Journal of Scientific Research & Reports

20(4): 1-13, 2018; Article no.JSRR.44150
ISSN: 2320-0227

Genomic Association between SNP Markers and QTLs for Protein and Oil Content in Grain Weight in Soybean (*Glycine max*)

Damião Nascimento[1], Leandra Regina Texeira Polo[2], Fabiane Lazzari[2], Glacy Jaqueline da Silva[1] and Ivan Schuster[3*]

[1]Biotechnology Applied to Agriculture, UNIPAR, Umuarama, Brazil.
[2]R&D Department, Coodetec, Cascavel, Brazil.
[3]R&D Department, Longping High-Tech, Cravinhos, Brazil.

Authors' contributions

This work was carried out in collaboration between all authors. Author DN performed the phenotypic analysis. Author LRTP wrote the protocol and supported the protein and oils analysis. Author FL provided the genotypic analysis. Author GJS managed the literature search and. Author IS designed the study, performed the statistical analysis and wrote the first draft of the manuscript. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JSRR/2018/44150
Editor(s):
(1) Dr. Surapong Pinitglang, Department of Food Science and Technology, School of Science and Technology, University of the Thai Chamber of Commerce, Thailand.
(2) Dr. Ozlem Sultan Aslanturk, Department of Biology, Faculty of Art and Science, Adnan Menderes University, Turkey.
Reviewers:
(1) Henrique Padilha, Federal University of Pelotas, Brazil.
(2) João Everthon da Silva Ribeiro, Universidade Federal da Paraíba, Brazil.
(3) O. F. Adewusi, Federal University of Technology, Akure, Nigeria.
Complete Peer review History: http://www.sciencedomain.org/review-history/26948

Original Research Article

Received 30 July 2018
Accepted 23 October 2018
Published 31 October 2018

ABSTRACT

Aims: This work aimed to identify SNP marker associated with QTLs controlling grain weight and protein and oil content in soybean grains, using genome wide association study (GWAS).
Study Design: Experimental design for field study was Complete Randomized Block.
Place and Duration of Study: Departments of Breeding and Biotechnology at Coodetec in Cascavel, Palotina, Rio Verde and Sorriso, Brazil, between July 2013 and July 2016,
Methodology: A set of 168 soybean varieties were evaluated in five environments and genotyped with a panel of 6,000 SNP markers. Protein and Oil content was obtained through NIR (Near

**Corresponding author: E-mail: ivanschuster.ivan@gmail.com;*

Infrared Reflectance). GWAS was made by mixed linear model and multiple regression analysis.
Results: Six QTLs in five chromosomes explained from 10.4% to 26.6% of protein variation in three environments. Eleven QTLs in seven chromosomes explained from 14.4% to 39% of oil variation in five environments. Six QTLs in five chromosome explained 32.5% and 37.1% of grain weight variation in two environments.
Conclusion: For the same trait, different QTLs was identified in different environments. It means that to use marker assisted selection for this traits in soybean breeding, the markers need to be validated or identified in the breeding population in each environment, before being used for selection.

Keywords: Genome wide association study; GWAS; mixed linear model; multiple regression; stepwise.

1. INTRODUCTION

Soybean is the main agricultural commodity in Brazil, having been cultivated in 35 million ha in the country in 2017/18, with a total production of 117 million tons [1]. Soybeans are grown mainly for their protein and oil contents, and these constituents comprise approximately 60% of the dry matter of the grains. In addition, soybean protein has high biological value [2], both for human and animal feed.

Variation in protein and oil content in soybean is genetically controlled, but also strongly influenced by the environment conditions during the filling of the grains [3]. The interaction between the genetic and environmental components results in different levels of variations in each QTL expression, as answer of environmental variations. The identification of QTLs expressed in each environment can provide tools for identifying QTLs that are subject to interaction with the environment and those that are stable in more than one environment.

Since oil and protein contents in soybeans are negatively correlated [4,5], the selection for simultaneous genetic gains for the two traits is difficult to perform successfully. As one of the causes of genetic correlation is a genetic linkage, the identification of recombinants can contribute to the simultaneous gain of both traits. Molecular markers associated with linked QTLs can be used to monitor the breakdown of binding between these QTLs [6].

A large number of QTLs associated with protein and oil contents have already been identified in the soybean genome [7]. Many of these QTLs have been mapped with rather large confidence intervals, which diminishes the accuracy of these markers for marker assisted selection (MAS) in soybean breeding programs. Nowadays, with the cost-effective high density genotyping platforms and improved analytical methods for big data analysis, genome-wide association studies (GWAS) are promising forecasts in improving complex genetic traits in soybean. GWAS have the advantage of detecting smaller chromosomal regions affecting the trait and gives precise estimates of the size and direction of the effects of alleles in known loci [8].

The objective of this work was to identify QTLs associated with oil and protein contents, as well as grain weight, in a panel of Brazilian soybean varieties evaluated in multiple environments using high density genotyping and GWAS approach.

2. MATERIALS AND METHODS

2.1 Study Material and Area

A total of 168 Brazilian soybean varieties were evaluated in five environments: Cascavel, PR (781m altitude, latitude 24°57'20"S, longitude 53°27'19"W) in 2013-14 and 2014-15; Rio Verde, GO (715m altitude, latitude 17º47'53"S, longitude 50º55'41"W) in 2013-14 and 2014-15; and Sorriso, MT (365m altitude, latitude 12º32 '43 "S, longitude 55º42'41"W) in 2014-15.

2.2 Experimental Design and Trait Analysis

At each site, the experiments were conducted in a complete randomized block design with two replicates. The plots consisted of four rows, 5m long, with spacing of 0.45m. The determination of protein and oil content was performed at Coodetec, in Cascavel, Parana, Brazil. Twenty grams of grains from each experimental unit were milled in a Cyclone Sample Mill, and 3.6g of the milled sample was used for analysis. Protein, oil and moisture content was obtained by NIR (Near Infrared Reflectance) method in Dickey

John FT-NIR equipment (model Instalab 600). The results were expressed as a percentage of the oil and protein contents in the dry matter, after correction as a function of the moisture content. One hundred grain weight was obtained by weighing the grains in electronic scale, correcting the weight to the humidity of 13%.

