QTL mapping for grain quality traits in testcrosses of a maize biparental population using genotyping-by-sequencing data

QTL kartiranje za svojstva kvalitete zrna kod testkrižanaca biparentalne populacije kukuruza korištenjem podataka genotipizacije sekvenciranjem

Galić, V., Franić, M., Jambrović, A., Zdunić, Z., Brkić, A., Šimić, D.

Poljoprivreda/Agriculture

ISSN: 1848-8080 (Online)
ISSN: 1330-7142 (Print)

http://dx.doi.org/10.18047/poljo.23.1.5

Poljoprivredni fakultet u Osijeku, Poljoprivredni institut Osijek

Faculty of Agriculture in Osijek, Agricultural Institute Osijek

ISSN 1330-7142
UDK: 631.523/.526:633.15
DOI: 10.18047/poljo.23.1.5

QTL MAPPING FOR GRAIN QUALITY TRAITS IN TESTCROSSES OF A MAIZE BIPARENTAL POPULATION USING GENOTYPING-BY-SEQUENCING DATA

Galić, V.[(1)], Franić, M.[(1)], Jambrović, A.[(1,2)], Zdunić, Z.[(1,2)], Brkić, A.[(1)], Šimić, D.[(1,2)]

Original scientific paper
Izvorni znanstveni članak

SUMMARY

We performed QTL mapping in testcrosses of maize population IBMSyn4 for three grain quality traits: oil and protein contents and test weight. 191 phenotyped and genotyped lines were used as a training set while 85 genotyped only lines comprised a validation set used to calculate best linear unbiased predictions (BLUP), making a total of 276 phenotypes for the QTL analysis. 92000 filtered Genotyping-By-Sequencing (GBS) SNP markers were used to calculate BLUPs, while a set of 2178 genetically mapped SSRs was used in QTL analysis. By simple QTL scan, we scored several minor effect QTLs: one for oil content (chromosome 1), one for protein content (chromosome 10) and four for test weight (chromosomes 1, 3, 5 and 10). QTLs associated with test weight were found to be additive, and 18.25% of phenotypic variance was explained by their joint effect. Only one QTL for test weight was found to be significant in composite interval mapping and it was mapped on chromosome 5. This QTL accounted for 9.97% of phenotypic variance. QTLs detected in this study represent monitoring of commercially most successful elite maize germplasm for grain quality traits.

Key-words: best linear unbiased predictions, IBM population, maize, quantitative trait loci, grain quality traits

INTRODUCTION

In developing countries, maize grains are the only source of proteins and caloric value for several million people (Nuss and Tanumihandrjo, 2011), as well as a valuable source of oil (Prasanna et al., 2001). Grain quality traits such as the content of oil and proteins and test weight play an important role in determining the value of produced maize. Test weight in maize reflects a degree of maturity, uniformity and integrity and presents an important parameter in grading grains for the different purposes (Quiang-Ding et al., 2011). There are several options of phenotyping these traits, but the one that offers the highest throughput with satisfactory reliability and repeatability of results is near infrared transmittance (NIT) analysis (Lee et al., 2007). Selection for grain oil and protein contents are some of the longest-spanning, still perpetuating selection strains with very high efficiency and gain per cycle of breeding (Bennetzen and Hake, 2009). Grain oil and protein contents, as well as test weight, are quantitative traits of high heritability, and associated quantitative trait loci (QTLs) were identified in a number of studies (Goldman et al., 1993; Goldman et al., 1994; Wassom et al., 2008; Zhang et al., 2008; Quiang-Ding et al., 2011). Enrichment of maize cultivars for grain oil and protein content presents an increase in their nutritional values, while test weight, being a complex trait, offers a number of benefits. So far, traditional breeding approaches have been used in

