Theor Appl Genet (2014) 127:2313–2331
DOI 10.1007/s00122-014-2379-7

ORIGINAL PAPER

Dent and Flint maize diversity panels reveal important genetic potential for increasing biomass production

R. Rincent · S. Nicolas · S. Bouchet · T. Altmann · D. Brunel · P. Revilla · R. A. Malvar · J. Moreno-Gonzalez · L. Campo · A. E. Melchinger · W. Schipprack · E. Bauer · C.-C. Schoen · N. Meyer · M. Ouzunova · P. Dubreuil · C. Giauffret · D. Madur · V. Combes · F. Dumas · C. Bauland · P. Jamin · J. Laborde · P. Flament · L. Moreau · A. Charcosset

Received: 23 May 2014 / Accepted: 15 August 2014 / Published online: 10 October 2014

Abstract
Key message **Genetic and phenotypic analysis of two complementary maize panels revealed an important variation for biomass yield. Flowering and biomass QTL were discovered by association mapping in both panels.**

Abstract The high whole plant biomass productivity of maize makes it a potential source of energy in animal feeding and biofuel production. The variability and the genetic determinism of traits related to biomass are poorly known. We analyzed two highly diverse panels of Dent and Flint lines representing complementary heterotic groups for Northern Europe. They were genotyped with the 50 k SNP-array and phenotyped as hybrids (crossed to a tester of the complementary pool) in a western European field trial network for traits related to flowering time, plant height, and biomass. The molecular information revealed to be a

Communicated by Michael Gore.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-014-2379-7) contains supplementary material, which is available to authorized users.

R. Rincent · S. Nicolas · S. Bouchet · D. Madur · V. Combes · F. Dumas · C. Bauland · P. Jamin · L. Moreau · A. Charcosset (✉)
UMR de Génétique Végétale, INRA, Université Paris-Sud, CNRS, AgroParisTech, Ferme du Moulon, 91190 Gif-Sur-Yvette, France
e-mail: alain.charcosset@moulon.inra.fr

R. Rincent · P. Dubreuil
BIOGEMMA, Genetics and Genomics in Cereals, 63720 Chappes, France

R. Rincent · N. Meyer · M. Ouzunova
KWS Saat AG, Grimsehlstr 31, 37555 Einbeck, Germany

R. Rincent · P. Flament
Limagrain, site d'ULICE, av G. Gershwin, BP173, 63204 Riom Cedex, France

Present Address:
S. Bouchet
Department of Agronomy, Throckmorton Plant Science Center, Kansas State University, Manhattan, KS 66506, USA

T. Altmann
Max-Planck Institute for Molecular Plant Physiology, 14476 Potsdam-Golm, Germany

T. Altmann
Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK), 06466 Gatersleben, Germany

D. Brunel
INRA, UR 1279 Etude du Polymorphisme des Génomes Végétaux, CEA Institut de Génomique, Centre National de Génotypage, 2, rue Gaston Crémieux, CP5724, 91057 Evry, France

P. Revilla · R. A. Malvar
Misión Biológica de Galicia, Spanish National Research Council (CSIC), Apartado 28, 36080 Pontevedra, Spain

J. Moreno-Gonzalez · L. Campo
Centro de Investigaciones Agrarias de Mabegondo, Apartado 10, 15080 La Coruna, Spain

A. E. Melchinger · W. Schipprack
Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, Fruwirthstr.21, 70599 Stuttgart, Germany

E. Bauer · C. Schoen
Plant Breeding, Technische Universität München, 85354 Freising, Germany

powerful tool for discovering different levels of structure and relatedness in both panels. This study revealed important variation and potential genetic progress for biomass production, even at constant precocity. Association mapping was run by combining genotypes and phenotypes in a mixed model with a random polygenic effect. This permitted the detection of significant associations, confirming height and flowering time quantitative trait loci (QTL) found in literature. Biomass yield QTL were detected in both panels but were unstable across the environments. Alternative kinship estimator only based on markers unlinked to the tested SNP increased the number of significant associations by around 40 % with a satisfying control of the false positive rate. This study gave insights into the variability and the genetic architectures of biomass-related traits in Flint and Dent lines and suggests important potential of these two pools for breeding high biomass yielding hybrid varieties.

Introduction

Maize is together with wheat and rice one of the three main sources of nutritional energy for humans and is extensively being used in animal feeding, either as grain or whole plant forage. The high efficiency of its C4 metabolism also makes it a resource for biofuel production, as attested by the recent development of BioGas in Germany (Herrmann and Rath 2012; Rath et al. 2013). In Europe, maize cultivation was adopted on a broad scale rapidly after the discovery of America (Rebourg et al. 2003) and a dramatic evolution of varieties occurred with the development of hybrids after World War 2. Dent lines from Northern American origin proved at that time to be highly complementary with Flint lines from European origin to combine productivity and environmental adaptation features for maize cultivation in Central and Northern Europe. These Flint × Dent hybrid varieties have proven to be extremely successful for both grain and silage production. Subsequent reciprocal selection of the two groups increased their differentiation and complementarity. However, potential of this material for biomass production remains poorly documented. Biomass quantitative trait loci (QTL) were detected in biparental crosses (Barrière et al. 2001, 2010), but to our knowledge there was no association genetics study for biomass yield on more diverse material. It is, therefore, of high interest to investigate the variability of this trait and the underlying genetic determinism within these two groups.

Panels of highly diverse materials have proven to be most useful to investigate the organization of diversity available for breeding at phenotypic and genotypic levels. The high density of molecular markers now available for many species makes it possible to discover major genes involved in the variation of traits of agronomic interest using Genome Wide Association Studies (GWAS) (Ozaki et al. 2002; Beló et al. 2007; Jones et al. 2008). Highly diverse panels have accumulated numerous historical recombination events, leading to a limited extent of linkage disequilibrium (LD), which is favorable to fine-map QTL. However, LD in association mapping panels is not only due to genetic linkage, but can also be caused by population structure, relatedness, drift, and selection (Jannink and Walsh 2003; Flint-Garcia et al. 2003). The contribution of these factors relative to linkage can be evaluated statistically (Mangin et al. 2012) and proved for instance to be substantial in grapevine and maize (Mangin et al. 2012; Bouchet et al. 2013). This component of LD due to population structure and relatedness can generate false positives and has thus to be taken into account in association mapping models to control false positives (Ewens and Spielman 1995; Thornsberry et al. 2001). Once these effects are correctly modeled, only marker-trait associations due to linkage should be detected. Efficient softwares were developed to infer population structure using genotypic data (Pritchard et al. 2000; Alexander et al. 2009), and several estimators of relatedness between individuals are available (VanRaden 2008; Astle and Balding 2009; Rincent et al. 2014). The estimated admixture (Q) and kinship (K) matrices can be included in the GWAS statistical model to control false positive efficiently (Yu et al. 2006).

The present work was conducted within the European Cornfed project (Rincent et al. 2012). Its objectives were to (1) investigate genotypic diversity in European and American Dent and Flint inbred lines (2) evaluate phenotypic variability within these two groups for traits related to biomass and flowering time, and (3) detect QTL for these traits by association mapping. For this, original Dent and Flint panels representing different periods of European maize breeding were assembled, characterized with the 50 k SNP array (Ganal et al. 2011) and evaluated per se and as hybrids using a tester line representative of the complementary group in a European field trial network. Association mapping was conducted using the approach recently developed by Rincent et al. (2014) to limit confounding between the tested marker effect and the random polygenic effect.

C. Giauffret
INRA/Université des Sciences et Technologies de Lille, UMR1281, Stress Abiotiques et Différenciation des Végétaux Cultivés, Estrées-Mons, B.P. 136, 80203 Péronne Cedex, France

J. Laborde
INRA Stn Expt Mais, 40590 St Martin De Hinx, France

Materials and methods

Genetic material and genotyping data

Within the "CornFed" project, we developed two new specific Dent and Flint panels (CF-Dent and CF-Flint) aiming at analyzing more precisely the two genetic groups of interest for maize hybrid breeding in Northern Europe, as briefly described in a methodological context by Rincent et al. (2012). Both panels are composed of 300 lines aiming at best representing the diversity of these groups and different generations of genetic materials. These include the first commercially used inbred lines created from open pollinated varieties (OPVs), further referred to as first cycle lines, and more recent lines developed by public institutes or, in the case of the CF-Dent panel, private companies. The CF-Dent panel (see list in Table S1) includes 124 lines from the C-K panel (Camus-Kulandaivelu et al. 2006) determined as belonging to the "Corn Belt Dent" and "Stiff Stalk" groups with an admixture coefficient above 0.5. These were complemented by 58 from the University of Hohenheim (Stuttgart, Germany), 25 from the Misión Biológica de Galicia and the Estación Experimental de Aula Dei (CSIC, Spain), 12 from Centro Investigacións Agrarias de Mabegondo (CIAM, Spain), 58 from the ex plant variety protection (ex-PVP) lines (Mikel 2006; Nelson et al. 2008), and 23 recent lines from Institut National de la Recherche Agronomique (INRA, France). Similarly, the Flint panel (CF-Flint, see list in table S2) includes 118 lines of the C-K panel determined as belonging to the European Flint and Northern Flint groups with an admixture coefficient above 0.5. These were complemented by lines derived from breeding programs of the following institutes: 70 from the University of Hohenheim (Riedelsheimer et al. 2012), 56 from CSIC, 23 from CIAM, 23 from the Eidgenössische Technische Hochschule Zürich (ETHZ, Switzerland), and 10 recent lines from INRA. Four lines (FP1, C105, F816 and EM1027) attributed by STRUCTURE to both Dent and Flint groups with probabilities close to 0.5 in Camus-Kulandaivelu et al. (2006) were assigned to both CF-Dent and CF-Flint panels.

These panels were genotyped with the Illumina MaizeSNP50 BeadChip described in Ganal et al. (2011), as presented in Rincent et al. (2012). Individuals which had marker missing rate and/or heterozygosity higher than 0.1 and 0.05, respectively, were eliminated. Markers which had missing rate and/or average heterozygosity higher than 0.2 and 0.15, respectively, were eliminated from the concerned panel. In each panel, few individuals were highly related. One individual was removed from each pair when the inbreds were identical for more than 98 % of the loci. Three Dent lines and nine Flint lines were eliminated for this reason. Missing genotypes (below 2 % in both panels) were imputed with the software BEAGLE (Browning and Browning 2009). In total 276 and 259 phenotyped individuals passed the genotyping filters for the CF-Dent and CF-Flint panels, respectively (Tables S1 and S2). The filtered markers with a Minor Allele Frequency (MAF) above 0.05 were tested for association (42,214 and 39,076 markers for the CF-Dent and CF-Flint panels, respectively).

Diversity analysis

In the diversity analysis (Q and K estimation), an additional filtering criteria was used to select the SNPs. To reduce the ascertainment bias noted by Ganal et al. (2011), we only used the markers that were developed by comparing the sequences of nested association mapping founder lines (PANZEA SNPs, Gore et al. 2009). In total, 29,418 and 28,513 markers which had a MAF above 0.01 were considered for the diversity analysis in the CF-Dent and CF-Flint lines, respectively. Genotypic data of each panel were organized as G matrices with N rows and L columns, N and L being the panel size and number of SNP loci, respectively. Genotype of individual i at marker l ($G_{i,l}$) was coded as 1, 0.5, or 0 for homozygote for an arbitrarily chosen allele, heterozygote, and the other homozygote, respectively.

Kinship was estimated following Astle and Balding (2009) as follows:

$K\_\mathrm{Freq}_{i,j} = \frac{1}{L}\sum_{l=1}^{L}\frac{(G_{i,l}-p_l)(G_{j,l}-p_l)}{p_l(1-p_l)}$, where p_l is the frequency of the allele coded 1 of PANZEA marker l in the panel of interest; subscripts i and j indicate the lines for which the kinship was estimated. Note that contrary to the Identity By State (IBS, the proportion of shared alleles) estimation, this formula gives a higher weight to loci with a low gene diversity. Also, similarity is higher if two individuals share rare alleles than common alleles.