Genomic DNA was extracted from leaf tissues collected from a mix of ten plants of each variety. DNA-easy Plant Kit (Qiagene) was used for DNA extraction, in the biotechnology lab of Coodetec, in Cascavel, Parana, Brazil. The samples were genotyped with 6,000 SNP (single nucleotide polymorphisms) using the Illumina BARCSoySNP6K BeadChip, which corresponds to a subset of SNPs from the SoySNP50K BeadChip [9]. Genotyping was conducted by Deoxi Biotechnology Ltda, in Aracatuba, Sao Paulo, Brazil. A total of 3,780 SNP markers, including polymorphic and non-redundant SNPs, SNP markers with greater than 10% minor allele frequency (MAF) and missing data value lower than 25% were used in subsequent analysis, with heterozygous markers treated as missing data according to Hwang et al. [10].

2.3 Data Analysis

The phenotypic data from each environment were submitted to the Lilliefors normality test using the software GENES [11]. Characteristics that deviated from the normal distribution were not used in genomic association analysis.

Population structure was inferred using the Bayesian clustering method implemented in the program InStruct [12]. The posterior probabilities were estimated using five independent runs of the Markov Chain Monte Carlo (MCMC) sampling algorithm for the numbers of groups genetically differentiated (k) varying from 2 to 10, without prior population information. The MCMC chains were run with 5,000 burn-in period, followed by 50,000 iterations. The convergence of the log likelihood was determined by the value of the Gelman-Rubin statistic. The best estimate of k groups was determined according to the lowest value of the average log-Likelihood and Deviance Information Criterion (DIC) values among the simulated groups [12]. The population structure Q matrix was used in association mapping analysis to avoid false positive SNP-phenotype association due to the population structure.

To account for the effects of population structure (Q) and genetic relatedness (K) among the cultivars, a mixed linear model (MLM) of association was employed. The structured association model (Q model), taking into account the genetic structure of the population was included in the association mixed model. The kinship coefficient matrix (K) that explain the most probable identity by state of each allele between cultivars was estimated using the program TASSEL [13,14].

To identify non-redundant markers, markers with -Log10(P) value higher than 2 was used in multiple regression analysis, with Stepwise variable selection procedure and 5% probability of entry and exit. Multiple regression analysis was performed with the JMP program [15].

To identify markers on linkage disequilibrium in regions containing QTLs, linkage disequilibrium analysis was performed using the Haploview program [16]. Linkage disequilibrium blocks were formed by markers whose D' value was greater than 80%.

3. RESULTS AND DISCUSSION

The three characteristics evaluated showed a great variability between and within the five environments (Fig. 1), which demonstrates the effects of genetic (variation within environmental) and environmental (variation between environments) effects on the expression of these characteristics.

Protein content in soybean grains ranged from 29% to 48.1%, considering the five environments evaluated. In the Central region of Brazil (Rio Verde 2013-14 and 2014-15, and Sorriso 2014-15), the frequency distribution of protein contents was similar (Fig. 1), with a higher frequency between 41% and 42% in Rio Verde 2013-14, and between 39% and 42% in Rio Verde and Sorriso in 2014-15. There was variation in the frequency distribution of protein contents in the two years of evaluation in Cascavel (2013-14 and 2014-15). In 2013-14 the protein content was higher in all varieties, with a higher frequency between 42% and 43%, and in 2014-15 it was lower, with a higher frequency between 36% and 37%.

Table 1. Significant markers in the multiple regression with the stepwise method of model selection for the traits protein and oil content and 100 grain weight, in soybean

Trait	Maker	Chrom	Effect	R^2
Protein Rio Verde 2013-14	Gm08_3272385_G_T	8	0.52	0.266
	Gm08_17579484_C_T	8	0.59	
	Gm19_37893995_G_A	19	0.87	
Protein Rio Verde 2014-15	Gm10_43840376_T_C	10	0.56	0.141
	Gm13_27212330_G_A	13	0.44	
Protein Sorriso 2014-15	Gm20_30417244_C_T	20	0.61	0.104
Oil Cascavel 2013-14	Gm01_1059407_T_C	1	0.91	0.390
	Gm08_21933156_G_A	8	0.75	
	Gm14_27937142_C_T	14	0.32	
Oil Rio Verde 2013-14	Gm03_46214163_C_T	3	0.51	0.359
	Gm04_40811025_C_A	4	0.27	
	Gm06_1328895_T_C	6	0.37	
	Gm07_43189903_A_G	7	0.66	
Oil Cascavel 2014-15	Gm14_9642828_T_C	14	0.71	0.206
	Gm06_44474853_A_G	6	0.66	
Oil Rio Verde 2014-15	Gm03_3334303_C_A	3	0.56	0.144
Oil Sorriso 2014-15	Gm07_35194991_A_G	7	0.32	0.156
	Gm08_21172458_T_C	8	0.30	
100 grain weight	Gm02_48874048_G_A	2	0.55	0.325
Rio Verde 2014-15	Gm04_42638631_T_C	4	0.74	
	Gm19_40053178_G_A	19	0.51	
100 grain weight	Gm13_22446130_C_T	13	1.00	0.371
2014-15	Gm08_2671408_T_C	8	0.47	
	Gm13_6811934_T_G	13	0.82	

