

10-6-2016

Genome-Wide Association Study Identifies Candidate Loci Underlying Agronomic Traits in a Middle American Diversity Panel of Common Bean

Samira Mafi Moghaddam

North Dakota State University, samira.mafi@ndsu.edu

Sujan Mamidi

North Dakota State University

Juan M. Osorno

North Dakota State University, Juan.Osorno@ndsu.edu

Rian Lee

North Dakota State University, Rian.Lee@ndsu.edu

Mark A. Brick

Colorado State University, Mark.Brick@ColoState.edu

See next page for additional authors

Follow this and additional works at: <http://digitalcommons.unl.edu/panhandleresext>

Moghaddam, Samira Mafi; Mamidi, Sujan; Osorno, Juan M.; Lee, Rian; Brick, Mark A.; Kelly, James; Miklas, Phillip N.; Urrea, Carlos A.; Song, Qijian; Cregan, Perry B.; Grimwood, Jane; Schmutz, Jeremy; and McClean, Phillip E., "Genome-Wide Association Study Identifies Candidate Loci Underlying Agronomic Traits in a Middle American Diversity Panel of Common Bean" (2016). *Panhandle Research and Extension Center*. 123.

<http://digitalcommons.unl.edu/panhandleresext/123>

This Article is brought to you for free and open access by the Agricultural Research Division of IANR at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Panhandle Research and Extension Center by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

Samira Mafi Moghaddam, Sujan Mamidi, Juan M. Osorno, Rian Lee, Mark A. Brick, James Kelly, Phillip N. Miklas, Carlos A. Urrea, Qijian Song, Perry B. Cregan, Jane Grimwood, Jeremy Schmutz, and Phillip E. McClean

Genome-Wide Association Study Identifies Candidate Loci Underlying Agronomic Traits in a Middle American Diversity Panel of Common Bean

Samira Mafi Moghaddam,* Sujan Mamidi, Juan M. Osorno, Rian Lee, Mark Brick, James Kelly, Phillip Miklas, Carlos Urrea, Qijian Song, Perry Cregan, Jane Grimwood, Jeremy Schmutz, and Phillip E. McClean

ABSTRACT

Common bean (*Phaseolus vulgaris* L.) breeding programs aim to improve both agronomic and seed characteristics traits. However, the genetic architecture of the many traits that affect common bean production are not completely understood. Genome-wide association studies (GWAS) provide an experimental approach to identify genomic regions where important candidate genes are located. A panel of 280 modern bean genotypes from race Mesoamerica, referred to as the Middle American Diversity Panel (MDP), were grown in four US locations, and a GWAS using >150,000 single-nucleotide polymorphisms (SNPs) (minor allele frequency [MAF] \geq 5%) was conducted for six agronomic traits. The degree of inter- and intrachromosomal linkage disequilibrium (LD) was estimated after accounting for population structure and relatedness. The LD varied between chromosomes for the entire MDP and among race Mesoamerica and Durango–Jalisco genotypes within the panel. The LD patterns reflected the breeding history of common bean. Genome-wide association studies led to the discovery of new and known genomic regions affecting the agronomic traits at the entire population, race, and location levels. We observed strong colocalized signals in a narrow genomic interval for three interrelated traits: growth habit, lodging, and canopy height. Overall, this study detected ~30 candidate genes based on a priori and candidate gene search strategies centered on the 100-kb region surrounding a significant SNP. These results provide a framework from which further research can begin to understand the actual genes controlling important agronomic production traits in common bean.

COMMON BEAN is the most important grain legume for direct human consumption (Broughton et al., 2003). It has a relatively small diploid genome ($2n = 22$; 521.1 Mb; Schmutz et al., 2014) and consists of two gene pools: Middle American and Andean. Domestication has occurred independently in each gene pool (Gepts and Bliss, 1986; Koenig and Gepts, 1989; Khairallah et al., 1990; Koinange and Gepts, 1992; Freyre et al., 1996; Mamidi et al., 2013). The two gene pools are strongly differentiated, and the Middle American gene pool has greater genetic diversity (Mamidi et al., 2013; Schmutz et al., 2014). Selection under domestication within each gene pool has generated distinct ecogeographical races and market classes (Singh et al., 1991; Beebe et al., 2000; Díaz and Blair, 2006; Mamidi et al., 2011b). Members of each race have specific morphological, agronomic, physiological, and molecular traits in common and can differ in the allele frequencies of the genes responsible for a trait (Singh et al., 1991). The

S.M. Moghaddam, S. Mamidi, P.E. McClean, Genomics and Bioinformatics Program and Dep. of Plant Sciences, North Dakota State Univ., Fargo, ND 58105-6050; Present address S. Mamidi, HudsonAlpha Institute of Biotechnology, Huntsville, AL 35806; J.M. Osorno and R. Lee, Dep. of Plant Sciences, North Dakota State Univ., Fargo, ND 58105-6050; M. Brick, Dep. of Soil and Crop Sciences, Colorado State Univ., Fort Collins, CO 80523-1170; J. Kelly, Dep. of Plant, Soil and Microbial Sciences, Michigan State Univ., East Lansing, MI 48824; P. Miklas, Vegetable and Forage Crops Research Unit, USDA-ARS, Prosser, WA 99350; C. Urrea, Dep. of Agronomy and Horticulture, Univ. of Nebraska, NE 69361-4939; Q. Song and P. Cregan, Soybean Genomics and Improvement Lab, USDA-ARS, Beltsville, MD 20705; J. Grimwood and J. Schmutz, HudsonAlpha Institute of Biotechnology, Huntsville, AL 35806; Received 8 Feb. 2016. Accepted 8 July 2016. *Corresponding author (samira.mm@gmail.com; samira.mafi@ndsu.edu).

Published in Plant Genome
Volume 9. doi: 10.3835/plantgenome2016.02.0012

© Crop Science Society of America
5585 Guilford Rd., Madison, WI 53711 USA
This is an open access article distributed under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Abbreviations: DJ, Durango–Jalisco; GWAS, genome-wide association studies; LD, linkage disequilibrium; MA, Mesoamerican; MAF, minor allele frequency; MDP, Middle American diversity panel; MLMM, multi-locus mixed model; PCA, principle component analysis; QTL, quantitative trait loci; SNP, single nucleotide polymorphism.

Middle American gene pool consists of races Durango–Jalisco (DJ) and Mesoamerica (MA). Pinto, great northern, medium red, and pink bean are major US market classes within race DJ, while navy and black bean are major race Mesoamerican market classes in the United States. (Mensack et al., 2010; Vandemark et al., 2014). Other studies investigated the population structure in the common bean gene pools and confirmed these classifications (Blair et al., 2009; Kwak and Gepts, 2009).

Cultivated forms of common bean show considerable morphological variation (Hedrick et al., 1931) for growth habit, seed size, and color (Leakey, 1988; Singh, 1989). Domestication and breeding selected for agronomorphological traits important for production and altered the genetic architecture underlying those traits. Breeding for higher yield in bean is affected by many interdependent traits including growth habit, seed size, and maturity (Kelly et al., 1998; Kornegay et al., 1992). Genetic factors controlling important agronomic traits have been mapped to common bean linkage groups using biparental populations and significant quantitative trait loci (QTL) mapped to wide intervals (Beattie et al., 2003; Blair et al., 2006, 2012; Checa and Blair, 2012; Pérez-Vega et al., 2010; Tar'an et al., 2002; Wright and Kelly, 2011). Nevertheless, our knowledge of genes controlling agronomic traits is limited. Quantitative trait loci analysis usually has low resolution because of the limited number of recombination events in a biparental population (Balasubramanian et al., 2009). Quantitative trait loci intervals can span a few centiMorgans (cM), which can indeed be megabases (Mb) long in physical distance and contain hundreds of candidate genes.

In contrast, GWAS considers more recombination events by using an association panel of individuals each with unique recombination histories. This provides higher mapping resolution, as a result of shorter LD blocks, than a biparental population. Thus, higher marker saturation is necessary to cover the whole genome. Higher mapping resolution, though, has only recently been possible with the development of high throughput genotyping in common bean (Hyten et al., 2010; Song et al., 2015). Resequencing GWAS populations and mapping the reads on to the reference genome sequence of common bean (Schmutz et al., 2014) lead to higher mapping resolution because of the greater depth of coverage. Multiple GWAS studies have led to the discovery of candidate genes affecting trait expression (Zhao et al., 2011; Korte and Farlow, 2013; Li et al., 2013; Appels et al., 2013; Verslues et al., 2014). Therefore, GWAS has recently been implemented in common bean (Cichy et al., 2015; Kamfwa et al., 2015a,b)

In this study, a GWAS was performed followed by candidate gene identification for six important agronomic traits that affect common bean production: days to flower, days to maturity, growth habit, canopy height, lodging, and seed weight. Over 240,000 SNPs were obtained from a collection of 280 diverse genotypes from a population of Middle American genotypes.

Materials and Methods

Middle American Diversity Panel and Single-Nucleotide Polymorphism Data Set

The MDP consists of 280 modern Middle American dry bean cultivars from among the 507 genotypes of the entire BeanCAP diversity panel (http://www.beancap.org/_pdf/DRY_BEAN_GENOTYPES_BEANCAP.pdf). This subpopulation was chosen to reduce the effect of population structure during the GWAS. The MDP itself consists of 100 race MA and 180 race DJ genotypes.

The SNP data set was obtained by genotyping the MDP with two Illumina iSelect 6K Gene Chip (BARC-BEAN6K_1 and BARCBEAN6K_2) sets (Song et al., 2015) and low-pass sequencing. The Gene Chips are described in Song et al. (2015). A total of 8526 SNPs from the two 6K Gene Chips were mapped to the common bean reference genome (Schmutz et al., 2014). Single-enzyme low-pass sequencing was performed by Institute for Genomic Diversity at Cornell University according to Elshire et al. (2011) using the enzyme *ApeKI* for digestion and yielded 17,611 mapped SNPs. The two-enzyme, low-pass sequencing SNP set was created according to Schröder et al. (2016) using *TaqI* and *MseI* enzymes. The depth of high-quality mapped reads ranged from 0.1× to 3.3× among the genotypes with an average of 0.7×. A total of 217,486 mapped SNPs were identified. The SNPs with <50% missing data were imputed. All SNP data sets were merged and imputed in fastPHASE (Scheet and Stephens, 2006). The chromosome distribution of SNPs that resulted from the three different platforms is summarized in Table 1. The unimputed HapMap data for each of the three platforms and the combined imputed data in numeric format can be found at <http://www.beancap.org/Research.cfm>.

Phenotypic Analysis

Data for days to flower, days to maturity, growth habit, canopy height, lodging, and seed weight were collected from field trials grown in Colorado, Michigan, Nebraska, and North Dakota. Days to flower were measured as the number of days from planting to when ~50% of the plants in a plot had at least one open flower. Days to maturity were measured as the number of days from planting until complete physiological development and senescence. Growth habit was recorded during flowering and verified at senescence as type I (determinate erect or upright), type II (indeterminate erect), or type III (indeterminate prostrate). Canopy height was measured at harvest and was recorded in centimeters from the base of the plant at the soil surface to the top of the canopy. Lodging was scored at harvest on a 1-to-5 scale, where 1 indicates 100% plants standing erect and 5 indicates 100% plants flat on the ground. Seed weight was measured as the weight of 100 randomly selected undamaged seeds recorded in grams at 16% moisture. Data for growth habit in Nebraska and growth habit and lodging in North Dakota were not available. The experimental design was an α -lattice design with two replicates. Trait

Table 1. Chromosome distribution of genotyping platforms.

Chromosome region	Physical size bp	Platforms								
		10 k chip			TaqI + MseI digestion			ApeKI digestion		
		No. SNP	Percentage SNP %	No. SNP Mb ⁻¹	No. SNP	Percentage SNP %	No. SNP Mb ⁻¹	No. SNP	Percentage SNP %	No. SNP Mb ⁻¹
Pv01										
Euchromatic 1	6,833,267	86	8	13	2143	9	314	555	24	81
Heterochromatic	31,514,565	698	65	22	17,118	75	543	745	32	24
Euchromatic 2	13,835,592	282	26	20	3541	16	256	1023	44	74
Pv02										
Euchromatic 1	4,498,885	78	8	17	1327	7	295	468	16	104
Heterochromatic	21,271,839	500	53	24	11,729	59	551	639	22	30
Euchromatic 2	23,262,928	372	39	16	6678	34	287	1750	61	75
Pv03										
Euchromatic 1	5,893,785	54	8	9	1327	7	225	461	19	78
Heterochromatic	23,491,976	463	67	20	11,729	59	499	551	22	23
Euchromatic 2	22,832,814	177	26	8	6678	34	292	1479	59	65
Pv04										
Euchromatic 1	7,965,625	310	28	39	3341	14	419	486	33	61
Heterochromatic	31,549,681	673	62	21	17,724	76	562	442	30	14
Euchromatic 2	6,277,826	107	10	17	2313	10	368	555	37	88
Pv05										
Euchromatic 1	3,818,224	61	6	16	1604	7	420	374	23	98
Heterochromatic	29,955,639	762	72	25	18,040	83	602	632	38	21
Euchromatic 2	6,463,603	241	23	37	2036	9	315	642	39	99
Pv06										
Heterochromatic	14,427,371	242	49	17	10,165	71	705	271	14	19
Euchromatic 1	17,545,758	250	51	14	4207	29	240	1643	86	94
Pv07										
Euchromatic 1	10,092,552	121	13	12	2767	13	274	872	36	86
Heterochromatic	27,468,866	618	65	22	14,640	70	533	471	19	17
Euchromatic 2	14,031,054	212	22	15	3604	17	257	1101	45	78
Pv08										
Euchromatic 1	9,747,519	131	11	13	2850	10	292	971	32	100
Heterochromatic	38,510,796	839	72	22	22,608	77	587	764	25	20
Euchromatic 2	11,376,202	193	17	17	3830	13	337	1290	43	113
Pv09										
Heterochromatic	5,767,213	109	17	19	2953	29	512	149	8	26
Euchromatic 1	31,632,295	522	83	17	7334	71	232	1804	92	57
Pv10										
Euchromatic 1	5,246,302	70	7	13	3351	13	639	271	20	52
Heterochromatic	29,290,594	717	72	24	19,464	78	665	518	38	18
Euchromatic 2	8,676,205	212	21	24	2224	9	256	578	42	67
Pv11										
Euchromatic 1	9,800,202	128	11	13	2742	9	280	916	41	93
Heterochromatic	33,323,703	843	69	25	22,500	77	675	699	31	21
Euchromatic 2	7,079,654	246	20	35	4155	14	587	615	28	87

heritability was calculated for each subpopulation (DJ and MA) using PROC MIXED in SAS 9.3 (SAS Institute, 2011) based on the method proposed by Holland et al. (2003). Trait correlations were calculated and plotted in R 3.0 (R Development Core Team, 2013) using cor.matrix() and corrplot() from the corrplot package (Wei and Wei, 2013). Histograms were created in R 3.0 (R Development Core Team, 2013) using the hist() function. Adjusted phenotypic values were calculated using the lsmeans

statement in SAS 9.3 (SAS Institute, 2011) PROC MIXED per location and across all locations. A total of 27 phenotypic datasets were analyzed. The lsmeans are available at http://www.beancap.org/_data/Adjusted-means-for-Agronomic-traits-with-race-and-market-calss-info.txt.

Population Structure and Kinship

Population structure was calculated with STRUCTURE 2.3 (Pritchard et al., 2000) using SNPs whose pairwise r^2

was <0.5 for all pairwise comparisons. STRUCTURE uses a model-based method to assign the subpopulation membership to each individual. The STRUCTURE parameters used were an admixture model with independent allele frequencies, a burn-in of 100,000, and an MCMC replication of 500,000 for $K = 1$ to 10 subpopulations with five replications. The Wilcoxon signed ranked test implemented in SAS 9.3 (SAS Institute, 2011) was used to select the optimum number of subpopulations (Rosenberg et al., 2001). The optimum number of subpopulations was the smallest K in the first nonsignificant Wilcoxon test. Distruct 1.1 (Rosenberg, 2003) was used for graphical display of the STRUCTURE output. Population structure for GWAS was determined using principal component analysis (PCA) using the `prcomp()` R 3.0 function (R Development Core Team, 2013; Price et al., 2006) with the complete SNP data set for loci with $MAF \geq 5\%$. To account for individual relatedness, an identity-by-state kinship matrix was generated by the EMMA algorithm (Kang et al., 2008) embedded in GAPIT (Lipka et al., 2012) using the complete SNP data set with $MAF \geq 0.05$.

Linkage Disequilibrium

Pairwise LD between markers in the null model was calculated as the squared allele frequency correlation in PLINK (Purcell et al., 2007) using SNPs with $MAF \geq 5\%$. To calculate genome-wide LD, a set of SNPs spaced every 50 Kb across the genome was used. The R package LDcorSV (Mangin et al., 2012) was used to calculate the pairwise LD when accounting for population structure, relatedness, or both. The LD heatmaps were generated for each chromosome in R 3.0 (R Development Core Team, 2013) using the LDheatmap package (Shin et al., 2006). The positions of the pericentromeric borders were obtained from Supplementary Table 6 in Schmutz et al. (2014). The LD between significant SNPs for the best model was calculated using kinship and PC matrices derived from the complete SNP data set with $MAF \geq 5\%$.

Genome-Wide Association Study

From a total of 243,623 SNPs, 159,884 with a $MAF \geq 5\%$ were used for the MDP GWAS. The DJ subpopulation ($n = 180$) GWAS analyses used 151,239 SNPs ($MAF \geq 5\%$), while the MA subpopulation GWAS analyses used 134,924 SNPs ($MAF \geq 5\%$). Analyses were performed using pooled data across all locations and for each location separately. More than 200 GWAS analyses were performed using GAPIT, an R function developed by Lipka et al. (2012). Multiple models were tested per trait in MDP: (i) null general linear model, (ii) general linear model with fixed effects to control for population structure, (iii) univariate unified mixed linear model (Yu et al., 2005) using the population parameters previously determined (Zhang et al., 2010) protocol to control for individual relatedness (random effect), and (iv) a model controlling for both individual relatedness and population structure. We used the first two principal components (controls for ~25% of variation) to account for

population structure. The best model was selected based on the mean squared difference as described by Mamidi et al. (2011a). The final Manhattan plots were created using the `mhtplot()` function from R package `gap` (Zhao, 2007). The phenotypic variation explained by significant markers in the best model was calculated based on the likelihood-ratio-based R^2 (R^2_{LR}) (Sun et al., 2010) using the `genABEL` package in R (Aulchenko et al., 2007). Multilocus mixed model (MLMM) (Segura et al., 2012) was used to evaluate the results for single-marker tests. The MLMM reduces the masking effect of population structure or selection on causative loci by forward and backward stepwise regression. Markers in complete LD ($r^2 = 1$) were excluded for this analysis. The best model was selected considering both multiple Bonferroni and extended Bayesian information criteria.

Significant Single-Nucleotide Polymorphisms and Candidate Gene Identification

The significance cutoff threshold affects candidate gene identification in GWAS. Although controlling for false positives is very crucial, the effect of false negatives should not be ignored. False negatives can certainly occur if the cutoff value is too stringent. Different approaches to setting significance cutoffs have included selecting the top 50 SNPs (Stanton-Geddes et al., 2013), considering two false discovery rate significance levels (Lipka et al., 2013), or adhering to the stringent Bonferroni cutoff. We defined significant markers by two cutoff levels. For the first, SNPs falling in the lower 0.01 percentile tail of the empirical distribution of P -values after 1000 bootstraps (Mamidi et al., 2014), which is a stringent cutoff but does not imply that if a SNP does not pass this criterion it does not affect the trait of interest. This was demonstrated by Atwell et al. (2010), where a SNP close to or even within a candidate gene previously validated by transgenic complementation experiments did not meet specific significance criteria. Therefore, the second cutoff used was more relaxed. SNPs falling in the lower 0.1 percentile tail of the empirical distribution of P -values after 1000 bootstraps we considered. Two approaches were used to define candidate genes: (i) a priori candidates based on the published literature and (ii) evaluation of genes within a 100-Kb window centered on a significant SNP. The primary focus was on SNPs falling in the more stringent cutoff, but SNPs falling in the less stringent cutoff were considered if there was a priori candidate support.

Results

Population Structure

With $K = 2$ subpopulations, the MDP was split into two subpopulations corresponding to races MA and Durango (Fig. 1a). When the subpopulation number was set to $K = 3$ to $K = 7$, genotypes clustered according to the market classes and admixture was observed between the market classes. Figure 1b shows that the first three principal components from PCA correspond to the STRUCTURE results. The first component separates race MA and

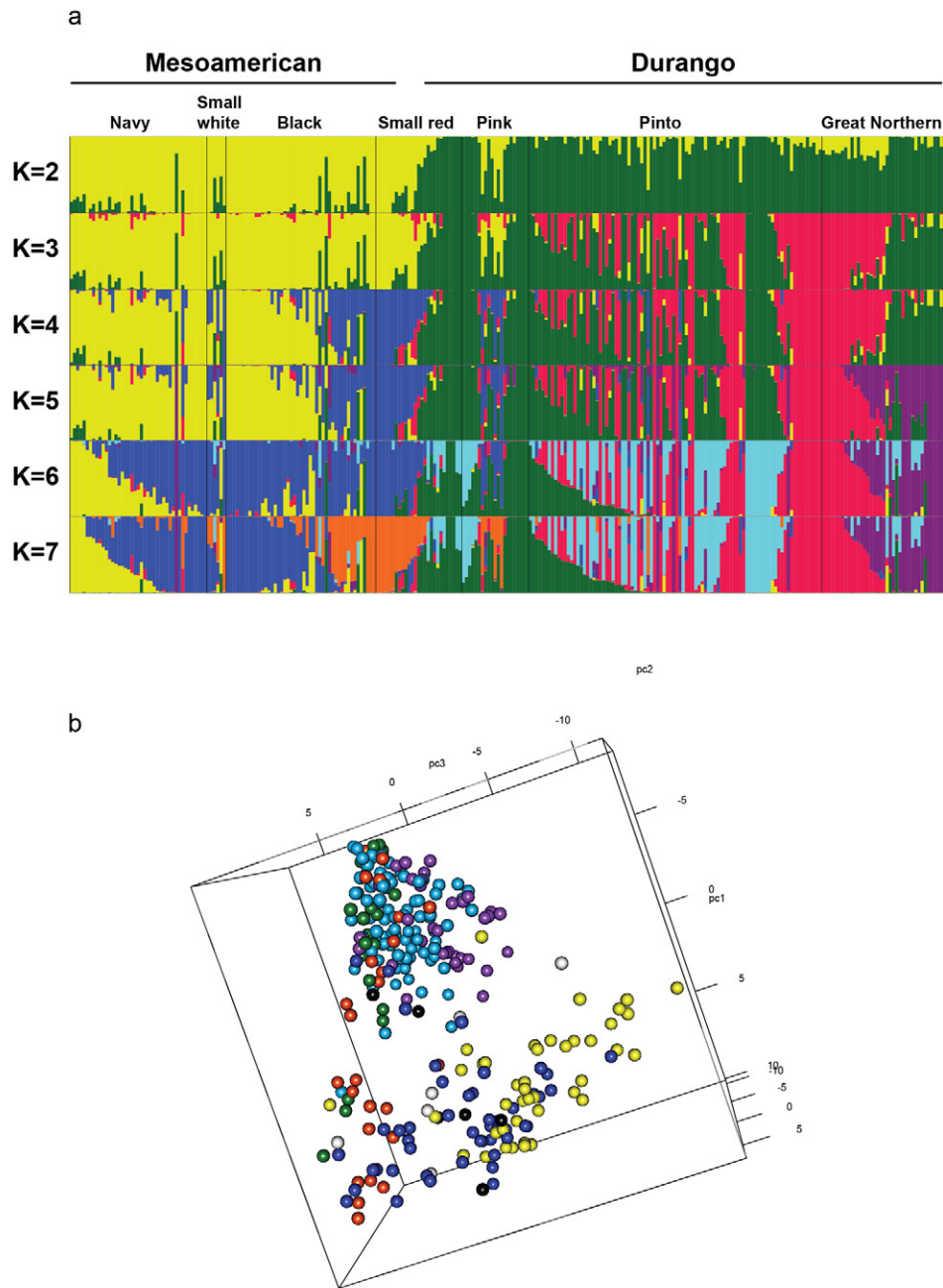


Fig. 1. STRUCTURE and principal component analysis. (a) STRUCTURE analysis of the Middle American diversity panel (MDP) from $K = 2$ to $K = 7$. Line $K = 2$ shows the split between the two major races, and from $K = 3$ to $K = 7$, each race subdivides into market classes. (b) Three-dimensional principal component (PC) analysis of the MDP. The first dimension separates the two major races. The color coding of bean market classes is as follow: great northern (purple); black (blue); navy (yellow); pink (green); pinto (light blue); small red (orange); small white (white); and black mottled, carioca, flor de mayo, and tan (black).

Durango. Pink and small red bean genotypes are distributed across both races, a result consistent with the admixture pattern in these market classes observed with the STRUCTURE analysis (Fig. 1a). The PC2 clusters some of the pinto and great northern bean genotypes closer to pink and small red bean market classes; the PC3 separates the race MA navy and black bean market classes and shows that small red and some pink bean are more closely related to the black bean market class.

