Development of a QTL-environment-based predictive model for node addition rate in common bean

Citation:

Zhang, Li; Gezan, Salvador A.; Vallejos, C. Eduardo; Jones, James W.; Boote, Kenneth J.; Clavijo-Michelangeli, Jose A.; Bhakta, Mehul; Osorno, Juan M.; Rao, Idupulapati; Beebe, Stephen; Roman-Paoli, Elvin; Gonzalez, Abiezer; Beaver, James; Ricaurte, Jaumer; Colbert, Raphael; Correll, Melanie J.. 2017. Development of a QTL-environment-based predictive model for node addition rate in common bean. Theoretical and Applied Genetics. 130(5): 1065-1079.

Publisher's DOI:
http://dx.doi.org/10.1007/s00122-017-2871-y

Access through CIAT Research Online:
http://hdl.handle.net/10568/80542

Springer

Dear Author,

Here are the proofs of your article.

- You can submit your corrections **online**, via **e-mail** or by **fax**.
- For **online** submission please insert your corrections in the online correction form. Always indicate the line number to which the correction refers.
- You can also insert your corrections in the proof PDF and **email** the annotated PDF.
- For fax submission, please ensure that your corrections are clearly legible. Use a fine black pen and write the correction in the margin, not too close to the edge of the page.
- Remember to note the **journal title**, **article number**, and **your name** when sending your response via e-mail or fax.
- **Check** the metadata sheet to make sure that the header information, especially author names and the corresponding affiliations are correctly shown.
- **Check** the questions that may have arisen during copy editing and insert your answers/ corrections.
- **Check** that the text is complete and that all figures, tables and their legends are included. Also check the accuracy of special characters, equations, and electronic supplementary material if applicable. If necessary refer to the *Edited manuscript*.
- The publication of inaccurate data such as dosages and units can have serious consequences. Please take particular care that all such details are correct.
- Please **do not** make changes that involve only matters of style. We have generally introduced forms that follow the journal's style.
  Substantial changes in content, e.g., new results, corrected values, title and authorship are not allowed without the approval of the responsible editor. In such a case, please contact the Editorial Office and return his/her consent together with the proof.
- If we do not receive your corrections **within 48 hours**, we will send you a reminder.
- Your article will be published **Online First** approximately one week after receipt of your corrected proofs. This is the **official first publication** citable with the DOI. **Further changes are, therefore, not possible.**
- The **printed version** will follow in a forthcoming issue.

Please note

After online publication, subscribers (personal/institutional) to this journal will have access to the complete article via the DOI using the URL: http://dx.doi.org/[DOI].
If you would like to know when your article has been published online, take advantage of our free alert service. For registration and further information go to: http://www.link.springer.com.

Due to the electronic nature of the procedure, the manuscript and the original figures will only be returned to you on special request. When you return your corrections, please inform us if you would like to have these documents returned.

Metadata of the article that will be visualized in OnlineFirst

ArticleTitle	Development of a QTL-environment-based predictive model for node addition rate in common bean	
Article Sub-Title		
Article CopyRight	Springer-Verlag Berlin Heidelberg (This will be the copyright line in the final PDF)	
Journal Name	Theoretical and Applied Genetics	
Corresponding Author	Family Name	**Correll**
	Particle	
	Given Name	**Melanie J.**
	Suffix	
	Division	Agricultural and Biological Engineering Department
	Organization	University of Florida
	Address	Gainesville, FL, 32611, USA
	Phone	352-392-1864
	Fax	
	Email	correllm@ufl.edu
	URL	
	ORCID	http://orcid.org/0000-0002-2432-6107
Author	Family Name	**Zhang**
	Particle	
	Given Name	**Li**
	Suffix	
	Division	Agricultural and Biological Engineering Department
	Organization	University of Florida
	Address	Gainesville, FL, 32611, USA
	Phone	
	Fax	
	Email	lzhang1210@ufl.edu
	URL	
	ORCID	
Author	Family Name	**Gezan**
	Particle	
	Given Name	**Salvador A.**
	Suffix	
	Division	School of Forest Resources and Conservation
	Organization	University of Florida
	Address	Gainesville, FL, 32611, USA
	Phone	
	Fax	
	Email	sgezan@ufl.edu
	URL	

	ORCID	
Author	Family Name	**Eduardo Vallejos**
	Particle	
	Given Name	**C.**
	Suffix	
	Division	Horticultural Sciences Department
	Organization	University of Florida
	Address	Gainesville, FL, 32611, USA
	Phone	
	Fax	
	Email	vallejos@ufl.edu
	URL	
	ORCID	
Author	Family Name	**Jones**
	Particle	
	Given Name	**James W.**
	Suffix	
	Division	Agricultural and Biological Engineering Department
	Organization	University of Florida
	Address	Gainesville, FL, 32611, USA
	Phone	
	Fax	
	Email	jimj@ufl.edu
	URL	
	ORCID	
Author	Family Name	**Boote**
	Particle	
	Given Name	**Kenneth J.**
	Suffix	
	Division	Agronomy Department
	Organization	University of Florida
	Address	Gainesville, FL, 32611, USA
	Phone	
	Fax	
	Email	kjboote@ufl.edu
	URL	
	ORCID	
Author	Family Name	**Clavijo-Michelangeli**
	Particle	
	Given Name	**Jose A.**
	Suffix	
	Division	Agronomy Department
	Organization	University of Florida
	Address	Gainesville, FL, 32611, USA
	Phone	

	Fax	
	Email	pepeclavijom@gmail.com
	URL	
	ORCID	
Author	Family Name	**Bhakta**
	Particle	
	Given Name	**Mehul**
	Suffix	
	Division	Horticultural Sciences Department
	Organization	University of Florida
	Address	Gainesville, FL, 32611, USA
	Phone	
	Fax	
	Email	bhaktamehul@ufl.edu
	URL	
	ORCID	
Author	Family Name	**Osorno**
	Particle	
	Given Name	**Juan M.**
	Suffix	
	Division	Department of Plant Sciences
	Organization	North Dakota State University
	Address	Fargo, ND, 58108, USA
	Phone	
	Fax	
	Email	Juan.Osorno@ndsu.edu
	URL	
	ORCID	
Author	Family Name	**Rao**
	Particle	
	Given Name	**Idupulapati**
	Suffix	
	Division	
	Organization	CIAT
	Address	Cali, Colombia
	Phone	
	Fax	
	Email	i.rao@cgiar.org
	URL	
	ORCID	
Author	Family Name	**Beebe**
	Particle	
	Given Name	**Stephen**
	Suffix	
	Division	

	Organization	CIAT
	Address	Cali, Colombia
	Phone	
	Fax	
	Email	s.beebe@cgiar.org
	URL	
	ORCID	
Author	Family Name	**Roman-Paoli**
	Particle	
	Given Name	**Elvin**
	Suffix	
	Division	
	Organization	University of Puerto Rico
	Address	Mayagüez, 00682, Puerto Rico
	Phone	
	Fax	
	Email	elvin.roman@upr.edu
	URL	
	ORCID	
Author	Family Name	**Gonzalez**
	Particle	
	Given Name	**Abiezer**
	Suffix	
	Division	
	Organization	University of Puerto Rico
	Address	Mayagüez, 00682, Puerto Rico
	Phone	
	Fax	
	Email	agonzalez111@gmail.com
	URL	
	ORCID	
Author	Family Name	**Beaver**
	Particle	
	Given Name	**James**
	Suffix	
	Division	
	Organization	University of Puerto Rico
	Address	Mayagüez, 00682, Puerto Rico
	Phone	
	Fax	
	Email	james.beaver@upr.edu
	URL	
	ORCID	
Author	Family Name	**Ricaurte**
	Particle	

	Given Name	**Jaumer**
	Suffix	
	Division	
	Organization	CIAT
	Address	Cali, Colombia
	Phone	
	Fax	
	Email	j.ricaurte@cgiar.org
	URL	
	ORCID	
Author	Family Name	**Colbert**
	Particle	
	Given Name	**Raphael**
	Suffix	
	Division	Department of Plant Sciences
	Organization	North Dakota State University
	Address	Fargo, ND, 58108, USA
	Phone	
	Fax	
	Email	Raphael.Colbert@my.ndsu.edu
	URL	
	ORCID	

Schedule	Received	14 September 2016
	Revised	
	Accepted	2 February 2017

Abstract

Key message:
This work reports the effects of the genetic makeup, the environment and the genotype by environment interactions for node addition rate in an RIL population of common bean. This information was used to build a predictive model for node addition rate.

Abstract:
To select a plant genotype that will thrive in targeted environments it is critical to understand the genotype by environment interaction (GEI). In this study, multi-environment QTL analysis was used to characterize node addition rate (NAR, node day^{-1}) on the main stem of the common bean (*Phaseolus vulgaris* L). This analysis was carried out with field data of 171 recombinant inbred lines that were grown at five sites (Florida, Puerto Rico, 2 sites in Colombia, and North Dakota). Four QTLs (Nar1, Nar2, Nar3 and Nar4) were identified, one of which had significant QTL by environment interactions (QEI), that is, Nar2 with temperature. Temperature was identified as the main environmental factor affecting NAR while day length and solar radiation played a minor role. Integration of sites as covariates into a QTL mixed site-effect model, and further replacing the site component with explanatory environmental covariates (i.e., temperature, day length and solar radiation) yielded a model that explained 73% of the phenotypic variation for NAR with root mean square error of 16.25% of the mean. The QTL consistency and stability was examined through a tenfold cross validation with different sets of genotypes and these four QTLs were always detected with 50–90% probability. The final model was evaluated using leave-one-site-out method to assess the influence of site on node addition rate. These analyses provided a quantitative measure of the effects on NAR of common beans exerted by the genetic makeup, the environment and their interactions.

Footnote Information

Communicated by Marcos Malosetti.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-017-2871-y) contains supplementary material, which is available to authorized users.

