Breakthroughs in the Genetics and Breeding of Capsicum and Eggplant

Torino, 2-4 September 2013

XV Meeting on Genetics and Breeding of Capsicum and Eggplant

EUCARPIA

EDITORS

Sergio Lanteri

Giuseppe Leonardo Rotino

Multivariate QTL analyses and predictions of yield related traits in pepper

Alimi N.A.[1,2], Bink M.C.A.M.[1], Palloix A.[2], Van Eeuwijk F.A.[1]

[1] Wageningen UR, Biometris, P.O. Box 100, 6700AC, Wageningen, The Netherlands

[2] INRA PACA, GAFL UR 1052, BP 94, 84143 Montfavet Cedex France

Abstract

Yield is a key trait in pepper and is often measured simultaneously with other traits over several environments. The understanding and genetic improvement of yield may benefit from the joint analysis of yield with its related traits simultaneously. Linear mixed models have emerged as a flexible approach that correctly model underlying variance-covariance structures among the traits and between environments simultaneously. In this study, we applied four different QTL approaches based on linear mixed model on five yield related pepper traits measured across four environments. We evaluated the performance of the approaches in terms of the number of QTLs detected for each trait and their explained variance. The QTL models are a single-trait single-environment approach (STSE), a multi-trait approach (MT), a multi-environment approach (ME), and a multi-trait multi-environment approach (MTME). We further compared prediction accuracies between STSE and MT models. The predictions were subjected to a five-fold cross validation. Our results showed that multi-trait and/or multi-environment QTL analyses are more powerful and effective to map pleiotropic QTL and QTL by environment interactions than performing STSE analysis. The multivariate models further showed improvement over STSE in terms of both number of QTLs and the explained variance. MTME clearly outperformed all the other methods. With MTME, nine QTLs explaining 51% of genetic variation were identified for yield in the autumn trial in Spain as against three, three and six QTLs explaining 37%, 29% and 43% from STSE, ME and MT analyses, respectively. The MT model for yield in SP2 had prediction accuracy of 0.53, against 0.42 from the STSE model. These results confirmed that multivariate analyses of traits have better capabilities to unravel complex traits than single trait approach. Our result showed that trait's prediction accuracy depends not only on prediction model of choice and traits genetic architecture but also on the environment.

Keywords: Pepper, Quantitative Trait Loci, Multi-Trait-Multi-Environment, Genetic Correlation, Pleiotropy, Prediction Accuracy.

Introduction

When measurements are obtained for several traits on a plant simultaneously, it is only proper to consider analysing such traits multivariately instead of univariate analyses. This is even more so as biological processes are usually interdependent. Also, good varieties are known to show combined optimal values for several traits simultaneously. Many of such traits are often genetically correlated and proper QTL mapping could show if the correlation is due to a QTL simultaneously affecting many traits i.e. pleiotropy. Also, for several uncorrelated traits, the overall experiment type-I error (α) can easily be computed. However, the formula for such computation is not correct if some of the traits are correlated (Weller et al. 1996). In univariate analysis of correlated traits, the sampling variances of estimated parameters tend to be higher and the hypothesis tests show lower statistical power. The joint analysis of multiple traits has been shown to improve the power and precision of QTL mapping. It has also helped in improving the selection of some primary traits with low heritabilities or difficult to measure by exploiting their genetic correlations with other traits (Jiang and Zeng, 1995).

Also, measurement for yield and other important traits of agronomic importance may be done across a number of environments and may thus exhibit genotype by environment interactions (GEI). When dealing with unravelling the genetic architecture of such traits, their QTLs have to be analyzed by considering the combination of the QTL under different environment using the so-called QTL by Environment Interaction (QEI) analysis (Boer et al. 2007). The use of QEI would allow QTL categorization according to the stability of their effects across different environments. A 'constitutive' QTL is consistently detected across most environments, while an 'adaptive' QTL is detected only in specific environmental conditions or increases in expression with the level of an environmental factor (Vargas et al. 2006).

