GRACE SUNSHINE DAVID

EFFICIENCY OF QUANTITATIVE TRAIT LOCI MAPPING UNDER GENOTYPE X ENVIRONMENT INTERACTION

Dissertation submitted to the Genetics and Breeding Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

Adviser: José Marcelo Soriano Viana

VIÇOSA - MINAS GERAIS
2022

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
D249e
2022

David, Grace Sunshine, 1985-
Efficiency of quantitative trait loci under genotype x environment interaction / Grace Sunshine David. – Viçosa, MG, 2022.
1 dissertação eletrônica (33 f.): il. (algumas color.).

Texto em inglês.
Orientador: José Marcelo Soriano Viana.
Dissertação (mestrado) - Universidade Federal de Viçosa, Departamento de Biologia Geral, 2022.
Referências bibliográficas: f. 31-33.
DOI: https://doi.org/10.47328/ufvbbt.2022.797
Modo de acesso: World Wide Web.

1. Marcadores genéticos. 2. Mapeamento cromossômico. 3. Interação genótipo-ambiente. I. Viana, José Marcelo Soriano, 1963-. II. Universidade Federal de Viçosa. Departamento de Biologia Geral. Programa de Pós-Graduação em Genética e Melhoramento. III. Título.

CDD 22. ed. 576.53

Bibliotecário(a) responsável: Euzébio Luiz Pinto CRB-6/3317

GRACE SUNSHINE DAVID

EFFICIENCY OF QUANTITATIVE TRAIT LOCI MAPPING UNDER GENOTYPE X ENVIRONMENT INTERACTION

Dissertation submitted to the Genetics and Breeding Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

APPROVED: December 15, 2022.

Assent:

Grace Sunshine David
Author

José Marcelo Soriano Viana
Adviser

ACKNOWLEDGEMENTS

To God almighty for the success of this program.

To the Universidade Federal de Viçosa, for such an opportunity to be part of this most amazing community in an institution of high repute. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

My deepest appreciation goes to my supervisor Prof. José Marcelo Soriano Viana for spending time to teach me, wise counsel, motivation, encouragement and constructive criticism, all endeared towards the success of this work. Thank you Prof., I have learnt so much and I am returning to my country better than I came.

To the Postgraduate Program in Plant Breeding and Genetics, for providing access to a course of excellence. Indeed, this has been an awesome experience and I am so glad for the opportunity to be here.

I specially acknowledge the coordinator of Post graduate Program in Genetics and Breeding, Prof. Felipe Silva, the seminar coordinator, Prof Rodrigo Lima and all the lecturers that taught me, Prof Guilherme da Silva among others. I specially acknowledge you Sir, thank you so much for your kind assistance and contributions towards the work. To Dr. Iara Santos, you are amazing, thank you for assisting me and responding at all times, helping me arrange the work, editting and the wonderful memories with your lovely family. I am so grateful and I pray for God's blessings upon you and your family.

To Prof. Kaio Olimpio Dias, thank you Sir for contributing significantly to my research, and the entire success of this work.

To the Forum for Agricultural research in Africa FARA / TETFUND / UFV partnership program for granting me a scholarship and giving me this opportunity to study in one of the best universities in Brazil. Sincerely, this program would not have been possible without the scholarship. I deeply appreciate this great initiative not just for researchers from Nigeria but also other parts of Africa.

To Dr. Yemi Akinbamijo the executive director of FARA and Dr. Abdulrazak Ibrahim, thank you for assisting me in various ways and your words of encouragement since the inception of this program. Despite the challenges and delay, your positive

attitude kept me going. Dr. Ibrahim, I appreciate all you did not just for me but for other scholars in Brazil who came through the scholarship program.

To the director of international affairs Prof Vladimir Di Iorio, who worked tirelessly for the success of this program. Thank you Prof. for all your efforts and help at all times. You made it so easy coming to Brazil, gave us all the necessary guidelines, assisted when our Visas were delayed by writing to the embassy, continual follow up and responding very promptly to mails. You are deeply appreciated Sir.

Worthy of acknowledgement is my husband Pst. Sunshine David, thank you dear for allowing me travel for this program and taking care of the kids in my absence. May God continually bless and keep you in safety.

To my kids, Jeffrey, Averter and Angel. Thank you for your prayers and understanding. I love you all.

To the management of the University of Calabar, Cross River state, Nigeria particularly the Vice Chancellor, Prof. Florence Obi for approving my study leave and granting me this opportunity to advance my career.

To Prof. Ausaji Ayuk (the former Dean of Faculty of Agriculture, University of Calabar, Nigeria), thank you so much Sir for everything, indeed words fail me. You followed up right from inception of this program till the end. I appreciate you Sir and may God grant the desires of your heart. I also acknowledge Prof Walter Binang, the present Dean Faculty of Agriculture, University of Calabar Nigeria.

To my mentor Prof. Donatus Uwah, I would not be where I am today in academics without you. You believed in me right from my undergraduate days and have always supported and encouraged me. I am grateful Prof. To Ass. Prof. Idorenyin Udo, you have always been there motivating and supporting all I do. I appreciate you.

To my mother of blessed memory, Mrs Grace Ukoha, your prayers brought me this far. I can never forget all your love and sacrifices; I miss you dearly. Keep resting in the bosom of your creator.

To my Dad, Mr Sunday Ukoha, my siblings Peace, Sarah and Samuel Ukoha my grandmother Mrs Charity Alom, Aunty Mercy and my nephews Donald and Gold, I

appreciate your support, prayers and motivation. To Mr Tom Godfrey, thank you so much for assisting in making this work a success.

Finally, to my friends, Ekaette Udoekong, Jumoke Kelechi, Vanessa Vitoriano, Mayara Silva, Eleniz Dias, Tarciso, Leticia Zamberlan, Leandro Henrique. Thank you all.

“O único lugar aonde o sucesso vem antes do trabalho é no dicionário”.

(Albert Einstein)

ABSTRACT

DAVID, Grace Sunshine, M.Sc., Universidade Federal de Viçosa, December, 2022. **Efficiency of quantitative trait loci (QTL) mapping under genotype x environment interaction.** Adviser: José Marcelo Soriano Viana.

Quantitative trait loci (QTL) mapping using simulation gives an opportunity to actually determine the number and positions of QTLs which cannot be done using field data. This study was carried out to assess the efficiency of quantitative trait loci mapping under genotype by environment interaction. In this investigation, we simulated 50 samples of 300 recombinant inbred lines (RILs) in six environments which were genotyped for 1000 single nucleotide polymorphism (SNPs) and phenotyped for grain yield. A total of six major and 190 minor QTLs (19 in each chromosome) were randomly distributed in the regions covered by the SNPs along ten chromosomes. The chromosome length was 200 cM and the average density was 2 cM. The average degree of dominance was 0.6. There were basically two scenarios for comparison in this investigation. The first (with genotype x environment effect) and the second (without genotype x environment effect). The QTL heritabilities ranged from 2.1 to 14% and the trait heritability across environments was within the range of 23 to 85%. The results across environments for the first scenario showed that QTL power of detection was 82% while bias and false positive rate (FPR) were 2.1 cM and 4.5% respectively. In the second scenario, power of detection was 86% while bias and FPR were 2.2 cM and 4.4% respectively. In the joint QTL mapping analysis, power of detection increased with higher QTL heritability and there was an effective control of false positive rate in the two scenarios. These results depict a real field data and shows the effectiveness of mapping QTLs across environment and its role in expression of quantitative traits.

