BIOSTATISTICAL TOOLS FOR PLANT BREEDING IN THE GENOMICS ERA

L Gutierrez[1], A Borges[1], G Quero[2], A Gonzalez-Reymundez[1], I Berro[1], B Lado[1,3], A Castro[4]

SUMMARY

Since the advent of agriculture, plant breeding has successfully improved plants for human benefit. Modern plant breeding activities consist in evaluating the genetic merit of lines discerning genetic from environment and noise components. To do so, modern plant breeding relies on the genetics foundations derived from Mendel's work and statistical tools (or biometry) generated afterwards. Plant breeding activities could be grouped in three categories: traditional, marker assisted (MAS), and genomic selection (GS). Traditional plant breeding uses either *per se* phenotypic information, or information from relatives to evaluate the genetic value. MAS on the other hand, involves the identification of markers linked to genes or quantitative traits loci (QTL) of relevant traits, and then selecting individuals based on their marker scores. Finally, GS involves the prediction of the genetic merit of individuals based on their marker scores and a statistical model. All of the three strategies require the evaluation of large number of individuals creating massive amounts of data that needs proper analyses. Our objective was to present some biostatistical strategies that are successfully being used in plant breeding programs. First, we used novel simulation approaches to compare the use of experimental design and spatial corrections in the context of genotypic evaluations. Second, we proposed some strategies for modeling and interpreting QTL by environment interaction for QTL mapping. Third, we compared models for Genome-wide Association Mapping (GWAS) using different strategies for accounting for population structure, and we evaluated the performance of models for mapping non-normal traits. Finally, we compared and evaluated strategies for implementing GS in national breeding programs. Statistics has therefore become a key component in plant breeding activities.

Keywords: QTL Mapping, GWAS, Genomic Selection, Genotype by Environment Interaction

Abbreviations: QTL, Quantitative Trait Loci; GWAS, Genome-wide Association Mapping; MAS, Marker Assisted Selection; GS, Genomic Selection; GEI, Genotype by Environment Interaction; QEI, QTL by Environment Interaction; CRD, completely randomized design; RCBD, randomized complete block design; IB-á, incomplete block design

[1]Departamento de Biometría, Estadística y Computación, Facultad de Agronomía, Universidad de la República, Garzón 780, CP 12900, Montevideo, Uruguay;

[2]Departamento de Biología Vegetal, Facultad de Agronomía, Universidad de la República, Garzón 780, CP 12900, Montevideo, Uruguay;

[3]Instituto Nacional de Investigación Agropecuaria (INIA), Estación Experimental La Estanzuela, Ruta 50 km 11, 500, CP 70000, Colonia, Uruguay;

[4]Departamento de Producción Vegetal, Facultad de Agronomía, Universidad de la República, R3 Km 373, Paysandú, Uruguay.

E-mail: luciag@fagro.edu.uy

INTRODUCTION

Since the advent of agriculture, plant breeding has successfully improved plants for human benefit (Allard 1960, Fehr 1984, Hallauer and Miranda Filho 1988, Duvick *et al.,* 2004). Modern plant breeding activities consist in evaluating the genetic merit of lines discerning genetic from environment and noise components. To do so, modern plant breeding relies on the genetics foundations derived from Mendel's work, and statistical tools (or biometry) generated afterwards (Sprague and Dudley, 1988, Lamkey and Lee, 2006). Plant breeding activities could be grouped in three categories: traditional, marker assisted (MAS; Tanksley, 1983), and genomic selection (GS; Meuwissen *et al.,* 2001). Traditional plant breeding uses either *per se* phenotypic information, or information from relatives to evaluate the genetic value (Fehr, 1987, Bernardo, 2010). MAS on the other hand, involves the identification of markers linked to genes or quantitative traits loci (QTL) of relevant traits, and then selecting individuals based on their marker scores (Tanksley 1993, Hospital and Charcosset 1997). Finally, GS, involves the prediction of the genetic merit of individuals based on their marker scores and a statistical model (Meuwissen *et al.,* 200, de los Campos, 2012). All of the three strategies require the evaluation of large number of individuals creating massive amounts of data that needs proper analyses.

A MAS program requires the identification of genes or genomic regions associated to the traits of interest. There are several strategies to identify quantitative trait loci (QTL) of relevant traits including bi-parental population or traditional QTL mapping (Hayes, 1993) and Genome-wide Association Mapping (GWAS; Jannink *et al.,* 2001). Traditional QTL Mapping requires first the construction of balanced populations with known recombination history. Later, a statistical association between a molecular marker and the trait of interest is sought through linkage disequilibrium using either linear regression models (Haley and Knott, 1992) or mixture distributions (Lander and Botstein, 1989). Since all the recombination occurred within the limits of the experiment, the linkage disequilibrium is expected to be caused by physical linkage of the molecular marker and the QTL, and therefore the location of the QTL can be inferred. The GWAS is also based on a statistical association between the molecular marker and the trait of interest. However, since diverse populations without a known recombination history are used, the cause of linkage disequilibrium could be physical linkage and other causes such as selection, genetic drift, mutation, admixture and population structure among other evolutionary forces (Jannink *et al.,* 2001). Therefore, controlling for population structure is crucial in GWAS. Advantages of GWAS, as compared to bi-parental QTL mapping include: assessment of genetically diverse germplasm stocks, higher resolution mapping, effective use of historical data, and immediate applicability to cultivar development because the genetic background in which QTL are estimated is directly relevant for plant breeding (Kraakman *et al.,* 2004, Dekkers and Hospital, 2002, Yu and Buckler, 2006). This strategy has successfully been used in plants (Kraakman *et al.,* 2004, 2006, Hayes and Szücs, 2006, Stracke *et al.,* 2009, Waugh *et al,.* 2009, Roy *et al.,* 2010, Bradbury *et al.,* 2011, von Zitzewitz *et al.,* 2011, Gutierrez *et al.,* 2011, Locatelli *et al.,* 2013).