Admixture was estimated in the CF-Dent and CF-Flint panels using the software ADMIXTURE (Alexander et al. 2009) with a number of groups varying from 2 to 8. This software is based on the same statistical model as STRUCTURE (Pritchard et al. 2000; Falush et al. 2003) but uses a fast numerical optimization algorithm, which permits to considerably reduce computational time. The groups identified by the software were interpreted using the available pedigree information. Differentiation among genetic groups (*Fst*, Nei 1973) was estimated at each locus using the R-package r-hierfstat (Goudet 2005) for each number of groups Q (from 2 to 8), using the individuals attributed to one subgroup with a probability above 0.7 (these individuals are then considered as representative of the corresponding subgroup). Gene diversity (Expected heterozygosity, He; Nei 1978) was also estimated at each marker as $2p_l(1 - p_l)$. The parameters *Fst* and *He* were averaged

on all the markers to characterize the panels more globally. A principal coordinate analysis (PCoA) was performed on the genetic distance matrices (Gower 1966), estimated as $1_{N,N} - \text{K\_Freq}$, where $1_{N,N}$ is a matrix of ones of the same size as K_Freq. We also represented each panel by a network, in which two individuals were linked when their relationship coefficient was above 0.2, unlinked otherwise. For this, the genomic relationship matrix was transformed in a matrix of booleans indicating if the coefficients were above 0.2 or not. These networks were drawn with a Fruchterman and Reingold's force-directed algorithm (Fruchterman and Reingold 1991) with the package « network » in R 3.0.0 (R development Core Team 2013).

Linkage disequilibrium (LD)

To estimate the minimum number of markers needed to cover the genome for GWAS, we estimated intra-chromosomic LD using all the markers. LD was first estimated as the squared correlation between the allelic doses at two markers (denoted by r^2) located on the same chromosome (Hill and Robertson 1968). As kinship has to be taken into account in the GWAS model to control false positives, we need to take it into account to estimate the number of markers required to cover the genome. For this reason, the approach of Mangin et al. (2012) was used to correct for kinship and estimate the part of LD only due to linkage (r^2K). To visualize major trends of LD variation along each chromosome, r^2K was averaged along the genome using a sliding window of 4 Mbp. This was represented on a graph together with marker diversity (*He*) and differentiation (*Fst*) after adjusting cubic smoothing splines along the genome using the R function smooth.spline (Hastie and Tibshirani 1990).

Genetic distances between loci were taken from the map of Ganal et al. (2011) based on the cross F2 × F252. Unmapped markers were positioned according to the local ratio between physical and genetic distances. The variation of LD with the genetic distances on each chromosome was adjusted to the model of Hill and Weir (1988), using only the pairs of markers separated by <4 cM. We estimated the LD decay for each chromosome as the abscissa of the intersection between the fitted curve and the horizontal line $y = 0.1$. Knowing the length of each chromosome (in cM), we could estimate the minimum number of markers required on each chromosome to get an average r^2 or r^2K of 0.1 between each pair of adjacent markers.

Phenotypic data

The Flint and Dent lines were, respectively, crossed to a Dent (F353) and a Flint (UH007) tester to produce hybrid progenies for phenotypic evaluation. These two lines were representative of advanced materials within their respective groups. The two hybrid panels were evaluated for flowering and biomass production-related traits in trials located in Germany, France, and Spain. Two separate experiments were conducted for the Dent and Flint hybrids, with five locations in 2010, and six (CF-Dent) and five (CF-Flint) locations in 2011 (see Table S3 for geographical coordinates). Within each panel, the hybrids were divided into two groups of precocity and each group was evaluated in a different block. Each block was composed of 9 sub-blocks of 20 plots. Within the 20 genotypes of a given sub-block, 4 were repeated in 4 other sub-blocks of the same block, and 2 were repeated in 2 sub-blocks of the other block to estimate experimental error, block and sub-block effects. Male and female flowering time, plant height (PLHT, cm), dry matter content at harvest (DMC, %), and dry matter yield (DMY, Mg/ha) were recorded for each plot. Male and female flowering times were converted into growing degree days, considering a base temperature of 6 °C, using the mean daily air temperature measured at each location (these measures were, respectively, denoted by Tass_GDD6 for male flowering and Silk_GDD6 for female flowering). The Anthesis to Silking Interval (ASI_GDD6) was obtained by subtracting Tass_GDD6 from Silk_GDD6. DMC and DMY were observed at only nine of the ten trials for the CF-Flint panel. DMC and DMY were corrected by flowering precocity (DMCcorr and DMYcorr) by regressing the raw data on Silk_GDD6 for each block for DMC or for each trial for DMY.

$$\text{DMC}_{ijkl} = \mu_{jk} + \alpha_{jk} \times \text{Silk\_GDD6}_{ijkl} + E_{ijkl} \text{ and}$$
$$\text{DMCcorr}_{ijkl} = \hat{E}_{ijkl}$$
$$\text{DMY}_{ijkl} = \mu_{j} + \alpha_{j} \times \text{Silk\_GDD6}_{ijkl} + E_{ijkl} \text{ and}$$
$$\text{DMYcorr}_{ijkl} = \hat{E}_{ijkl}$$

with subscripts i, j, k and l the indices indicating, respectively, the genotype, the trial, the block, and the repetition in the block; μ_{jk} and μ_j indicate the intercepts for block jk and trial j, respectively; α_{jk} and α_j are the block and trial-specific regression coefficients on silking for DMC and DMY, respectively.

Outlier plots with extreme phenotypes showing a discontinuity from the rest of the distribution were excluded from the study (<2.5 % of the observations were removed in both panels). Least-squares means of genotypes over the global network were calculated with the GLM procedure (SAS 9.3, SAS Institute 2011) by adjusting for block and trial effects, except for DMY adjusted means, which were not corrected by block effects. Such a correction would indeed rely on the performances of the genotypes common to the two blocks, which are likely to be affected by competition effects (early genotypes being penalized in the "late" block and late genotypes favored in the "early block"). All effects

were considered fixed at this step. Considering the important difference of residual variance among trials, we took heteroscedasticity into account by estimating a residual variance for each trial. For all the traits except DMY and DMYcorr, weighted least squares were computed using the following model:

$Y_{ijkl} = \mu + G_i + T_j + T(B)_{jk} + E_{ijkl}$, with E_{ijkl}~ N(0,σ_j^2), and for DMY and DMYcorr with the model: $Y_{ijkl} = \mu + G_i + T_j + E_{ijkl}$, with E_{ijkl}~ N(0,σ_j^2), where Y_{ijkl} is the phenotype of the repetition l of genotype i in block k of trial j; μ is the global mean; G_i is the fixed genotype effect of individual i; T_j is the effect of trial j; and $T(B)_{jk}$ is the effect of block k within trial j. Other field effects (sub-block, row and column effects) were not included in the model as they were not significant.

Trait heritability was estimated at the level of the experimental design. For traits other than DMY and DMYcorr, variance components of heritability were estimated in two steps. In a first step, genotypes were considered as fixed effect to get block effect estimates based only on the repetitions across blocks.

$$Y_{ijkl} = \mu + G_i + T_j + B_{k(j)} + \mathbf{E}_{ijk}$$

In a second step, phenotypes were corrected by block effects and were analyzed considering genotype and genotype x trial effects as random:

$$Y_{ijkl} - \widehat{B_{k(j)}} = \mu + \mathbf{G}_i + T_j + \mathbf{G} \times \mathbf{T}_{ij} + \mathbf{E}_{ijk},$$

where $\mathbf{GxT}_{ij}$ is the random interaction effect between genotype i and trial j. For DMY and DMYcorr, variance components of heritability were estimated in one step only to prevent confounding block effects with competition between early and late lines:

$$Y_{ijkl} = \mu + \mathbf{G}_i + T_j + \mathbf{G} \times \mathbf{T}_{ij} + \mathbf{E}_{ijk}$$

Heritabilities were then estimated as $h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2/r + \sigma_{g\times t}^2/L}$, where σ_g^2, σ_e^2 and $\sigma_{g\times t}^2$ are the variance estimates of the random effects $\mathbf{G}_i$, $\mathbf{E}_{ijk}$ and $\mathbf{G} \times \mathbf{T}_{ij}$, respectively; L is the mean number of environments; and r is the average total number of repetitions per genotype across environments. We also computed adjusted means and heritabilities for each trial by simplifying accordingly the above-described statistical models.

The lines per se were also evaluated at one trial for Tass_GDD6, Silk_GDD6 (Dent and Flint lines), and PLHT (only Flint lines). The Dent and Flint lines per se were evaluated in 2011 in Saint-Martin de Hinx and Gif-sur-Yvette (France), respectively. Least-squares genotype means of per se lines were calculated with the GLM procedure by adjusting for block effects. Variances of the per se experiment were estimated with the same mixed model used to estimate heritabilities at the trial level in the hybrid experiments. Analyses were conducted using SAS 9.3 (SAS Institute 2011).

Phenotypic characterization of the genetic groups within each panel

Genetic groups defined by admixture were compared within each panel for their phenotypic performance by estimating the genetic average of each group (denoted by μ_q) using the following model:$Y_i = \sum_{q=1}^{N_Q} \mu_q F_{i,q} + \mathbf{E}_i$, where Y_i is the adjusted mean of individual i, $F_{i,q}$ is the admixture coefficient of individual i in group q, and N_Q is the number of groups. $N_Q = 8$ was considered for both panels to characterize the groups at the finer level that is informative to breeders.

Statistical model for association mapping

Mixed models are classically used to detect QTL while controlling false positive rate in GWAS (Yu et al. 2006). Relatedness among individuals is taken into account by considering that the random polygenic effects are not independent, with a covariance matrix determined by K. A fixed structure effect (associated to a structure matrix Q) can also be included if the dataset is highly structured. Comparison of p values obtained with different $(Q + K)$ models revealed that K was sufficient to control both structure and relatedness (Fig. S1).

We tested each SNP with a MAF above 0.05 (42,214 and 39,076 SNPs in the CF-Dent and CF-Flint panels, respectively) in the following model: $Y = X\beta + U + E$, where Y is the vector of phenotypes (adjusted means of the per se performances, or of the hybrid performances at one trial or in the whole trial network); X includes a vector of 1 and the genotypes at the tested marker (coded as 0, 0.5, or 1 as mentioned above); β includes the intercept and the additive effect of the tested marker (β_l), defined as the difference between the two homozygous genotypes; $U \rightsquigarrow N(0, K.\sigma_{gl}^2)$ is the vector of random polygenic effects, K being the kinship estimate and σ_{gl}^2 the residual polygenic variance; $E \rightsquigarrow N(0, I.\sigma_e^2)$ is the vector of remaining residual effects with variance σ_e^2; I is an identity matrix of size equal to the number of individuals (N), U and E are supposed to be independent. We used two different estimates of K in the model: K_Freq as presented above, and K_Chr (Rincent et al. 2014) which is computed only with the markers physically unlinked to the tested SNP: $\text{K\_Chr}_{i,j,c} = \frac{1}{L_{-c}} \sum_{l \notin c} \frac{(G_{i,l} - p_l)(G_{j,l} - p_l)}{p_l(1-p_l)}$, where c is the considered chromosome, L_{-c} is the number of markers not located on chromosome c. This second estimator was developed to take into account the fact that including

markers in high LD with the tested SNP in the kinship estimation decreased power (Listgarten et al. 2012; Rincent et al. 2014). Each marker was tested for association with the different traits using a Wald test (Wald 1943) in ASReml-R (Gilmour et al. 2009). The scripts were written in R 3.0.0 (R development Core Team 2013). The statistical significance threshold was set to 0.05/Meff, which corresponds to a Bonferroni correction on Meff tests, Meff being the number of independent tests estimated as in Li and Ji (2005). This procedure evaluated 3,638 and 3,527 independent tests in the CF-Dent and CF-Flint panels, respectively, which led to a $-\log_{10}(p \text{ value})$ threshold of 4.9 in both panels. Significant SNPs separated by <100 kb were considered as a single QTL for interpretation of the results. QTL size was estimated as the proportion of variance explained by the most significant SNP (Bouchet et al. 2013) of the QTL region after removing the structure effect (admixture being estimated with ADMIXTURE for $N_Q = 6$).