Keywords: Power of QTL detection, false positive rate, mapping precision, genotype by environment interaction.

RESUMO

DAVID**,** Grace Sunshine, M.Sc., Universidade Federal de Viçosa, Dezembro de 2022. **Eficiência do mapeamento de locos de caracteres quantitativos (QTL) sob interação genótipo x ambiente**. Orientador: José Marcelo Soriano Viana.

O mapeamento de loci de traços quantitativos (QTL) usando simulação oferece uma oportunidade para realmente determinar o número e as posições dos QTLs, o que não pode ser feito usando dados de campo. Este estudo foi realizado para avaliar a eficiência do mapeamento quantitativo de locos de caracteres sob interação genótipo por ambiente. Nesta investigação, simulamos 50 amostras de 300 linhagens endogâmicas recombinantes (RILs) que foram genotipadas para 1000 polimorfismo de nucleotídeo único (SNPs) e fenotipadas para rendimento de grãos em seis ambientes. Um total de seis e 190 QTL de efeitos maiores e menores (19 em cada cromossomo), respectivamente, foram distribuídos aleatoriamente nas regiões cobertas pelos SNPs ao longo de dez cromossomos. O comprimento do cromossomo foi de 200 cM e a densidade média foi de 2 cM. O grau médio de dominância foi de 0,6. Havia basicamente dois cenários para comparação nesta investigação. A primeira (com efeito genótipo x ambiente) e a segunda (sem efeito genótipo x ambiente). As herdabilidades de QTL variaram de 2,1 a 14% e a herdabilidade de características entre os ambientes estava dentro da faixa de 23 a 85%. Os resultados entre ambientes para o primeiro cenário mostraram que o poder de detecção de QTL foi de 82%, enquanto o viés e a taxa de falsos positivos (FPR) foram de 2,1 cM e 4,5%, respectivamente. No segundo cenário, o poder de detecção foi de 86%, enquanto o viés e o FPR foram de 2,2 cM e 4,4%, respectivamente. Na análise conjunta do mapeamento de QTL, o poder de detecção aumentou com maior herdabilidade de QTL e houve um controle efetivo da taxa de falsos positivos nos dois cenários. Esses resultados retratam dados de campo reais e mostram a eficácia do mapeamento de QTLs sobre o ambiente e seu papel na expressão de caracteres quantitativos.

Palavras-chave: Poder de detecção de QTL, taxa de falsos positivos, precisão de mapeamento, interação genótipo por ambiente.

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

GxE	Genotype by environment interaction.
SNP	Single nucleotide polymorphism.
RILs	Recombinant inbred lines
DH	Double haploids
QTL	Quantitative trait loci.
FPR	False positive rate.
P_1	Parent 1.
P_2	Parent 2.
F_1	First filial generation.
F_2	Second filial generation.
Idx	Environment Index.
h^2	Heritability.
IM	Interval mapping.
CIM	Composite interval mapping.

LIST OF SYMBOLS

cM	Centimorgan.
%	Percentage.
g / plant	Grams per plant.
σ_G^2	Genotypic variance.
σ^2_P	Phenotypic variance.
σ^2	Error variance.
σ^2_E	Environment variance.
σ^2_{GXE}	Genotype by environment interaction variance.

SUMMARY

1. INTRODUCTION

Quantitative trait loci (QTL) mapping provides a beginning for dissecting the genetic architecture of complex traits. It helps us know how many QTLs significantly contribute to the variation of a trait in a given population. It also aids in our understanding of how much variation is due to the additive, dominance effects of QTL and also if there is an interaction between QTLs and the environment. Studies in this area contribute to the efficient use of marker assisted selection in breeding and genotype by environment interactions (El-Soda *et al*., 2014). QTL mapping method also helps to quantify the effects of QTL alleles as well as enhancing the breeding process (Dev Paudel *et al.,* 2020). Many quantitative traits have already been investigated using conventional methods but the quantitative trait loci approach results from a combination of recent crop genomics and conventional breeding, which accounts for a more advanced method of crop breeding. Marker assisted selection using QTL information evaluated in multiple environments is gaining popularity and it does not only yield higher response, but the response obtained is vast.

Different mapping populations such as $F_{2:3}$, recombinant inbred lines (RILs) and double haploid (DH) schemes have been used for QTL mapping studies. Several molecular markers have also been used for QTL mapping studies but the most frequently used is the single nucleotide polymorphism (SNP) (Wang *et al*., 2019). To detect QTLs and associations between molecular markers and traits of interest, statistical approaches such as interval mapping (IM) (Zeng, 1994), composite interval mapping (CIM) (Li *et al*., 2007) among others have been used.

Some authors have proposed simulations as a complement to statistical QTL mapping for interpretation of QTL by environment interaction (Van Eeuwijk *et al*., 2010; Wurschum, 2012). There may be some challenges when using actual mapping population for example, the true QTL positions and effects are usually unknown. Another problem with using real data to compare methods is that the parameter ranges are usually limited by experimental conditions, genetic architectures and crop species (Li *et al*., 2012). This therefore, justifies the use of simulations. Research works have been done on efficiency of QTL mapping using simulation studies. An efficient QTL mapping method should have

high QTL power of detection, a reduced false positive rate (FPR), and less bias in estimates of QTL effects (Li *et al*., 2007). Likewise, separation of effects of possible multiple linked quantitative trait loci (QTLs) on mapping QTLs is the key to increasing the precision of QTL mapping (Zeng, 1994). Nobari *et al*. (2012) assessed the efficiency of QTLs, assuming different levels of dominance. Their result showed that the power of QTL detection was strongly affected by the QTL heritability and the precision of QTL location was also proportional to the dominance effect. Wang *et al*. (2012) reported that the bias in the QTL position was influenced by the sample size, marker density, QTL effect, and true QTL location.

Some authors have also reported their findings on QTL studies across environments. Dev Paudel *et al*. (2020) reported that QTLs that are detected in multiple environments are called stable QTL and are reliable for marker assisted selection. Ren *et al*. (2021) in their study on 371 recombinant inbred lines (RILs) reported low heritability among the yield-related traits of wheat which shows that they were more affected by environmental factors. Their major findings were that they mapped two major QTLs for the kernel circumference and kernel area which were detected in multiple environments explaining 22.25 and 20.34% of the phenotypic variation respectively. Maniruzzaman *et al*. (2022) identified novel QTLs for development of new salt tolerant varieties in rice using $F_{2:3}$ populations. Some QTLs have been identified to improve drought tolerance with strong genotype by environment interactions in maize (Li *et al*., 2018, Hu *et al*. 2021) and common bean (Nabateregga *et al*., 2019). Wang *et al*. (2019) in their research on QTL by environment interaction of multi seed pod of peanut reported that multi-seed pod trait of peanut was mainly controlled by minor genes but the QTLs detected in multiple environments could be used for further studies. Hudson *et al.* (2022) in their study on maize also reported that genotype by environment interactions contributed much genotypic effects to the variation in some agronomic traits. They also identified a small number of significant markers associated with genotype by environment. Likewise, in a recent study on potato (Getahun *et al*., 2022) the authors observed that genotype by environment interaction was found to be important in nitrogen use efficiency of potato, they detected stable and environment specific QTLs for the related traits which could be used as indirect assessment for tuber yield. Also, novel regions underlying the variation

of bio energy related traits have been discovered in sorghum (Souza *et al*., 2021), the authors also reported the existence of genotype by environment interaction across environments.