One of the main limitations of QTL studies is that not all of the QTL can successfully be identified mainly due to population sizes and the number and size of the QTL effects (*Beavis effect,* Beavis, 1998). Furthermore, the QTL that are identified have small effect and explain a small portion of the total variation (*missing heritability*, Manolio *et al.,* 2009). These make it challenging to actually use the QTL results in breeding programs. Some alternatives include GS that use all the markers to predict the performance of the individuals skipping the significance test for any marker (Meuwissen *et al.,* 2001). The principle consists in developing a prediction model based on a large population thoroughly studied for both molecular and phenotypic information (i.e. the training population) and using the model to predict phenotypic performance in instances where phenotyping is

47

not suitable (i.e. early generation testing, off-season nurseries and others; Heffner *et al.,* 2009). Several models have been developed for GS **(**Meuwissen *et al.,* 2001, Gianola *et al.,* 2006, Bernardo and Yu, 2007, Lorenzana and Bernardo, 2009, de los Campos *et al.,* 2012): based on mixed models (G-BLUP), bayesian models (Bayesian LASSO, Bayesian-RR, Bayes A, B, C and others) and semi-parametric models (RKHS, PNN, etc.).

Quantitative traits are affected by the environment making phenotyping crucial in any plant breeding activity. This creates two challenges. First, field trials and experimental designs for large number of genotypes should be carefully chosen to reduce spatial heterogeneity and experimental error and to increase heritability (Cullis *et al.,.* 1998). Second, Genotype by Environment Interaction (GEI) is widespread in plants, and affects especially quantitative traits that are of main importance for plant breeding (Mathews *et al.,* 2008). Mixed models have been used for modeling GEI and QTL by Environment Interaction (QEI; Piepho, 2000, Verbyla *et al.,* 2003, Malosetti *et al.,* 2004, van Eeuwijk *et al.,* 2005, Boer *et al.,* 2007, Mathews *et al.,* 2008).

Our objective was to present some biostatistical strategies that are successfully being used in plant breeding programs. First, we used novel simulation approaches to compare the use of experimental design and spatial corrections in the context of genotypic evaluations of large number of genotypes. Second, we proposed some strategies for modeling and interpreting QEI for both traditional QTL mapping and GWAS. Third, we compared models for GWAS using different strategies for accounting for population structure, and we evaluated the performance of models for mapping non-normal variables. Finally, we compared and evaluated strategies for implementing GS in national breeding programs.

48

EXPERIMENTAL DESIGN AND SPATIAL CORRECTION

Materials and Methods

Yield data from 15, 50 or 200 genotypes was simulated using real spatial variability and genotypic effects randomized with three experimental designs: completely randomized design (CRD), randomized complete block design (RCBD), and incomplete block design (IB-á). Afterwards, each simulation was analyzed using models with different levels of spatial correction: no spatial correction model, spatially correlated error model with one-dimensional auto-regressive process [AR(1)] and two-dimensional auto-regressive process [AR(1)×AR(1)]. Models were compared by goodness of fit, accuracy, recovery of superior genotypes and percentage of rejection of the null hypothesis in the ANOVA.

Results and Discussion

Spatial variation in environmental and soil factors commonly occurs in field conditions (Grondona *et al.,* 1996, Legendre, 1993). Therefore using experimental designs that incorporate local control is especially beneficial in experiments with large number of treatments (Legendre *et al.,* 2004, Gonçalves *et al.,* 2010, Masood *et al.,* 2008). We found that experimental designs with larger local control performed better (see Borges *et al.,* 2014). For moderate to small experiment size, the IB-á obtained the best results in fit, precision and recovery of superior genotypes. In this situation, the CRD showed the worst performance for almost all statistics, with very low efficiency, reaching only 1.49%, 19.1% and 10% (15, 50 and 200 genotypes respectively) of rejection of the null hypothesis.

Design deficiencies to control spatial variability could be somewhat compensated by using a model that includes spatial variation (Casler and Undersander, 2000, Qiao *et al.,* 2000). In most of the situations models that include spatial correlation are more efficient (Brownie *et al.,* 1993, Kravchenko *et al.,* 2006, Mallarino *et al.,*

2000). We found that modeling spatial heterogeneity in our study improved design performances. The models that included spatial correlation were generally better than those that did not in terms of model fitness. This improvement was clearer in the cases of the CRD and RCBD than in the IB-α. However, modeling spatial heterogeneity was not enough in the CRD design.