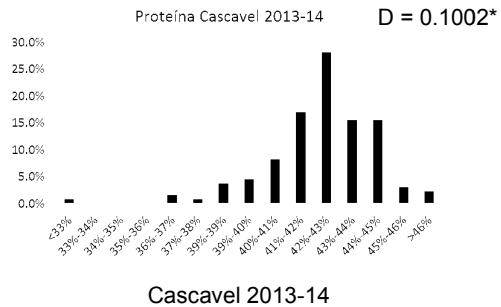
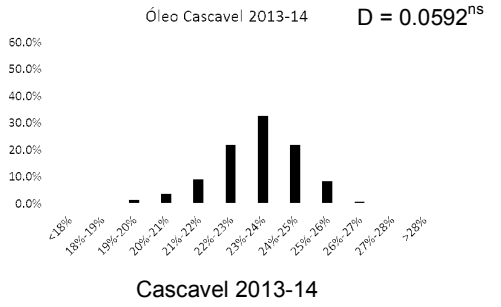
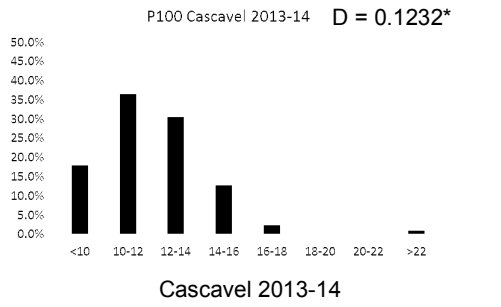
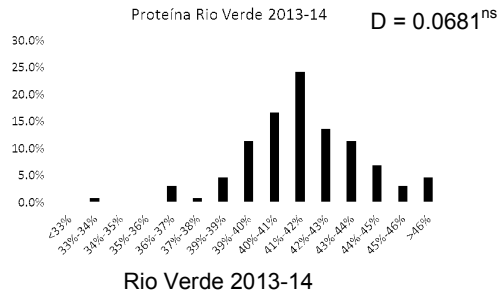
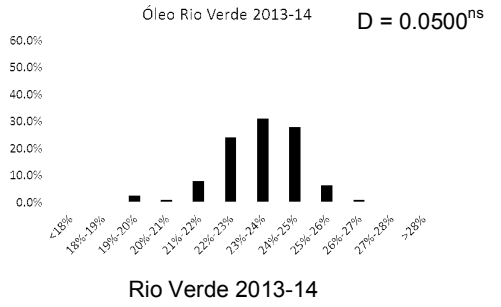
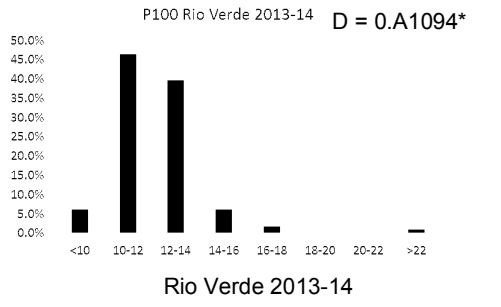
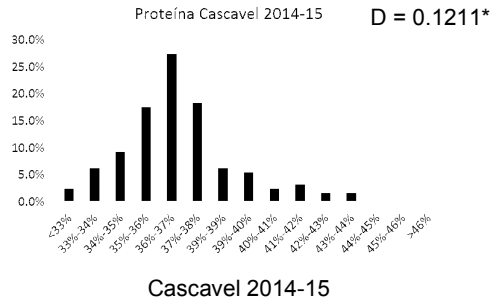
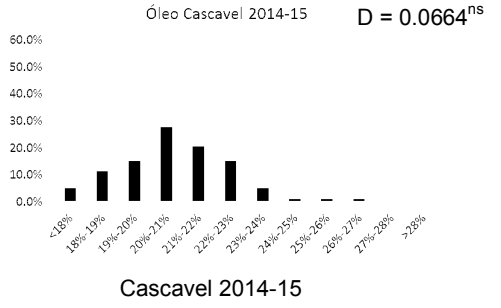
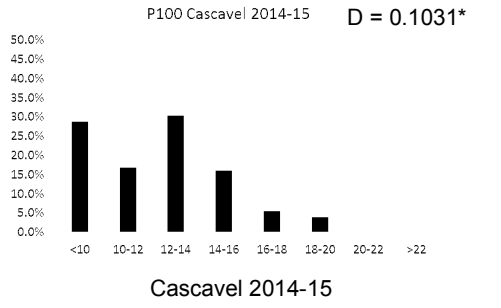
The oil content ranged from 14.9% to 29.3% in the five environments. The lowest oil content was observed in Cascavel 2014-15, with a higher frequency of varieties producing between 20% and 21% of oil (Fig. 1). The highest content was observed in Sorriso 2014-15, with the highest frequency of varieties producing between 25% and 26% of oil. In Rio Verde 2014-15 the highest frequency of oil content was between 24% and 25%, and in Cascavel 2013-14 and Rio Verde 2013-14, the highest frequency of oil content was between 23% and 24%.

The weight of 100 grains varied from 6.6g to 25.0g. Heavier grains were obtained in Rio Verde 2014-15 and Sorriso 2014-15, with a higher frequency between 14g and 16g per 100 grains (Fig. 1). The lighter grains were obtained in Rio Verde 2013-14, with a great frequency of varieties with less than 10g per 100 grains, due to water stress in the period and grain filling in this environment. In Cascavel 2013-14 and 2014-15, the highest frequency of the varieties had weight of 100 grains between 10g and 12g.

Quantitative characteristics are influenced by environmental effects, and this influence can be observed in the results presented in Fig. 1. The same set of variances, evaluated in different environments, may present different values for the quantitative characteristics. And this influence of the environment on the expression of the quantitative characteristics can be as much between different places in the same year of cultivation, as between different years of cultivation in the same place.

The frequency distributions of protein contents in Cascavel in 2013-14 and in 2014-15, and the weight of 100 grains in Cascavel 2013-14, 2014-15 and in Rio Verde 2013-14 deviated from the frequencies expected under normal distribution (Fig. 1). These data were not used for GWAS.

