University of São Paulo
"Luiz de Queiroz" College of Agriculture

Genomic analysis applied to tomato improvement: from genetic architecture to genomic selection

Roberta Luiza Vidal

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding

Piracicaba
2023

Roberta Luiza Vidal
Agronomist

Genomic analysis applied to tomato improvement: from genetic architecture to genomic selection

versão revisada de acordo com a Resolução CoPGr 6018 de 2011

Advisor:
Prof. Dr. **FERNANDO ANGELO PIOTTO**

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding

Piracicaba
2023

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Vidal, Roberta Luiza

Genomic analysis applied to tomato improvement: from genetic architecture to genomic selection / Roberta Luiza Vidal. - - Piracicaba, 2023. Versão revisada de acordo com a Resolução CoPGr 6018 de 2011.

85 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Melhoramento de porta enxertos 2. Genética do sistema radicular 3. Crescimento da copa 4. Predição genômica 5. Produção de tomate para processamento I. Título

DEDICATÓRIA

À todos os meus professores ao longo da pós-graduação.
Dedico

AGRADECIMENTOS

À minha família, em especial aos meu pais por todo amor e também por possibilitarem que eu cursasse o doutorado.

Ao meu orientador, Professor Dr. Fernando Angelo Piotto, por toda orientação e paciência durante o curso.

Aos amigos e colaboradores do Laboratório de Melhoramento de Hortaliças por toda ajuda e companheirismo durante a execução dos projetos.

Ao Professor Dr. David Francis e toda equipe do The Tomato Lab por me receberem durante o intercâmbio e ensinamentos passados.

À Escola Superior de Agricultura "Luiz de Queiroz" e a The Ohio State University pela minha formação acadêmica.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa para cursar o doutorado.

À todos os meus amigos, antigos e recentes, pela amizade e bons momentos compartilhados. Em especial ao meu amigo Eduardo Carvalho, que mesmo a distância me deu grande suporte na execução deste trabalho.

EPÍGRAFE

“Entretanto, durante muito tempo, eles cantaram cada um sozinho ou apenas alguns juntos, enquanto os outros escutavam; pois cada um compreendia apenas aquela parte da mente de Ilúvitar da qual havia brotado e evoluía devagar na compreensão de seus irmãos. Não obstante, de tanto escutar, chegaram a uma compreensão mais profunda, tornando-se mais consonantes e harmoniosos.”

(J. R. R. Tolkien, 1977)

CONTENTS

RESUMO

Análises genômicas aplicados ao melhoramento do tomateiro: da arquitetura genética até a seleção genômica

Tecnologias moleculares com grande potencial para auxiliar estudos genéticos se tornaram disponíveis para cultura do tomateiro (*Solanum lycopersicum*) recentemente. O que permitiu a aplicação de análises modernas ao estudo de caracteres complexos, especialmente dentro do contexto de populações de melhoramento. Variâncias fenotípicas e genotípicas identificadas nestas populações poderiam ser prontamente exploradas pelos melhoristas. Desta forma, a identificação de regiões genômicas e de marcadores associados a características de interesse possui enorme potencial para auxiliar o melhoramento do tomateiro. Neste trabalho diferentes tipos de análises genômicas foram aplicados a populações de tomate, tanto indústria quanto *in natura*, buscando estudar deste o controle genético de caracteres relevantes até a aplicabilidade de modelo de predição. O objetivo do primeiro capítulo foi fornecer conhecimento básico para auxiliar programas de melhoramento de porta enxerto. Para tal, um estudo de mapeamento associativo foi realizado em um painel de diversidade buscando desvendar o controle genético ligado a performance como porta enxerto e a características do sistema radicular. Polimorfismos associados a caracteres quantitativos foram identificados para quase todas as características avaliadas, bem como genótipos com potencial para serem usados como parentais. O segundo capítulo foi desenvolvido na The Ohio State University e utilizou diferentes populações de tomate indústria para reportar um *quantitative trait locus* (QTL) relacionado a produção total, validá-lo e incorporá-lo em modelos de predição genômica. Um QTL associado a produção total no cromossomo cinco foi identificado, validado e a incorporação de informações de marcadores ligados a este QTL e seu efeito gênico aumentou a capacidade preditiva de modelos de seleção genômica.

Palavras-chave: Melhoramento de porta enxertos, Genética do sistema radicular, Crescimento da copa, Predição genômica, Produção de tomate para processamento

ABSTRACT

Genomic analysis applied to tomato improvement: from genetic architecture to genomic selection

Molecular technologies that can greatly assist genetic studies are currently available for tomato crop (*Solanum lycopersicum*). This allowed the application of modern analysis to study complex traits, especially within a breeding population context. Phenotypic and genotypic variance found in this scenario could be readily explored by breeders. Thereby, identifying genomic regions and markers associated with important traits could greatly assist tomato breeding. Here different genomic analysis were applied to fresh and processing tomato populations aiming to study from the genetic control of relevant traits to the feasibility of prediction models. The goal of the first chapter of this thesis was to provide base knowledge to rootstock breeding programs. A genome-wide association study was performed on a diversity panel to uncover the genetic control of rootstock performance and root system features. Quantitative traits nucleotides associated with most traits evaluated were identified, as well as genotypes with the potential to be used as rootstock parents. The second chapter was developed at The Ohio State University and used different processing tomato populations to report a yield-related quantitative traits locus (QTL), validate it and incorporate it into genomic prediction models. A yield-related QTL on chromosome five was identified and validated, and adding linkage and gene effect information about it improved prediction accuracies.

Keywords: Rootstock breeding, Root system genetics, Scion growth, Genomic prediction, Processing tomato yield

1. INTRODUCTION

Tomato (*Solanum lycopersicum*) stands as one of the main vegetable crop worldwide. It industry can be divided into two basic types, the fresh-market tomato that is grown for fresh consumption, and the processing tomato that is destined for processed foods. Each type possesses its own characteristics of cultivation and ideotypes to be achieved by breeding programs.

The use of molecular markers has leveraged huge advances in plant breeding. Mapping experiments have been widely used to identify genes/quantitative traits locus (QTL) related to interest traits and markers associated with it that can further be used for selection. Indeed, marker-assisted selection (MAS) strategies have proven to increase the efficiency and cost-effectiveness of breeding programs (Vivek et al., 2017). In tomato crop there is a long history of using MAS, especially to guide the introduction of novel disease resistance from a wild relative into an improved variety and to pyramid resistance genes (Rick, 1974; Hanson et al., 2016).

As a model and well-study specie, many markers associated with genes or QTLs were reported for numerous economically important traits in tomato (Foolad and Panthee, 2012). It is important to highlight though, that most of the genetic inheritance research done so far focused on differences between domesticated and wild tomatoes, and most of this phenotypic/genotypic variance is already fixed in contemporary varieties (Foolad and Panthee, 2012; Bhandari et al., 2023).

Currently, more modern tools were developed and are available for the research community, including the tomato reference genome (The Tomato Genome Consortium, 2012), high-density genetic maps, and high-throughput molecular markers (Cappetta et al., 2020). These tools allowed the application of modern genomic analysis approaches, such as genome-wide association studies (GWAS) and genomic selection (GS), especially within a breeding population context.

In this thesis, different genomic analyses were applied to tomato (*S. lycopersicum*) populations aiming to study the genetic control of relevant traits and the feasibility of prediction models. The goal of the first chapter was to provide base knowledge that can assist rootstock breeding programs. A GWAS was performed on a diversity panel to uncover the genetic architecture of initial rootstock performance and root system features, and potential genotypes were identified to be further used as rootstock parents. The second chapter was developed at The Ohio State University and divided into three stages. The first stage used a recombinant inbred line (RIL) population to identify yield-related QTLs, and derived generations to validate those. Next, an independent Testcross population was used to estimate the gene effects of markers linked to the QTLs. Finally, genomic prediction models that incorporated information on linkage and gene effects

were developed using three training populations and tested through F1 prediction in an independent population of hybrids

References

Bhandari, P., Kim, J., & Lee, T. G. (2023). Genetic architecture of fresh-market tomato yield. BMC Plant Biology, 23:18. https://doi.org/10.1186/s12870-022-04018-5

Cappetta, E., Andolfo, G., Di Matteo, A., Barone, A., Frusciante, L., & Ercolano, M. R. (2020). Accelerating tomato breeding by exploiting genomic selection approaches. Plants, 9, 1236. https://doi.org/10.3390/plants9091236

Foolad, M. R. & Panthee, D. R. (2012). Marker-assisted selection in tomato breeding. Critical reviews in plant science, 31, 93-123. https://doi.org/10.1080/07352689.2011.616057

Hanson, P., S.-F. Lu, J.-F. Wang, W. Chen, L. Kenyon, C.-W. Tan, et al. (2016). Conventional and molecular marker-assisted selection and pyramiding of genes for multiple disease resistance in tomato. Scientia Horticulturae, 201, 346-354.

Rick, C. (1974). Association of an allozyme with nematode resistance. Tomato Genet. Coop. Rep 24: 25.

The Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature, 485, 635–641. https://doi.org/10.1038/nature11119

Vivek, B.S., G.K. Krishna, V. Vengadessan, R. Babu, P.H. Zaidi, L.Q. Kha, et al. (2017). Use of genomic estimated breeding values results in rapid genetic gains for drought tolerance in maize. The Plant Genome, 10: plantgenome2016.2007.0070. https://doi.org/10.3835/plantgenome2016.07.0070.

