



Prediction of QTL genotypes and trait phenotypes using FlexQTL™: A pedigree-based analysis approach

Sujeet Verma*; Vance M Whitaker

IFAS Gulf Coast Research and Education Center, University of Florida, USA

***Corresponding Author(s): Sujeet Verma,**

IFAS Gulf Coast Research and Education Center, University of Florida, Wimauma, Florida 33598 USA

Email: sujeet.verma@ufl.edu

Abstract

Predicting phenotypes and QTL genotypes is of great value to breeding programs, especially those that make hundreds of crosses every year. Determining parents to make cross combinations is a complex process, and a breeder will use all available information to make that determination. Likewise, any available predictions of performance or of QTL genotypes for seedlings can be used in selection. Several well-established statistical methodologies are utilized to predict phenotypes and breeding values using genome-wide markers [1-3], but predicting unknown QTL genotypes and phenotypes based on specific QTL is rarely reported. However, a Pedigree-Based Analysis (PBA) software called FlexQTL™ has the statistical capability to predict QTL genotypes and unknown trait phenotypes. Although the theoretical foundation of this approach was laid years ago [4,5], no application has been reported in the literature to our knowledge. Rather, FlexQTL™ has primarily been used for QTL discovery and validation in multi-parental, pedigree-connected populations. Yet PBA can be used to predict QTL genotypes (QQ , Qq , qq) and phenotypes for individuals having marker data only.

The goal of predicting unknown phenotypes is the same for established genome-wide selection approaches and for the FlexQTL™ approach, but pedigree connectivity is at the core of the analysis, and locus-specific markers (as opposed to genome-wide markers) are utilized. In both approaches, datasets are divided into *training* populations with phenotypic and marker data and *test* populations with marker data only. Of course, a high degree of relatedness between training and test population is essential using either approach.

Received: Mar 14, 2018

Accepted: May 11, 2018

Published Online: May 17, 2018

Journal: Journal of Plant Biology and Crop Research

Publisher: MedDocs Publishers LLC

Online edition: <http://meddocsonline.org/>

Copyright: © Verma S (2018). *This Article is distributed under the terms of Creative Commons Attribution 4.0 International License*

Overview of an example breeding program

In a typical strawberry breeding program, the pool of elite parents used to make crosses is updated rapidly according to industry and environmental needs. Because of its octoploid genome, determining the genetic architecture underlying traits in cultivated strawberry is challenging. In addition, strawberry

varieties are mostly asexually propagated, and maintaining a strawberry clone from first year field trial to its release is expensive. Every year in the University of Florida (UF) strawberry breeding program, about a hundred crosses are made in the anticipation of superior fruit quality and better resistance to diseases. It takes a year to obtain performance data on progeny and make informed crosses for the next year. If the genetic po-



Cite this article: Verma S, Whitaker VM. Prediction of QTL genotypes and trait phenotypes using FlexQTL™: A pedigree-based analysis approach. J Plant Biol Crop Res. 2018; 2: 1006.

tential of future progeny can be estimated prior to field evaluations, increased genetic gains can be realized and resources can be saved.

An older breeding program like that at UF can utilize years of accumulated phenotypic data and many generations of pedigree information. Most of the elite parents descend from a few common progenitors in the UF strawberry breeding program, which helps magnify pedigree relatedness between training and test populations. In this article, we refer to four training populations, T1/2013 (Trial 1: year 2013-14), T2/2013 (Trial 2: year 2013-14), T1/2014 (Trial 1: year 2014-15) and T2/2014 (Trial 2: year 2014-15) [3], and a test population: T2/2015 (Trial 2: year 2014-15). In general, T1 trials are of unselected seedlings arising from complex mating designs, and T2 trials are collections of advanced selections that represent the parent pool. Here we utilize this data to demonstrate the ability of FlexQTL™ to predict unknown QTL genotypes and phenotypes for a moderate-effect Soluble Solids Content (SSC) locus on Linkage Group (LG) 6A (unpublished).

Objectives of this editorial are to (1) describe the ability of FlexQTL™ to predict unknown QTL genotypes and phenotypes using the SSC locus on LG 6A as an example, and (2) demonstrate predictive ability by correlating predicted and observed SSC data for the test population.