In the three environments in which the protein content data were considered (Rio Verde 2013-14, 2014-15 and Sorriso 2014-15), six QTLs associated with protein content were identified in the chromosomes 8, 10, 13, 19 and 20 (Table 1). In Rio Verde 2013-14, two markers were identified on chromosome 8. These two markers are located at a distance of 14.3Mb, and in addition, are in linkage equilibrium (Fig. 2), and are two distinct QTLs on chromosome 8.

Protein	**Oil**	**P100**
 Cascavel 2013-14	 Cascavel 2013-14	 Cascavel 2013-14
 Rio Verde 2013-14	 Rio Verde 2013-14	 Rio Verde 2013-14
 Cascavel 2014-15	 Cascavel 2014-15	 Cascavel 2014-15

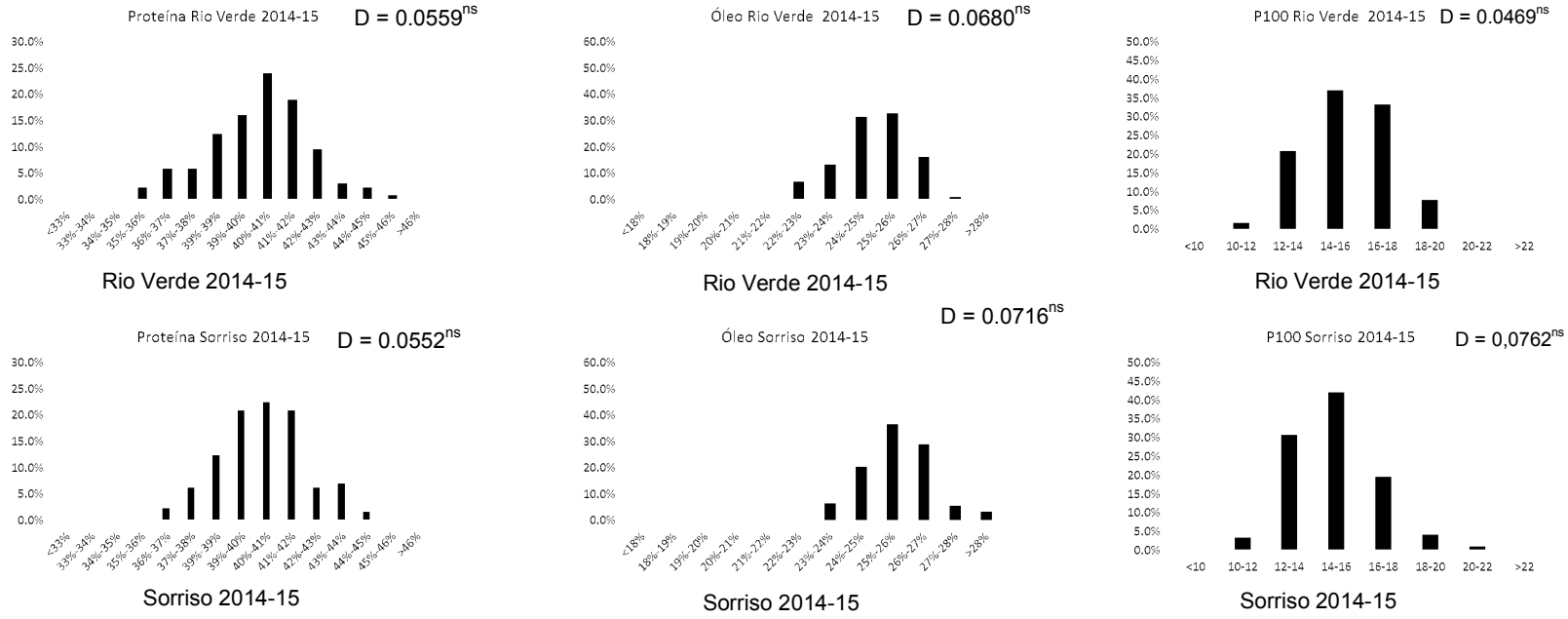


Fig. 1. Distribuition of frequency of protein and oil content, and 100 grain waight (P100) in soybean, in multiple environments. The graphics for the same trait are in the same scale. D = Dcalculated in the Lilliefors normality test. *significative at 5% probability. nsnon-significative. D tabelated for 5% probability = 0.0771

No QTL associated with protein content was repeated in more than one environment (Table 1). The three QTLs identified in Rio Verde 2013-14 explained 26.6% of the observed variation in protein content. The two QTLs identified in Rio Verde 2014-15 explained 14.1% and the QTL identified in Sorriso 2014-15 explained 10.4% of the variation observed in protein content.

The difference in the average protein content of the varieties containing the haplotypes associated with the lowest protein content and the varieties containing the haplotypes associated with the highest protein content was 4.3, 2.1 and 1.2 percentage points in Rio Verde 2013-14, Rio Verde 2014 -15 and Sorriso 2014-15, respectively (Fig. 3). It is noted that the difference in protein content between haplotypes is proportional to the number of markers, and consequently of QTLs, in each haplotype. This demonstrates the additive effect of QTLs associated with protein content. Each QTL contributes a small part of the variation, and the total variation is the sum of the effects of each QTL. Fig. 3a shows this continuous variation in protein contents as a function of the allelic composition of the haplotypes.

For the oil content, 11 QTLs were identified on 7 chromosomes (1, 3, 4, 6, 7, 8 and 14). In chromosome 8, two closely related markers associated with the oil content were identified (Table 1). The marker Gm08_21933156_G_A was significantly associated with the oil content in Cascavel 2013-14 and the marker Gm08_21172458_T_C in Sorriso 2014-15. These markers are 760kb apart. The marker Gm08_21933156_G_A is located in a block of linkage disequilibrium with four other markers (Fig. 2), and the marker Gm08_21172458_T_C is closely linked to this block of linkage disequilibrium.