Sustainability of production and food security largely depends on the adaptability of crops to various environments and their maximal performance across these environments. Considering a polygenic trait such as yield, evaluating genotypes in different environments is important because a high yielding variety in one environment may not necessarily be high yielding in another environment. Worthy of note in plant breeding is also the identification of environment-specific and stable QTLs having consistent genetic effects across a wide range of environments (Li *et al*., 2015).

Although some research works have been carried out on QTL mapping, methods, advantages and limitations, there is no adequate information on efficiency of QTL under genotype by environment interaction using simulations. Hence, the need for this study. Also simulation gives an opportunity to actually determine the number and positions of QTLs which cannot be done using field data. This therefore, justifies the use of simulations. It is important to continually evaluate the efficiency of QTLs and responses under genotype and environments. The outcome of this study will be used to acknowledge the roles of genotype by environment interactions particularly for complex polygenic characters.

1.1. Objective

To assess the efficiency of QTL mapping under genotype x environment interaction.

2. MATERIALS AND METHODS

2.1. Simulation

For this study, the parental inbred lines, F_1 and F_2 phenotypic and genotypic data was simulated using the *REALbreeding* software (Viana *et al.,* 2017). This software has been used in studies on genomic selection (Viana *et al*., 2018), QTL mapping (Viana, *et*

al., 2017) among others. The software simulates individual genotypes for genes, molecular markers and phenotypes in three steps using user inputs. The first step (genome simulation) is the specification of the number of chromosomes, molecular markers, and genes, as well as marker type and density. The second step (population simulation) is the specification of the population(s) and sample size or progeny number and size. A population is characterized by the average frequency for the genes (biallelic) and for the first marker. The last step (trait simulation) is the specification of the traits. In this stage, the user informs the minimum and maximum genotypic values for homozygotes, the minimum and maximum phenotypic values, the direction and degree of dominance, and the broad sense heritability. The phenotypic values (P) was computed assuming error effects (E) sampled from a normal distribution.

The mapping population was developed from a cross between two contrasting inbred parents; Inbred AA is high grain yielding while inbred aa shows low grain yielding. This generated a segregating population of 300 recombinant inbred lines (RILs) developed by single-seed descent (SSD) after selfing of individual F_2 plants. The number of environments was six. In this investigation, linkage map was constructed across the ten linkage groups using 1000 Single Nucleotide Polymorphism (SNPs). The length per chromosome was 200 cM at an average density of 2 cM (one SNP every 2 cM). The trait under consideration is grain yield of maize in g / 0.18 m^2. The QTLs and minor genes were randomly distributed in the regions covered by the SNPs. The number of QTLs were six and minor genes 190 (19 in each chromosome). The QTL positions were as follows: one at chromosome 1, two at chromosome 5 and three at chromosome 6. The QTL heritabilities ranged from 2.1 to 14% while the trait heritabilities over the environments ranged from 23 to 85%. The number of simulations was 50 and the average degree of dominance was 0.6.

2.2 Efficacy of QTL mapping

The efficiency of the QTL mapping was assessed based on the power of detection per segregating QTL, observed false positive rate (FPR) and bias in QTL position (mapping precision). Power of detection and FPR for most QTL mapping methods can be properly evaluated using computer simulations. (Li *et al*., 2012). In QTL mapping, power

tells the likelihood that a real QTL was detected and is therefore the most important indicator of a method's efficiency. The power of detection per QTL was calculated as the ratio of the number of times that a true QTL was declared to the total number of simulations. The FPR was also calculated as the ratio of false QTLs to the total number of QTLs declared by simulation. Two types of error can occur in QTL mapping, type I and type II, (Li *et al*., 2010). Type I error is a false positive, in which a segregating QTL is detected when in the real sense it is not present whereas type II error is a false negative, in which a QTL is not detected when it actually exists. The bias in the QTL position (mapping precision) per simulation was the average of the differences between the QTL estimated position and the true QTL position.

We assumed a positive evidence against null hypothesis H_0. To read each output file with the result from 50 simulations and to classify each estimated QTL, we used a program developed in the *REALbreeding* software. The classification of each QTL as true or false was performed using the difference between the position of the estimated QTL (peak) and the position of the true QTL (candidate gene). The estimated QTL was declared to be true when the difference was less than or equal to 20 cM. Li *et al*. (2012) used 10 and 20 cM and defined this difference as support interval, they also stated that the methods were compared based on the power of QTL detection (probability of rejecting null hypothesis H_0 when H_0 is false; control of type II error), FPR (control of type I error), and bias in the estimated QTL position (mapping precision).

2.3. Statistical analysis

The analyses were performed using the R/qtl packages software (Broman *et al.* 2003). For the maximum likelihood and least squares approaches we defined a threshold of 3.0 for the LOD score. In similar studies, Li *et al*. (2007; 2012) defined a LOD threshold of 2.5 or 3.0. We processed two data sets, one including the environment effects only and the other including environment, and genotype by environment effects. A total of 90,000 data was processed for each scenario. Environmental index was calculated as the difference between the phenotypic value in each environment and the mean value across environments. Individual and multi environment analysis was performed to test the

significance of genotypes and genotype by environment interaction. Average grain yield mean values across environments was used for QTL mapping.

3. RESULTS

In this investigation, we used 50 simulations of 300 RILs in six environments. This was done basically for two scenarios, the first with genotype x environment effect while the second was without genotype x environment effect. This data set gave very impressive results and depicts a real field investigation data. The trait heritability in both scenarios ranged from approximately 23 to 85% with the lowest value in environment six and the highest in environment four (Table 1). Environments one, two and five had positive environmental index value while four and six had negative values in both scenarios (Table 1). The phenotypic variance was generally higher than the genotypic variance across environments in both scenarios (Table 1). In the first scenario, phenotypic variance ranged from 479.2 to 1772.5 and in the second scenario the phenotypic variance ranged from 482.4 to 1482.3 with the lowest and highest values recorded in environment four and six respectively in both scenarios. Although, environment six had the highest phenotypic variance, it also recorded the lowest heritability of 23% and 27.5% in the first and second scenario. This clearly shows the influence of the environment on the expression of the trait. The genotype x environment variance was only recorded in the first scenario with values ranging from 61.0 to 562.6. It was also observed that there was no difference in the genotypic variance across the six environments in both scenarios.