Some reports argue that spatial methods of analysis provide more accurate and precise estimates of genotypic effects than either complete or incomplete blocks analysis (Cullis *et al.,* 1998, Cullis and Gleeson, 1991). However, we observed that the improvements achieved with the design were greater than those obtained with the inclusion of spatial correction in the analysis models. Therefore, spatial modelling aid but does not substitute experimental design.

QTL BY ENVIRONMENT INTERACTION FOR ABIOTIC STRESS

Materials and Methods

An inter-specific population of recombinant inbred lines (RIL) was used to identify QTL associated to abiotic stress.

Briefly, RIL from the cross between *Lotus japonicus* Gifu and *L. burtii* were genotyped with molecular markers that cover all the linkage groups and phenotyped in hydroponic conditions at three stress-conditions: ionic-stress, osmotic-stress, and control. A multi-QEI was conducted using mixed models on shoot, root, and total relative growth to identify stress-specific QTL. For more details, see Quero *et al.* (2014).

Results and Discussion

Plants phenotypic expression is the result of an interaction between the genome and the environmental conditions (Quero *et al.* 2014). However, efforts in genomic analysis have not been followed by proper understanding of the phenotype, creating what is called as phenotype gap (Miflin, 2000, Verslues *et al.,* 2006). Quantitative traits are affected by GEI (Mathews *et al.,* 2008) and therefore modeling GEI and QEI provides a more natural interpretation of GEI (Piepho, 2000, Verbyla *et al,.* 2003, Malosetti *et al.,* 2004, van Eeuwijk *et al.,* 2005, Boer *et al.,* 2007, Mathews *et al.,* 2008). By using mixed models for QEI analysis, we demonstrated that RIL from *Lotus* have different responses to ionic and osmotic stresses, and that we could map genomic
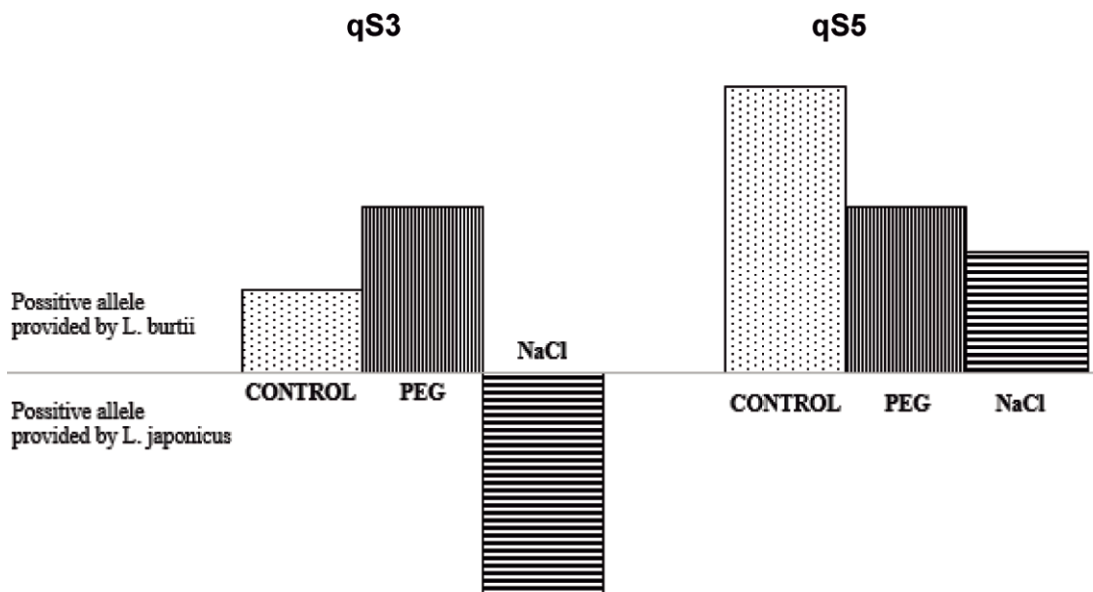
49

Figure 1. Magnitude of QTL effect in stress environments for the relative growth rate of the shoot of a *Lotus japonicus_L. burttii* RIL population where the length of the bar on the *y*-axis indicates the magnitude of the effect. Modified from Quero *et al.*, 2014.

regions associated to each response (Figure 1; Quero *et al.,* 2014). Furthermore, the favorable allele for osmotic stress on Chromosome-3 was provided by *L. burtii*, while the favorable allele for ionic stress was provided by *L. japonicus* (Figure 1; Quero *et al.,* 2014). We showed how QEI strategies could be implemented to get a better understanding of the GEI and to better understand the phenotype.