If considered in pairs, the marker Gm08_21443387_G_T is in linkage disequilibrium with both the marker Gm08_21933156_G_A and Gm08_21172458_T_C (Fig. 2). Given the proximity of these markers to the soybean genome, and both markers are in linkage disequilibrium with a third marker (Gm08_21443387_G_T), these two markers will be associated with the same QTL, and this QTL was identified in two environments: Cascavel 2013-14 and Sorriso 2014-15. None of the other 10 QTLs associated with oil content was repeated in more than one environment.

In each of chromosomes 3, 6, 7 and 14, two markers associated with the oil content were identified. In these chromosomes, the markers associated with the oil content are in linkage equilibrium (Fig. 2), and therefore, each marker is associated with a different QTL. That is, in each of the chromosomes 3, 6, 7 and 14 two QTLs associated with the oil content was identified.

The three QTLs identified in Cascavel 2013-14 explained 39% of the observed variation in oil content. In Rio Verde 2013-14, four QTLs accounted for 35.9% of the variation. In Cascavel 2014-15, two QTLs explained 20.6% and in Rio Verde 2014-15 one QTL explained 14.4% of the variation observed. In Sorriso 2014-15, two QTLs explained 15.6% of the variation (Table 1).

The differences in the average oil content among the varieties containing the haplotypes associated with the lowest and the highest oil content were 3.4, 5.0, 2.4, 1.1 and 1.3 percentage points in Cascavel 2013-14, Rio Verde 2013-14, Cascavel 2014 -15, Rio Verde 2014-15 and Sorriso 2014-15, respectively (Fig. 3).

Six QTLs associated with the weight of 100 grains of soybean were identified, three in Rio Verde 2014-15, which explained 32.5% of the variation, and three in Sorriso 2014-15, which explained 37.1% of the variation (Table 1). In the other environments, genomic association analysis was not performed since the weight of 100 grains had no normal distribution. The two markers significantly associated with the weight of 100 grains identified on chromosome 13 in Sorriso 2014-15 are associated with two distinct QTLs, as they are distant in the soybean genome (15.6Mb), and are in linkage disequilibrium (Fig. 2).

No QTL associated with the weight of 100 grains was significant in both environments simultaneously. The difference between the average of 100 grains weight of the varieties with the haplotypes associated with the lowest weight of grains and the varieties with the haplotypes associated with the highest grain weight was 3.8g in Rio Verde 2014-15 and 4.2g in the 2014-15.

Protein and oil contents in soybean grains are negatively correlated [4, 5]. Pleiotropy and gene linkage are the cause of genetic correlation. No pleiotropic QTL was identified in this study. QTLs associated with protein and oil contents were

observed on chromosome 8. In the chromosome 8, QTLs associated with the three evaluated traits were identified (Figure 2), demonstrating the genetic linkage of these characteristics, and consequently, the correlation that may exist between them.

QTLs for the protein content was also identified linked to QTLs for the weight of 100 grains on chromosomes 13 and 19. For the oil content and weight of 100 grains, linked QTLs were identified on chromosome 4 (Fig. 2).

In the soybean genome, 83 QTLs for protein content, 78 QTLs for oil content and 85 QTLs for grain weight are mapped [7]. QTLs for these three traits are mapped on all the 20 soybean chromosomes. Given the number of QTLs and the number of chromosomes of soybean, it is expected that it has several linked QTLs associated with these characteristics in the soybean genome.

Several research have reported QTLs for protein and oil content in soybean varieties in Brazil. Soares et al. [17] mapped five QTLs on chromosomes 1, 3, 6, 15, and 18, which explained between 7.34% and 14.37% of the variation in protein content. Rodrigues et al. [18] mapped four QTLs that explained from 6.24% to 18.94% of the variation in protein content in Brazilian soybean varieties, on chromosomes 1, 5, 18, and 20. The same authors also identified three QTLs that explained 17.26% to 25.93% of the variation for the oil content in chromosomes 5, 10 and 20.

All these studies were carried out on mapping populations derived from controlled crosses. This is the first work to identify QTLs associated with protein and oil contents using greater germplasm variability, represented by a collection of varieties, and analysis of broad genomic association. At our knowledge, for 100 grain weight there are no mapping results from Brazilian varieties until now.

In this work, in the GWAS using a mixed linear model, few significant markers were identified at 0.1% probability (Fig. 4). This is the level of confidence typically used in GWAS jobs that use mixed linear models. However, at this level of probability it is expected that only QTLs of greater effect is significant. Due to the restricted level of significance, the chance of type II error (not identifying true QTLs) is greater. QTLs of lower effect associated with quantitative traits may possibly be identified at a lower level of significance. But using a lower level of significance can increase the chance of type I error (assume with true QTLs that do not exist).

To balance error types (I and II), we used sequential analysis of MLM at 1% confidence level, followed by multiple regression analysis using only significant markers in the previous analysis. The use of all markers in multiple regression analysis could lead to the identification of spurious associations due to the presence of subpopulations or family structure within the population. Using only the markers selected by the MLM analysis, these spurious associations were avoided, since the population structure was considered in the MLM assessment.

Using this sequential analysis of MLM and multiple regression with stepwise method of model choice, more markers were identified. By comparing the haplotypes formed by these markers it was possible to observe considerable differences between the averages of each haplotype, especially among the haplotypes associated with the highest and lowest values. This means that the use of this MLM and multiple regression analysis strategies was efficient in reducing type II error, without an increase in type I error.

The QTLs identified in each environment were different for the same characteristic. This difference in expression of QTLs as a function of the environment illustrates the GxE interaction. Considering that in each environment different QTLs are responsible for the expression of the contents of protein and oil, and weight of 100 soybean grains, and even in the same place, there is a difference between QTLs expressed between the years, the conventional strategy of MAS cannot be used. Instead, a preliminary association analysis should be performed in each generation of selection and in each environment.