Linkage Disequilibrium

The LD heatmaps were created for each chromosome for the MDP and the MA and DJ race subpopulations to illustrate the extent of LD at different levels of population structure. The LD pattern not only varied between subpopulations but also varied among chromosomes within a subpopulation. Pv06, Pv07, and Pv11 showed dramatically different LD patterns among the DJ and MA race subpopulations (Fig. 2a, 2b).

a

Durango/Jalisco Subpopulation

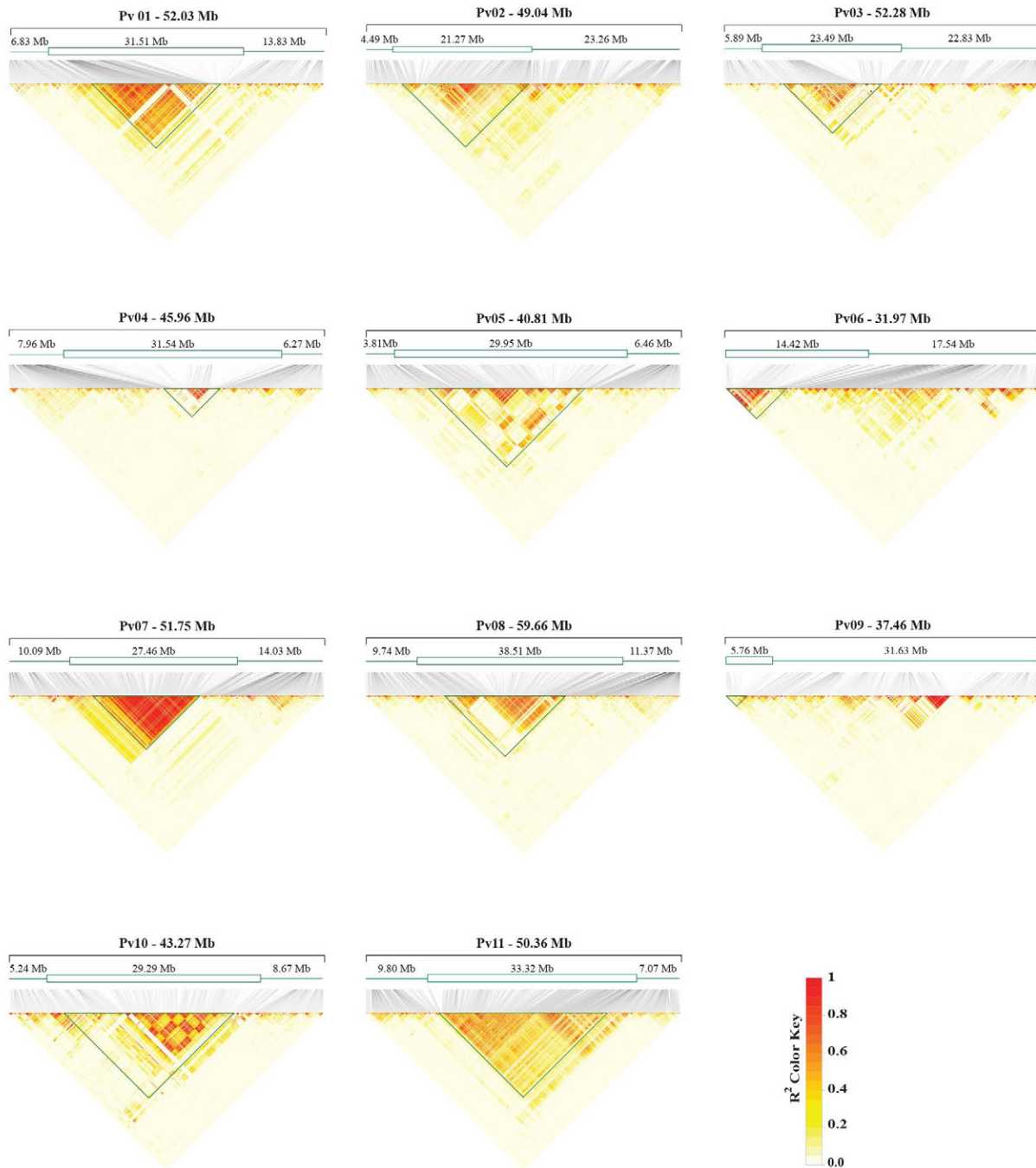


Fig. 2. (continued on next page) Linkage disequilibrium heatmaps of 11 chromosomes in (a) race DJ and (b) race MA after controlling for population relatedness. The green lines denote the boundaries of the pericentromeric region. A subset of single-nucleotide polymorphisms with minor allele frequency $\geq 5\%$ were used.

The majority of interchromosomal LD ($r^2 \geq 0.6$ after controlling for population structure and relatedness in MDP) occurred among pericentromeric regions. The largest interchromosomal LD is between Pv07 and other chromosomes. A ~ 30 -Mb region of Pv07 that encompasses the complete pericentromeric region (27 Mb) and

a small amount of neighboring euchromatic DNA is in LD with either a single SNP or narrow regions on other chromosomes. Only Pv05 and Pv09 do not exhibit any LD with Pv07. Although the mixed model significantly reduced the overall interchromosomal LD, some long-range LD blocks persisted in the DJ subpopulation, but

b

Mesoamerican Subpopulation

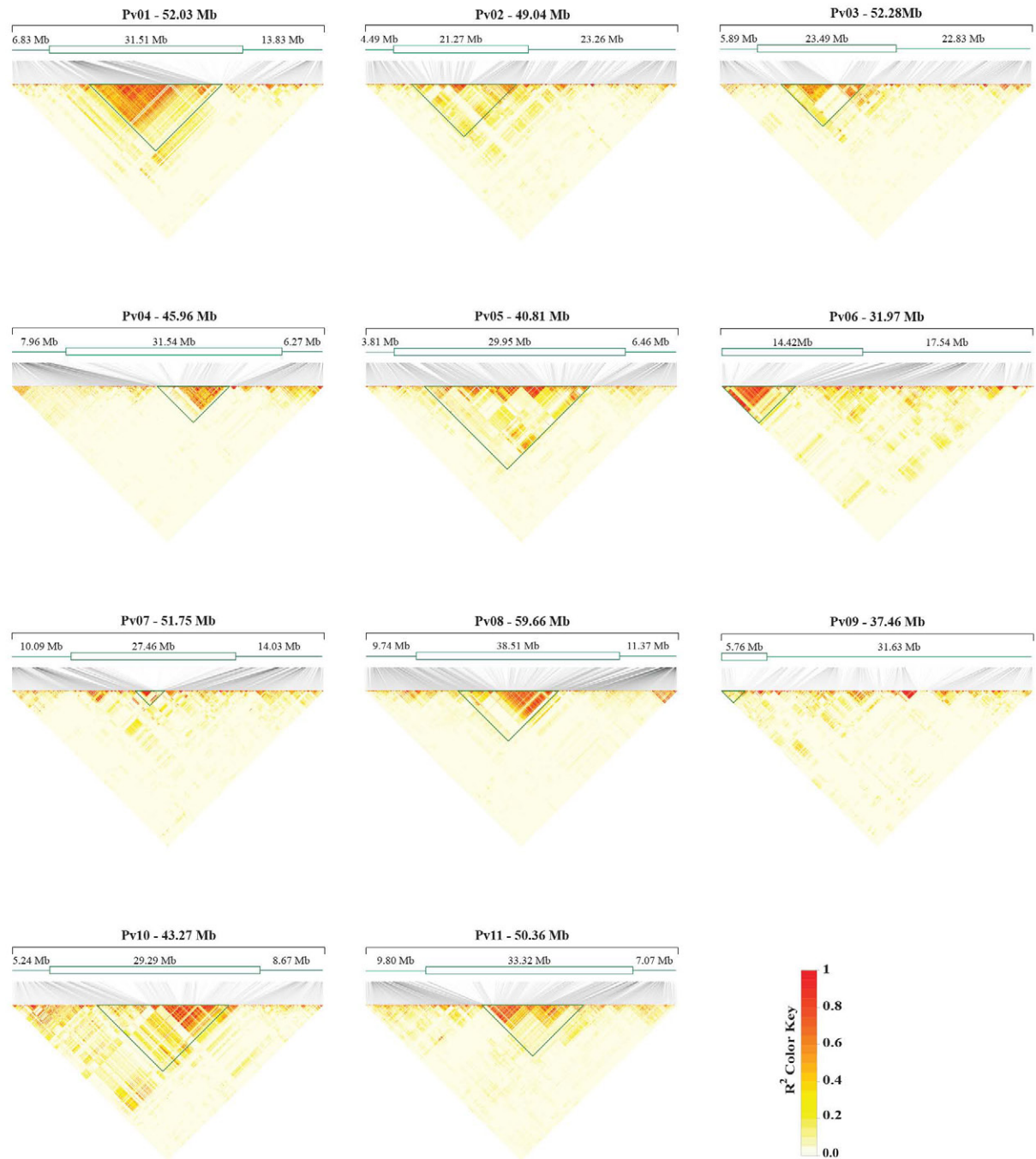


Fig. 2 (continued from previous page).

very little interchromosomal LD was detected in the MA subpopulation (Fig. 3a, 3b).

Phenotypic Analysis

The phenotypic expression of all agronomic traits varied between locations and subpopulations. The longest and shortest days to flower occurred in North Dakota (49–68 d) and Michigan (35–55 d), respectively. The average days

to flower across all the locations was 46 and 49 d in the MA and DJ subpopulations, respectively. Days to maturity ranged from 60–125 d. The plants matured earlier in Michigan (96–110 d) than in North Dakota (75–125 d). Seed weight showed a bimodal distribution with values ranging from 14.1 to 39.7 g 100 seeds⁻¹ for race MA genotypes and 21.1 to 53.7 g 100 seeds⁻¹ for race DJ genotypes. Figure 4 shows the distribution of each trait across

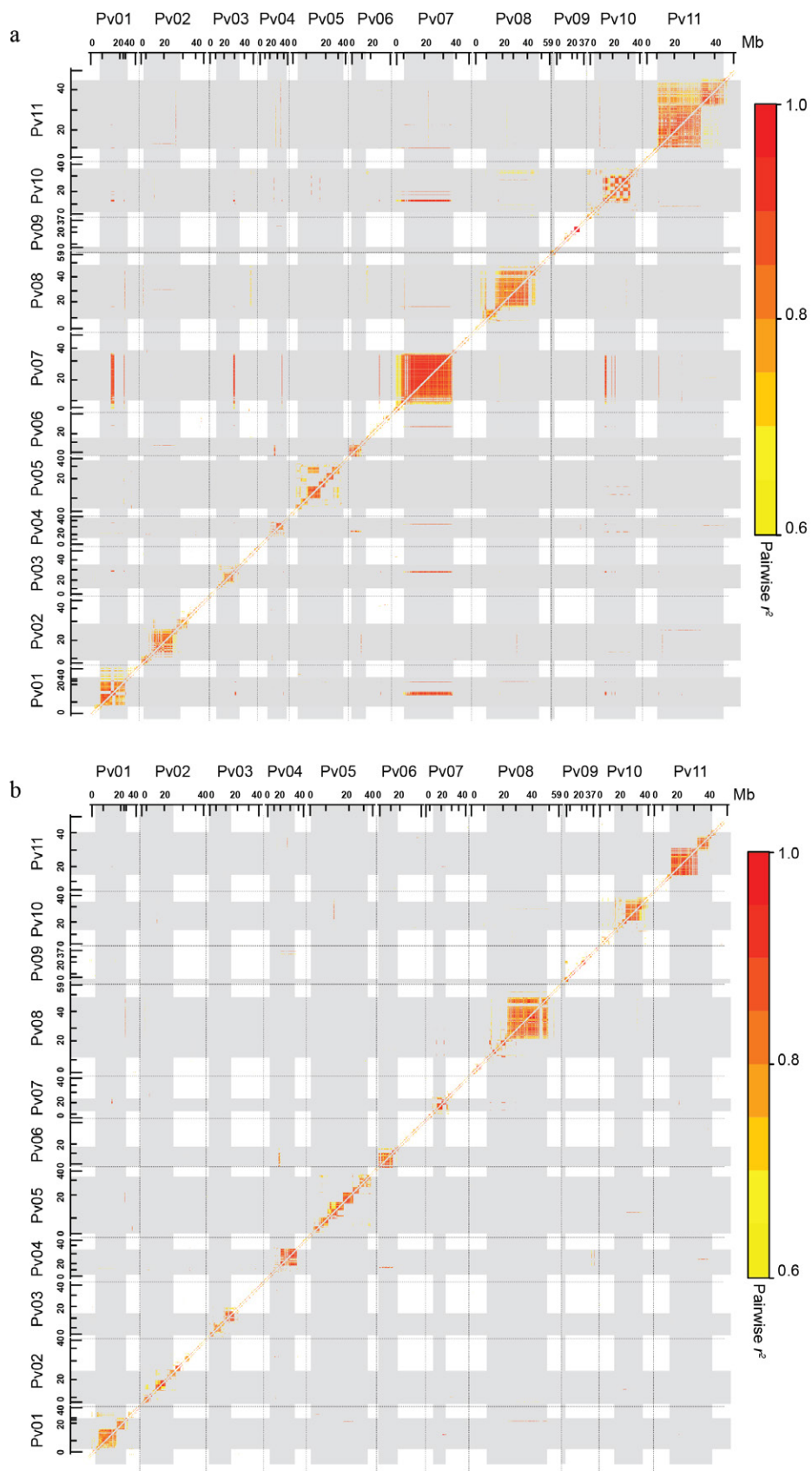


Fig. 3. Genome-wide linkage disequilibrium heatmap in races (a) Durango–Jalisco and (b) Mesoamerican. Data above the diagonal represent the null model, and data below the diagonal image represents the model that accounts for individual relatedness. A subset of single-nucleotide polymorphisms spaced every 50 kb was used, and only pairwise $r^2 > 0.6$ are shown. The gray rectangular areas show the pericentromeric regions. Gray dashed lines define the chromosome boundaries.

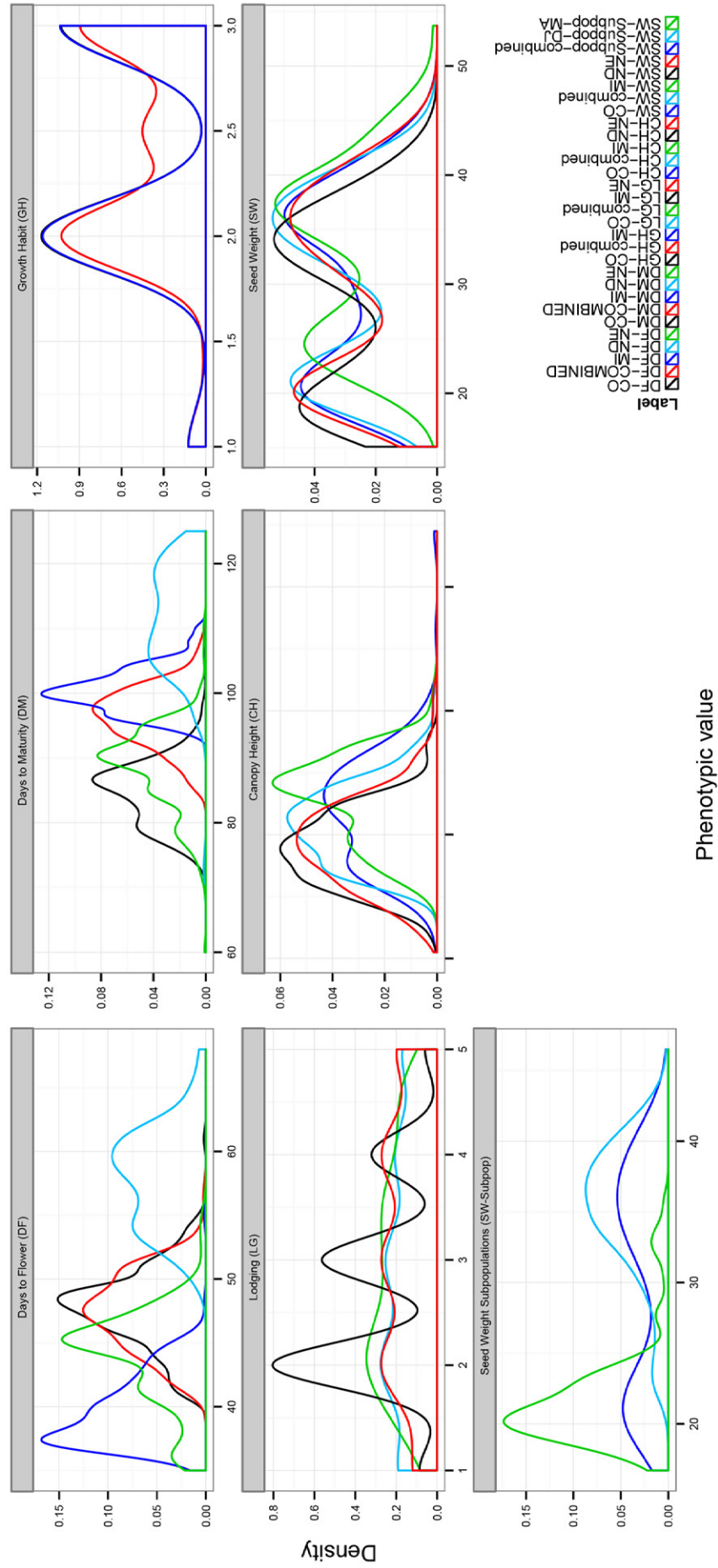


Fig. 4. Density plot of phenotypic distribution of six agronomic traits. Color coding is provided in the label section. CO, Colorado; MI, Michigan; NE, Nebraska; ND, North Dakota; Combined, across all the locations; DJ, Durango/Jalisco; MA, Mesoamerican.

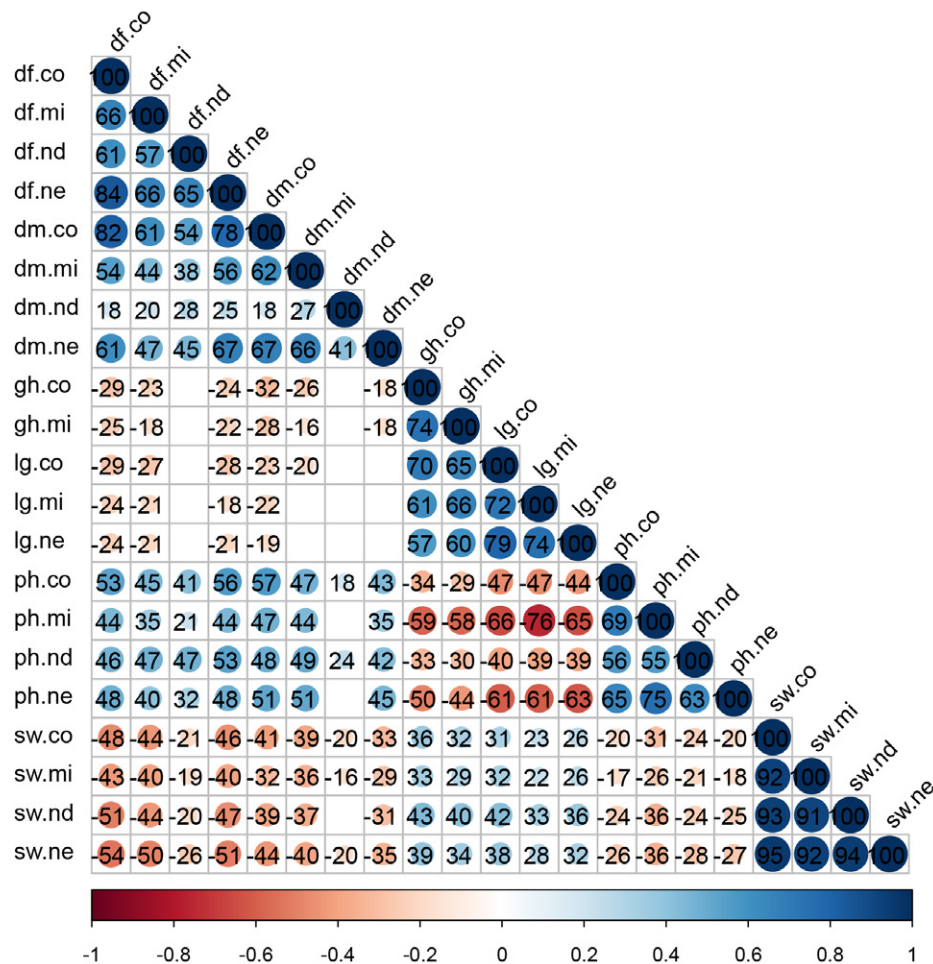


Fig. 5. Correlation between traits and locations based on adjusted means. DF, days to flower; DM, days to maturity; GH, growth habit; LG, lodging; CH, canopy height; SW, seed weight; CO, Colorado; MI, Michigan; ND, North Dakota; NE, Nebraska. Cells with correlation values not significant at P -value < 0.01 are left blank.

Table 2. Heritability of agronomic traits across four locations. The heritability values were calculated by the method proposed by Holland et al. (2003).

Trait	Durango/Jalisco		Mesoamerican	
	Plot basis	Family mean basis	Plot basis	Family mean basis
Days to flower	0.54	0.87	0.53	0.88
Days to maturity	0.35	0.71	0.16	0.53
Growth habit	0.68	0.88	0.65	0.86
Lodging	0.56	0.85	0.44	0.78
Canopy height	0.51	0.87	0.40	0.81
Seed weight	0.74	0.94	0.81	0.96

different locations and combined across locations. Trait values were highly correlated among locations except for days to maturity in North Dakota. Days to maturity was positively correlated ($r > 0.5$) with days to flower among different locations. Growth habit and lodging were positively correlated, while canopy height was negatively correlated with growth habit and lodging (Fig. 5). Table 2 shows that heritability was moderate to high for each trait in races DJ and MA.

Genome-Wide Association Study

Multiple genomic regions were found to be associated with the six agronomic traits. The growth habit data was analyzed with and without type I determinate genotypes because the strong Pv01 signal associated with the determinacy locus reduced the number of loci at other positions associated with the type II and III phenotypes that passed the cutoff value. The P -values for the best model varied among traits, locations, and subpopulations. This suggested that the number and effect size varied among the loci that affected the different traits. Therefore, a single P -value cutoff would not be a suitable approach to identify relevant SNPs among all traits. Rather, the P -values were bootstrapped 1000 times for each trait, and SNPs falling in the top 0.01 percentile of the empirical distribution were considered significant (Mamidi et al., 2014). This led to a different cutoff threshold for each trait and sometimes the cutoff was even more stringent than the Bonferroni value. Supplemental Table S1 shows the significant SNPs and the candidate gene models within 50 Kb upstream and downstream of the significant marker. For each trait, the amount of variation explained by a SNP was only calculated for those SNPs for which a candidate gene was identified.

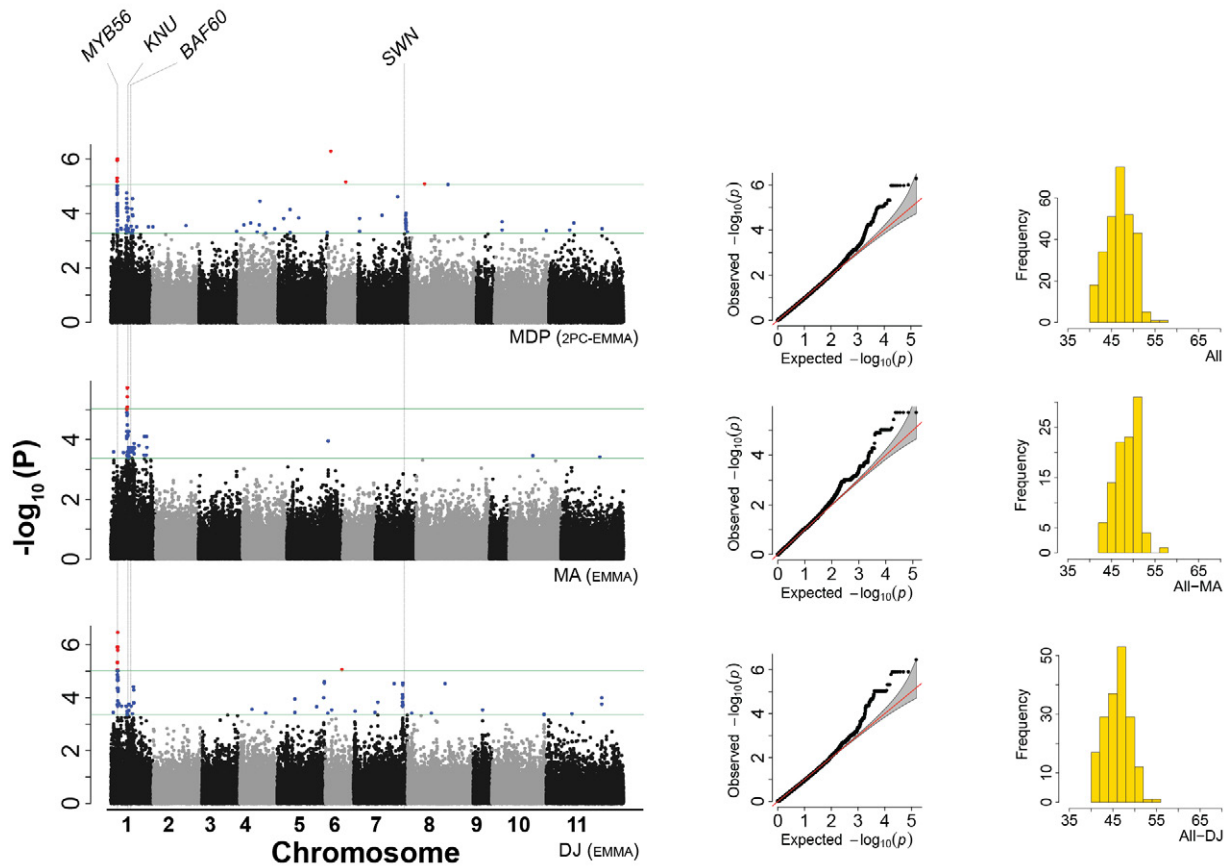


Fig. 6. Manhattan plots of the best models for flowering time in the Middle American diversity panel (MDP) and Mesoamerican (MA), Durango-Jalisco (DJ) races across all the locations. The two major peaks on Pv01 (8 and 15 Mb) vary between races. The best model is indicated in parenthesis. The green lines are the cutoff values to call a significant peak. Single-nucleotide polymorphisms that pass the 0.01 percentile are highlighted in red, and those falling between 0.01 and 0.1 percentile are highlighted in blue. The quantile–quantile plot shows the goodness of fit of the best model, and the histogram shows the trait distribution for adjusted means across all the locations in that population.