GWAS MODELS FOR POPULATION STRUCTURE

Materials and Methods

We studied the association between five malting quality traits and 3072 single nucleotide polymorphisms (SNPs) from the barley oligonucleotide pool assay (BOPA) 1 and 2, assayed in advanced inbred lines from the Oregon State University breeding program from three germplasm arrays (CAP I, CAP II, and CAP III). We compared 16 models to account for population structure that included all the combinations of population control (i.e. principal component, non-metric

multidimensional scaling, population structure, and no-control) with different kinship estimation methods (i.e. EMMA, TASSEL, SPAGeDI, and none). For more details, see Gutierrez *et al.,* 2011.

Results and Discussion

Population structure and genetic relatedness are one of the main causes of spurious association in GWAS studies (Jannink *et al.,* 2001, Yu *et al.,* 2006, Cappa *et al.,* 2013). However, having a good control for population structure is not straightforward. Several strategies have been proposed for controlling population structure including using Bayesian inferred population structure (Pritchard *et al.,* 2000); kinship relationship matrix (Parisseaux and Bernardo, 2004); using both population structure and kinship (Yu *et al.,* 2006); using other multivariate approaches to account for population structure like principal component analysis (Patterson *et al.,* 2006, Price *et al.,* 2006) or non-metric multi-dimensional scaling (Zhu and Yu, 2009); or using genome-wide markers (Bernardo, 2013). We found that the best
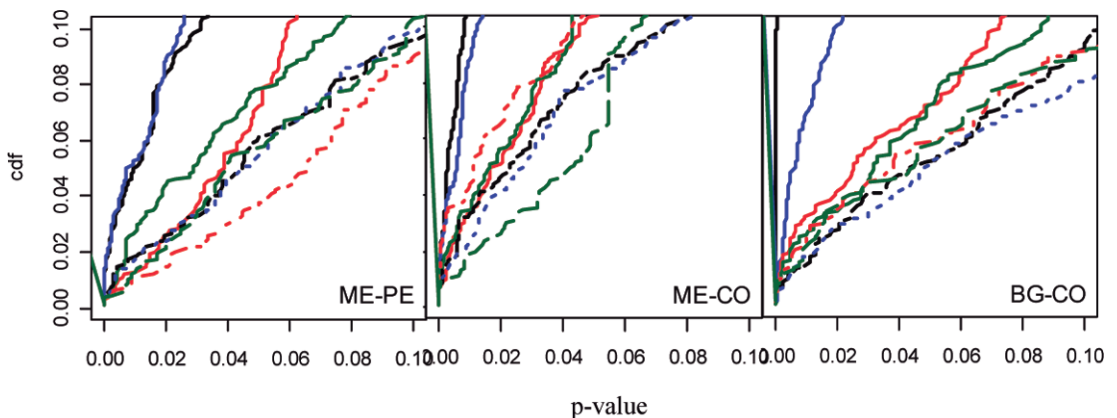

Figure 2. Cumulative distribution function (cdf) of *p*-values in genome-wide scans for a barley array for malt extract in two environments (ME-PE and ME-CO) and wort β. glucan in one environment (BG-CO). The different curves correspond to different models compared: Naive, marker regression without correction for population structure; **Q**, posterior probabilities matrix inferred from software STRUCTURE (Pritchard *et al*., 2000); **P**, fixed-effects matrix from principal component analysis; **M**, fixed-effects matrix from nonmetric multidimensional scaling; **K**, mixed models using kinship matrix as implemented by efficient mixed model association (EMMA [Kang *et al*., 2008]); **QK**, mixed models with **Q** matrix as fixed effects and kinship matrix as random effects; **PK.E**, mixed models with **P** matrix as fixed effects and **K** matrix as random effects; and MK, mixed models with **M** matrix as fixed effects and **K** matrix as random effects. Modified from Gutierrez *et al*., 2011.

model was population, trait, and environment dependent (Figure 2; Gutierrez *et al.,* 2011). However, all of the mixed models recover the QTL of large effect (Gutierrez *et al.,* 2011, Gutierrez *et al.,* 2012).

GWAS MODELS FOR NON-GAUSSIAN DATA

Materials and Methods

Five methods for GWAS for ordinal variables including generalized linear models and transformations were compared in terms of their relative efficacy in QTL detection and estimation. Simulations were conducted for a wide range of population sizes, number of QTL, and heritabilities. We used both real genotypic data from a barley population, and *de novo* simulated data. Phenotypic values for ordinal variables were simulated according to different genetic models and QTL were recovered using different GWAS models. Power, false discovery rate, and bias in QTL effect estimation were compared.

