation of Δp. Indeed, considering the overestimation of Δp, an overestimation of QTL effects was expected. This selection bias was equivalent to that in classical QTL detection although the overestimation could be greater due to the large random variation of Δp.

Global genetic differentiation between populations

A preliminary two-way analysis of variance on all marker frequencies in C2, for both conditions of selection, showed a significant effect of the N-level of selection, with a higher fixation of *Io* alleles in N1-condition than in N0 (data not shown). However, this approach was not rigorous as the markers were not independent. The pairwise F_{ST} test and the exact test of population differentiation were more accurate, and both led to the conclusion of absence of divergence between populations, except between RILs and populations resulting from the second cycle of selection in N0 (C2N0) and N1 (C2N1), if we accepted a threshold of 10% for the exact test of Fisher (Table 5).

General trends in marker allele frequency changes

Selected markers for which the Waples test was significant are reported in Table 6. It illustrates that 19 genomic regions exhibited selection signature, with 12 identified for N1-selection. This was consistent with the agronomic evaluation in C1 displaying that N1-selection led to population adaptation to high N-input whereas both N-selections gave similar population performance at low N-input. Therefore, agronomic and genotypic results displayed the low efficiency of the N0-selection. Nearly half of the genomic regions identified

Table 4 Results of QTLs detection based on means of Bertin and Gallais (2001)

Traits corrected by earliness	Chr	N-condition	LOD	Left_marker	Abs_pos	IC	r^2_P
cor_GY	ch1	N1	2.87	bnlg1643	192	186–202	12.6
	ch2	N0	3.55	umc16a	118	114–122	15.2
	ch4	N0	2.21	gsy474_a_HHU524	22	4–40	11.1
		N1	2.94	bnlg572	180	150–192	13.5
	ch6	N1	4.24	bnlg249	16	12–22	18.3
			2.64	bnlg426	20	12–40	11.7
	ch8	N1	2.56	gsy224b_SPS	164	142–188	11.6
cor_KW	ch1	N1	2.55	bnlg2295	94	80–102	11.2
		N1	3.37	umc1335	118	110–126	14.5
		N0	5.33	umc1335	118	110–124	22
		N0	3.33	umc1035	140	128–152	14.4
	ch2	N1	2.64	bnlg1175	106	100–114	11.5
		N0	2.08	bnlg1175	106	96–118	9.2
	ch4	N1	3.79	phi079	96	82–110	16.2
		N0	3.39	phi079	94	80–108	14.6
	ch5	N1	3.05	gsy258a_ANTI	158	146–168	13.2
cor_KN	ch1	N1	2.65	bnlg2204	30	22–36	11.8
		N1	4.72	bnlg1590	112	106–122	19.9
		N0	3.76	umc1590	110	104–116	16.1
		N1	3.86	umc1035	146	134–178	16.6
		N1	2.27	umc1421	248	238–258	10.3
	ch3	N0	2.59	umc1135	146	136–158	11.4
	ch6	N1	2.64	bnlg249	14	2–24	11.8
		N0	3.04	bnlg426	20	12–30	13.3
		N0	2.13	umc1897	182	156–204	9.8
	ch8	N0	2.21	bnlg2235	38	24–60	9.8
		N0	2.8	gsy224b_SPS	164	140–186	12.6

The table displays results for grain yield (cor_GY), kernel number (cor_KN) and kernel weight (cor_KW) corrected by earliness
Chr chromosome, *LOD* value of the LOD score test, *Left_marker* name of the flanking left marker, *Abs_pos* absolute position of the QTL, *IC* confidence interval of the position of the QTL, r^2_P percentage of phenotypic variance explained by the QTL

were specifically selected during the second cycle (9/19 with significant C1 vs. C2 *t* test), and only one region appeared to be specifically selected in C1 cycle (region 5). Meanwhile, as could be expected, 7 out of the 19 regions under selection were selected in the first cycle and confirmed in the second selection cycle. Unfortunately, none of the *t* tests comparing marker allele frequencies in populations N0 and N1 were significant. Therefore, we cannot conclude with confidence on the N-level specificity of any marker selection, even if we can give a tendency, particularly for loci strongly selected in only one N-condition. Oppositely, two genomic regions appeared to be selected in both N-conditions of selection. For selection under high N-input, nearly half of the regions were selected since the C1 cycle, whereas for selection under low N-input only two regions out of seven were selected in C1.

Markers selected only in C1 cycle

Only two regions targeted by selection in C1 had not been confirmed in C2. The first one was the region number 5 on chromosome 2. It was under selection in N1-condition ($\alpha = 0.5\%$) and would be specific to C1. QTL coincidences and markers' ANOVA led to the assumption that this region is involved in grain yield production, kernel weight, kernel number and root architecture (LI8 trait), with a positive effect of the allele coming from the parental line F2. The second region was defined by marker *bnlg1401*, weakly selected on chromosome 9 ($\alpha = 5\%$). Its role in grain yield determinism was not clear not only because of the low Δp (−0.25) but also because the QTL coincidences were not very strict.

