

RESEARCH ARTICLE

Identification of candidate markers associated with frying colour trait of potato

■ PUNAM GHARGE AND I.R.H.J. (HERMAN) VAN ECK

SUMMARY

The demand for potato products like chips and French fries is steadily increasing all over the world. Consumers and potato processing industries have become more stringent for quality along with higher demand of products. Hence, potato breeders are stimulated to develop new potato cultivars with more emphasis on quality traits. Potato breeding is mainly based on crossing two heterozygous parents or complementary parental clones and multi-year clonal selection to identify candidate cultivars with excellent quality. Hence, modern breeders like to use DNA/molecular markers to speed up the selection process by screening large numbers of genotypes at a time. To end up with a shortlist of candidate markers, three criteria; consistency, redundancy and multiple testing corrections were used for removal of false positive and redundant associations. In total, 62 marker-trait associations for frying colour were found to be informative after consistency over several sub traits with threshold level >3.3 in at least three sub traits. Finally, replacement analysis was performed to replace unmapped markers with mapped markers. 22 markers for frying colour trait were selected as a set of marker which could be used in Marker assisted breeding. It is clear that statistical approach provides a quick way of analyzing vast amounts of marker-trait associations to end up with short list of candidate markers. However, conformation is still needed to validate the markers.

Key Words : Potato, Candidate markers, Marker assisted selection, Marker-trait associations, Frying colour etc.

How to cite this article : Gharge, Punam and Van Eck, I.R.H.J. (Herman) (2016). Identification of candidate markers associated with frying colour trait of potato. *Internat. J. Plant Sci.*, 11 (2): 291-298, DOI: 10.15740/HAS/IJPS/11.2/291-298.

Article chronicle : Received : 13.03.2016; Revised : 10.05.2016; Accepted : 16.06.2016

Potato is an autotetraploid ($2n=4x=48$), displaying tetrasomic inheritance (Bradshaw *et al.*, 2008). As a result, genetic analysis is more complex as compared to diploid species and potato breeding is

genetically not easy. The selection of improved cultivars with DNA-based markers will become more easy with the knowledge of genes that control the inheritance of agronomically important traits (Li *et al.*, 2008). In commercial breeding companies, time to market is a key factor. Depending on plant type and traditional breeding methods, it takes about 16-18 years from pre-breeding to commercializing a new variety and 12-15 years from original cross to commercializing a new variety (Dejong, 1983). Therefore, modern breeders like to use DNA/molecular markers to speed up the selection process by screening large numbers of genotypes at a time. Since,

MEMBERS OF THE RESEARCH FORUM

Author to be contacted :

PUNAM GHARGE, Department of Agricultural Botany, College of Agriculture (M.P.K.V.) BARAMATI (M.S.) INDIA

Address of the Co-authors:

I.R. H.J. (HERMAN) VAN ECK, Laboratory of Plant Breeding, Plant Sciences Group, Wageningen University and Research Center, NETHERLANDS

DNA can be harvested early in the year from crop. Genotypes can be assessed for frying quality before harvest based on DNA markers, which saves a lot of time because the alternative is that potatoes have to be stored for 2 to 6 months in order to determine frying quality.

At present, there are numerous cost effective methods to generate huge amount of molecular marker data due to development in DNA technology and genetics. As a consequence, constructions of dense linkage maps or genome wide marker scans are cheaper per data point. Therefore, breeders look forward to identify candidate markers for interesting traits and thereby adopt marker assisted breeding (MAS) in introgression breeding, progenitor development and also for speeding up the selection process during the breeding programme. Apart from the application of marker assisted selection, the identification of candidate genes and characterization of the allelic variation for genes involved in interesting traits is the basic prerequisite for enhancement of potato varietal development (Visser *et al.*, 2009). In recent years, association mapping is the most widely used technique to identify marker-trait associations in various crops, for example in rice for agronomic traits (Zhang *et al.*, 2005). Identification of marker-trait associations is the first step towards the identification of candidate genes and the study of allelic variation associated with the traits of interest.