For days to flower, collectively, all candidate SNPs explained 14% of the phenotypic variation for the entire MDP and across all the locations analyzed. The strongest signal for days to flower was located on Pv01 and includes two major peaks in a block that spans from 8 to 20 Mb. The most significant GWAS peak on Pv01 differed among the races. The most significant GWAS peak for race DJ occurred at Pv01 (8 Mb), while for race MA, the peak SNP was located at Pv01 (15 Mb) (Fig. 6). It is difficult to determine whether both regions are equally important in both subpopulations or one of the regions is the main peak in each race and the other peak is due to extensive LD. Indeed, the 8-Mb region is in LD ($r^2 \approx 0.4$) with the 15-Mb region after controlling for both population structure and relatedness. To further investigate this region, we used the MLM (Segura et al., 2012) for the MDP population and the DJ and MA subpopulations. The most significant SNP controlled all the variation on Pv01 when used as a cofactor in the MLM analysis. The same result was observed at the subpopulation level. The interaction effect between the candidate SNPs at these two regions was not significant.

For days to maturity, GWAS peaks were noted across the genome and with different peaks at each location and for each subpopulation. Collectively, all candidate SNPs explained 21% of the variation. The most noticeable peaks are on Pv04 and Pv11. The peak at the end of Pv11 is from North Dakota, and the peak at the beginning of Pv11 is from Nebraska and Michigan. The peak on Pv04 originates from Michigan and has mainly a DJ origin. There is a Nebraska-specific peak on Pv01 and a Colorado-specific peak on Pv07 (Supplemental Fig. S1d–f). Most of the candidate genes found for this trait are associated with flowering time, senescence, or nutrient remobilization.

For the entire MDP population, a major GWAS signal was noted on Pv01 for growth habit where the flowering time candidate genes *LWD1*, *SPY*, and *TFL-1* are located. All candidate SNPs on Pv01 collectively explained 42% of the phenotypic variation. When the determinate type I genotypes were excluded from the growth habit analysis, the Pv01 peak disappeared and major peaks appeared at Pv04 (3 Mb), Pv06 (30 Mb), Pv07 (46 Mb), and Pv11 (10 and 43 Mb) (Fig. 7). The candidate SNPs across the genome and those on Pv07

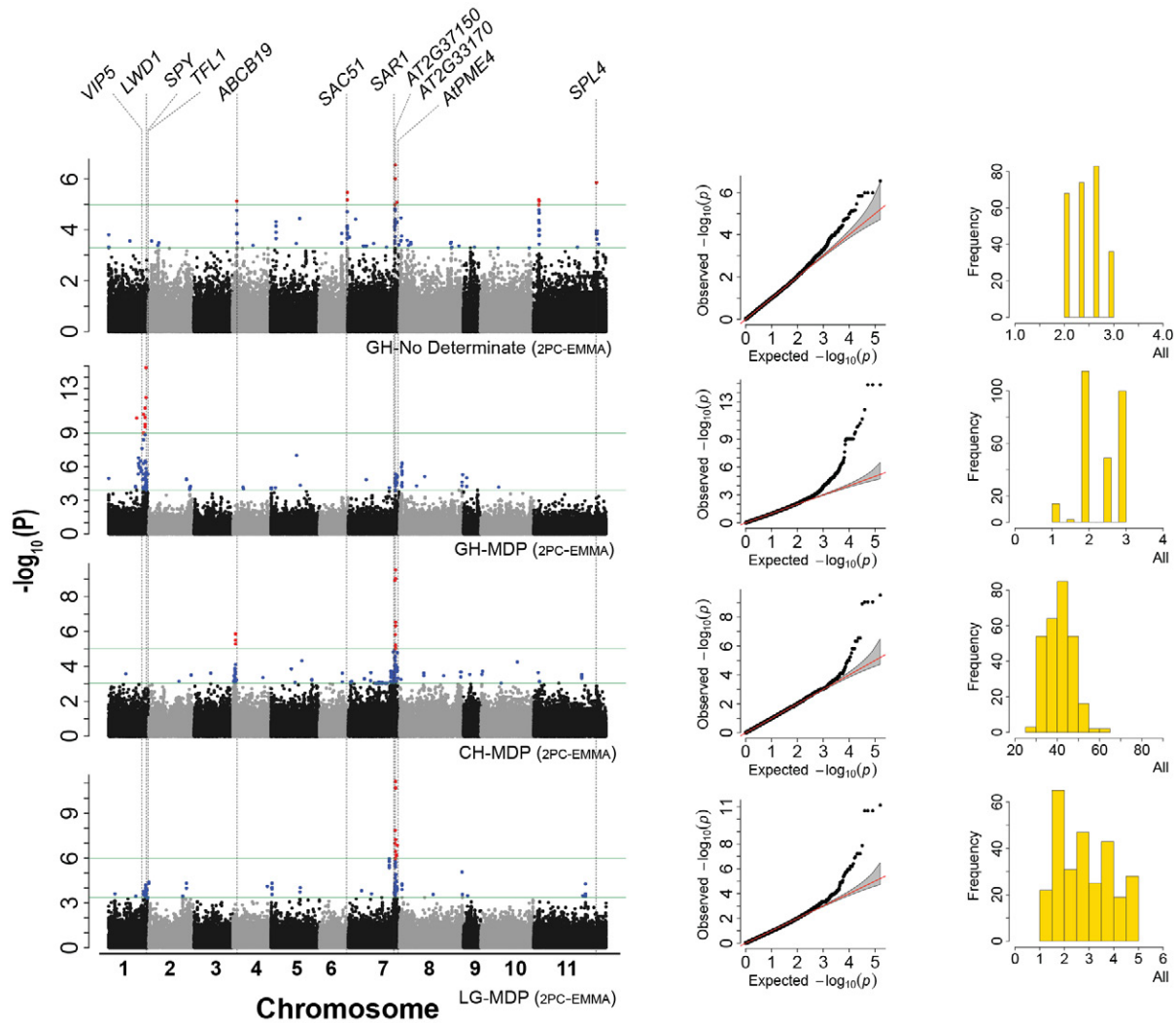


Fig. 7. Manhattan plots of the best models for plant architecture related traits (LG, lodging; CH, canopy height; GH, growth habit). The peak on Pv07 (46 Mb) is in common among all three traits when the determinate genotypes are excluded from the population for growth habit analysis. Growth habit encompasses major peaks from both lodging and canopy height. Growth habit using the entire Middle American diversity panel (MDP) shows a strong signal at the end of Pv01, which captures the determinacy characteristic of genotypes with type I growth habit and masks the peak on Pv07. This peak is shared with a flowering time peak in Michigan and is close to flowering time candidate genes such as *fin* locus (*TFL-1*). The best model is indicated in parenthesis. The green lines are the cutoff values to call a peak significant. Single-nucleotide polymorphisms that pass the 0.01 percentile are highlighted in red and those falling between 0.01 and 0.1 percentile are highlighted in blue. The quantile–quantile plot shows the goodness of fit of the best model, and the histogram shows the trait distribution for adjusted means across all locations in that population.

explained 33 and 12% of the phenotypic variation, respectively.

For lodging, a very strong and consistent Pv07 (46 Mb) peak was observed at all locations as well as across all locations. The peak is 68.6 Kb wide and consists of multiple SNPs, with pairwise r^2 LD values >0.5 . There are two smaller signals at Pv07 (45 Mb) and Pv07 (48 Mb) that are not in LD with the Pv07 (46 Mb) region (Fig. 8). There is another peak at 35.42 Mb, which falls in the 0.1 percentile tail of the empirical distribution of the P -values. All of the Pv07 variation is accounted for when SNPs at both Pv07 (46.11 Mb) and Pv07 (35.42 Mb) are included as cofactors in a MLM analysis. Using SNPs at 48 or 45 Mb as cofactors did not control for all the Pv07 variation. The

candidate SNPs on Pv07 collectively explained 21% of the phenotypic variation. A significant SNP at the proximal end of Pv05 was observed in Michigan, but no candidate genes were identified (Supplemental Fig. S1m–o).

The same lodging Pv07 (46 Mb) region was significant for canopy height across all locations (Fig. 7) and in each location except for North Dakota (Supplemental Fig. S1p–r). We also found a major peak on Pv04 for this trait that was not present for the lodging analysis. However, this peak did not remain significant when the SNP at Pv07 (46 Mb) was used as a cofactor in MLM.

Major peaks for seed weight were observed on Pv10 across all locations and individually in Colorado and Michigan (Supplemental Fig. S1s–u). This signal is mainly

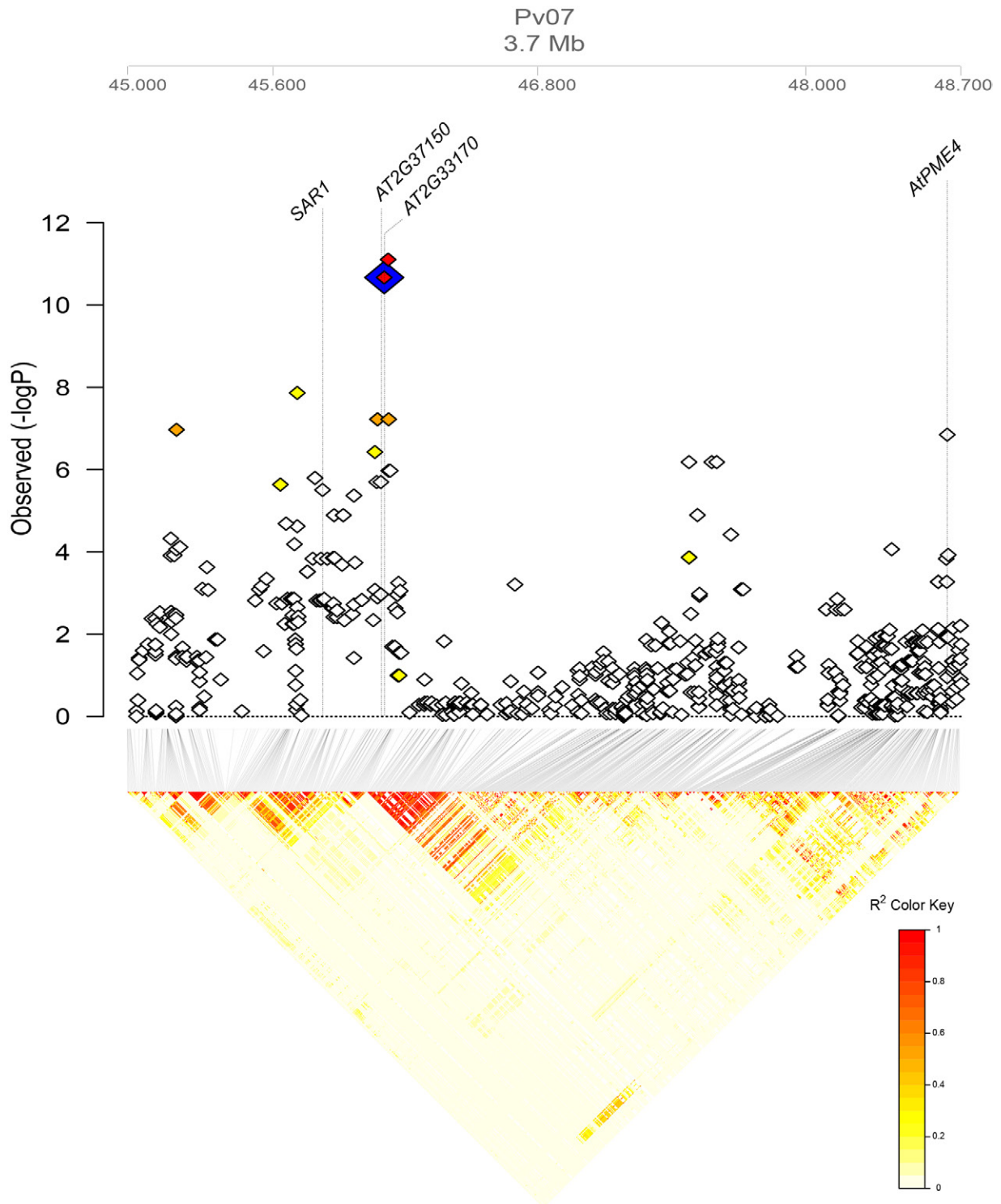


Fig. 8. Manhattan plot and linkage disequilibrium (LD) heatmap over 3.7-Mb region (45.03 – 48.37 Mb) on Pv07 for lodging. This region includes the three significant peaks (Pv07 [45.18 Mb], Pv07 [46.11 Mb], and Pv07 [48.63 Mb]) and their associated candidate gene homologs in *Arabidopsis*. The marker at 46.11 Mb, which resides in *Phvul.007G221700*, is colored in blue in the Manhattan plot and the rest of the markers in this region are color coded based on their pairwise r^2 value with this marker. Single-nucleotide polymorphisms (SNPs) with $r^2 < 0.25$ are white colored, those with $0.25 \leq r^2 \leq 0.5$ are orange, and SNPs with $r^2 > 0.5$ are colored in red. The r^2 values in Manhattan plot and LD heatmap are corrected for population structure and relatedness based on the best model using all the SNPs with minor allele frequency $\geq 5\%$.

from the DJ subpopulation, while no significant peak in the MA subpopulation was observed. The MLM analysis did not select any signals for seed weight. This might imply

the presence of other environmental or confounding variables that affect the causal loci nonlinearly and therefore make signal detection difficult.

Discussion

Population Structure

The STRUCTURE analysis, when clustered according to race and market classes, revealed patterns of admixture. Market classes in common bean are distinguished based on the seed size, shape, and color. Therefore, admixed lines within a market class would occur based on only a few traits. Admixture observed among the genotypes within market classes at $K = 7$ reflects the breeding strategies used in bean improvement. During the first half of the 21st century, bean production, and therefore breeding in United States, was regional, and crosses were made within market classes or between closely related market classes in that region. At that time, pinto and great northern bean were grown in the Central Great Plains, and in the far West, pink, great northern, red Mexican, and pinto bean were the main bean market classes (McClellan et al., 1993). Even today, breeders commonly make crosses among black, small red, and navy bean in race Mesoamerica and between pinto and great northern bean in race Durango. Black bean has tropical origin and are more recent to North America. In Central America, black bean has been used to improve red bean by introgressing yield and pest-resistance traits. Thus, for some red bean lines, the majority of their pedigree consists of black bean parents (Beebe et al., 1995). Moreover, breeding for new red and pink bean varieties in recent decades involved the integration of plant architectural traits from the Mesoamerican black and navy bean market classes (Kelly, 2001; Vandemark et al., 2014). The pink, red, and pinto bean were landraces in the early 1900s that were crossed among each other primarily to introgress disease resistance traits. Even though some lines are admixed, general clustering occurs by market class (Miklas, 2000).

Patterns of Linkage Disequilibrium

Linkage disequilibrium patterns vary by specific chromosome regions and between populations (Taillon-Miller et al., 2000; Huttley et al., 1999; Laan and Pääbo, 1997; Huang et al., 2010). Evidence shows that smaller populations show higher level of LD (Laan and Pääbo, 1997; Flint-Garcia et al., 2003). Within the MDP, race MA is a smaller subpopulation than DJ, which could have resulted in higher levels of LD as well as lower polymorphism rate within that race. Many markers with long-range interchromosomal LD in the MDP and DJ subpopulation were monomorphic in MA subpopulation or had a MAF < 5%. On the other hand, the differences in LD between specific regions of the genome might be stronger than the effect of population on LD (Zavattari et al., 2000). For example, both subpopulations exhibited higher LD in the pericentromeric than the ends of the chromosomes. In contrast to populations in linkage equilibrium, where LD only depends on the relative rate of mutation and recombination, the patterns of LD within the MA and DJ subpopulations depend on (i) the evolutionary steps that led to the divergence within

the wild Mesoamerican gene pool, (ii) domestication events within that gene pool, (iii) the development of the landraces, and (iv) the effects of artificial selection for beneficial alleles specific to each race during variety development (Gepts and Debouck, 1991; Kwak and Gepts, 2009; Hamblin et al., 2011; Bitocchi et al., 2012; Mamidi et al., 2013; Schmutz et al., 2014). All these factors result in changes in allele frequencies as a result of genetic drift or selection. Some allelic combinations can be lost, leading to higher LD between the remaining alleles. Strong selection during breeding in conjunction with local adaptation can create LD between unlinked loci, which does not decay with a longer genetic distance (Long et al., 2013). The SNP allele frequencies and LD patterns within the two subpopulations in the MDP clearly varied. Any conclusion on how the demographic factors shaped the LD pattern in this population needs in-depth population genetic studies using data from higher-pass sequencing of genotypes.

Linkage Disequilibrium affects the Interpretation of Genome-Wide Association Study Results

The GWAS for days to flower within the MDP identified two SNP peaks that mapped near each other (~8 and ~15 Mb on Pv01). These two peaks are located in the low-recombination pericentromeric region of the chromosome, and the LD value was $r^2 = 0.8$ with the null model, a value that dropped to $r^2 = 0.4$ when LD was corrected for population structure and individual relatedness. An important question is whether these Pv01 peaks were significant because of LD or whether they both might contain genes associated with the trait. To evaluate the possibilities, we compared the GWAS results with those obtained with MLM. At the 0.01% cutoff, a group of linked SNPs were significant at ~15 Mb in the MA subpopulation, while within the DJ subpopulation, a group of linked SNPs were significant at ~8 Mb. The best MLM forward and backward stepwise regression model for the MA subpopulation only included the 15-Mb peak SNPs, while the best MLM model for the DJ subpopulation only included the 8-Mb peak SNPs. Therefore, the result observed with the full MDP was actually the union of the effects of the two subpopulations and is not a spurious result resulting from LD.

Interrelated Traits show Colocalized Genome-Wide Association Study Peaks

Multiple traits exhibited colocalized GWAS signals. Most prominent among these was a Pv07 peak observed in nearly all test locations for lodging, canopy height, and growth habit (when determinate genotypes are excluded). This region spans the 45- to 48-Mb region and is the main peak for lodging and canopy height. Stepwise regression defined the 46-Mb region on Pv07 as the most significant region for these three interrelated traits. This could indicate the existence of a gene with a pleiotropic effect for the three traits or that these are interrelated subtraits associated with plant architecture. The Pv07 (46

Mb) peak contains a cluster of SNPs within two genes that could be considered as candidate genes. A single SNP within *Phvul.007G221800*, a leucine-rich repeat receptor-like protein kinase, causes a nonsynonymous substitution (S740T). This gene model is homologous to *BRASSINOSTEROID INSENSITIVE 1 PRECURSOR* (29904.t000125) in castor bean (*Ricinus communis* L.) (76% similarity; Goodstein et al., 2012) and *Arabidopsis thaliana* (L.) Heynh. *At2g33170* (63% identity). *BRI1* is a membrane-localized leucine-rich repeat receptor-like kinase that perceives and conveys the brassinosteroid signal (Clouse, 2002; Peng and Li, 2003; Thummel and Chory, 2002; Wang and He, 2004). Plants defective in *BRI1* express dwarfism. The *Arabidopsis* homolog (*At2g33170*) affects root length through cytokinin (ten Hove et al., 2011), but there are no reports on its effect on shoot length. The second gene at this location encodes a *RING/U-box* superfamily protein (*Phvul.007G221700*) with multiple SNPs, but the function of this gene has not been shown to be associated with plant architecture.

Another case of colocalized GWAS peak was observed for days to flower and growth habit in Michigan when determinate type I genotypes were included in the growth habit analysis. A Pv01 (45 Mb) GWAS peak is shared between both traits and is near a *TFL-1* homolog (*Phvul.001G189200*) that maps to the *fin* locus for determinacy in common bean (Repinski et al., 2012). The most significant SNP in this peak has a MAF of 0.05 for growth habit. This implies the peak is mainly capturing the effect of the 19 determinate genotypes and not the type II and III growth habit genotypes. The syntenic gene to the *fin* locus also affected plant height in a GWAS in soybean [*Glycine max* (L.) Merr.] (Zhang et al., 2015). The only additional growth habit candidate near this region is *VIP5*, which, in *Arabidopsis*, affects flowering time by promoting *FLOWER LOCUS C* (*FLC*) activation, which in turn represses flowering (Oh et al., 2004). When determinate genotypes were excluded from the growth habit analysis, five major GWAS peaks appeared: Pv07 (46 Mb), Pv11 (10 and 43 Mb), Pv06 (30 Mb), and Pv04 (3 Mb). The Pv07 (46 Mb) peak and its colocalization with other traits is discussed above and the candidate genes found on the other chromosomes are discussed in the Candidate Genes section that follows.

Candidate Genes

Days to Flower

Flowering is controlled by a complex network of genes that collectively integrate the photoperiod, vernalization, gibberellin, autonomous, and flowering time clock pathways. These pathways merge into two floral integrator genes, *FLOWERING LOCUS T* (*FT*) and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*) that promote flowering in *Arabidopsis* (Fornara et al., 2010; Chen et al., 2015). Although neither of these genes was associated with a days-to-flowering GWAS peak, multiple genes involved in the flowering pathways were identified as candidate genes. One candidate, *Phvul.001G064600* (Pv01 [8 Mb]), is an ortholog of *MYB56*, a gene that negatively

regulates *FT* (Chen et al., 2015). The peak marker for the df1.1 days-to-flowering QTL (Blair et al., 2006) also maps ~1 Mb from the *MYB56* candidate. Another candidate, *Phvul.001G192200*, is located near the Pv01 (45 Mb) GWAS peak and is an ortholog of *LIGHT-REGULATED WDI* (*LWD1*). Alternate alleles at this gene change the expression periodicity of genes in the circadian rhythm pathway leading to earlier flowering (Wu et al., 2008). This acceleration of flowering is mediated through *CONSTANS* (*CO*), which activates the integrator *FT*. A second days-to-flowering candidate gene at the same Pv01 (45 Mb) GWAS peak, *Phvul.001G192300*, is a *SPINDLY* (*SPY*) ortholog and is located immediately adjacent to *LWD1*. *SPINDLY* delays flowering by suppressing the gibberellic acid pathway and directly interacts with *GIGANTEA* (*GI*), a gene within the photoperiod pathway (Tseng et al., 2004). *Phvul.001G094300*, an ortholog of *Arabidopsis BRG-1 ASSOCIATED FACTOR 60* (*AtBAF60*), maps near the Pv01 (20 Mb) GWAS peak. *AtBAF60* is a subunit of the *SWI/SNF* polycomb complex, a complex that regulates the photoperiod and vernalization pathway by directly interacting with and suppressing *FLC* (Jégu et al., 2014). *FLC* is a flowering-time suppressor that functions upstream of *FT* and *SOC1*. *Phvul.007G213900*, a homolog of *Arabidopsis SWINGER* (*SWN*), maps near the Pv07 (45 Mb) peak. *SWINGER* is a component of the polycomb repressive complex 2 (*PRC2*) that promotes flowering via *FT* through the suppression of the flowering suppressor *FLC* (Wood et al., 2006; Jiang et al., 2008). Polycomb repressive complex 2 not only affects flowering time, but also affects floral meristem development. The release of *PRC2* from a repressive complex that includes *KNUCKLES* (*KNU*) leads to determinate flower development. The bean ortholog of *KNU*, *Phvul.001G087900*, maps near the Pv01 (15 Mb) days-to-flowering GWAS peak. When transitioning to a determinate floral meristem, *KNU* is activated and functions in a feed-back loop that promotes determinate development. This results in flower set over a defined time period (Lenhard et al., 2001; Payne et al., 2004; Sun et al., 2009, 2014; Liu et al., 2011; Wu et al., 2012; Mozgova et al., 2015). Figure 9 maps the common bean days-to-flowering candidate genes relative to the *Arabidopsis* flowering pathway.