Theor Appl Genet
DOI 10.1007/s00122-017-2871-y

ORIGINAL ARTICLE

Development of a QTL-environment-based predictive model for node addition rate in common bean

Li Zhang[1] · Salvador A. Gezan[2] · C. Eduardo Vallejos[4] · James W. Jones[1] · Kenneth J. Boote[3] · Jose A. Clavijo-Michelangeli[3] · Mehul Bhakta[4] · Juan M. Osorno[5] · Idupulapati Rao[6] · Stephen Beebe[6] · Elvin Roman-Paoli[7] · Abiezer Gonzalez[7] · James Beaver[7] · Jaumer Ricaurte[6] · Raphael Colbert[5] · Melanie J. Correll[1]

Received: 14 September 2016 / Accepted: 2 February 2017

Abstract

Key message **This work reports the effects of the genetic makeup, the environment and the genotype by environment interactions for node addition rate in an RIL population of common bean. This information was used to build a predictive model for node addition rate.**

Abstract To select a plant genotype that will thrive in targeted environments it is critical to understand the genotype by environment interaction (GEI). In this study, multi-environment QTL analysis was used to characterize node addition rate (NAR, node day^{-1}) on the main stem of the common bean (*Phaseolus vulgaris* L). This analysis was carried out with field data of 171 recombinant inbred lines that were grown at five sites (Florida, Puerto Rico, 2 sites in Colombia, and North Dakota). Four QTLs (Nar1, Nar2, Nar3 and Nar4) were identified, one of which had significant QTL by environment interactions (QEI), that is, Nar2 with temperature. Temperature was identified as the main environmental factor affecting NAR while day length and solar radiation played a minor role. Integration of sites as covariates into a QTL mixed site-effect model, and further replacing the site component with explanatory environmental covariates (i.e., temperature, day length and solar radiation) yielded a model that explained 73% of the phenotypic variation for NAR with root mean square error of 16.25% of the mean. The QTL consistency and stability was examined through a tenfold cross validation with different sets

Communicated by Marcos Malosetti.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-017-2871-y) contains supplementary material, which is available to authorized users.

✉ Melanie J. Correll
correllm@ufl.edu

Li Zhang
lzhang1210@ufl.edu

Salvador A. Gezan
sgezan@ufl.edu

C. Eduardo Vallejos
vallejos@ufl.edu

James W. Jones
jimj@ufl.edu

Kenneth J. Boote
kjboote@ufl.edu

Jose A. Clavijo-Michelangeli
pepeclavijom@gmail.com

Mehul Bhakta
bhaktamehul@ufl.edu

Juan M. Osorno
Juan.Osorno@ndsu.edu

Idupulapati Rao
i.rao@cgiar.org

Stephen Beebe
s.beebe@cgiar.org

Elvin Roman-Paoli
elvin.roman@upr.edu

Abiezer Gonzalez
agonzalez111@gmail.com

James Beaver
james.beaver@upr.edu

Jaumer Ricaurte
j.ricaurte@cgiar.org

Raphael Colbert
Raphael.Colbert@my.ndsu.edu

[1] Agricultural and Biological Engineering Department, University of Florida, Gainesville, FL 32611, USA

of genotypes and these four QTLs were always detected with 50–90% probability. The final model was evaluated using leave-one-site-out method to assess the influence of site on node addition rate. These analyses provided a quantitative measure of the effects on NAR of common beans exerted by the genetic makeup, the environment and their interactions.

Introduction

AQ1 Developing cultivars that are well adapted to particular environments requires accurate characterization of genotype-by-environment interactions (GEI). In order to make the best breeding decisions given the complexity of mechanisms and biological processes underlying a phenotypic AQ2 trait including its responses to multiple environments, it is helpful to use adequate strategies to understand GEI. GEI has been defined in different ways. Often, it is regarded as the differential phenotypic performance of a genotype from others to different environments (Griffiths et al. 2000). GEI can also be identified by observing the matrix of genetic variance–covariances of the phenotypic trait across all environments (Malosetti et al. 2013). Here, GEI is considered as the presence of heterogeneity of genetic variance across environments, or as non-perfect genetic correlations of the genotypic performances across environments. The estimation of this genetic matrix requires the fitting of complex statistical models that simultaneously combine information from all tested field experiments by using a linear mixed-model framework together with genetic information (i.e., marker data) for the population studied (Beeck et al. 2010). A full understanding of the dynamics of GEI is a critical component of the best breeding strategies, which strive to identify genotypes best adapted to the targeted environments.

Traits that show continuous variation are usually complex and controlled by several genes. The genetic complexity of these traits can be dissected with the assistance of dense molecular-based linkage maps that allow scanning of the genome to identify loci exhibiting large effects for the trait, commonly referred to as quantitative trait loci (QTLs; Lander and Botstein 1989). Furthermore, conducting phenotyping experiments under multiple environments provides an opportunity to identify the sources of variations in a segregating population based on genetic (QTL), environmental, and QTL-by-environment interactions (QEI). These sources can be adequately identified and quantitatively characterized by a mixed model that can utilize explicit genotypic information (Boer et al. 2007). Combining genotypic and phenotypic data from multiple environments into this type of model represents a powerful approach to more accurately estimate the contribution to variation by the different sources affecting a trait. For example, multi-environment QTL mixed models have been used to identify QTLs for different traits in several species. For maize, application of these models led to the detection of drought resistance QTLs and QEI in the CIMMYT (International Maize and Wheat Improvement Center) drought stress trial by introducing genotypic and environmental covariates to explain genetic and GEI (Malosetti et al. 2004, 2013); van Eeuwijk et al. 2010). Similar analyses have been carried out in wheat (Mathews et al. 2008), sorghum (Sabadin et al. 2012), and in pepper where a multi-trait and multi-environment (MTME) model explained about 83% of the variation for total fruit dry weight from each plant (Alimi et al. 2013). Recently, Heslot et al. (2014) demonstrated the advantages of integrating environmental covariates and crop modeling into a genomic selection framework to predict GEI using a large winter wheat dataset. These new approaches provide insight into the architecture of GEI, and may improve the prediction of genotype performance based on climatic conditions.

The growth rate and duration of the vegetative phase of development of a plant may be important determinants of the success of the reproductive phase and thus affect crop yield. The developmental rate during the vegetative phase can be measured through the accumulation of plastochrons, a developmental unit that measures the time interval between the initiations of two successive leaves on the shoot apical meristem. The successive generation of leaves at the shoot apical meristem results in the production of successive phytomers; these are repeating units comprised of a node with an attached leaf, a subtending internode and an axillary meristem at the base of the internode (Sussex 1989). Thus, the number of flowers/inflorescences and branches may be directly proportional to the number of nodes produced in a plant. The rate of node production has been reported to depend on temperature, the genotype, and CO_2 levels (Vallejos et al. 1983; Reddy et al. 1994). Furthermore, the node addition rate (NAR) has also been associated with levels of miR156, squamosal-like proteins and cytochrome P450 genes in *Arabidopsis* (Schwarz et al. 2008; Wang et al.

[2] School of Forest Resources and Conservation, University of Florida, Gainesville, FL 32611, USA

[3] Agronomy Department, University of Florida, Gainesville, FL 32611, USA

[4] Horticultural Sciences Department, University of Florida, Gainesville, FL 32611, USA

[5] Department of Plant Sciences, North Dakota State University, Fargo, ND 58108, USA

[6] CIAT, Cali, Colombia

[7] University of Puerto Rico, Mayagüez 00682, Puerto Rico

Journal : Large 122 | Article No : 2871 | Pages : 15 | MS Code : TAAG-D-16-00400 | Dispatch : 13-3-2017

2008). In summary, previous research indicates that NAR is under genetic control, and can be influenced by the environment. Therefore, simultaneous characterization of the environmental effect on NAR along with the identification and characterization of the genes that control NAR can help lay the foundation for more complex growth traits, particularly those associated with leaf area index and the reproductive phase.

Common bean (*Phaseolus vulgaris*) is one of the world's most important food legumes representing a major source of protein and fiber for human consumption. Given current trends in population growth, demand for this crop, particularly in Latin America and sub-Saharan Africa, can be expected to grow at unprecedented levels (Echeverría 2014). To meet the growing demand for beans, plant breeders will have to develop high yielding new cultivars best adapted to sometimes harsh environments. Towards this goal, this study was conducted to identify the QTL exhibiting large effects on NAR, the environmental covariates that have a significant effect on NAR, and the GEI, which are all challenging factors to characterize in breeding programs. These objectives were achieved through the analysis of a recombinant inbred population grown in multi-environment trials (MET) using mixed-effect models. We present the construction of a QTL- and environment-based predictive model, in which the QTL consistency was evaluated through cross validation by genotypes and the final model was assessed through leave-one-site-out method.

Materials and methods

Field experiments

An extensive, multisite, phenotyping experiment was carried out with a recombinant inbred (RI) family of common bean. This family was obtained from a cross between the determinate Andean cultivar Calima (SINCS 2003) and the indeterminate Mesoamerican cultivar Jamapa (Voysest 2000). The parental genotypes have contrasting growth habits, organ sizes, branching patterns, and photoperiod sensitivity. This RI family comprises 188 lines, which were propagated by single seed descent to the 11th generation, and in bulk to the 14th generation ($F_{11:14}$), and it was planted at five different field sites, along with the parental lines, using a Latinized, row-column design with three blocks (3 plots for each genotype, and 6 to 9 plots for each parental line depending on the site). Details of the field sites are presented in Table 1. The field experiments were located in: Citra, FL (CT); Palmira, Colombia (PA); Popayan, Colombia (PO); Isabela, Puerto Rico (PR); and Prosper, North Dakota (ND). These were planted between March 2011 and May 2012. For each site, 50 seeds of each genotype were sown in each plot at a depth of 2.5–3 cm

Table 1 Summary of site and management practices at five sites where 128–188 genotypes of common bean were grown

	CT	PA	PR	PO	ND
Site	Citra, Florida, USA	Palmira, Colombia	Isabela, Puerto Rico	Popayan, Colombia	Prosper, North Dakota, USA
Latitude	29°39′N	03°29′N	18°28′N	02°25′N	47°00′N
Longitude	82°06′W	76°81′W	61°02′W	76°62′W	96°47′W
Growing season	Mar 24th, 2011–Jun, 2011	Nov 10th, 2011–Jan, 2012	Feb 6th, 2012–May, 2012	Mar 23rd, 2012–Jun, 2012	May 19th, 2012–Aug, 2012
Previous culture	Fallow	Beans	Beans	Fallow	Wheat
Soil texture	Sand	Clay	Clayey Kaolinite	Medium Loam	Silt Clay Loam
Fertilization (kg ha^{-1})	136 (N); 60 (P);112 (K)	40 (Urea)	55 (N-P-K:10-10-10)	96 (P); 129 (N); 80.3 (K)	No fertilizer
Irrigation	Central pivot sprinkler system	Rain fed	Drip	Rain fed	Rain fed
Plant density (plants m^{-2})	4.3	3.0	3.9	4.3	3.3
Row spacing (cm)	90	120	100	90	150
# Of blocks	3	3	3	3	3
Total # of genotypes[a]	188	174	128	178	176
Measurement frequency	Weekly	Weekly	Weekly	Weekly	Weekly

CT, PA, PR, PO, and *ND* represent corresponding sites of Citra, *FL* Palmira, *Colombia* Isabela, *Puerto Rico* Popayan, Colombia and Prosper, North Dakota, respectively

[a]Number of genotypes (RILs) for each site was limited by seed availability

Journal : Large 122 | Article No : 2871 | Pages : 15 | MS Code : TAAG-D-16-00400 | Dispatch : 13-3-2017

with planting density ranging from 3.0 to 4.3 plants m^{-2}. An average germination rate of about 80% was observed in the field, resulting in approximately 40 plants per plot that were available for data collection. All experiments were carried out under well-managed conditions with sufficient fertilizer and water, and pesticides and herbicides were applied as needed.