2. GENOMIC REGIONS FROM ROOTSTOCK ARE ASSOCIATED WITH IMPROVED EARLY STEM GROWTH AND ROOT TRAITS IN GRAFTED TOMATO

Abstract

Grafting has become a widespread tool for enhancing tomato growth, leading many companies to seek out superior rootstock materials. Despite this trend, there is limited information available on how rootstock affects scion development, or on the genetics of root morphology that could support breeding programs searching for superior materials. Here, we evaluated a diverse panel of tomatoes (*Solanum lycopersicum* L.) for their potential use as rootstock, as well as root morphology traits. Our aim was to obtain knowledge that could assist breeding programs. We conducted three experiments in different seasons to identify stable genotypes, investigate the relationship between root and shoot traits, and identify loci related to key traits using genome-wide association studies (GWAS). The panel was genotyped using a single nucleotide polymorphism (SNP) panel derived from the "SolCAP" Infinium array and through genotyping by sequencing, resulting in a total of 3,305 SNPs. We tested two GWAS models, FarmCPU and MLMM, and found that the panel showed high diversity for all evaluated traits, enabling us to select superior materials. We also found that vigorous root systems were highly correlated with improved early stem growth. We identified 7 shoot dry mass-related, 5 shoot height-related, 6 ratio of shoot/root-related, 7 root dry mass-related, and 4 root volume-related quantitative trait nucleotides (QTNs), with some QTNs associated with more than one trait. The FarmCPU model identified most of the associations. To our knowledge, this is the first population-level study focusing on rootstock traits, which opens possibilities for further research that can assist rootstock breeding.

Keywords: *Solanum lycopersicum*, genome-wide association studies (GWAS), genetic control, marker-assisted breeding, scion development

2.1. Introduction

Tomato is one of the most appreciated and produced vegetable crops worldwide. As a crop highly susceptible to several biotic and abiotic stressors, the use of the rootstock throughout grafting became a widespread tool in cultivations under greenhouses and open fields. Grafting can be used to improve plant vigor, increase yield and/or fruit quality, obtain resistance/tolerance to diseases, and obtain tolerance to restrictive environmental conditions (Kubota et al. 2008). Reports of increased yield in grafted tomatoes are widely reported (Venema et al. 2008; Barrett et al. 2012; Djidonou et al. 2013; Bayindir and Kamdemir, 2022). Increased scion vigor and yield are attributed to improved uptake of water and minerals conferred by the rootstock through a more vigorous root system (Martínez-Ballesta et al. 2010; Paez-Garcia et al. 2015; Bayindir and Kamdemir, 2022).

The increasing use of grafting tools has also affected breeding programs, by allowing them to focus independently on different traits for rootstock and scion genotypes (Mudge et al., 2009). Indeed, the interest of vegetable seed companies in obtaining superior rootstock genotypes has been growing (Lee et al., 2010). Since the rootstock contributes to the grafted complex through its root system, breeding strategies require a deeper understanding of how root morphology features influence scion development.

Research on rootstock performance has focused on measuring scion traits, with poor or no evaluation of root traits. Additionally, few studies have been conducted to uncover the

underlying genetic architecture modulating root system traits (Alaguero-Cordovilla et al. 2018; Xie et al., 2019). Understanding the variability and genetic contribution of rootstock features may help to give directions in developing genotypes with superior performance (Paez-Garcia et al., 2015; Khan et al., 2016; Bayindir and Kamdemir, 2022).

In addition to phenotypic assessment and quantification, genetic evaluation may provide additional informations about the characteristics of superior rootstocks. Molecular markers have been increasingly used in breeding programs to assist selection of interesting alleles. Due to the difficulty of phenotyping root systems, marker-based approaches are even more attractive for selection (Coudert et al., 2010). Genome-wide association studies (GWAS) provide an efficient method for identifying single nucleotide polymorphisms (SNPs) associated with specific traits, or quantitative trait nucleotides (QTNs). The availability of high-throughput genotyping of SNPs facilitates high-resolution genetic analysis (Cortes et al., 2021). In addition, GWAS is appropriate for unstructured germplasm collections and exploits genetic recombination events that occurred over several generations in the populations (Yu and Buckler, 2006) as a powerful tool for studying simple and complex traits (Cortes et al., 2021). Coupling GWAS with the several available tomato reference genomes allows more accurate QTL localization and elucidation of potential candidate genes (Sim et al., 2012; Tomato Genome Consortium, 2012).

Based on that, in this work we evaluated a diversity panel of tomatoes (*Solanum lycopersicum* L.) for their performance as rootstock, measured by scion growth after grafting, and root system morphology. So, this study had three main goals: a) Understand how root traits affect initial growth of the scion; b) Search for quantitative trait nucleotides (QTNs) associated with improved scion growth and root system features through GWAS analysis; and, c) Identify promising genotypes for use as rootstocks or parents in a rootstock breeding program.

2.2. Materials and Methods

Plant material and grafting

A tomato panel of 249 accessions was evaluated as rootstock. These accessions were selected from the germplasm collection belonging to the Horticultural Breeding Laboratory, Genetics Department of the University of São Paulo, including a diversity of accessions developed and/or collected in Brazil and in other countries. The collection contains accession of *Solanum lycopersicum* L. var. *esculentum* and *Solanum lycopersicum* L. var. *cerasiforme* (Supplementary Table 1).

The tomato seedlings were produced in coconut fiber substrate, using the nutrition protocol from the Horticultural Breeding Laboratory. At 20 days after sowing, the seedlings were

grafted using a scalpel and silicone clips to keep the plants together, using the splice method. After grafting, the plants were kept in a humid chamber and then acclimatized for 10 days before being ready for installation in the experiment. The panel genotypes were grafted under the common scion the cultivar "Santa Clara". As an additional control, cv. "Santa Clara" was self-grafted, totaling 250 treatments.

Experimental design and phenotypic evaluations

We conducted three experiments to evaluate the rootstock morphologic traits and changes in early scion growth under grafting. Prior to establishing hydroponic production we carefully washed grafted seedling roots to remove all substrates. We choose to grow the plants in a static hydroponic system (Supplementary Figure 1) to allow free root growth and to avoid any loss that could be caused by root cleaning prior to phenotypic evaluation and quantification. The trials were set up in a randomized block design, with three replications/blocks, each block corresponding to one pool, and plots composed of three plants. Each pool consisted of 600 liters of a modified Hoagland and Arnon (1950) diluted to 50% concentration relative to complete nutrient solution (Supplementary Table 2). This volume corresponded to a pool depth of 14 centimeters. Two water circulation pumps were installed in each pool to keep the nutrient solution homogeneous. We used polystyrene plates to suspend the seedlings over the nutrient solution. Further detail on pool structure can be found in Supplementary Figure 1. Nutrient solution levels (height relative to the depth) was measured daily and the volume was maintained by adding more water as needed. The experiments were conducted in three time periods, November to December 2020, February to March 2021, and July 2021, to account for potential environmental x genotype interaction and allow the selection of stable genotypes. The weather conditions can be found in Supplementary Figure 2. A total of 196 genotypes were evaluated in experiment 1, and 241 genotypes in experiments 2 and 3.

The plants were grown for two weeks before evaluation. At the end of the experiment, the plants were cut at the point of grafting, separating the shoot from the root system. We measured the shoot height (SH, cm) from the grafting point to the last branch with a ruler and then stored it in identified paper bags. The roots were stored in plastic pots with a 25% ethanol solution to preserve them until evaluation.

The shoot samples were put into a forced air circulation oven at 65 °C and, after the samples were completely dry (constant mass), they were weighed to obtain the shoot dry mass (SDM, g). We used an Epson LA2400 scanner to acquire root images and processed these images using the WinRHIZO (Reagent Instruments Inc., Quebec, Canada) to obtain root volume (RV,

cm^3), root average diameter (RAD, mm), and root surface area (cm^2). Next, the samples were dried out and weighed to determine the root dry mass (RDM, g). The ratio of shoot/root (RSR, $g.g^{-1}$) was obtained by dividing the SDM by the RDM, and the root-specific area (RSA, $cm^2.g$) by dividing the root surface area by the RDM.

DNA extraction, genotyping, and quality filtering

DNA extraction was performed on young leaflets with the DNeasy Plant Mini Kit (Qiagen). The DNA was quantified based on the fluorescence method using Qubit (Invitrogen) and normalized at 30 $\eta g.uL^{-1}$. Two types of marker sets were used. First, we genotyped the panel using an optimized set of 384 single nucleotide polymorphisms (SNPs). This SNP panel was derived from the validated polymorphisms in the "SolCAP" Infinium array (Illumina Inc., San Diego, CA)(Sim et al. 2012a) by filtering the 7,000 SNPs based on genome coverage, recombination and polymorphic information content (Sim et al. 2012b). The 384 SNPs were called using an amplicon-based sequencing assay (PlexSeqTM, AgriPlex Genomics, Cleveland, Ohio, USA).

Additionally, we obtained a genotyping by sequencing (GBS) marker set (Elshire et al. 2011). Genomic libraries were built according to Poland et al. (2012) with some modifications, using the restriction enzymes PstI and MseI. Then, the libraries were sequenced on an Illumina HighSeq2500. More details on genotyping, alignment, and SNP calls are described in Niederheitmann (2021).

We identified the physical position for both marker sets using the tomato reference genome SL.3.0 (The Tomato Genome Consortium, 2012). Merging positional and genetic information and data filtering was accomplished using R v. 3.6.1 (R Development Core Team, 2020). Retained SNPs with a Call rate > 90% and removed markers with a Minor allele frequency (MAF) lower than 3%. The initial set composed of 384 PlexSeqTM SNPs and 7,247 GBS SNPs was reduced to 3,305 total SNPs after filtering and was the set used for GWAS analysis.