Quality is also one of the important aspects related to a crop as it determines the value in terms of price and consumer acceptance of final product. Quality parameters are different according to the end use of products. Potatoes are used for direct consumption, processed food products like chips and French fries, for industrial starch production and for animal feed (Gebhardt *et al.*, 2007). During the breeding of new potato cultivars, the breeding goals differ for these three main areas. According to these goals, there are different agronomic and quality characteristics that are considered in the selection process during breeding. In this section, more effort will be put into the description of the background of important traits to be analysed in this research for candidate marker identification. Moreover, an attempt is made to describe other crucial traits which also express the quality of tuber and are interlinked with our interested traits *i.e.*, frying colour.

Frying colour of potato chips or French fries mainly depends on the amount of reducing sugars (glucose and

fructose) in the tuber. There are various factors which influence reducing sugar contents and frying colour such as: genotype of potato (Coffin *et al.*, 1987), duration of potatoes in cold storage (Stevenson and Cunningham, 1961), minimization of storage temperature (Gould *et al.*, 1979) and tuber maturity at harvest (Hope *et al.*, 1960). For the production of potato chips and French fries, high frying temperature is used which activates the non-enzymatic Maillard reaction between free aldehyde groups of reducing sugars (glucose and fructose) and free α -amino groups of amino acids and proteins (Li *et al.*, 2005). Due to this Maillard reaction, acrylamide is formed which causes unattractive chips with a bitter flavour. Darker frying colour of chips or fries is positively correlated with higher reducing sugar content. Storage of tubers at low temperature below 4-8 °C, delays the sprouting and leads to a breakdown of starch into reducing sugars (cold sweetening) in response to cold stress (Burton, 1969). Accumulation of sugars starts when there is an imbalance between starch degradation, synthesis and respiration of carbohydrates. Breeding for improved frying colour can be effective because frying colour is found to be a heritable trait. Heritability values for chip colour are reasonably high: 0.81 to 0.87 following cold storage (Cunningham and Stevenson, 1963) and 0.47 to 0.63 for sugar contents (Pereira *et al.*, 1994). Consequently, knowing the genes responsible for reducing sugar contents and frying colour and developing markers near the gene possibly will make potato breeding more efficient.

MATERIAL AND METHODS

Plant material :

Five phenotypic datasets (FT2006, FT2008, FT2009, MYML and a joint dataset FT2008_2009) recorded on 430 potato genotypes were used for identification of markers. These 430 cultivars were a representative set of the worldwide available potato germplasm and collected from five breeding companies and several gene banks. The core set contained 221 tetraploid potato cultivars and progenitor clones. The remaining 209 genotypes comprising parents of SH x RH mapping population, 17 extra tetraploid potato cultivars and 190 advanced breeding clones (approximately 40 of each company: Agrico research, Averis seeds, C Meijer, HZPC Research and Van Rijn). All five datasets contained different subsets of those 430 potato genotypes. All 430 genotypes were genotyped with AFLP markers.

Additionally a 384 SNP set was scored on the core set of 220 potato cultivars by (D'hoop *et al.*, 2010). In this research, the both marker sets were used.

Marker-trait associations :

For the removal of false positive and redundant associations, 99,225 marker-trait associations were analyzed. These associations were resulted from 27 phenotypic sub traits (27 frying colour) and 3675 (3364 AFLP + 311 SNP) markers [$27 \times 3675 = 99,225$ data points]. The marker-trait association data was used from an analysis performed by D'Hhoop (2009). The *P*-values obtained for marker-trait associations were very small (as small as 10^{-13}) and, therefore, were transformed into $-\log_{10} p$ -values for better visibility of the significance level. Frying colour were measured several times per dataset. The combinations of year and time points are hereafter referred to as sub traits.