Phenotypic data

One plant per plot from each block, for each RIL and the two parental lines, was harvested on a weekly basis. Harvests were initiated after emergence of the first true leaf, and were carried out a total of 7–10 times at each site depending on plant availability. The number of nodes bearing trifoliates larger than 2.5 cm on the main stem was recorded for each plant at each harvest time. Nodes on the main stem emerged in a predictable pattern such that when plotted vs. calendar days after planting, an approximate linear segmented relationship was obtained for each genotype at each site. Node addition proceeded approximately in a linear fashion until it transitioned into a stationary plateau phase when the final node number had been reached. Hence, data of the first 3 harvests or 4 harvests were used to estimate the slope of the linear model for determinate and indeterminate RILs, respectively, in the R statistical software (V 3.1.2, 2014), and the final node numbers were obtained from the last few harvests. The slope corresponds to the average rate of node addition (NAR, node day^{-1}). To obtain the slope for each genotype at each site, nodes on main stem from harvested plants over the 3 plots were combined to calculate the average NAR (i.e., one NAR value per genotype per site), and they were used for multi-environment QTL mapping. However, NAR for each plot of a given genotype at each site (i.e., 3 NAR values per genotype per site) was also obtained for heritability analysis (Eq. 1).

Environmental covariates

Temperature, solar radiation, and rainfall were recorded at weather stations located near each field. Day length was obtained from United States Naval Observatory (USNO) (http://aa.usno.navy.mil/data/docs/RS_OneYear.php) and solar radiation data that were not recorded or missing from local weather stations were obtained from the NASA POWER database (http://power.larc.nasa.gov/). The environmental covariates considered were average day length (DL, hr), average daily temperature (TEMP, °C), maximum temperature (TMAX, °C), minimum temperature (TMIN, °C) and solar radiation (SRAD, MJ m^{-2} day^{-1}) as presented in Fig. 1 for the growing season at each site. However, for analytical purposes, all covariates were averaged for the time period from the 1st trifoliate to end of vegetative phase (3rd or 4th harvests depending on the site) at each site (i.e., one value for each environmental covariate per genotype per site, Supplemental Table S1).

Genotypic data

The QTL analyses on the MET data was carried out with a high-density linkage map (Bhakta et al. 2015), which was constructed with the experimental RI population using the genotyping-by-sequencing method described by Elshire et al. (2011). This linkage map contains 513 molecular markers distributed on 11 linkage groups with an average marker distance of 1.84 cM, which is considered as a good coverage for the population we are using.

Multiple-environment and QTL analyses

As a first step, broad-sense heritability (H^2) of NAR trait was calculated by fitting the following specific linear mixed model using individual site data:

$$NAR = \mu + block + row(block) + column(block) + g + e \quad (1)$$

where μ is the population mean of NAR (node day^{-1}) at each site; *block, row(block)* and *column(block)* are random effects of block, rows and columns within blocks, respectively; g is the random genotype effect of a RIL; and e is the random residual effect. The heritability was estimated by utilizing the generalized equation proposed by Cullis et al. (2006) as the experimental design was not a random complete block design. The equation is expressed as $H^2 = 1AvgDIFF/(2*Vg\text{^}2)$, with *AvgDIFF* the average of the variance for differences between genotypes, and *Vg* the estimate of the genetic variance, as implemented in Genstat v.18 (VSN International Ltd Hemel Hempstead, UK).

The second step consisted in fitting a MET model that combined experimental data from all sites. Row-column design effect was first excluded because it had little impact on NAR (data not shown). In addition, the heritability at PR site was found to be zero (Table 2) so the PR site data were removed for subsequent analyses. Therefore, the MET model was initially fitted using only NAR data for each genotype from the remaining four sites (CT, ND, PA, and PO; i.e., one data point per genotype per site). The methodology was described in Malosetti et al. (2013). Briefly, this consisted of modifying the model from Eq. 1 to consider the four sites by including a fixed effect of site, s, and a random effect of genotype-by-site, $g \times s$, as shown in Eq. 2.

$$NAR = \mu + s + g + g \times s + e \quad (2)$$

A factor analytic order 1 matrix was selected according to the Akaike Information Criterion (AIC) (Akaike 1974)

Journal : Large 122 | Article No : 2871 | Pages : 15 | MS Code : TAAG-D-16-00400 | Dispatch : 13-3-2017

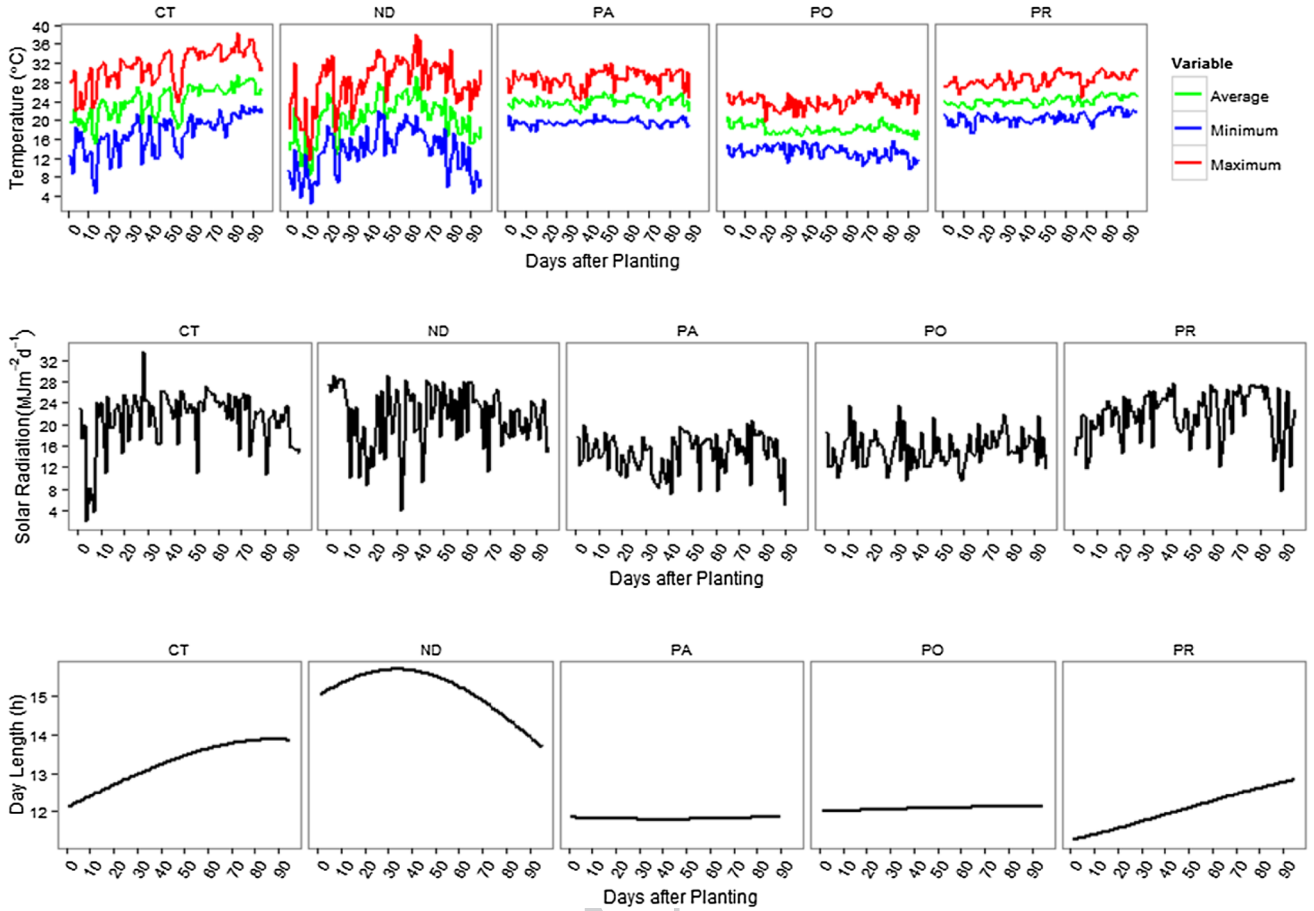

Fig. 1 The average, minimum, maximum temperatures (°C), day length (hrs) and solar radiation (MJ m^{-2} day^{-1}) at Citra, FL (CT), Palmira, Colombia (PA), Isabela, Puerto Rico (PR), Popayan, Colombia (PO) and Prosper, North Dakota (ND)

to model the variance–covariance component and was later used to specify the data structure in a genome-wide QTL scan using simple interval mapping, which evaluates each marker individually for significance (Lander and Botstein 1989). Then, the QTLs identified from simple interval mapping were specified as cofactors in composite interval mapping (Zeng 1994), and composite interval mapping was run at least three times consecutively to confirm stability of the fitted statistics profile.