(1) Vlatko Galić, mag. ing. agr (vlatko.galic@poljinos.hr), Mario Franić, prof. biol. i kem., dr. sc. Antun Jambrović, dr. sc. Andrija Brkić, dr. sc. Domagoj Šimić - Agricultural Institute Osijek, Department of Maize Breeding and Genetics, Južno predgrađe 17, 31103 Osijek, Croatia, (2) Dr. sc. Antun Jambrović, doc. dr. sc. Zvonimir Zdunić, dr. sc. Domagoj Šimić - Centre of Excellence for Biodiversity and Molecular Plant Breeding, Svetošimunska 25, 10000 Zagreb, Croatia

breeding for these traits (Dorsey – Redding et al., 1991), although their high heritability and quantitative nature make them a perfect fit for genomic selection (GS). GS allows predictions of phenotypic and breeding values of individuals that have been only genotyped (validation set and other lines derived from the population) based on genetic marker data that has been accounted for marker effects (trained) by best linear unbiased prediction (BLUP) in a training set (genotyped and phenotyped) (Meuwissen et al. 2001). BLUP calculations have been used extensively in maize breeding (Bernardo, 1996; Balint-Kurti et al., 2010; Lian et al., 2015). The prerequisite for BLUP is genotypic covariance between lines in genotype space. At the core of its calculation is a maximum-likelihood algorithm for mixed models able to account for a single variance component (marker effects) besides residual error (Endelman, 2011).

Objectives of our study were: a) to perform best linear unbiased predictions (BLUPs) to scan for QTLs in full set of lines from IBM*Syn4* population; b) to detect QTLs for oil content, protein content and test weight in testcrosses of IBMSyn4 population.

MATERIAL AND METHODS

Plant material, phenotyping and experimental design

A total of 191 intermated recombinant inbred lines (IRILs) of the maize biparental IBM*Syn4* population (the cross between the inbred lines B73 and Mo17 intermated for four generations, Lee *et al*. 2002) were testcrossed to the Agricultural Institute Osijek proprietary line 84-28A of Iodent genetic background. The experiment was set as an unreplicated incomplete block design in the growing season of 2015 in Osijek. Ears were hand-harvested and shelled at approximately 20% moisture. 500 g samples were collected and the measuring of oil and protein contents, and test weight were performed with NIT Infratec Grain Analyzer machine model 1241. Each sample was measured three times and average of three measurements was used for further analysis.

Genotyping

Genotyping By Sequencing (GBS) was performed by Panzea (panzea.org) with the enzyme *Ape*KI according to Elshire *et al.* (2011). The total of 955,690 SNPs was extensively filtered and finally, ~92000 SNP markers were chosen. These markers were not genetically mapped, and they were used only for calculation of BLUPs since prediction accuracy is directly influenced by marker density. 2178 SSRs (Simple Sequence Repeat, Andorf et al., 2010) comprised another set of genetically mapped markers that was used for QTL mapping.

Statistical analysis

Filtering of GBS data was performed in TASSEL version 5 software (Bradbury et al., 2007) by exclusion of rare alleles and residual heterozygosity. All other statistical analyses were performed in R programming language (R Core Team, 2012). Package {rrBLUP} was used for BLUP calculations. 191 genotyped and phenotyped IRILs were used as a training set, while the other 85 lines that were genotyped only were used as a validation set for predictions. BLUPs were calculated using the equation:

$$y = \mu + Xg + \varepsilon \quad (1)$$

y = phenotypic mean, μ = overall mean of training set, X = marker matrix, g = marker effects, ε = residual effects

QTL analysis was performed with {qtl} package (Broman et al., 2003) in a set of 276 lines comprised of 191 observed entries and 85 predictions obtained by BLUP. Initial QTL scan was performed with scanone function assuming one QTL per chromosome and offering a loose implementation of WinQTL "simple interval mapping" function. LOD score confidence intervals were calculated with bootstrap procedure running 1000 permutations. Putative QTLs scored with scanone were used as cofactors for further analysis and composite interval mapping (CIM) forward selection procedure. CIM was performed using Haley-Knott regression and window size of 10 cM. 1000 permutations test was also run for CIM results.