The result from the multi environment analysis showed that the trait heritability was 90.2% for the first and 89% for the second scenario (Table 1). This clearly shows that the environment did not really affect the expression of the trait. The correlations between the environmental index and trait heritabilty was -0.03 and 0.36 for the first and second scenarios respectively. Regardless of QTL heritability which ranged from 2.1 to 14%, the power of detection, false positive rate and mapping precision all assumed positive dominance. In the single environment analysis for the first scenario, the power of detection for the six major QTLs ranged from 2 to 100% in environment one, 14 to 86% in environment two, 0 to 98% in environment three and four and 0 to 100% in environment

five and six (Table 2). The mapping precision ranged from 1.1 to 2.9 cM with the lowest value in environment five. The FPR ranged from 2.0 to 8.5%. Environments three and four had the lowest and highest FPR respectively.

There was a significant effect of QTL heritability on the environment. The correlation coefficient between power of detection and QTL heritabilities in all the environments was 0.03 for both scenarios (Figures 1 and 2). This shows that there was no association. This could have been because some of the QTLs were linked. Plotting the averages of both power of detection and QTL heritabilities gave a positive correlation of 0.62 and 0.66 for scenarios one and two respectively (Figures 3 and 4). This means that the power of detection was directly proportional to the QTL heritability. For the second scenario, power of detection for environment one ranged from 18 to 94%, and as high as 100% for environment two, four and six (Table 2). The bias ranged from 1.9 to 2.7 cM with the lowest and highest values in environments five and one respectively. FPR ranged from 3.2 to 11.1%.

Table 1. Parents (P_1 and P_2), F_1, and RILs average genotypic values (g/plant), environment (Env), environment index (Idx), genotypic (σ_G^2), environment (σ_E^2), GxE interaction (σ_{GxE}^2), error (σ^2), and phenotypic (σ_P^2) variances in each environment, and heritabilities (h^2; %), assuming GxE interaction and no GxE interaction.

Scenario	P_1	P_2	F_1	RIL	Env.	Idx	σ_G^2	σ_E^2	σ_{GxE}^2	σ^2	σ_P^2	h^2
GxE	160.0	40.0	144.0	99.9	1	3.2	407.1	57.5	186.6	901.5	1355.1	30.0
					2	1.1	407.1	75.3	61.0	607.2	954.0	42.7
					3	−0.6	407.1	199.6	87.7	394.3	1032.0	39.5
					4	−3.2	407.1	124.1	63.7	270.0	479.2	84.9
					5	22.7	407.1	28.4	562.6	167.3	1453.5	28.0
					6	−23.2	407.1	683.7	104.0	129.4	1772.5	23.0
				Across		-	407.1	-	-	-	451.7	90.2
No GxE	160.0	40.0	144.0	99.9	1	2.8	407.1	57.5	-	891.0	1351.9	30.1
					2	1.1	407.1	75.3	-	599.9	1037.2	39.2
					3	9.1	407.1	199.6	-	394.6	831.0	49.0
					4	−1.7	407.1	124.1	-	271.0	482.4	84.0
					5	11.2	407.1	28.4	-	171.0	700.5	58.1
					6	−22.5	407.1	683.7	-	98.8	1482.3	27.5
				Across		-	407.1	-	-	-	457.8	89.0

Table 2. QTL heritability (h^2; %) and average QTL power of detection (%), FPR (%), and bias (cM) in the positioning the true QTLs, in each environment and across the environments, assuming GxE interaction and no GxE interaction.

Env.	QTL	GxE				No GxE			
		h^2	Power	FPR	Bias	h^2	Power	FPR	Bias
1	1	2.1	48	6.3	2.9	2.1	18	5.4	2.7
	2	3.5	4			3.5	22		
	3	3.0	2			3.0	22		
	4	4.7	62			4.7	94		
	5	4.1	100			4.1	72		
	6	2.5	2			2.5	40		
2	1	3.0	50	8.2	2.9	2.7	62	7.7	2.5
	2	5.0	14			4.6	86		
	3	4.3	86			3.9	14		
	4	6.6	86			6.1	100		
	5	5.8	82			5.3	14		
	6	3.6	14			3.3	4		
3	1	2.7	2	2.0	1.2	3.4	2	3.2	2.1
	2	4.6	98			5.7	78		
	3	3.9	2			4.9	22		
	4	6.1	4			7.6	78		
	5	5.4	2			6.6	4		
	6	3.3	0			4.1	34		
4	1	5.9	96	8.5	2.0	5.8	12	11.1	2.0
	2	9.9	94			9.9	0		
	3	8.5	8			8.4	100		
	4	13.2	98			13.2	96		
	5	11.5	0			11.5	22		
	6	7.1	6			7.1	0		
5	1	1.9	0	4.5	1.1	4.0	6	4.7	1.9
	2	3.3	0			6.8	58		
	3	2.8	100			5.8	42		
	4	4.4	100			9.1	62		
	5	3.8	0			7.9	62		
	6	2.3	78			4.9	2		
6	1	1.6	28	2.3	1.8	1.9	94	3.2	2.1
	2	2.7	96			3.2	36		
	3	2.3	4			2.7	52		
	4	3.6	2			4.3	0		
	5	3.1	100			3.7	100		
	6	1.9	0			2.3	0		
Across	1	2.1	82	4.5	2.1	2.6	86	4.4	2.2
	2	3.5	48			4.3	44		
	3	3.0	52			3.7	56		
	4	4.7	46			5.8	52		
	5	4.1	72			5.0	74		
	6	2.5	14			3.1	16		

Comparatively, it was generally observed that the mapping precision and QTL power of detection were higher in the analysis done without genotype x environment effect (Table 2). In the multi environment analysis for the first scenario, QTL power of detection was as high as 82% while bias and FPR had values of 2.1 cM and 4.5% respectively while for the second scenario, QTL power of detection was as high as 86% while bias and FPR had values of 2.2 cM and 4.4% respectively.The result of the analysis for the second scenario without genotype x environment effects was generally superior for power of detection but the bias and FPR in both scenarios had differences of just about 0.1.