In this study, the minimum separation distance for selected QTL was set to 3 cM and the minimum cofactor proximity was set to 50 cM based on current linkage map information, and significance level was set at $P<0.001$. The fitted model (QTL mixed site-effect model) from Eq. 2 would contain all significant QTL and QEI model terms as exemplified in Eq. 3,

$$NAR = \mu + s + \mathrm{QTL}_1 + \mathrm{QTL}_2 + \cdots + \mathrm{QTL}_1 \times s + \mathrm{QTL}_2 \times s + \cdots + e \quad (3)$$

where μ is the population mean across sites, s represents the site effect; QTL_1, QTL_2 etc. are assumed QTLs identified as having main effect; $\mathrm{QTL}_1 \times s$, $\mathrm{QTL}_2 \times s$ etc. are assumed QTL by site (Environment) interactions (i.e., QEI), and e represents the error term that was modelled by considering a factor analytic of order 1 variance–covariance error structure. A backward selection procedure was used to retain significant fixed terms ($P<0.05$). The above procedure was done using the QTL mapping framework described by Malosetti et al. (2013) as implemented in GenStat v.18.

QTLs and QEI consistency

The consistency and stability of the identified QTLs together with their interaction with site was examined through a tenfold cross validation by using different sets of genotypes. Specifically, all 171 RILs were randomly distributed into 10 groups, then QTLs analysis was conducted by dropping one group of genotypes at a time, that is, the QTLs would be identified based on 9 groups of genotypes each time and this process would be repeated 10 times to assess the reliability of these identified QTLs and their possible interactions with site.

Table 2 The average rate of node addition (NAR, node day^{-1}) and final number of nodes on main stem at each location for the two parental lines (Jamapa and Calima) and recombinant inbred lines (RILs) separated by indeterminate and determinate genotypes (n=3 for RILs; n=6-9 for parental lines)

Variables/site	CT*	PA	PR	PO	ND
NAR					
Jamapa	0.26 ± 0.04	0.35 ± 0.02	0.31 ± 0.03	0.20 ± 0.01	0.24 ± 0.03
Calima	0.17 ± 0.04	0.23 ± 0.04	0.20 ± 0.05	0.14 ± 0.02	0.13 ± 0.06
RILs*	0.26 ± 0.06^{ab}	0.34 ± 0.07^{a}	0.28 ± 0.08^{ab}	0.18 ± 0.04^{b}	0.24 ± 0.08^{ab}
Indeterminate**	0.29 ± 0.05	0.35 ± 0.04	0.29 ± 0.06	0.20 ± 0.03	0.28 ± 0.05
Determinate	0.22 ± 0.06	0.32 ± 0.09	0.26 ± 0.10	0.16 ± 0.03	0.20 ± 0.08
Final number of nodes					
Jamapa	14.7 ± 2.1	16.1 ± 2.2	15.2 ± 3.2	11.7 ± 2.3	18.5 ± 0.71
Calima	5.0 ± 1.0	7.8 ± 0.9	4.8 ± 1.3	6.3 ± 2.5	4.0 ± 0.0
RILs	8.8 ± 3.5	11.1 ± 3.5	8.9 ± 4.7	8.7 ± 3.0	10.8 ± 5.3
Indeterminate**	11.9 ± 1.9	13.8 ± 2.3	12.5 ± 3.7	10.9 ± 1.8	14.0 ± 2.7
Determinate	6.1 ± 2.1	8.03 ± 1.4	5.2 ± 2.0	6.2 ± 1.8	5.9 ± 2.5
Genetic variance (×10E-2)	0.36	0.30	0.00	0.08	0.33
Residual variance ×10E-2)	0.24	0.30	1.12	0.18	0.30
Heritability (H^2)	0.64	0.73	0.00	0.58	0.72

The genetic variance, residual variance and broad-sense heritability (H^2) are reported at each location. NARs were estimated as the slope with final number of nodes as the plateau of the segmental linear relationship between nodes on the main stem and calendar days

CT, PA, PR, PO and *ND* represent corresponding sites of Citra, *FL* Palmira, *Colombia* Isabela, *Puerto Rico* Popayan, Colombia and Prosper, North Dakota, respectively

*Different letters in a row indicated the significance (one-way ANOVA, $P < 0.05$) of the average values of the rate of node addition (*NAR*, node day^{-1}) across RILs

**The node addition rates (*NAR*, node day^{-1}) and final node number on the main stem between indeterminates and determinates were significantly different at all five sites ($P < 0.05$)

Integration of environmental information in modeling QTL effect

The specific covariates included in a linear form were DL, SRAD, TEMP, TMAX and TMIN. These were averaged over all three plots for each genotype at each site during the node addition period (Supplement Table S1). They were incorporated in the model by replacing the site term *"s"* and any of its specific QEI in Eq. 3. Briefly, the procedure initially consisted of including each environmental covariate, one at a time, and then performing a pre-selection of relevant terms that have biological meaning for NAR such as temperature and day length. Later, a full model (see Eq. 5 in Results) was fitted that followed a backwards selection that retained terms that were statistically significant ($P<0.05$) or that contributed to the understanding of the biological mechanisms. All covariates were evaluated to determine whether they had a main effect or interaction with one or more QTLs, and the factor analytic order 1 variance–covariance matrix was always retained as the genetic structure for NAR during this fitting process. Hence, the final model corresponds to a QTL-environment-based predictive model that depends uniquely on some QTLs, environmental covariates and, whenever it exists and is significant, their interactions. This model allows prediction of the NAR phenotype of specific genotypes (those described by the QTLs selected) for similar site with known weather/climatic conditions that are within the range of our experimental conditions.

Model evaluation

To assess the influence of each site on the final QTL-environment-based model, leave-one-site-out method was used to evaluate the quality of the crop model obtained based on environmental variables. Specifically, a QTL-environment-based model would be built up using three of the four datasets (CT, ND, PA and PO) and the remaining dataset would be used to evaluate the performance of the model. The evaluation of the prediction ability of the model was based on the correlation coefficient between the observed NAR and predicted NAR, the % bias and the RMSE.

Results

Phenotypic data analysis

The average NAR values and final node numbers on the main stem of the parental lines and all RILs at all five sites are presented in Table 2. As expected, the NARs of the two parental lines, Jamapa and Calima, were significantly different from each other ($P<0.05$) at all five sites, and the

Journal : Large 122 | Article No : 2871 | Pages : 15 | MS Code : TAAG-D-16-00400 | Dispatch : 13-3-2017

RILs showed clear continuous variation for this trait suggesting polygenic control (Table 2; Fig. 2). Overall, these results indicated that NAR is strongly affected by both the genotype and the environment. For instance, plants grown in PA had the highest average NAR (0.34 ± 0.07 node day^{-1}) with one node being added every 2.9 days, while those grown in PO had the lowest NAR (0.18 ± 0.04 node day^{-1}) with one node being added every 5.6 days. The NAR of Jamapa was about 50% higher in CT, PA and PR, but only 40% higher in PO than that of Calima, while ND had a disproportionate effect where NAR of Jamapa was over 80% higher than that of Calima. Although there was a clear difference between indeterminates and determinates at all five sites and the rates of indeterminates were higher than those of the determinate RILs in general, the differences in NAR between these two growth habits do not resemble the differences observed between the parents, suggesting that growth habit had an effect on NAR but could not explain all the variations in the RIL family. The lack of uniformity in the environmental NAR responses among the parents and RILs suggested the existence of GEI. The final number of nodes on the main stem also varied among sites, had a trend similar to that of NAR. NAR and final node number were highly correlated (correlation = 0.45, $P < 0.05$) across all 5 sites, with the specific correlations of 0.55, 0.59, 0.58. 0.52 and 0.21 at CT, ND, PA, PO and PR, respectively.

The frequency distributions (i.e., density plots) of NAR are presented in Fig. 2 for the entire population (left panel) and for the population separated by growth habit (right panel) at each site. These histograms display three important features: (1) continuous variation, (2) transgressive behavior in the RIL family, as many lines have NARs that are outside the parental range, with many significantly different from either parent ($P < 0.05$), and (3) the shape of the distributions appears to be influenced by the site. For example, at one extreme are CT and PA where the distributions approximated normality while at ND, PO and PR they appeared to be bimodal for the entire population with some