Markers selected in C1 and confirmed in C2

This pattern of response to selection represented most of the regions under C1 selection (seven out of nine). It may correspond to loci conferring general abilities for grain yield elaboration, with low genotype by environment interactions and low epistatic effect. In this situation we expected direct selection of grain yield (or component) QTLs detected on the RIL population.

Table 5 Test of global differentiation between populations

	RIL	C1N0	C1N1	C2N0
C1N0	0.419 (0.121) 1			
C1N1	1.697 (0.0005) 1	0.234 (0.34) 1		
C2N0	1.842 (>0.00001) 0.06	0.145 (0.35) 0.25	0.544 (0.2) 1	
C2N1	4.521 (>0.00001) 0.09	4.227 (0.002) 0.24	0.163 (0.34) 1	1.448 (0.072) 1

The first line displays F_{ST} (multiplied by 100) and its associated *P* value in brackets; the second line is the *P* value of the exact test of population differentiation computed

This was the case of marker *bnlg1643* on chromosome 1, selected with relatively low intensity ($\alpha = 3\%$) in N1-condition. This marker colocalized with QTLs of grain yield and N-uptake detected in N1, all with positive effect of the *Io* allele which is selected. The ANOVA results confirmed the positive effect of the *Io* allele for grain yield in N1 and also in N0-condition. It was consistent with the significant reduction of the *F2* allele frequency between C1 and C2 in N0-selection.

On chromosome 5, two regions exhibited highly significant C1 and C2 selection signatures ($\alpha = 1\%$). First, the region defined by *umc1060* was selected for the *Io* allele under N1-condition. This region seemed to be involved in grain yield determinism because of its colocalization with a QTL of grain yield detected in N1 with positive effect of the selected *Io* allele. This putative function under N1-condition was also corroborated by ANOVA. The second region of chromosome 5 defined by loci *mmc0481* and *bnlg278* was under a strong selection for the *Io* haplotype in N1-condition ($\Delta p = -0.40$). This region is of great interest because it was one of the two genomic zones where numerous QTLs for root morphology traits overlapped (Guingo et al. 1998). A QTL of leaf glutamine synthetase (GS) activity, a key gene for NUE, was also found in this zone with positive effect of the allele coming from the parental line Io. Furthermore, considering the ANOVA results, the region seemed to be involved in grain yield production under N1-condition. As suggested by Guingo et al. (1998), we may deduce that this region was involved in the control of the general development and growth process governed by the quantity of available nutrients.

Finally, the three regions of chromosome 6 with selection signature in C1 were confirmed during C2. The region defined by loci *y1ssr*, *bnlg249* and *bnlg426* was strongly selected ($\Delta p = -0.40$) in both cycles in N1-condition for the *Io* haplotype. Based on QTL coincidences, this region could be involved in grain yield and kernel number at both N-levels. The second region of chromosome 6 defined by markers *umc1143* and *bnlg161* was near the first one and also selected in N1 for the *Io* haplotype but with lower intensity, and it was in the confidence interval of previous QTLs. Interestingly, Khavkin and Coe (1997) have shown that this chromosomal segment (bin 6.01) displayed a developmental gene cluster combining naked-eye polymorphisms (e.g. gene for hormone sensors), homeotic genes (e.g. transcription factor) and QTLs for traits such as grain yield. The last marker under selection on this chromosome was *umc1653* in N0-selection. This marker coincided with QTLs for grain yield and kernel weight specifically detected in N0-condition on line per se. Therefore, this region could be involved in the grain yield setting, probably through an efficient grain filling.

The case of marker *bnlg1175* on chromosome 2 was attractive because it was significantly selected ($\Delta p = 0.30$) in C1 cycle in both N-selections and in C2 cycle but only for N0-level. It colocalized with QTLs of kernel weight