Results and Discussion

Because GWAS models are variations of the linear mixed model, they assume normality of residuals (Henderson, 1984). When this assumption does not hold, inference on QTL position and effects could be negatively affected, causing a bias in the QTL estimated effect, out-of range predictions, or inaccurate hypothesis tests results (Casella and Berger, 1990, Wu *et al.,* 2010). However, some of the relevant traits being mapped are not normally distributed (i.e. disease resistance, water deficit resistance, and grain quality, which are ordinal variables). Some strategies have been implemented for mapping non-normal traits including the use of normal error models in balanced populations (Visscher *et al.,* 1996, Rebai *et al.,* 1997) and generalized linear models in balanced (Spyrides-Cunha *et al.,* 2000, Diao and Lin, 2006) and GWAS populations (Iwata *et al.,* 2009). We compared the use of five different methods for dealing with non-normal error data including general linear models (i.e. no-transformation, squared-root transformation, and logarithmic transformation) and generalized linear models (i.e. probit and logit regression). Under a wide range of population sizes, number of QTL, and heritabilities, no differences in power and false positives rate were detected across methods while similar bias were obtained for all methods (Figure 3). This suggests that the choice of the
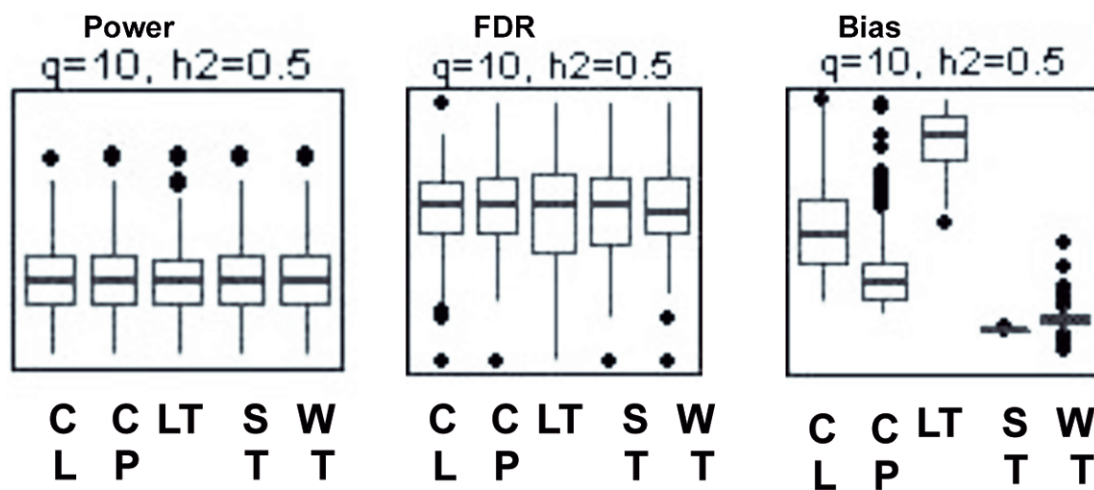
Figure 3. Power and FDR of QTL detection and bias of QTL effect for simulated genotypic set evaluated with five different methods: WT (simple linear regression without transformation), ST (simple linear regression on square root transformed data), LT (simple linear regression on logarithm transformed data), CP (cumulative simple regression with probit link) and CL (cumulative simple regression with logit link).

method for dealing with ordinal variables does not have a major impact on GWAS results.

GS IMPLEMENTATION IN WHEAT

Materials and Methods

A total of 1044 advanced inbred lines from the National Wheat Breeding Program (NWBP) were used to train GS models. Genotyping by sequencing (GBS; Elshire *et al.,* 2011 modified by Poland and Rife, 2012 for wheat) was used with the Tassel pipeline (Glaubitz *et al.,* 2014) as in Lado *et al.* (2013) to obtain 81,999 filtered SNP. Multi-environment trials were used for phenotypic information; the lines were evaluated in multiple years and locations in Uruguay. Additive-Main Interaction Models (AMMI; Gauch, 1992), GGE biplots (Yan, 2000), and correlations across environments were used to establish mega-environments. Mixed models were used to estimate Breeding Values of the lines and model performance was evaluated with the prediction accuracy. We compared prediction accuracy within and among mega-environments, as well as the prediction accuracy modeling the GEI (Lado *et al.,* 2014). Additionally, we compared strategies to establish the training population.

Results and Discussion

Several strategies have been proposed to handle GEI in plant breeding context, to ignore, to avoid, and to exploit it (Bernardo, 2010). Which strategy to follow will depend entirely on the breeding objectives and the targeted environments for the breeding program. However, understanding the nature of the GEI is crucial in order to make informed decisions. Multiplicative models were initially used to study GEI, and AMMI models (Gauch, 1992) as well as GGE models (Yan, 2000) have been widely used. Mixed models, on the other hand, provide a natural way to model the correlation across environments due to GEI (Malosetti *et al.,* 2013, Cooper *et al.,* 2014). Since genomic prediction tools come from systems where GEI is not an important issue, little attention has been paid into incorporating GEI into prediction models.

However, Burgueño *et al.* (2012) used Mega-environments to make within-mega-environment predictions, while Heslot *et al.* (2014) used environmental co-variables to improve predictions. We used a large population of advanced inbred lines from the INIA-Uruguay Wheat National Breeding Program combined with extensive genotyping and phenotyping. A set of meteorological data from environments within the range of targeted environments for the Wheat National Breeding Program was used. We found that modeling GEI data produce higher prediction accuracy than using average data (Lado *et al.,* 2014). Additionally, modeling within mega-environments was beneficial as long as population sizes were maintained (Lado *et al.,* 2014). Constructing the training population with a larger population size, even at the expense of genetic relatedness, was more beneficial than using fewer and more related individuals.