Genetic control of a trait and prediction methodology

The genetic control or architecture of a trait may range from single major locus to a couple of moderate effect loci to many small-effect loci, or any combination thereof. Before predicting unknown phenotypes of individuals, knowledge of the genetic control of a trait is vital. A genome-wide prediction approach is usually effective for a trait control by many small-effect loci as compared to Quantitative Trait Loci (QTL) based Marker-Assisted Selection (MAS) approach with known large-effect marker-trait associations [6]. Several well-established statistical methodologies have been recommended for making genome-wide predictions including Genomic Best Linear Unbiased Prediction (GBLUP) and Bayesian methods [3,7]. However, when the genetic architecture of a trait includes at least one identifiable locus, it is valuable to have predictions of QTL alleles, their combinations (genotypes) and their patterns of segregation. In our experience, such a QTL-based approach is particularly effective for a trait controlled primarily by single locus or a couple of loci explaining half or more of the phenotypic variation for the trait.

The PBA approach using FlexQTL™ software not only detects QTLs but also assigns predicted QTL alleles (*QQ*, *Qq*, *qq*) based on molecular marker allele frequency, Identity By Descent (IBD) probabilities, and available phenotypic data accounting for all known pedigree relationships including immediate parents, grandparents, offspring, siblings and more distant ancestors [4,5,8-10] (Figure 1). FlexQTL™ also helps visualize segregation of QTL genotypes via pedigrees using PEDI map software. Where phenotypic data is missing but marker data is available, phenotypes can be predicted based on the predicted QTL genotypes.

Prediction with FlexQTL™

The advantages of the PBA approach, both for QTL detection and QTL allele and phenotype prediction, are most apparent in complex population structures [11-19]. Bi-parental experimental designs are inefficient for a large breeding program like the UF strawberry breeding program where the number of parents

in a crossing cycle is large and constantly changing. Most QTL analysis methods assume that parents are unrelated, but in reality, most parents in breeding populations are related. FlexQTL™ efficiently utilizes pedigree connections and establishes relatedness among parents in the most complex of population structures [5]. The ability of FlexQTL™ to identify inherited relatedness and re-evaluate pedigree relationships using IBD probability matrices offers superior statistical ability to estimate functional genotypes at a locus and their effects on a trait.

FlexQTL™ implements phasing of marker alleles based on founder alleles using Linkage Disequilibrium (LD). The tracing of segregating marker alleles from founders to connected parents and progeny generates phased haplotype information for each individual. In addition, where marker information is missing, LD-based estimation is utilized to impute marker alleles [20]. FlexQTL™ efficiently conducts linkage phasing between QTL genotypes (*QQ*, *Qq*, *qq*) and marker genotypes (A/B or A/T/G/C) over diverse genetic backgrounds, providing vital information on the segregation of functional QTL genotypes throughout a breeding population. QTL alleles homozygous for positive effect are represented as *QQ* [+ +], for heterozygous effects represented as *Qq* [+ -], and homozygous for negative effects represented as *qq* [- -]. If an individual's alleles are not represented sufficiently in the training population, the prediction will be poor. Thus, when separating datasets into training and test populations, relatedness and representation of alleles between the two sets must be carefully considered.

A PBA prediction example

Soluble solids content is partly controlled by a moderate-effect QTL on LG 6A in the UF breeding program, explaining around 8-15% of phenotypic variation. T1/2013, T2/2013, T1/2014, and T2/2014 trials conducted in two different years are considered together as a training population (phenotypic and marker data included in analysis), and T2/2015 is considered as the test population (marker data only included in analysis). The training population includes more than 200 full-sib families with over 1,500 individuals in total. The parents and/or grandparents of the test population individuals are pedigree-connected to the training population. This allows dynamic multi-directional flow of allele information between training and test populations. An example of multi-directional flow of allele information is presented in Figure 1.

The test population T2/2015 is comprised of approximately 200 selections, the parents of which were represented by numerous half-sibs in the training population. FlexQTL™ simulations were implemented according to [12] using 13 SNP markers spanning the QTL region, and SSC QTL genotypes and phenotypes predicted. In order to estimate Predictive Ability (PA), a Pearson correlation was conducted between observed and predicted SSC data. A positive ($r=0.21$) and significant ($p=0.0027$) correlation was observed (Figure 2).