Table 6 Characterization of regions under selection

Chr	Markers	Abs_dist	Region	Waples test						Region description				Locus-by-locus ANOVA		
				N1-selection			N0-selection									
				RIL versus C1N1	RIL versus C2N1	C1N1 versus C2N1	RIL versus C1N0	RIL versus C2N0	C1N1 versus C2N1	N-condition	Allele	QTL (or gene) coincidence	r_P^2	Allele	N-condition	Traits
1	dupssr26	65	1	–	−0.40***	−0.23***	–	–	–	N1	Io	GS activity		–	–	–
	bnlg2295	67.5		–	−0.30(*)	−0.16(*)	–	–	–							
	bnlg1643	185.6	2	−0.28*	−0.30(*)	–	–	–	−0.17*	N1?	Io	N1: grain yield, N-uptake	11.7	Io	N1 and N0	Grain yield, kernel number
2	bnlg2248	66.9	3	–	−0.33*	−0.32***	–	−0.30(*)	−0.27***	N1 and N0	Io	C-metabolism		F2	N0	Grain yield
	umc1580	91	4	–	–	–	0.25(*)	–	–							
	bnlg1175	103.6		0.34**	–	–	0.28*	0.32*	–	N1 and N0	F2	N1 and N0: kernel weight; N1: N-remobilization	13.3	F2	N1 and N0	Grain yield, kernel weight
	phi092	119.1	5	0.33***	–	−0.16(*)	–	–	–	N1	F2	N1 and N0: grain yield; N1: roots architecture	14.1	F2	N1 and N0	Grain yield, kernel number
	umc1079	123.6		0.28**	–	–	–	–	–					F2	N1 and N0	Grain yield, kernel weight
	bnlg371	124.1		0.25(*)										F2	N1 and N0	Grain yield, kernel weight
	bnlg1940	203.7	6	–	–	–	–	0.40**	0.34***	N0	F2	Stress-inducible enzyme		–	–	–
3	bnlg1601	76.8	7	–	0.30*	–	–	–	–	N1	F2	?		–	–	–
	mmc0022	79.9		–	0.30*	–	–	–	–							
	dupssr23	99.1	8	–	–	0.31***	–	0.50***	0.39***	N0?	F2	Stress-inducible enzyme		Io	N1	Grain yield
4	bnlg2291	121.8	9	–	–	–	–	−0.30(*)	−0.19*	N0	Io	N1: roots GS activity; structural GS gene		Io	N1	Kernel weight
5	umc1060	92	10	−0.33***	−0.38**	–	–	–	–	N1	Io	N1: grain yield; N0: precocity	18.8	Io	N1	Grain yield
	dupssr10	94.6		–	−0.31(*)	–	–	–	–							
	mmc0481	123.6	11	−0.28*	−0.40**	–	–	–	–	N1	Io	N1: roots architecture and GS activity		Io	N1	Grain yield
	bnlg278	125.6		−0.25(*)	−0.38**	–	–	–	–							
	bnlg389	240.1	12	–	0.34*	0.34***	–	–	0.34***	N1?	F2	?		ne	ne	ne
6	y1ssr	8	13	−0.29*	−0.43***	–	–	–	–	N1	Io	N1 and N0: grain yield, kernel number	24	Io	N1	Grain yield, kernel number
	bnlg249	11.6		−0.33***	−0.40***	–	–	–	–							
	bnlg426	16.3		−0.39***	−0.40***	–	–	–	–							
	umc1143	26	14	−0.28*	–	–	–	–		N1	Io	N0: grain yield, kernel number	10.9	Io	N1	Grain yield, kernel number
	bnlg161	30.4		−0.31**	−0.29(*)	–	–	–	–							
	umc1653	204.8	15	–	–	–	0.25(*)	0.29(*)	–	N0	F2	N0: grain yield, kernel weight	10.9	–	–	–
8	bnlg2082	53.6	16	–	–	–	–	−0.33(*)	−0.16(*)	N0	Io	N0: plant grain number	14.1	Io	N0	Grain yield
	bnlg666	96.6	17	–	−0.30(*)	−0.22***	–	–	–	N1?	Io	N1 and N0: precocity	11.7	–	–	–

Table 6 (Contd.)

Chr	Markers	Abs_dist	Region	Waples test						Region description			Locus-by-locus ANOVA			
				N1-selection			N0-selection									
				RIL versus C1N1	RIL versus C2N1	C1N1 versus C2N1	RIL versus C1N0	RIL versus C2N0	C1N1 versus C2N1	N-condition	Allele	QTL (or gene) coincidence	r^2_P	Allele	N-condition	Traits
bnlg162	96.6		–	−0.30(*)	−0.30***	–	–	−0.16(*)								
9	bnlg1401	36	18	−0.25(*)	–	–	–	–	–	N1	Io	N1and N0: N-utilization efficiency		–	–	–
	umc1366	123.6	19	–	0.45***	0.48***	–	–	0.16(*)	N1?	F2	Structural GS gene		–	–	–

The table displays loci with signature of selection according to the selection cycle (C1 and C2) and the N-condition of selection (N1 or N0-selection) based upon results of Waples test. For this test both the *F2* frequency variation (Δp) and the level of significance are specified. Several information about regions are also mentioned: the selected allele, the putative N-condition of selection, QTL (or gene) coincidence, the percentage of phenotypic variance explained by the marker (r^2_P) estimated with the Δp at C2 (formula 1) in the case of grain yield or component QTL coincidences. Results of locus-by-locus analysis of variance (ANOVA) performed on population C0: the allele with positive effect and the associated traits are mentioned. For the threshold: *ne* non-estimated, – test non-significant, (*) test significant at $P = 0.05$, *test significant at $P = 0.03$, **test significant at $P = 0.01$, ***test significant at $P = 0.005$

Chr chromosome, *Abs_dist* absolute distance of markers along chromosome in cM, *Region* genomic regions under selection taking into account hitchhiking effect between close markers

and total amount of N remobilized, all with positive effect of the *F2* allele which was selected in the region. Besides, ANOVA outlined the positive effect of the *F2* allele in yield production in both N-conditions. Therefore, we could suppose that loci under selection in this region contributed to grain yield genetic determinism, whatever the N-fertilization, probably through a better N-utilization leading to an efficient grain filling.