To do this, part of the soybean lines under evaluation must be assessed phenotypically and also genotyped with a set of thousands of molecular markers. In the initial stages of evaluation, when thousands of lines are being evaluated, less than 5% of the lines need to be evaluated phenotypically, to allow analysis of genomic association. Once the genotypic and phenotypic data from this sample of the lines are available, the genomic association analysis can

be performed in each environment, and the significant markers in each environment can be used to select the rest of the lines, which do not need to be evaluated phenotypically.

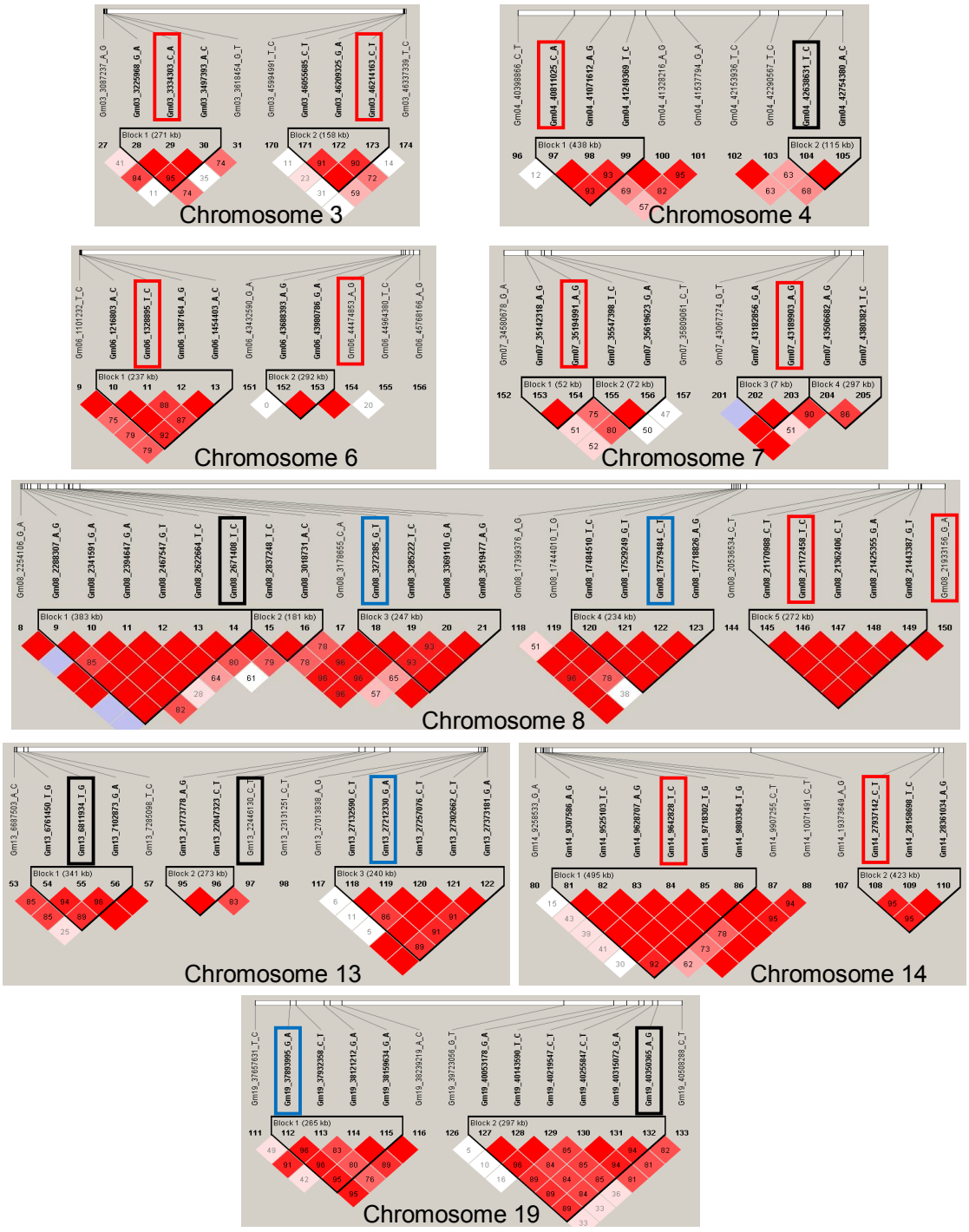


Fig. 2. Linkage disequilibrium map of the regions containing QTLs in chromosomes with more than one QTL. Blue rectangle identifies markers associated with protein content, red rectangle identifies markers associated with oil content and black rectangles, markers associated with 100 seed weight