Days to Maturity

Since the biological pathway that controls progress toward maturity is not fully elucidated, we searched for candidate genes known (or hypothesized) to be involved in senescence, nutrient remobilization, and seed development. Flowering time candidates were also considered because of the positive correlation between days to flowering and days to maturity (Córdoba et al., 2010; Blair et al., 2012). The GWAS peak at Pv11 (4.3 Mb) maps within the bean ortholog of the *Arabidopsis TARGET of RAPAMYCIN* (*TOR*), *Phvul.011G050300*. *TOR* is a member of a signaling pathway that perceives the nutrient and energy status of the plant (Wullschleger et al., 2006; Kim and Guan, 2011) and controls senescence and nutrient recycling by regulating *SERINE/THREONINE*

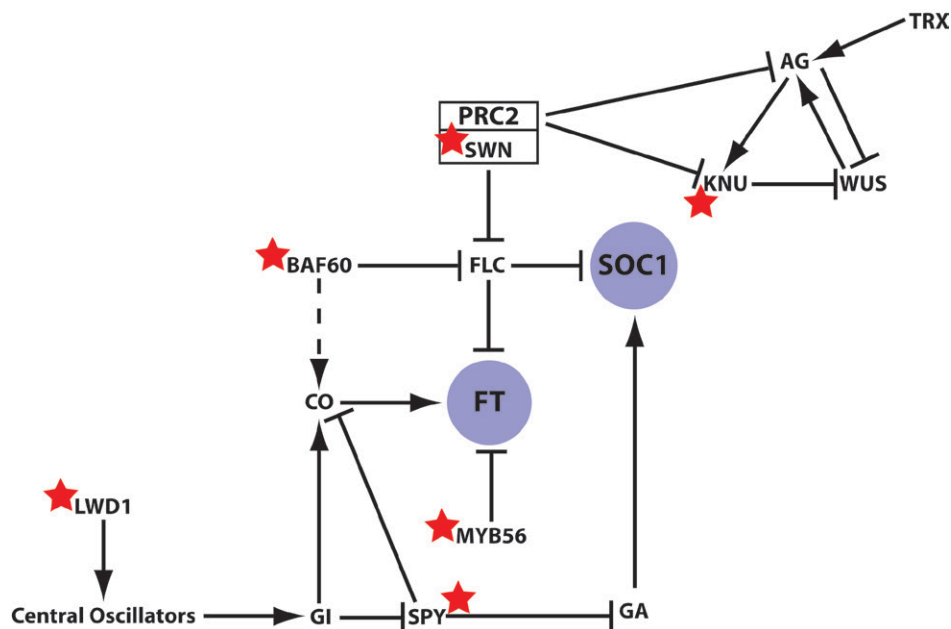


Fig. 9. A simplified schematic illustration of the *Arabidopsis* flowering time pathways. Only pathways and genes related to our candidate genes are shown. The common bean orthologs are denoted with a star.

PROTEIN PHOSPHATASE 2A (PP2A) activity (MacKintosh, 1992). Enhanced TOR expression increases organ and cell size and seed production, two biological processes required for the plants migration toward maturity (Deprost et al., 2007). *Phvul.004G011400*, an ortholog of *ATMYB118*, is a candidate gene at the Pv04 (20 Mb) GWAS peak. *ATMYB118* is an *Arabidopsis* transcription factor that regulates seed maturation at the spatial level by suppressing maturation related genes within the endosperm of seeds and maintains the embryo as the major nutrient sink (Barthole et al., 2014).

Three days-to-maturity candidate genes function in flowering pathways. *Phvul.011G053300*, maps near the Pv11 (4 Mb) peak and is an ortholog of the *Arabidopsis* *FPA* gene. *FPA* is a member of the autonomous flowering pathway (Koornneef et al., 1991), and genotypes overexpressing *FPA* flower earlier in short days and are insensitive to photoperiod (Schomburg, 2001). Its role in maturation is unclear, but early flowering is often correlated with early maturity. The bean ortholog of *REDUCED VERNALIZATION RESPONSE 1 (VRN1)*, *Phvul.011G050800*, maps to the same Pv11 (4 Mb) peak as *FPA*. Levy et al. (2002) reported that constitutive overexpression of *VRN1* led to early flowering without vernalization in *Arabidopsis*. Another days-to-maturity candidate gene related to flowering is *Phvul.011G158300*, which maps at the Pv11 (41 Mb) peak. This gene is homologous to one of the two *Arabidopsis* *SHL* genes (*At4g22140*), which encode a transcriptional regulator and chromatin remodeling protein. Genotypes overexpressing *SHL* led to early flowering and senescence (Müssig and Altmann, 2003).

Growth Habit, Lodging, and Canopy Height

Plant architecture is a complex trait that involves many hormone and shoot branching pathways that affect main stem

length. A major signal in the Pv11 (43 Mb) peak contains four significant SNPs located inside *Phvul.011G164800*, an ortholog of *Arabidopsis* *SPL4*. *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL)* is one regulator of *TB1/BRC1* transcription factor that integrates the hormone and branching pathways. Three SNPs are in complete LD with each other, and two of these cause nonsynonymous substitutions (Cys145Gly and Ser146Asn). In *Arabidopsis*, *SPL3*, *SPL4*, and *SPL5* act redundantly to affect the timing of the juvenile to adult phase transition (Wu and Poethig, 2006; Schwarz et al., 2008; Wang et al., 2009; Poethig, 2009; Yamaguchi et al., 2009). *SPL* genes in rice (*Oryza sativa* L.) and alfalfa (*Medicago sativa* L.) have been reported to cause architectural changes, such as lodging, internode length, branching, and biomass (Jiao et al., 2010; Aung, 2014), and have a pleiotropic effect on shoot branching, plant height, and floral induction. These pleiotropic effects of the candidate gene might explain why early efforts to create early flowering type II architecture in common bean were not successful (Kelly, 2001). The *Arabidopsis* and bean homologs both contain a *miR156* cleavage site in their 3' untranslated region, and branching may be regulated by the *miR156-SPL* pathway.

Three other candidate genes act in the auxin pathway upstream of cytokinin and strigolactone to regulate the auxiliary bud outgrowth through *TB1/BRC1* (Müller and Leyser, 2011; Rameau et al., 2014). *Phvul.006G203400*, at the Pv06 (30 Mb) peak, is an ortholog of *SUPPRESSOR OF ACAULIS 51 (SAC51)*, which encodes a bHLH (basic helix-loop-helix) transcription factor that controls stem elongation in *Arabidopsis*. *SAC51* uORF-mediated translational regulation is affected by thersmospermine, which limits the auxin signaling that promote xylem differentiation. The GWAS peak at Pv07 (45.8 Mb) includes a candidate gene (*Phvul.007G218900*) homologous to *SUPPRESSOR OF AUXIN RESISTANCE 1 (SARI)* and

appears to control stem thickness. *Sar1* is epistatic to *AUXIN RESISTANCE 1 (axr1)* and together the two genes increase plant height and internode distance in *Arabidopsis* (Cernac et al., 1997; Parry et al., 2006). The GWAS peak at Pv04 (2.9 Mb) is located in the candidate gene *Phvul.004G027800*, a homolog to *ABCBI9 (MDR1)* an ABC transporter protein that mediates polar auxin transport in stems and roots (Noh et al., 2001; Kaneda et al., 2011). This gene affects apical dominance (Noh et al., 2001), hypocotyl gravitropism, and inflorescence stem phototropism and curvature, all factors associated with plant architecture (Noh et al., 2003; Kumar et al., 2011).

As noted above, several lodging and canopy-height peaks were collocated. Another example is the Pv07 (48 Mb) GWAS peak that is located in *Phvul.007G246700*. This gene encodes a homolog of *AtPME41*, an invertase–pectin methylesterase inhibitor. This protein changes the methylesterification level of pectin, a modification that affects cell wall rigidity (Micheli, 2001; Castillejo et al., 2004). This may be a factor that favors an upright, nonlodged architecture. Similarly, a GWAS on soybean agronomic traits reported a pectin lyase-like superfamily protein as a candidate gene for plant height (Zhang et al., 2015).

Seed Weight

Seed weight is a complex trait that encompasses many developmental and physiological processes, such as plant architecture, seed development, and metabolic efflux, to enhance sink strength (Van Camp, 2005). There was no clear signal in the MA subpopulation for seed weight, but Pv08 and Pv10 showed strong peaks in the DJ subpopulation. The Pv10 candidate gene, *Phvul.010G017600*, affects metabolic efflux and protein storage in seed. This gene is a homolog of *Arabidopsis ALPHA-AMYLASE LIKE GENE (AMY1)* that is associated with the onset of suspensor and endosperm programmed cell death and early nutrient mobilization to nourish the growing embryo (Mansfield and Briarty, 1994; Sreenivasulu and Wobus, 2013). The significant peak at Pv08 (1 Mb) within the DJ subpopulation contains *Phvul.008G013300*, a subtilase gene family member that is homologous to *SBT1.1* in *Arabidopsis*. In legumes, this gene is specifically expressed during the early stages of endosperm development when the embryo cells are still dividing. It is hypothesized that it provides developmental signals or regulates metabolic flux to control embryo growth by affecting the embryo cell number and therefore might have a role in regulating seed size (Abirached-Darmency et al., 2005; Melkus et al., 2009; D'Erfurth et al., 2012). *Phvul.006G069300*, at the Pv06 (18 Mb) GWAS peak in the Michigan environment, is homologous to *Arabidopsis ASN1*, a glutamine-dependent asparagine synthase 1 and a domestication gene in the Middle American gene pool (Schmutz et al., 2014). *Arabidopsis* transgenic plants overexpressing *ASN1* show increased N content, higher seed-soluble protein content, and increased seed weight (Lam et al., 2003). Many QTL analyses in common bean have located seed-weight QTL on the same chromosomes as described here (Beattie et al., 2003; Blair

et al., 2006, 2012; Pérez-Vega et al., 2010; Wright and Kelly, 2011; Linares-Ramirez, 2013).

Conclusion

Genome-wide association studies using a diverse panel and a large set of SNPs was able to dissect the genetic architecture of important economic traits in common bean and define promising candidate genes for six agronomic traits. These candidates are involved in many different molecular biochemical, hormone, developmental, and transcriptional regulation pathways. Adding phenotypic data from multiple years to this data set can improve the accuracy of the results. Moreover, increasing the number of individuals in the MA subpopulation can yield more polymorphic SNPs and hence facilitate the discovery of MA specific signals for some traits. It is noteworthy that the GWAS yielded more promising results for traits with higher heritability than those that are affected more by the environment. Moreover, the extent of LD, especially for signals falling in the heterochromatic region, is wide in this population, which made candidate gene discovery more challenging. The population and genotyping tools developed during this project are now available to study many other phenotypes that are important for the productivity of common bean. These tools also can provide an additional starting point for the discovery, validation, and functional investigation of genes that control the many unique features of common bean that are also shared by other legume species.

Acknowledgments

Financial support provided by USDA, National Institute of Food and Agricultural (NIFA), Agriculture and Food Research Initiative (AFRI), Project #2009-01929

References

- Abirached-Darmency, M., M.R. Abdel-Gawwad, G. Conejero, J.L. Verdeil, and R. Thompson. 2005. In situ expression of two storage protein genes in relation to histo-differentiation at mid-embryogenesis in *Medicago truncatula* and *Pisum sativum* seeds. *J. Exp. Bot.* 56:2019–2028. doi:10.1093/jxb/eri200
- Appels, R., R. Barrero, and M. Bellgard. 2013. Advances in biotechnology and informatics to link variation in the genome to phenotypes in plants and animals. *Funct. Integr. Genomics* 13:1–9. doi:10.1007/s10142-013-0319-2
- Atwell, S., Y.S. Huang, B.J. Vilhjálmsson, G. Willems, M. Horton, Y. Li, D. Meng, A. Platt, A.M. Tarone, T.T. Hu, R. Jiang, N.W. Mulyati, X. Zhang, M.A. Amer, I. Baxter, B. Brachi, J. Chory, C. Dean, M. Debieu, J. de Meaux, J.R. Ecker, N. Faure, J.M. Kniskern, J.D.G. Jones, T. Michael, A. Nemri, F. Roux, D.E. Salt, C. Tang, M. Todesco, M.B. Traw, D. Weigel, P. Marjoram, J.O. Borevitz, J. Bergelson, and M. Nordborg. 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465:627–631. doi:10.1038/nature08800
- Aulchenko, Y.S., S. Ripke, A. Isaacs, and C.M. van Duijn. 2007. GenABEL: An R library for genome-wide association analysis. *Bioinformatics* 23:1294–1296. doi:10.1093/bioinformatics/btm108
- Aung, B. 2014. Effects of microRNA156 on flowering time and plant architecture in *Medicago sativa*. MS thesis. Univ. of Western Ontario, London, ON, Canada.
- Balasubramanian, S., C. Schwartz, A. Singh, N. Warthmann, M.C. Kim, J.N. Maloof, O. Loudet, G.T. Trainer, T. Dabi, J.O. Borevitz, J. Chory, and D. Weigel. 2009. QTL mapping in new *Arabidopsis thaliana*

- advanced intercross-recombinant inbred lines. *PLoS One* 4:e4318. doi:10.1371/journal.pone.0004318
- Barthole, G., A. To, C. Marchive, V. Brunaud, L. Soubigou-Taconnat, N. Berger, B. Dubreucq, L. Lepiniec, and S. Baud. 2014. MYB118 represses endosperm maturation in seeds of *Arabidopsis*. *Plant Cell* 26:3519–3537. doi:10.1105/tpc.114.130021
- Beattie, A.D., J. Larsen, T.E. Michaels, and K.P. Pauls. 2003. Mapping quantitative trait loci for a common bean (*Phaseolus vulgaris* L.) ideotype. *Genome* 46:411–422. doi:10.1139/g03-015
- Beebe, S.E., I. Ochoa, P. Skroch, J. Nienhuis, and J. Tivang. 1995. Genetic diversity among common bean breeding lines developed for Central America. *Crop Sci.* 35:1178–1183. doi:10.2135/cropsci1995.0011183X003500040045x
- Beebe, S., P.W. Skroch, J. Tohme, M.C. Duque, F. Pedraza, and J. Nienhuis. 2000. Structure of genetic diversity among common bean landraces of Middle American origin based on correspondence analysis of RAPD. *Crop Sci.* 40:264–273. doi:10.2135/cropsci2000.401264x
- Bitocchi, E., L. Nanni, E. Bellucci, M. Rossi, A. Giardini, P.S. Zeuli, G. Logozzo, J. Stougaard, P. McClean, G. Attene, and R. Papa. 2012. Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proc. Natl. Acad. Sci. USA* 109:E788–E796. doi:10.1073/pnas.1108973109
- Blair, M.W., L.M. Díaz, H.F. Buendía, and M.C. Duque. 2009. Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 119:955–972. doi:10.1007/s00122-009-1064-8
- Blair, M.W., C.H. Galeano, E. Tovar, M.C. Muñoz Torres, A.V. Castrillón, S.E. Beebe, and I.M. Rao. 2012. Development of a Mesoamerican intra-genepool genetic map for quantitative trait loci detection in a drought tolerant × susceptible common bean (*Phaseolus vulgaris* L.) cross. *Mol. Breed.* 29:71–88. doi:10.1007/s11032-010-9527-9
- Blair, M.W., G. Iriarte, and S. Beebe. 2006. QTL analysis of yield traits in an advanced backcross population derived from a cultivated Andean × wild common bean (*Phaseolus vulgaris* L.) cross. *Theor. Appl. Genet.* 112:1149–1163. doi:10.1007/s00122-006-0217-2
- Broughton, W.J., G. Hernández, M. Blair, S. Beebe, P. Gepts, and J. Vanderleyden. 2003. Beans (*Phaseolus* spp.): Model food legumes. *Plant Soil* 252:55–128. doi:10.1023/A:1024146710611
- Castillejo, C., J.I. de la Fuente, P. Iannetta, M.A. Botella, and V. Valpuesta. 2004. Pectin esterase gene family in strawberry fruit: Study of FaPE1, a ripening-specific isoform. *J. Exp. Bot.* 55:909–918. doi:10.1093/jxb/erh102
- Cernac, A., C. Lincoln, D. Lammer, and M. Estelle. 1997. The *SARI* gene of *Arabidopsis* acts downstream of the *AXR1* gene in auxin response. *Development* 124:1583–1591.
- Checa, O.E., and M.W. Blair. 2012. Inheritance of yield-related traits in climbing beans (*Phaseolus vulgaris* L.). *Crop Sci.* 52:1998. doi:10.2135/cropsci2011.07.0368
- Chen, L., A. Bernhardt, J. Lee, and H. Hellmann. 2015. Identification of *Arabidopsis* MYB56 as a novel substrate for CRL3^{BPM} E3 ligases. *Mol. Plant* 8:242–250. doi:10.1016/j.molp.2014.10.004
- Cichy, K.A., J.A. Wiesinger, and F.A. Mendoza. 2015. Genetic diversity and genome-wide association analysis of cooking time in dry bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 128:1555–1567. doi:10.1007/s00122-015-2531-z
- Clouse, S.D. 2002. Brassinosteroid signal transduction. *Mol. Cell* 10:973–982. doi:10.1016/S1097-2765(02)00744-X
- Córdoba, J.M., C. Chavarro, J.A. Schlueter, S.A. Jackson, and M.W. Blair. 2010. Integration of physical and genetic maps of common bean through BAC-derived microsatellite markers. *BMC Genomics* 11:436. doi:10.1186/1471-2164-11-436
- D'Erfurth, I., C. Le Signor, G. Aubert, M. Sanchez, V. Vernoud, B. Darchy, J. Lherminier, V. Bourion, N. Bouteiller, A. Bendahmane, J. Buitink, J.M. Prospero, R. Thompson, J. Burstin, and K. Gallardo. 2012. A role for an endosperm-localized subtilase in the control of seed size in legumes. *New Phytol.* 196:738–751. doi:10.1111/j.1469-8137.2012.04296.x
- Deprost, D., L. Yao, R. Sormani, M. Moreau, G. Leterreux, M. Nicolai, M. Bedu, C. Robaglia, and C. Meyer. 2007. The *Arabidopsis* TOR kinase links plant growth, yield, stress resistance and mRNA translation. *EMBO Rep.* 8:864–870. doi:10.1038/sj.embor.7401043
- Díaz, L.M., and M.W. Blair. 2006. Race structure within the Mesoamerican gene pool of common bean (*Phaseolus vulgaris* L.) as determined by microsatellite markers. *Theor. Appl. Genet.* 114:143–154. doi:10.1007/s00122-006-0417-9
- Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi:10.1371/journal.pone.0019379
- Flint-Garcia, S.A., J.M. Thornsberry, and E.S. Buckler. 2003. Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54:357–374. doi:10.1146/annurev.arplant.54.031902.134907
- Fornara, F., A. de Montaigu, and G. Coupland. 2010. SnapShot: Control of flowering in *Arabidopsis*. *Cell* 141:550.e1–2. doi:10.1016/j.cell.2010.04.024
- Freyre, R., R. Ríos, L. Guzmán, D.G. Debouck, and P. Gepts. 1996. Ecogeographic distribution of *Phaseolus* spp. (Fabaceae) in Bolivia. *Econ. Bot.* 50:195–215. doi:10.1007/BF02861451
- Gepts, P., and F.A. Bliss. 1986. Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. *Econ. Bot.* 40:469–478. doi:10.1007/BF02859660
- Gepts, P., and D. Debouck. 1991. Origin, domestication, and evolution of the common bean (*Phaseolus vulgaris* L.). In: A. van Schoonhoven and O. Yovest, editors, *Common beans: Research for crop improvement*. CAB International, Wallingford, UK. p. 7–53.
- Goodstein, D.M., S. Shu, R. Howson, R. Neupane, R.D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam, and D.S. Rokhsar. 2012. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* 40:D1178–D1186. doi:10.1093/nar/gkr944
- Hamblin, M.T., E.S. Buckler, and J.L. Jannink. 2011. Population genetics of genomics-based crop improvement methods. *Trends Genet.* 27:98–106. doi:10.1016/j.tig.2010.12.003
- Hedrick, U.P., W.T. Tapley, G.P. van Eseltine, and W.D. Enzie. 1931. The vegetables of New York. Vol. 1, Part II. Beans of New York. J.B. Lyons & Co., New York.
- Holland, J.B., Nyquist, W.E., and C.T. Cervantes-Martínez. 2003. Estimating and interpreting heritability for plant breeding: An update. *Plant Breed. Rev.* 22:9–112.
- Huang, X., X. Wei, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, C. Li, C. Zhu, T. Lu, Z. Zhang, and M. Li. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42:961–967. doi:10.1038/ng.695
- Huttley, G.A., M.W. Smith, M. Carrington, and S.J. O'Brien. 1999. A scan for linkage disequilibrium across the human genome. *Genetics* 152:1711–1722.
- Hyten, D.L., Q. Song, E.W. Fickus, C.V. Quigley, J.S. Lim, I.Y. Choi, E.Y. Hwang, M. Pastor-Corrales, and P.B. Cregan. 2010. High-throughput SNP discovery and assay development in common bean. *BMC Genomics* 11:475. doi:10.1186/1471-2164-11-475
- Jégu, T., D. Latrasse, M. Delarue, H. Hirt, S. Domenichini, F. Ariel, M. Crespi, C. Bergounioux, C. Raynaud, and M. Benhamed. 2014. The BAF60 subunit of the SWI/SNF chromatin-remodeling complex directly controls the formation of a gene loop at FLOWERING LOCUS C in *Arabidopsis*. *Plant Cell* 26:538–551. doi:10.1105/tpc.113.114454
- Jiang, D., Y. Wang, Y. Wang, and Y. He. 2008. Repression of FLOWERING LOCUS C and FLOWERING LOCUS T by the *Arabidopsis* Polycomb repressive complex 2 components. *PLoS One* 3:e3404. doi:10.1371/journal.pone.0003404
- Jiao, Y., Y. Wang, D. Xue, J. Wang, M. Yan, G. Liu, G. Dong, D. Zeng, Z. Lu, X. Zhu, Q. Qian, and J. Li. 2010. Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. *Nat. Genet.* 42:541–544. doi:10.1038/ng.591
- Kamfwa, K., K.A. Cichy, and J.D. Kelly. 2015a. Genome-wide association analysis of symbiotic nitrogen fixation in common bean. *Theor. Appl. Genet.* 128:1999–2017. doi:10.1007/s00122-015-2562-5
- Kamfwa, K., K.A. Cichy, and J.D. Kelly. 2015b. Genome-wide association study of agronomic traits in common bean. *Plant Genome* 8. doi:10.3835/plantgenome2014.09.0059