RESULTS AND DISCUSSION

Weak to moderate correlations were observed for all three analysed traits (data not shown). Weak to moderate correlations for these traits were observed earlier (Dorsey – Redding et al., 1991; Li et al., 2009). In initial QTL analysis of grain oil content, a single QTL was identified at position 386.4 cM of chromosome 1 (Table 1). Putative QTL explained 5.69% of phenotypic variance and was highly significant at $P=<0.001$ in model testing. Comparable QTL was reported by Zhang et al. (2008), although positions on genetic maps used in the present study cannot be directly compared due to high resolution and greater size of IBM map (Falque et al., 2005) compared to an F_2 map used by Zhang et al. (2008). QTL for protein content was detected on chromosome 10, at 53 cM. Putative QTL accounted for 6.19% of phenotypic variance for the trait and it was highly significant at the $P=<0.001$ level (Table 1). Interestingly, Li et al. (2009) identified the same QTL at nearly the same position (53.2 cM), though this position was in bin 10.02, while on our map it is 10.01. Possibly, it was the same QTL we identified. Zhang et al. (2008) also mapped QTLs for grain protein content, but none of QTLs were detected in our study.

As expected, a larger number of smaller effect QTLs was identified for test weight which is a rather complex trait. QTL on chromosome 1, at position 405 cM (csu3) explained 4.13% of phenotypic variance and was highly significant ($P=<0.001$). A QTL was identified at chromosome 3, although with nonsignificant phenotypic effect. A QTL on chromosome 5 at position 307

cM (umc2298) had highly significant effect at $P=0.003$ and was accounted for explaining 2.67% of phenotypic variance (Table 1). All three QTLs found were previously identified by Quiang-Ding et al. (2011), although QTL on chromosome 3 was mapped at a different position. Another QTL for test weight on chromosome 10 at position 445.7 was accounted for 3.31% of phenotypic variance of trait and its effect was significant at $P=0.001$. This might be a QTL for kernel weight identified by Prado *et al*. (2014), mapped on bin 10.03 across all investigated populations. In our study of one biparental population, it was found at bin 10.06/10.07. QTLs for test weight were tested for additive effects with model y = Q1 + Q2 + Q3 + Q4 and "drop one" procedure. The percentage of phenotypic variance explained by their additive action was found to be 18.25%, which was highly significant at $P = <0.001$ (Table 1, last row) and exceeded sum of phenotypic variances attributed to each QTL individually.

Table 1. Results of initial QTL scan assuming single QTL per chromosome

Tablica 1. Rezultati inicijalnog QTL skena pod pretpostavkom jednog QTL-a po kromosomu

Trait *Svojstvo*	Chr *Krom.*	1.5 LOD interval	LOD	R^2 (%)[1]	*P* value *P vrijednost*	Flanking markers *Dodirni markeri*
Oil /*Ulje*	1	229.6 - 934.5	3.51	5.69	0.000	AY110052 - uaz130a
Protein / Proteini	10	0.0 - 260.5	3.83	6.19	0.000	mmp48a - mgs1
Test weight / *Hektolitarska masa*	1	391.8 - 445.5	3.97	4.13	0.000	umc2228 - AY110396
	3	203.4 - 828.9	3.69	0.21	0.404	mmp36 - mmp191
	5	229.0 - 332.7	6.34	2.67	0.003	bcd207b - mmp19
	10	442.2 - 450.8	4.05	3.31	0.001	agrr37c - AY110016
				18.25	0.000	

[1] Percentage (%) of the variance explained by QTL / *Udio variance objašnjen QTL-om*

In composite interval mapping (CIM) procedure, putative QTLs identified via the scanone function were set as cofactors to perform the procedure of forward selection of QTLs by multiple regression. Only one QTL of moderate effect for test weight was identified (Table 2) at position 307 cM (umc2298). Phenotypic variance explained by the QTL was 9.97 % and it was shown to be significant at $P=<0.001$ level. It was probably the same QTL identified by Quiang-Ding et al. (2011).

Generally, our QTL results should be interpreted with caution since our study was conducted in only one environment and using genetically very narrow population. The major impediment to the implementation of QTL analysis results is the lack of consistency in results due to large QTL x environment interactions (Liu et al., 2014), especially for important traits such as test weight, oil and protein contents for which the selection has already been done in both parental lines. Both parental lines of IBM population, B73 and Mo17, have undergone several cycles of selection and have probably had the highest commercial success in the history of maize seed business (Hallauer et al., 2010), so these results represent monitoring rather than implementation for use in marker assisted selection. Results of composite interval mapping can be seen in Figure 1 indicating other two possible QTLs on Chromosomes 1 and 4 with LOD scores higher than 3 which were not significant.