Markers selected in C2 cycle only

This type of selection concerned half of the regions under selection (10/19) and two groups of regions were observed. The first one was defined by markers selected in C2 cycle but with non-significant C1 versus C2 *t* test. Only one region, weakly selected ($\alpha = 3\%$), on chromosome 3 was observed but the putative interest of this region for NUE cannot be determined because of the lack of QTL (or gene) colocalization.

The second group included markers specifically selected in C2 cycle, with significant C1 versus C2 *t* test. That represented 9 out of the 19 regions with selection signature. On chromosome 1, marker *dupssr26* displayed a strong selection signature for the *Io* allele in N1-condition ($\Delta p = -0.40$). In this region, we have detected a QTL of leaf GS activity at young stages with favourable effect of the *Io* allele, precisely selected. GS is one of the main enzymes involved in the assimilation and recycling of mineral N (Lea and Ireland 1999; Cren and Hirel 1999) and its function in NUE has already been demonstrated (Gallais and Hirel 2004).

The *Io* allele of markers *bnlg2291* and *bnlg2082*, respectively, on chromosomes 4 and 8 was slightly selected ($\alpha = 5\%$) under N0-condition. For *bnlg2291* on chromosome 4, QTLs of GS activity in young roots and GS activity at early stage of germination overlapped with the line Io as the more efficient parent. Interestingly, this region contains one of the six GS genes (*gln5* or *GS1-4*) which seems to be much more induced in roots compared with other GS isoforms (B. Hirel, personal communication). Therefore, we can propose *gln5* as a putative candidate to explain this response to selection. The locus on chromosome 8 could be putatively targeted in N0-condition because of its involvement in kernel number setting and grain yield.

The other regions specifically selected in C2 did not display any QTL colocalizations or significant effect on grain yield or components through markers' ANOVA. Meanwhile, some of them showed interesting gene coincidences. For instance, markers *bnlg1940* on chromosome 2 and *dupssr23* on chromosome 3 were both strongly selected in stress N-condition ($\alpha = 0.5\%$) and were in the close vicinity of stress-inducible enzymes: an ascorbate peroxydase (*gsy309f_Apx*) for *bnlg1940* and a lipoxygenase (*gsy461_Lipox*) for *dupssr23*. On chromosome 9, the region defined by *umc1366*, and strongly targeted by selection in N1 ($\Delta p = 0.45$) for the *Io* allele, was also worth mentioning because of the quasi-comapping between

umc1366 and a GS gene (*gln3* or *GS1-5*). Lastly, regions 12 and 17, respectively, on chromosomes 5 and 8 were under C2 selection in N1-condition and tended to be in N0, but the putative interest of these regions for grain yield cannot be determined because of the lack of QTL (or gene) colocalization with advantageous effect of the allele selected.

The region of chromosome 2 defined by *bnlg2248* was attractive because it was another case of common response to N0 and N1-selections with a highly significant change in marker frequency from C1 to C2 ($\Delta p = 0.30$). It may be involved in grain filling because Prioul et al. (1999) have detected in this region a QTL for grain amylopectin content, with positive effect of the *Io* allele selected (data not shown). They have hypothesized that the candidate gene underlying this trait was a vacuolar invertase (*gsy348c_Ivr1*), but the involvement of other genes like a transporter protein near *bnlg2248* could not be excluded.

General discussion

Lack of power of the design

According to the single locus test of Waples, in the first cycle the change in marker frequencies must be greater than 0.23 to be significant at 0.05, meaning that observed frequencies must be greater than 0.73. Such a large interval of expected random variation is due to the low number of selected plants contributing to a low effective size. To halve this interval, we would have been required to select 64 lines, meaning an RIL population size of about 360 lines if the same selection intensity was applied ($i = 1.458$). In the second cycle, the situation would have been worse because it cumulated two sampling levels. Then, the observed frequency must be greater than 0.77 to be different from the C0 frequency. In the same way frequency comparisons between N0 and N1 were never significant because, on average, a Δp of 0.40 was needed. This lack of power will be a property of all unidirectional selective genotyping design with inadequate size of initial population and low number of selected genotypes.

Observed versus predicted responses to selection

We have shown that 5.5 and 16% of markers surveyed within populations N0 and N1, respectively, rejected the null hypothesis that genetic drift was solely responsible for their allele frequency changes since C0. Therefore, we could conclude that allele frequency of those markers had changed in response to selection via selective sweep because they were associated with segregating QTLs, either directly or, more likely, indirectly because of physical linkage. However, these results must be interpreted with caution because such selection signature depends on the origin of linkage disequilibrium. In theory, interchromosomal linkage disequilibrium does not exist in an RIL population but it could be developed by the process of selection. Meanwhile, it would have been halved at each generation of intercrossing developed for the production of C2 populations. In such a way, interchromosomal linkage disequilibrium as intrachromosomal linkage disequilibrium between distant loci may be weak. Consequently, we presumed that significant variations of allele frequencies between C0, C1 and C2 populations revealed marker–QTL associations for grain yield determinism. However, expected marker–QTL associations were different according to the selection cycle.