- Kaneda, M., M. Schuetz, B.S.P. Lin, C. Chanis, B. Hamberger, T.L. Western, J. Ehrling, and A.L. Samuels. 2011. ABC transporters coordinately expressed during lignification of *Arabidopsis* stems include a set of ABCBs associated with auxin transport. *J. Exp. Bot.* 62:2063–2077. doi:10.1093/jxb/erq416
- Kang, H.M., N.A. Zaitlen, C.M. Wade, A. Kirby, D. Heckerman, M.J. Daly, and E. Eskin. 2008. Efficient control of population structure in model organism association mapping. *Genetics* 178:1709–1723. doi:10.1534/genetics.107.080101
- Kelly, J.D. 2001. Remaking bean plant architecture for efficient production. *Adv. Agron.* 71:109–143. doi:10.1016/S0065-2113(01)71013-9
- Kelly, J.D., J.M. Kolkman, and K. Schneider. 1998. Breeding for yield in dry bean (*Phaseolus vulgaris* L.). *Euphytica* 102:343–356. doi:10.1023/A:1018392901978
- Khairallah, M.M., M.W. Adams, and B.B. Sears. 1990. Mitochondrial DNA polymorphisms of Malawian bean lines: Further evidence for two major gene pools. *Theor. Appl. Genet.* 80:753–761. doi:10.1007/BF00224188
- Kim, J., and K.L. Guan. 2011. Amino acid signaling in TOR activation. *Annu. Rev. Biochem.* 80:1001–1032. doi:10.1146/annurev-biochem-062209-094414
- Koenig, R., and P. Gepts. 1989. Allozyme diversity in wild *Phaseolus vulgaris*: Further evidence for two major centers of genetic diversity. *Theor. Appl. Genet.* 78:809–817. doi:10.1007/BF00266663
- Koinange, E.M.K., and P. Gepts. 1992. Hybrid weakness in wild *Phaseolus vulgaris* L. *J. Hered.* 83:135–139.
- Koornneef, M., C.J. Hanhart, and J.H. van der Veen. 1991. A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Mol. Gen. Genet.* 229:57–66. doi:10.1007/BF00264213
- Kornegay, J., J.W. White, and O.O. de la Cruz. 1992. Growth habit and gene pool effects on inheritance of yield in common bean. *Euphytica* 62:171–180. doi:10.1007/BF00041751
- Korte, A., and A. Farlow. 2013. The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* 9:29. doi:10.1186/1746-4811-9-29
- Kumar, P., K.D.L. Millar, and J.Z. Kiss. 2011. Inflorescence stems of the *mdr1* mutant display altered gravitropism and phototropism. *Environ. Exp. Bot.* 70:244–250. doi:10.1016/j.envexpbot.2010.09.019
- Kwak, M., and P. Gepts. 2009. Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theor. Appl. Genet.* 118:979–992. doi:10.1007/s00122-008-0955-4
- Laan, M., and S. Pääbo. 1997. Demographic history and linkage disequilibrium in human populations. *Locus* 238:260.
- Lam, H.M., P. Wong, H.K. Chan, K.M. Yam, L. Chen, C.M. Chow, and G.M. Coruzzi. 2003. Overexpression of the *ASN1* gene enhances nitrogen status in seeds of *Arabidopsis*. *Plant Physiol.* 132:926–935. doi:10.1104/pp.103.020123
- Leakey, C.L.A. 1988. Genetic resources of *Phaseolus* beans. In: P. Gepts, editor, Genetic resources of *Phaseolus* beans. Springer, Netherlands, Dordrecht. p. 245–327. doi:10.1007/978-94-009-2786-5_12
- Lenhard, M., A. Bohnert, G. Jürgens, and T. Laux. 2001. Termination of stem cell maintenance in *Arabidopsis* floral meristems by interactions between *WUSCHEL* and *AGAMOUS*. *Cell* 105:805–814. doi:10.1016/S0092-8674(01)00390-7
- Levy, Y.Y., S. Mesnage, J.S. Mylne, A.R. Gendall, and C. Dean. 2002. Multiple roles of *Arabidopsis VRN1* in vernalization and flowering time control. *Science* 297:243–246. doi:10.1126/science.1072147
- Li, H., Z. Peng, X. Yang, W. Wang, J. Fu, J. Wang, Y. Han, Y. Chai, T. Guo, N. Yang, J. Liu, M.L. Warburton, Y. Cheng, X. Hao, P. Zhang, J. Zhao, Y. Liu, G. Wang, J. Li, and J. Yan. 2013. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.* 45:43–50. doi:10.1038/ng.2484
- Linares-Ramirez, A. 2013. Selection of dry bean genotypes adapted for drought tolerance in the northern Great Plains. Ph.D. diss. North Dakota State Univ., Fargo
- Lipka, A.E., M.A. Gore, M. Magallanes-Lundback, A. Mesberg, H. Lin, T. Tiede, C. Chen, C.R. Buell, E.S. Buckler, T. Rocheford, and D. DellaPenna. 2013. Genome-wide association study and pathway-level analysis of tocopherol levels in maize grain. *G3: Genes, Genomes, Genet.* 3:1287–1299. doi:10.1534/g3.113.006148
- Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, M.A. Gore, E.S. Buckler, and Z. Zhang. 2012. GAPIT: Genome association and prediction integrated tool. *Bioinformatics* 28:2397–2399. doi:10.1093/bioinformatics/bts444
- Liu, X., Y.J. Kim, R. Müller, R.E. Yumul, C. Liu, Y. Pan, X. Cao, J. Goodrich, and X. Chen. 2011. *AGAMOUS* terminates floral stem cell maintenance in *Arabidopsis* by directly repressing *WUSCHEL* through recruitment of Polycomb Group proteins. *Plant Cell* 23:3654–3670. doi:10.1105/tpc.111.091538
- Long, Q., F.A. Rabanal, D. Meng, C.D. Huber, A. Farlow, A. Platzer, Q. Zhang, B.J. Vilhjálmsson, A. Korte, V. Nizhynska, and V. Voronin. 2013. Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nat. Genet.* 45:884–890. doi:10.1038/ng.2678
- MacKintosh, C. 1992. Regulation of spinach-leaf nitrate reductase by reversible phosphorylation. *Biochim. Biophys. Acta, Mol. Cell Res.* 1137:121–126.
- Mamidi, S., S. Chikara, R.J. Goos, D.L. Hyten, D. Annam, S.M. Moghaddam, R.K. Lee, P.B. Cregan, and P.E. McClean. 2011a. Genome-wide association analysis identifies candidate genes associated with iron deficiency chlorosis in soybean. *Plant Genome* 4:154–164. doi:10.3835/plantgenome2011.04.0011
- Mamidi, S., R.K. Lee, J.R. Goos, and P.E. McClean. 2014. Genome-wide association studies identifies seven major regions responsible for iron deficiency chlorosis in soybean (*Glycine max*). *PLoS One* 9:e107469.
- Mamidi, S., M. Rossi, D. Annam, S. Moghaddam, R. Lee, R. Papa, and P. McClean. 2011b. Investigation of the domestication of common bean (*Phaseolus vulgaris*) using multilocus sequence data. *Funct. Plant Biol.* 38:953. doi:10.1071/FP11124
- Mamidi, S., M. Rossi, S.M. Moghaddam, D. Annam, R. Lee, R. Papa, and P.E. McClean. 2013. Demographic factors shaped diversity in the two gene pools of wild common bean *Phaseolus vulgaris* L. *Heredity* 110:267–276. doi:10.1038/hdy.2012.82
- Mangin, B., A. Siberchicot, S. Nicolas, A. Doligez, P. This, and C. Cierco-Ayrolles. 2012. Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity* 108:285–291. doi:10.1038/hdy.2011.73
- Mansfield, S.G., and L.G. Briarty. 1994. Endosperm development. In: J. Bowman, editor, *Arabidopsis: An atlas of morphology and development*. Springer-Verlag, New York. p. 385–397.
- McClean, P.E., J.R. Myers, and J.J. Hammond. 1993. Coefficient of parentage and cluster analysis of North American dry bean cultivars. *Crop Sci.* 33:190–197. doi:10.2135/cropsci1993.0011183X003300010034x
- Melkus, G., H. Rolletschek, R. Radchuk, J. Fuchs, T. Rutten, U. Wobus, T. Altmann, P. Jakob, and L. Borisjuk. 2009. The metabolic role of the legume endosperm: A noninvasive imaging study. *Plant Physiol.* 151:1139–1154. doi:10.1104/pp.109.143974
- Mensack, M.M., V.K. Fitzgerald, E.P. Ryan, M.R. Lewis, H.J. Thompson, and M.A. Brick. 2010. Evaluation of diversity among common beans (*Phaseolus vulgaris* L.) from two centers of domestication using “omics” technologies. *BMC Genomics* 11:686. doi:10.1186/1471-2164-11-686
- Micheli, F. 2001. Pectin methylesterases: Cell wall enzymes with important roles in plant physiology. *Trends Plant Sci.* 6:414–419. doi:10.1016/S1360-1385(01)02045-3
- Miklas, P.N. 2000. Use of *Phaseolus* germplasm in breeding pinto, great northern, pink, and red beans for the Pacific Northwest and Inter-Mountain Region. In: S. Singh, editor, *Bean research, production, and utilization*. Proceedings of the Idaho Bean Workshop. Twin Falls, ID. 3–4 Aug. 2000. University of Idaho, Moscow, ID. p. 13–29.
- Mozgova, I., C. Köhler, and L. Hennig. 2015. Keeping the gate closed: Functions of the polycomb repressive complex PRC2 in development. *Plant J.* 83:121–132. doi:10.1111/tpj.12828
- Müller, D., and O. Leyser. 2011. Auxin, cytokinin and the control of shoot branching. *Ann. Bot. (London, UK)* 107:1203–1212. doi:10.1093/aob/mcr069
- Müssig, C., and T. Altmann. 2003. Changes in gene expression in response to altered *SHL* transcript levels. *Plant Mol. Biol.* 53:805–820. doi:10.1023/B:PLAN.0000023661.65248.4b
- Noh, B., A. Bandyopadhyay, W.A. Peer, E.P. Spalding, and A.S. Murphy. 2003. Enhanced gravi- and phototropism in plant *mdr* mutants

- mislocalizing the auxin efflux protein PIN1. *Nature* 423:999–1002. doi:10.1038/nature01716
- Noh, B., A.S. Murphy, and E.P. Spalding. 2001. Multidrug resistance-like genes of *Arabidopsis* required for auxin transport and auxin-mediated development. *Plant Cell* 13:2441–2454. doi:10.1105/tpc.13.11.2441
- Oh, S., H. Zhang, P. Ludwig, and S. van Nocker. 2004. A mechanism related to the yeast transcriptional regulator Paf1c is required for expression of the *Arabidopsis* *FLC/MAF* MADS box gene family. *Plant Cell* 16:2940–2953. doi:10.1105/tpc.104.026062
- Parry, G., S. Ward, and A. Cernac. 2006. The *Arabidopsis* SUPPRESSOR OF AUXIN RESISTANCE proteins are nucleoporins with an important role in hormone signaling and development. *Plant Cell* 18:1590–1603. doi:10.1105/tpc.106.041566
- Payne, T., S.D. Johnson, and A.M. Koltunow. 2004. *KNUCKLES (KNU)* encodes a C2H2 zinc-finger protein that regulates development of basal pattern elements of the *Arabidopsis* gynoecium. *Development* 131:3737–3749. doi:10.1242/dev.01216
- Peng, P., and J. Li. 2003. Brassinosteroid signal transduction: A mix of conservation and novelty. *J. Plant Growth Regul.* 22:298–312. doi:10.1007/s00344-003-0059-y
- Pérez-Vega, E., A. Pañeda, C. Rodríguez-Suárez, A. Campa, R. Giraldez, and J.J. Ferreira. 2010. Mapping of QTLs for morpho-agronomic and seed quality traits in a RIL population of common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 120:1367–1380. doi:10.1007/s00122-010-1261-5
- Poethig, R.S. 2009. Small RNAs and developmental timing in plants. *Curr. Opin. Genet. Dev.* 19:374–378. doi:10.1016/j.gde.2009.06.001
- Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38:904–909. doi:10.1038/ng1847
- Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller, P. Sklar, P.I.W. de Bakker, M.J. Daly, and P.C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–575. doi:10.1086/519795
- Rameau, C., J. Bertheloot, N. Leduc, B. Andrieu, F. Foucher, and S. Sakr. 2014. Multiple pathways regulate shoot branching. *Front. Plant Sci.* 5:741.
- R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Stat. Comput., Vienna, Austria.
- Repinski, S.L., M. Kwak, and P. Gepts. 2012. The common bean growth habit gene *PvTFL1y* is a functional homolog of *Arabidopsis* *TFL1*. *Theor. Appl. Genet.* 124:1539–1547. doi:10.1007/s00122-012-1808-8
- Rosenberg, N.A. 2003. *distruct*: A program for the graphical display of population structure. *Mol. Ecol. Notes* 4:137–138. doi:10.1046/j.1471-8286.2003.00566.x
- Rosenberg, N.A., T. Burke, K. Elo, M.W. Feldman, P.J. Freidlin, M.A. Groenen, J. Hillel, A. Mäki-Tanila, M. Tixier-Boichard, A. Vignal, K. Wimmers, and S. Weigend. 2001. Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics* 159:699–713.
- SAS Institute. 2011. *Base SAS 9.3 procedures guide*. SAS Institute Inc. Cary, NC.
- Scheet, P., and M. Stephens. 2006. A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78:629–644. doi:10.1086/502802
- Schmutz, J., P.E. McClean, S. Mamidi, A.G. Wu, S.B. Cannon, J. Grimwood, J. Jenkins, S. Shu, Q. Song, C. Chavarro, M. Torres-Torres, V. Geffroy, S.M. Moghaddam, D. Gao, B. Abernathy, K. Barry, M. Blair, M.A. Brick, M. Chovatia, P. Gepts, D. Goodstein, M. Gonzales, U. Hellsten, D.L. Hyten, G. Jia, J.D. Kelly, D. Kudrna, R. Lee, M.M.S. Richard, P.N. Miklas, J.M. Osorno, J. Rodrigues, V. Thareau, C.A. Urrea, M. Wang, Y. Yu, M. Zhang, R.A. Wing, P.B. Cregan, D.S. Rokhsar, and S.A. Jackson. 2014. A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46:707–713. doi:10.1038/ng.3008
- Schomburg, F.M. 2001. *FPA*, a gene involved in floral induction in *Arabidopsis*, encodes a protein containing RNA-recognition motifs. *Plant Cell* 13:1427–1436. doi:10.1105/tpc.13.6.1427
- Schröder, S., S. Mamidi, R. Lee, M.R. McKain, P.E. McClean, and J.M. Osorno. 2016. Optimization of genotyping by sequencing (GBS) data in common bean (*Phaseolus vulgaris* L.). *Mol. Breed.* 36:6. doi:10.1007/s11032-015-0431-1
- Schwarz, S., A.V. Grande, N. Bujdosó, H. Saedler, and P. Huijser. 2008. The microRNA regulated SBP-box genes *SPL9* and *SPL15* control shoot maturation in *Arabidopsis*. *Plant Mol. Biol.* 67:183–195. doi:10.1007/s11103-008-9310-z
- Segura, V., B.J. Vilhjálmsson, A. Platt, A. Korte, Ü. Seren, Q. Long, and M. Nordborg. 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* 44:825–830. doi:10.1038/ng.2314
- Shin, J.H., S. Blay, B. McNeney, and J. Graham. 2006. LDheatmap: An R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J. Stat. Softw.* 16:1–10. doi:10.18637/jss.v016.c03
- Singh, S.P. 1989. Patterns of variation in cultivated common bean (*Phaseolus vulgaris*, Fabaceae). *Econ. Bot.* 43:39–57. doi:10.1007/BF02859324
- Singh, S.P., P. Gepts, and D.G. Debouck. 1991. Races of common bean (*Phaseolus vulgaris*, Fabaceae). *Econ. Bot.* 45:379–396. doi:10.1007/BF02887079
- Song, Q., G. Jia, D.L. Hyten, J. Jenkins, E.Y. Hwang, S.G. Schroeder, J.M. Osorno, J. Schmutz, S.A. Jackson, P.E. McClean, and P.B. Cregan. 2015. SNP assay development for linkage map construction, anchoring whole-genome sequence, and other genetic and genomic applications in common bean. *G3: Genes, Genomes, Genet.* 5:2285–2290. doi:10.1534/g3.115.020594
- Sreenivasulu, N., and U. Wobus. 2013. Seed-development programs: A systems biology-based comparison between dicots and monocots. *Annu. Rev. Plant Biol.* 64:189–217. doi:10.1146/annurev-arplant-050312-120215
- Stanton-Geddes, J., T. Paape, B. Epstein, R. Briskine, J. Yoder, J. Mudge, A.K. Bharti, A.D. Farmer, P. Zhou, R. Denny, G.D. May, S. Erlandson, M. Yakub, M. Sugawara, M.J. Sadowsky, N.D. Young, and P. Tiffin. 2013. Candidate genes and genetic architecture of symbiotic and agronomic traits revealed by whole-genome, sequence-based association genetics in *Medicago truncatula*. *PLoS One* 8:e65688. doi:10.1371/journal.pone.0065688
- Sun, B., L.S. Looi, S. Guo, Z. He, E.S. Gan, J. Huang, Y. Xu, W.Y. Wee, and T. Ito. 2014. Timing mechanism dependent on cell division is invoked by Polycomb eviction in plant stem cells. *Science* 343:1248559. doi:10.1126/science.1248559
- Sun, B., Y. Xu, K.H. Ng, and T. Ito. 2009. A timing mechanism for stem cell maintenance and differentiation in the *Arabidopsis* floral meristem. *Genes Dev.* 23:1791–1804. doi:10.1101/gad.1800409
- Sun, G., C. Zhu, M.H. Kramer, S.S. Yang, W. Song, H.P. Piepho, and J. Yu. 2010. Variation explained in mixed-model association mapping. *Heredity* 105:333–340. doi:10.1038/hdy.2010.11
- Taillon-Miller, P., I. Bauer-Sardiña, N.L. Saccone, J. Putzel, T. Laitinen, A. Cao, J. Kere, G. Pilia, J.P. Rice, and P.Y. Kwok. 2000. Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. *Nat. Genet.* 25:324–328. doi:10.1038/77100
- Tar'an, B., T.E. Michaels, and K.P. Pauls. 2002. Genetic mapping of agronomic traits in common bean. *Crop Sci.* 42:544–556. doi:10.2135/cropsci2002.0544
- ten Hove, C.A., Z. Bochdanovits, V.M.A. Jansweijer, F.G. Koning, L. Berke, G.F. Sanchez-Perez, B. Scheres, and R. Heidstra. 2011. Probing the roles of *LRR* *RLK* genes in *Arabidopsis thaliana* roots using a custom T-DNA insertion set. *Plant Mol. Biol.* 76:69–83. doi:10.1007/s11103-011-9769-x
- Thummel, C.S., and J. Chory. 2002. Steroid signaling in plants and insects-common themes, different pathways. *Genes Dev.* 16:3113–3129. doi:10.1101/gad.1042102

- Tseng, T.S., P.A. Salomé, C.R. McClung, and N.E. Olszewski. 2004. SPINDLY and GIGANTEA interact and act in *Arabidopsis thaliana* pathways involved in light responses, flowering, and rhythms in cotyledon movements. *Plant Cell* 16:1550–1563. doi:10.1105/tpc.019224
- Van Camp, W. 2005. Yield enhancement genes: Seeds for growth. *Curr. Opin. Biotechnol.* 16:147–153. doi:10.1016/j.copbio.2005.03.002
- Vandemark, G.J., M.A. Brick, J.M. Osorno, J.D. Kelly, and C.A. Urrea. 2014. Yield gains in major U.S. field crops. ASA, CSSA, and SSSA, Milwaukee, WI.
- Verslues, P.E., J.R. Lasky, T.E. Juenger, T.W. Liu, and M.N. Kumar. 2014. Genome-wide association mapping combined with reverse genetics identifies new effectors of low water potential-induced proline accumulation in *Arabidopsis*. *Plant Physiol.* 164:144–159. doi:10.1104/pp.113.224014
- Wang, R., S. Farrona, C. Vincent, A. Joecker, H. Schoof, F. Turck, C. Alonso-Blanco, G. Coupland, and M.C. Albani. 2009. *PEPI* regulates perennial flowering in *Arabidopsis alpina*. *Nature* 459:423–427. doi:10.1038/nature07988
- Wang, Z.Y., and J.X. He. 2004. Brassinosteroid signal transduction: Choices of signals and receptors. *Trends Plant Sci.* 9:91–96. doi:10.1016/j.tplants.2003.12.009
- Wei, T., and M.T. Wei. 2013. Package “corrplot.” *Statistician* 56:316–324.
- Wood, C.C., M. Robertson, G. Tanner, W.J. Peacock, E.S. Dennis, and C.A. Helliwell. 2006. The *Arabidopsis thaliana* vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. *Proc. Natl. Acad. Sci. USA* 103:14631–14636. doi:10.1073/pnas.0606385103
- Wright, E.M., and J.D. Kelly. 2011. Mapping QTL for seed yield and canning quality following processing of black bean (*Phaseolus vulgaris* L.). *Euphytica* 179:471–484. doi:10.1007/s10681-011-0369-2
- Wu, G., and R.S. Poethig. 2006. Temporal regulation of shoot development in *Arabidopsis thaliana* by *miR156* and its target *SPL3*. *Development* 133:3539–3547. doi:10.1242/dev.02521
- Wu, J.F., Y. Wang, and S.H. Wu. 2008. Two new clock proteins, LWD1 and LWD2, regulate *Arabidopsis* photoperiodic flowering. *Plant Physiol.* 148:948–959. doi:10.1104/pp.108.124917
- Wu, M.F., Y. Sang, S. Bezhani, N. Yamaguchi, S.K. Han, Z. Li, Y. Su, T.L. Slewinski, and D. Wagner. 2012. SWI2/SNF2 chromatin remodeling ATPases overcome polycomb repression and control floral organ identity with the LEAFY and SEPALLATA3 transcription factors. *Proc. Natl. Acad. Sci. USA* 109:3576–3581. doi:10.1073/pnas.1113409109
- Wulfschleger, S., R. Loewith, and M.N. Hall. 2006. TOR signaling in growth and metabolism. *Cell* 124:471–484. doi:10.1016/j.cell.2006.01.016
- Yamaguchi, A., M.F. Wu, L. Yang, G. Wu, R.S. Poethig, and D. Wagner. 2009. The microRNA-regulated SBP-Box transcription factor SPL3 is a direct upstream activator of *LEAFY*, *FRUITFULL*, and *APETALA1*. *Dev. Cell* 17:268–278. doi:10.1016/j.devcel.2009.06.007
- Yu, J., G. Pressoir, W.H. Briggs, I. Vroh Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler. 2005. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38:203–208. doi:10.1038/ng1702
- Zavattari, P., E. Deidda, M. Whalen, R. Lampis, A. Mulargia, M. Loddo, I. Eaves, G. Mastio, J.A. Todd, and F. Cucca. 2000. Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: Demography, chromosome recombination frequency and selection. *Hum. Mol. Genet.* 9:2947–2957. doi:10.1093/hmg/9.20.2947
- Zhang, J., Q. Song, P.B. Cregan, R.L. Nelson, X. Wang, J. Wu, and G.L. Jiang. 2015. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics* 16:217. doi:10.1186/s12864-015-1441-4
- Zhang, Z., E. Ersoz, C.Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, P.J. Bradbury, J. Yu, D.K. Arnett, J.M. Ordovas, and E.S. Buckler. 2010. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42:355–360. doi:10.1038/ng.546
- Zhao, J.H. 2007. gap: Genetic Analysis Package. *J. Stat. Softw.* 23. doi:10.18637/jss.v023.i08
- Zhao, K., C.W. Tung, G.C. Eizenga, M.H. Wright, M.L. Ali, A.H. Price, G.J. Norton, M.R. Islam, A. Reynolds, J. Mezey, A.M. McClung, C.D. Bustamante, and S.R. McCouch. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* 2:467. doi:10.1038/ncomms1467

Supplementary Table S1. Candidate genes in the 100 Kb region centered on a significant marker. The P -value is from the first population/location in column nine. The negative sign in column five indicates that the marker is downstream of the candidate gene and no sign indicates the marker is upstream of the candidate. Markers highlighted in gray fall in the 0.01 percentile tail of the empirical distribution of P -value in at least one population/location. The R^2_{LR} values are reported for the entire population (MDP) and across all the locations. MDP, Middle American Diversity panel; DJ, Durango/Jalisco subpopulation; MA, Mesoamerican subpopulation; CO, Colorado; MI, Michigan; NE, Nebraska; ND, North Dakota. If no location is specified for a population in column nine, all locations are considered in the analysis.

Pv	SNP position (bp)	$-\log_{10}$ (P -value)	Bean candidate gene	Marker distance from candidate gene (bp)	Arabidopsis gene model	Arabidopsis gene symbol	Arabidopsis annotation	Population/location in which is significant	R^2_{LR} in the best model
<i>Days to flowering</i>									
1	8,123,404	3.3	PhvuL001G064600	5,375	AT5G17800	AtMYB56,MYB56	-	MDP-CO,DJ,DJ-CO, DJ-NE	-
1	15,827,862	4.7	PhvuL001G087900	-5,006	AT5G14010	KNU	C2H2 and C2HC zinc fingers superfamily protein	MDP, MDP-CO, MDP-MI, MDP-ND, MA, MA-MI, MA-ND, DJ-CO	0.07
1	20,358,909	3.3	PhvuL001G094300	28,765	AT5G14170	CHC1 ,AtBAF60, AtSWP73B	SWIB/MDM2 domain superfamily protein	MDP, MDP-MI, DJ, DJ-CO, DJ- MI	0.04
1	45,815,561	3.9	PhvuL001G192200	6,809	AT1G12910	ATAN11,LWD1	Transducin/WD40 repeat-like superfamily protein	MDP-MI, DJ-MI	-
1	45,815,561	3.9	PhvuL001G192300	0	AT3G11540	SPY	Tetratricopeptide repeat (TPR)-like superfamily protein	MDP-MI,MI-DJ	-
7	45,299,031	3.5	PhvuL007G213900	0	AT4G02020	EZA1,SDG10,SWN	SET domain-containing protein	MDP, MDP-CO, MDP-NE, DJ, DJ-ND, DJ-NE	0.04

Supplementary Table S1. Candidate genes in the 100 Kb region centered on a significant marker (continued).

Pv	SNP position (bp)	$-\log_{10}$ (P-value)	Bean candidate gene	Marker distance from candidate gene (bp)	Arabidopsis gene model	Arabidopsis gene symbol	Arabidopsis annotation	Population/location in which is significant	R^2_{LR} in the best model
<i>Days to maturity</i>									
4	1,105,556	4.4	PhvuL004G011400	-20,714	AT3G27785	ATMYB118,MYB118,PGA37	myb domain protein 118	MDP, MDP-MI, DJ-MI	0.06
11	4,316,178	3.4	PhvuL011G050300	0	AT1G50030	TOR	target of rapamycin	MDP, MDP-NE	0.06
11	4,358,407	3.7	PhvuL011G050800	5,073	AT3G18990	REM39,VRN1	AP2/B3-like transcriptional factor family protein	MDP, MDP-NE	0.06
11	4,481,167	3.5	PhvuL011G053300	44,868	AT2G43410	FPA	RNA binding	MDP	0.06
11	41,693,254	4.0	PhvuL011G158300	0	AT4G22140	EBS	PHD finger family protein / bromo-adjacent homology (BAH) domain-containing protein	MDP, MDP-ND, DJ, DJ-ND	0.05
<i>Growth habit: with determinate genotypes</i>									
1	42,857,105	11.0	PhvuL001G167200	0	AT1G61040	VIP5	plus-3 domain-containing protein	MDP, MDP-CO, MDP-MI	0.19
1	45,593,025	5.8	PhvuL001G189200	29,699	AT5G03840	TFL-1,TFL1	PEBP (phosphatidylethanolamine-binding protein) family protein	MDP, MDP-CO, MDP-MI, MA, MA-CO, MA-MI	0.09
1	45,732,679	14.8	PhvuL001G191500	0	AT2G36200	-	P-loop containing nucleoside triphosphate hydrolases superfamily protein	MDP, MDP-CO, MDP-MI	0.26
1	45,795,950	4.4	PhvuL001G192200	-9,924	AT1G12910	ATAN11,LWD1	Transducin/WD40 repeat-like superfamily protein	MDP, MDP-CO, MA, MA-CO, MA-MI	0.07
1	45,795,950	4.4	PhvuL001G192300	19,239	AT3G11540	SPY	Tetratricopeptide repeat (TPR)-like superfamily protein	MDP, MDP-CO, MA, MA-CO, MA-MI	0.07
7	45,836,294	5.0	PhvuL007G218900	0	AT1G33410	ATNUP160,NUP160,SAR1	SUPPRESSOR OF AUXIN RESISTANCE1	MDP, MDP-CO, MDP-MI, DJ, DJ-CO, DJ-MI	0.07

Supplementary Table S1. Candidate genes in the 100 Kb region centered on a significant marker (continued).