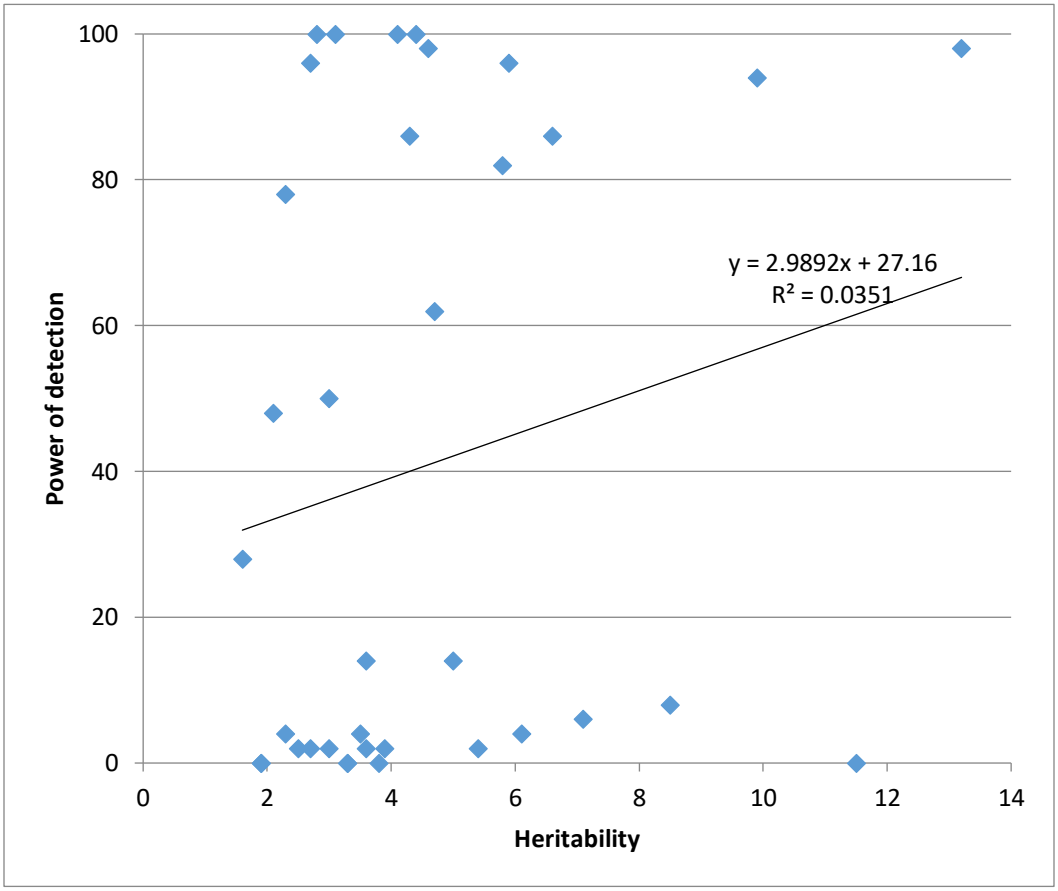

Figure 1: Correlation between the power of detection and QTL heritabilities in all the environments **assuming genotype by environment interaction**.

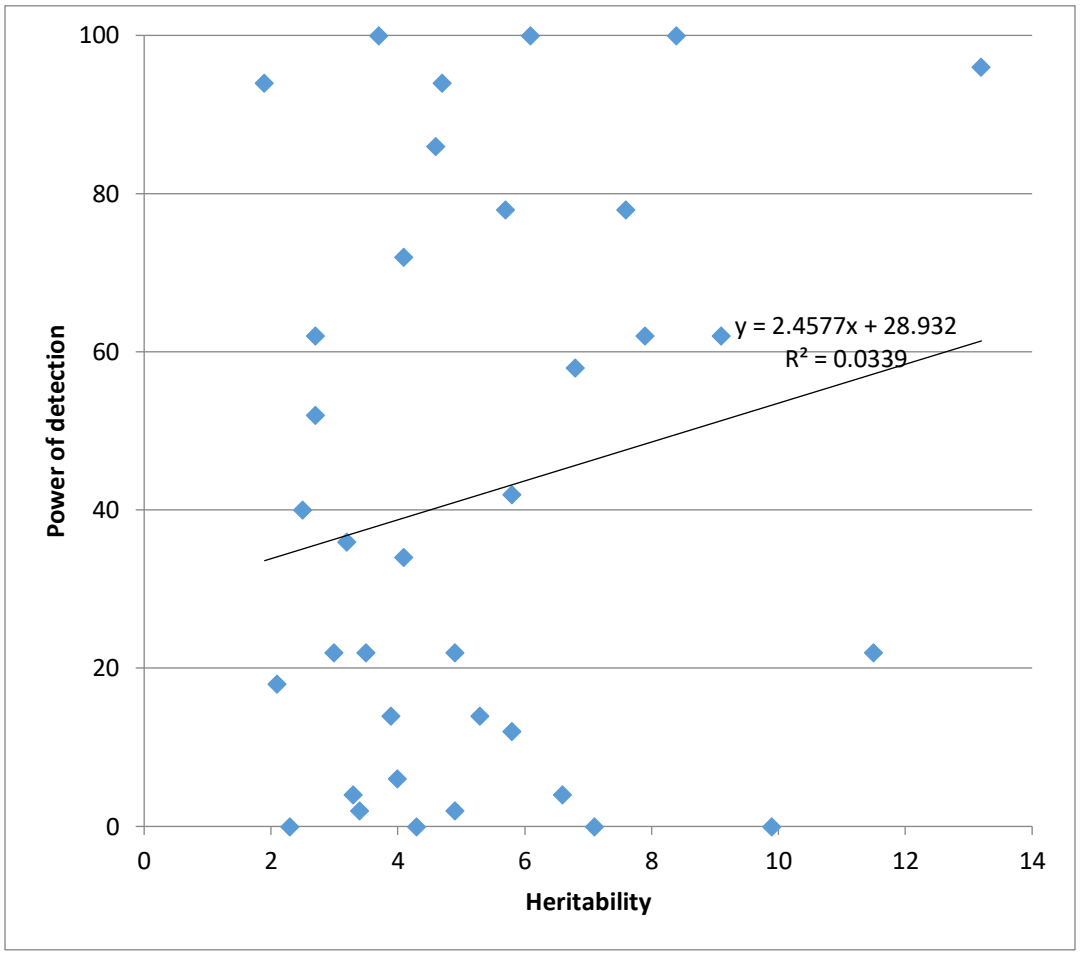

Figure 2: Correlation between the power of detection and QTL heritabilities in all the environments **assuming no genotype by environment interaction**.