C1 cycle

Half of the selected regions (9/19) exhibited selection signatures after C1 selection cycle. Five regions out of the 9 coincided with grain yield QTLs of major effects detected in the RIL population, and they represented only 4 out of the 10 QTLs detected in C0 for grain yield corrected or not for earliness. However, when ANOVA was computed on selected markers in C0 population, seven out of the nine regions selected were significantly associated with grain yield and/or components at a threshold of 0.05, without contradiction on allelic effect. This better correspondence with ANOVA results was expected because the main parameter influencing marker selection is the percentage of variance that marker explained $r_{\mathrm{M}}^2 = (1-2c)^2 r_{\mathrm{P}}^2$, i.e. the product of the percentage of phenotypic variance explained by the QTL and the square of the linkage parameter $\lambda = 1-2c$, as for the ANOVA. This means that selection occurred on QTLs of major effects but marker density is also an important factor for the identification of such genomic regions. The other factor which could affect the detection was the selection intensity. Increasing selection intensity would increase the expected change in selected marker frequency, but at the same time, it will increase random variation. By selecting only 10% of the C0 population (10 plants) we identified only one more grain yield QTL on chromosome 4 ($\Delta p = 0.28$). Our results are quite consistent with those of Moreau et al. (2004), for a similar selection design, displaying that in the first cycle of phenotypic selection the higher changes in marker frequencies were associated in direction with the QTLs of major effects.

C2 cycle

At the level of the second cycle three situations were observed. The first one was the confirmation of regions selected during C1, which included seven out of the nine regions selected in C1. Fortunately, it involved four of the five regions coinciding with grain yield QTL in C1. The second situation observed in C2 is the non-confirmation of regions selected in C1 (only two out of nine). The phenomenon of QTL-by-environment interactions can be invoked, because C2 evaluation was based on new environmental conditions that preclude the confirmation of C1 regions involved in such interactions.

Furthermore, the risk of false detection in C1 due to a low population size (Bernardo 2004; Melchinger et al. 2004) and the large sampling effect at the level of selected plants have also to be considered. The third observed situation in C2 was the selection of new regions. To explain such a situation, we can once again consider the low power of QTL detection in C0 and C1 populations led by the low population size. In this way, it is interesting to note that among these new regions three were detected by ANOVA in C0. Furthermore, the C0 phenotyping for 2 years in only one location and the more accurate evaluation of C1S1 for 2 years in two locations may lead to the selection of new loci not only because of an increase of the power of detection but also because of new QTL by environment interactions. New QTLs, with lower effects, could also contribute to response to selection due to the increase in frequency of favourable alleles at QTLs with larger effects, leading to a lower variance for such QTLs. A last factor could be the change in genetic background from C0 to C2, so that epistasis could play a certain role. However, it could be expected to be low, because at the level of testcross progenies, epistasis cannot contribute very much to the variation (Gallais 1991). In the Moreau et al. (2004) experiment, the same situations were observed in the second cycle of selection: new regions of the genome were affected by phenotypic selection whereas some regions detected in C1 disappeared in C2.

Genetic differentiation between populations

A low global genetic differentiation occurred between C0 and C2 populations, and no global genetic differentiation was observed between N0 and N1 populations. However, investigation of locus' temporal variation through the Waples test revealed that 21% of surveyed loci exhibited signature of selection, but without significant divergence between N0 and N1 populations. However, more loci appeared to be under selection in N1-condition which was in accordance with results of Bertin and Gallais (2001), who had detected fewer QTLs for grain yield and its components in stress N-condition. Furthermore, two genomic regions responding to selection were common to N0 and N1-conditions. It corroborates the observation that grain yield QTLs detected in N0 were very often a subset of N1 QTLs, but probably differentially expressed (Bertin and Gallais 2001).

Validation of grain yield QTL

Four genomic regions were identified as directly associated with grain yield exhibiting low QTL by environment or background interactions because of their selection in C1 and C2. On chromosome 1, the region defined by locus *bnlg1643* may be involved in grain yield determinism in optimal N-fertilization probably through an efficient plant N-uptake. Another interesting zone was on chromosome 5 (no. 10) and may be an illustration of pleiotropy, inducing that according to the environment, some QTLs can appear as QTLs for grain yield, which was the case in N1, or as QTLs for earliness, which was the case in N0. At the end of chromosome 6, locus *umc1653* would play a role in the adaptation to stress N-condition through an efficient grain filling. Finally, it appeared that a large part of bin 6.01 was involved in grain yield determinism probably through kernel number setting. It was strongly selected in N1, and QTLs were detected in both N-levels of fertilization. Such a region involved in genetic determinism of grain yield, with probably low N-fertilization interactions, could be of great interest for MAS.