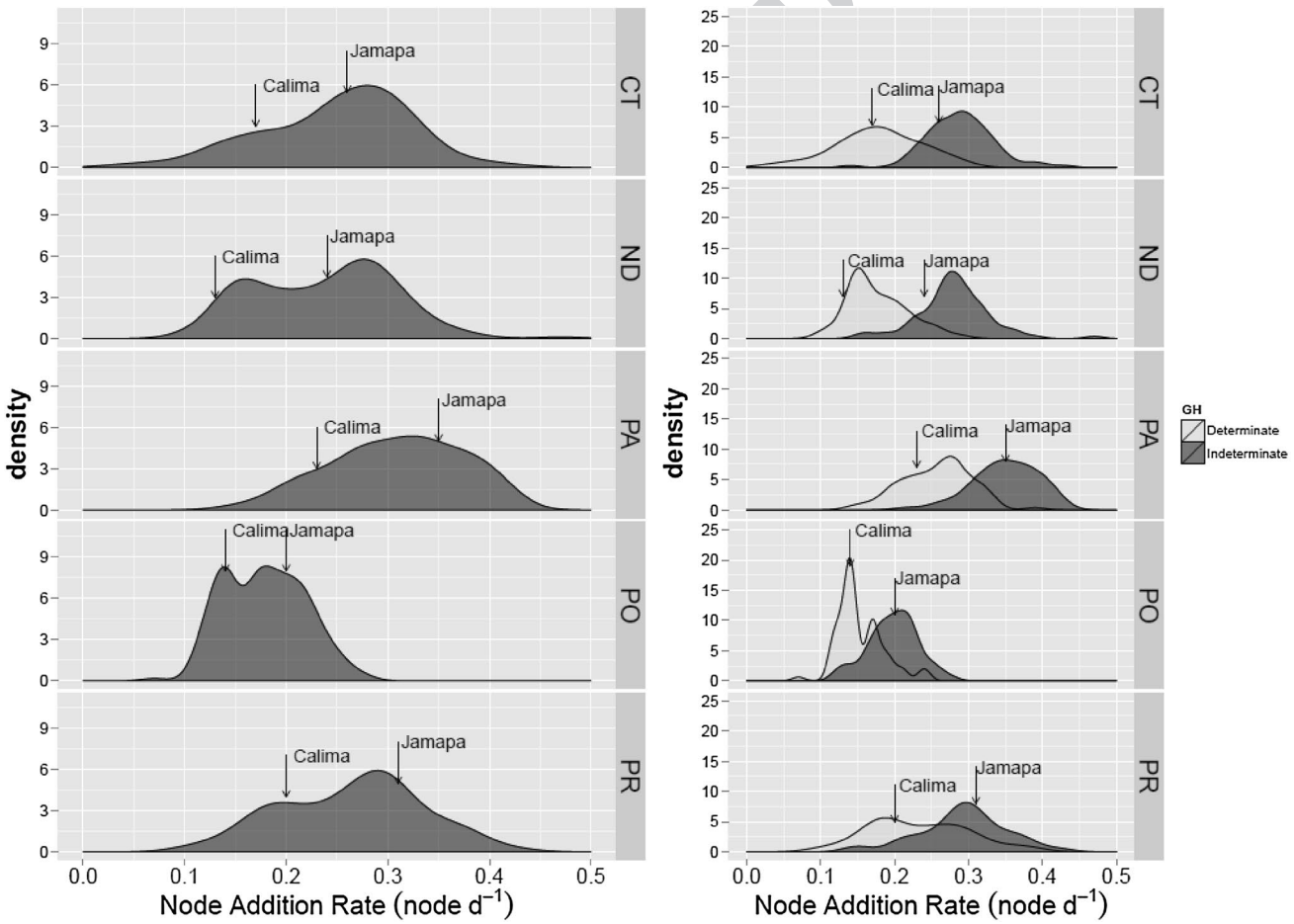

Fig. 2 Density plots of main stem node addition rate (NAR, node day^{-1}) for the entire recombinant inbred lines (RIL) population (*left panel*) and for the different growth habit (*right panel*) at each site. *CT, PA, PR, PO* and *ND* represent Citra, *FL* Palmira, *Colombia* Isabela, *Puerto Rico* Popayan, Colombia and Prosper, North Dakota, respectively

Journal : Large 122 | Article No : 2871 | Pages : 15 | MS Code : TAAG-D-16-00400 | Dispatch : 13-3-2017

overlapping between determinate and indeterminate genotypes (Fig. 2). Closer examination by separating the population according to growth habit revealed that the bimodality was caused to a great extent by the different growth habits. These density plots (Fig. 2) also show the heterogeneous variances across sites.

The broad-sense heritability (H^2) of the five sites averaged 0.53 with values of 0.64, 0.73, 0.000, 0.58 and 0.72 (Table 2) for CT, PA, PR, PO and ND, respectively (Table 2), while the row and column factors had little impact on NAR. Results from 4 out of 5 sites implied that NAR may be under genetic control indicating the appropriateness of conducting further QTL analysis. Since the heritability at PR site was zero, the PR data was dropped and not used for further QTL analysis. Subsequently, the correlations between sites were examined among the 4 sites (CT, PA, PO and ND) ranging from 0.53 to 0.72. Also, the biplot (Fig. 3) showed the high correlation among these sites, with the highest correlation between CT and PA (smallest angle, correlation of 0.72) and lowest correlation between CT and PO (largest angle, correlation of 0.53); similar length of the arrow for each site indicated similar phenotypic variances within each site; additionally, some RILs (light grey dots) had similar responses at each site (negative values on first dimension) while other RILs had different responses (positive values on first dimension), implying potential genotype by environment interactions.

QTL mixed site-effect model

Analysis of the multi-environment data (four sites data, CT, ND, PA and PO) revealed the presence of four NAR QTLs ($P < 0.001$). Three separate QTLs were detected on chromosome 1 (Nar1, Nar2, Nar3), and one on chromosome 7 (Nar4) (Fig. 4). This analysis also indicated that site had a significant impact on NAR. Furthermore, one QTL (Nar2) presented significant interactions with site. The QTL site-effect model is described by the following equation:

$$NAR = \mu + s + \text{Nar1} + \text{Nar2} + \text{Nar3} + \text{Nar4} + \text{Nar2} \times s + e \quad (4)$$

Details of the site effects, marker positions where the highest $-\log(P)$ values associated with the QTL region were found, the QTL main effects, and QEI effects are listed in Table 3. The QTL effects shown in the table are those for the Jamapa alleles. The effect of Jamapa alleles was not uniform across sites for QTL Nar2, displaying significant QEI. Jamapa alleles at Nar2 increased NAR at all sites, but their effects varied between 17% of the parental difference in PO to 6% in ND. Nar1, Nar3 and Nar4 did not show QEI and contributed to NAR uniformly across sites by 0.010, 0.009 and 0.008 node day^{-1}, respectively. The contribution by these loci represented approximately 6 to 16% of the parental differences. Overall, this QTL mixed site-effect model (Eq. 4) explained 73% of the phenotypic variation of NAR, with RMSE of 16.16% and bias of 0.00% (Fig. 5a).

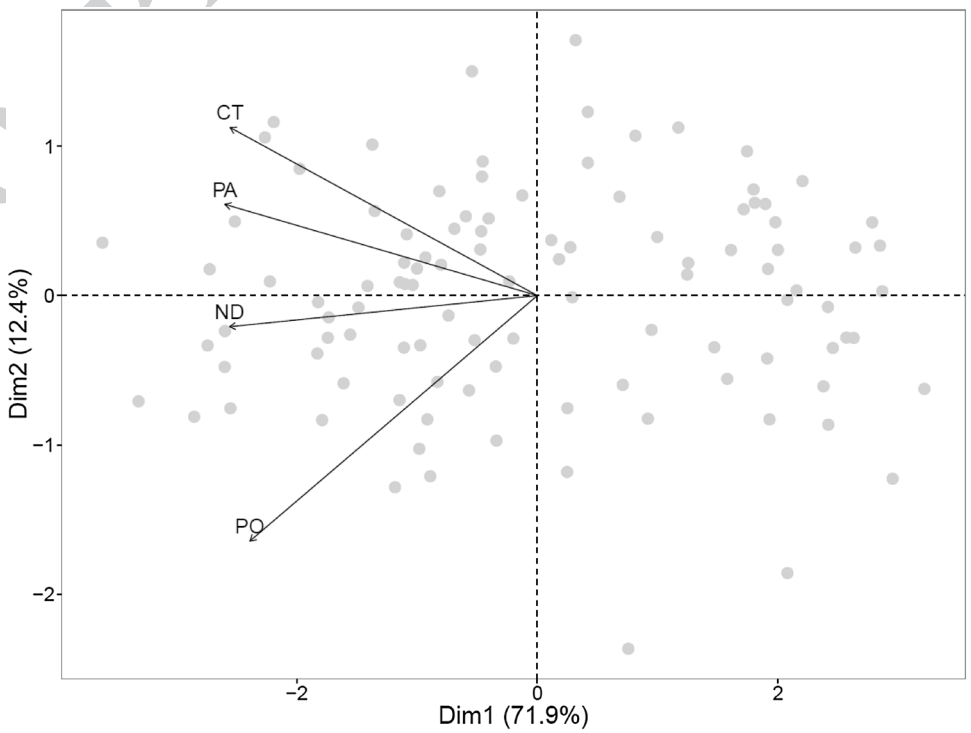

Fig. 3 Biplot between sites and genotypes showing the genetic correlation among the four sites (*arrows*) for NAR. Each grey dot represents one genotype, Dim1 and Dim2 are the first dimension and second dimension of the data, respectively, explaining the phenotypic variation. *CT, PA, PO* and *ND* represent Citra, *FL* Palmira, *Colombia* Popayan, Colombia and Prosper, North Dakota, respectively

Journal : Large 122 | Article No : 2871 | Pages : 15 | MS Code : TAAG-D-16-00400 | Dispatch : 13-3-2017

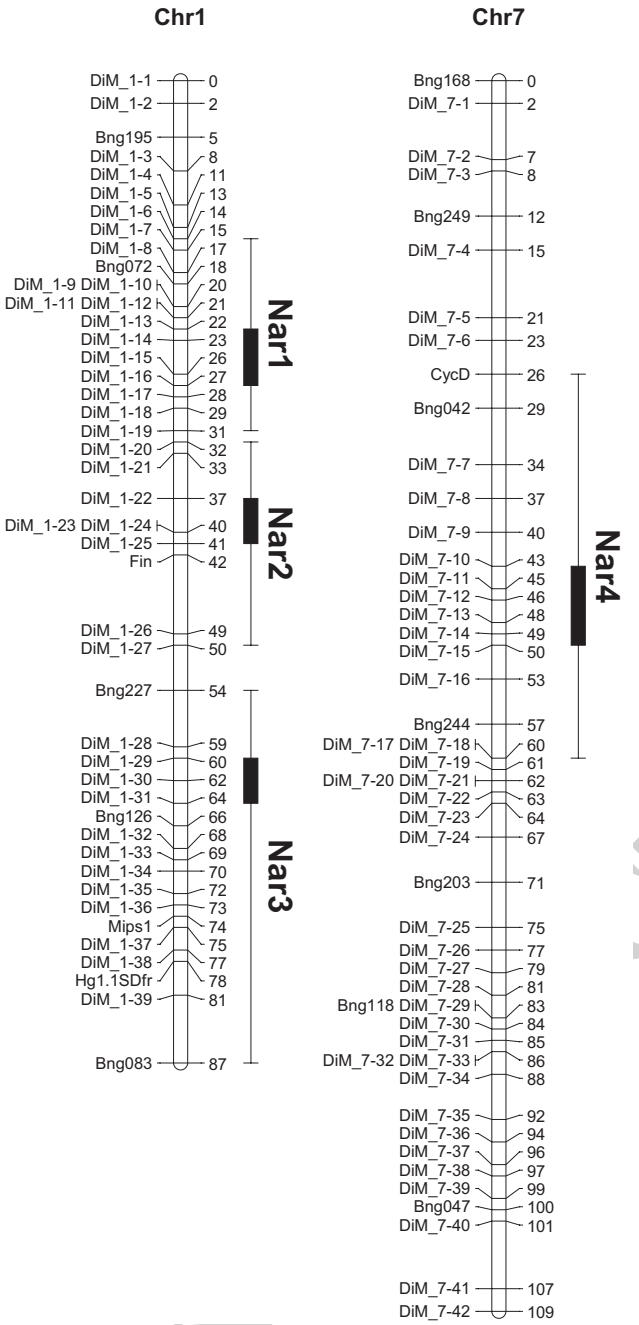

Fig. 4 QTL result from multi-environment QTL mapping for main stem node addition rate (NAR, node day^{-1}) in the common bean RIL population. Four QTLs were detected with Nar1, Nar2, Nar3 on chromosome 1 (Chr1) and Nar4 on chromosome 7 (Chr7). *Error bars* represent the corresponding marker positions within ±1.5 of the −log(P) value of the QTL peak (*black box*)