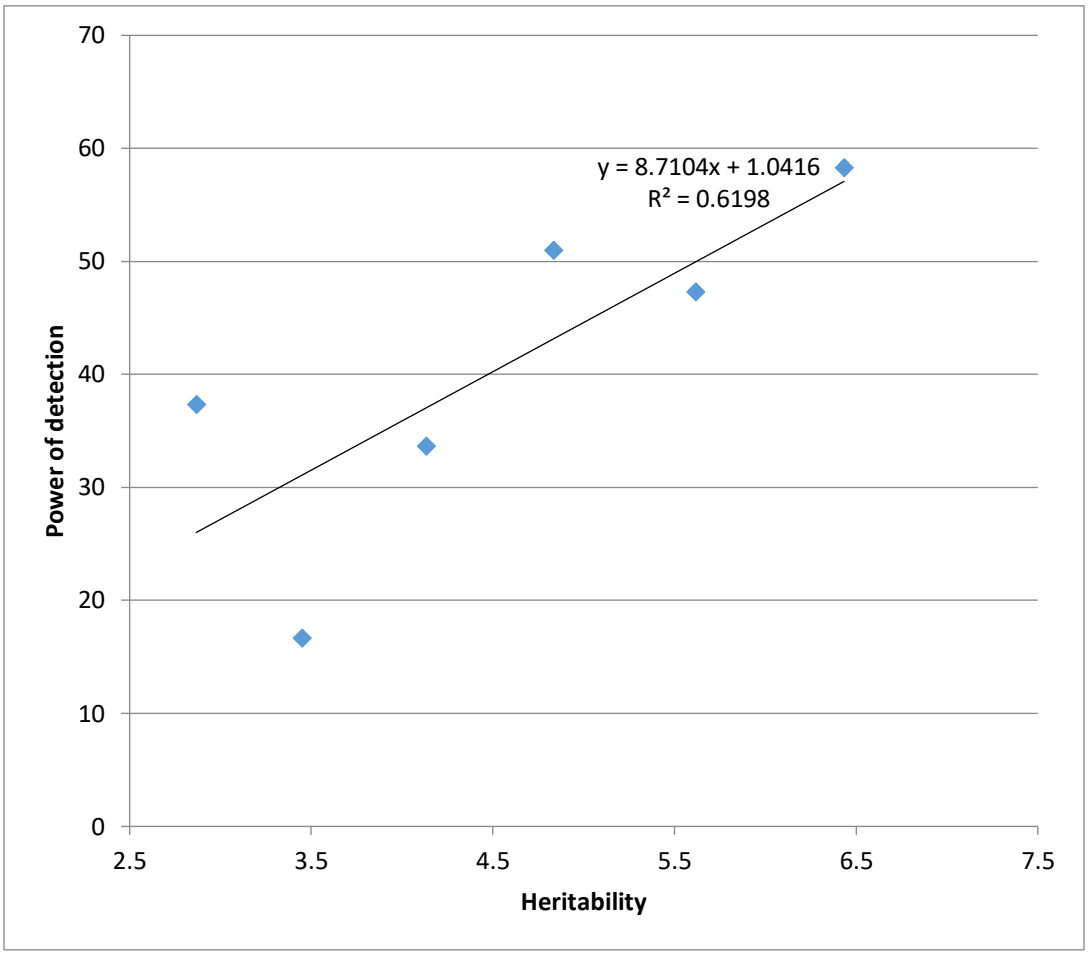

Figure 3: Correlation between the average power of detection and QTL heritabilities in each environment **assuming genotype by environment interaction**.

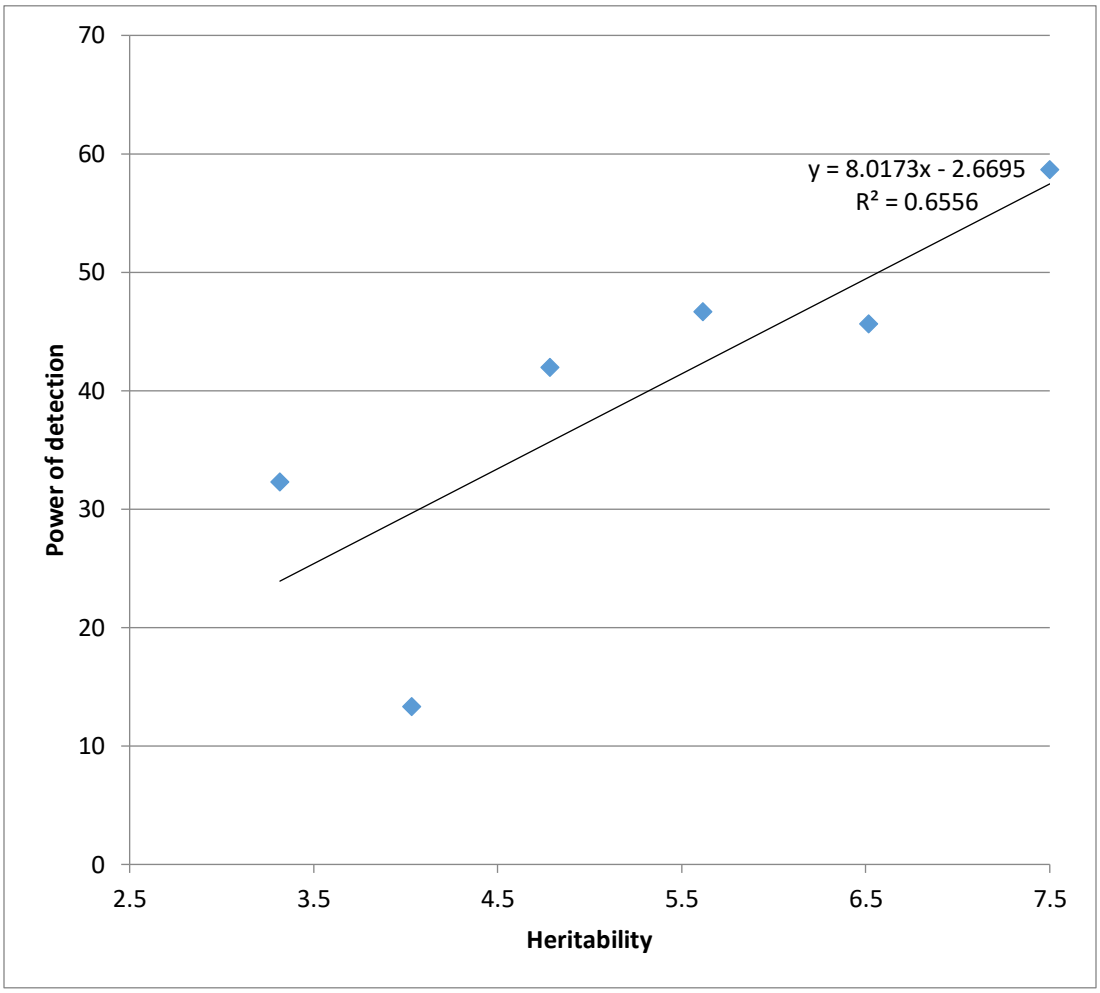

Figure 4: Correlation between the average power of detection and QTL heritabilities in each environment **assuming no genotype by environment interaction**.

4. DISCUSSION

Quantitative trait loci mapping studies have been done for various crops and across environments but information on assessing the efficiency of QTL mapping under genotype by environment interaction remains scarce. Comparing the results obtained from this study and the results from some field investigations, we can confidently say that the data set used in this study depicts a real field data (Li *et al.*, 2018; Xie *et al.*, 2020; Smith *et al.,* 2022). Interestingly, in the studies on rape seed by Xie *et al.,* (2020), the authors recorded a positive correlation between some yield related traits and the environment which also agrees with the results from this study. The trait heritability for their study was within the range of 31 to 89% which is also within the range of the values in this study. This result is

also similar to the findings of Li *et al.* (2018) in their study on maize using 204 recombinant inbred lines as they observed a positive correlation between some traits and the environment with heritability within the range of 30 to 87%. In another study by Smith *et al.* (2022) on mapping aflatoxin content of maize, they recorded a correlation coefficient of -0.5 between environmental index and trait heritability. The values of heritability in their study was within the range of 43 to 63%. This result is different from the one in this investigation probably because the number of environments was lower than that used for this study, they used four environments while we used six. Futhermore, there was a wide gap in range of heritability (23 to 85%) in this study compared to theirs. Also, the heritability across their environments was as high as 90% which is the same with the trait heritability across environments in the first scenario of this study. They also reported that QTLs with small effect that were significant in one population may have been present but undetected in the other population. QTL with small effects are not useful as breeding targets. This is why it is important to assess efficiency of QTL mapping.