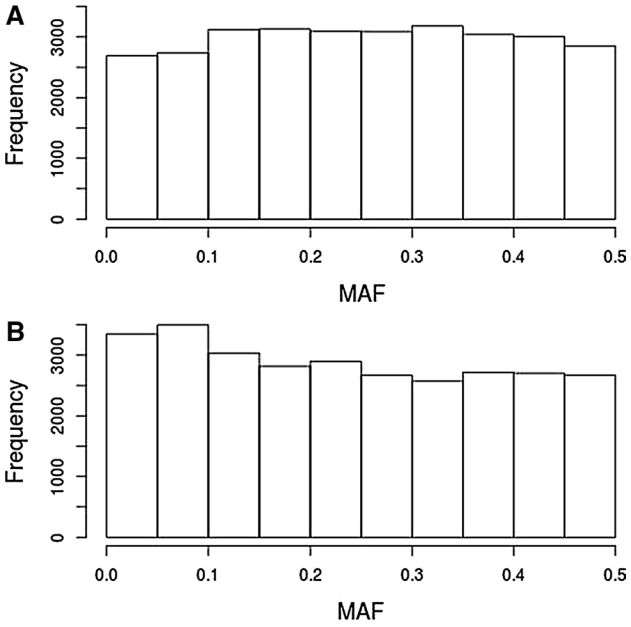

Fig. 1 Histograms of the minor allele frequencies of the polymorphic PANZEA markers in the CF-Dent (**a**) and CF-Flint (**b**) panels

Results

Diversity and structure analysis

The histograms of the minor allele frequencies (MAF) of the polymorphic PANZEA markers showed a slight deficit in rare alleles in the CF-Dent panel and a slight excess in the CF-Flint panel, compared to a uniform distribution (Fig. 1). This is consistent with typical flint lines (assignment to a given group above 0.95 at the structure level $N_Q = 8$) showing a higher proportion of monomorphic PANZEA markers compared to dent lines (18 and 15 %, respectively; data not shown). MAF was on average slightly higher in the CF-Dent (0.25) than in the CF-Flint panel (0.24), which resulted in a lower index of gene diversity ($\overline{He}$) in the CF-Flint than in the CF-Dent panel (0.36 and 0.37, respectively). These differences between CF-Dent and CF-Flint panels for MAF and $\overline{He}$ were highly significant (*t* test). Gene diversity *He* was variable along the genome (Fig. 2), with generally lower values in centromeric regions.

The cross-validation criterion proposed by ADMIXTURE suggested the presence of at least 4 main groups in each of the panels, and the criterion always improved with the number of groups (results not shown). For an expected number of genetic groups ranging from 2 to 8, all the subgroups identified by ADMIXTURE were interpretable in terms of pedigree and/or geographical origins. The genetic groups were composed of lines sharing a common recent ancestor (for example F252 in the CF-Dent panel), or a common ancestral origin (for example Northern Flint in the CF-Flint panel). We noted that groups at level N_Q could generally be related to groups at level $N_Q + 1$ by the subdivision of one subgroup into two (see Fig. S2 for an empirical synthesis). For a given number of groups, the differentiation among groups was higher in the CF-Dent than in the CF-Flint panel (Table 1). The mean *Fst* over the genome increased with the number of groups in both panels, but it reached a plateau at 7 groups in the CF-Flint panel. *Fst* was variable along the genome; in particular, when considering four groups (Fig. 2), peaks of *Fst* were clearly visible in the CF-Dent panel (chromosomes 7 and 10) and in the CF-Flint panel (chromosome 8).

The first two axes of the PCoA explained 16.1 and 15.7 % of the variability in the CF-Dent and CF-Flint panels, respectively (Fig. 3). The different groups identified by ADMIXTURE were clearly identifiable on the PCoA graphs. The first axis separated the Iodent from the non-Iodent lines in the CF-Dent panel and the Northern Flint from the other Flint lines in the CF-Flint panel. Note that extreme positions along the axes were observed for the key founders of these groups (e.g. PH207 for Iodent, B73 for Stiff Stalk, Mo17 for Lancaster, D105 for Northern Flint).

Network representations of the CF-Dent and the CF-Flint panels revealed clusters of related individuals and isolated lines (Fig. 3). The shape of the network was different in the two panels: the CF-Dent panel was composed of isolated lines having no or only limited relatedness to any other line and few clusters of related individuals. The network of the CF-Flint panel also revealed clusters of related individuals but was much looser than the network of the CF-Dent panel. Groups identified with ADMIXTURE at $N_Q = 4$ were in good agreement with the network

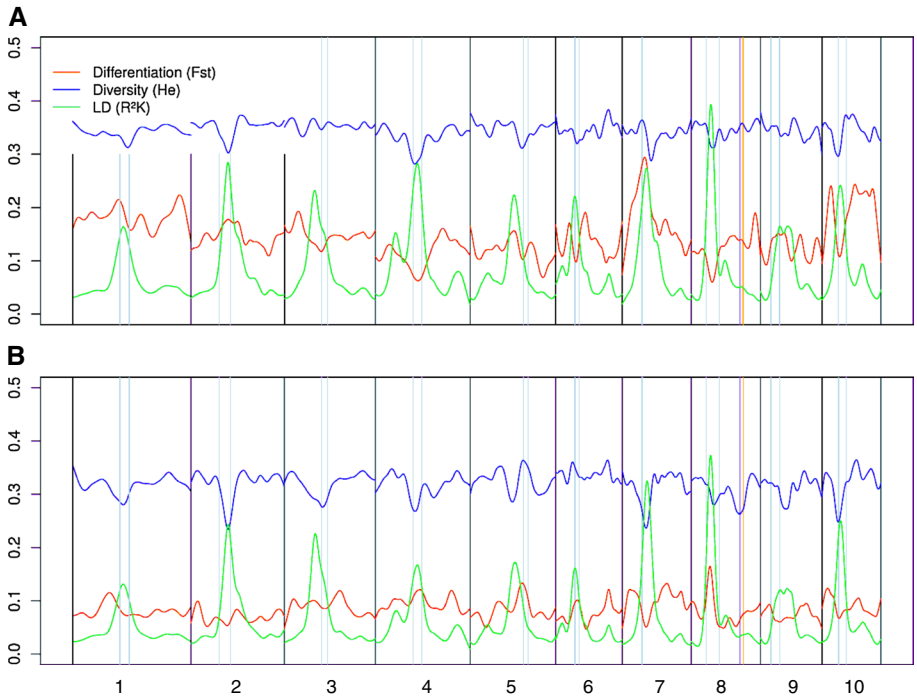

Fig. 2 Differentiation among groups (*Fst*, estimated at Q = 4), gene diversity *(He)* and linkage disequilibrium along the genome (physical distance) in the CF-Dent (**a**) and CF-Flint (**b**) panels. For each parameter a cubic smoothing spline was adjusted along the genome (with a smoothing parameter of 0.9). Centromere limits, *Vgt1,* and *Vgt2* are indicated by *blue*, *pink,* and *purple lines*, respectively

Table 1 Average differentiation index *Fst* among the genetic groups for different numbers of groups (N_Q) varying from 2 to 8

	$N_Q = 2$	$N_Q = 3$	$N_Q = 4$	$N_Q = 5$	$N_Q = 6$	$N_Q = 7$	$N_Q = 8$
Fst							
CF-Dent	0.07	0.15	0.24	0.26	0.30	0.32	0.35
CF-Flint	0.07	0.13	0.15	0.18	0.21	0.23	0.22

Lines were attributed to a given group if their admixture was above a threshold of 0.7

visualization. In each panel, one of the four groups (called "Others" in Fig. 3) was composed of more heterogeneous material including many 1st cycle lines and appeared fragmented in the network.

Linkage disequilibrium

The LD was on average more extended in the CF-Dent than in the CF-Flint panel (0.21 and 0.12 cM to reach an r^2 of 0.1 on average over all chromosomes, respectively, see Table 2). Inter-chromosomal LD was observed in both panels (Fig. S3). When considering physical distances, LD extent was highly variable between chromosomes and along chromosomes (Fig. 2), being more extended in centromeric regions. Taking relatedness into account substantially reduced the extent of LD in both panels, particularly in the CF-Dent panel (Table 2), and considerably reduced inter-chromosomal LD (Fig. S3). For intra-chromosomal LD, the decrease observed when considering relatedness was particularly strong for chromosomes 3 and 8 in both panels, and chromosomes 4 and 7 in the CF-Dent panel only (Table 2). The chromosomes 3, 4, and 8 in the CF-Dent panel had a more extended LD (0.16, 0.18 and 0.18 cM to reach a r^2K of 0.1, respectively) than the others (between 0.09 and 0.13 cM, Table 2). In the CF-Flint panel, all the chromosomes displayed similar r^2K except chromosome 8 for which LD was more extended (0.14 cM to reach a r^2K of 0.1 for chromosome 8, only 0.09–0.10 cM for the other chromosomes). Knowing the length of the chromosomes (in cM), these statistics allowed the estimation of the minimum number of markers required to cover the genome (assuming evenly spaced markers on the genetic map): more markers are needed in the CF-Flint (24,387) than in the CF-Dent panel (19,000) to get a r^2K of 0.1 between evenly spaced adjacent markers (Table 2).

Phenotypic variation

We observed a substantial variability for all the traits in both panels and in both hybrid and per se evaluations