Phenotypic analysis

Principal component analysis and Pearson correlation were performed to study how the traits interact with each other. We then used the ASReml-R 4.0 package (Butler et al. 2017) to fit the following model and extract the adjusted means:

$$y = X_1 g + X_2 e + X_3 ge + Z_1 b + \varepsilon \qquad (1)$$

where y refers to the phenotypic observation; X_1, X_2, and X_3 are the incidence matrices of fixed effects; and Z_1 is the incidence matrix for random effect. g is the genotype fixed effect; e is the experiment fixed effect; ge is the fixed effect of interaction between genotype and experiment; b is the random effect of blocks within experiments, where $b \sim N(0, I\sigma_b^2)$; and ε is the residual random effect, where we used the command ~dsum to account for different variances between experiments. A completely random model was fitted to extract variances components, and calculate broad sense heritabilities using the following formula:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{e} + \frac{\sigma_r^2}{be}} \qquad (2)$$

Where, h^2 refers to heritability; σ_g^2 is the genotype variance component; σ_{ge}^2 is the variance component due to interaction between genotype and experiments; σ_r^2 is the residual variance component; e is the number of experiments (e=3); and b is the number of blocks (b=3).

GWAS analysis and QTNs interval confidence

GWAS was performed on the adjusted means of all traits evaluated using the GAPIT package (Lipka et al. 2012). We fit two multi-locus models to search for trait-associated QTNs, Multiple Loci Linear Mixed Model (MLMM) (Segura et al. 2012) and Fixed and Random Circulating Probability Unification (FarmCPU) (Liu et al. 2016). MLMM is a modified Mixed Linear Model (MLM) (Yu et al. 2006), that in addition to incorporating population structure and kinship as an adjustment to control false positives on association tests, also includes markers as covariates in a stepwise process that intends to partially remove the confounding effects of kinship and testing markers. FarmCPU divides MLMM into two parts that are iteratively used: a fixed effect model that tests the markers one at a time using multiple associated markers as covariates aiming to control false positives, and a random effect model that estimates and uses the associated markers to define kinship, avoiding over-fitting. We used three PCA to account for the population structure for both models. A threshold of 0.0005 was used to declare a significant association with a QTN to balance type I and type II error as described (Nguyen et al. 2021).We used the CMplot package (Yin et al. 2021) to build the Manhattan plots, QQ plots, and a figure representing the genome coverage for the markers used in GWAS.

QTN confidence intervals surrounding each trait-associated SNP were defined based on linkage disequilibrium (LD) as described previously (Bineau et al. 2021). We first compute the marker pairwise LD within each chromosome using the Sommer package (Covarrubias-Pazaran

2016). We next defined the confidence interval based on LD r^2 higher than 0.5 in a 2 Mbp region around the SNP (1 Mb downstream and upstream). Multiple SNPs within the confidence interval were treated as the same QTN.

2.3. Results

Phenotypic data analysis

Exploratory analysis and heritabilities

The minimum, mean, and maximum values for the phenotypic traits are presented in Table 1, and demonstrate the panel diversity for all traits evaluated. Heritabilities ranged from 0.22, for shoot height (SH), to 0.34, for root average diameter (RAD), with indices showing higher values. The interaction between genotypes and experiments contributed variance to the denominator which decreased the heritabilities of all traits.

Table 1. Range of values and heritabilities for phenotypic traits. SDM, shoot dry mass; SH, shoot height; RSR, ratio of shoot/root; RDM, root dry mass; RAD, root average diameter; RV, root volume; RSA, root specific area

Trait	SDM	SH	RSR	RDM	RAD	RV	RSA
	g	cm	$g\ g^{-1}$	g	mm	cm^3	$cm^2\ g^{-1}$
Minimum	0,35	11,93	4,31	0,07	0,46	0,71	784,4
Mean	0,69	18,58	7,11	0,09	0,52	1,49	1116,1
Maximum	1,13	23,6	11,98	0,16	0,61	2,45	1461,5
Heritability	0,26	0,22	0,30	0,32	0,34	0,24	0,33

We observed a positive correlation between almost all traits, except RAD and ratio of shoot/root (RSR) (Figure 1). Indeed, RAD was negatively correlated with all others traits. RSR can be considered a measure of carbon partitioning between shoot and roots, and although it was negatively correlated with root dry mass (RDM), it was positively correlated with shoot dry mass (SDM) and shoot SH. The trait most correlated with SDM was SH, followed by root volume (RV) and RDM. RDM and RV were also highly correlated with each other. RSA can be interpreted as a contact surface measure that reflects how much root system carbon was truly converted in root contact area, and it was moderately correlated with the other traits, except RAD.

PCA analysis also demonstrated the high correlation between traits, with the first two principal components accounting for 73% of data variation (Figure 2). Visualization of the first two PCA also illustrated the panel's phenotypic diversity relative to both root and scion traits as influenced by grafting.

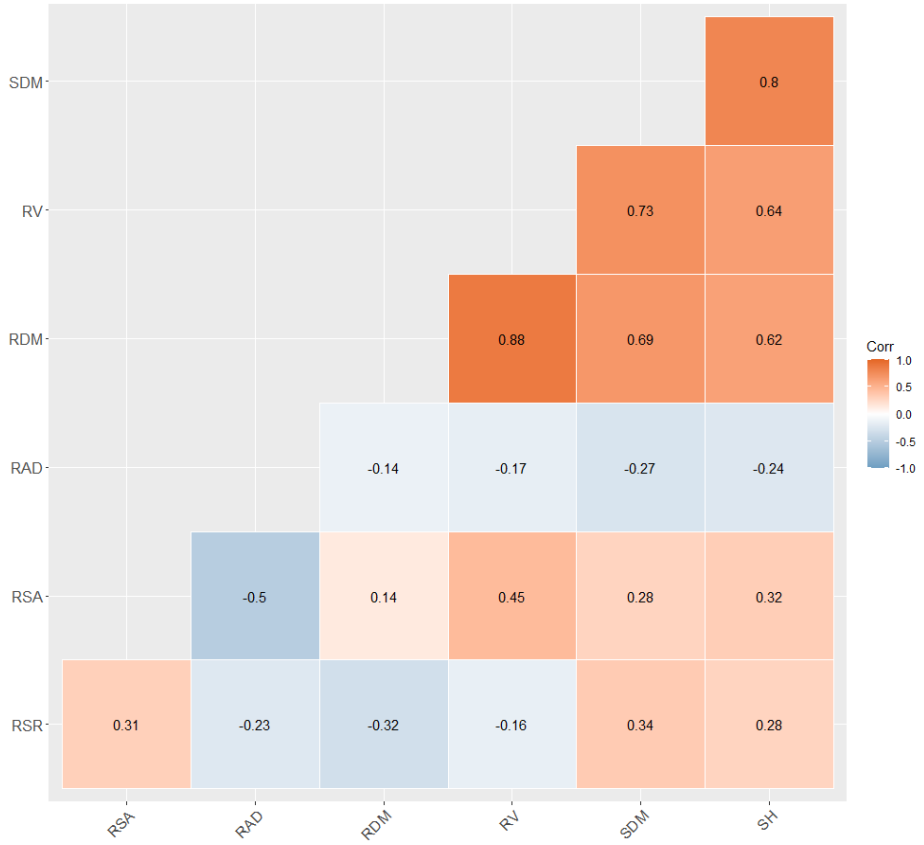


Figure 1. Pearson´s Correlation between the seven evaluated phenotypic traits. SDM, shoot dry mass; SH, shoot height; RSR, ratio of shoot/root; RDM, root dry mass; RAD, root average diameter; RV, root volume; RSA, root specific area. Non-significant correlations are omitted

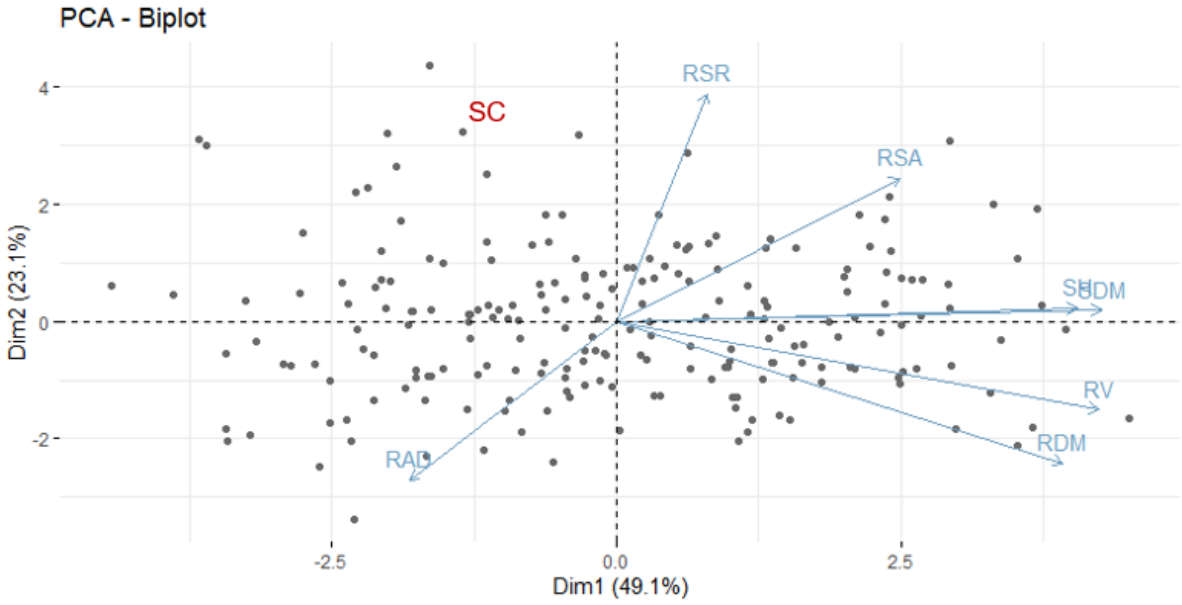


Figure 2. Principal components analysis using 7 phenotypic traits and 250 treatments. SDM, shoot dry mass; SH, shoot height; RSR, ratio of shoot/root; RDM, root dry mass; RAD, root average diameter; RV, root volume; RSA, root specific area

Panel rootstock potential

Considering SDM as a direct measure of scion development, we found that most panel genotypes were able to impart greater scion growth compared to the "Santa Clara" self-grafted control (Figure 3). This result can also be observed in the PCA plot (Figure 2), where the self-grafted control is placed further to the left of the first component axis, which mostly accounts for SDM and SH variation.