Table 2. Results of composite interval mapping for test weight

Tablica 2. Rezultati složenog intervalnoga mapiranja za hektolitarsku masu

Trait *Svojstvo*	Chr *Krom.*	1.5 LOD interval	LOD	R^2 (%)[1]	*P* value *P vrijednost*	Flanking markers *Dodirni markeri*
Test weight *Hektolitarska masa*	5	301.6 - 310.0	8.21	9.97	0.000	bnl8.33 - umc86b

[1] Percentage (%) of the variance explained by QTL / *Udio variance objašnjen QTL-om*

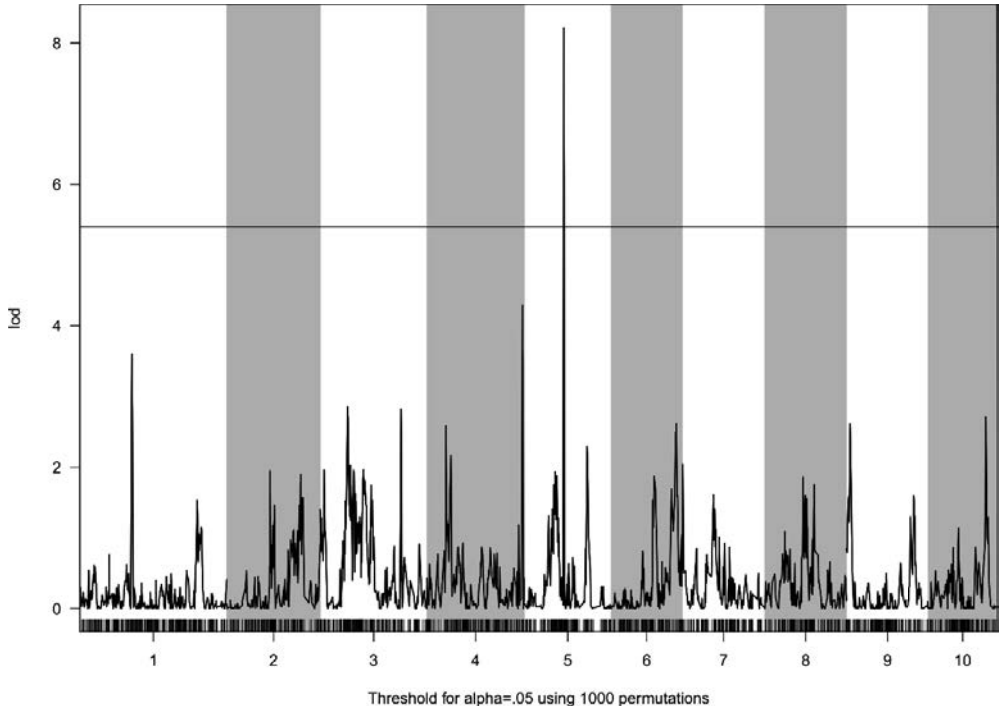

Figure 1. QTL for test weight scored with composite interval mapping procedure (the solid line denotes threshold at alpha = *0.05*)

Slika 1. QTL za hektolitarsku masu izračunat kompozitnim intervalnim kartiranjem (puna linija predstavlja prag značajnosti pri alfa = 0,05)

CONCLUSION

QTLs detected via simple QTL scan were not completely consistent with composite interval mapping. Only one QTL for test weight on chromosome 5 has been identified in both mapping procedures. All putative QTLs identified in this study can be useful guidelines in breeding for grain oil and protein contents and test weight when using marker assisted selection. Applying best linear unbiased predictions was proved to be a worthwhile tool for QTL mapping in maize biparental populations using genotyping-by-sequencing data.

ACKNOWLEDGEMENTS

This research was funded by the Croatian Science Foundation (the project no. 5707: „Genetics and physiology of multiple stress tolerance in maize“).