In addition, Fisher's Least Significant Difference (LSD) test was conducted to separate observed mean values of each of the predicted QTL genotype classes, and a significant separation ($p<0.05$) between phenotypic means of each of the QTL genotype classes was observed (Figure 2) [21]. A predictive ability (PA) of 0.21 in genome-wide depends on the context such as population size and diversity, trait heritability and the goal of the study. Given the low to moderate heritability of this trait, we consider 0.21 high enough to warrant further study.

We believe that as the UF strawberry program accumulates more genotypic and phenotypic data over the years, training populations will grow in size and scope and prediction of unknown QTL genotypes and phenotypes will be more precise. These results hold great value to the UF strawberry breeding program as UF no longer conducts stage-1 replicated trials [3] and information on genetic potential of parents will inform cross combinations. This method is being explored for other traits as well.

Figures

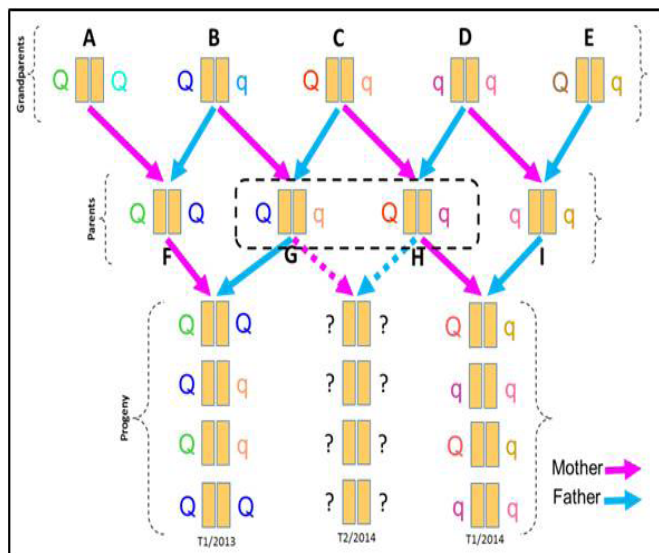


Figure 1: A hypothetical representation of hierarchical transmission of QTL genotypes (QQ, Qq, qq) via small pedigrees of five grandparents (top panel; A to E), three crosses among four parents (middle panel; F to I), and their progeny (bottom panel). Dotted lines indicate progeny from a test population. A black dotted box in the middle panel highlights parents “G” and “H” used to generate the test population and their QTL genotypes. Pink and blue arrow represent maternal and paternal sources, respectively. Q and q are QTL alleles and different colors of QTL alleles help trace them via pedigrees.

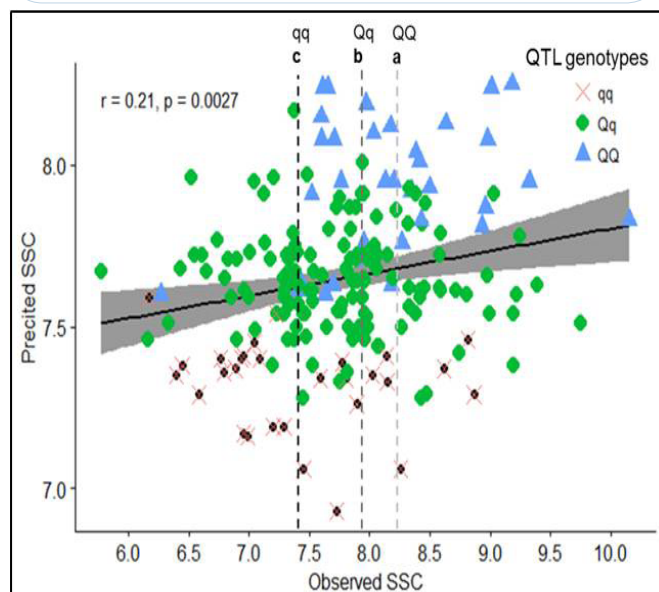


Figure 2: Pearson correlation between predicted and observed soluble solids content (SSC) for the test population, T2/2015. Dashed lines represent observed means of individuals grouped by predicted QTL genotype, and letters represent separations of observed means.