Earlier studies in pepper breeding focused mostly on univariate analyses of traits in single environments (Alimi et al. 2013a; Barchi et al. 2009; Ben Chaim et al. 2006; Mimura et al. 2010; Rao et al. 2003). In this study, aside from the univariate analysis, we implemented three different multivariate QTL modelling strategies to analyse data on a recombinant inbred line (RIL) pepper population (Alimi et al. 2013b; Voorrips et al. 2010; www.spicyweb.eu). These QTL modelling strategies are multi environment (ME), multi trait (MT) and multi-trait multi-environment (MTME) analyses. We modelled genetic correlations within (between traits in a given environment) and between environments, and explicitly test the presence of QEI and pleiotropic QTLs. Furthermore, we investigated and compared the accuracies of predictions from STSE and MT models. The QTL analyses and prediction accuracies were evaluated with five yield related traits measured across four environments. The traits and environments were selected from the EU-SPICY data (Alimi et al. 2013a).

Materials and Methods

Genotypic and Phenotypic Data

The traits selected here are taken from the four SPICY project phenotypic experiments (Alimi et al. 2013a). The mapping population is from sixth generation (F_6) of the segregating recombinant inbred lines (RILs) of an intraspecific cross between the large – fruited inbred cultivar 'Yolo Wonder' (YW) and the pungent small-fruited cultivar 'Criollo de Morelos 334' (CM 334) of pepper plant. DNA was extracted from 149 RILs to produce information for 455 markers assembled into 12 pepper chromosomes, covering 1705cM. The majority of markers used are SNP and SSR markers.

Phenotypic evaluations of the RILs were carried out via designed experiments across two locations; Spain (SP) and Netherlands (NL). The trials were done under both spring (1) and autumn (2) weather conditions in 2009. This gave a total of four trials (i.e. environments); Netherland trial in spring (NL1), Netherland trial in autumn (NL2), Spain trial in spring (SP1) and Spain trial in autumn (SP2). The five studied traits are total dry weight fruit (DWF) which represents fruit yield for these experiments, mean increase in leaf area index per unit time (LAI) and light use efficiency (LUE) which is a measure of dry matter production per megajoule of intercepted radiation. Other traits include total number of harvested fruits (NF) and the proportion of total biomass due to fruit (PF). Phenotypic characteristics for these traits including trait distributions and heritabilities are given in Alimi et al. (2013a).

Univariate QTL Model

The single-trait single-environment (STSE) model was of the form:

$$y_i = \mu + \sum_{j=1}^{m} x_{ij}\beta_j + e_i, \qquad (1)$$

where y_i was the phenotypic response of genotype i, μ the population mean, β_j was the additive effect of marker j. Genetic predictors were calculated at all marker positions and intermediate positions for those marker intervals that were larger than 5cM, giving a total of 639 evaluation

points. The genetic predictor for genotype i at genomic evaluation point j is denoted by x_{ij}, and e_i was the residual term.

Multivariate QTL Model

Three different types of multivariate QTL models were implemented. These include

Multi-Environment (ME) QTL model where each trait was evaluated over the four trials with the aim of investigating genotype-by-environment interaction (GEI) and QTL-by-environment interaction (QEI). The ME model was of the form:

$$y_{ik} = \mu + E_k + \sum_{j=1}^{m} x_{ij}\beta_{kj} + g_{ik} + \varepsilon_{ik}, \quad (2)$$

where E_k was the environmental mean deviation from the population mean, β_{kj} was the environment-specific effect of the additive genetic predictor at evaluation point j, g_{ik} represented the genetic effect of genotype i for environment k, and ε_{ik} represented the non-genetic component. We assumed that the vectors $g_i = (g_{i1}, \ldots, g_{ij})$ follow a multivariate normal distribution with zero mean and an unstructured variance-covariance (VCOV) matrix G i.e. $g_i \sim N(0, G)$. This model accounts for genetic correlations between traits.

Multi-Trait (MT) QTL model: The MT model is a joint analysis of the five traits within each environment. The model specification is similar to that of ME except that E_k in equation 2 was replaced by T_p which represented the trait mean deviation from the population mean i.e.

$$y_{ip} = \mu + T_p + \sum_{j=1}^{m} x_{ij}\beta_{pj} + g_{ip} + \varepsilon_{ip}, \quad (3)$$

This model allows us to explicitly model genetic correlations between traits in each environment via pleiotropy by specifying unstructured VCOV matrix among each pair of traits.

Multi-Trait Multi-Environment (MTME) QTL model: This is a joint analysis combining all the traits across the four environments in a single mixed model analysis. ME and MT models were extended by allowing the response (y) to be a vector of trait-environment (TE) combinations.

$$y_{ir} = \mu + TE_r + \sum_{j=1}^{m} x_{ij}\beta_{rj} + g_{ir} + \varepsilon_{ir}, \quad (4)$$

The trait-environment mean deviation from the population mean was represented as TE_r while other model parameters are as explained above. With the MTME model, GEI and genetic correlations between traits were simultaneously modelled.