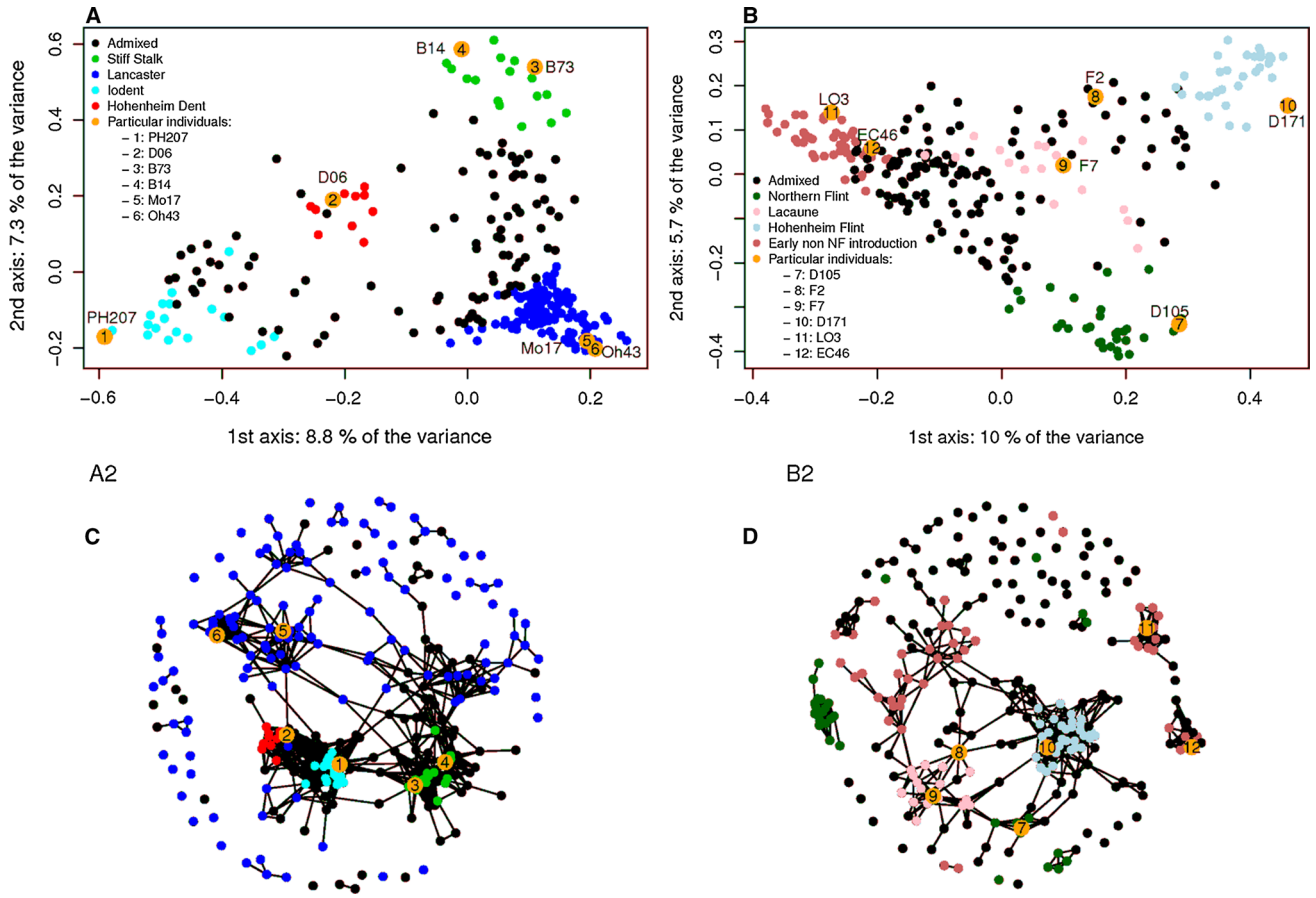

Fig. 3 Principal coordinate analysis (**a**, **b**) and network (**c**, **d**) representations of the CF-Dent (**a**, **c**) and CF-Flint (**b**, **d**) panels. Both representations are based on the covariance matrix K_Freq. The most representative individuals of each subgroup at $N_Q = 4$ (admixture above 0.7) were colored. Few key individuals are indicated in each panel (PH207, D06, B73, B14, Mo17, OH43, respectively, *numbered* from *1* to *6* in CF-Dent and D105, F2, F7, D171, L03, and EC46, respectively, numbered from *7* to *12* in CF-Flint). EC46 is a typical line of the group "Aranga". In the network representation, individuals are linked if their covariance is above 0.2, unlinked otherwise. In these networks, distances are not informative

(Tables 3 and S4), with for instance least-squares means of DMY of the hybrids over the trial network varying between 11 and 20 Mg/ha in both panels. Averages and heritabilities were variable between trials for a given trait (Table S3), but high heritabilities were observed at the trial network level (over 0.73 and 0.65 in the CF-Dent and CF-Flint panels, respectively). For most traits, heritability was higher in the CF-Dent than in the CF-Flint panel. This was related to higher residual variances in the CF-Flint panel. Tass_GDD6 and Silk_GDD6 were the most heritable traits (0.96 and 0.97 in the CF-Dent and CF-Flint panels, respectively). ASI_GDD6 and yield traits (DMC, DMCcorr, DMY and DMYcorr) were the less heritable traits. The lowest heritability was 0.65 for ASI_GDD6 in the CF-Flint panel. The heritabilities of the per se evaluations were close to the heritabilities of the hybrid trial network (Table 4), although inbred lines were evaluated at only one trial. This was due to much higher genetic variances in the per se evaluation (up to 6.4 times higher). The correlations between the hybrid and the per se adjusted means were quite high for Tass_GDD6 and Silk_GDD6 (between 0.68 and 0.87), but lower for ASI_GDD6 (between 0.22 and 0.43). These correlations were higher in the CF-Dent than in the CF-Flint panels for the three traits that were common to both panels (Table 4).

Phenotypic characterization of the genetic groups within each panel

For hybrid performances, we observed differences between the genetic groups identified within the two panels. Adjusted R^2 explained by the groups were between 0.11 and 0.47 in CF-Dent and between 0.05 and 0.41 in CF-Flint when considering 8 groups (Table 5). In the CF-Dent panel, the lines related to D06 or to F252 displayed the earliest flowering time and the highest DMC and DMCcorr (Table 5). The

Table 2 Extent of linkage disequilibrium and number of markers needed to reach an average r^2 or r^2 K of 0.1 for each chromosome (Chrom) in the two panels (CF-Dent and CF-Flint)

Chrom.	CF-Dent				CF-Flint			
	r^2		r^2K		r^2		r^2K	
	r^2 extent (cM)[a]	N markers[b]	r^2K extent (cM)[a]	N markers[b]	r^2 extent (cM)[a]	N markers[b]	r^2K extent (cM)[a]	N markers[b]
1	0.12	2,740	0.09	3,605	0.09	3,636	0.09	3,841
2	0.10	2,461	0.09	2,715	0.09	2,702	0.09	2,838
3	0.32	786	0.16	1,572	0.16	1,441	0.10	2,578
4	0.27	853	0.18	1,299	0.10	2,209	0.09	2,497
5	0.20	1,179	0.13	1,749	0.10	2,390	0.09	2,526
6	0.20	968	0.14	1,332	0.10	1,924	0.09	2,037
7	0.28	741	0.11	1,877	0.10	2,086	0.09	2,299
8	0.25	937	0.18	1,349	0.23	1,008	0.14	1,736
9	0.19	993	0.11	1,725	0.10	1,924	0.09	2,113
10	0.19	892	0.10	1,778	0.10	1,761	0.09	1,923
Total		12,552		19,000		21,081		24,387

r^2 is the squared correlation between markers and r^2K is r^2 corrected by kinship

The genetic position of the markers was derived from the genetic map LHRE Ganal et al. (2011). r^2 and r^2K were calculated with the R package LDcorSV

[a] genetic distance (in cM) to reach r^2 or r^2K equal to 0.1, after fitting the Hill and Weir (1988) model

[b] Number of markers required to reach an average r^2 of 0.1 between adjacent markers

Table 3 Variances in the hybrid experimental design at the network level

	Tass_GDD6	Silk_GDD6	ASI_GDD6	PLHT	DMC	DMCcorr	DMY	DMYcorr
CF-Dent panel								
Trial network								
$\hat{\sigma}^2_g$	1,322.3	1,515.0	68.9	133.1	10.3	2.5	2.0	1.5
$\hat{\sigma}^2_{g\times t}$	295.7	436.9	87.0	51.7	3.6	2.6	1.9	1.6
$\hat{\sigma}^2_e$	324.5	375.0	218.7	129.7	5.1	5.1	3.3	3.5
Nb of trials	11	11	11	10	11	11	11	11
h^2	0.96	0.96	0.73	0.89	0.93	0.80	0.82	0.78
CF-Flint panel								
Trial network								
$\hat{\sigma}^2_g$	1,623.5	1,558.6	53.5	193.9	6.8	2.2	1.9	1.4
$\hat{\sigma}^2_{g\times t}$	181.2	143.5	74.5	116.0	4.5	3.3	1.8	1.5
$\hat{\sigma}^2_e$	345.4	363.7	218.9	196.3	6.5	6.6	3.8	4.2
Nb of trials	10	10	9	9	9	9	9	9
h^2	0.97	0.97	0.65	0.86	0.86	0.69	0.76	0.71

The different traits are male (Tass_GDD6), female (Silk_GDD6) flowering times, Anthesis to Silking interval (ASI_GDD6) expressed in growing degree days considering a base temperature of 6 °C, plant height (PLHT, cm), dry matter content (DMC, %), and dry matter yield (DMY, Mg/ha). DMCcorr and DMYcorr are DMC and DMY corrected by Silk_GDD6. The genetic variance, genotype × trial variance, residual variance and heritability estimates are denoted by $\hat{\sigma}^2_g$, $\hat{\sigma}^2_{g\times t}$, $\hat{\sigma}^2_e$ and h^2, respectively. All variances were highly significant in both panels

Lancaster and Stiff Stalk groups displayed the latest flowering time and were also the most productive (DMY of up to 17.6 Mg/ha). In the CF-Flint panel, the Lacaune descendants (Fv7 related), the Northern Flints, and the Hohenheim Flints displayed the earliest flowering time and the highest DMC and DMCcorr. Groups from Southern Europe (related to CIAM Aranga and descendants from Italian open pollinated varieties (OPV) or from other non-Northern Flints introductions into Europe) displayed the latest flowering. The lines related to CIAM Aranga, to UH006 or to Fv7 (Lacaune)

Table 4 Variances in the line per se experimental design (one trial), and correlation between the line per se and the hybrid adjusted means (Corr Hyb/PerSe)

	CF-Dent			CF-Flint			
	Tass_GDD6	Silk_GDD6	ASI_GDD6	Tass_GDD6	Silk_GDD6	ASI_GDD6	PLHT
$\hat{\sigma}_g^2$	7,382	8,361	433	10,440	8,666	653	728.3
$\hat{\sigma}_e^2$	604	429	223	1,186	461	1,118	39.4
h^2	0.93	0.96	0.68	0.91	0.96	0.40	0.96
Corr Hyb/PerSe[a]	0.85	0.87	0.43	0.68	0.77	0.22	0.58

The genetic variance, residual variance and heritability estimates are denoted by $\hat{\sigma}_g^2$, $\hat{\sigma}_e^2$ and h^2, respectively

[a] Correlation between hybrid and per se adjusted means

Table 5 Characterization of the different genetic groups at $N_Q = 8$ in the CF-Dent and CF-Flint panels

Panel	Genetic groups	Frequency	Tass_GDD6	Silk_GDD6	ASI_GDD6	DMC	DMY	PLHT	DMCcorr	DMYcorr
CF-Dent	Stiff Stalk (B73 type)	0.07	906	916	10	31.9	17.5	267	−1.2	1.4
	Lancaster (MO17 type)	0.09	940	963	21	30.1	17.6	273	−1.0	1.0
	D06 family (mostly Iodent at K = 3)	0.09	854	866	12	38.2	16.1	253	2.0	0.5
	Iodent (PH207 type)	0.15	870	887	18	36.1	16.1	252	1.0	0.3
	Stiff stalk (B14 type)	0.12	913	927	13	33.0	17.3	263	0.2	1.1
	Minnesota13 (Wf9, A3 type)	0.27	890	916	25	32.6	15.0	253	−0.9	−1.1
	Lancaster (OH43 type)	0.09	903	920	18	31.4	16.1	254	−1.5	−0.1
	F252 family	0.11	840	853	15	39.1	14.7	241	2.1	−0.8
	Adj. R^{2a}		0.33	0.33	0.13	0.47	0.23	0.20	0.34	0.23
	LSD[b]		0.6	0.7	0.6	0.1	0.1	0.5	0.1	0.1
CF-Flint	Hohenheim Flint (D171 type)	0.13	848	876	21	33.7	14.7	247	1.1	0.1
	UH_006 family	0.10	867	893	20	32.3	15.2	257	0.3	0.3
	Lacaune (Fv7 type)	0.11	843	874	24	33.5	15.4	239	0.8	0.7
	CIAM Aranga and EC18 related	0.06	899	931	22	30.6	16.6	258	0.0	1.3
	Descendants from italian OPVs (numerous 1st cycles)	0.09	912	942	21	29.9	14.9	257	−0.3	−0.5
	Descendants from non-NF introductions in Europe (Spanish and others)	0.17	952	984	21	28.2	15.8	270	−0.4	−0.2
	Pyrenean (Numerous 1st cycle)	0.16	876	908	26	30.7	14.7	250	−0.8	−0.3
	NF (numerous 1st cycle)	0.18	855	891	27	32.2	14.5	253	0.0	−0.4
	Adj. R^{2a}		0.38	0.41	0.05	0.27	0.06	0.12	0.06	0.07
	LSD[b]		0.9	0.9	0.7	0.1	0.1	0.6	0.1	0.1

The group means of each trait were obtained by regressing the adjusted means on the admixture coefficients. The different traits are male (Tass_GDD6), female (Silk_GDD6) flowering times, anthesis to silking interval (ASI_GDD6) expressed in growing degree days with a base temperature of 6 °C, plant height (PLHT, cm), dry matter content (DMC, %), and dry matter yield (DMY, Mg/ha). DMCcorr and DMYcorr are the DMC and DMY corrected by Silk_GDD6

[a] Adjusted R^2 of the regression on the admixture at $N_Q = 8$

[b] Fisher least significant difference for the comparison of the two smallest groups

were the most productive when crossed to the Dent tester (DMY of up to 16.6 Mg/ha). Despite the negative correlation between flowering precocity and productivity in both panels (results not shown), we could observe different levels of productivity at a given flowering time in some cases. For example, lines related to B73 and those related to OH43 both displayed late flowering but the first group was more productive. In the CF-Flint panel, the group "CIAM Aranga and EC18 related" was by far the most productive, although earlier than other groups. We observed that the three Flint groups which had the highest contribution in first cycle lines, namely Italian OPVs, Pyrenean and Northern Flints, were the least productive, with DMY below 15 Mg/ha (Table 5). A similar trend was found for Dents, with most first cycle lines grouped in the "Minnesota13" group, which displayed the lowest value for DMYcorr. We also noted substantial

variation within genetic groups (see, for example, Iodent and Italian OPVs in Tables S1 and S2, respectively), consistent with the limited proportion of variance explained by admixture for all the traits. Within a given group, the most typical lines (admixture above 0.98) could differ by up to 5 Mg/ha (e.g. non-admixed individuals of the "UH006 family" group ranged from 12.7 to 17.7 Mg/ha, Table S2). A formal analysis of genetic gain over breeding generations could not be conducted due to the complexity of the pedigrees but some interesting trends could be noted. For instance, within the PH207 group, most lines derived from the founder PH207 appear superior to it in terms of performance (Table S1).