Conclusion

Until now the PBA approach using FlexQTL™ has been employed only for the detection of QTLs and QTL allele analysis for various traits using full phenotypic and marker data [10,22-30]. Here we have explored the ability of FlexQTL™ to predict QTL genotypes and unknown phenotypes where only marker data is available. Preliminary results indicate that it can be a valuable tool to make DNA-informed breeding decisions.

Based on our experience, several points should be considered:

1) Traditional genome-wide prediction approaches do not usually isolate the effects of discrete QTL or predict QTL genotypes; however, QTL genotype information from FlexQTL™ can be used as fixed effect for better genome wide predictions using traditional genomic selection models.

2) Predicted QTL genotypes can be utilized, not only for parent selection to achieve the highest progeny mean, but also to predict phenotypic variances of crosses based on predicted QTL genotype ratios of progenies. If effective, prediction of cross variance would be a unique contribution to the field of genomic prediction.

3) An important point to mention is that FlexQTL™ requires a genetic or physical map, which is a limitation for some crop species as compared to GBLUP or Bayes B methods, which do not require marker positions.

4) This methodology will be most advantageous for QTL with moderate effects as opposed to major loci explaining most of the genetic variance for a trait. In the case of a major locus, a single marker in the QTL region is likely to explain the phenotypic variance. In the case of a moderate-effect QTL such as the SSC QTL in this example, FlexQTL™ accounts for the segregation of the 13 SNPs in the QTL region across pedigrees, making QTL predictions that would be almost impossible manually.

5) Continuing with this line of reasoning, the value of FlexQTL™ should be highest in cases of multiple discrete QTL controlling a trait, which can be simultaneously predicted using this software.

In summary, FlexQTL™ provides exciting new possibilities for predictive breeding applications that should be especially valuable for traits controlled by one or more moderate-effect QTL.

References

- Bernardo R, Yu J. Prospects for genome wide selection for quantitative traits in maize. *Crop Sci.* 2007; 47: 1082.
- Kumar S, Chagné D, Bink MCAM, Volz RK, Whitworth C, Carlisle C. Genomic selection for fruit quality traits in apple (*Malus domestica* Borkh.). *PLoS ONE.* 2012; 7: e36674.
- Gezan SA, Osorio LF, Verma S, Whitaker VM. An experimental validation of genomic selection in octoploid strawberry. *Hortic Res.* 2017; 4: 16070.
- Bink M, Uimari P, Sillanpää M, Janss L, Jansen R. Multiple QTL mapping in related plant populations via a pedigree-analysis approach. *Theor Appl Genet.* 2002; 104: 751–762.
- Bink MCAM, Totir LR, Braak CJF ter, Winkler CR, Boer MP, Smith OS. QTL linkage analysis of connected populations using ancestral marker and pedigree information. *Theor Appl Genet.* 2012; 124: 1097–1113.