Removal of false positive and redundant associations :

In order to create a short list of candidate markers, marker-trait association were evaluated for removal of possible false positive and redundant associations. False positive associations create ambiguity to identify real associations as well as it increases time and cost of analysis. Therefore, three filtering steps were used to remove possible false positive associations of markers.

Consistency :

A marker is considered to be a false positive when it is significant (*p*-value < 0.05) in only one dataset. Consistency was assumed when a marker was significant in at least two dataset with $-\log_{10} p \geq 1.301$

Redundancy :

Redundancy refers to pairwise relationships between all pairs of markers in a given predictor dataset (Ooi *et al.*, 2006). Redundancy was assumed when two markers displayed an identical band pattern, mapped to the same position in mapping population and linked to the same gene in a QTL governing trait of interest. In that case, from both selected markers, one good marker will be used for analysis instead of both. Consequently, removal of redundant marker resulted in smaller number of markers, which reduces cost and also save time for further analysis. The redundancy of marker-trait associations were analysed by calculating marker-marker correlations. Intensity values of AFLP and theta

values of SNP markers were used to calculate Pearson's correlation co-efficient (*r*) between a pair of markers (M-M). For this purpose, the bicorrelate function in Genstat (software) was used. However, high correlations do not always be a sign of true linkage between markers because sometimes high correlations between marker pair resulted from markers that were mapped on different chromosomes. Therefore, an additional criterion was applied to remove redundant markers *i.e.*, map position of markers.

Multiple testing correction :

A multiple testing correction was performed on remaining candidate marker-trait associations to reduce the number of false positive significant associations. The threshold for significance considering multiple testing (0.0005) was obtained by Bonferroni type of multiple testing correction threshold. Selections of informative markers were performed by using consistency and multiple testing analysis together to reduce the number of marker and retain only highly consistent markers across sub traits. Finally Replacement of unmapped selected markers with highly correlated mapped markers based on a marker which has larger fragment size, a map position, consistent $-\log_{10} p$ -value and high correlation with selected marker. To identify highly correlated and mapped makers with these selected markers (top 30 markers), all significant (at $r=0.05$) markers resulted from filtering false positive associations were paired. A table was produced by using marker-marker correlations between all significant ($p < 0.05$) and selected markers. Map positions for these marker pairs were also given. If the marker pairs had a Pearson correlation co-efficient >0.3 they were selected for further analysis. If a marker has a small fragment size, then it is present at the bottom of the gel. Hence, cutting of the band is very complex. In addition, such small sized fragments have less sequence information. The $-\log_{10} p$ -value from ChipCol_Oct8c_2008 and FrCol_Dec8c_2008 were used for better judgement of these selected markers because these two sub trait contain more significant markers. Only unmapped markers were replaced with mapped ones, which have good fragment size, good $-\log_{10} p$ -value, high correlation (*r*).

RESULTS AND DISCUSSION

The present study was carried out to identify candidate marker-trait association for frying colour of

potato. Therefore, several filtering steps were used to remove false positive and redundant marker-trait associations to come up with a short list of candidate markers. In addition, finally selected markers were analyzed for a possible replacement with better markers which have a large fragment size, good *p*-value, map position, and high marker-marker correlation.

Removal of false positive associations :

From our finding the number of markers was decreased after each filtering step. A too stringent significance level for marker-trait associations might obscure false positives association. The first filtering step used to remove false positive marker-trait associations was consistency of markers across the datasets. The consistency criterion did not consider the markers significant at a *p*-value of 0.05 if it occurred in only one dataset. Less than half of the total markers were observed as significant in at least two datasets (consistent) with a *p*-value <0.05. The frying colour sub trait February 8C has the highest number of consistent markers while the frying colour sub trait April 8C has less consistent markers. The number of significant consistent markers per trait ranged from 666 to 1400 (Table 1).