Pv	SNP position (bp)	$-\log_{10}$ (P-value)	Bean candidate gene	Marker distance from candidate gene (bp)	Arabidopsis gene model	Arabidopsis gene symbol	Arabidopsis annotation	Population/location in which is significant	R^2_{LR} in the best model
7	46,132,395	5.3	PhvuI.007G221800	0	AT2G33170	-	Leucine-rich repeat receptor-like protein kinase family protein	MDP, MDP-CO, MDP-MI, DJ, DJ-CO, DJ-MI	0.08
7	48,630,699	5.0	PhvuI.007G246700	0	AT4G02330	ATPMEPCRB,AtPME41	Plant invertase/pectin methylesterase inhibitor superfamily	MDP, MDP-CO, MDP-MI, DJ, DJ-CO, DJ-MI	0.07
8	3,166,255	5.2	PhvuI.008G037900	10,802	AT5G61430	ANAC100,ATNAC5,NA C100	NAC domain containing protein 100	MDP, MDP-CO, MDP-MI, DJ, DJ-CO	0.07
<i>Growth habit: determinate genotypes excluded</i>									
4	2,923,391	5.1	PhvuI.004G027800	0	AT3G28860	ABCB19,ATABCB19,ATMDR1,ATMDR11,ATPGP19,MDR1,MDR11,PGP19	ATP binding cassette subfamily B19	MDP, MDP-CO, MDP-MI, DJ, DJ-CO	0.08
6	30,675,879	5.4	PhvuI.006G203400	34,372	AT5G64340	SAC51	sequence-specific DNA binding transcription factors; transcription regulators	MDP, MDP-CO, MDP-MI, DJ, DJ-CO, DJ-MI	0.08
7	45,836,294	4.3	PhvuI.007G218900	0	AT1G33410	ATNUP160,NUP160,SAR1	SUPPRESSOR OF AUXIN RESISTANCE1	MDP, MDP-MI, DJ, DJ-CO, DJ-MI	0.06
7	46,112,355	6.0	PhvuI.007G221700	0	AT2G37150	-	RING/U-box superfamily protein	MDP, MDP-CO, MDP-MI, DJ, DJ-CO, DJ-MI	0.10
7	46,132,395	4.1	PhvuI.007G221800	0	AT2G33170	-	Leucine-rich repeat receptor-like protein kinase family protein	MDP, MDP-MI, DJ, DJ-CO, DJ-MI	0.06
7	48,630,699	4.2	PhvuI.007G246700	0	AT4G02330	ATPMEPCRB,AtPME41	Plant invertase/pectin methylesterase inhibitor superfamily	MDP, MDP-CO, MDP-MI, DJ, DJ-CO	0.06
11	43,284,551	5.8	PhvuI.011G164800	0	AT1G53160	SPL4	squamosa promoter-binding protein-like 4	MDP, MDP-CO, MDP-MI, DJ, DJ-MI	0.09

Supplementary Table S1. Candidate genes in the 100 Kb region centered on a significant marker (continued).

Pv	SNP position (bp)	$-\log_{10}$ (<i>P</i> -value)	Bean candidate gene	Marker distance from candidate gene (bp)	Arabidopsis gene model	Arabidopsis gene symbol	Arabidopsis annotation	Population/location in which is significant	R^2_{LR} in the best model
<i>Lodging</i>									
7	45,836,294	5.5	PhvuI.007G218900	0	AT1G33410	ATNUP160,NUP160,SAR1	SUPPRESSOR OF AUXIN RESISTANCE1	MDP, MDP-CO, MDP-NE, DJ, DJ-CO, DJ-MI, DJ-NE	0.08
7	46,112,355	10.6	PhvuI.007G221700	0	AT2G37150	-	RING/U-box superfamily protein	MDP, MDP-CO, MDP-MI, MDP-NE, DJ, DJ-CO, DJ-MI, DJ-NE	0.18
7	46,131,994	7.2	PhvuI.007G221800	0	AT2G33170	-	Leucine-rich repeat receptor-like protein kinase family protein	MDP, MDP-CO, MDP-MI, MDP-NE, DJ, DJ-CO, DJ-MI, DJ-NE, MA, MA-CO, MA-NE	0.11
7	48,630,699	6.8	PhvuI.007G246700	0	AT4G02330	ATPMEPCRB,AtPME41	Plant invertase/pectin methyltransferase inhibitor superfamily	MDP, MDP-CO, MDP-NE, DJ, DJ-CO, DJ-NE, DJ-MI	0.11
8	12,646,413	3.1	PhvuI.008G108600	-22,921	AT2G38050	ATDET2,DET2,DWF6	3-oxo-5-alpha-steroid 4-dehydrogenase family protein	MA, MA-NE, MA-CO	-
<i>Canopy height</i>									
7	45,836,294	4.1	PhvuI.007G218900	0	AT1G33410	ATNUP160,NUP160,SAR1	SUPPRESSOR OF AUXIN RESISTANCE1	MDP, MDP-MI, DJ, DJ-CO, DJ-MI,	0.05
7	46,112,355	9.0	PhvuI.007G221700	0	AT2G37150	-	RING/U-box superfamily protein	MDP, MDP-CO, MDP-MI, MDP-NE, DJ, DJ-CO, DJ-MI, DJ-NE, MA, MA-NE	0.15
7	46,131,994	5.0	PhvuI.007G221800	0	AT2G33170	-	Leucine-rich repeat receptor-like protein kinase family protein	MDP, MDP- MI, MDP-NE, DJ, DJ- MI, DJ-NE, MA-NE	0.07
7	48,630,699	4.7	PhvuI.007G246700	0	AT4G02330	ATPMEPCRB, AtPME41	Plant invertase/pectin methyltransferase inhibitor superfamily	MDP, MDP-CO, MDP-MI, MDP-NE, DJ, DJ-CO, DJ-MI, DJ-NE	0.06

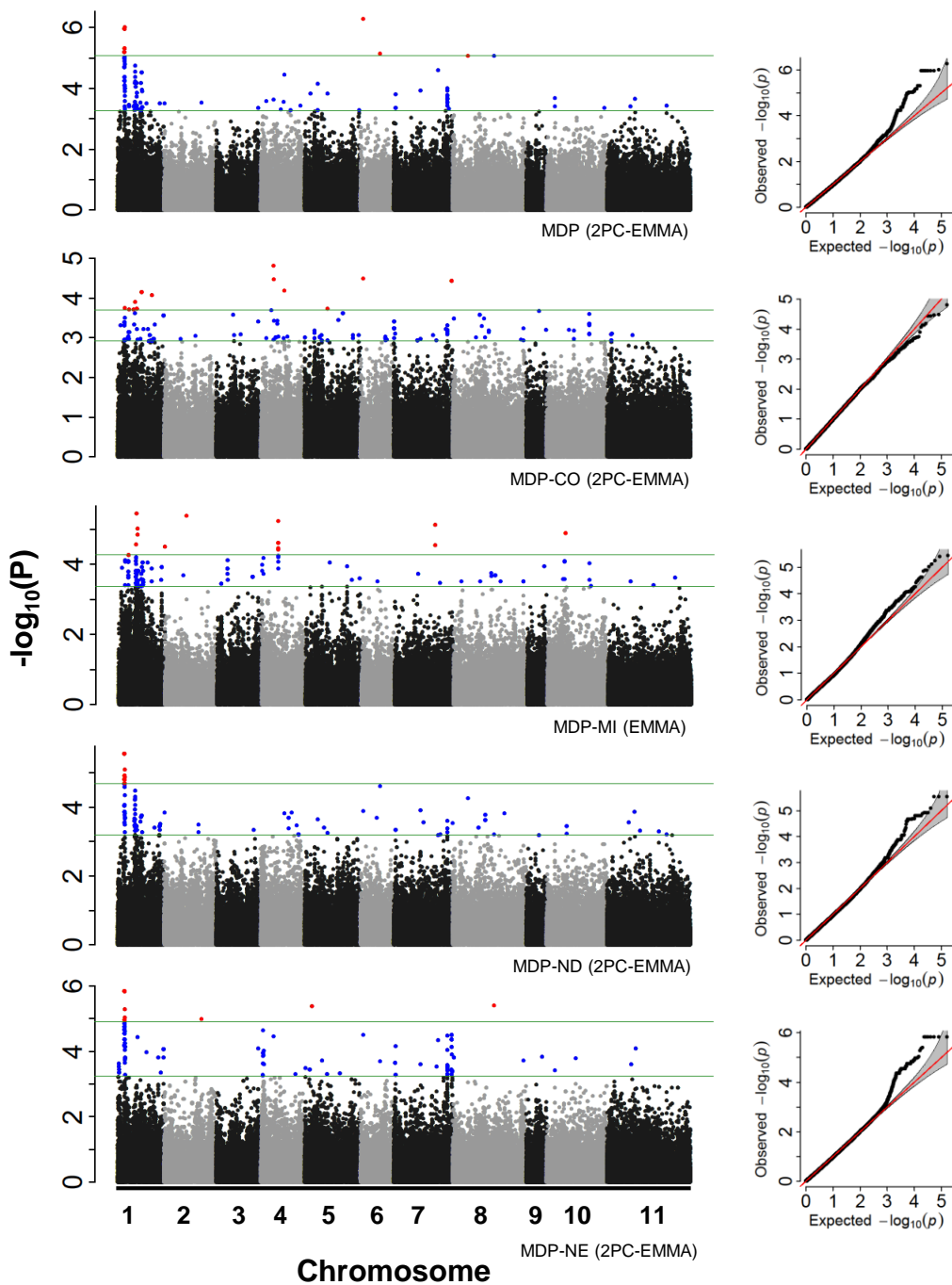
Supplementary Table S1. Candidate genes in the 100 Kb region centered on a significant marker (continued).

Pv	SNP position (bp)	-log10 (P-value)	Bean candidate gene	Marker distance from candidate gene (bp)	Arabidopsis gene model	Arabidopsis gene symbol	Arabidopsis annotation	Population/location in which is significant	R^2_{LR} in the best model
<i>Seed weight</i>									
3	32,853,141	3.8	PhvuL003G136500	28,340	AT5G56970	ATCKX3,CKX3	cytokinin oxidase 3	DJ, MI-DJ, DJ-ND, ND	-
8	1,146,498	4.5	PhvuL008G013300	21,900	AT1G01900	ATSBT1.1,SBTI1.1	subtilase family protein	DJ, DJ-CO, DJ-MI, DJ-ND, DJ-NE	-
10	2,756,229	4.8	PhvuL010G017600	22,158	AT4G25000	AMY1,ATAMY1	alpha-amylase-like	MDP, MDP-CO, MDP-MI, MDP-ND	0.07
10	2,762,081	3.6	PhvuL010G017600	16,306	AT4G25000	AMY1,ATAMY1	alpha-amylase-like	DJ, DJ-CO, DJ-NE	-

Supplementary Figure S1. Manhattan plots, and QQ plots of the best models for agronomic traits. MDP, Middle American Diversity panel; DJ, Durango/Jalisco subpopulation; MA, Mesoamerican subpopulation; CO, Colorado; MI, Michigan; NE, Nebraska; ND, North Dakota. If no location is specified for a population/subpopulation, all locations are considered in the analysis. The best model for each analysis is indicated in parenthesis.