QTL mapping has been used in plant breeding to identify regions on the chromosome responsible for the expression of quantitative traits. Interestingly, this method is very efficient especially for traits such as yield which is controlled by many genes. The main reason for QTL mapping for almost three decades is to dissect the genetic architecture of quantitative traits (Viana *et al*., 2017). However, one setback of QTL mapping is low precision. The use of simulations has provided an effective way of assesing efficiency of QTL mapping by providing high power of detection, low false positive rates and bias in precision mapping depending on population size, trait heritability and QTL effect (Viana *et al*., 2017). This has been demonstrated using both low (10 to 20 cM) and high marker density of one marker every cM. This is very similar to the marker density used for this study which was one marker every 2 cM. For this study, assuming LOD of 3, FPR varied across environments for the single site analysis and was within the range of 4.4 to 4.5% for the multi environment analysis. This result does not agree with Li *et al.* (2012); Viana *et al*. (2017) as they recorded values of 0.5 to 1.2% and 1.5 to 2.5% for different mapping methods. The reason could be because of the differences in heritability values, method of estimating QTL and sample size. The FPR result obtained from this study was slightly higher than the former but agrees with the findings of Yang *et*

al. (2009) as they recorded FPR values of 4 to 6% with population size of 300 and 150 respectively.

The correlation value of -0.03 observed between the environmental index and trait heritability across environments in the first scenario clearly shows that there was no association between them but this was not the case in the second scenario which was positive with value of 0.36. Regarding the QTL power of detection, the results are very impressive. In both scenarios under high trait heritability value of about 80%, power of detection was as high as 100% in environment four. In some investigations carried out on efficiency of QTL mapping, power of detection ranged from 60 to 100%. This result agrees with findings from other researchers (Nobari *et al.,* 2012; Li *et al.,* 2012; Viana *et al.,* 2017; Andrade *et al.,* 2020). In QTL mapping, power indicates the probability that a real QTL is detected and is therefore the most important indicator of the method's efficiency. Correlation between the average power of detection and QTL heritabilities in both scenarios was positive which shows that power was directly proportional to QTL heritability. An interesting result also observed in this study is that the power of detection was as high as 100% for some low QTLs. This is due to genotypic, phenotypic and genotype by environment interaction variance effects.

Simulations have proven to be effective because power and FPR can not be assumed theoretically. There are also challenges one may encounter when using actual mapping populations for example, the true QTL positions and effects are usually unknown. Another problem with using real data to compare methods is that the parameter ranges are usually limited by experimental conditions, genetic architectures and species of crops (Li *et al*., 2012). It provides a unique way to calculate the power and FPR of different QTL mapping methods to evaluate and compare their efficiencies. Prior to any costly experiment, a researcher would like to be certain that the design ensures sufficiently high power based on the objectives of the study. Statistically, power depends not only on sample size and the actual values of the unknown distribution parameters being estimated, but also on the assumed level of significant threshold. In QTL mapping, power tells the likelihood that a real QTL was detected and is therefore the most important indicator of a method's efficiency. Li *et al*., (2012) reported that a FPR value of 5% means

that among all the features identified as significant, 5% on average are truly false. However, precisely estimating QTL position and effect is also an important factor to be considered.

On bias (mapping precision) from the multi environment analysis, the value from this study was 2.1 and 2.2 cM for the first and second scenarios respectively. This result agrees with previous studies as bias generally ranged from 0 to 4 cM. Liu *et al.* (2017) also observed average power of detection of 97.3% for a sample size of 300 but bias was in the range of 0.4 to 0.6 cM which was lower than that obtained from this study. Regarding mapping precision, QTL mapping generally provides precise localization of candidate genes with bias in the QTL position that is lower than the high density intervals of 4 to 6 cM. The FPR for the single site analysis ranged from 2.0 to 8.5% and 3.2 to 11.1% for the first and second scenarios respectively. This result agrees with the findings of Andrade *et al.,* (2020) as they recorded FPR of 15 and 9% for low and high marker density. This investigation was done under high marker density of 2 cM. The non significant difference in genotypic variance across environments was because the genotypic file for both scenarios were the same while the genotype x environment interaction effect recorded in the first scenario only was because the second did not have any interaction effect.

An increasing number of QTLs and genes that affect yield, quality among others have been identified using molecular markers (Ren *et al.,* 2021). QTL mapping efficiency is affected by the QTL heritabilty, mapping population, size, number of markers and marker density (Wang *et al.,* 2012; Andrade *et al.,* 2020). It should be noted that increase in the prior expected number of QTLs increased the power of QTL detection and mapping precision. Breeders should note that this is not an important factor compared to QTL heritability and population size. Our simulation based study used 50 replications and this gives very confident inferences on QTL power of detection, precision mapping and control of false positives.

Genotypic main effects and genotype x environment interactions contributed a significant amount to the variance of the traits. Genotype x environment interactions are known to be important for many agronomical traits (Hudson *et al*., 2022) while grain yield

have large genotype x environment variance components relative to main genotype effects (Rogers *et al.,* 2021). Grain yield is a highly integrated trait dependent on the interaction of many other traits with the environment. If those traits have a complex basis within different environments, then it would not be surprising to observe large genotype x environment variance at the level of genotype (Hudson *et al*., 2022). From this investigation, QTLs were declared even under low QTL heritability. This was due to genotypic, phenotypic and genotype by environment interaction variance effects.

5. CONCLUSIONS

Based on the results obtained from this investigation, environment four had high heritability in the two scenarios. It was observed that the higher the QTL heritability, the higher the power of detection as recorded in environment four.

For QTL power of detection we can draw the following inferences:

1. In the first scenario, genotype by environment effect decreased the correlation between heritability and power.
2. The correlation of averages across environments between QTL heritability and power was 0.62 and 0.66 for first and second scenario respectively. This clearly shows that on the average, we declared more times the QTL of higher heritability than the ones with lower heritability.
3. Mapping QTLs over environments allows detection of low heritability in the range of 2 to 7% for most QTLs.
4. The joint QTL mapping analysis provided for most QTLs a higher power of detection compared with the average power assuming genotype by environment and no genotype by environment interactions.

For false positive rate (FPR), we can infer that:

1. The FPR was greater than 5% in the environment of higher trait heritability. In environment four assuming genotype by environment interaction, correlation was 0.64 and in no genotype by environment interaction it was 0.69.

2. In the joint QTL mapping analysis, there is an effective control of false positive rate in the two scenarios.

For mapping precision. We can conclude that:

1. Mapping precision in a single environment or joint analysis is precise once the difference between the estimated and true QTL position is in the range of 1 to 3 cM approximately irrespective of genotype by environment interaction.

Finally, based on the methods of assessing efficiency of QTL under genotype by environment interaction and comparing the two scenarios, it is important to highlight that the best environment in this study was environment four, we also declared QTLs even under very low heritability across environments which shows that QTL mapping using simulations is an effective process.

6. REFERENCES

Andrade, A.C.B, Viana, J.M.S, Pereira, H.D, Fonseca e Silva, F. (2020). Efficiency of Bayesian quantitative trait loci mapping with full-sib progeny. *Agronomy Journal*.112:2759–2767.

Broman, K.W., Wu, H., Sen, S., Churchill, G.A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19: 889–890.

Dev Paudel, Smit Dhakal, Saroj Parajuli, Laxman Adhikari, Ze Peng, You Qian, Dipendra Shahi, Muhsin Avci, Shiva O. Makaju, Baskaran Kannan (2020). Use of quantitative trait loci to develop stress tolerance in plants. In: Plant Life Under Changing Environment. pp 917 -965.