The number of significant QTLs and their explained variance were compared among the four models. QEI and pleiotropic effects were also investigated. The QTL analyses were performed using the QTL facilities in GenStat 15 (VSNi, 2012).

Finally, prediction accuracies were estimated for STSE and MT models. Prediction accuracy was defined as the correlation between phenotypic values and predicted breeding values. The predicted breeding values were obtained through a five-fold cross validation process with ten replications each. For each prediction runs, four of the five subsets were used as the training set for model fitting while the last subset were predicted based on the fitted individuals. This was repeated until all the subsets have been predicted.

Results and Discussion

Genetic Correlations and Comparison of results from QTL models

The pepper traits considered showed positive and mostly uniform correlations between environments (Table 1). These between-environment correlations were generally moderate to high, ranging from 0.34 for LUE between NL2 and SP1 (i.e. NL2.SP1) to 0.79 for LAI between SP1 and SP2 (i.e. SP1.SP2), with the majority of the correlations above 0.6. The results from the ME

analysis showed very significant non-crossover (quantitative) QEI (i.e. QTL effects between environments differ only in magnitude but not in direction) which supported the uniform correlations observed among environments. As an example, the ME analysis revealed three QTLs for DWF in most of the environments (except NL1 with two QTLs). Two of these QTLs were constitutive i.e. they showed consistent significant effects across the four environments.

Table 1 Trait genetic correlations between environments.

	NL1.NL2	NL1.SP1	NL1.SP2	NL2.SP1	NL2.SP2	SP1.SP2	Mean
DWF	0.72	0.60	0.61	0.53	0.62	0.58	0.61
LAI	0.73	0.76	0.70	0.67	0.75	0.79	0.73
LUE	0.64	0.45	0.60	0.34	0.64	0.36	0.51
NF	0.70	0.55	0.54	0.49	0.65	0.41	0.56
PF	0.69	0.65	0.67	0.54	0.72	0.57	0.64
Mean	0.70	0.60	0.62	0.51	0.68	0.54	0.61

Table 2 *Genetic correlation of traits within each environment*

Trait	DWF	LAI	LUE	NF
NL1				
LAI	0.07			
LUE	0.13	0.11		
NF	0.85	0.01	0.23	
PF	0.90	-0.18	-0.11	0.76
NL2				
LAI	0.19			
LUE	0.26	0.32		
NF	0.86	0.09	0.36	
PF	0.91	-0.07	-0.03	0.76

Trait	DWF	LAI	LUE	NF
SP1				
LAI	0.09			
LUE	0.40	-0.34		
NF	0.80	0.12	0.37	
PF	0.93	-0.18	0.34	0.74
SP2				
LAI	-0.04			
LUE	0.22	0.01		
NF	0.60	-0.01	0.36	
PF	0.89	-0.37	0.03	0.51

The within-environment correlations among traits were mostly consistent in sign and magnitude among the environments (Table 2). DWF, NF and PF were highly correlated. This is expected since there is usually a direct relationship between number of harvested fruits and fruit weight. Also, PF was computed from total fruit weight and total plant biomass. The MT analysis revealed pleiotropic QTLs which are consistent with genetic correlations among the traits. As an example, five QTLs were detected for yield in SP1 environment. All the five QTLs were also detected for PF while three of the QTLs also influenced NF, LAI and LUE. However, two of the QTLs showed crossover pleiotropy between yield and LAI and between yield and LUE.

Factorial combinations of traits and environments and their joint analysis through the MTME revealed a total number of 17 regions as harbouring putative QTLs (Figure 1). As the MTME model fully utilizes covariance structures between environments and among traits within environments, it thus increased the power of QTL detection with increased precision. Table 3 displays number of QTLs together with their explained variance for each of the five traits in the four environments using the four QTL methods. There is a clear gain in the number of QTLs and their explained variances using multivariate QTL methods over univariate method. Also, jointly accounting for correlations among environments and among traits within an environment result in detection of far more QTLs than simply accounting for correlation among environments or among traits within an environment alone. As an example, 3, 3, 5 and 8 QTLs were identified for DWF in SP1 using SE, ME, MT and MTME QTL methods respectively explaining about 25%, 40%, 41% and 42% of genetic variations respectively. Most of the QTLs picked up in simpler methods were also detected in more complex methods. The three QTLs picked up for DWF in SP1 by STSE method were also detected by ME, MT and MTME methods. Many of the yield increasing alleles originated from the large fruited YW parental line. Many of the additional QTLs detected in MTME were of small effects. Also the QTLs already detected from simpler methods had reduced explained variances in

MTME. This might be related to the so called "Beavis effect" as simpler models resulted in overestimation of some effect sizes (Beavis, 1997).