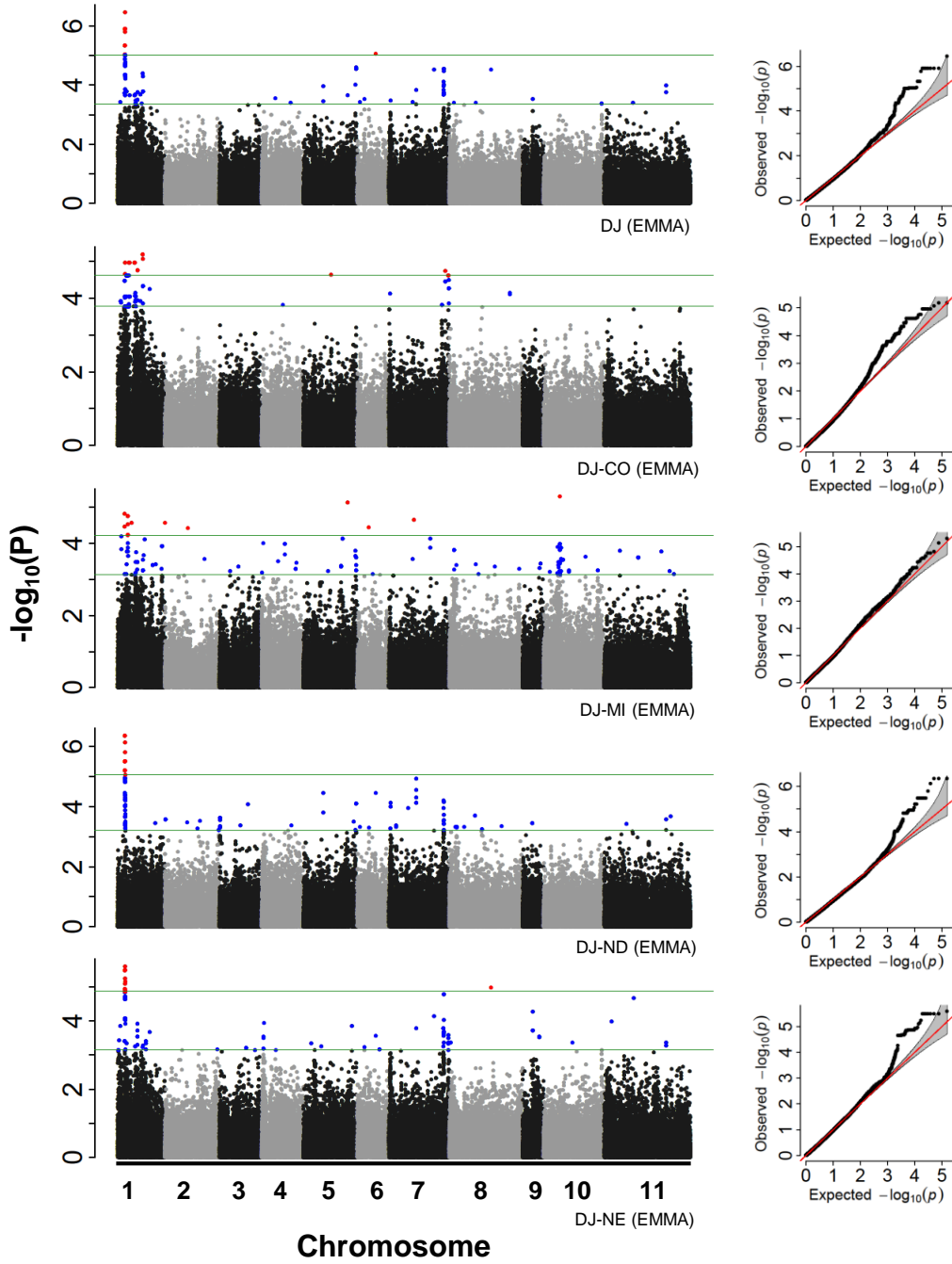
a

Days to Flower MDP



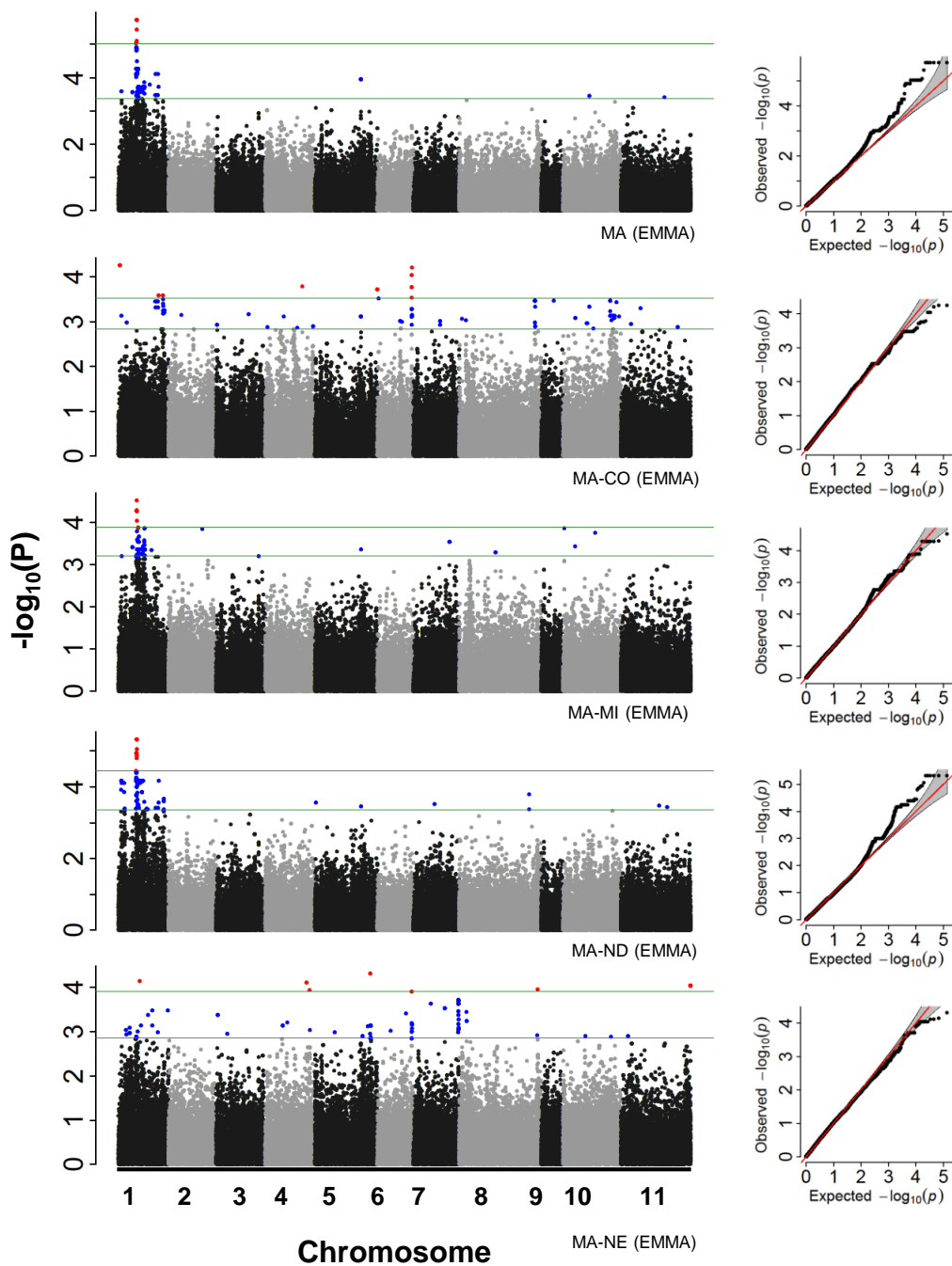
b

Days to Flower Durango/Jalisco



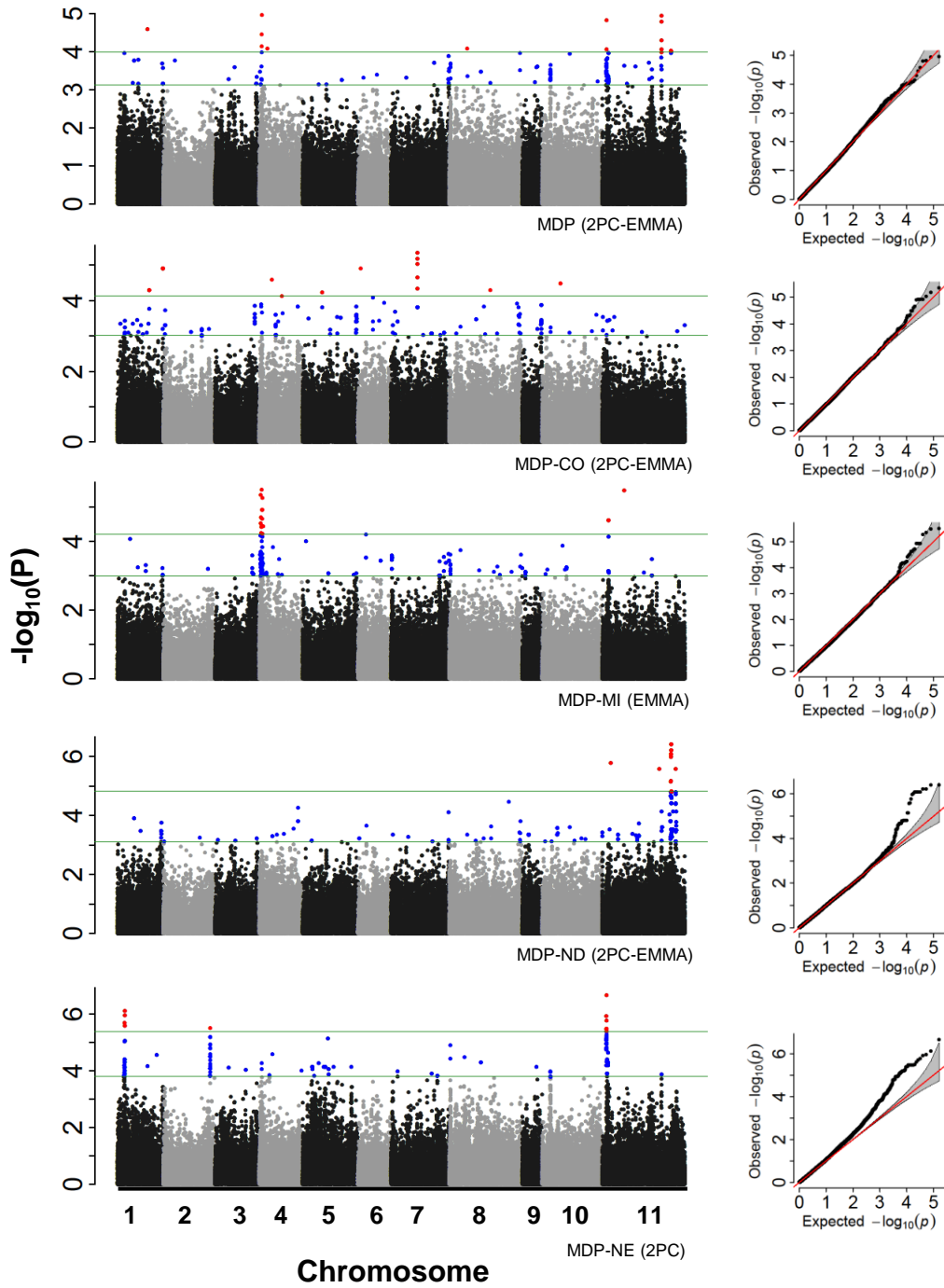
c

Days to Flower Mesoamerica



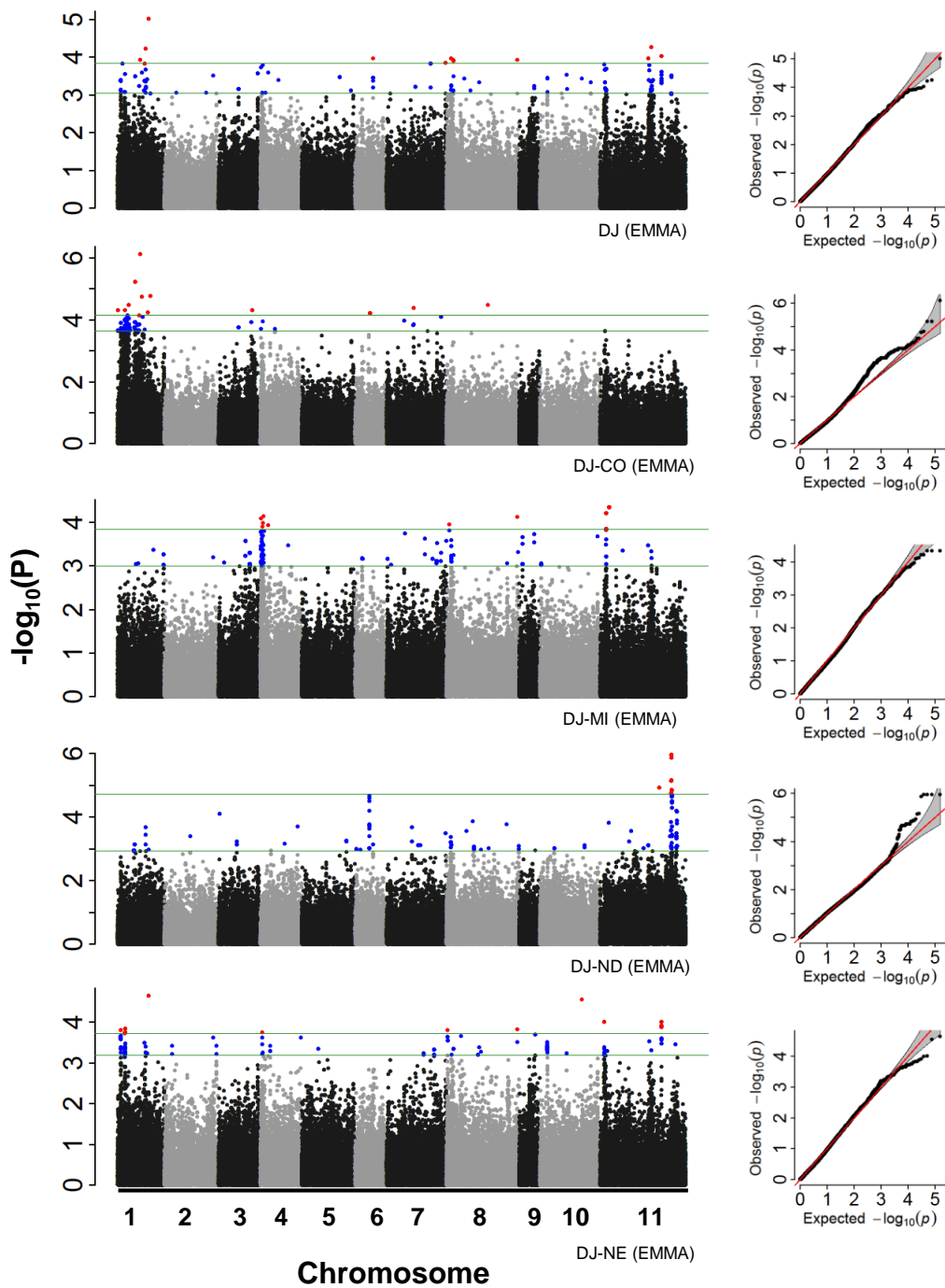
d

Days to Maturity MDP



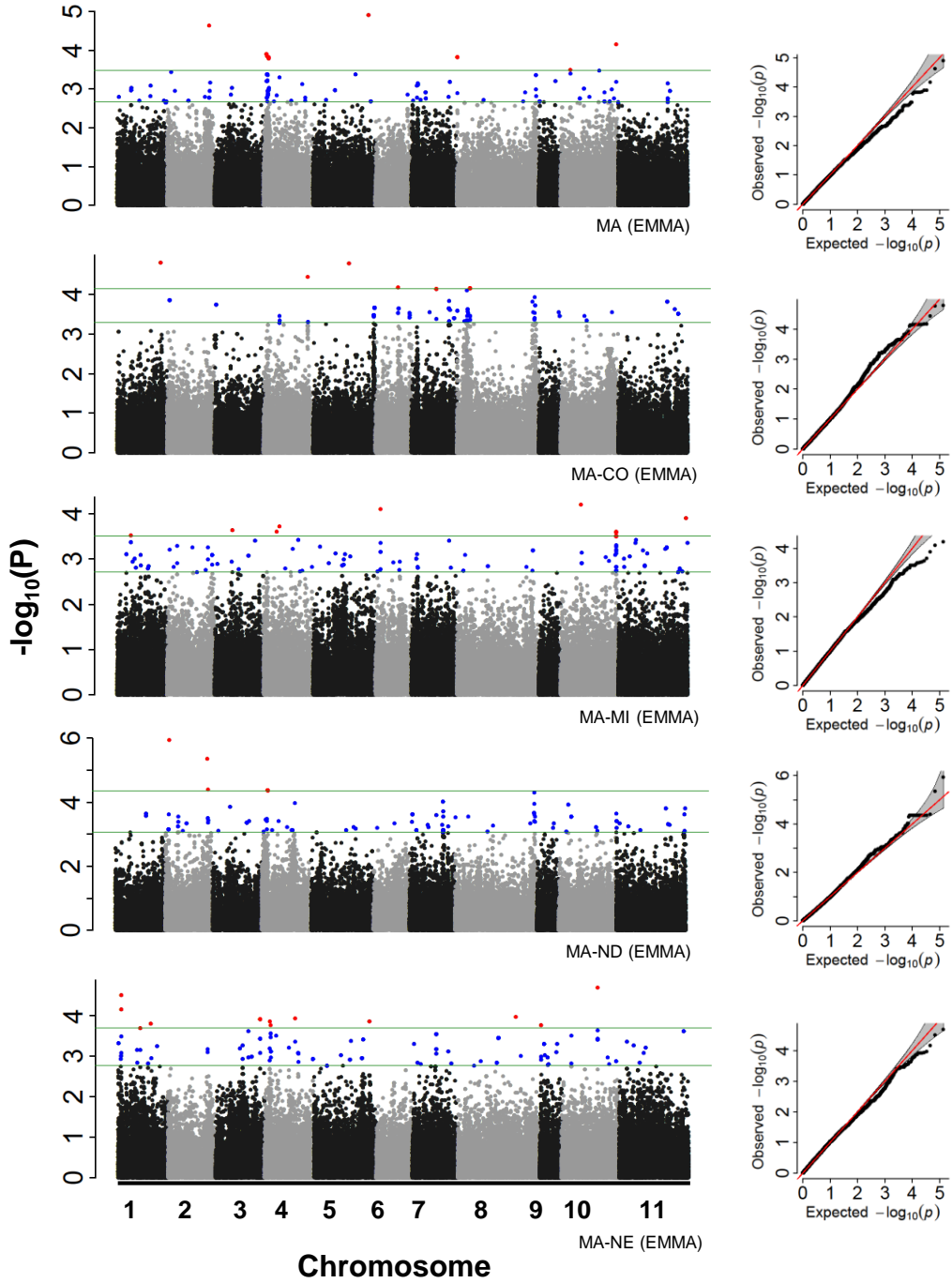
e

Days to Maturity Durango/Jalisco



f

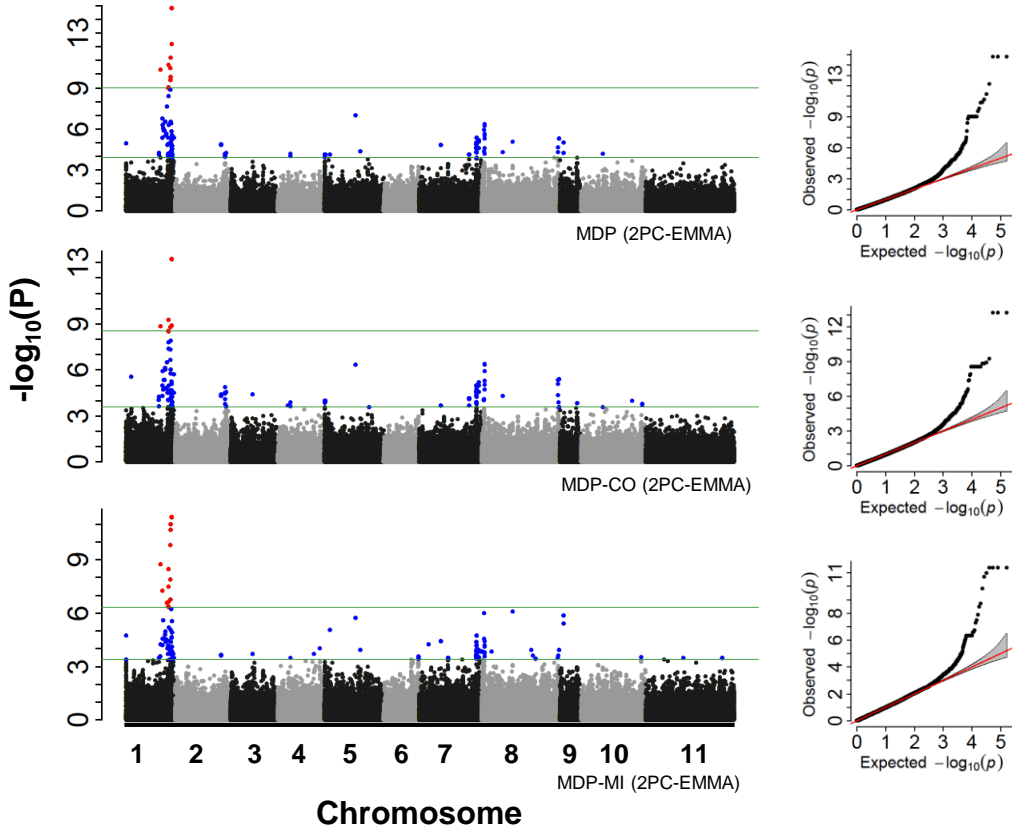
Days to Maturity Mesoamerica



Growth Habit

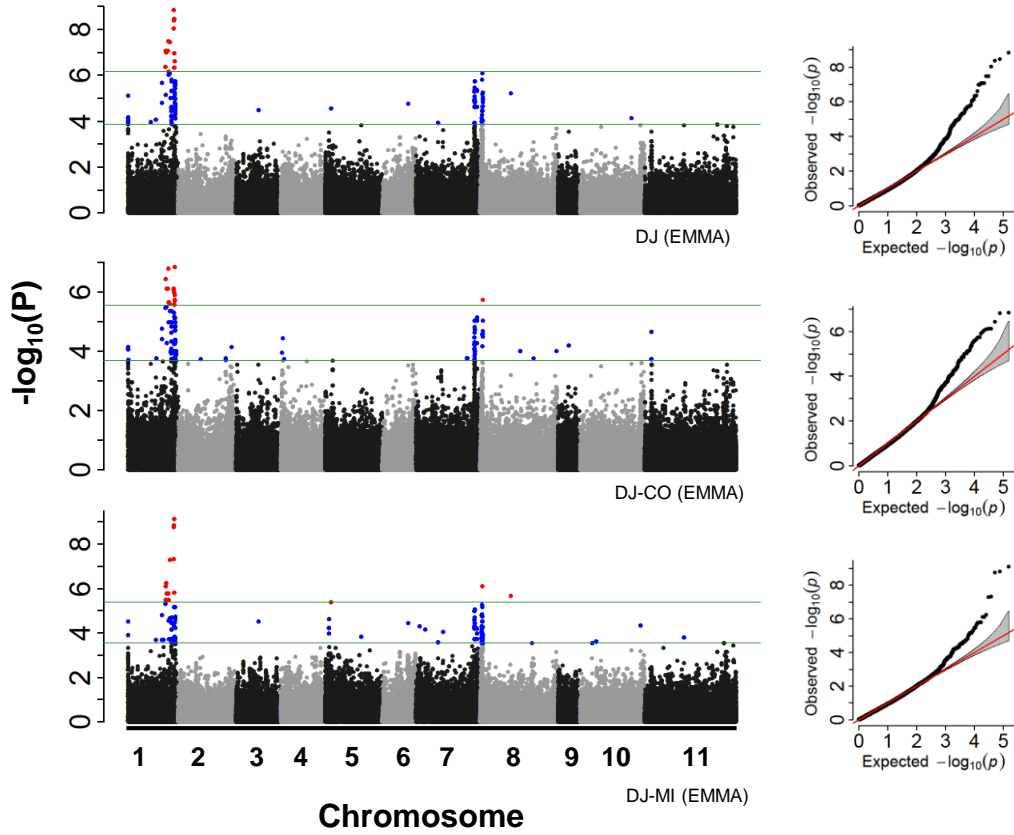
With determinate types

MDP



h

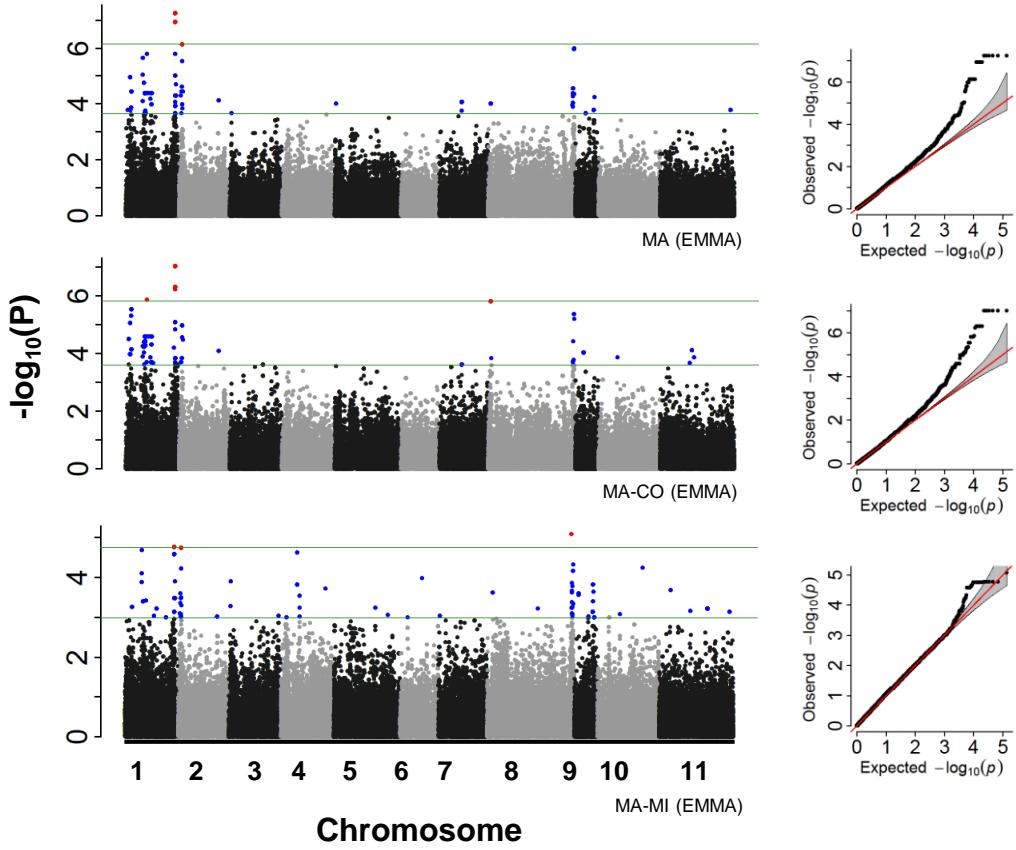
Growth Habit
With determinate types
Durango/Jalisco