QTLs and QEI consistency

The QTL and QEI consistency and stability were checked through a tenfold cross validation by genotypes by randomly creating 10 groups of genotypes (i.e., 17 genotypes for 9 groups and one group of 18 genotypes), then dropping one group at a time and identifying QTLs for NAR based on the remaining 9 groups (Table 4). When dropping one group of genotypes from the analyses, the previously identified QTLs (Eq. 4 and Table 3) or the QTL regions near these identified QTLs were detected for most of the group analyses. For example, Nar2 and Nar3 were accurately detected 8 out of 10 (80%) and 9 out of 10 (90%) group analyses, respectively. It should be noted that while it seems new QTLs were identified with cross validation, some of these markers are closely located to the region of QTLs identified using all genotypes. For instance, a QTL identified on marker positions at 40.3 cM on chromosome 1 was very close to or in the range of the QTL position of Nar2 (marker position 42.1 cM). If considering the nearby region (i.e., ±1.5 of thelog$_{10}$(P) value of the highest peak, Fig. 3) as the same QTL regions, Nar2 was considered to be identified every single group analysis together with its interaction with site. Nar1, Nar4 were detected with 90% and 50% of the group analyses, respectively, when considering nearby QTL regions as the same (Fig. 3). However, one QTL on chromosome 11 and its interaction with site was also detected but in only 2 of the tenfold analyses (20%), and another QTL on chromosome 10 was detected just once. This is likely due to the fact that the weight of each group of genotypes was different allowing for the detection of a specific QTL. Overall, the four QTLs and the QEI effect identified across the four locations with all genotypes (Table 3) were considered to be consistent and were supported by the tenfold cross validation (Table 4).

QTL-environment-based predictive model with the inclusion of environmental covariates

The prediction of the QTL mixed site-effect model (Eq. 4) is restricted to the sites and conditions recorded during the experiment. To extend the predictability of the model, we used the mixed effects approach to extract information about the effects of specific environmental covariates for which we had collected data during the growing periods at each site. The objective of this analysis was to replace the "*s*" term effect with those significant environmental covariates. TEMP successfully explained part of the variation of Nar2 effects; partitioning the site effect ("*s*" term in Eq. 4) in the mixed site-effect model revealed that TEMP had the largest environment effect contributing 0.020 to NAR per °C (Eq. 5); in other words, an increase of 1 °C in the mean of daily average temperature would cause an increase of 0.020 nodes per day in NAR (i.e., 8% increase of the overall mean of 0.243 node day^{-1}). In contrast, SRAD and DL had minor, but significant effects. Details of the effects for significant environmental covariates, QTLs and QEI are

Journal : Large 122 | Article No : 2871 | Pages : 15 | MS Code : TAAG-D-16-00400 | Dispatch : 13-3-2017

Table 3 Marker genetic locations (in cM) with the highest −log(P) value associated with QTL region, site effects with standard error, and QTL by environment interactions (QEI) from the QTL mixed site-effect model (Eq. 4) for node addition rate (NAR, node day^{-1}) at each site

	Chr	Linkage position (cM)	$-\log_{10}$ (P)	CT	PA	PO	ND	QEI[a]
s (site effect)[b]				0.000 (0.004)	0.065 (0.004)	−0.063 (0.004)	−0.012 (0.004)	
Nar1[c]	1	17.80	4.979	0.010 (0.002)	0.010 (0.002)	0.010 (0.002)	0.010 (0.002)	No
Nar2	1	42.10	54.110	0.045 (0.004)	0.041 (0.004)	0.017 (0.003)	0.046 (0.004)	Yes
Nar3	1	87.10	5.503	−0.009 (0.002)	−0.009 (0.002)	−0.009 (0.002)	−0.009 (0.002)	No
Nar4	7	50.00	4.113	0.007 (0.002)	0.007 (0.002)	0.007 (0.002)	0.007 (0.002)	No

The site mean of 0.024 (0.004) node day^{-1} at Citra, FL is used as the baseline μ in the Eq. 4

CT, PA, PO and *ND* represent corresponding sites of Citra, *FL* Palmira, *Colombia* Popayan, Colombia and Prosper, North Dakota, respectively

[a]EI represents whether there exists a significant QTL by environment interaction based on AIC criteria

[b]Site effects were using CT mean as baseline

[c]Nar1, Nar2, Nar3 and Nar4 are the identified QTLs for node addition rate in common bean. For the same QTL, positive sign indicates that the Jamapa alleles accelerated NAR while negative sign indicates the Calima alleles accelerated NAR

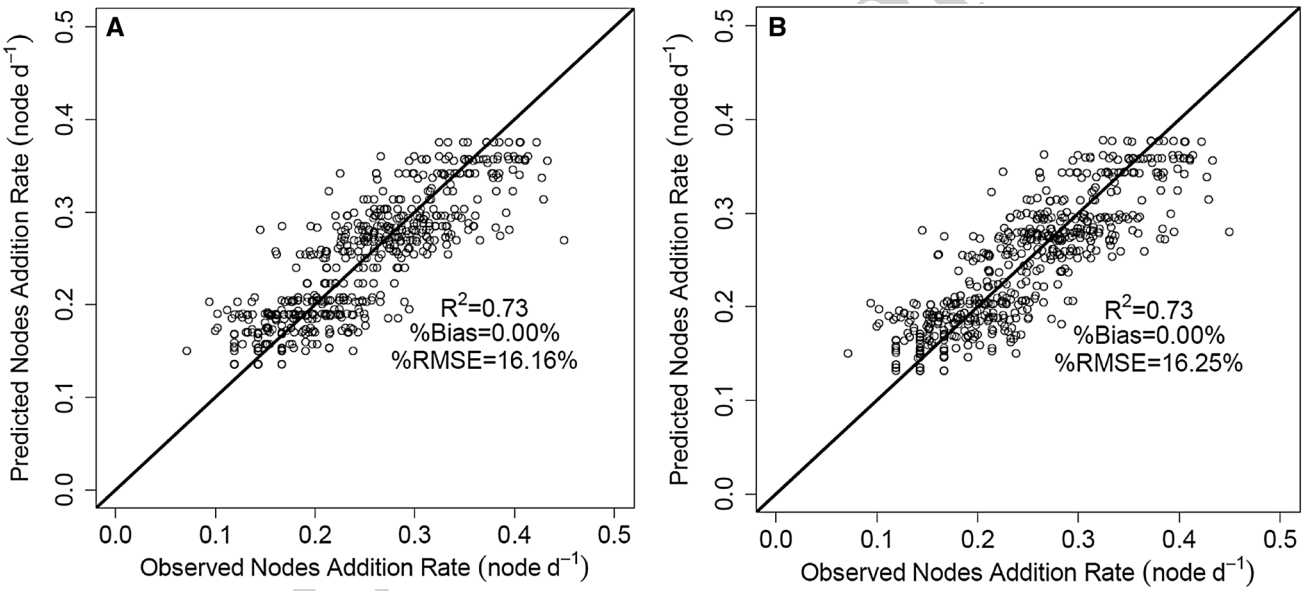

Fig. 5 **a** Predicted vs. observed main stem node addition rate (NAR, node day^{-1}) from the QTL mixed site-effect model (Eq. 4) for each RIL at all four sites. **b** Predicted vs. observed NAR from the QTL-and environment-based predictive model (Eq. 5) for each RIL at all four sites. Points on the solid 1:1 diagonal line represent equal predicted and observed NAR values for **a** and **b**

presented in Table 5. The final QTL-environment-based predictive model attained was as follows with fixed effects:

$$\widehat{NAR} = 0.243 + 0.020(\text{TEMP} - 20.95) - 0.005(\text{SRAD} - 16.26) - 0.004(\text{DL} - 13.05) - 0.009\text{Nar1} - 0.032\text{Nar2} + 0.009\text{Nar3} - 0.008\text{Nar4} - 0.004\text{Nar2}(\text{TEMP} - 21.51) \quad (5)$$

where all the terms were previously described and Nar1, Nar2, Nar3 and Nar4 are genotypic variables (QTLs) that take the value "1" for Calima alleles and "-1" for Jamapa alleles. The values 20.95, 16.26 and 13.05 are the average values of the environmental covariates of TEMP (°C), SRAD (MJ m^{-2} day^{-1}) and DL (hr) across all four sites during the node addition period. This QTL-environment-based predictive model represents the QTL by environment covariate interaction (i.e., QEI) as shown with Nar2×TEMP model terms. As mentioned earlier, an increase of 0.020 nodes per day in NAR would be expected per °C increase in the daily average temperature. Additionally, the NAR would increase by 0.008 units when the Calima alleles are replaced with the Jamapa alleles and vice

Journal : Large 122 | Article No : 2871 | Pages : 15 | MS Code : TAAG-D-16-00400 | Dispatch : 13-3-2017

Table 4 Marker locations (in cM) on a chromosome (Chr) that were identified through the tenfold cross validation by genotypes

Chr	Linkage position (cM)	1st[a]	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Percentage[c]
1	13.2			Y[b]								10
1	**17.8**		Y		Y					Y	Y	40
1	20.7	Y						Y				20
1	21.4					Y			Y			20
1	28.1						Y					10
1	40.3					Y			Y			20
1	**42.1**	Y	Y	Y	Y		Y	Y		Y	Y	80
1	78.2					Y						10
1	**87.1**	Y	Y	Y	Y		Y	Y	Y	Y	Y	90
7	34.4					Y						10
7	47.7			Y								10
7	**50.0**							Y	Y		Y	30
7	52.7	Y										10
10	13.6								Y			10
11	6.5			Y			Y					20

The chromosome number (Chr) and linkage position in bold are the QTLs identified using all genotypes across four locations., while the chromosome number (Chr) and linkage position with underline are the QTL by environment interactions (QEI) identified

[a]He 171 genotypes were randomly distributed into 10 groups, 1st represents dropping the 1st group of genotypes (10% of RILs) and performing the QTL analyses and continues through the 10th group

[b]"Y" indicates the QTL or QEI was identified when dropping the corresponding group of genotypes

[c]The percentage of group analyses (out of 10) that a QTL was identified

Table 5 Environmental covariates, QTLs and QTL by environment interactions (QEI) included in the QTL-environment-based predictive model (Eq. 5) and their effects with standard errors (SE) for node addition rate (NAR, node day^{-1}) in common bean

Significant term[a]	*F* statistic	*P* value	Coefficients (SE)
TEMP × Nar2[b]	63.23	<0.001	−0.004 (0.001)
Nar4	16.01	<0.001	−0.008 (0.002)
Nar3	21.82	<0.001	0.009 (0.002)
Nar2	171.79	<0.001	−0.032 (0.002)
Nar1	18.88	<0.001	−0.010 (0.002)
DL	13.90	<0.001	−0.005 (0.001)
SRAD	95.40	<0.001	−0.005 (0.001)
TEMP	1349.62	<0.001	0.020 (0.001)

[a]TEMP, DL, SRAD represent average daily temperature (°C), day length (hr) and solar radiation (MJ m^{-2} d^{-1}), respectively

[b]Nar1, Nar2, Nar3 and Nar4 are the identified QTLs for node addition rate in common bean

versa for the term Nar2 × (TEMP-20.95), that is twice the value of the coefficient (0.004) for the term. The model still explained 73% of the phenotypic variation of NAR with a RMSE of 16.25% and a bias of 0.00% (see Fig. 5b). The final model (Eq. 5) can be used to inform crop models on the effects of different environments on NAR.