Association mapping results

The complete lists of significant SNPs are presented in Tables S5 and S6, and the most significant associations (−log(p value) above 5) are summarized in Tables 6 and 7. The highest −log(p value) was 9.98 on chromosome 8 in CF-Dent and 6.71 on chromosome 1 in the CF-Flint panel, corresponding both to associations with flowering traits (Tass_GDD6 or Silk_GDD6).

Regarding the two statistical methods which were used, both kinship estimators (K_Freq and K_Chr) appeared efficient to control false positive rate, as revealed by QQ-plots (Fig. S1). At the chosen Bonferroni threshold, the kinship estimator K_Chr permitted the discovery of more SNPs than K_Freq for all the traits in both panels, at the trial or at the network level, except DMYcorr in the CF-Flint panel (Table 8). K_Chr permitted the discovery of 62 additional SNPs in the CF-Dent panel, and 15 in the CF-Flint panel (11 and 7 at the network level, corresponding to an increase of 41 and 39 % in the CF-Dent and CF-Flint panels, respectively). Only 1 and 3 SNPs were identified with K_Freq but not with K_Chr in the CF-Dent and CF-Flint panels, respectively.

Table 6 Most significant associations in the CF-Dent panel at the network level

Trait	Chr	Pos	MAF[a]	−log_K_Freq[b]	−log_K_Chr[c]	R2[d]	Rare allele effect[e]	Rare allele[f]	Closest gene	Gene descr.
Tass_GDD6	2	178262299	0.22	7.07	7.22	0.15	11.68	C	GRMZM2G098828	ATP binding
Tass_GDD6	3	150832948	0.48	5.11	5.23	0.10	−8.86	T	GRMZM2G082387	Transcr. factor
Tass_GDD6	4	233828118	0.47	5.78	6.29	0.07	8.36	T	GRMZM2G064023	Citrate synthase
Tass_GDD6	7	122130497	0.05	5.71	5.79	0.08	17.30	A	GRMZM2G075348	Uncharacterized
Tass_GDD6	8	115446396	0.14	5.37	5.93	0.08	11.58	A	GRMZM2G111396	Unknown
Tass_GDD6	8	118188472	0.39	5.36	5.76	0.07	7.80	A	GRMZM2G047842	Uncharacterized
Tass_GDD6	8	123506141	0.27	8.81	9.98	0.11	11.65	C	GRMZM2G179264	ZCN8 protein
Tass_GDD6	8	126077120	0.37	4.56	5.74	0.02	6.88	A	GRMZM2G380515	Zinc ion bind
Tass_GDD6	8	126287026	0.36	3.88	5.00	0.01	6.41	C	GRMZM2G118834	Uncharacterized
Silk_GDD6	2	178262299	0.22	7.42	7.57	0.14	12.97	C	GRMZM2G098828	ATP binding
Silk_GDD6	7	122130497	0.05	5.22	5.31	0.09	17.80	A	GRMZM2G075348	Uncharacterized
Silk_GDD6	8	115446396	0.14	5.42	5.90	0.08	12.59	A	GRMZM2G111396	Unknown
Silk_GDD6	8	123506141	0.27	7.83	8.86	0.11	11.97	C	GRMZM2G179264	ZCN8 protein
PLHT	2	178262299	0.22	4.88	5.10	0.10	4.11	C	GRMZM2G098828	ATP binding
PLHT	2	186447969	0.14	4.59	5.19	0.06	4.28	C	GRMZM2G381059	Protein binding
DMYcorr	5	190732112	0.19	6.00	6.07	0.14	0.49	C	GRMZM2G031952	Cytoskeleton
DMY	5	190732112	0.19	6.54	6.70	0.15	0.56	C	GRMZM2G031952	Cytoskeleton
DMC	3	150832948	0.48	5.26	5.41	0.10	0.74	C	GRMZM2G082387	Transcr. factor
DMC	10	31219126	0.11	4.95	5.35	0.09	−1.05	T	AC189796.3	Unknown

The different traits are male (Tass_GDD6), female (Silk_GDD6) flowering times, Anthesis To Silking Interval (ASI_GDD6) expressed in growing degree days with a base temperature of 6 °C, plant height (PLHT, cm), dry matter content (DMC, %), and dry matter yield (DMY, Mg/ha). DMCcorr and DMYcorr are the DMC and DMY corrected by Silk_GDD6

[a] Minor allele frequency

[b] −log(p value) with K_Freq

[c] −log(p value) with K_Chr

[d] Proportion of variance explained by the SNP after removing structure effect (admixture for $N_Q = 6$)

[e] Effect of the rare allele (half difference between the two homozygotes)

[f] Rare allele at the SNP (forward)

Table 7 Most significant associations in the CF-Flint panel at the network level

Trait	Chr	Pos	MAF[a]	−log_K_Freq[b]	−log_K_Chr[c]	R2[d]	Rare allele effect[e]	Rare allele[f]	Closest gene	Gene descr.
Tass_GDD6	1	53414468	0.24	5.37	5.75	0.10	12.14	A	GRMZM2G031001	DNA binding
Silk_GDD6	1	53414468	0.24	6.15	6.72	0.12	12.41	A	GRMZM2G031001	DNA binding
Silk_GDD6	1	300441295	0.36	5.10	5.44	0.03	−8.52	C	GRMZM2G377487	Unknown
PLHT	1	53414468	0.24	4.63	5.05	0.08	5.30	C	GRMZM2G031001	DNA binding
PLHT	1	153344342	0.25	4.62	5.64	0.07	5.11	T	GRMZM2G422631	Cell wall modification
PLHT	1	154077833	0.17	5.31	6.10	0.08	6.03	G	GRMZM2G056039	Heat shock protein
PLHT	8	84808001	0.06	4.83	5.03	0.06	8.79	A	GRMZM2G128809	RNA binding
PLHT	8	101237704	0.14	5.96	6.12	0.08	6.43	A	GRMZM2G055667	Fatty ac. biosynth. process
PLHT	9	119310870	0.13	5.88	5.78	0.11	−7.25	A	GRMZM2G098179	Response to freezing
DMYcorr	1	17966974	0.23	5.60	5.76	0.11	−0.52	C	GRMZM2G059102	Transcription factor
DMYcorr	1	154077833	0.17	5.20	5.45	0.07	0.55	G	GRMZM2G056039	Heat shock protein
DMY	1	153344342	0.25	4.91	5.39	0.07	0.53	T	GRMZM2G422631	Cell wall modification
DMY	1	154077833	0.17	6.00	6.42	0.08	0.65	G	GRMZM2G056039	Heat shock protein
DMC	4	152972399	0.17	5.30	5.33	0.08	−0.97	A	GRMZM2G406313	Cortical cell delineating
ASI_GDD6	7	32478358	0.08	5.40	5.68	0.09	−4.56	G	GRMZM2G472146	Signaling pathway
ASI_GDD6	7	99894530	0.24	4.79	5.09	0.07	−2.77	A	GRMZM2G166692	Unknown

The different traits are male (Tass_GDD6), female (Silk_GDD6) flowering times, anthesis to silking interval (ASI_GDD6) expressed in growing degree days with a base temperature of 6 °C, plant height (PLHT, cm), dry matter content (DMC, %), and dry matter yield (DMY, Mg/ha). DMCcorr and DMYcorr are the DMC and DMY corrected by Silk_GDD6

[a] Minor allele frequency

[b] −log(*p* value) with K_Freq

[c] −log(*p* value) with K_Chr

[d] Proportion of variance explained by the SNP after removing structure effect (admixture for $N_Q = 6$)

[e] Effect of the rare allele (half difference between the two homozygotes)

[f] Rare allele at the SNP (forward)

Comparing the two panels, the total number of SNPs (Table 8) significant in at least one environment or at the network level with one of the two methods (K_Freq or K_Chr) was more than two times higher in the CF-Dent (258 SNPs) than in the CF-Flint panel (116 SNPs). This difference was less pronounced when considering regions (QTL) instead of SNPs (173 and 108 QTL identified in the CF-Dent and CF-Flint panels, respectively). The only exception to this global trend was PLHT, for which more QTL were discovered in the CF-Flint than in the CF-Dent panel. The QTL identified in these panels explained between 1 and 15 % and between 3 and 12 % of the phenotypic variance that was not explained by the admixture groups in the CF-Dent and CF-Flint panels, respectively.

Considering traits, more SNPs were discovered for DMY, DMYcorr, Tass_GDD6. and Silk_GDD6 than for DMC, DMCcorr, and ASI_GDD6 (Table 8). However, most of the DMY and DMYcorr SNPs (96–100 %) were significant in only one environment, whereas some Tass_GDD6 and Silk_GDD6 QTL were stable across most of the environments (Fig. 4, chromosomes 2, 3, 4, 7, and 8; Fig. 5 chromosome 1). The proportion of SNP significant in only one environment was higher in the CF-Flint than in the CF-Dent panel, with the exception of PLHT. At the network level, more SNPs were declared significant for Tass_GDD6 and Silk_GDD6 in the CF-Dent panel, and for DMY and DMYcorr in the CF-Flint panel. Note that some of the significant SNPs were associated with more than one

Table 8 Statistics on the number of significant SNPs and QTL regions in the CF-Dent and CF-Flint panels evaluated on tester, for all traits, using K_Freq or K_Chr as kinship

	Estimation of K	Tass_GDD6	Silk_GDD6	ASI_GDD6	PLHT	DMC	DMCcorr	DMY	DMYcorr	Sum
CF-Dent										
Network[a]	K_Freq	12	8	0	2	3	0	1	1	27
Trials[b]	K_Freq	35	27	22	12	14	5	48	33	196
Per_trial[c]	K_Freq	6.18	4.09	2	1.09	1.45	0.45	4.36	3.09	
Prop_specific[d]	K_Freq	0.69	0.74	1	1	0.93	1	1	0.97	4
Network	K_Chr	16	10	0	4	5	1	1	1	38
Trials	K_Chr	45	39	26	14	23	8	56	47	258
Per_trial	K_Chr	7.91	5.82	2.36	1.36	2.27	0.73	5.18	4.45	
Prop_specific	K_Chr	0.71	0.77	1	0.93	0.96	1	0.98	0.96	2
QTL[e]	K_Freq or K_Chr	29	24	22	8	21	9	33	27	173
CF-Flint										
Network	K_Freq	1	2	1	4	1	0	4	5	18
Trials	K_Freq	16	17	12	15	8	5	12	16	101
Per_trial	K_Freq	2.3	2.3	1.2	1.6	0.8	0.5	1.2	1.6	
Prop_specific	K_Freq	0.69	0.88	1	0.93	1	1	1	1	5
Network	K_Chr	2	3	2	8	1	0	4	5	25
Trials	K_Chr	18	19	14	19	13	6	12	15	116
Per_trial	K_Chr	2.9	2.5	1.4	2.1	1.3	0.6	1.2	1.5	
Prop_specific	K_Chr	0.72	0.89	1	0.89	1	1	1	1	5
QTL[e]	K_Freq or K_Chr	14	16	15	18	11	6	13	15	108

[a] Number of significant SNPs when considering the trial network adjusted means

[b] Number of significant associations when considering the trial-specific adjusted means

[c] Mean number of significant associations per trial

[d] Proportion of significant associations specific to one trial

[e] Total number of regions (QTL) detected

trait (Tables S5 and S6). These pleiotropic effects particularly concerned the following couples of traits: Tass_GDD6 and Silk_GDD6, Tass_GDD6 (or Silk_GDD6) and DMC (or DMY, or PLHT), PLHT and DMY (or DMC, or Tass_GDD6, or Silk_GDD6). In the CF-Flint panel, one SNP was associated with Tass_GDD6, Silk_GDD6, PLHT, and DMC.