Figure 1 Significant QTL positions from MTME model for five yield related traits (DWF, LAI, LUE, NF & PF) across the four environments. The top section shows the *P*-values of tests for QTL main effects. The bottom section shows heat maps along the genome for each trait, where blue indicates QTLs with significant effect from YW allele while red indicates QTLs with significant effect from CM334 allele. Most of the QTLs showed pleiotropic and QEI effects.

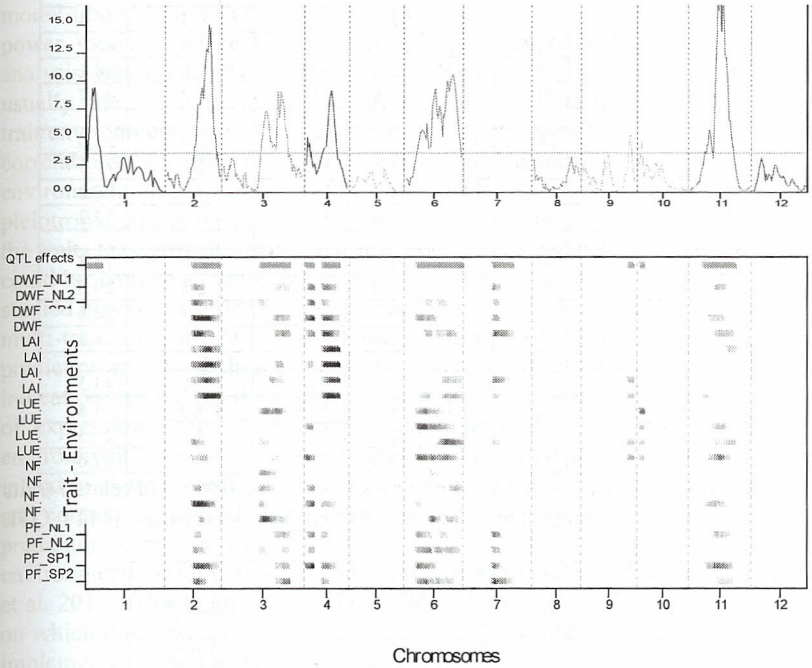


Table 3 Comparison of Number of QTLs (#QTL) and Explained Variance ($H^2_{(qtl)}$) from STSE, ME, MT and MTME models for five yield related traits measured across four environments.

		Number of QTLs (#QTL)				QTL Variance Explained ($H^2_{(qtl)}$)					
Trait	Method	NL1	NL2	SP1	SP2	NL1	NL2	SP1	SP2	Avg. #QTL	Avg. $H^2_{(qtl)}$
DWF	STSE	1	2	3	3	18	18	25	37	2.3	24.5
	ME	2	3	3	3	21.8	24.9	39.6	28.7	2.8	28.8
	MT	2	2	5	6	18.9	13.1	40.8	43.2	3.8	29.0
	MTME	5	7	8	9	26.1	34.7	42.3	50.6	7.3	38.4
LAI	STSE	2	3	2	2	33	48	22	42	2.3	36.3
	ME	4	4	6	5	37.4	49.7	57.8	42.7	4.8	46.9
	MT	2	4	5	4	30	39.9	31.6	50.9	3.8	38.1
	MTME	4	6	7	8	34.2	42.1	47.8	54.7	6.3	44.7
LUE	STSE	2	1	1	4	26	14	17	31	2.0	22.0
	ME	3	2	2	3	25.8	19.5	11.6	22.5	2.5	19.9
	MT	4	2	4	4	26.3	13.4	21.4	27.8	3.5	22.2
	MTME	5	5	7	8	26	22.4	41.6	38.9	6.3	32.2
NF	STSE	0	1	4	3	0	10	31	34	2.0	18.8
	ME	4	2	2	3	20.4	14	15.2	33.5	2.8	20.8
	MT	2	3	4	4	7.4	20	16.3	38.6	3.3	20.6
	MTME	2	4	6	4	10.6	24.3	33.4	27.9	4.0	24.1
PF	STSE	0	0	4	3	0	0	32	26	1.8	14.5
	ME	4	3	6	3	42.1	34.1	66.6	36.4	4.0	44.8
	MT	3	4	5	5	15.7	28.9	32.5	24.1	4.3	25.3
	MTME	6	7	7	7	32.5	41.3	37.3	39.3	6.8	37.6