CONCLUSIONS

Modern plant breeding requires the use of large data sets combining the information from hundreds or thousands of individuals evaluated in multiple environments and screened for thousands of molecular markers. Managing these kind of data could therefore be challenging. But more importantly, exploiting this information to advance breeding cycles and produce better cultivars requires the intense use of biostatistical tools and asking the right questions. We showed some examples where biostatistical methods aid in the analysis and interpretation of the results to advance genetic gain. Statistics has therefore become a key component in plant breeding activities.

ACKNOWLEDGEMENTS

REFERENCES

Allard RW. 1960. Principles of plant breeding. John Wiley and Sons, Inc. New York.

Beavis WD. 1998. QTL analyses: power, precision, and accuracy, pp. 145–162 in Molecular Dissection of Complex Traits, edited by Paterson A. H.. CRC Press, New York.

Bernardo R. 2010. Breeding for Quantitative Traits in Plants. 2nd Ed. Stemma Press, Minnesota.

Bernardo R. 2013. Genomewide markers for controlling background variation in association mapping. The Plant Genome, 6(1).

Bernardo R, Yu J. 2007. Prospects for genomewide selection for quantitative traits in maize. Crop Science, 47(3), 1082-1090.

Boer M, Wright D, Feng L, Podlich D, Luo L, Cooper M, Van Eeuwijk F. 2007. A mixed model QTL analysis for multiple environment trial data using environmental covariables for QTLxE with an example in maize. Genetics 177:1801–1813.

Borges A, Gonzalez A, Ernst O, Cadenazzi M, Gutierrez L. 2014. Comparison of experimental designs and spatial models based on simulated data using real field spatial heterogeneity in agricultural experimentation. In: Proceedings of the XXVII International Biometrics Conference. Florence, 2014.

Bradbury P, Parker T, Hamblin MT, Jannink JL. 2011. Assessment of power and false discovery rate in genome-wide association studies using the BarleyCAP germplasm Crop Sci. 51: 52-59.

Brownie C, Bowman DT, Burton JW. 1993. Estimating Spatial Variation in Analysis of Data from Yield Trials: A Comparison of Methods. Agronomy Journal, *85*, 1244–1253.

Burgueño J, de los Campos G, Weigel K, Crossa J. 2012. Genomic prediction of breeding values when modeling genotype x environment interaction using pedigree and dense molecular markers. Crop Science 52: 707-719.

Cappa EP, El-Kassaby YA, Garcia MN, Acuña C, Borralho NMG, et al. 2013. Impacts of Population Structure and Analytical Models in Genome-Wide Association Studies of Complex Traits in Forest Trees: A Case Study in *Eucalyptus globulus*. PLoS ONE 8(11): e81267. doi:10.1371/journal.pone.0081267

Casella G, Berger R. 1990. Statistical inference, Duxbury.

Casler MD, Undersander DJ. 2000. Forage Yield Precision, Experimental Design, and Cultivar Mean Separation for Alfalfa Cultivar Trials. Agronomy Journal, *92*, 1064–1071.

Cooper M, Messina CD, Podlich D, Totir LR, Baumgarten A, Hausmann NJ, .and Graham G. 2014. Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. Crop and Pasture Science, 65(4): 311-336.

Cullis B, Gogel B, Verbyla A, Thompson R. 1998. Spatial analysis of multi-environment early generation variety trials. Biometrics, 54(1), 1–18.

Cullis BR, Gleeson AC. 1991. Spatial Analysis of Field Experiments-An Extension to

Two Dimensions. Biometrics, 47(4), 1449–1460.

Dekkers JC, Hospital F. 2002. The use of molecular genetics in the improvement of agricultural populations. Nature Reviews Genetics, 3(1), 22-32.

de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL. 2012. Whole Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. Genetics (June 28). doi:10.1534/genetics.112.143313.

Diao G, Lin DY. 2006. Improving the power of association tests for quantitative t r a i t s in family studies. Genet. Epidemiol, 30, 301-313.

Duvick DN, Smith JSC, Cooper M. 2004. Long-term selection in a commercial hybrid maize breeding program. In 'Plant Breeding Reviews 24: Long term selection: Crops, animals, and bacteria'. Vol. 24, Part 2. (Ed. J Janick), pp. 109–151. (John Wiley & Sons: New York)

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*, *6*(5), e19379.

Fehr WR. 1987. Principles of cultivar development. Vol. 1, Theory and technique. Macmillan Publishing Company: London.

Gauch Jr HG. 1992. Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Elsevier Science Publishers.

Gianola D, Fernando RL, Stella A. 2006. Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics, 173(3), 1761-1776.

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES. 2014. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PloS one*, *9*(2), e90346.

Gonçalves E, Aubyn A, Martins, A. 2010. Experimental designs for evaluation of genetic variability and selection of ancient grapevine varieties: a simulation study. Heredity, *104*(6), 552-62.