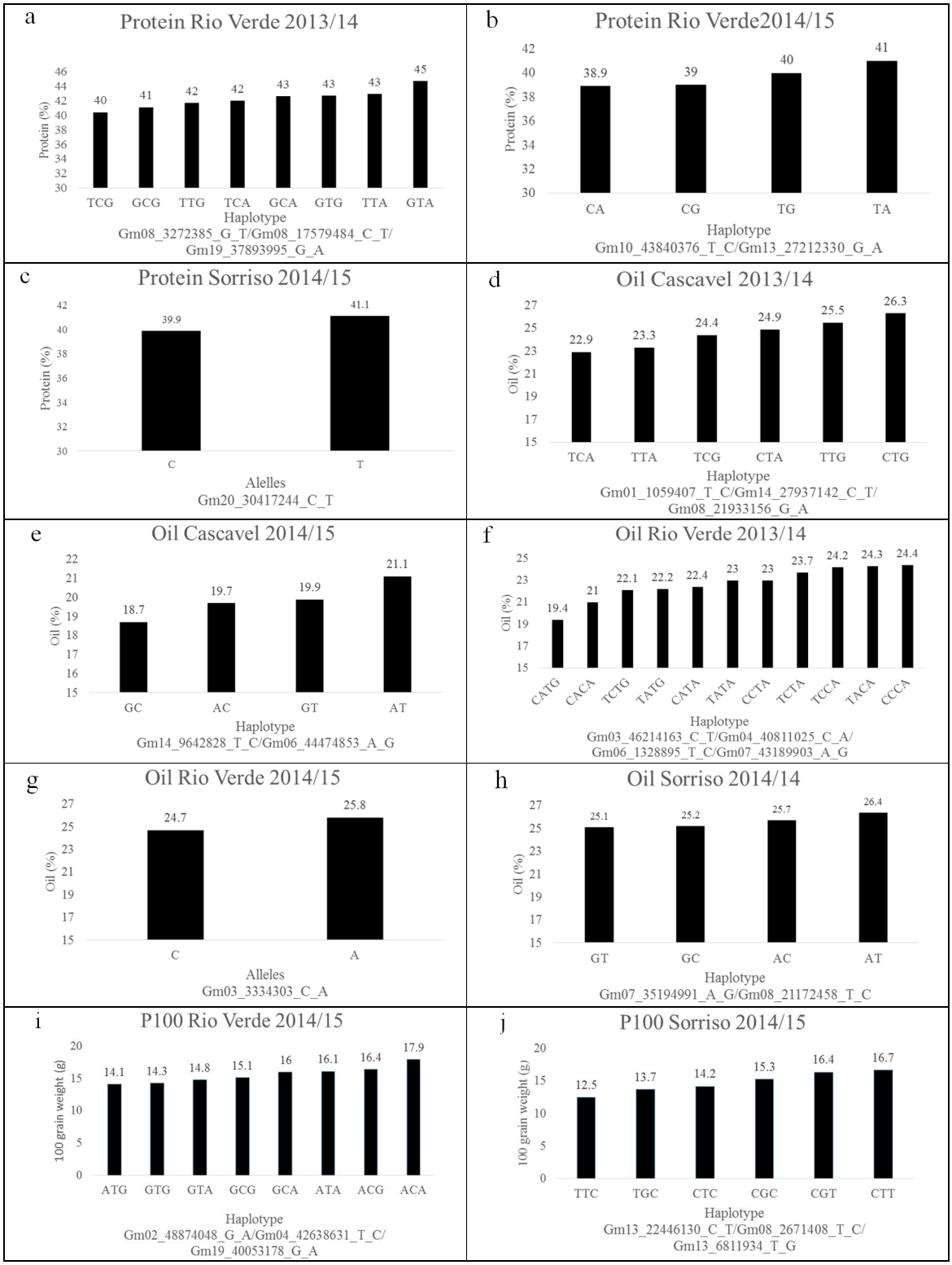


Fig. 3. Average of protein and oil content, and 100 grain weight in soybean varieties grouped by haplotypes based on markers associated with QTLs identified in each environment