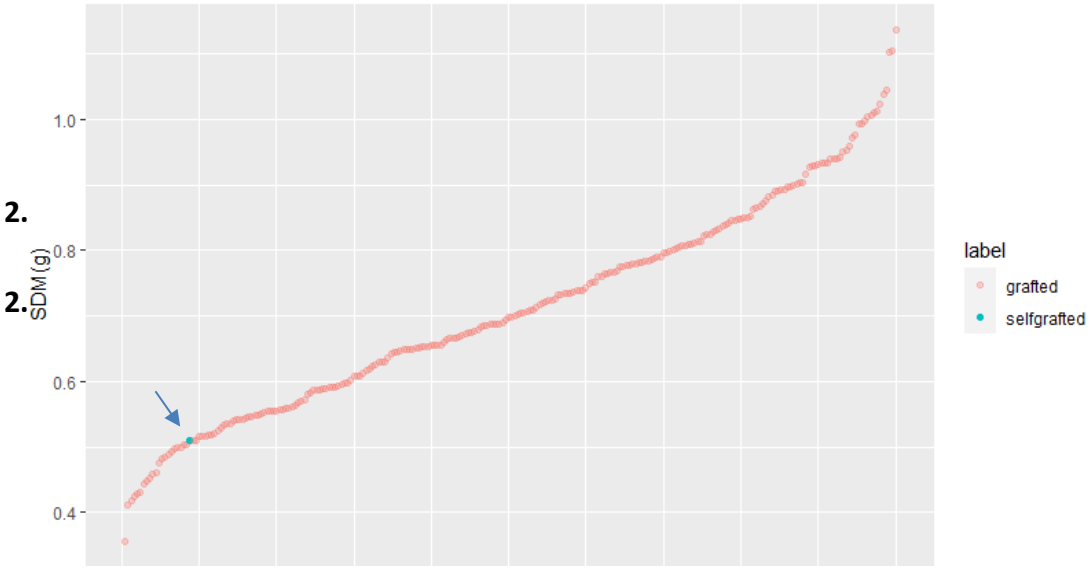


Figure 3. Genotypes grafting potential to increase shoot dry mass (SDM)

Selection of best rootstock genotypes

We chose to use the adjusted mean model (equation 1) to select the top twenty SDM genotypes due to the heterogeneity of variances between experiment 1 and experiments 2 and 3. Genotypes panel numbers 21, 22, 25, 34, 85, 96, 128, 199, 205, 217, 234, 236, 244, 249, 275, 289, 297, 336, 337, and 359 were selected by the model, and most of them displayed SDM performance that exceeded the mean for all experiments (Figure 4).

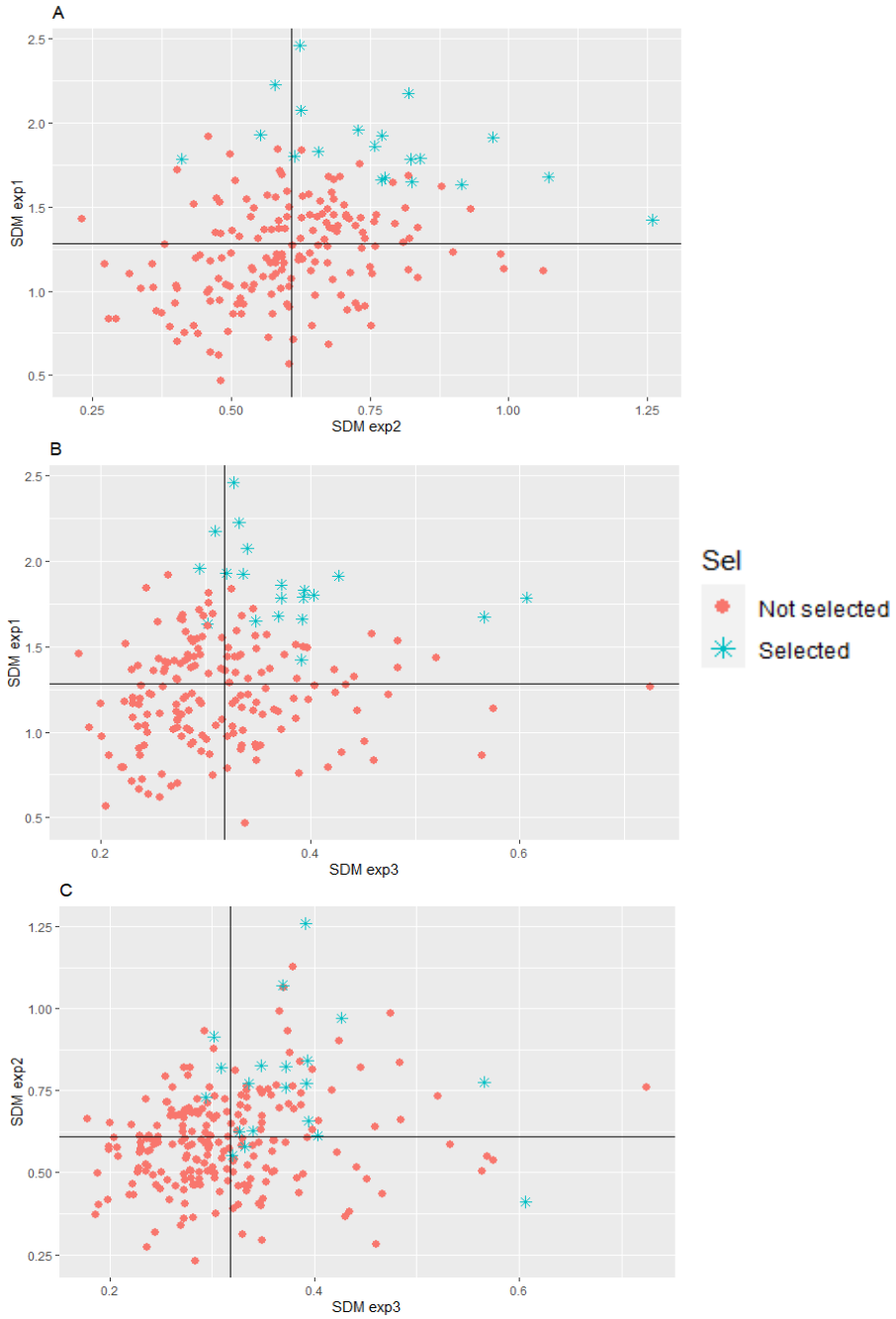


Figure 4. Biplots of shoot dry mass (SDM) adjusted means for the 249 genotypes evaluated, including the selected individuals. (A) environment 1 x environment 2; (B) environment 1 x environment 3; (C) environment 2 x environment 3. The black lines represent the experiment's means

Genotypic data analysis

Genome coverage

Following quality control we retained 3,305 SNPs. These markers were unevenly distributed across the genome. In most chromosomes, the markers were placed mainly on the distal ends with few SNPs on the pericentromeric regions. Chromosomes 4, 5, and 9 presented

more markers distributed along these. The average number of markers per chromosome was 275, with a low of 86 on chromosome 8 and a high of 556 on chromosome 9 (Supplementary Table 3). A graphical representation of genome coverage can be found in Supplementary Figure 3.

Genome-wide association study

A total of 30 QTNs were identified. At least one QTN was identified for most traits evaluated, with the exception of RAD and RSA. We detected eight markers associated with SDM (Figure 5), nine with SH, six with RSR, eight with RDM, and four with RV (Supplementary Figures 4-7). QTNs were found on almost all chromosomes, suggesting complex inheritance for most of these traits.

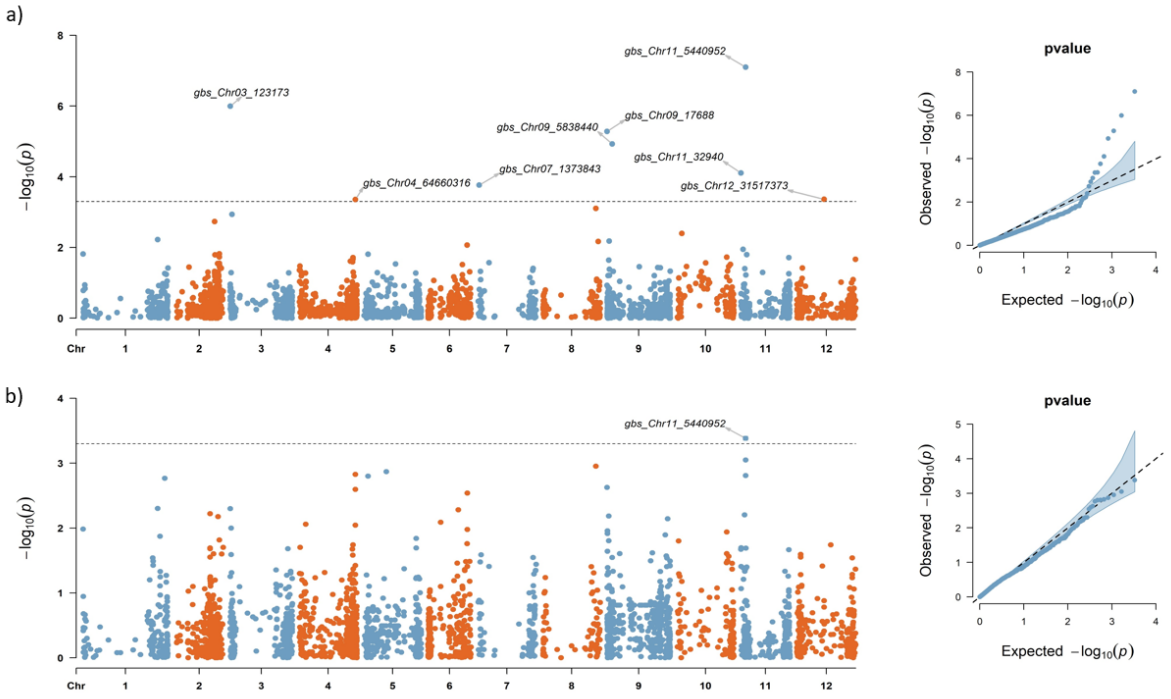


Figure 5. Manhattan and QQ plots for shoot dry mass (SDM). (A) FarmCPU model; (B) MLMM model. The grey dotted lines on Manhattan plots represent the defined cut ($p > 0.0005$)

QTNs were further evaluated through boxplot visualization (Figure 6 for SDM, and Supplementary Figures 8, 9 and 10 for the other traits). Boxplot visualization revealed the possibility false positives associations, such as gbs_Chr07_1373843 for SDM, gbs_Chr01_2140956 and gbs_Chr11_5693979 for SH, and gbs_Chr02_53445405 for RDM. These markers did not show a clear effect of an allele substitution and thus were removed for further steps.