REFERENCES

1. Andorf, C.M., Lawrence, C.J., Harper, L.C., Schaffer, M.L., Campbell, D.A., Sen, T.Z. (2010): The Locus Lookup tool at MaizeGDB: identification of genomic regions in maize by integrating sequence information with physical and genetic maps. Bioinformatics, 26(3): 434-436.
   doi: http://doi.org/10.1093/bioinformatics/btp556
2. Balint-Kurti, P.J., Yang, J., Van Esbroeck, G., Jung, J., Smith, M. (2010): Use of a maize advanced intercross line for mapping of QTL for Northern Leaf Blight resistance and multiple disease resistance. Crop Science, 50: 458-466.
   doi: http://doi.org/10.2135/cropsci2009.02.0066
3. Bernardo, R. (1996): Best linear unbiased prediction of maize single-cross performance. Crop Science, 36: 50-56.
4. Bennetzen, J.L., Hake, S.C. (Eds.) (2009): Handbook of maize: genetics and genomics. Springer, NY, USA.
   doi: http://doi.org/10.1007/978-0-387-77863-1
5. Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S. (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633-2635.
6. Broman, K.W., Wu, H., Sen, S., Churchill, G.A. (2003): R/qtl: QTL mapping in experimental crosses. Bioinformatics, 19(7): 889-890.
   doi: http://doi.org/10.1093/bioinformatics/btg112
7. Dorsey-Redding, C., Hurburgh, C.R. Jr., Johnson, L.A., Fox, S.R. (1991): Relationships among maize quality factors. Cereal Chemistry, 68(6): 602-605.
8. Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E. (2011): A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS one, 6(5): e19379.
   doi: http://doi.org/0.1371/journal.pone.0019379
9. Endelman, J.B. (2011): Ridge regression and other kernels for genomic selection with R package rrBLUP. The Plant Genome, 4(3): 250-255.
   doi: http://doi.org/103835/plantgenome2011.08.0024

10. Falque, M., Decousset, L, Dervins, D., Jacob, A.M., Martinat, J.P., Raffoux, X., Ribière N., Ridel C., Samson D., Charcosset A., Murigneux A. (2005): Linkage mapping of 1454 new maize candidate gene Loci. Genetics, 170(4): 1957-1966.
doi: http://doi.org/10.1534/genetics.104.040204
11. Goldman, I.L., Rocheford, T.R., Dudley, J.W. (1993): Quantitative trait loci influencing protein and starch concentration in the Illinois Long Term Selection maize strains. Theoretical and Applied Genetics, 87: 217-224.
12. Goldman, I.L., Rocheford, T.R., Dudley, J.W. (1994): Molecular markers associated with maize kernel oil concentration in an Illinois High Protein X Illinois Low Protein cross. Crop Science, 34: 908-915.
13. Hallauer, A.R., Carena, M.J., Miranda Filho, J.B. (2010): Quantitative genetics in maize breeding. Springer – Verlag, NY, USA.
doi: http://doi.org/ 10.1007/978-1-4419-0766-0
14. Lee, M., Sharopova, N., Beavis, W.D., Grant, D., Katt, M., Blair, D., Hallauer, A. (2002): Expanding the genetic map of maize with the intermated B73 × Mo17 (IBM) population. Plant Molecular Biology, 48(5): 453-461.
doi: http://doi.org/10.1023/A:1014893521186.
15. Lee, K.M., Herrman, T.J., Rooney, L., Jackson, D.S., Lingenfelser, J., Rausch, K.D., McKinney, J., Iiams, C., Byrum, L., Hurburgh, C.R., Johnson, L.A., Fox, S.R. (2007): Corroborative study on maize quality, dry-milling and wet-milling properties of selected maize hybrids. Journal of Agricultural and Food Chemistry, 55(26): 10751-10763.
doi: http://doi.org/10.1021/jf071863f
16. Li, Y., Wang, Y., Wei, M., Li, X., Fu, J. (2009): QTL identification of grain protein concentration and its genetic correlation with starch concentration and grain weight using two populations in maize (Zea mays L.). Journal of Genetics, 88(1): 61-67.
doi: http://doi.org/10.1007/s12041-009-0008-z
17. Liu, Y., Wang, L., Sun, C., Zhang, Z., Zheng, Y., Qiu, F. (2014): Genetic analysis and major QTL detection for maize kernel size and weights in multi-environments. Theoretical and Applied Genetics, 127: 1019-1037.
doi: http://doi.org/10.1007/s00122-014-2276-0
18. Lian, L., Jacobson, A., Zhong, S., Bernardo, R. (2015): Genomewide prediction accuracy within 969 maize biparental populations. Crop Science, 54: 1514-1522.
doi: http://doi.org/10.2135/cropsci2013.12.0856
19. Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E. (2001): Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157: 1819-1829.
20. Nuss, E.T., Tanumihardjo, S.A. (2011): Quality Protein Maize for Africa: Closing the Protein Inadequacy Gap in Vulnerable Populations. Advances in Nutrition, 2: 217-214.
doi: http://doi.org/10.3945/an.110.000182
21. Prado, S.A., López, C.G., Senior, M.L., Borrás, A. (2014): The genetic architecture of maize (zea mays l.) kernel weight determination. G3, 4: 1611-1621.
doi: http://doi.org/10.1534/g3.114.013243
22. Prasanna, B.M., Vasal, S.K., Kassahun, B., Singh, N.N. (2001): Quality protein maize. Current Science, 81(01): 1308-1319.
23. Quiang-Ding, J., Ma, J.L., Zhang, C.R., Dong, H.F., Xi, Z.Y., Xia, Z.L., Wu, J.Y. (2011): QTL mapping for test weight by using F2:3 population in maize. Journal of Genetics, 90: 75-80.
doi: http://doi.org/10.1007/s12041-011-0036-3
24. R Core Team (2012): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.
25. Wassom, J.J., Wong, J.C., Martinez, E., King, J.J, DeBaene, J., Hotchkiss, J.R., Mikkilineni, V., Bohn, M.O., Rocheford, T.R. (2008): QTL Associated with Maize Kernel Oil, Protein, and Starch Concentrations; Kernel Mass; and Grain Yield in Illinois High Oil × B73 Backcross-Derived Lines. Crop Science, 48(1): 243-252.
doi: http://doi.org/10.2135/cropsci2007.04.0205
26. Zhang, J., Lu, X.Q., Song, X.F., Yan, J.B., Song, T.M., Dai, J.R., Rocheford, T., Li, J.S. (2008): Mapping quantitative trait loci for oil, starch, and protein concentrations in grain with high-oil maize by SSR markers. Euphytica, 162: 335-344.
doi: http://doi.org/10.1007/s10681-007-9500-9