El-Soda, M., Malosetti, M., Zwaan, B.J., Koornneef, M., Aarts, M.G.M. (2014). Genotype × environment interaction QTL mapping in plants: lessons from Arabidopsis. *Trends in Plant Science*. 19: 390–398.

Getahun, B. B., Tiruneh, M. A., Aliche, E., Malossetti, M., Visser, R. and Linden, G. (2022). Genotype-by-Environment Interaction for Quantitative Trait Loci Affecting Nitrogen Use Efficiency and Associated Traits in Potato. *Potato Research* 65:777–807.

Hu, X., Wang, G., Du, X., Zhang, H., Xu, Z., Wang J., Chen, G., Wang, B., Li, X., Chen, X., Fu, J., Zheng, J., Wang, J., Gu, R. and Wang, G. (2021). QTL analysis across multiple environments reveals promising chromosome regions associated with yield-related traits in maize under drought conditions. *The Crop Journal* 9 :759 -766.

Hudson, A. I., Odell, S. G., Dubreuil, P., Tixier, M. H., Praud, S., Runcie, D. E. and Ross-Ibarra J. (2022). Analysis of genotype-by-environment interactions in a maize mapping population. *G3,* 12 (3).

Li, H., Ye, G., and Wang, J. (2007). A modified algorithm for the improvement of composite interval mapping. *Genetics* 175: 361–374.

Li, H., Hearne, S., Banziger, M., Li, Z., and Wang, J. (2010). Statistical properties of QTL linkage mapping in biparental genetic populations. *Heredity* 105: 257–267.

Li, H., Zhang, L., and Wang, J. (2012). Estimation of statistical power and false discovery rate of QTL mapping methods through computer simulation. *Chin. Sci. Bull* 57:2701–2710.

Li, S., Wang, J., Zhang, L. (2015). Inclusive composite interval mapping of QTL by environment interactions in biparental populations. *PLoS ONE* 10(7): e0132414.

Li, P., Zhang, Y., Yin, S., Zhu, P., Pan, T., Xu, Y., Wang, J., Hao, D., Fang, H., Xu, C. and Yang, Z. (2018). QTL-by-environment interaction in the response of maize root and shoot traits to different water regimes. *Frontiers Plant Science.* 9:229

Liu, F. X., Tong, C. F., Tao, S. T.,Wu, J. Y., Chen, Y. H., Yao, D., ... Shi, J. (2017). MVQTLCIM: Composite interval mapping of multivariate traits in a hybrid F-1 population of outbred species. *Bmc Bioinformatics*, *18*(1), 515

Maniruzzaman, S., Rahman, M.A., Hasan, M., Rasul, M.G., Molla, A.H., Khatun, H., Akter, S. (2022). Genetic Mapping to Detect Stringent QTLs Using 1k-RiCA SNP Genotyping Platform from the New Landrace Associated with Salt Tolerance at the Seedling Stage in Rice. *Plants* 11, 1409

Nabateregga, M., Mukankusi C., Raatz, B., Edema, R., Nkalubo, S. and Alladassi, B. (2019). Quantitative trait loci (QTL) mapping for intermittent drought tolerance in BRB 191 × SEQ 1027 Andean Intra-gene cross recombinant inbred line population of common bean (*Phaseolus vulgaris* L.). *African Journal of Biotechnology.* 18(21): 452-461.

Nobari, K., Nassiry, M.R., Aslaminejad, A.A., Tahmoorespur, M., Esmailizadeh, A. K (2012). Effects of QTL parameters and marker density on efficiency of Haley-Knott regression interval mapping of QTL with complex traits and use of artificial neural network for prediction of the efficiency of HK method in livestock. *Journal of Applied Animal Research* 40:247–255.

Ren, T., Fan, T., Chen, S., Ou, X., Chen, Y., Jiang, Q., Diao, Y., Sun, Z., Peng, W., Ren, Z., Tan, F. and Li, Z. (2021). QTL mapping and validation for kernel area and circumference in common wheat via high-density SNP-based genotyping. *Frontiers Plant Science* 12:713890

Rogers, A.R., Dunne, J.C., Romay, C., Bohn, M., Buckler, E.S., Ciampitti, I.A., Edwards, J., Ertl, D., Flint-Garcia, S., Gore M.A...(2021). The importance of dominance and genotype –by –environment interaction on yield variation in a large – scale public cooperative maize experiment. *G3* (Bethesda) 11:jkaa050

Smith, J.S., Williams, W.P., Warburton, M., Windham, G., Xu, W., Bhattramakki, D. (2022). Mapping quantitative trait loci for aflatoxin accumulation resistance in two populations containing resistant maize inbred Mp717.*Crop Science* 62: 780 -791.

Souza, V. F., Pereira, G. S., Pastina, M. M., Parrella, R. A., Simeone, M. L., Barros, B. A., Noda, R. W., Costa e Silva, L., Magalhaes, J., Schaffert, R. E., Garcia, A. A. and Damasceno, C. M. (2021). QTL mapping for bioenergy traits in sweet sorghum recombinant inbred lines *G3,* 11 (11).

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 (2):193-205.

Viana, J. M. S., Silva, F. F., Mundim, G. B., Azevedo, C. F. and Jan, H. U. (2017). Efficiency of low heritability QTL mapping under high SNP density. *Euphytica,* 213, 13.

Viana, J. M. S., Pereira, H. D., Mundim, G. B., Piepho, H. P. and Silva, F. F. E. (2018). Efficiency of genomic prediction of non-assessed single crosses. *Heredity*, *120*, 283–295.

Wang, X., Gilbert, H., Moreno, C., Filangi, O., Elsen, J-M., Le Roy, P. (2012). Statistical properties of interval mapping methods on quantitative trait loci location: impact on QTL/eQTL analyses. *BMC Genetics* 13:29.

Wang, L., Yang, X., Cuib, S., Mub, G., Suna, X., Liub, L., Lia, Z. (2019). QTL mapping and QTL × environment interaction analysis of multi-seed pod in cultivated peanut (*Arachis hypogaea* L.). *Crop Journal* 7: 249-260.

Wurschum, T., Liu, W., Gowda, M., Maurer, H.P., Fischer, S., Schechert, A., Reif, J.C. (2012). Comparison of biometrical models for joint linkage association mapping. *Heredity* 108: 332–340.

Xie, Y., Xu, J., Tian, G., Xie, L., Xu, B., Liu, K., Zhang, X. (2020). Unraveling yield-related traits with QTL analysis and dissection of QTL x environment interaction using a high – density bin map in rape seed (*Brassica napus* L.). *Euphytica* 216: 171.

Xu, Y., Li, P., Yang, Z., Xu, C. (2017). Genetic mapping of quantitative trait loci in crops. *The Crop Journal* 5 (2):175-184.

Yang, R., Wang, X., Li, J., Deng, H. (2009). Bayesian robust analysis for genetic architecture of quantitative traits. *Bioinformatics* 25:1033–1039.

Zeng, Z. B (1994). Precision mapping of quantitative trait loci. *Genetics* 136:1457-1468.