When testing associations with per se adjusted means, QTL of Tass_GDD6 and Silk_GDD6 were discovered but only one QTL of ASI_GDD6 (in the CF-Dent panel) and no QTL of PLHT. Again, more QTL were found in the CF-Dent panel (25) than in the CF-Flint panel (14). Most of these QTL were located on chromosomes 3 and 8 in the CF-Dent panel and on chromosomes 1, 3, and 9 in the CF-Flint panel. Five QTL of Tass_GDD6 and four of Silk_GDD6 were found associated with both hybrid and per se performances (including *ZCN8,* see discussion) in the CF-Dent panel. In the CF-Flint panel, only one QTL of Tass_GDD6 and one QTL of Silk_GDD6 were found in both hybrid and per se evaluations.

Discussion

In the present study two highly diverse panels (a flint and a dent) were characterized genotypically and phenotypically for traits related to flowering time and biomass production. This allowed the evaluation of the available diversity and the detection of QTL. This is a first step to analyze the breeding potential of this material for traits related to biomass.

Genetic diversity organization

Both panels displayed a high genetic diversity attested by a high proportion of polymorphic PANZEA-markers in both panels (85 and 82 % in the CF-Dent and CF-Flint panels, respectively). Average gene diversity ($\overline{He}$) at polymorphic loci was high (0.37 and 0.36 in the CF-Dent and CF-Flint panels, respectively). These ($\overline{He}$) values are in the upper range of those reported in diversity studies based on SNPs (Hamblin et al. 2007; Lu et al. 2009; Truntzler et al. 2012;

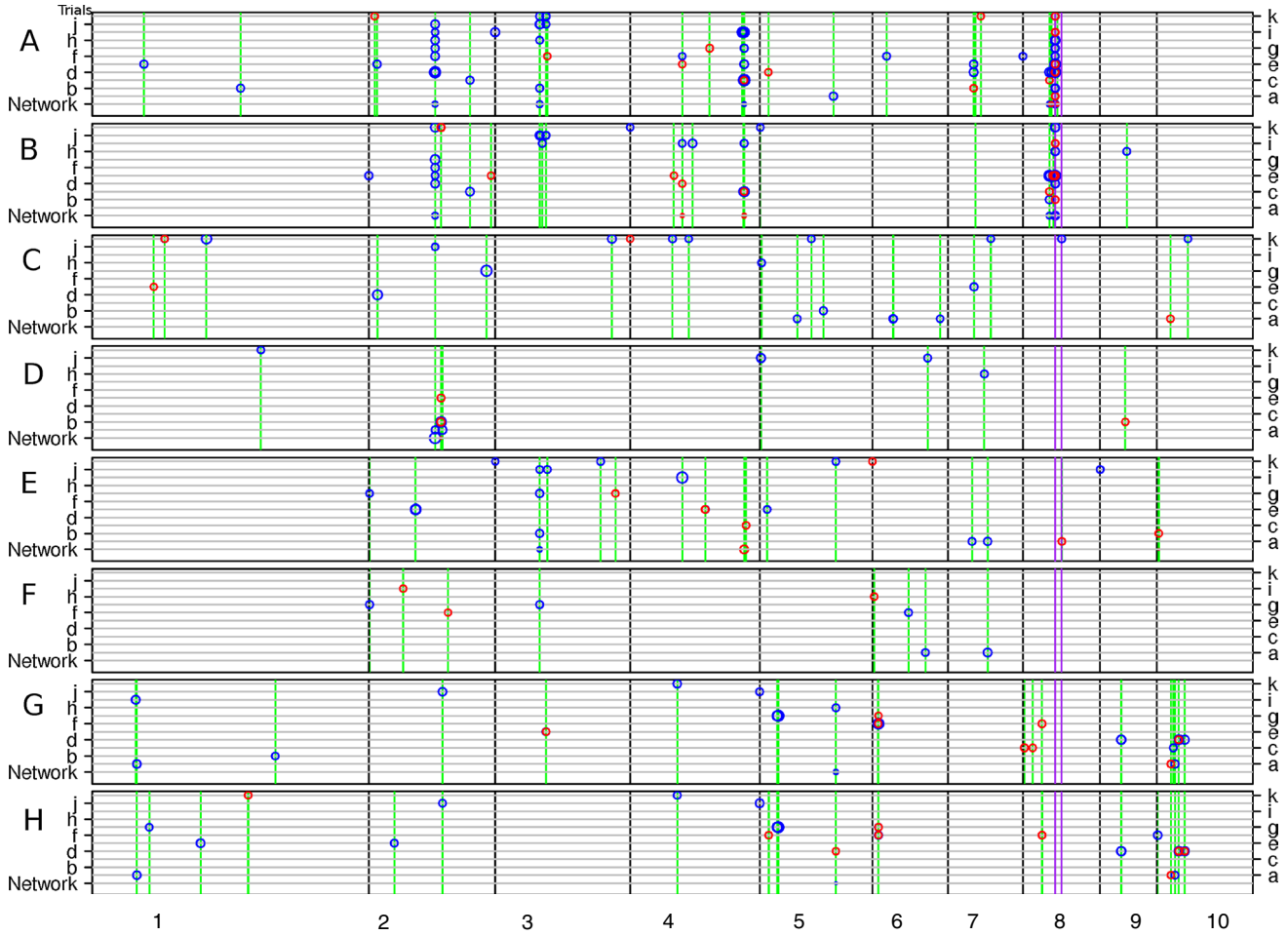

Fig. 4 Significant SNPs identified in the CF-Dent panel for the different traits in the different environments and for the global adjusted means. Circle diameters are proportional to the −log10 (*p* value), and the *red color* indicates the additional significant SNPs when using K_Chr as covariance matrix (the markers physically linked to the tested SNP are not used to estimate kinship). Chromosomes are separated by *black lines; Vgt1* and *Vgt2* (from *right* to *left*) are indicated by *purple lines*. The different traits are *A* Tass_GDD6, *B* Silk_GDD6, *C* ASI_GDD6, *D* PLHT, *E* DMC, *F* DMCcorr, *G* DMY, *H* DMYcorr. The trials are: *a* Mons 2010, *b* Pontevedra 2010, *c* La Coruna 2010, *d* Roggenstein 2010, *e* Einbeck 2010, *f* Mons Late Planting 2011, *g* Moulon 2011, *h* Mons Early Planting 2011, *i* Pontevedra 2011, *j* La Coruna 2011, *k* Pocking 2011

Van Inghelandt et al. 2010; Bouchet et al. 2013). The lower polymorphism and slightly lower gene diversity observed in the CF-Flint panel are consistent with the observations of Bouchet et al. (2013), who hypothesized that this could be the consequence of the severe bottleneck encountered by the Flint material when diverging from tropical germplasm. As in our study (Fig. 1), Bouchet et al. (2013) observed more rare alleles in the Flints. Based on this, they hypothesized that the initial bottleneck was followed by an expansion involving migrations from different origins, each with limited contributions. This hypothesis is consistent with the more complex population structure observed for the Flints (see below). The grouping based on the molecular information revealed the complex structure of both panels. From $N_Q = 2$ to $N_Q = 8$, all the identified groups could be interpreted using the pedigree information and/or known assignation to heterotic groups (Fig. S2). The groups identified in the CF-Flint panel appear to be related to the ancient history of this material. In particular, the first introductions of maize into Europe (from the Caribbean into Southern Europe by Columbus in 1493, and from Northern America into Northern Europe before 1539, Rebourg et al. 2003; Revilla et al. 2003) are still clearly visible in our results (Southern OPVs vs. Northern Flint, respectively, Fig. 3). The CF-Dent panel does not show such ancient historical patterns, consistent with the fact that this group originated from Corn-Belt Dent open pollinated varieties which displayed limited population structure (Camus-Kulandaivelu et al. 2006). Admixture groups observed in our study appear to be the result of the diverse breeding strategies which have been applied since the early development of hybrid maize in the USA. The network and PCoA