Common putative function of selected regions

For grain yield, our results pointed out the consistency of the involvement of some morpho-physiological traits whatever the N-fertilization. First of all, our findings underlined the putative involvement of root architecture and activity in response to grain yield selection in 4 regions (out of the 19). In particular the targeted region on chromosome 5 outlined the possible interaction between root and shoot compartments governed by nutrient availability and hormones as proposed by Lemaire and Millard (1999). The genetic variability of some root morpho-physiological traits could be directly linked to genetic variability for grain yield whatever the nitrogen fertilization. This type of relationship has already been reported but in water-stressed type experiments (Lebreton et al. 1995; Tuberosa et al. 2002, 2003) rather than in NUE studies (Kamara et al. 2002, 2003). Finally, it appeared that efficient activity of GS would be selected in yielding genotype in both N-conditions. Indeed, four selected regions were targeted because they probably increase GS activity either in leaves or in roots and probably because of a GS gene vicinity. The positive correlation between grain yield and GS activity has already been observed by Gallais and Hirel (2004) and also demonstrated by experiences of GS overexpression (Gallardo et al. 1999; Migge et al. 2000). Finally, we can suppose that plant traits controlling productivity under stress N-conditions are constitutive rather than stress adapted, as underlined for drought stress (Blum 1988; Passouria 1996; Tuberosa et al. 2003); but they may be determined by different genomic regions and/or by different mechanisms of regulation (e.g. induction of stress-inducible enzyme under low nitrogen conditions).

Conclusion

The power of our phenotypic selection experiment which was aimed to confirm genomic regions with grain yield QTLs appeared to be low. However, the results are consistent with previous QTLs detection: more genomic

regions were detected with selection under high N-input, where grain yield response to selection was higher. Moreover, we have validated some QTLs for grain yield and for less complex traits, such as root architecture and glutamine synthetase activity, which may be determinants for grain yield setting whatever the N-fertilization. Reciprocally, selection on markers of such QTLs will be expected to increase grain yield in both N-conditions. Therefore, we may wonder whether a MAS using markers–QTL associations for traits less complex than grain yield, such as N-uptake, root architecture and key enzyme activity, would not be more efficient than a MAS using markers of grain yield QTLs only. In conclusion, the study of change in marker frequency by selection from a population where linkage disequilibrium is only due to physical linkage appears to be sufficiently efficient to detect markers linked to QTLs and useful for MAS. Indeed, even if it is less powerful than classical QTL detection, with a greater risk of false QTLs, the essential thing is to detect QTLs with the largest effects. Furthermore, it is a less expensive method requiring the genotyping of a low number of individuals.

Acknowledgments We are grateful to our experimental and molecular biology teams, particularly to Philippe Jamin and Denis Coubriche involved in agronomic evaluation and multiplication of the plant material and also to Valerie Combes, Delphine Madur and Fabrice Dumas in charge of the marker genotyping task of this program. We also thank Dr. Laurence Moreau and Dr. Isabelle Goldringer for their reading and critical review of the manuscript.

Appendix: Derivation of var(p_x–p_y)

In what follows, p_x represents the frequency of allele from F2 parent in population x (with $x=0$, 1 or 2). As we can write var(p_x–p_y) = var p_x + var p_y −2 cov($p_x p_y$), we have to then derive var p_x and cov($p_x p_y$).

Derivation of the variance of marker frequencies assuming only genetic drift

Our experiment corresponds to a sampling in three steps:

- From an infinite RIL population where $p=0.5$, 99 RIL have been drawn leading to the population C0 with a frequency p_0, and var $p_0=p(1-p)/N_0$, with $p=0.5$ and N_0 the number of RIL in C0.
- From C0 to C1: N_1 RIL have been selected (already genotyped), leading to $p_1|p_0$ (p_1 conditional to p_0), and var $p_1|p_0=p(1-p_0)/N_1$.
- From C1 to C2: N_2 non-inbred individuals ($2N_2$ genes) were selected and genotyped, leading to $p_2|p_1$ and var $p_2(p_1=p_1(1-p_1)/2N_2$.

To derive var p_1 and var p_2, we used the relationship var Y = E[var($Y|X$)] + var[E($Y|X$)], which means that the total variance is equal to the sum of intraclass variance and interclass variance. Then

$$\mathrm{var}p_n = E[\ \mathrm{var}(p_n|p_{n-1})] + \ \mathrm{var}[E(p_n|p_{n-1})]; \tag{3}$$

with $E(p_n|p_{n-1}) = p_{n-1}$ then

$$\mathrm{var}p_n = E[\ \mathrm{var}(p_n|p_{n-1})] + \ \mathrm{var}p_{n-1}. \tag{4}$$

Furthermore,

$$E[\ \mathrm{var}(p_n|p_{n-1})] = E\left[\frac{p_{n-1}(1-p_{n-1})}{\theta N_n}\right],$$

N_n being the number of sampled plants to develop generation n, and $\theta=1$ in C0 and C1, and $\theta=2$ in C2, with

$$E[p_{n-1}(1-p_{n-1})] = E(p_{n-1}) - E(p_{n-1}^2) \\ = p - (\ \mathrm{var}p_{n-1} + p^2),$$

results in

$$E[\ \mathrm{var}(p_1|p_0)] = \frac{p(1-p)}{N_1} - \frac{\mathrm{var}p_0}{N_1} \\ = p(1-p)\frac{1}{N_1}\left(1-\frac{1}{N_0}\right), \tag{5}$$

with

$$\mathrm{var}p_1 = p(1-p)\left[1-\left(1-\frac{1}{N_0}\right)\left(1-\frac{1}{N_1}\right)\right]$$

and

$$\mathrm{var}p_2 = p(1-p)\left[1-\left(1-\frac{1}{N_0}\right)\left(1-\frac{1}{N_1}\right)\left(1-\frac{1}{2N_2}\right)\right]. \tag{6}$$

More generally

$$\mathrm{var}p_n = p(1-p)\left[1-\prod_{i=0}^{i=n}\left(1-\frac{1}{\theta N_i}\right)\right],$$

which is indeed the Fisher–Wright formula when $\theta N_i=2N$.