Model evaluation

Leave-one-site-out analyses were conducted to assess the quality of the final QTL-environment-based models in the sense of crop modeling. For each of the QTL-environment-based models built upon three sites, temperature was always found playing a major role in affecting NAR not only as a main environmental covariate but also as interacting with the QTL Nar2; the four QTLs were also identified for most of the analyses when dropping individual site to detected QTLs (data not shown); in addition, these models explained 65–77% phenotypic variation of NAR. Figure 6 shows the comparisons between simulated NAR from the QTL-environment-based models based on three sites vs. the observed NAR from the fourth location. For instance, leave-one-site-out evaluation at CT means that, the QTL-environment-based model was built upon PA, PO and ND data, and CT data was used to evaluate the model. The average RMSE and bias are at 35.4 and 25.9%, respectively, with PO having the poorest performance.

Discussion

This study focused on the identification and characterization of the genetic and environmental factors that affect

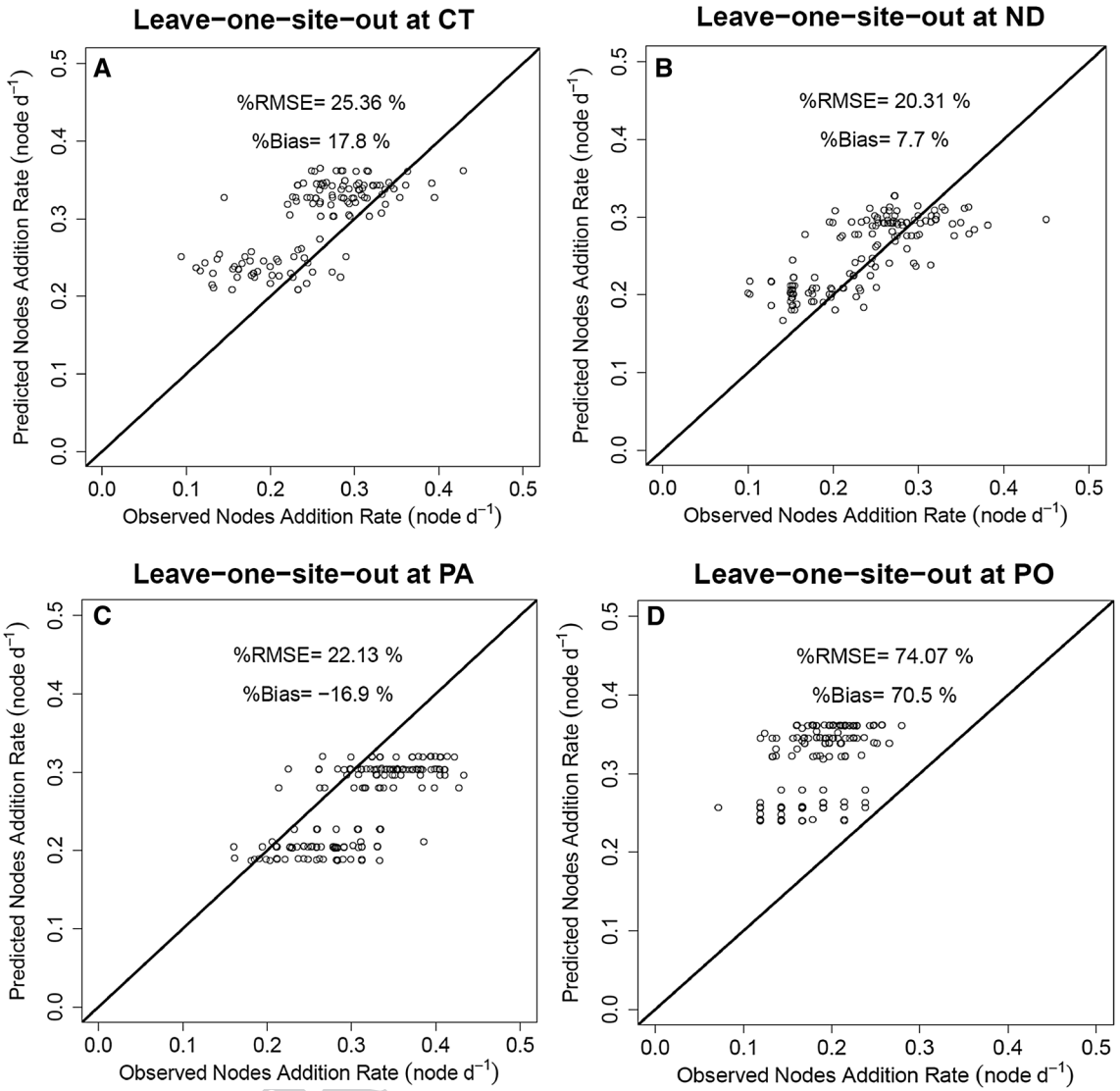

Fig. 6 Simulated vs. observed main stem node addition rate through leave-one-site-out method. "Leave-one-site-out at CT" indicates that, the QTL-environment-based model was built up using data from PA, PO and ND but was evaluated using data from CT (**a**). Similar interpretation applies for other locations based on what site was being evaluated, i.e. evaluated at ND (**b**), PA (**c**), or PO (**d**). *CT, PA, PO* and *ND* represent Citra, *FL* Palmira, *Colombia* Popayan, Colombia and Prosper, North Dakota, respectively

NAR in common bean. This is an important trait because it reflects the rate of development during the vegetative phase of the plant. The factors that control this rate can have an impact on the rate of dry matter accumulation and the onset of reproductive development. Density plots of NAR at each site (Fig. 2) and the genetic correlation among sites (Fig. 3) indicated potential polygenic control, strong environmental responses and transgressive segregation.

In the mixed site-effect QTL model (Eq. 2–4), Nar2 displayed significant interactions with temperature. Nar2 co-segregated with the *FIN* gene, which controls growth habit (Norton 1915), and has been identified as a homolog of the *Arabidopsis TFL1* gene (Repinski et al. 2012). *FIN* has been associated with the domestication syndrome including determinacy, number of nodes on main stem, number of pods and number of days to flowering in common bean (Koinange et al. 1996). Determining whether Nar2 and *FIN* are the same gene would provide new insight into the mechanism of node addition in common bean. In general, the bimodal distribution could be explained by the effect of a single gene. The bimodality observed within determinate plants in PO suggested there may exist another gene that interacts with the recessive allele of the potential *FIN* gene, however, no other QTLs were detected within the determinate plants, which may be caused by the small number of determinate

Journal : Large 122 | Article No : 2871 | Pages : 15 | MS Code : TAAG-D-16-00400 | Dispatch : 13-3-2017

RILs within the entire population, or the error propagation from data collection and NAR estimation which may contribute to a possible false peak detected within determinates.

Transitioning from the QTL-site effect model (Eq. 4) to the QTL-environment-based model (Eq. 5) did not lower the model efficiency (73%) indicating that almost all of the site effects were extracted as individual environmental covariates. The predictive ability of our model compares well with similarly constructed models of other species. For instance, in a recent study of the genetic architecture of maize height (Peiffer et al. 2014), the family-nested QTL models explained approximately 51–86% of the variation of maize height. As mentioned above, the unexplained variation of our model (Fig. 5) could be explained in part by genetic effects that were not detected by our analysis, or environmental variables such as soil moisture content that were not measured in these experiments.

The QTL-environment-based model showed that temperature was the main factor driving NAR in common bean. This is consistent with other studies on node addition in indeterminate common bean plants under different temperature and photoperiods (Wallace et al. 1991; Yourstone and Wallace 1990a, b), and is also supported by studies where temperature was used as the main or only factor affecting node/leaf addition for modeling purposes (Jones et al. 1999; Sinclair 1984; Soltani et al. 2006; Soltani and Sinclair 2012). For beans, an increase in temperature from 17 to 23 and to 29 °C caused increased node addition rate under 12 h of day length (Yourstone and Wallace 1990b). Day length plays a critical part in flowering and NAR in the common bean, an increase in day length from 12 to 14 h caused an increase in NAR under constant 23 °C but further increase to 16 h did not enhance NAR for selected indeterminate beans (Yourstone and Wallace 1990b). However, their results were based on removal of branches which may have altered the source sink relationship of the plant and have a secondary effect on node addition rate. Here, day length only had a minor effect on NAR. Solar radiation also had a minor effect with a small reduction in NAR (0.005 node day^{-1} per MJ m^{-2} day^{-1}). Nevertheless, as of now, it appears that for predictive models or crop models, temperature can be used as the main environmental factor affecting NAR in common bean.