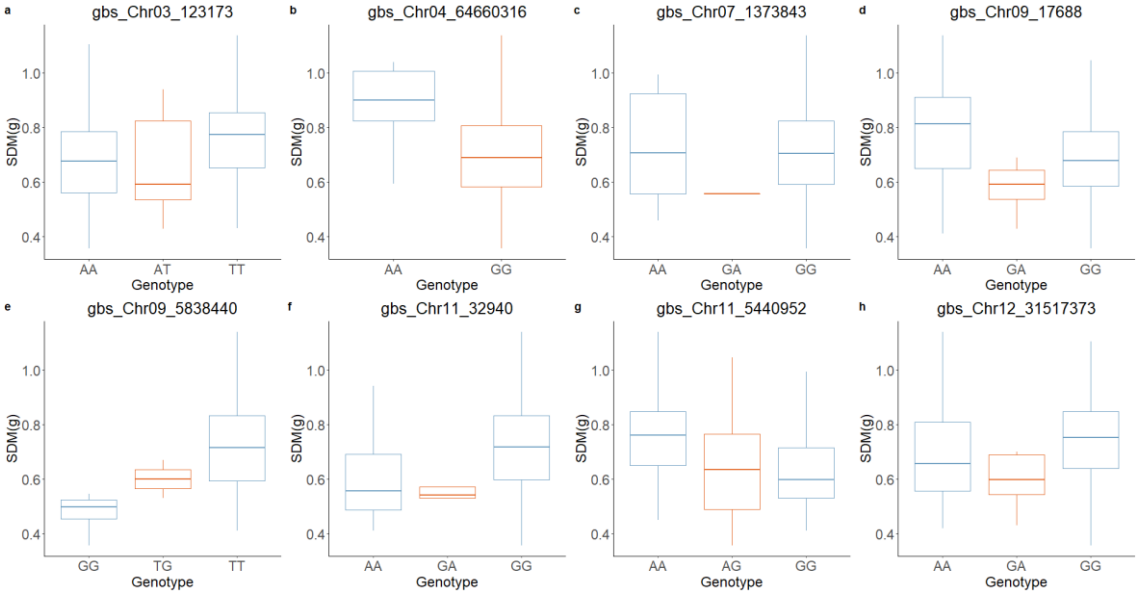


Figure 6. Boxplot visualization of the effect of allele substitution for shoot dry mass (SDM) detected QTNs

Based on SNP association and confidence interval we defined seven reliable SDM-related, five SH-related, six RSR-related, seven RDM-related, and four RV-related QTNs. The QTNs are summarized in Table 2. Markers gbs_Chr03_123173, gbs_Chr04_64660316, gbs_Chr06_43560142, and gbs_Chr11_5440952 were detected for more than one trait, though always between the highly correlated traits such as SDM, SH, RDM, and RV. In some cases, although different markers were detected for correlated traits, they are located inside the confidence interval. This is the case for markers gbs_Chr09_5599363 and gbs_Chr09_5838440, which are placed in the interval between 5573820 and 6776714 bp on chromosome 9, markers gbs_Chr10_57346179 and gbs_Chr10_57351889, located between 57197730 and 58350349 on chromosome 10, and markers gbs_Chr11_5440952, gbs_Chr11_5507348, gbs_Chr11_5542428, and gbs_Chr11_5693979, which accounted for the same region on chromosome 11.

FarmCPU showed greater ability in detecting associations than MLMM model and most markers identified by the latter were also identified by FarmCPU. We only found one marker

exclusively detected by MLMM and that represents a unique chromosome region, gbs_Chr03_2603399.

Table 2. Quantitative trait nucleotides (QTNs) detected through GWAS analysis, chromosome, their positions, related trait, model of detection, minor allele frequency (MAF), and alleles ([1]reference allele; [2]alternative allele)

Markers	Chr	Position (Mb)	Trait	Model	MAF
solcap_snp_sl_6255	2	20824592	RSR	FarmCPU	A[1]/G[2] (0.30)
gbs_Chr02_45884979	2	45884979	RDM	FarmCPU	G[1]/A[2] (0.20)
gbs_Chr03_123173	3	123173	SDM, RDM, and RV	FarmCPU	A[1]/T[2] (0.35)
gbs_Chr03_2603399	3	2603399	RSR	MLMM	T[1]/G[2] (0.09)
gbs_Chr04_12360258	4	12360258	RSR	FarmCPU	C[1]/T[2] (0.06)
gbs_Chr04_55420736	4	55420736	RSR	FarmCPU	G[1]/A[2] (0.21)
gbs_Chr04_64660316	4	64660316	SDM and RDM	FarmCPU	G[1]/A[2] (0.06)
solcap_snp_sl_2629	6	33633429	SH	FarmCPU	T[1]/C[2] (0.43)
gbs_Chr06_43560142	6	43560142	RDM and RV	FarmCPU and MLMM	A[1]/G[2] (0.15)
gbs_Chr07_63050211	7	63050211	SH	FarmCPU	C[1]/T[2] (0.19)
gbs_Chr08_61010457	8	61010457	RDM	FarmCPU and MLMM	A[1]/G[2] (0.09)
gbs_Chr09_17688	9	17688	SDM	FarmCPU	G[1]/A[2] (0.30)
gbs_Chr09_5599363	9	5599363	SH	FarmCPU	G[1]/A[2] (0.04)
gbs_Chr09_5838440	9	5838440	SDM	FarmCPU	T[1]/G[2] (0.06)
gbs_Chr09_28487419	9	28487419	RV	FarmCPU	C[1]/T[2] (0.07)
gbs_Chr09_57177772	9	57177772	RSR	FarmCPU	G[1]/T[2] (0.08)
gbs_Chr10_57346179	10	57346179	RV	FarmCPU	G[1]/T[2] (0.43)
gbs_Chr10_57351889	10	57351889	RDM	FarmCPU	G[1]/C[2] (0.42)
solcap_snp_sl_8859	10	63956636	RDM	FarmCPU	C[1]/G[2] (0.11)
gbs_Chr11_32940	11	32940	SDM	FarmCPU	G[1]/A[2] (0.07)
gbs_Chr11_5440952	11	5440952	SDM and SH	FarmCPU and MLMM	A[1]/G[2] (0.32)
gbs_Chr11_5507348	11	5507348	SH	MLMM	T[1]/C[2] (0.22)
gbs_Chr11_5542428	11	5542428	SH	MLMM	G[1]/C[2] (0.22)
gbs_Chr11_37771444	11	37771444	RSR	FarmCPU	A[1]/T[2] (0.25)
solcap_snp_sl_12664	12	2301635	SH	FarmCPU	A[1]/T[2] (0.27)
gbs_Chr12_31517373	12	31517373	SDM	FarmCPU	G[1]/A[2] (0.49)

QTN genotype of selected individuals

An effective way to design crosses that aim to further develop superior rootstock material to improve scion growth could be pursued by combining beneficial alleles at the associated QTNs. Seven reliable SDM-related and four RDM-related QTNs were identified and their genotypes for the twenty superior accessions are presented in Table 3. Markers gbs_Chr06_43560142, gbs_Chr08_61010457, and gbs_Chr09_5838440 are monomorphic for the beneficial allele in the selected set and can be ignored. Genotypes GENO199 and GENO25 have almost all associated-QTNs with the beneficial allele and are potential parents for crosses, such as GENO199 x GENO25, GENO199 x GENO128, and GENO25 x GENO205.