QTL KARTIRANJE ZA SVOJSTVA KVALITETE ZRNA KOD TESTKRIŽANACA BIPARENTALNE POPULACIJE KUKURUZA KORIŠTENJEM PODATAKA GENOTIPIZACIJE SEKVENCIRANJEM

SAŽETAK

Proveli smo QTL kartiranje u test križancima IBMSyn4 populacije za tri svojstva kvalitete zrna: sadržaj ulja i proteina te hektolitarsku masu. 191 fenotipizirana i genotipizirana linija korištena je kao "training" populacija, dok je 85 linija, koje su bile samo genotipizirane, činilo „validacijsku" populaciju pri izračunavanju najboljih linearnih nepristranih predviđanja (BLUP). QTL analiza provedena je na ukupno 276 fenotipova. 92000 filtriranih SNP markera, dobivenih genotipizacijom sekvenciranjem (GBS), korišteno je za izračun BLUP-ova, dok je set od 2178 genetski kartiranih SSR markera korišten za QTL analizu. Pri jednostavnom QTL skeniranju, detektirali smo nekoliko QTL-ova slabijega fenotipskoga učinka: jedan za sadržaj ulja (kromosom 1), jedan za sadržaj proteina (kromosom 10) i četiri za hektolitarsku masu (kromosomi 1, 3, 5 i 10). QTL-ovi povezani s hektolitarskom masom imali su aditivno djelovanje te je njihovim združenim djelovanjem objašnjeno 18.25% fenotipske varijance. Samo jedan QTL za hektolitarsku masu na petome kromosomu bio je statistički značajan pri kompozitnom intervalnom kartiranju. Navedeni QTL podržava 9.97% fenotipske varijance. QTL-ovi detektirani u ovom istraživanju predstavljaju monitoring komercijalno najuspješnije elitne germplazme za svojstva kvalitete zrna.

Ključne riječi: najbolja linearna nepristrana predviđanja, IBM populacija, kukuruz, lokusi kvantitativnih svojstava, svojstva kvalitete zrna

(Received on 23 February 2017; accepted on 10 May 2017 – *Primljeno 23. veljače 2017.; prihvaćeno 10. svibnja 2017.*)