i

Growth Habit

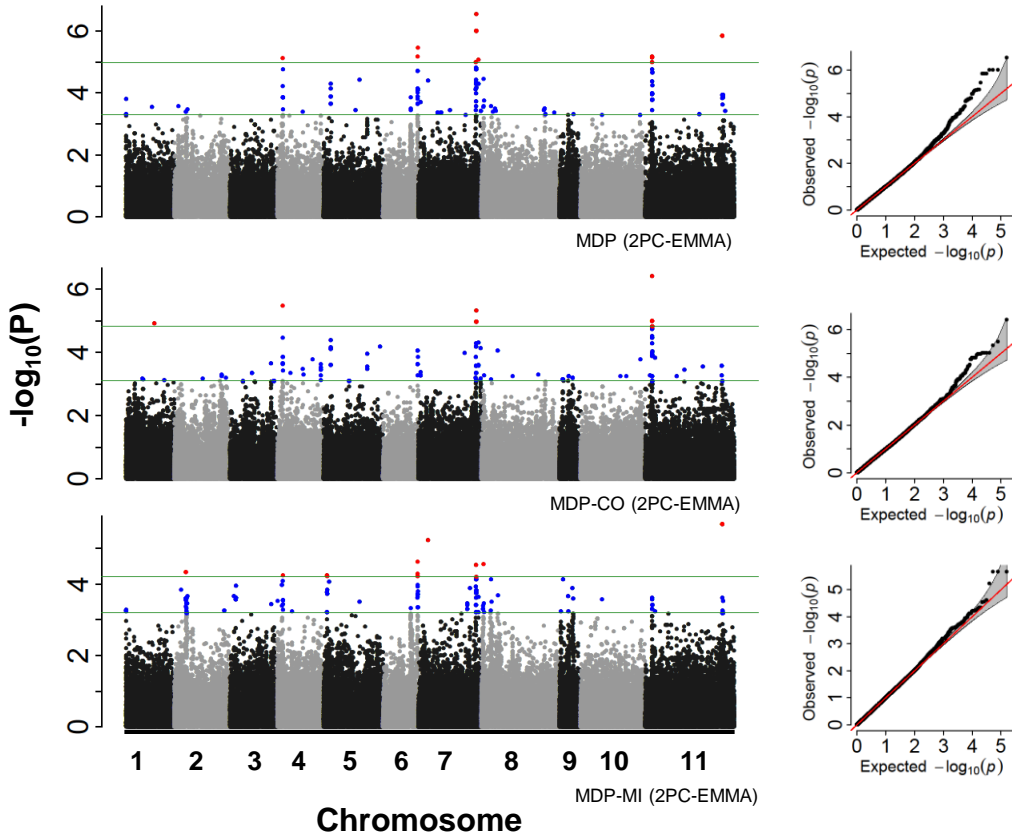
With determinate types
Mesoamerica



j

Growth Habit

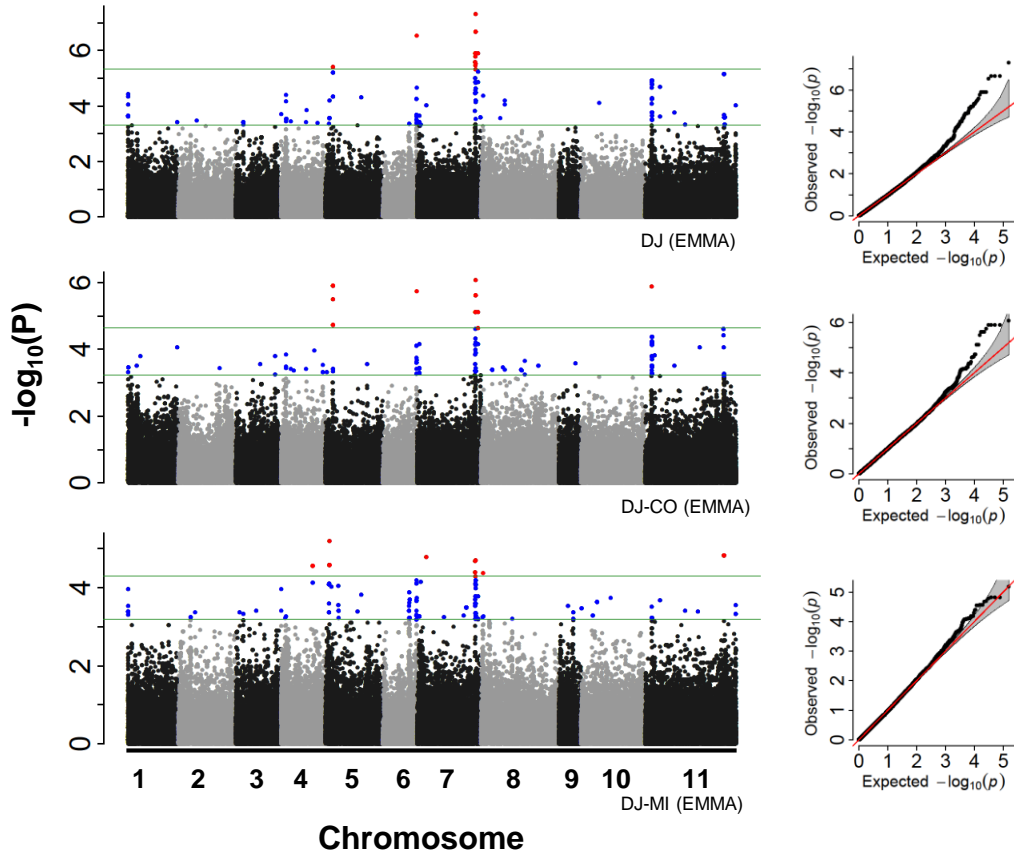
No determinate types
MDP



k

Growth Habit

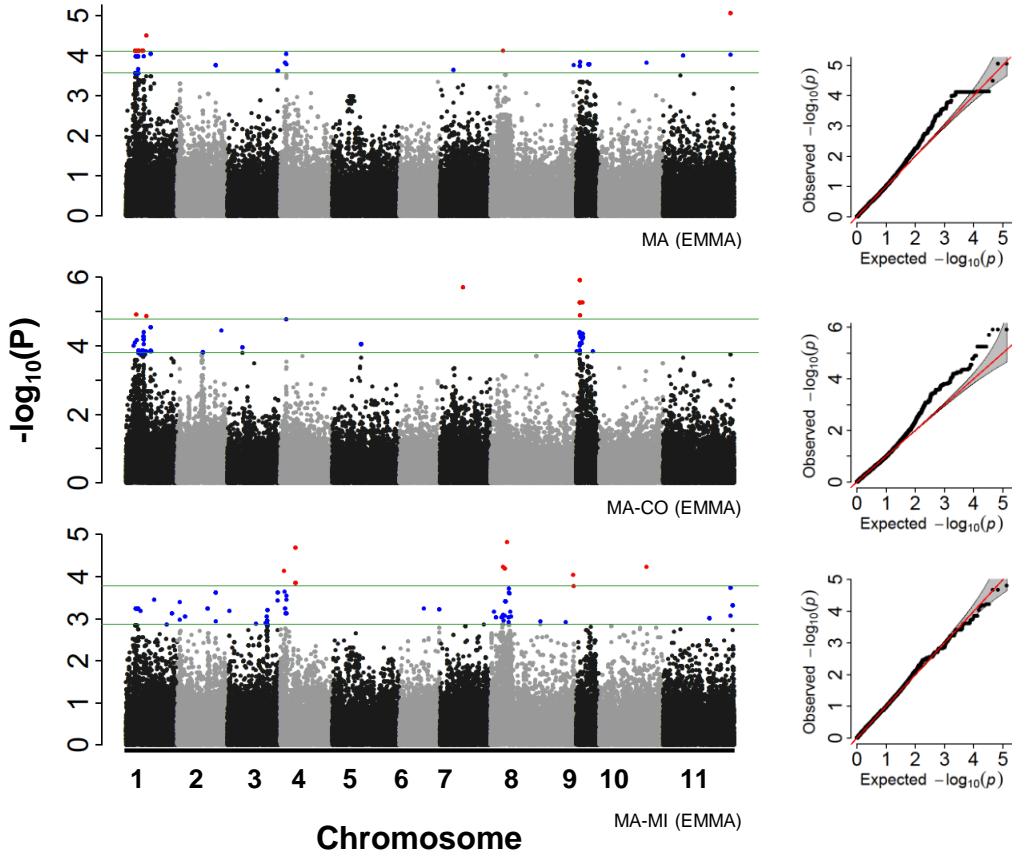
No determinate types
Durango/Jalisco



I

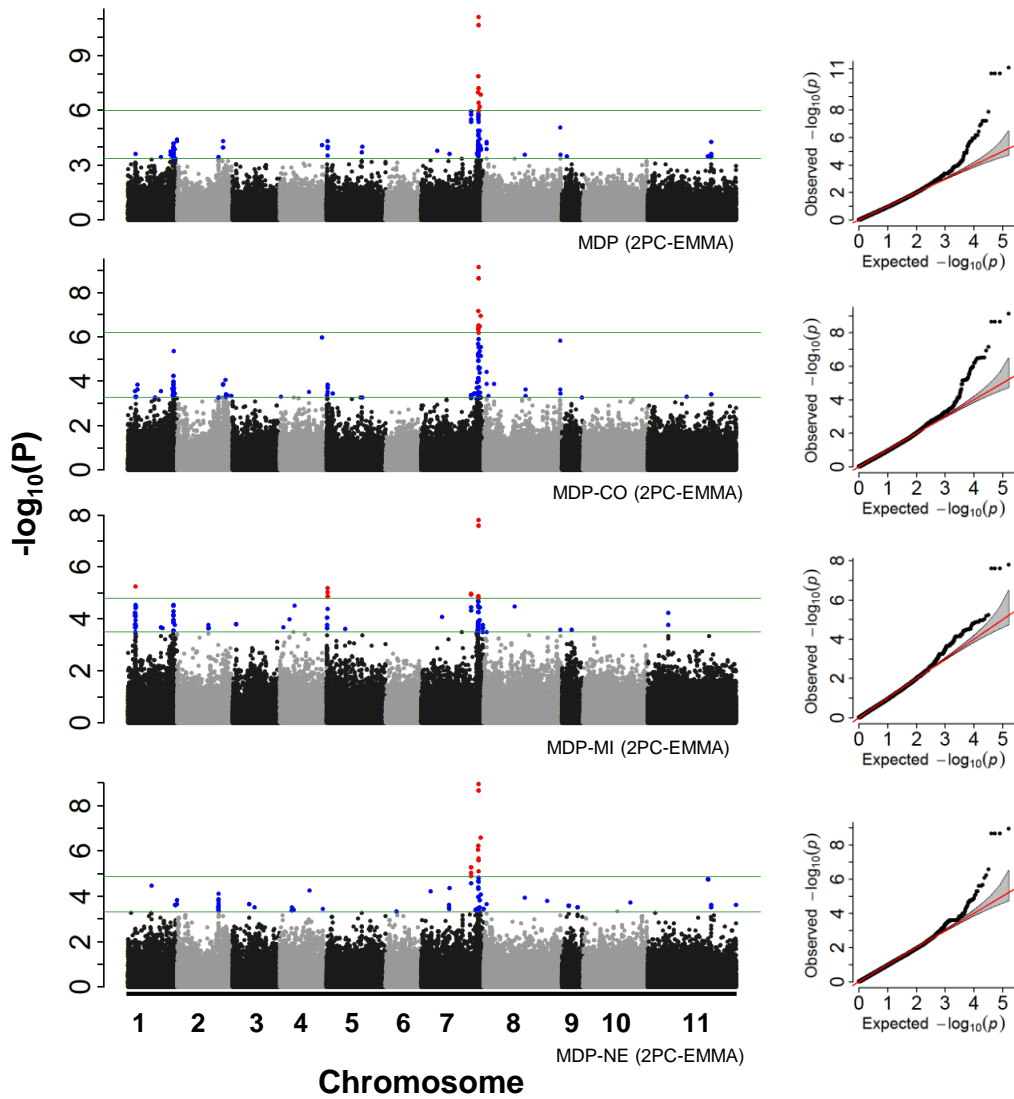
Growth Habit

No determinate types
Mesoamerica



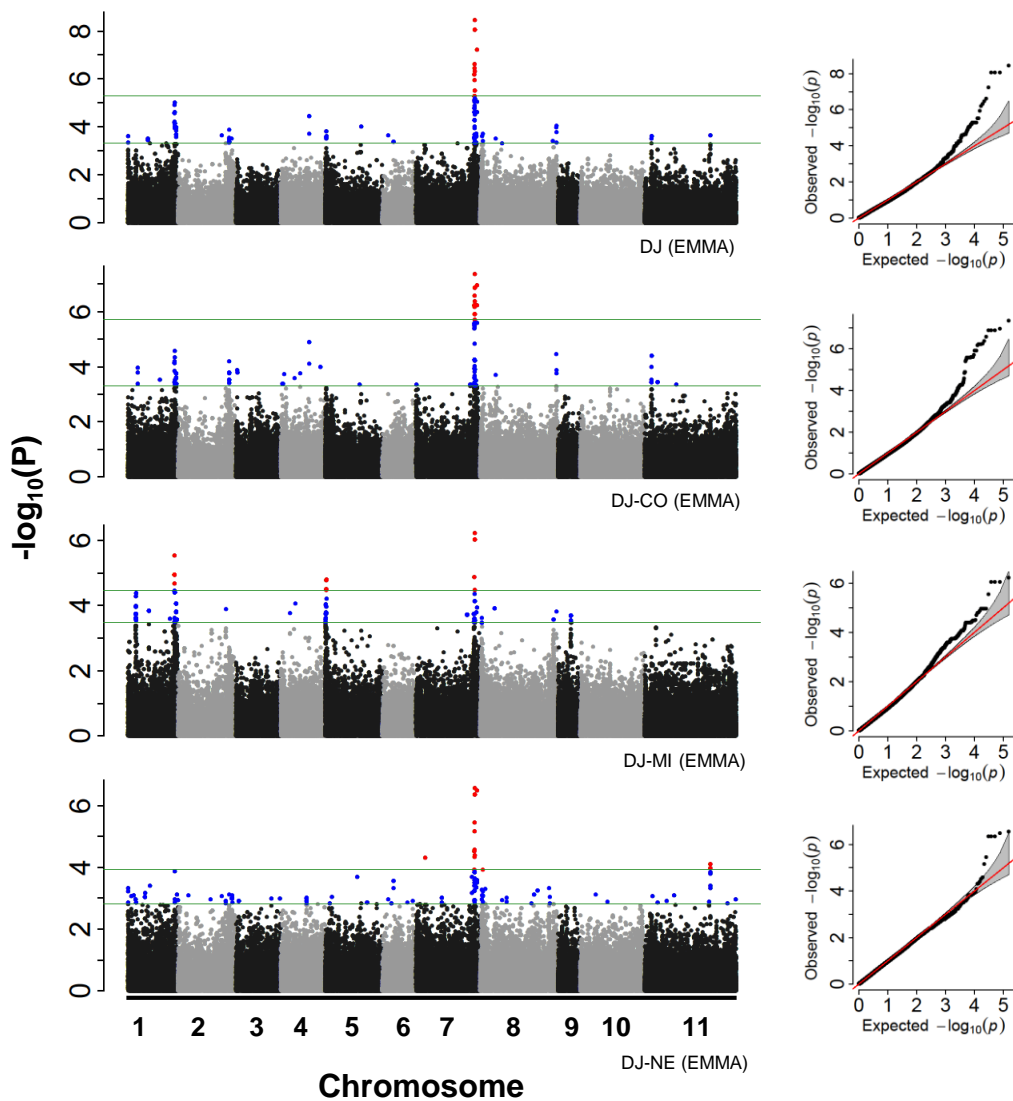
m

Lodging MDP



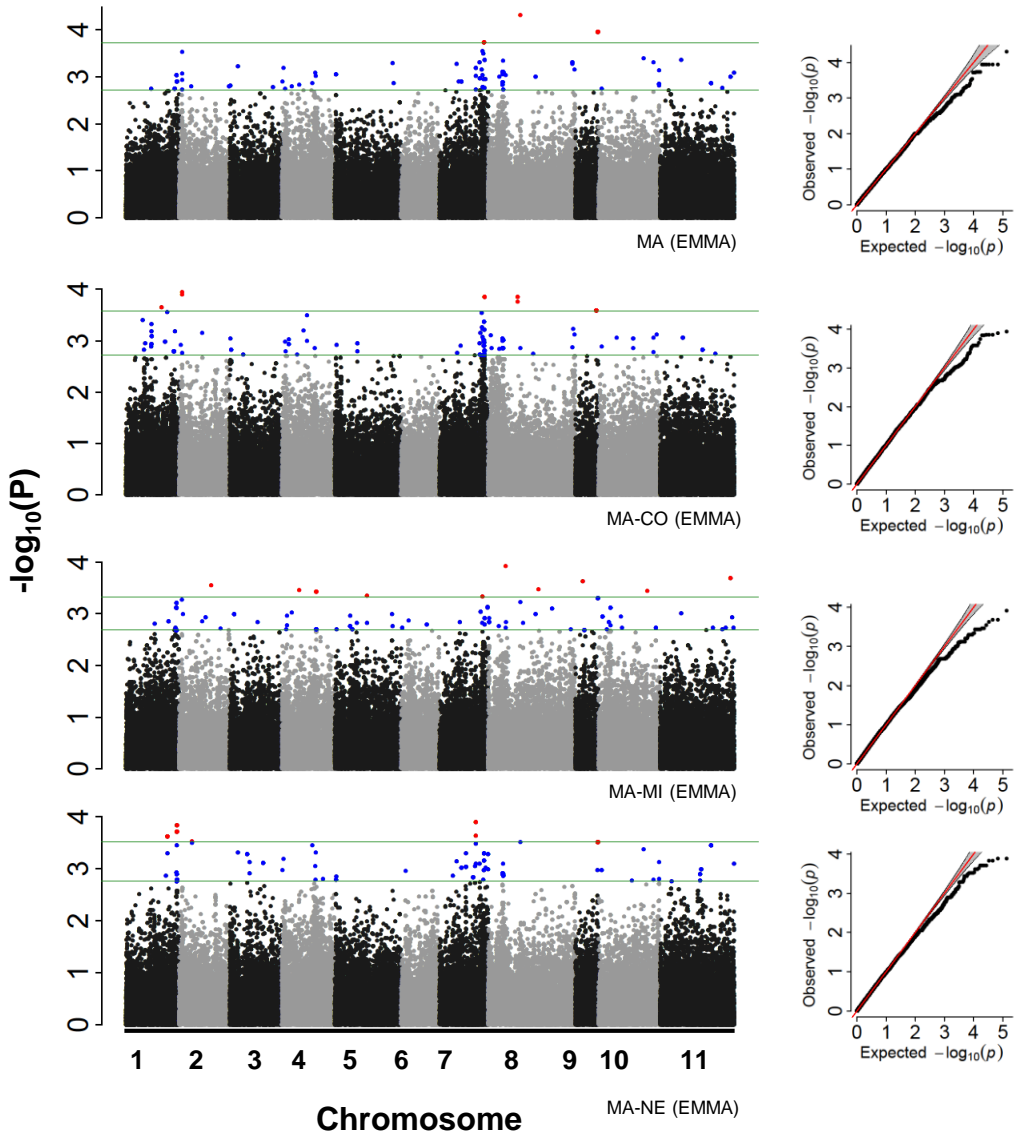
n

Lodging Durango/Jalisco



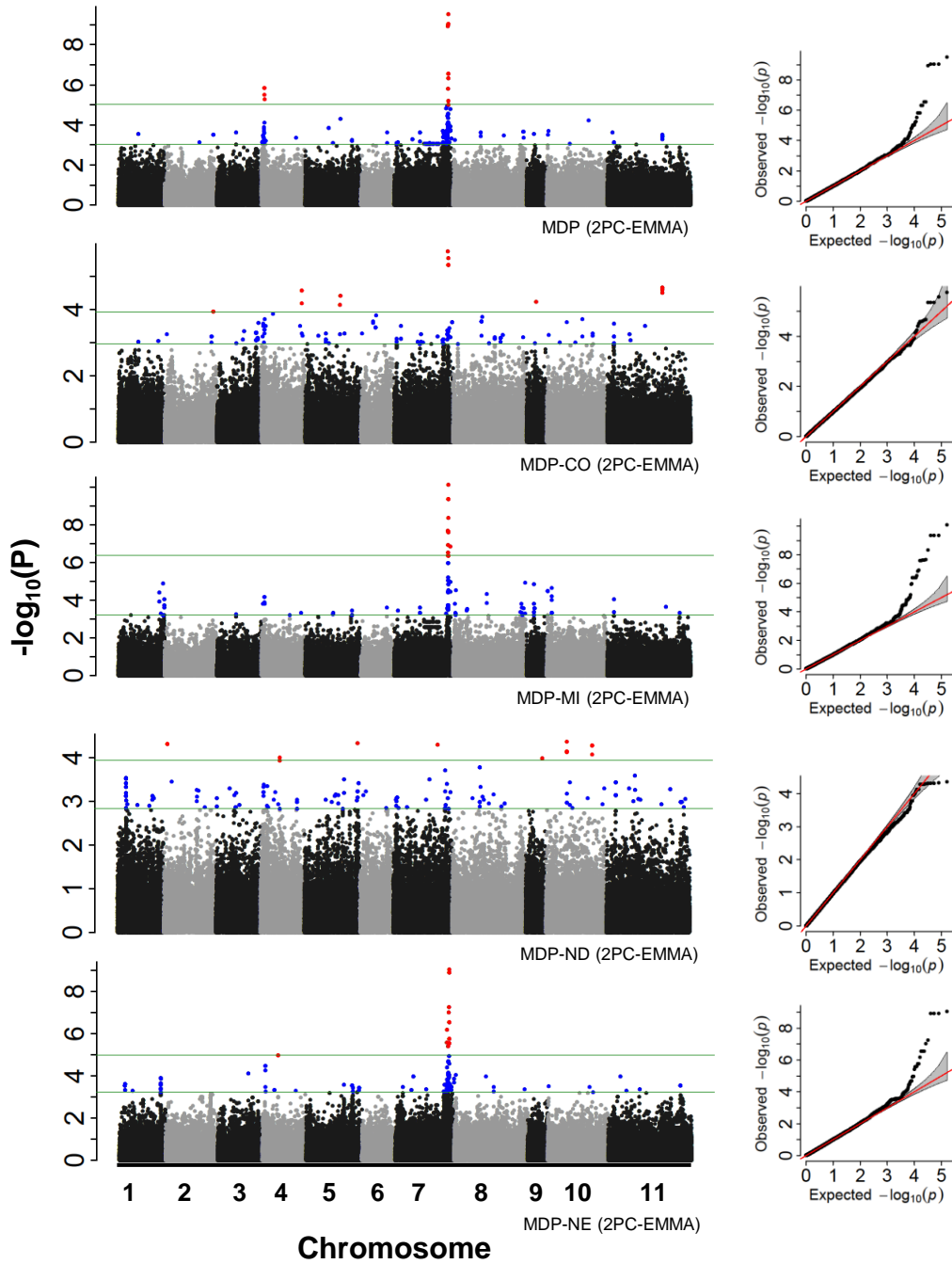
0

Lodging Mesoamerica



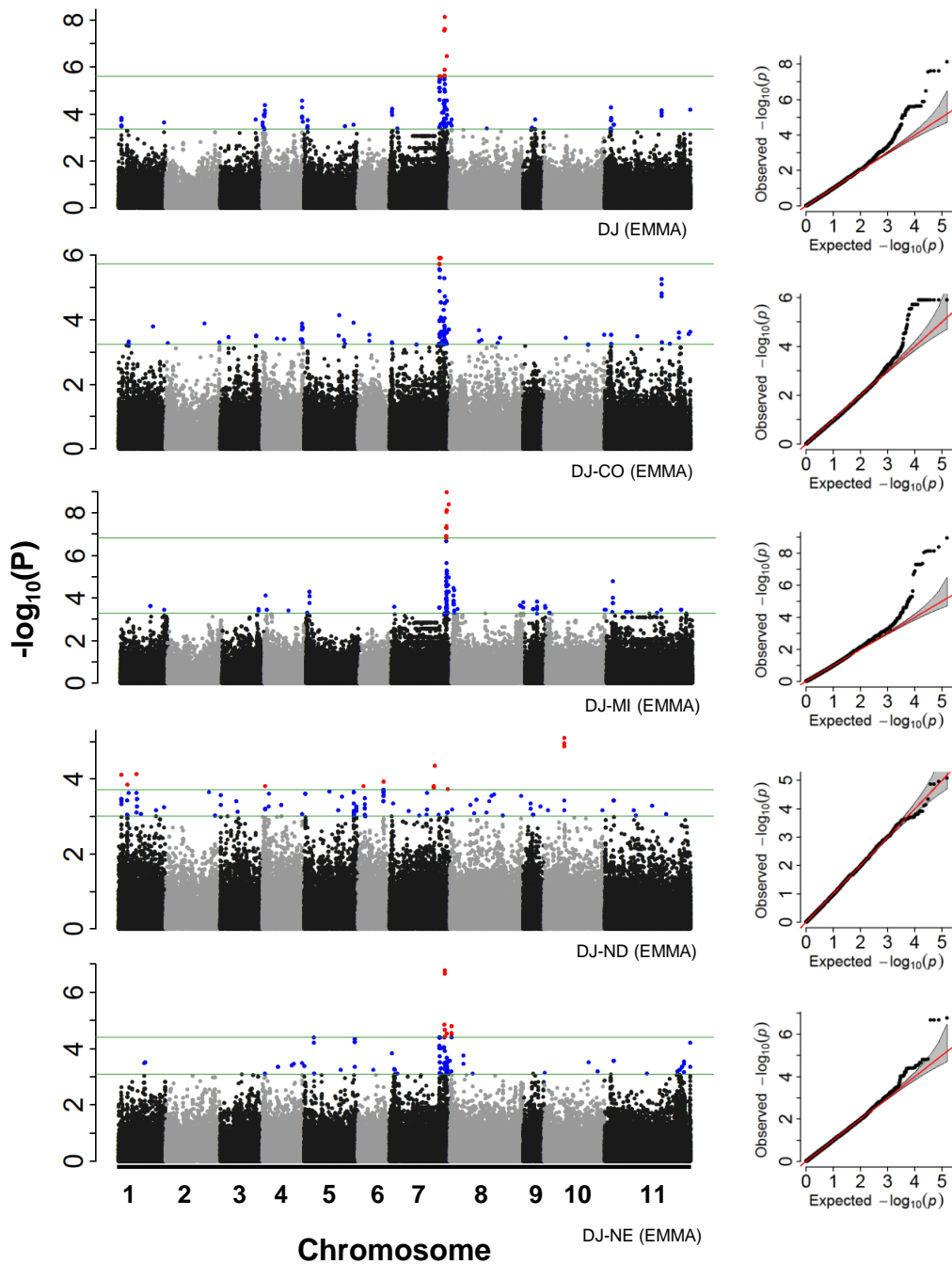
p

Canopy Height MDP



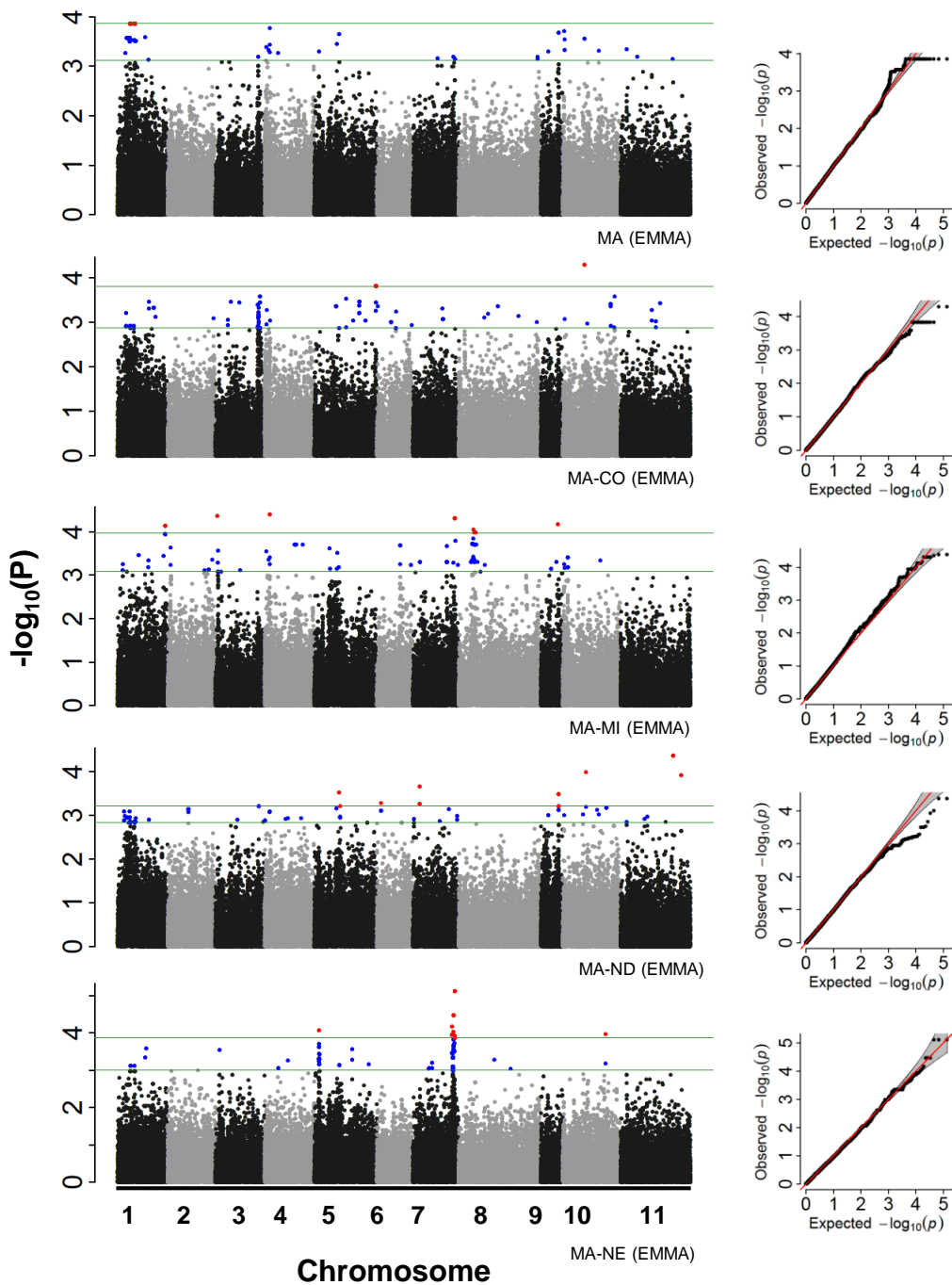
q

Canopy Height Durango/Jalisco



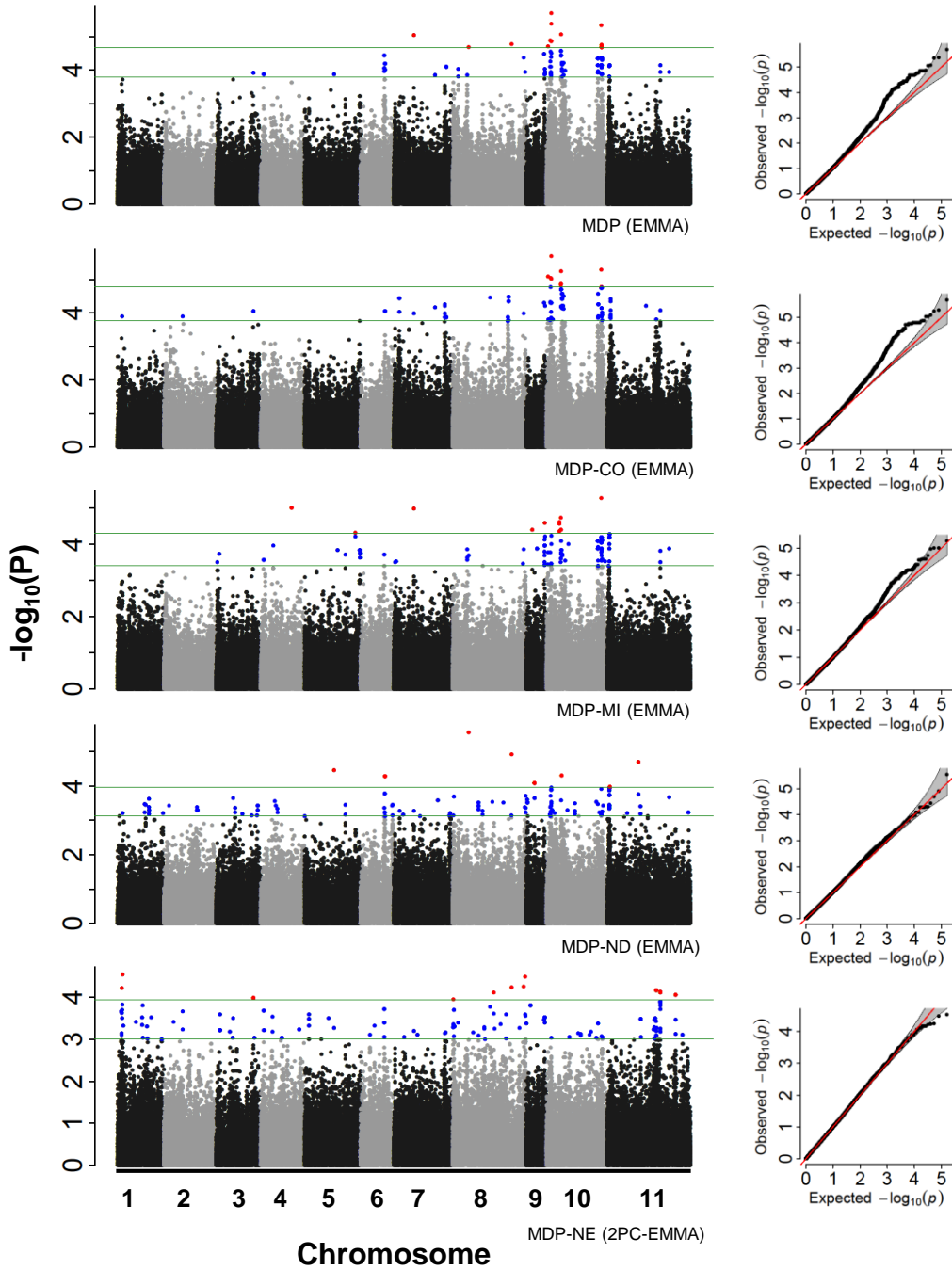
r

Canopy Height Mesoamerica



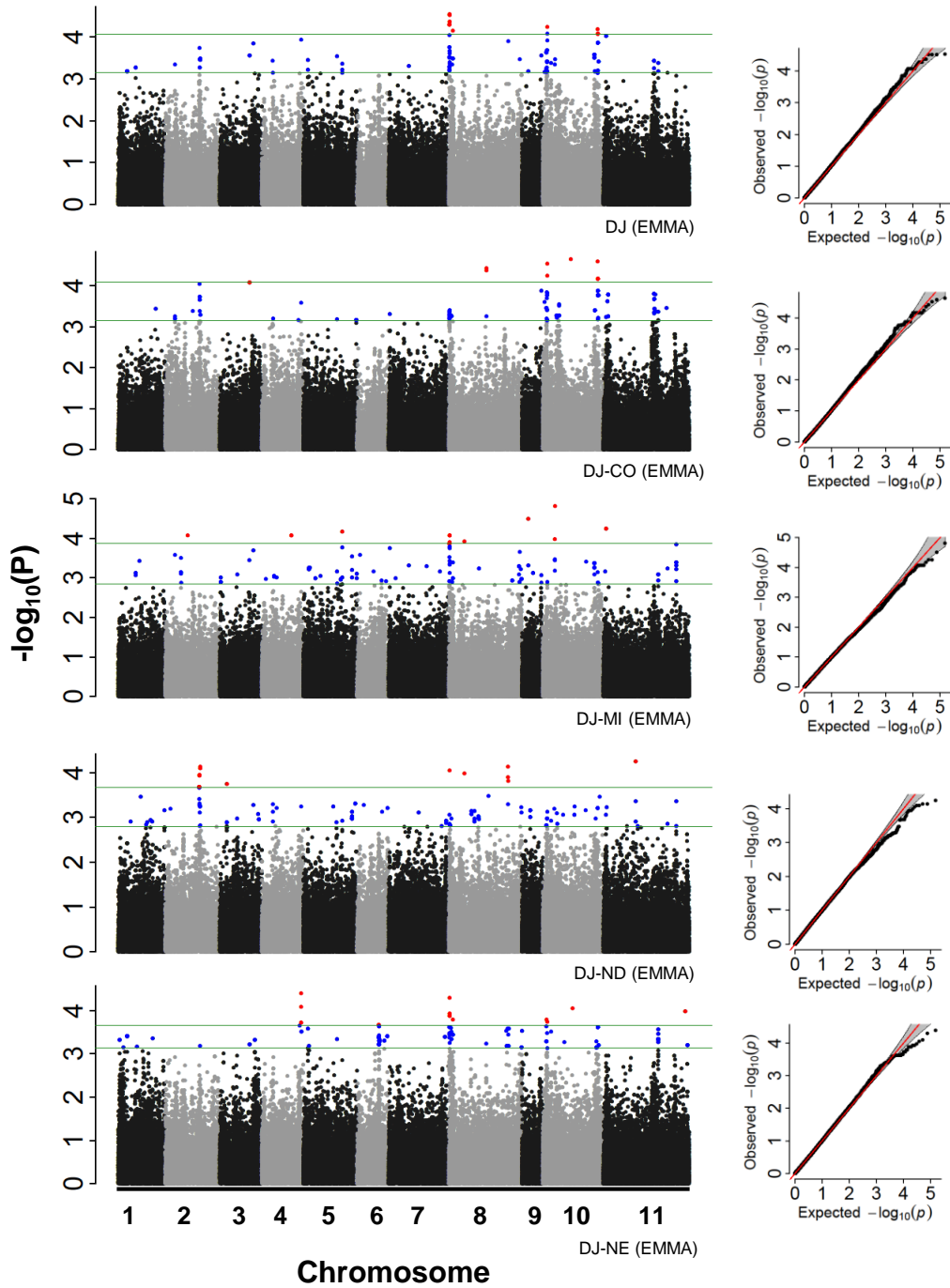
s

Seed Weight MDP



t

Seed Weight Durango/Jalisco



u

Seed Weight Mesoamerica