Table 3. Shoot dry mass (SDM) and root dry mass (RDM)-related quantitative traits nucleotides (QTN) genotypes for the twenty best performing accessions

Chromosome	QTN*s*										
	2	3	4	6	8	9	9	10	11	11	12
Position (bp)	45884979	123173	64660316	43560142	61010457	17688	58384 40	57351889	32940	5440952	31517373
Alleles	G/A	A/T	G/A	A/G	A/G	G/A	T/G	G/C	G/A	A/G	G/A
Beneficial allele	A	T	A	A	A	A	T	G	G	A	G
GENO21	AA	AA	GG	AA	AA	GG	TT	GG	GG	AA	GG
GENO22	GG	AA	GG	AA	AA	AA	TT	GG	GG	AA	GG
GENO25	GG	TT	GG	AA	AA	AA	TT	GG	GG	AA	GG
GENO34	GG	AA	GG	AA	AA	GG	TT	CC	GG	AA	GG
GENO85	GG		GG	AA	AA	AA	TT	GG	GG	GG	AA
GENO96		AT		AA	AA	AA	TT	GC	AA	AA	GG
GENO128	GG	TT		AA	AA	AA	TT	GG	GG	GG	AA
GENO199	GG	AA	AA	AA	AA	AA	TT	GG	GG	AA	GG
GENO205	GG	TT	GG	AA	AA	AA	TT	GG	GG	GG	GG
GENO217	GG	TT	GG	AA	AA	AA	TT	GG	GG	AA	AA
GENO234	GG	AA	GG	AA	AA	AA	TT	GG	GG	GG	GG
GENO236	AA	AA	GG	AA	AA	GG	TT	GG	GG	AA	GG
GENO244	GG	TT	GG	AA	AA	GG	TT	GG	GG	AA	GG
GENO249	GG	TT	GG	AA	AA	GG	TT	GC	GG	AG	GG
GENO275	GG	AA	GG	AA	AA	AA	TT	GG	GG	GG	GG
GENO289	GG	AA	AA	AA	AA	GG	TT	CC	GG	AA	AA
GENO297	GG	AA	AA	AA	AA	GG	TT	GG	GG	GG	AA
GENO336	GG	AA	GG	AA	AA	GG	TT	GG	GG	AA	GG
GENO337	AA	AA	GG	AA	AA	GG	TT	CC	GG	AA	GG
GENO359	GG		AA	AA	AA	AA	TT	GG	GG	GG	AA

2.4. Discussion

We evaluated rootstock performance based on scion growth and root morphology in a large and diverse population of *S. lycopersicum*. We found an uneven distribution of markers on the genome with a concentration of SNPs on the chromosome's distal ends. This is common coverage patter reported in tomato (Chen et al., 2014; Xie et al., 2019), and it is probably due to the narrow genetic diversity resulting from the high inbreeding rate that occurred during the tomato domestication process (Foolad, 2007). Although tomato is known to be a crop with a narrow genetic base, the germplasm panel demonstrated high phenotypic diversity for both the ability to improve scion growth and desirable root morphology traits. Superior accessions were identified based on phenotypic performance.

The performance of grafted plants is usually attributed to an improved ability to absorb water and nutrients conferred through a more vigorous root system (Martínez-Ballesta et al., 2010; Lovelli et al., 2012; Paez-Garcia et al., 2015; Suchoff et al., 2017; Xie et al., 2019; Bayindir and Kamdemir 2022). A comparison of self-grafted and non-grafted plants suggested a potentially beneficial effect on scion development caused by the grafting process itself (Khah et al, 2006; Sánches-Rodríguez et al., 2014; Kabas & Kucukaydin, 2022). A robust root system has also been related to tolerance to edaphic stress, such as low availability of nutrients (Hill et al., 2006; Lambers et al., 2006; Suchoff et al., 2017).

We did not investigate the grafting effect *per se*, but our results showed that most genotypes evaluated were able to improve scion growth compared to the self-grafted check. We attribute improved scion growth to a robust root system. We found a high correlation between SDM and RV/RDM, and two QNTs (gbs_Chr03_123173 and gbs_Chr04_64660316) which were associated with both SDM and RDM. Indeed, this high positive correlation between traits is beneficial for breeding purposes as it allows for the simultaneous selection of traits.

Tomato has a long history of genetic breeding, though few programs have addressed root morphology traits as a key breeding objective, even for rootstock materials. The main reasons for this gap of information are attributed to the difficulty of phenotyping and complex or unknown genetic control of root traits (Kuijken et al., 2015). QTL mapping in an F2 population, derived from a cross between *S. lycopersicum* and *S. cheesmanie*, and evaluated as rootstock in early seedlings stages identified two QTLs hotspot regions for root and shoot traits (Xie et al., 2019). Although none of these QTLs overlapped with those we identified here, the authors also reported a high positive correlation between root and shoot traits.

Our work took a first step in filling this gap by providing basic information on the genetic background of key shoot/root traits. We were able to select twenty superior genotypes to be used as rootstock parents and, by looking at their genotypes, to design crosses that could join together most of the identified QTNs for both shoot and root traits. Selection root traits must be tagged if ones want to succeed in rootstock breeding programs (Bayindir and Kamdemir 2022).

The evaluation of rootstock materials should be performed within the context of grafting since there is evidence scion may also affect root growth (Kakita et al., 2015). In addition, roots present plasticity depending on the medium in which they are grown (Bao et al., 2014; Robbins et al., 2015). Research will be needed to assess whether the hydroponic evaluation presented here can be extrapolated to soil or substrate media. Finally, further studies are also necessary to validate the QTNs and discover if they continue to affect the growth and vigor of older plants.

We started with a set of 7,631 SNPs that was reduced to 3,305 after filtering for quality and MAF. This number provided excellent genome coverage, especially given the reduced recombination found in populations of *S. lycopersicum* (Sim et al., 2012a). For GWAS analysis, the FarmCPU model showed greater statistical power relative to MLMM, with almost all QTNs detected with MLMM also detected by FarmCPU. FarmCPU accounts for the confounding effects of kinship by only using the associated markers to derive the kinship matrix used as a co-variate. The nature of crop improvement programs creates a situation where kinship estimates may differ across the genome due to selection. GWAS model comparison using simulated and real data from maize and soybean indicated FarmCPU reduced both false positives and false negatives relative to other models (including MLMM) (Kaler et al. 2020). The superiority of FarmCPU over MLM model was also reported in tomatoes (Rodriguez et al. 2020).

2.5. Conclusion

A robust root system is associated with improved early growth in grafted tomatoes. Stable QTNs related to shoot dry mass (SDM), shoot height (SH), the ratio of shoot/root (RSR), root dry mass (RDM), and root volume (RV) were identified. We select twenty superior rootstock materials to be used as parents and allow the combination of many of the detected QTNs.

References

Alaguero-Cordovilla A, Gran-Gómez FJ, Tormos-Moltó S, Pérez-Pérez JM (2018) Morphological characterization of root system architecture in diverse tomato genotypes during early growth. International Journal of Molecular Sciences 19:3888. https://doi.org/10.3390/ijms19123888

Bao Y, Aggarwal P, Robbins NE, Sturrock CJ, Thompson MC, Tan HQ, Tham C, Duan L, Rodriguez PL, Vernoux T (2014) Plant roots use a patterning mechanism to position lateral root branches toward available water. PNAS 111:9319–9324. https://doi.org/10.1073/pnas.1400966111

Barrett CE, Zhao X, Mcsorley R (2012) Grafting for root-knot nematode control and yield improvement in organic heirloom tomato production. HortScience 47:614–620. https://doi.org/10.21273/HORTSCI.47.5.614

Bayındır S, Kandemir D (2022) Root system architecture of interspecific rootstocks and its relationship with yield components in grafted tomato. Gesunde Pflanzen. https://doi.org/10.1007/s10343-022-00704-4

Bineau E, Rambla JL, Priego-Cubero S, Hereil A, Bitton F, Plissonneau C, Granell A, Causse M. (2021) Breeding tomato hybrids for flavour: comparison of GWAS results obtained on lines and f1 hybrids. genes 12(9):1443. https://doi.org/10.3390/genes12091443

Butler DG, Cullis BR, Gilmour AR et al (2017) ASReml-R reference manual version 4. VSN International Ltd

Chen A-L, Liu C-Y, Chen C-H, Wang J-F, Liao Y-C, Chang C-H, et al (2014) Reassessment of QTLs for Late Blight Resistance in the Tomato Accession L3708 Using a Restriction Site Associated DNA (RAD) Linkage Map and Highly Aggressive Isolates of Phytophthora infestans. PLoS ONE 9(5): e96417. https://doi.org/10.1371/journal.pone.0096417

Cortes LT, Zhang Z, Yu, J (2021) Status and prospects of genome-wide association studies in plants. Plant Genome 14(1):e20077. http://doi.org/ 10.1002/tpg2.20077

Coudert Y, Perin C, Courtois B, Khong NG, Gantet P (2010). Genetic control of root development in rice, the model cereal. Trends in Plant Science 15(4):219–226. http://doi.org/10.1016/j.tplants.2010.01.008

Covarrubias-Pazaran G (2016) Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer. PLoS ONE 11(6): e0156744. https://doi.org/10.1371/journal.pone.0156744

Djidonou D, Zhao X, Simonne EH, Koch KE, Erickson JE (2013) Yield, water-, and nitrogen-use efficiency in fieldgrown, grafted tomatoes. HortScience 48:485–492. https://doi.org/10.21273/HORTSCI.48.4.485.

E Sánchez-Rodríguez, R Leyva, C Constán-Aguilar, L Romero, JM Ruiz (2014) How does grafting affect the ionome of cherry tomato plants under water stress?. Soil Science and Plant Nutrition. http://doi.org/10.1080/00380768.2013.870873

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6(5):e19379. http://doi.org/10.1371/journal.pone.0019379

Foolad MR (2007) Genome mapping and molecular breeding of tomato. International Jounal of Plant Genomics 64358. http://doi.org/10.1155/2007/64358

Hoagland DR, Arnon DI (1950) The water-culture method for growing plants without soil. California: Agricultural Experiment Station.