Another criterion used to remove redundant marker-trait associations is redundancy. Redundancy involves two criterions: one is correlation between markers, so that redundant markers mapped to the same haplotype can be removed. The other is map position to remove redundant markers mapped to different haplotypes. We assumed that true linkage always give high correlation. Literature studies reported that the linkage yielded high correlation. Therefore, removal of one marker resulted in no loss of information (Cho and Dupuis, 2009). A high positive correlation between two markers means both markers present on same haplotype

(coupling phase) and they are not segregating in another parent. Likewise, negative correlation between markers indicated that the marker-QTL associations present on different haplotypes (repulsion phase). If both highly correlated markers were found to be located within 1cM region of QTL of interest then removal of one of a set of linked markers would cause no loss of information. The correlation analysis yielded several highly correlated markers which were mapped at different chromosomes. These were called ‘mismatches’ by us. These mismatches suggested that something went wrong, either in the map position of one of the markers or in the correlation calculation. Therefore, map position information was used to detect first a possible mismatch of a marker pair. The first mismatch of a marker pair was observed at a Pearson’s correlation co-efficient (r) = 0.85. This analysis was also resulted in identification of conflicting map position for 21 markers with high correlation co-efficient value. It is concluded that higher correlation was not always the result of true linkage. This conclusion was supported by studies of Nielsen *et al.* (2008) on how well linkage disequilibrium between markers predicts redundant associations. He also concluded that, the high linkage disequilibrium between markers does not indicate that the one marker is a direct substitute for another marker.

The third filtering step to remove false positive marker-trait associations is a multiple testing correction threshold. Our study ended with 16-220 markers per trait after applying the multiple testing correction thresholds. The highest number of markers was retained for the trait frying colour February 8c. This may be due to frying colour has enormous number of significant marker-trait associations (Table 1). Tatonetti *et al.* (2010) used multiple testing correction to score the candidate gene from association studies for warfarin doses. He

Table 1 : Overview of markers retained after Consistency (at least in two datasets with $-\log_{10}p \geq 1.301$) and multiple testing correction threshold analysis for frying color traits.

Filtering criteria	Consistent markers (at least in 2 datasets with $p < 0.05$ or $-\log_{10} p \geq 1.3$)			Multiple testing correction (at least in 2 dataset with $p < 0.0005$ or $-\log_{10} p \geq 3.3$)		
	AFLP	SNP	Total	AFLP	SNP	Total
Frying colour						
Oct_FrCol_8c	605	69	674	7	9	16
Dec_FrCol_8c	908	80	988	38	8	46
Feb_FrCol_8c	1291	109	1400	160	10	170
Apr_FrCol_8c	629	37	666	22	1	23
May_FrCol_8c	674	64	738	34	2	36
May_FrCol_4c	755	65	820	19	1	36

invertase gene *invGE* and *invGF* which colocalizes with cold induced sweetening QTL *sug9*. This QTL present in between 3–8cM distance which is not a centromeric region. Another thing is that *InvGE/GF* gene is a berry specific gene and thus, cannot influence chip quality (Anne-marie Wolters, personal communication). Therefore, there may be other genes in centromeric region which might be associated with chip quality.

Final selection of markers :

The markers selected from consistency across sub trait analysis were analyzed for replacement by using several criterion like map position, high M-M correlation, good *p*-values, larger fragment size. Because these criteria saves lot of our further work like mapped marker easily find back, higher fragment size is easy to excise from gel and *p*-values provides assurance about quality of marker. Short size of a marker fragment resulted in short DNA sequence which can affect designing of PCR primers. Another reason to select large fragment size markers is less chances of multiple fragment extraction due to clear separation of band at the top of gel. In total, 22 markers are finally selected for frying colour. For example, marker E32_M51_086_91_GST was replaced with E32_M54_430_74_GST because marker E32_M54_430_74_GST was highly correlated ($r = 0.62$) and has a map position on chromosome 9 at 16.6 cM. The fragment size of 430 nucleotide of E32_M54_430_74_GST marker is much bigger as compared to marker E32_M51_086_91_GST. Hence, the marker fragment present on the top level on a gel which is easy to excise due to clear separation between different allelic bands of a marker. It reduces the chances of multiple fragment extraction during conversion of marker into simple easy to use marker. Another reason to select marker E32_M54_430_74_GST instead of E32_M51_086_91_GST was more significant association of E32_M54_430_74_GST with frying colour. Some markers were not replaced with other markers because of their good fragment size, allele frequency and good *p*-values. For example, Marker E36_M62_165_10_GST was not replaced with another markers because marker E36_M62_165_10_GST have map position 53.9cM on chromosome 7, fragment size of 165 nucleotides and good *p*-values (Table 3).