6. Jannink JL, Lorenz AJ, Iwata H. Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomics*. 2010; 9: 166–177.
7. Habier D, Fernando RL, Kizilkaya K, Garrick D. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics*. 2011; 12: 186.
8. Bink MCAM, Anderson AD, Weg WE van de, Thompson EA. Comparison of marker-based pairwise relatedness estimators on a pedigreed plant population. *Theor Appl Genet*. 2008; 117: 843–855.
9. Bink MCAM, Boer MP, ter Braak CJF, Jansen J, Voorrips RE, van de Weg WE. Bayesian analysis of complex traits in pedigreed plant populations. *Euphytica*. 2008a; 161: 85–96.
10. Bink MCAM, Jansen J, Madduri M, Voorrips RE, Durel CE, Kouassi AB, et al. Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. *Theor Appl Genet*. 2014; 127: 1073-1090.
11. Van de Weg E, Di Guardo M, Jansch N, Socquet-Juglard D, Costa F, Baumgartner I, et al. Epistatic fire blight resistance QTL alleles in the apple cultivar 'Enterprise' and selection X-6398 discovered and characterized through pedigree-informed analysis. *Mol Breeding*. 2018; 38: 5.
12. Mangandi J, Verma S, Osorio L, Peres NA, van de Weg E, Whitaker VM. Pedigree-based analysis in a multiparental population of octoploid strawberry reveals QTL alleles conferring resistance to *Phytophthora cactorum*. *G3 Genes Genomes Genet*. 2017; 7: 1707–1719.
13. Cai L, Voorrips RE, van de Weg E, Peace C, Iezzoni A. Genetic structure of a QTL hotspot on chromosome 2 in sweet cherry indicates positive selection for favorable haplotypes. *Mol Breeding*. 2017; 37: 85.
14. Cai L, Voorrips RE, van de Weg E, Peace C, Iezzoni A. Erratum to: Genetic structure of a QTL hotspot on chromosome 2 in sweet cherry indicates positive selection for favorable haplotypes. *Mol Breeding*. 2017a; 37: 100.
15. Durand JB, Allard A, Guitton B, van de Weg E, Bink MCAM, Costes E. Predicting Flowering Behavior and Exploring Its Genetic Determinism in an Apple Multi-family Population Based on Statistical Indices and Simplified Phenotyping. *Frontiers*. 2017; 8: 858.
16. Hernández Mora JR, Micheletti D, Bink M, Van de Weg E, Bassi D, Nazzicari N, et al. Integrated QTL detection for key breeding traits in multiple peach progenies. *BMC Genomics*. 2017; 18: 404.
17. Hernández Mora JR, Micheletti D, Bink M, Van de Weg E, Bassi D, Nazzicari N, et al. Discovering peach QTLs with multiple progeny analysis. *Acta Hort*. 2017a; 1172: 405-410.
18. Di Guardo M, Micheletti D, Bianco L, Koehorst-van Putten HJJ, Longhi S, Costa F, et al. ASSIST: An Automatic SNP Scoring Tool for in- and out-breeding species. *Bioinformatics*. 2015; 31: 3873-3874.
19. Di Guardo M, Bink M, Guerra W, Letschka T, Lozano L, Busatto N, et al. Deciphering the genetic control of fruit texture in apple by multiple-family based analysis and genome-wide association. *J Exp Bot*. 2017; 68.
20. Meuwissen TH, Goddard ME. Prediction of identity by descent probabilities from marker-haplotypes. *Genet Sel Evol*. 2001; 33: 605–634.
21. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2017.
22. Howard NP, van de Weg E, Tillman J, Tong CBS, Silverstein KAT, Luby JJ. Two QTL characterized for soft scald and soggy breakdown in apple (*Malus × domestica*) through pedigree-based analysis of a large population of interconnected families. *Tree Genet Genomes*. 2018; 14.
23. Verma S. A predictive genetic knowledge for apple. Washington State University. 2014.
24. Verma S, Zurn JD, Salinas N, Mathey MM, Denoyes B, Hancock JF, et al. Clarifying sub-genomic positions of QTLs for flowering habit and fruit quality in U.S. strawberry (*Fragaria ×ananassa*) breeding populations using pedigree-based QTL analysis. *Hortic Res*. 2017; 4: 17062.
25. Roach JA, Verma S, Peres NA, Jamieson AR, van de Weg WE, Bink MCAM, et al. FaRXf1: a locus conferring resistance to angular leaf spot caused by *Xanthomonas fragariae* in octoploid strawberry. *Theor Appl Genet*. 2016; 129: 1191-1201.
26. Allard A, Legave JM, Martinez S, Kelner JJ, Bink MCAM, Di Guardo M, et al. Detecting QTLs and putative candidate genes involved in budbreak and flowering time in an apple multiparental population. *J Exp Bot*. 2016; 67: 2875-2888.
27. Guan Y, Peace C, Rudell D, Verma S, Evans K. QTLs detected for individual sugars and soluble solids content in apple. *Mol Breed*. 2015; 35.
28. Fresnedo-Ramírez J, Bink MCAM, Van de Weg WE, Famula TR, Crisosto CH, Gasic K, et al. QTL mapping of pomological traits in peach and related species breeding germplasm. *Mol Breeding*. 2015; 35: 166.
29. Fresnedo-Ramírez J, Frett TJ, Sandefur PJ, Salgado-Rojas A, Clark JR, Gasic K, et al. QTL mapping and breeding value estimation through pedigree-based analysis of fruit size and weight in four diverse peach breeding programs. *Tree Genet. Genomes*. 2016; 12: 25.
30. Rosyara UR, Bink MCAM, van de Weg E, Zhang G, Wang D, Sebolt A, et al. Fruit size QTL identification and the prediction of parental QTL genotypes and breeding values in multiple pedigreed populations of sweet cherry. *Mol Breeding*. 2013.