Kabas A, Kucukaydin H (2020) Effect of tomato interspecific hybrid (f1) rootstocks on yield and fruit quality traits. Gesunde Pflanzen. https://doi.org/10.1007/s10343-022-00725-z

Kakita T, Abe A, Ikeda T (2015) differences in root growth and permeability in the grafted combinations of dutch tomato cultivars (Starbuck and Maxifort) and japanese cultivars (Reiyo, Receive, and Magnet). American Journal of Plant Sciences 6:2640-2650. http://doi.org/10.4236/ajps.2015.616266

Kaler S, Gillmand JD, Beissinger T, Purcell LC (2020) Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. Frontiers in Plant Science. http://doi.org/10.3389/fpls.2019.01794

Khan MA, Gemenet DC, Villordon A (2016) Root system architecture and abiotic stress tolerance: current knowledge in root and tuber crops. Frontiers in Plant Science 7:1584. https://doi.org/10.3389/fpls.2016.01584

Kubota C, McClure MA, Kokalis-Burelle N, Bausher MG, Rosskopf EN (2008) Vegetable grafting: History, use, and current technology status in North America. HortScience 43(6):1664-1669. https://doi.org/10.21273/HORTSCI.43.6.1664

Kuijken RC, van Eeuwijk FA, Marcelis LF, Bouwmeester HJ (2015) Root phenotyping: from component trait in the lab to breeding. Journal of Experimental Botany 66(18):5389-401. http://doi.org/10.1093/jxb/erv239

Lee JM, Kubota C, Tsao SJ, Bie Z, Echevarria PH, Morra L, Oda M (2010) Current status of vegetable grafting: Diffusion, grafting techniques, automation. Scientia Horticulturae 127:93-105. http://doi.org/10.1016/j.scienta.2010.08.003

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28(18):2397–2399. https://doi.org/10.1093/bioinformatics/bts444

Liu X, Huang M, Fan B, Buckler ES, Zhang Z (2016) Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLoS Genet 12(2): e1005767. https://doi.org/10.1371/journal.pgen.1005767

Lovelli S, Perniola M, Di Tommaso T, Bochicchio R, Amato M (2012) Specific root length and diameter of hydroponically grown tomato plants under salinity. Agronomy 11:101–1106. https://doi.org/10.3923/ja.2012.101.106

Martínez-Ballesta MC, López-Pérez L, Hernández M, López-Berenguer C, Fernández-García N, Carvajal M (2008) Agricultural practices for enhanced human health. Phytochem Rev 7:251–260. https://doi.org/10.1007/s11101-007-9071-3

Mudge K, Janick J, Scofield S, Goldschmidt EE (2009) A history of grafting. In: Janick J (ed) Horticultural reviews John Wiley & Son, pp 437-493

Niederheitmann M (2021) Fenotipagem e associação genômica ampla para resistência à mancha bacteriana do tomateiro (Solanum lycopersicum L.). Dissertation, University of São Paulo

Nguyen TT, Ngoc TL, Sim S-C (2021) Genome-wide association study and marker development for bacterial wilt resistance in tomato (Solanum lycopersicum L.). Scientia Horticulturae 110418. https://doi.org/10.1016/j.scienta.2021.110418

Paez-Garcia A, Motes CM, Scheible WR, Chen R, Blancaflor EB, Monteros MJ (2015) Roots traits and phenotyping strategies for plant improvement. Plants 4:334-355. http://doi.org/10.3390/plants4020334

Robbins NE, Dinneny JR (2015) The divining root: Moisture-driven responses of roots at the micro-and macro-scale. Journal of. Experimental Botany 2015, http://doi.org/10.1093/jxb/eru496

Rodriguez M, Scintu A, Posadinu CM, Xu Y, Nguyen CV, Sun H, Bitocchi E, Bellucci E, Papa R, Fei Z, Giovannoni JJ, Rau D, Attene G (2020) GWAS based on rna-seq snps and high-throughput phenotyping combined with climatic data highlights the reservoir of valuable genetic diversity in regional tomato landraces. Genes 11(11):1387. https://doi.org/10.3390/genes11111387

Segura V, Vilhjálmsson B, Platt A, Korte A, Seren U, Long Q, Nordborg M (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nature Genetics 44:825–830. https://doi.org/10.1038/ng.2314

Sim S-C, Surstewitz G, Plieske J, Ganal MW, Van Deynze A, Stoffel K, Hamilton J, Buell CR, Zarka D, Douches DS, Francis DM (2012a). Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. PLoS ONE 7(7):e40563. https://doi.org/10.1371/journal.pone.0040563

Sim S-C, Van Deynze A, Stoffel K, Douches DS, Zarka D, Ganal MW, Chetelat RT, Hutton SF, Scott JW, Gardner RG, Panthee DR, Mutschelr M, Myers JR, Francis DM (2012b) High-density SNP genotyping of tomato (Solanum lycopersicum L.) reveals patterns of genetic variation due to breeding. PLoS ONE 7(9): e45520. https://doi.org/10.1371/journal.pone.0045520

Singh H, Kumar P, Chaudhari S, Edelstein M (2017) Tomato grafting: a global perspective. HortScience 52 (10):1328–1336. https://doi.org/10.21273/HORTSCI11996-17

Suchoff DH, Gunter CC, Louws FJ (2017) Comparative analysis of root system morphology in tomato rootstock. Hortechnology 27(3):319-324. http://doi.org/10.21273/HORTTECH03654-17

The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635–641. https://doi.org/10.1038/nature11119

Venema JH, Dijk BE, Bax JM, Van Hasselt PR, Elzenga JTM (2008) Grafting tomato (Solanum lycopersicum) onto the rootstock of a high-altitude accession of Solanum habrochaites improves suboptimal-temperature tolerance. Environmental and Experimental Botany 63:359–367. https://doi.org/10.1016/j.envexpbot.2007.12.015

Xie L, Klein P, Crosby, K (2019) A genotyping-by-sequencing single nucleotide polymorphism–based map and genetic analysis of root traits in an interspecific tomato population. Journal of American Society of Horticultural Science 144(6):394-404. https://doi.org/10.21273/JASHS04565-19

Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, Yuan X, Zhu M, Zhao S, Li X, Liu X (2021) rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. Genomics Proteomics Bioinformatics 19(4):619-628. https://doi.org/10.1016/j.gpb.2020.10.007

Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. Current Opinion in Biotechnology 17:155–160. https://doi.org/10.1016/j.copbio.2006.02.003

Yu J, Pressoir G, Briggs W et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38:203–208. https://doi.org/10.1038/ng1

Supplementary Information

Supplementary Table 1. Tomato (*Solanum lycopersicum* L.) diversity panel description and its occurrence on all trials

Panel number	USP entry	Entry	Varieties	Experiment		
				1	2	3
1	USP001		Olena Ukrainien	Present	Present	Present
2	USP002		Pusa Ruby	Present	Present	Present
3	USP003		Early cherry	Present	Present	Present
4	USP004		Sebastopol	Present	Present	Present
5	USP005		Cereja FC ("Perinha")	Present	Present	Present
6	USP006		Santa Cruz Kada Gigante	Present	Present	Present
7	USP007R		Corrogo	Present	Present	Present
8	USP007M		Corrogo	Present	Present	Present
9	USP008		San Marzano	Present	Present	Present
10	USP009		Yoshimatsu-L3	Present	Present	Present
11	USP010		Azure	Present	Present	Present
12	USP011		Tecoh Tepee	Present	Present	Present
13	USP012		Santa Clara	Present	Present	Present
14	USP013		High Country	Present	Present	Present
15	USP014		Immune	Present	Present	Present
16	USP015		Indigo Rose	Present	Present	Present
17	USP016		Banana Legs	Present	Present	Present
18	USP017		Calabash Rouge	Present	Present	Present
19	USP018		Coeur de Boeuf Jaune	Present	Present	Present
20	USP019		Tasty Evergreen	Present	Present	Present
21	USP020		Tropic Two Orders	Present	Present	Present
22	USP021		Géante D´Orenburg	Present	Present	Present
23	USP022		Solymari	Present	Present	Present
24	USP023		Tomate Laranja Salada (Top Seed)	Present	Present	Present
25	USP024		White Wonder	Present	Present	Present
26	USP025		Pêache Rouge	Present	Present	Present
27	USP026A		Prize of the Trial	Absent	Present	Present
28	USP026V		Prize of the Trial	Absent	Present	Present
29	USP027		Tomate Cereja Laranja (Top Seed)	Present	Present	Present
30	USP028		Olirose de St Domingue	Present	Present	Present
31	USP029		Persimmon	Present	Present	Present
32	USP030		IPA-6	Present	Present	Present
33	USP031		Santa Adélia	Present	Present	Present
34	USP032		Amalia	Present	Present	Present
35	USP033		INCA-945	Present	Present	Present
36	USP034		Mara	Present	Present	Present
39	USP038	LA4393		Absent	Present	Present
40	USP039	LA2414	Cal Ace	Present	Present	Present
41	USP040	LA4348		Present	Present	Present
42	USP041	LA2661	Nagcarlang	Present	Absent	Present
43	USP042	LA2706	Moneymaker	Present	Present	Present

44	USP043	LA4104		Present	Present	Present
45	USP044	LA1787		Absent	Present	Present
46	USP045	LA4285		Present	Absent	Present
47	USP046	LA2086		Present	Present	Present
49	USP048	LA4451	Black Cherry	Absent	Present	Present
50	USP049	LA4449	Black Plum	Present	Present	Present
51	USP050	LA2413		Present	Present	Present
52	USP051	LA3667		Present	Present	Present
53	USP052	LA2644		Absent	Present	Present
54	USP053	LA3736		Present	Present	Present
56	USP055	LA2830		Absent	Present	Present
57	USP056	LA3847	NC HS-1	Present	Present	Absent
58	USP057	LA3043		Present	Present	Present
59	USP058	LA3471		Absent	Present	Present
60	USP059	LA0806		Present	Present	Present
61	USP060	LA4425		Absent	Absent	Present
62	USP061	LA2399		Present	Present	Present
63	USP062	LA3273		Present	Present	Present
65	USP064	LA2445		Present	Present	Present
66	USP065	LA2531A		Present	Present	Present
68	USP067	LA2939		Present	Present	Present
69	USP068		Balkonstar	Absent	Present	Present
70	USP069		Saint Pierre	Present	Present	Present
71	USP070		Motelle (PCTM)	Present	Present	Present
72	USP072	LA4026		Present	Present	Present
73	USP073	LA3151		Present	Present	Present
74	USP074	LA3475	M82	Present	Present	Present
75	USP082	CNPH0527		Absent	Present	Present
76	USP083	CNPH0511		Present	Present	Present
77	USP084	CNPH0263		Present	Present	Present
78	USP085	CNPH0616		Present	Present	Present
79	USP086	CNPH0357		Present	Present	Present
80	USP087	CNPH0083		Present	Present	Present
81	USP088	CNPH0082		Absent	Present	Present
82	USP089	CNPH0633		Absent	Present	Present
83	USP090	CNPH0899		Absent	Present	Present
84	USP091	CNPH0422		Present	Present	Present
85	USP092	CNPH0883		Present	Present	Present
87	USP094	CNPH0529		Present	Present	Present
88	USP095	CNPH0650		Present	Present	Present
90	USP097	CNPH0500		Present	Present	Present
91	USP098	CNPH0443		Present	Present	Present
92	USP099	CNPH0525		Absent	Present	Present
93	USP100	CNPH1137		Present	Present	Present
94	USP101	CNPH1218		Present	Present	Present
95	USP102	CNPH1200		Present	Present	Present
96	USP103	CNPH0512		Present	Present	Present
97	USP104	CNPH0528		Present	Present	Present