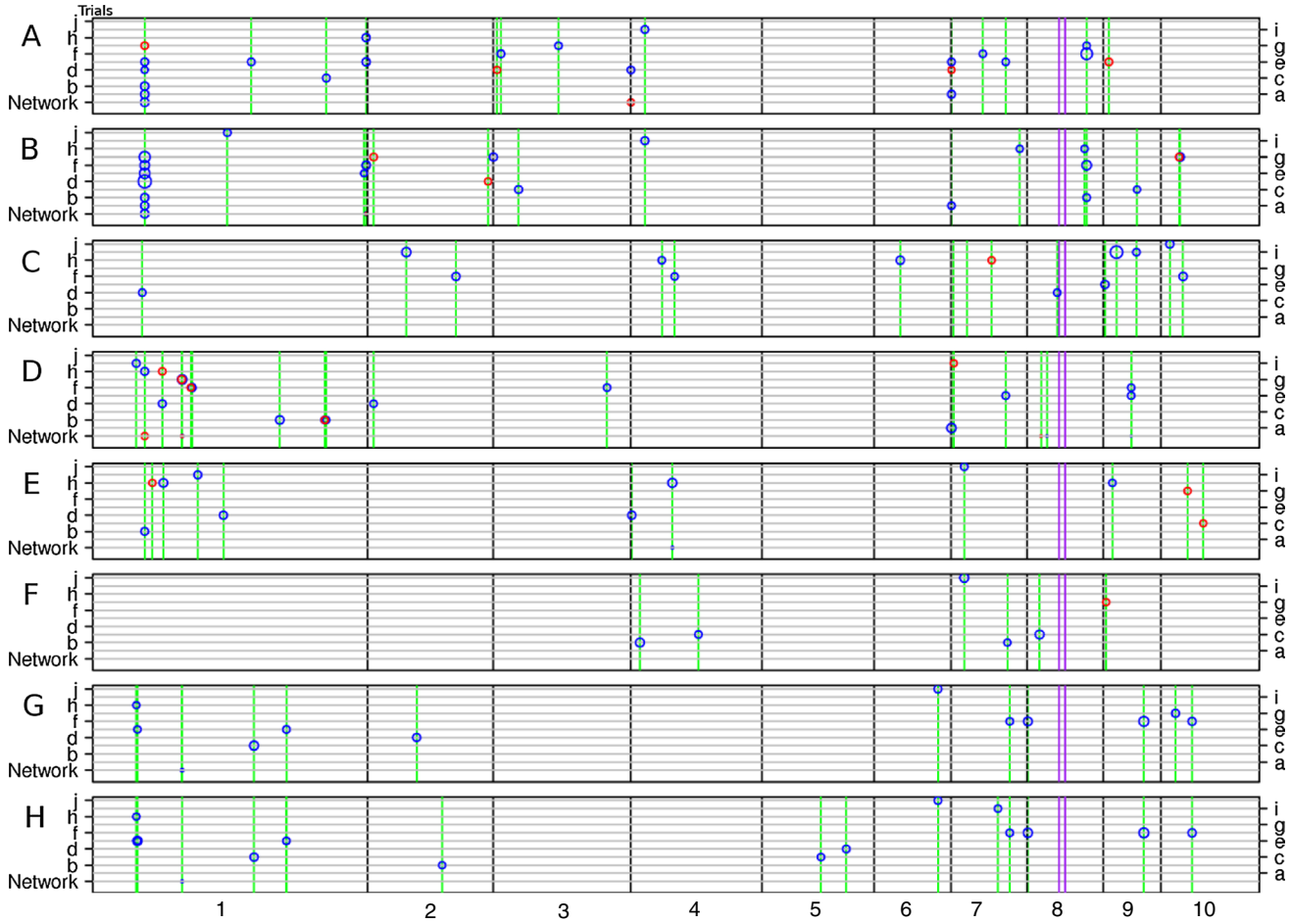

Fig. 5 Significant SNPs identified in the CF-Flint panel for the different traits in the different environments and for the global adjusted means. Circle diameters are proportional to the −log10 (*p* value), and the *red color* indicates the additional significant SNPs when using K_Chr as covariance matrix (the markers physically linked to the tested SNP are not used to estimate kinship). Chromosomes are separated by *black lines*; *Vgt1* and *Vgt2* (from *right* to *left*) are indicated by *purple lines*. The different traits are *A* Tass_GDD6, *B* Silk_GDD6, *C* ASI_GDD6, *D* PLHT, *E* DMC, *F* DMCcorr, *G* DMY, *H* DMYcorr. The trials are: *a* Mons 2010, *b* Pontevedra 2010, *c* La Coruna 2010, *d* Roggenstein 2010, *e* Einbeck 2010, *f* Moulon 2011, *g* Ploudaniel 2011, *h* Pontevedra 2011, *i* La Coruna 2011, *j* Pocking 2011

visualizations revealed that the material available relates to a large extent to a limited number of key lines, in particular in the CF-Dent panel (Fig. 3). Each key line and the material derived from it generated structure groups which were also clearly visible in the network and PCoA visualizations. This clustering around key lines (B73, Mo17, and PH207) corresponds to the three main Dent groups (Stiff Stalk, Lancaster and Iodent, respectively) and was also shown for instance by Dubreuil et al. (1996), and Romay et al. (2013) using genotyping by sequencing (GBS) data. The fact that the CF-Flint panel appeared less structured by modern breeding than the CF-Dent panel is consistent with the fact that Flint genetic group was submitted to less breeding cycles. Hybrids involving Flint parents are indeed recent (1960s) compared to the first Dent hybrids (1930s) developed in the USA. Also, as contrary to Dents, hybrids between two Flint lines only had a very limited success, this group was not sub-structured into complementary heterotic groups (Revilla et al. 2002). The different history of the genetic groups resulted in higher differentiation of the groups in the CF-Dent than in the CF-Flint panel (this was true for $N_Q = 2$ to $N_Q = 8$, Table 1). Despite efforts made to assemble materials from different institutes, it appeared that some heterotic groups or families were common to these institutes. There are, however, some noticeable exceptions like CIAM-Aranga and Hohenheim Flints which appear specific for the institutes which created the corresponding lines.

Relatedness between individuals greatly influenced LD between pairs of markers (in particular between unlinked markers, Fig. S3) in both panels, particularly in CF-Dent. When taking kinship into account, LD remained higher in the CF-Dent panel. As observed in previous studies (Van Inghelandt et al. 2010; Bouchet et al. 2013), LD decreased

with the physical (and genetic) distance. LD decay was variable between chromosomes, with a higher extent on chromosomes 3, 4, and 8 in the CF-Dent panel, and chromosome 8 in the CF-Flint panel. This is in accordance with Djemel et al. (2013), Romay et al. (2013) and with Khobragade et al. (pers. com.), who identified particularly long haplotypes on chromosome 4, in regions including important domestication genes. Other important genes related to flowering time (*Vgt1* and *Vgt2*) are located on chromosome 8 (Chardon et al. 2005; Salvi et al. 2007, 2009; Veyrieras et al. 2007; Ducrocq et al. 2008; Van Inghelandt et al. 2012; Bouchet et al. 2013). The slight drop of diversity in the region of *Vgt*1 and *Vgt*2 in the CF-Flint panel (Fig. 2) may be due to the fixation of the early alleles during adaptation to short growing seasons. The higher LD extent in the CF-Dent panel resulted in a reduced number of SNPs required for a minimum coverage of the genome (19,000 markers in comparison to 24,387 markers in the CF-Flint). The number of SNPs available in GWAS in the panels (42,214 and 39,076 in the CF-Dent and CF-Flint panels, respectively) makes it possible to conduct a first genome-wide analysis. However, these available markers are not evenly spaced along the genetic map, and a LD of 0.1 between adjacent pairs of SNP is insufficient to detect QTL of small to intermediate effect in our panels. In the CF-Flint panel, fewer markers were available for GWAS, whereas more markers were needed to cover the genome than in the CF-Dent panel. This could lead to a lower power in the CF-Flint panel in some regions of the genome. In both panels, we expect that a substantial gain in power could be obtained by increasing the number of markers (by combining GBS, sequencing, and imputation for example).

Also, one of the main limitations in the dissection of quantitative traits is the size of the population under study, which affects GWAS power and the reliability of genomic predictions. For this reason, the panel size should be as large as possible. However, we showed in this study that, at some point, the sampling of additional individuals often results in relatedness (possibly high, Fig. 3), which may decrease GWAS marginal gain of power. This highlights the importance of screening collections of landraces and of first cycle lines, which can probably be used to increase panel size and diversity without increasing too much relatedness, and as a result increase the potential of the panels for QTL detection, and for the inference of evolutionary events.

Trait variation within and among genetic groups

All traits in both panels showed high genetic variability, which resulted in high heritabilities at the trial network levels. Male and female flowering time (Tass_GDD6 and Silk_GDD6) were the most heritable traits (above 0.96 at the network level), and Anthesis to Silking Interval (ASI_GDD6), which is highly sensitive to environmental stresses, was the least heritable trait (0.73 in CF-Dent and 0.65 in CF-FFlint). Heritabilities at the trial-network level were in the range that is expected for the observed traits. Trial heritabilities of yield traits (DMC, DMCcorr, DMY and DMYcorr) were highly variable between trials, likely because of different environmental conditions and different management practices. The hybrid heritabilities were slightly higher for the Dent than for the CF-Flint panel except for flowering time. This is mostly due to higher residual variances in the CF-Flint panel, partly explained by plant lodging in some of the trials.

The heritabilities in the per se single location evaluations are close to the heritabilities at the trial-network level in the hybrid evaluation for Tass_GDD6 and Silk_GDD6. This is due to a genetic variance 5.5–6.4 times higher, and a residual variance only 1.2–3.4 times higher than in the hybrid experimental design (Tables 3, 4). This difference of genetic variability between per se and hybrid evaluation is higher than the fourfold increase expected under an additive model. This suggests the existence of a substantial amount of non-additive genetic effects. The range of correlations between per se and hybrid adjusted means revealed the importance to evaluate biomass production potential of the lines in hybrid progenies and not per se only.

The high genetic diversity and phenotypic variability of these two panels encourage the development of more productive biomass maize. Comparison of group materials revealed by population structure analysis showed a significant effect on all traits (Table 5). It highlighted groups with original characteristics like the "CIAM Aranga and EC18 related" group in the Flints, or the Stiff Stalk lines (particularly those related to B73) in the Dents, which displayed a high productivity relative to their earliness (Table 5). High variances nevertheless exist within genetic groups. Although a formal analysis was not possible due to the complexity of pedigrees, we observed some groups for which recent materials were more productive than that of ancestral founder lines (e.g. PH207 is considered by breeders as the most representative of the related founders of the Iodent group, and displays a DMYcorr value below that of its three direct descendants PHH93, PHG50 and PHG83, see Table S1). This reveals that both Flint and Dent groups have undergone genetic progress (Tables S1 and S2). However, substantial variability remains in the more recent lines (e.g. group "CIAM Aranga" in Table S2), which is promising for further breeding.

Association mapping results

The distribution of the *p* values (QQ-plot, Fig. S1) illustrates that a random polygenic effect was required to

control false positive rate efficiently and that both K_Freq and K_Chr were efficient for this (distribution near diagonal for *p* values above 0.01). However, the use of K_Chr instead of K_Freq substantially increased the number of significant SNPs (increase of around 40 % in both panels). Mixed model with K_Freq as kinship appeared indeed to be too conservative in simulations, and as a result less powerful (Rincent et al. 2014). This confirms the importance of removing markers in LD with the tested marker from the kinship estimation to limit "proximal contamination" (Listgarten et al. 2012). As expected from simulations in Rincent et al. 2014, the gain of power was less important in the CF-Flint than in the CF-Dent panel.

QTL were identified for all the traits in both panels (Tables 6, 7, 8, S5 and S6; Figs. 4, 5). Globally, more QTL were discovered in the CF-Dent than in the CF-Flint panel in the hybrid evaluation (173 and 108 QTL, respectively) and in the *per se* evaluation (25 and 14 QTL, respectively). This is consistent with the higher MAF, number of markers and LD extent in the CF-Dent panel (see above).

As expected based on knowledge of trait complexity and consequences on power, more QTL were found for Tass_GDD6 and Silk_GDD6 than for more complex traits (ASI_GDD6 or DMC), and these flowering QTL were more stable across environments. In particular, four regions of the genome in the CF-Dent panel (Fig. 4, chromosomes 2, 3, 4, and 8) and one region of the genome in the CF-Flint panel (Fig. 5, chromosome 1), were associated with flowering time in most of the environments. Polymorphism in the vicinity of *ZCN8* gene appeared as the most significant in both hybrid and per se evaluations in the CF-Dent panel. It corresponds to the *Vgt2* QTL found in numerous studies (Chardon et al. 2005; Salvi et al. 2007, 2009; Veyrieras et al. 2007; Ducrocq et al. 2008; Van Inghelandt et al. 2012; Bouchet et al. 2013; Romay et al. 2013). Note that it was not significant in the CF-Flint panel, neither in hybrid nor in per se evaluations, consistent with the quasi fixation of the early allele in Flint material reported by Bouchet et al. (2013). None of four other regions for flowering time appeared as strongly significant in Bouchet et al. (2013). Also, the strong association with days to silking corresponding to gene *ZmCCT* (Romay et al. 2013) on Chromosome 10 was not detected in our study, probably because the late allele at this locus (Ducrocq et al. 2009) is underrepresented in our panels, or marker density was too low in this region for capturing this effect.

PLHT was an exception to the global trend, as more QTL were found in the CF-Flint than in the CF-Dent panel (18 and 8 QTL, respectively), probably because it is the only trait (with Tass_GDD6 to some extent) which had a much higher genetic variance in the CF-Flint than in the CF-Dent panel (Table 3). We found common associations with the study of Peiffer et al. (2014), in particular in the CF-Flint panel (e.g. the QTL close to position 249 Mb on chromosome 1 near the gene *brassinosteroid-deficient dwarf1*, Pettem 1956). Interestingly, most of the PLHT QTL are not associated with flowering traits, as also found by Peiffer et al. (2014). As both flowering time and plant height are increasingly documented in the literature and less subject to GxE interactions than yield, a formal meta-analysis of our study and literature investigations would be highly beneficial to go beyond these preliminary trends.