Table 4: 5-fold with 10 replications cross validated predictive accuracies for five yield related traits in four environments using single-trait single-environment (STSE) and multi-trait (MT) QTL models.

	NL1		NL2		SP1		SP2		
Trait	STSE	MT	STSE	MT	STSE	MT	STSE	MT	Mean
DWF	0.21	0.16	0.28	0.11	0.47	0.52	0.42	0.53	0.34
LAI	0.48	0.40	0.67	0.61	0.37	0.46	0.59	0.64	0.53
LUE	0.44	0.46	0.32	0.27	0.27	0.39	0.28	0.45	0.36
NF	0.05	0.07	0.01	0.04	0.37	0.42	0.41	0.43	0.23
PF	0.10	0.15	0.33	0.16	0.38	0.51	0.29	0.47	0.30
Mean	0.26	0.25	0.32	0.24	0.37	0.46	0.40	0.50	0.35

Prediction Accuracies of STSE and MT models

Prediction accuracies of each of the traits (Table 4) showed direct relationship to QTL genetic architecture of the traits. Traits with higher explained variance in the QTL study were also better predicted and vice versa. Using STSE method in NL1, no significant QTL was picked up for NF, hence the trait was very poorly predicted (0.05), while two QTLs with 33% explained variability were picked up for LAI, hence the trait was highly predicted (0.48). Also, for each of the five traits, prediction accuracies differed among environments, irrespective of the method employed. Most of the traits were better predicted in Spanish trials than in Netherlands trials. For example, yield was better predicted in Spanish trials (0.42 - 0.53) than in NL trials (0.11 - 0.28). This could be caused by poor fruit set in NL trials. Furthermore, for each of the traits, prediction accuracies differ between STSE and MT models. It is expected that multi-trait model should have better predictive power than single-trait model. This is mostly true in our case for SP trials but not for NL trials especially NL2. This showed that in situations where phenotypic data were simultaneously collected on a large number of traits, using multivariate QTL method that properly model underlying variance-covariance structures among the traits would lead to improved predictive

power than performing single trait analyses. In multi-trait model, information sharing among correlated traits helped to increase prediction accuracies for traits with hitherto low accuracies e.g. prediction accuracy for LUE in SP2 increased from 0.28 to 0.45 when handled multivariately.

Concluding remarks

We showed that in situations such as the EU-SPICY project, where phenotypic data on a number of traits have been collected in multiple environments, using QTL methods that properly model underlying VCOV structures among the traits and between environments led to improved power to detect more QTLs than performing individual trait/environment analyses. The joint analysis was especially suitable for complex traits (such as yield) whose genetic variations are usually due to a large number of QTLs of smaller effects which might go undetected with single trait/environment analysis. The five traits considered showed positive and mostly uniform correlations between environments. Many of the detected QTLs showed quantitative QTL-by-environment interactions which corroborated uniform correlations between environments. Also pleiotropic effects were observed for many of the QTLs, which resulted from relationships between the traits they govern. Pleiotropy may also suggest redundancy between the measured traits, which could be avoided to decrease the cost of experiments. Such pleiotropic effects were more accurately studied by explicit modelling of the correlation/covariance structure among the traits through a joint multi-trait analysis. We also showed that predictive accuracy of traits depends not only on prediction model of choice and traits genetic architecture but also on the environment. This is an indication that though these traits are genetically determined in any given environment, their degree of expression differs from one environment to the other indicating presence of genotype-by-environment interaction and QTL-by-environment interaction. In furtherance of this study, we intend to explore prediction accuracies from univariate and multivariate genomic prediction models and compare with prediction accuracies from QTL models. We would also investigate an indirect prediction approach where yield is predicted from underlying component traits coupled with environmental information through crop growth modelling strategies (Chapman, 2008; van Eeuwijk et al. 2010). This approach has the ability to help in dissecting QTLs responsible for complex traits on which direct selection may be difficult Yield predictions from the crop growth model would be implemented within and across environments.