According to expression 4 the variance in gene frequency at a generation m can also be written as the sum of all sampling variances from generation 0:

$$\mathrm{var}p_1 = E(\ \mathrm{var}p_1|p_0) + \ \mathrm{var}p_0,$$
$$\mathrm{var}p_2 = E(\ \mathrm{var}p_2|p_1) + \ \mathrm{var}p_1 \\ = E(\ \mathrm{var}p_2|p_1) + E(\ \mathrm{var}p_1|p_0) + \ \mathrm{var}p_0,$$

and more generally

$$\mathrm{var}p_m = \sum_{i=1}^{i=m} E(\ \mathrm{var}p_i|p_{i-1}) + \ \mathrm{var}p_0. \tag{7}$$

Derivation of cov p_1p_2, cov p_0p_1, cov p_0p_2

At a generation n deriving from an ancestor generation m, it is possible to write

$$p_n = p_m + \varepsilon,$$

where ε is a random deviation due to sampling process from m to n, resulting in

$$\text{cov} p_m p_n = \text{var} p_m.$$

Thus, cov p_1p_0 = var p_0, cov p_2p_0 = var p_0 and cov p_2p_1 = var p_1.

References

Agrama H, Zakaria A, Said F, Tuinstra M (1999) Identification of quantitative trait loci for nitrogen use efficiency in maize. Mol Breed 5:187–195

Albini G, Falque M, Joets J (2003) ActionMap: a web-based software that automates loci assignments to framework maps. Nucleic Acids Res 31:3815–3818

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

Ayoub M, Mather D (2002) Effectiveness of selective genotyping for detection of quantitative trait loci: an analysis of grain and malt quality traits in three barley populations. Genome 45:1116–1124

Balko LG, Russel WA (1980) Effects of rate of nitrogen fertilizer on maize inbred lines and hybrid progeny. I. Prediction of yield response. Maydica 25:65–79

Beauchamp EG, Kannenberg LW, Hunter RB (1976) Nitrogen accumulation and translocation in corn genotypes following silking. Agron J 68:418–422

Beavis WD (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. In: Annual corn sorghum research conference 49th. American Seed Trade Association, Washington, pp 252–268

Bernardo R (2004) What proportion of declared QTL in plants are false? Theor Appl Genet 109:419–424

Bertin P, Gallais A (2000) Genetic variation for nitrogen use efficiency in a set of recombinant maize inbred lines I. Agrophysiological results. Maydica 45:53–66

Bertin P, Gallais A (2001) Physiological and genetic basis of nitrogen use efficiency. II. QTL detection and coincidences. Maydica 46:53–68

Blum A (1988) Plant breeding for stress environments. CRC press, USA

Causse M, Santoni S, Damerval C, Maurice A, Charcosset A, Deatrick J, de Vienne D (1996) A composite map of expressed sequences in maize. Genome 39:418–432

Cren M, Hirel B (1999) Glutamine synthetase in higher plants: regulation of gene and protein expression from the organ to the cell. Plant Cell Physiol 40:1187–1193

Darvasi A, Soller M (1992) Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. Theor Appl Genet 85:353–359

Dubois F, Terce-Laforgue T, Gonzalez-Moro M-B, Estavillo J-M, Sangwan R, Gallais A, Hirel B (2003) Glutamate dehydrogenase in plants: is there a new story for an old enzyme? Plant Physiol Biochem 41:565–576

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. Longman, Halow

Gallais A (1983) Use of indirect selection in plant breeding. In: Lange ZAW, Hogenboom NG (eds) Efficiency in plant breeding. Proceedings of the eucarpia congress, Wageningen, The Netherlands

Gallais A (1991) A general approach for the study of a population of test cross progenies and consequences for recurrent selection. Theor Appl Genet 81:493–503

Gallais A, Hirel B (2004) An approach to the genetics of nitrogen use efficiency in maize. J Exp Bot 55:295–306

Gallardo F, Fu J, Canton F, Garcia-Gutierrez A, Canovas F, Kirby E (1999) Expression of a conifer glutamine synthetase gene in transgenic poplar. Planta 210:19–26

Griffing B (1961) Theoretical consequences of truncation selection based on the individual phenotype. Aust J Biol Sci 13:307–343

Guingo E, Hebert Y, Charcosset A (1998) Genetic analysis of root traits in maize. Agronomie 18:225–235

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 Physiol 125:1258–1270

Kamara AY, Kling JG, Ajala SO, Menkir A (2002) The relationship between vertical root-pulling resistance and nitrogen uptake and utilization in maize breeding lines. Maydica 47:135–140