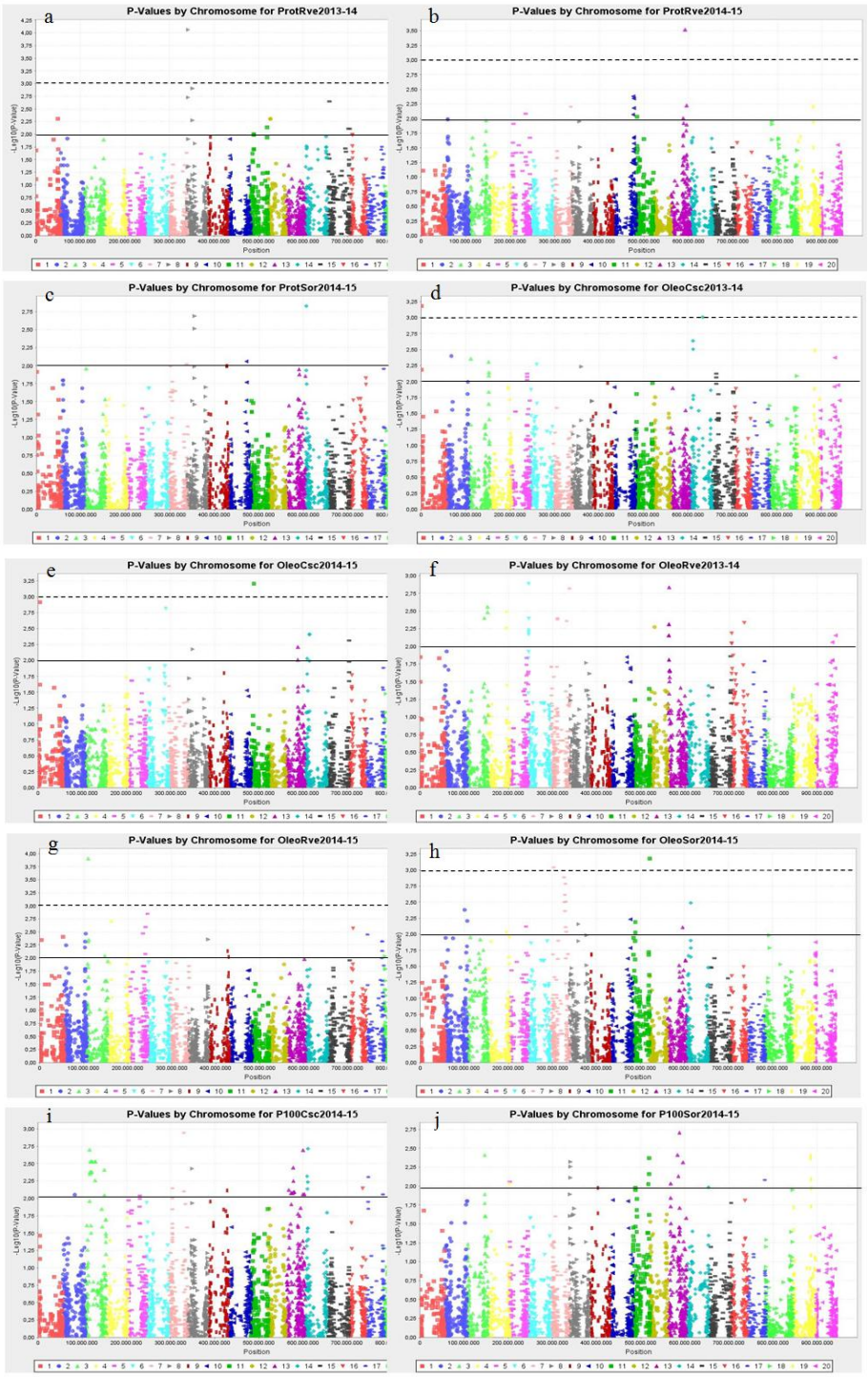


Fig. 4. Manhatan plot of –Log10(P) from GWAS analysis in soybean using mixed linear model. a) Protein Rio Verde 2013-14. b) Protein Rio Verde 2014-15. c) Protein Sorriso 2014-15. d) Oil Cascavel 2013-14. e) Oil Cascavel 2015-15. f) Oil Rio Verde 2013-14. g) Oil Rio Verde 2014-15. h) Oil Sorriso 2014-15. i) 100 grain weight Rio Verde 2014-15. j) 100 grain weight Rio Verde 2014-15. Continuous horizontal line is the threshold for 1% probability (-Log10(P)=2). Dotted horizontal line, when present, is the threshold for 0.1% probability (-Log10(P)=3).

4. CONCLUSION

SNP haplotypes identified in this work explained almost the third part of phenotypic variation in protein and oil content in soybean, and in 100 seed weight. Due the GxE interaction, the SNP haplotypes associated with this traits are different in different environments. For breeding application of SNP haplotypes for this traits, the SNP discovery needs to be performed each year, in each environment. A sample of less than 5% of the breeding population needs to be evaluated phenotypically and also genotyped with some thousands of SNP markers, for genomic association study. The markers identified in this association study can be used to select the rest of the population, without phenotypic evaluation.

ACKNOWLEDGEMENTS

To COODETEC, for the support for lab analysis and field evaluation. This research was supported by the CNPq (National Council for Scientific and Technological Development) grants no. 560465/2010-6 and 305900/2011-0.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. CONAB, Companhia Nacional de Abastecimento. Safras: Séries históricas – Milho Total. Accessed April, 15.
   Available:https://www.conab.gov.br/info-agro/safras/serie-historica-das-safras?start=20. Portuguese.
2. Messina M, Messina V, Jenkins DJ. Can breast cancer patients use soyafoods to help reduce risk of CHD. Brasil. Journal of Nutrition. 2012;5(108):810-819.
3. Pípolo AE, Sinclair TR, Camara GMS. Effects of temperature on oil and protein concentration in soybean seeds cultured in vitro. Annals of Applied Biology. 2004;144:71-76.
4. Bandillo N, Jarquin D, Song Q, Nelson R, Cregan P, Specht J, Lorenz AA. Population structure and genome-wide association analysis on the USDA soybean germplasm collection. The Plant Genome. 2015;8:1-13.
5. Dias DA, Polo LRT, Lazzari F, Silva GJ, Schuster I. Genome-wide association for mapping QTLs linked to protein and oil contents in soybean. Pesquisa Agropecuaria Brasileira. 2017;52:896-904.
6. Schuster I and Mora, F. Biometria aplicada ao melhoramento de espécies anuais e seus desafios. In: Ludke WH, et al. Editors. Desafios Biométricos no Melhoramento Genético. Gen Melhor, Universidade Federal de Viçosa, Viçosa; 2017. Portuguese.
7. Schuster I. Soybean genetic mapping. In Silva FL, Borem A, Sediyama T, Ludke WH, editors. Soybean breeding. Springer, New York; 2017.
8. Contreras-Soto RI, Mora F, Oliveira MAR, Higashi W, Scapim CA, Schuster I. A Genome-Wide association study for agronomic traits in soybean using SNP markers and SNP-based haplotype analysis. PLoS ONE. 2017;12(2): e0171105.
9. Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL, Cregan PB. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS ONE. 2013;8: e54985.
10. Hwang EY, Song Q, Jia G, Specht JE, Hyten DL, Costa J, Cregan PB. A genome-wide association study of seed protein and oil content in soybean. BMC Genomics. 2014;15:1.
11. Cruz CD. GENES - a software package for analysis in experimental statistics and quantitative genetics. Acta Scientiarum. 2013;35:271-276.
12. Gao H, Williamson S, Bustamante CD. An MCMC Approach for Joint Inference of Population Structure and Inbreeding Rates from Multi-Locus Genotype Data. Genetics. 2007;176(3):1635-1651.
13. Endelman JB, J-L Jannink. Shrinkage estimation of the realized relationship matrix. G3:Genes, Genomes, Genetics. 2012;2(11):1405-1413.
14. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23:2633-2635.
15. SAS, Institute INC. SAS/STAT user's guide, version 6.0. 3rd ed. Cary, NC; 1990.

16. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21(2):263-265.

17. Soares TCB, Good-God PIV, Miranda FD, Soares, YJB, Schuster I, Piovesan ND, Barros EG, Moreira MA. QTL mapping for protein content in soybean cultivated in two tropical environments. Pesquisa Agropecuaria Brasileira. 2008;43:1533-1541.

18. Rodrigues JIS, Miranda FD, Ferreira A, Borges LL, Ferreira MFS, Good-God PIV, Piovesan ND, Barros EG, Cruz CD, Moreira MA. Mapeamento de QTL para conteúdos de proteína e óleo em soja. Pesquisa Agropecuaria Brasileira. 2010; 45(5):472-480. Portuguese.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history/26948