Gonzalez-Reymundez A, Castro AJ, Speranza P, Gutierrez L. Relative efficacy of GWAS models for ordinal variables. In review Theoretical and Applied Genetics.

Grondona M, Crossa J, Fox P, Pfeiffer, W. 1996. Analysis of variety yield trials using two-dimensional separable ARIMA processes. Biometrics, *52*(2), 763-770.

Gutierrez L, Cuesta-Marcos A, Castro AJ, von Zitzewitz J, Schmitt M, Hayes PM. 2011. Association mapping of malting quality quantitative trait loci in winter barley: positive signals from small germplasm arrays. The Plant Genome 4:256-272

Gutierrez L, Berberian N, Capettini Flavio, Falcioni E, Fros D, Germán S, Hayes PM, Huerta-Espino, J, Herrera S, Pereyra S, Pérez C, Sandoval-Islas S, Singh R, Castro AJ. 2013. Genome-Wide Association Mapping Identifies Disease-Resistance QTLs in Barley Germplasm From Latin America. In: Advance in Barley Breeding. Springer, Netherlands. Pp 209 – 216.

Haley CS, Knott SA. 1992. A Simple Regression Method For Mapping Quantitative Trait Loci in Line Crosses Using Flanking Markers. Heredity 69:315-324

Hallauer AR, Miranda Filho JB. 1988. Quantitative genetics in maize breeding.' 2nd edn. Iowa State University Press: Ames, IA.

Hayes PM, Liu BH, Knapp SJ, Chen F, Jones B, Blake T, Franckowiak J, Rasmusson D, Sorrells M, Ullrich SE, Wesenberg D, Kleinhofs A. 1993. Quantitative trait locus effects and environmental interaction in a sample of North American barley germplasm. Theor. Appl. Genet. 87:392-401.

Hayes P, Szûcs P. 2006. Disequilibrium and association in barley: thinking outside the glass. Proceedings of the National Academy of Sciences, 103(49), 18385-18386.

Heffner EL, Sorrells ME, Jannink JL. 2009. *Genomic selection for crop improvement.* Crop Sci. 49**:** 1*–12.*

Henderson CR. 1984. Applications of linear models in animal breeding. University of Guelph.

Heslot N, Akdemir D, Sorrels ME, Jannink J-L. 2014. Integrating environmental

54

covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theoretical and Applied Genetics 127:463-480.

Hospital F, Charcosset A. 1997. Marker-assisted introgression of quantitative trait loci. Genetics, *147*(3), 1469-1485.

Iwata H, Ebana K, Fukuoka S. 2009. Bayesian multilocus association mapping on ordinal and censored traits and its application to the analysis of genetic variation among Oryza sativa L. germplasms». Theor. Appl. Genet., 18: 865-880.

Jannink J-L, Bink MCAM, Jansen RC. 2001. Using complex plant pedigrees to map valuable genes. Trends in Plant Science 6:337-342.

Kraakman ATW, Martinez F, Mussiraliev B, van Eeuwijk FA, Niks RE. 2006. Linkage disequilibrium mapping of morphological, resistance, and other agronomically relevant traits in modern spring barley cultivars. Molecular Breeding 17: 41-58.

Kraakman AT, Niks RE, Van den Berg PM, Stam P, Van Eeuwijk FA. 2004. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics, *168*(1), 435-446.

Kravchenko AN, Robertson GR, Snap SS, Smucker AJM. 2006. Using Information about Spatial Variability to Improve Estimates of Total Soil Carbon. Agronomy Journal, 98(3), 823–829.

Lado B, Matus I, Rodríguez A, Inostroza L, Poland JA, Belzile F, del Pozo A, Quincke M, Castro M, von Zitzewitz J. 2013. Increased Genomic Prediction Accuracy in Wheat Breeding Through Spatial Adjustment of Field Trial Data. G3 3:2105–2114.

Lado B, Quincke M, Silva P, Gutierrez L. 2014. Modelado de la Interacción Genotipo por Ambiente en Selección Genómica. In: Proceedings of the XV Jornadas de la SUB. Piriapolis, 2014.

Lamkey KR, Lee M (Eds). 2006 Plant breeding: The Arnel R. Hallauer International Symposium. Blackwell Publishing Ltd: Oxford, UK.

Lander ES, Botstein D. 1989. Mapping Mendelian Factors Underlying Quantitative Traits Using Rp Linkage Maps. Genetics 121:185-199.

Legendre P. 1993. Spatial autocorrelation: trouble or new paradigm? Ecology, 74(6), 1659–1673.

Legendre P, Dale MRT, Fortin M-J, Casgrain P, Gurevitch J. 2004. Effects of Spatial Structures on the Results of Field Experiments. Ecology, *85*(12), 3202–3214.

Locatelli A, Cuesta-Marcos A, Gutiérrez L, Hayes PM, Smith KP, Castro AJ. 2013. Genome-wide association mapping of agronomic traits in relevant barley germplasm in Uruguay. Molecular Breeding, *31*(3), 631-654.