Acknowledgement

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 211347.

References

Alimi N.A., Bink M., Dieleman J., Magan J., Wubs, M., Palloix A., van Eeuwijk F. 2013b. *Multi-trait and Multi-environment QTL analyses of yield and a set of physiological traits in pepper.* Theoretical and Applied Genetics (Accepted).

Alimi N.A., Bink M. C. A. M., Dieleman J. A., Nicolaï M., Wubs M., Heuvelink E., Magan J., Voorrips R. E., Jansen J., Rodrigues P. C., Heijden G. W. A. M., Vercauteren A., Vuylsteke M., Song Y., Glasbey C., Barocsi A., Lefebvre V., Palloix A., Eeuwijk F. A. 2013a. *Genetic and QTL analyses of yield and a set of physiological traits in pepper.* Euphytica 190, 181-201.

Barchi L., Lefebvre V., Sage-Palloix A.-M., Lanteri S., and Palloix A. 2009. *QTL analysis of plant development and fruit traits in pepper and performance of selective phenotyping.* TAG Theoretical and Applied Genetics 118, 1157-1171.

Beavis W.D. 1997. *QTL analyses: Power, precision, and accuracy* In Molecular Dissection of Complex Traits ed. A. H. Paterson (Boca Raton, FL, CRC Press, 1997), 145–162.

Ben Chaim A., Borovsky Y., Rao G., Gur A., Zamir D., Paran I. 2006. *Comparative QTL mapping of fruit size and shape in tomato and pepper.* Israel Journal of Plant Sciences 54, 191-203.

Boer M.P., Wright D., Feng L. Z., Podlich D. W., Luo L., Cooper M., van Eeuwijk F. A. 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.

Chapman S. 2008. *Use of crop models to understand genotype by environment interactions for drought in real-world and simulated plant breeding trials.* Euphytica 161, 195-208.

Jiang C.J., Zeng Z. B. 1995. *Multiple-trait analysis of genetic-mapping for quantitative trait loci.* Genetics 140, 1111-1127.

Mimura Y., Minamiyama Y., Sano H., Hirai M. 2010. *Mapping for Axillary Shooting, Flowering Date, Primary Axis Length, and Number of Leaves in Pepper (Capsicum annuum).* Journal of the Japanese Society for Horticultural Science 79, 56-63.

Rao G.U., Ben Chaim A., Borovsky Y., Paran I. 2003. *Mapping of yield-related QTLs in pepper in an interspecific cross of Capsicum annuum and C. frutescens.* Theoretical and Applied Genetics 106, 1457-1466.

Smart tools for Prediction and Improvements of Crop Yield – KBBE 211347. 2009. www.spicyweb.eu (F. A. van Eeuwijk, ed.).

van Eeuwijk F.A., Bink M., Chenu K., Chapman S.C. 2010. *Detection and use of QTL for complex traits in multiple environments.* Current Opinion in Plant Biology 13, 193-205.

Vargas M., van Eeuwijk F., Crossa J., Ribaut J.M. 2006. *Mapping QTLs and QTL × environment interaction for CIMMYT maize drought stress program using factorial regression and partial least squares methods.* TAG Theoretical and Applied Genetics 112, 1009-1023.

Voorrips R.E., Palloix A., Dieleman J. A., Bink M.C.A.M., Heuvelink E., Heijden G.W.A. M. v. d., Vuylsteke M., Glasbey C., Barócsi A., Magán J., Eeuwijk F.A. v. 2010. *Crop growth models for the -omics era: the EU-SPICY project.* In "Advances in Genetics and Breeding of Capsicum and Eggplant : Proceedings of the XIVth EUCARPIA Meeting on genetics and breeding of Capsicum and Eggplant", pp. 315-321. Editorial Universidad Politécnica de Valencia, Valencia, Spain, Valencia, Spain.

VSNi. 2012. *GenStat for Windows 15th Edition.* VSN International, Hemel Hempstead, UK.

Weller J. I., Wiggans G. R., Vanraden P. M., Ron M. 1996. *Application of a canonical transformation to detection of quantitative trait loci with the aid of genetic markers in a multi-trait experiment.* TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik 92, 998-1002.