Mostly AFLP markers were replaced with SNP markers although SNP markers were unmapped. It might be due to the SNP markers have 50 nucleotides known

sequence. During blasting, a very few matching segments with SNP marker sequence from genome sequence will be obtained as compared to AFLP marker. Therefore, it is easy to find back SNP marker on the potato genome sequence. For AFLP markers, only the primer sequences with 3 additional nucleotides were known. Therefore, too many matching segments of DNA sequence will be obtained from blasting to the potato genome sequence. Furthermore, AFLP markers are less suitable for marker-assisted selection, allele frequencies studies (Brugmans *et al.*, 2003). Also, AFLP markers are too expensive and laborious.

Conclusion :

A statistical approach provided a quick way of analyzing marker-trait associations for candidate markers associated with frying colour traits. The consistency across sub traits combined with multiple testing correction threshold analysis is helped to identify markers with consistent associations of candidate markers for frying colour trait. As our result is only based on statistical analysis. Therefore, for assurance, there is need to sequence these markers and blast the sequencing result against potato genome to obtain the correct map position.

Acknowledgement :

Author are thankful to Dr. Bjorn Hoop, Peter Vos, Dr. Chris Maliepaard, Dr. Herman van Eck Professors, Department of Plant Sciences, Wageningen University and Research Center (WUR), The Netherlands, for their valuable guidance, advice and motivation for this work.

REFERENCES

- Bradshaw, J., Hackett, C., Pande, B., Waugh, R. and Bryan, G. (2008). Qtl mapping of yield, agronomic and quality traits in tetraploid potato (*solanum tuberosum subsp tuberosum*). *TAG Theoretical & Appl. Genet.*, **116**(2): 193-211.
- Brugmans, B., van der Hulst, R.G.M., Visser, R.G.F., Lindhout, P. and van Eck, H.J. (2003). A new and versatile method for the successful conversion of *afp*TM markers into simple single locus markers. *Nucleic Acids Res.*, **31**(10): e55.
- Burton, W. (1969). The sugar balance in some british potato varieties during storage. Ii. The effects of tuber age, previous storage temperature, and intermittent refrigeration upon low-temperature sweetening. *Potato Res.*, **12**(2): 81-95.