The leave-one-site-out evaluation showed relatively low %RMSE and %bias for CT, ND and PA (Fig. 6) analyses, particularly when considering the fact that only a few QTLs were taken into account. However, the model when evaluated at PO was largely over predicted, which is likely caused by the much cooler temperatures at PO. Therefore, when building the model based on CT, ND and PA data, the low temperature effects were not captured in the model. These analyses demonstrate the need to have data from a broad range of environments to build the NAR model. Another limitation on these models are the fact they are linear and do not take into account the nonlinear relationships that often occur with a trait and the environmental covariates. Others have demonstrated that analyses that use nonlinear models for a trait can detect QTLs that are not found with a set of time-point analyses and that these non-linear models provide better understanding of the biological mechanisms of the trait (Malosetti et al. 2006; van Eeuwijk et al. 2010). The use of a nonlinear model for NAR will likely improve the model's predictability of this trait.

Predicting the phenotype from the genotype (G2P) is considered to be an essential outcome of the next generation of crop models. Models that can accurately predict plant growth and development based on the plant genotype and environmental data as inputs can solve the G2P problem. These types of models will not only help us understand the underlying mechanism of a trait or of an environmental response, but also help plant breeders design and identify suitable cultivars adapted to specific environments (White 2009; White and Hoogenboom 2003). GEI is a constant challenge in plant breeding programs. Thus, defining the means to identify and quantify this phenomenon will certainly facilitate breeding programs. Recently, research on GEI has been expanded to genotype-by-environment-by-management (GEM) (Asseng and Turner 2007; Montesino-San Martin et al. 2014). Such an expansion will help us both understand complex traits and improve final yield by combining QTL with proper management (e.g. row spacing, irrigation, etc.) under certain environments. Overall, the model we have developed could be incorporated into existing crop simulation models (Hoogenboom et al. 2012) AQ3 in an effort to convert them into gene-based simulation models that can provide a more comprehensive account of plant processes from planting to harvest using genotype and environmental data. AQ4

Author contribution statement LZ, JAC, MB, JMO, IR, SB, ERP, AG, JB, JR, RC collected and analyzed the data; SAG, JAC, MB, JWJ, KJB, CEV, JMO, IR, SB, ERP, AG, JB, JR, RC and MJC conceived and designed the project and experiment; LZ, SAG, JWJ, KJB, CEV and MJC interpreted the data; All authors were involved in writing and editing the paper.

Acknowledgements This research was funded by a National Science Foundation award IOS-0923975, the Institute of Food and Agricultural Science Seed Grant at the University of Florida, the USDA National Institute of Food and Agriculture Hatch Project (Correll #218999/005454), and the Chinese Scholarship Council (CSC). Special thanks to Mellissa Carvalho for providing help with statistical analysis, Christopher Hwang for assistance in cross validations, and all the people who helped collect the data or manage the fields.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

Akaike H (1974) New look at statistical-model identification. IEEE Trans Automat Control AC 19:716–723

Alimi NA, Bink MCAM, Dieleman JA, Magan JJ, Wubs AM, Palloix A, van Eeuwijk FA (2013) Multi-trait and multi-environment QTL analyses of yield and a set of physiological traits in pepper. Theor Appl Genet 126:2597–2625

Asseng S, Turner NC (2007) Modelling genotype x environment x management interactions to improve yield, water use efficiency and grain protein in wheat. Scale Complex in Plant Syst Res 21:93–103

Beeck CP, Cowling WA, Smith AB, Cullis BR (2010) Analysis of yield and oil from a series of canola breeding trials. Part I. Fitting factor analytic mixed models with pedigree information. Genome 53:992–1001

Bhakta M, Jones V, Vallejos CE (2015) Punctuated distribution of recombination hotspots and demarcation of pericentromeric regions revealed by a high resolution linkage map of *Phaseolus vulgaris* L. PLoS ONE 10(1): e0116822. doi:10.1371/journal.pone.0116822

Boer MP, Wright D, Feng LZ, Podlich DW, Luo L, Cooper M, van Eeuwijk FA (2007) A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. Genetics 177:1801–1813

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

Denis JB (1988) Two-way analysis using covariables. Statistics 19: 123–132

Echeverría RG (2014) Latin America and global food security: helping the whole world learn to fish. Ciatnews. http://www.ciatnews.cgiar.org/2014/04/22/latin-america-and-global-food-security-helping-the-whole-world-learn-to-fish/. Accessed 5 Fe 2015

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. Plos ONE. doi:10.1371/journal.pone.0019379

Griffiths AJF, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM (2000) An introduction to genetic analysis. New York

Heslot N, Akdemir D, Sorrells ME, Jannink J-L (2014) Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theor Appl Genet 127:463–480

Jones JW, Kenig A, Vallejos CE (1999) Reduced state-variable tomato growth model. Trans ASAE 42:255–265

Koinange EMK, Singh SP, Gepts P (1996) Genetic control of the domestication syndrome in common bean. Crop Sci 36:1037–1045

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

Malosetti M, Voltas J, Romagosa I, Ullrich SE, van Eeuwijk FA (2004) Mixed models including environmental covariables for studying QTL by environment interaction. Euphytica 137:139–145

Malosetti M, Visser RGF, Celis-Gamboa C, van Eeuwijk FA (2006) QTL methodology for response curves on the basis of non-linear mixed models, with an illustration to senescence in potato. Theor Appl Genet 113:288–300

Malosetti M, Ribaut J-M, van Eeuwijk F (2013) The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. Front Physiol doi:10.3389/fphys.2013.00044

Mathews KL, Malosetti M, Chapman S, McIntyre L, Reynolds M, Shorter R, van Eeuwijk F (2008) Multi-environment QTL mixed models for drought stress adaptation in wheat. Theor Appl Genet 117:1077–1091

Montesino-San Martin M, Olesen JE, Porter JR (2014) A genotype, environment and management (GxExM) analysis of adaptation in winter wheat to climate change in Denmark. Agric For Meteorol 187:1–13

Norton JB (1915) Inheritance of habit in the common bean. Am Nat 49:547–561

Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang Z, Millard MJ, Gardner CAC, McMullen MD, Holland JB, Bradbury PJ, Buckler ES (2014) The Genetic Architecture Of Maize Height. Genetics doi:10.1534/genetics.113.159152

R Core Team (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. ISBN 3-900051-07-0, URL http://www.R-project.org/

Reddy VR, Reddy KR, Acock B (1994) Carbon dioxide and temperature effects on cotton leaf initiation and development. Biotronics 23:59–74

Repinski SL, Kwak M, Gepts P (2012) The common bean growth habit gene PvTFL1y is a functional homolog of Arabidopsis TFL1. Theor Appl Genet 124:1539–1547

Sabadin PK, Malosetti M, Boer MP, Tardin FD, Santos FG, Guimaraes CT, Gomide RL, Andrade CLT, Albuquerque PEP, Caniato FF, Mollinari M, Margarido GRA, Oliveira BF, Schaffert RE, Garcia AAF, van Eeuwijk FA, Magalhaes JV (2012) Studying the genetic basis of drought tolerance in sorghum by managed stress trials and adjustments for phenological and plant height differences. Theor Appl Genet 124:1389–1402

Schwarz S, Grande AV, Bujdoso N, Saedler H, Huijser P (2008) The microRNA regulated SBP-box genes SPL9 and SPL15 control shoot maturation in Arabidopsis. Plant Mol Biol 67:183–195

Servicio Nacional de Inspección y Certificación de Semillas (SNICS) (2003) Catálogo nacional de variedades factibles de certificación. México. [In Spanish]

Sinclair TR (1984) Leaf-area development in field-grown soybeans. Agron J 76:141–146

Soltani A, Sinclair TR (2012) Modeling physiology of crop development, growth and yield. UK , Wallingford

Soltani A, Hammer GL, Torabi B, Robertson MJ, Zeinali (2006) Modeling chickpea growth and development: Phenological development. Field Crops Res 99:1–13

Sussex IM (1989) Developmental programming of the shoot meristem. Cell 56:225–229

Vallejos CE, Lyons JM, Breidenbach RW, Miller MF (1983) Characterization of a differential low-temperature growth-response in 2 species of lycopersicon - the plastochron as a tool. Planta 159:487–496

van Eeuwijk FA, Bink MCAM, Chenu K, Chapman SC (2010) Detection and use of QTL for complex traits in multiple environments. Curr Opin Plant Biol 13:193–205

Voysest OV (2000) Mejoramiento genético del frijol (Phaseolus vulgaris L.): legado de variedades de América Latina 1930–1999. Centro Internacional de Agricultura Tropical (CIAT), Cali. [In Spanish]

VSN International (2014) GenStat for Windows 17th Edition. VSN International, Hemel Hempstead

Wallace DH, Gniffke PA, Masaya PN, Zobel RW (1991) Photoperiod, temperature, and genotype interaction effects on days

Journal : Large 122 | Article No : 2871 | Pages : 15 | MS Code : TAAG-D-16-00400 | Dispatch : 13-3-2017

and nodes required for flowering of bean. J Am Soc Hortic Sci 116:534–543

Wang JW, Schwab R, Czech B, Mica E, Weigel D (2008) Dual effects of miR156-targeted SPL genes and CYP78A5/KLUH on plastochron length and organ size in Arabidopsis thaliana. Plant Cell 20:1231–1243

White JW (2009) Combining ecophysiological models and genomics to decipher the GEM-to-P problem. Njas-Wageningen J Life Sci 57:53–58

White JW, Hoogenboom G (2003) Gene-based approaches to crop simulation: Past experiences and future opportunities. Agron J 95:52–64

Yourstone KS, Wallace DH (1990a) Application of plastochron index to common bean grown in controlled environments. J Am Soc Hortic Sci 115:820–823

Yourstone KS, Wallace DH (1990b) Effects of photoperiod and temperature on rate of node development in indeterminate bean. J Am Soc Hortic Sci 115:824–828

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468

Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: genetic interactions create phantom heritability. PNAS doi:10.1073/pnas.1119675109

Journal : **Large 122**	Article No : **2871**	Pages : **15**	MS Code : **TAAG-D-16-00400**	Dispatch : **13-3-2017**

Journal:	**122**
Article:	**2871**

Springer
the language of science

Author Query Form

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

Query	Details Required	Author's Response
AQ1	Author: Please check and confirm that the authors and their respective affiliations have been correctly identified and amend if necessary.	
AQ2	Author: Please check and confirm the city, state, postcode and contry for all affiliations.	
AQ3	Author: References Denis (1988), R Core Team (2012), Zuk et al. (2012) are given in list but not cited in text. Please cite in text or delete from list.	
AQ4	Author: Reference Hoogenboom et al. (2012) is cited in text but not provided in the reference list. Please provide references in the list or delete these citations.	