Lorenzana RE, Bernardo R. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theoretical and Applied Genetics, 120(1), 151-161.

Mallarino A, Bermudez M, Wittry D, Hinz P. 2000. Alternative Data Managements and Interpretations. Fifth International Conference of Precision Agriculture. Minneapolis, MN.

Malosetti M, Ribaut J-L, van Eeuwijk FA. 2013. The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. Frontiers in Physiology. doi: 10.3389/fphys.2013.00044

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.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Visscher PM. 2009. Finding the missing heritability of complex diseases. Nature, *461*(7265), 747-753.

Masood AM, Farooq K, Mujahid Y, Anwar MZ. 2008. Improvement in Precision of Agricultural field Experiments through Design and analysis. Pakistan Journal of Life and Social Science, *6*(2), 89–91.

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. Theoretical and Applied Genetics, 117(7), 1077-1091

Meuwissen THE, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Miflin B. 2000. Crop improvement in the 21st century. Journal of Experimental Botany 51, 1–8. doi:10.1093/jexbot/51.342.1

Parisseaux B, Bernardo R. 2004. Insilico mapping of quantitative trait loci in maize. Theor. Appl. Genet. 109: 508-514.

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. PLoS genetics, *2*(12), e190.

Piepho H-P. 2000. A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. Genetics 156:2043–2050.

Poland JA, Rife TW. 2012. Genotyping-by-Sequencing for Plant Breeding and Genetics. Plant Genome 5:92–102.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38: 904–909.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Qiao CG, Basford KE, DeLacy IH, Cooper M. 2000. Evaluation of experimental designs and spatial analyses in wheat breeding trials. Theor Appl Genet, *100*, 9–16.

Quero G, Gutierrez L, Lascano R, Monza J, Sandal N, Borsani O. 2014. Identification of QTL for shoot and root growth under ionic-osmotic stress in Lotus, using a RIL population. Crop and Pasture Science http://dx.doi.org/10.1071/CP13222.

Rebai A. 1997. Comparison of methods for regression interval mapping in QTL analysis with non-normal traits, Genet. Res., 69: 69–74.

Roy JK, Smith KP, Muehlbauer GJ, Chao S, Close TJ, Steffenson BJ. 2010. Association mapping of spot blotch resistance in wild barley, Molecular Breeding 26: 243-256.

Sprague GF, Dudley JW (Eds). 1988. Corn and corn improvement. 3rd edn. American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc., Publishers: Madison, WI, USA.

Spyrides-Cunha M, Demétrio C, Camargo L. 2000. Proportional odds model applied to mapping of disease resistance genes in plants. Genet. Mol. Biol., 23, 223–227.

Stracke S, Haseneyer G, Veyrieras J-B, Geiger HH, Sauer S, Graner A, Piepho H-P. 2009. Association mapping reveals gene action and interactions in the determination of flowering time in barley, Theor. Appl. Genet. 118: 259-273.

Tanksley D. 1983. Molecular markers in plant breeding. Plant. Mol. Biol. Rep. 1: 3-8.

van Eeuwijk FA, Malosetti M, Yin X, Struik PC, Stam P. 2005. Statistical models for genotype by environment data: from conventional ANOVA models to eco-physiological QTL models. Aust J Agric Res 56:883–894.

Verbyla AP, Eckermann PJ, Thompson R, Cullis BR. 2003. The analysis of quantitative trait loci in multi-environment trials using a multiplicative mixed model. Crop and Pasture Science, 54(12), 1395-1408.

Verslues PE, Agarwal M, Katiyar-Agarwal S, Zhu J, Zhu JK. 2006. Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status. The Plant Journal 45, 523–539. doi:10.1111/j.1365-313X.2005.02593.x

Visscher PM, Haley CS, Knott SA. 1996. Mapping QTLs for binary traits in backcross and F2 populations. Genet. Res., 68: 55–63.

von Zitzewitz J, Cuesta-Marcos A, Condon F, Castro AJ, Chao S, Corey A, Filichkin T, Fisk SP, Gutierrez L, Haggard K, Karsai I, Muehlbauer GJ, Smith KP, Veisz O, Hayes PM. 2011. The genetics of winterhardiness in barley: perspectives from genome-wide association mapping. The Plant Genome 4(1): 76-91.

Waugh R, Jannink J-L, Muehlbauer GJ, Ramsay L. 2009. The emergence of

whole genome association scans in barley, Curr. Opin. Plant Biol. 12: 218-222.

Wu R, Ma C, Casella G. 2010. Statistical genetics of quantitative traits: Linkage, maps and QTL. Springer.

Yan W, Hunt LA, Sheng Q, Szlavnics Z. 2000. Cultivar evaluation and mega-environment investigation based on the GGE biplot. Crop Science, 40(3), 597-605.

Yu J, Buckler ES. 2006. Genetic association mapping and genome organization of maize. Current Opinion in Biotechnology, 17(2), 155-160.

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

Zhu CS, Yu JM. 2009. Nonmetric multidimensional scaling corrects for population structure in association mapping with different sample types. Genetics 182:875–888.

57