- Cho, K. and Dupuis, J. (2009). Handling linkage disequilibrium in qualitative trait linkage analysis using dense snps: A two-step strategy. *BMC Genet.*, **10**(1): 44.
- Coffin, R., Yada, R., Parkin, K., Grodzinski, B. and Stanley, D. (1987). Effect of low temperature storage on sugar concentrations and chip colour of certain processing potato cultivars and selections. *J. Food Sci.*, **52**(3): 639-645.
- Cunningham, C. and Stevenson, F. (1963) Inheritance of factors affecting potato chip colour and their association with specific gravity. *American J. Potato Res.*, **40**(8): 253-265.
- D'hoop, B.B. (2009). Marker-trait associations in potato. Association mapping in tetraploid potato, Ph.D. Thesis, Wageningen University, The Netherlands, 71-87pp.
- D'hoop, B.B., Paulo, M., Kowitwanich, K., Sengers, M., Visser, R., van Eck, H. and van Eeuwijk, F. (2010). Population structure and linkage disequilibrium unravelled in tetraploid potato. *TAG Theoretical & Appl. Genet.*, **121**(6): 1151-1170.
- Dejong, H. (1983). Inheritance of sensitivity to the herbicide metribuzin in cultivated diploid potatoes. *Euphytica*, **32**(1): 41-48.
- Gebhardt, C., Li, L., Pajeroska-Mukthar, K., Achenbach, U., Sattarzadeh, A., Bormann, C., Ilarionova, E. and Ballvora, A. (2007). Candidate gene approach to identify genes underlying quantitative traits and develop diagnostic markers in potato.
- Gould, W., Hair, B. and Baroudi, A. (1979). Evaluation of potato cultivars before and after storage regimes for chipping. *American J. Potato Res.*, **56**(3): 133-144.
- Hope, G., MacKay, D. and Townsend, L. (1960). The effect of harvest date and rate of nitrogen fertilization on the maturity, yield and chipping quality of potatoes. *American J. Potato Res.*, **37**(1): 28-33.
- Li, L., Paulo, M.J., Strahwald, J., Lübeck, J., Hofferbert, H.-R., Tacke, E., Junghans, H., Wunder, J., Draffehn, A., van Eeuwijk, F. and Gebhardt, C. (2008). Natural DNA variation at candidate loci is associated with potato chip colour, tuber starch content, yield and starch yield. *TAG Theoretical & Appl. Genet.*, **116**(8): 1167-1181.
- Li, L., Strahwald, J., Hofferbert, H.R., Lubeck, J., Tacke, E., Junghans, H., Wunder, J. and Gebhardt, C. (2005). DNA variation at the invertase locus *invge/gf* is associated with tuber quality traits in populations of potato breeding clones. *Genet.*, **170**(2): 813-821.
- Menendez, C.M., Ritter, E., Schafer-Pregl, R., Walkemeier, B., Kalde, A., Salamini, F. and Gebhardt, C. (2002). Cold sweetening in diploid potato: Mapping quantitative trait loci and candidate genes. *Genetics*, **162**(3): 1423-1434.
- Nielsen, D.M., Suchindran, S. and Smith, C.P. (2008). Does strong linkage disequilibrium guarantee redundant association results? *Genet. Epidemiol.*, **32**(6): 546-552.
- Ooi, C., Chetty, M. and Teng, S. (2006). Differential prioritization between relevance and redundancy in correlation-based feature selection techniques for multiclass gene expression data. *BMC Bioinform.*, **7**(1): 320.
- Pereira, A.D.S., Tai, G.C.C., Yada, R.Y., Tarn, T.R., Souza-Machado, V. and Coffin, R.H. (1994). Effect of selection for chip colour on some economic traits of potatoes. *Plant Breed.*, **113**(4): 312-317.
- Stevenson, F. and Cunningham, C. (1961). Chip colour in relation to potato storage. *American J. Potato Res.*, **38**(4): 105-113.
- Tatonetti, N., Dudley, J., Sagreiya, H., Butte, A. and Altman, R. (2010). An integrative method for scoring candidate genes from association studies: Application to warfarin dosing. *BMC Bioinform.*, **11**(Suppl 9): S9.
- Visser, R., Bachem, C., de Boer, J., Bryan, G., Chakrabati, S., Feingold, S., Gromadka, R., van Ham, R., Huang, S., Jacobs, J., Kuznetsov, B., de Melo, P., Milbourne, D., Orjeda, G., Sagredo, B. and Tang, X. (2009). Sequencing the potato genome: Outline and first results to come from the elucidation of the sequence of the world's third most important food crop. *American J. Potato Res.*, **86**(6): 417-429.
- Zhang, N., Xu, Y., Akash, M., McCouch, S. and Oard, J.H. (2005). Identification of candidate markers associated with agronomic traits in rice using discriminant analysis. *TAG Theoretical & Appl. Genetics*, **110**(4): 721-729.

11th
Year
★★★★★ of Excellence ★★★★★