Kamara AY, Kling JG, Menkir A, Ibikunle O (2003) Agronomic performance of maize (*Zea mays L.*) breeding lines derived from a low nitrogen maize population. J Agric Sci 141:221–230

Khavkin E, Coe EH (1997) Mapped genomic locations for developmental functions and QTLs reflect concerted groups in maize (*Zea mays L.*). Theor Appl Genet 95:343–352

Landbeck MV (1995) Untersuchungen zur genetischen verbesserung der anbaueigung von körnermais unter produktionsbedingungen mit verringerter sticksoffversorgung, University of Hohenheim, 109 pp

Lea P, Ireland R (1999) Nitrogen metabolism in higher plants. In: Singh B (ed) Plant amino acids. Marcel Dekker, New York, pp 1–47

Lebreton C, Lazic-Janvic V, Steed A, Pekic S, Quarrie S (1995) Identification of QTL for drought responses in maize and their use in testing causal relationships between traits. J Exp Bot 46:853–865

Lemaire G, Millard P (1999) An ecophysiological approach to modelling resource fluxes in competing plants. J Exp Bot 50:15–28

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 Physiol 130:1860–1870

Lin H, Yamamoto T, Sasaki T, Yano M (2000) Characterization and detection of epistatic interactions of 3 QTLs, *Hd1*, *Hd2* and *Hd3*, controlling heading date in rice using nearly isogenic lines. Theor Appl Genet 101:1021–1028

Melchinger AE, Utz HF, Schön CC (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics 149:383–403

Melchinger AE, Utz HF, Schön CC (2004) QTL analyses of complex traits with cross validation, bootstrapping and other biometric methods. Euphytica 137:1–11

Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics 142:1061–1064

Migge A, Carrayol E, Hirel B, Becker T (2000) Leaf-specific overexpression of plastidic glutamine synthetase stimulates growth of transgenic tobacco seedlings. Planta 210:252–260

Moll RH, Kamprath EJ, Jackson WA (1987) Development of nitrogen efficient prolific hybrids of maize. Crop Sci 27:181–186

Moreau L, Charcosset A, Gallais A (2004) Experimental evaluation of several cycles of marker-assisted selection in maize. Euphytica 137:111–118

Passouria J (1996) Drought and drought tolerance. Plant Growth Regul 20:79–83

Paterson AH, DeVerna JW, Lanini B, Tanksley SD (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. Genetics 124:735–742

Pollmer WG, Eberhard D, Klein D, Dhillon BS (1979) Genetic control of nitrogen uptake and translocation in maize. Crop Sci 19:82–85

Presterl T, Seitz G, Schmidt W, Geiger HH (2002) Improving nitrogen use efficiency in European maize comparison between line per se and testcross performance under high and low soil nitrogen. Maydica 47:83–91

Prioul J, Pelleschi S, Sene M, Theevenot C, Causse M, de Vienne D, Leonardi A (1999) From QTLs for enzyme activity to candidate genes in maize. J Exp Bot 50:1281–1288

Raymond M, Rousset F (1995) An exact test for population differentiation. Evolution 49:1280–1283

Reed AJ, Below FE, Hageman RH (1980) Grain protein accumulation and relationship between leaf nitrate reductase and proteases activities during grain development in maize. I. Variation between genotypes. Plant Physiol 66:164–170

Russel WA (1984) Further studies on the response of maize inbred lines to N fertilizer. Maydica 29:141–150

Schnell FW (1961) Some general formulations of linkage effects in inbreeding. Genetics 46:947–957

Smith M, Haigh J (1974) The hitch-hiking effect of a favorable gene. Genet Res 23:23–35

Stuber CW, Moll RH, Goodman MM, Schaffer HE, Weir BS (1980) Allozyme frequency changes associated with selection for increased grain yield in maize (*Zea mays L.*). Genetics 95:225–236

Stuber CW, Goodman MM, Moll RH (1982) Improvement of yield and ear number resulting from selection at allozyme loci in a maize population. Crop Sci 22:737–740

Tuberosa R, Salvi S, Sanguineti MC, Landi P, MacCaferri M, Conti S (2002) Mapping QTLs regulating morpho-physiological traits and yield: case studies, shortcomings and perspectives in drought-stressed maize. Ann Bot 89:941–963

Tuberosa R, Salvi S, Sanguineti MC, Maccaferri M, Giuliani S, Landi P (2003) Searching for quantitative trait loci controlling root traits in maize: a critical appraisal. Plant Soil 255:35–54

Utz HF, Melchinger AE (1995) PLABQTL: a computer program to map QTLs. JQTL 1:2

Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. Genetics 121:379–391

Weir B, Cockerham C (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370

Wingbermuehle WJ, Gustus C, Smith KP (2004) Exploiting selective genotyping to study genetic diversity of resistance to Fusarium head blight in barley. Theor Appl Genet 109:1160–1168

Zhang LP, Lin GY, Niño-Liu D, Foolad MR (2003) Mapping QTLs conferring early blight (*Alternaria solani*) resistance in a *Lycopersicon esculentum* * *L. hirsutum* cross by selective genotyping. Mol Breed 12:3–19