For DMY or DMYcorr, many significant associations were discovered, but they were highly instable between environments (more than 96 % of these SNPs were significant in only one environment). The genetic determinism of these traits is more difficult to investigate because of interactions with the environment and/or because they are highly integrative. We noted that some associations for DMY were common to flowering time, suggesting a pleiotropic effect of the corresponding QTL. Note, however, that the QTL observed at network level for DMYcorr and DMY at position 154,077,833 bp on chromosome 1 (Table 7) in the Flint and position 190,732,112 bp on chromosome 5 in the CF-Dent panel (Table 6) do not belong to this category and, therefore, would be particularly interesting to select for biomass yield without modifying flowering time. These two QTL are all the more promising that the favorable alleles were both found with a frequency below 0.2. Note that the favorable alleles of these two QTL were absent for most of the groups, except in the Lancaster, the Stiff Stalk and the D06-related groups for the dent QTL, and in the Pyrenean group and the descents from non Northern Flint introductions for the flint QTL. Finally, DMC, DMCcorr, and ASI displayed the fewest number of detected QTL, highlighting that they are most likely affected by numerous factors of small effects and/or strong environmental effects.

Most of the significant SNPs identified with the hybrid adjusted means were different from those identified with the per se adjusted means. This could be due to the fact that lines were observed per se in only one environment, or to interactions between alleles (dominance and possibly epistasis), which was also shown by the genetic variance higher than expected in the per se evaluations. This effect was more pronounced in the CF-Flint than in the CF-Dent panel. As the proportion of SNPs significant in only one environment was also higher in the CF-Flint panel, we can hypothesize that the CF-Flint panel is probably submitted to more gene × gene and gene × environment interactions. This probably relates to the fact that flint inbreds were specifically selected in different European environments, while the dent inbreds were mainly introduced from the USA because they had good performance over a large range of environments.

Conclusions

We could illustrate, using genotypic and phenotypic information, that Dent and Flint groups have a different history and that this has strong consequences on diversity, variability, and LD extent, which in turn influence detection power. The combination of phenotypic and genotypic data permitted the identification of flowering time and biomass-related QTL in both panels. This study would probably profit from increasing the number of markers and population size. Analyzing this dataset with statistical models including interaction terms would also probably be beneficial. Although further analyses are required, the favorable biomass QTL alleles detected in this study are potentially of considerable interest, because they could be introgressed in elite material to increase productivity.

Author contributions AC, CB, PR, JMG, AEM, MO, PD, PF, CCS conceived the experiments; PR, RAM, JMG, AEM, WS, JL, CB, AC provided genetic materials; DM, VC, FD, SN, TA, DB managed DNA extraction and genotyping; CG, EB, CCS, NM, MO, CB, PJ, LC, JMG conducted field experiments; RR, SB and SN conducted statistical analyses; RR, LM, AC wrote most of the MS with the help of JMG, PR, EB, AEM, CB, SB.

Acknowledgments We are very grateful to all who made possible the gathering of their inbred lines to our panels. In particular, Mark Millard from United States Department of Agriculture North Central Regional Plant Introduction Station (NCRPIS) of Ames, USA; Natalia de Leon from University of Wisconsin, USA; Geert Kleijer from Agroscope Changins-Wädenswil of Nyon (ETH Zurich) Switzerland; Wolfgang Schipprack from Universität Hohenheim (UH) of Eckartsweier, Germany; Rita Redaelli from Unita Di Ricerca per la Maiscoltura of Bergamo (ISC), Italy; Amando Ordás from Misión Biológica de Galicia of Pontevedra (CSIC), Spain; Ángel Álvarez from Estacion Experimental de Aula Dei of Zaragoza, Spain; José Ignacio Ruiz de Galarreta from Centro Neiker de Arkaute of Vitoria, Spain; colleagues from Centro de Investigaciones Agrarias de Mabegondo (CIAM), Spain and colleagues from Institut National de la Rercherche Agronomique of (INRA) Saint Martin de Hinx, France. This research was jointly supported as "Cornfed project" by the French National Agency for Research (ANR), the German Federal Ministry of Education and Research (BMBF), and the Spanish ministry of Science and Innovation (MICINN). R. Rincent is jointly funded by Limagrain, Biogemma, KWS, and the French ANRt. L. Moreau, S. Nicolas and A. Charcosset conducted this research in the framework of Amaizing Investissement d'Avenir program. The authors thank the reviewers and the editor for their comments which improved the manuscript.

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

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

References

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655–1664

Astle W, Balding DJ (2009) Population structure and cryptic relatedness in genetic association studies. Stat Sci 24:451–471

Barrière Y, Gibelin C, Argillier O, Méchin V (2001) Genetic analysis and QTL mapping in forage maize based on recombinant inbred lines descended from the cross between F288 and F271. I—yield, earliness, starch and crude protein content. Maydica 46:253–266

Barrière Y, Méchin V, Denoue D, Bauland C, Laborde J (2010) QTL for yield, earliness and cell wall digestibility traits in topcrossexperiments of F838xF286 RIL progenies. Crop Sci 50:1761–1772

Beló A, Zheng P, Luck S, Shen B, Meyer DJ, Li B, Tingey S, Rafalski A (2007) Whole genome scan detects an allelic variant of fad2 associated with increased oleic acid levels in maize. Mol Genet Genomics 279:1–10

Bouchet S, Servin B, Bertin P, Madur D, Combes V, Dumas F, Brunel D, Laborde J, Charcosset A, Nicolas S (2013) Adaptation of maize to temperate climates: mid-density genome-wide association genetics and diversity patterns reveal key genomic regions, with a major contribution of the *Vgt2* (*ZCN8*) locus. PLoS One 8:e71377

Browning BL, Browning SR (2009) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet 84:210–223

Camus-Kulandaivelu L, Veyrieras J-B, Madur D, Combes V, Fourmann M et al (2006) Maize adaptation to temperate climate: relationship between population structure and polymorphism in the Dwarf8 gene. Genetics 172:2449–2463

Chardon F, Hourcade D, Combes V, Charcosset A (2005) Mapping of a spontaneous mutation for early flowering time in maize highlights contrasting allelic series at two-linked QTL on chromosome 8. Theor Appl Genet 112:1–11

Djemel A, Romay MC, Revilla P, Khelifi L, Ordás A, Ordás B (2013) Genomic regions affecting fitness of the sweetcorn mutant sugary1. J Agric Sci 151:396–406

Dubreuil P, Dufour P, Krejci E, Causse M, De Vienne D, Gallais A, Charcosset A (1996) Organization of RFLP diversity among inbred lines of maize representing the most significant heterotic groups. Crop Sci 36:790–799

Ducrocq S, Madur D, Veyrieras J-B, Camus-Kulandaivelu L, Kloiber-Maitz M, Presterl T, Ouzunova M, Manicacci D, Charcosset A (2008) Key impact of *Vgt1* on flowering time adaptation in maize: evidence from association mapping and ecogeographical information. Genetics 178:2433–2437

Ducrocq S, Giauffret C, Madur D, Combes V, Dumas F, Jouanne S, Coubriche D, Jamin P, Moreau L, Charcosset A (2009) Fine mapping and haplotype structure analysis of a major flowering time quantitative trait locus on maize chromosome 10. Genetics 183:1555–1563

Ewens W, Spielman R (1995) The transmission disequilibrium test—history, subdivision and admixture. Am J Hum Genet 57:455–464

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Fruchterman TMJ, Reingold EM (1991) Graph drawing by force-directed placement software. Pract Exp 21:1129–1164

Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD, Graner E-M, Hansen M, Joets J, Le Paslier M-C, McMullen MD, Montalent P, Rose M, Schön C-C, Sun Q,

Walter H, Martin OC, Falque M (2011) A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS One 6:e28334

Gilmour AR, Gogel B, Cullis BR, Thompson R (2009) ASREML user guide release 30VSN International Ltd, Hemel Hempstead, UK

Gore MA, Chia J-M, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J, Ware DH, Buckler ES (2009) A first-generation haplotype map of Maize. Science 326:1115–1117

Goudet J (2005) Hierfstat, a package for r to compute and test hierarchical F-statistics. Mol Ecol Notes 5:184–186

Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–338

Hamblin M, Warburton M, Buckler E (2007) Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. PLoS One 2:e1367

Hastie TJ, Tibshirani RJ (1990) Generalized additive models. Chapman and Hall, London

Herrmann A, Rath J (2012) Biogas production from maize: current state, challenges and prospects 1 Methane yield potential. Bio Energy Res 5(4):1027–1042

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 33:226–231

Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. Theor Popul Biol 33:54–78

Jannink JL, Walsh B (2003) Association mapping in plant populations p 59–68. In: Kang MS (ed) Quantitative genetics, genomics and plant breeding. CAB Int, New York

Jones P, Chase K, Martin A, Davern P, Ostrander EA, Lark KG (2008) Single-nucleotide-polymorphism-based association mapping of dog stereotypes. Genetics 179:1033–1044

Li J, Ji L (2005) Adjusting multiple testing in multilocus analyses using the Eigen values of a correlation matrix. Heredity 95:221–227

Listgarten J, Lippert C, Kadie CM, Davidson RI, Eskin E et al (2012) Improved linear mixed models for genome-wide association studies. Nat Method 9:525–526

Lu Y, Yan J, Guimaraes CT, Taba S, Hao Z et al (2009) Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. Theor Appl Genet 120:93–115

Mangin B, Siberchicot A, Nicolas S, Doligez A, This P, Cierco-Ayrolles C (2012) Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. Heredity 108:285–291

Mikel MA (2006) Availability and analysis of proprietary Dent corn inbred lines with expired US plant variety protection. Crop Sci 46:2555

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci 70:3321–3323

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583

Nelson PT, Coles ND, Holland JB, Bubeck DM, Smith S et al (2008) Molecular characterization of Maize inbreds with expired US plant variety protection. Crop Sci 48:1673

Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Sato H, Hori M, Nakamura Y, Tanaka T (2002) Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. Nat Genet 32:650–654

Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang Z et al (2014) The genetic architecture of maize height. Genetics 196(4):1337–1356

Pettem F (1956) Dwarfs. Maize genetics cooperation. Newsletter 30:9–10

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945

R development Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Rath J, Heuwinkel H, Herrmann A (2013) Specific biogas yield of maize can be predicted by the interaction of four biochemical constituents. BioEnergy Res 6(3):939–952

Rebourg C, Chastanet M, Gouesnard B, Welcker C, Dubreuil P et al (2003) Maize introduction into Europe: the history reviewed in the light of molecular data. Theor Appl Genet 106:895–903

Revilla P, Malvar RA, Cartea ME, Soengas P, Ordás A (2002) Heterotic relationships among European maize inbreds. Euphytica 126:259–264

Revilla P, Soengas P, Cartea ME, Malvar RA, Ordás A (2003) Isozyme variability among European maize populations and the introduction of maize in Europe. Maydica 48:141–152

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F et al (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet 44:217–220

Rincent R, Laloe D, Nicolas S, Altmann T, Brunel D et al (2012) Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of Maize inbreds (*Zea mays* L.). Genetics 192:715–728

Rincent R, Moreau L, Monod H, Kuhn E, Melchinger AE et al (2014) Recovering power in association mapping panels with variable levels of linkage disequilibrium. Genetics 197:375–387

Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL et al (2013) Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol 14:R55

Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev EV, Svitashev S, Bruggemann E (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proc Natl Acad Sci 104:11376–11381

Salvi S, Castelletti S, Tuberosa R (2009) An updated consensus map for flowering time QTLs in maize. Maydica 54:501

SAS Institute (2011) Release 9.3. SAS Inst., Cary, NC, USA

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet 28:286–289

Truntzler M, Ranc N, Sawkins M, Nicolas S, Manicacci D, Lespinasse D, Ribière V, Galaup P, Servant F, Muller C et al (2012) Diversity and linkage disequilibrium features in a composite public/private dent maize panel: consequences for association genetics as evaluated from a case study using flowering time. Theor Appl Genet 125(4):731–747

Van Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor Appl Genet 120:1289–1299

Van Inghelandt D, Melchinger AE, Martinant J-P, Stich B (2012) Genome-wide association mapping of flowering time and northern corn leaf blight (Setosphaeria turcica) resistance in a vast commercial maize germplasm set. BMC Plant Biol 12:56

VanRaden P (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423

Veyrieras J-B, Goffinet B, Charcosset A (2007) MetaQTL: a package of new computational methods for the meta-analysis of QTL mapping experiments. BMC Bioinf 8:49

Wald A (1943) Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans Am Math Soc 54:426–482

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