98	USP105	CNPH0080		Present	Present	Present
99	USP106	CNPH0523		Present	Present	Present
100	USP107	CNPH1100		Present	Present	Present
101	USP108	CNPH0846		Absent	Present	Present
102	USP109	CNPH0357		Present	Present	Present
103	USP110A	CNPH1138R		Present	Present	Absent
104	USP110B	CNPH1138C		Present	Present	Present
105	USP111	CNPH0920		Present	Present	Absent
106	USP112	CNPH0442		Present	Present	Present
107	USP113	CNPH0923		Present	Present	Present
108	USP114	CNPH0390		Present	Present	Present
109	USP115	CNPH0081		Present	Present	Present
110	USP116	CNPH0635		Present	Present	Present
111	USP117	CNPH0079		Present	Present	Present
113	USP119	CNPH0634		Present	Present	Present
114	USP120		Mangymakuh	Present	Present	Present
115	USP121		Idyll	Present	Present	Present
117	USP123		Hellfrucht	Present	Present	Present
118	USP124		Tomate Gaúcho laranja (top seed)	Present	Present	Present
119	USP125		Santa Cruz Kada (Paulista) - ISLA	Present	Present	Present
120	USP126	LA0330		Present	Present	Present
121	USP127		Lingüiça Polonesa	Present	Present	Present
122	USP128		Tomato tree	Present	Present	Present
124	USP130		Des Andes	Present	Present	Present
125	USP131		Prune Noire	Present	Present	Present
126	USP132		Beauté Blanche	Present	Present	Present
127	USP133		Striped Cavern	Present	Present	Present
128	USP134		Poire Rouge	Present	Present	Present
129	USP135		Tigerella	Present	Present	Present
130	USP136		De Barao Gold	Present	Present	Present
131	USP137		Noire de Crimée	Present	Present	Present
132	USP138		Orange Queen	Present	Present	Present
133	USP139		Podland Pink	Present	Present	Present
134	USP140		Black Prince	Present	Present	Present
135	USP141		Green Sausage	Present	Present	Present
136	USP142		Eva's Purple Ball	Present	Present	Present
137	USP143		Coeur de Boeuf Orange	Present	Present	Present
138	USP144		Burbank	Present	Present	Present
139	USP145		Ester Hess Yellow	Present	Present	Present
140	USP146		Peasant	Present	Present	Present
142	USP148	LA0791	Long Jhon	Present	Absent	Absent
143	USP149	FAP0002		Present	Present	Present
144	USP150	FAP0001		Absent	Present	Absent
145	USP153	FAP0005		Absent	Present	Present
146	USP154		Tomate Italiano para molhos (Topseed)	Present	Present	Present
147	USP155		Tomate Pêra Amarelo (Topseed)	Present	Present	Present

148	USP156	PNZ006	TSW-10 CNPH	Present	Present	Present
149	USP157	PNZ009	AG 45 (Ohio 8145)	Present	Present	Present
151	USP158B	PNZ102	Rotam-4	Present	Present	Present
152	USP159A	PNZ100	Stevens	Present	Present	Present
153	USP160	PNZ103	Rodade	Present	Absent	Present
154	USP161	PNZ223	Vietnamita-BWR	Present	Present	Present
155	USP162		PCV-01	Present	Present	Present
156	USP163		Hawaii 7996	Present	Present	Present
161	GT0012		CAL J, LOTE: 478 1041	Present	Present	Present
163	GT0027		SEED 062 CAMARILLO	Present	Absent	Absent
164	GT0028		SEED 062 PETOEARLY	Present	Present	Present
166	GT0033		FPA-4	Present	Present	Present
170	GT0056		NEMADORO	Present	Present	Present
177	GT0084		EARLY CASCADE 312	Present	Present	Present
179	GT0087		SAINT PIERRE	Present	Present	Present
180	GT0092		VF 90	Present	Present	Present
186	GT0102		CARAÍBA	Present	Present	Present
188	GT0104		SEED VF 198	Present	Present	Present
190	GT0106		SEED NAPOLI VF	Present	Present	Present
194	GT0117		PETO 13	Present	Present	Present
195	GT0127		FARLYSTONE, LOTE: 5921006	Present	Present	Present
197	GT0130		PETOEARLY, LOTE: 6721067	Present	Present	Present
198	GT0131		RIO GRANDE, LOTE: 7457042	Present	Present	Present
199	GT0135		SANTA CRUZ KADA	Present	Present	Present
200	GT0137		PIRACÓ	Absent	Present	Present
201	GT0138		PIRACÓ PROJETO III	Absent	Present	Present
205	GT0150		SANTA CRUZ ANODA	Present	Present	Present
206	GT0162		CASTLONG, LOTE: 2237-69	Present	Present	Present
207	GT0163		CASTLEBLOCK, LOTE: 2299-245	Present	Present	Present
208	GT0164		CASTLESTAR EHV, LOTE: 2627	Absent	Present	Present
209	GT0165		UC - 82-A, LOTE: 2556	Absent	Present	Present
210	GT0166		CASTLEMOR IMP., LOTE: 2416-15	Absent	Present	Present
211	GT0167		CALYPSO, LOTE: 2697	Absent	Present	Present
214			Tomate Cereja Samambaia (Tradicional Hortaliças)	Absent	Present	Present
215			Santa Clara (Hortec)	Absent	Present	Present
216			San Marzano PCS 9/2001 (LG)	Absent	Present	Present
217			Minitomate Acesso 21 - IAC (melhor)	Present	Present	Present
219			Cuor di bue (Hortus Sementi)	Present	Present	Present
220			Saint Pierre (Hortus Sementi)	Present	Present	Present
221			Santa Cruz Kada Gigante (Top Seed)	Present	Present	Present
222			San Marzano (Landen)	Present	Present	Present
223		IAC-1689	Homesweet Heirloom	Absent	Present	Absent
224		IAC-1693	Aussie Heirloom	Absent	Absent	Present

225	IAC-1691	Black Prince Heirloom	Absent	Present	Present
226	IAC-1692	Costoluto genovese (multiplicação)	Absent	Present	Present
227	IAC-1692	Costoluto genovese	Present	Present	Present
228	IAC-1696	Caspian Pink Heirloom	Present	Present	Present
229	IAC-1695	Purple Russian Heirloom	Present	Present	Present
230	IAC-1697	Cherokee Purple Heirloom	Present	Present	Present
231	IAC-1694	Goliath Tomato	Present	Present	Present
232	IAC-1693	Aussie Tomato	Present	Present	Present
233	IAC-1690	Black Krim Heirloom	Absent	Present	Present
234		Tomate Santa Cruz + Brinde	Present	Present	Present
235	IAC-1612	Peacevine	Present	Present	Present
236	IAC-1613	S.T. Pierre	Present	Present	Present
237	IAC-1614	Money Marker	Absent	Present	Present
239	IAC-1615	Matina	Present	Present	Present
240	IAC-1617	Arkansas Trveller	Present	Present	Present
241	IAC-1619	Stupice	Present	Present	Present
242	IAC-1618	Bruno Simonetti	Present	Present	Present
243	IAC-1605	Uco Plata	Present	Present	Present
244	IAC-1606	TSW-10	Present	Present	Present
245	IAC-1607	Mars	Present	Present	Present
246	IAC-1608	ILDI Naranja	Absent	Present	Present
247	IAC-1610	Black Plum Paste	Present	Present	Present
248	IAC-1611	Cradwich	Absent	Present	Present
249	CGT-01	Linha de origem desconhecida	Present	Present	Present
250	CGT-03	Missouri 91	Present	Present	Present
251	CGT-04	Missouri 93	Present	Present	Present
253	CGT-06	Romitel	Present	Present	Present
255	CGT-14	Rio Fuego	Present	Present	Present
257	CGT-21-1	Hirol (Vermelho)	Absent	Present	Present
259	CGT-22	Rio Grande	Absent	Present	Present
260	CGT-23	UC 105	Present	Present	Present
261	CGT-25	Rotec	Present	Present	Present
262	CGT-26	Rossol	Present	Present	Present
263	CGT-27	Rio Fuego	Present	Present	Present
264	CGT-28	Hoffit	Present	Present	Present
265	CGT-29	Mecline	Absent	Present	Present
266	CGT-32	Heinz 1548	Absent	Present	Present
267	CGT-33	M 204	Absent	Present	Present
269	CGT-35-1	Dela Plata	Absent	Present	Present
270	CGT-36	M 145 (Saladette)	Absent	Present	Absent
271	CGT-37	PU 7328 (bu)	Present	Present	Present
272	CGT-38	Romitel	Present	Present	Present
273	CGT-43	Santa Adélia Super	Absent	Present	Present
275	CGT-45	Roquesso (Ag. 591)	Present	Present	Present
276	CGT-47	Santa Clara Albino	Absent	Present	Present
277	CGT-48	Olho Roxo	Present	Present	Present
279	CGT-52	Floradade (F2 Ve Sm)	Absent	Present	Present
280	CGT-54	Calypso	Absent	Present	Present

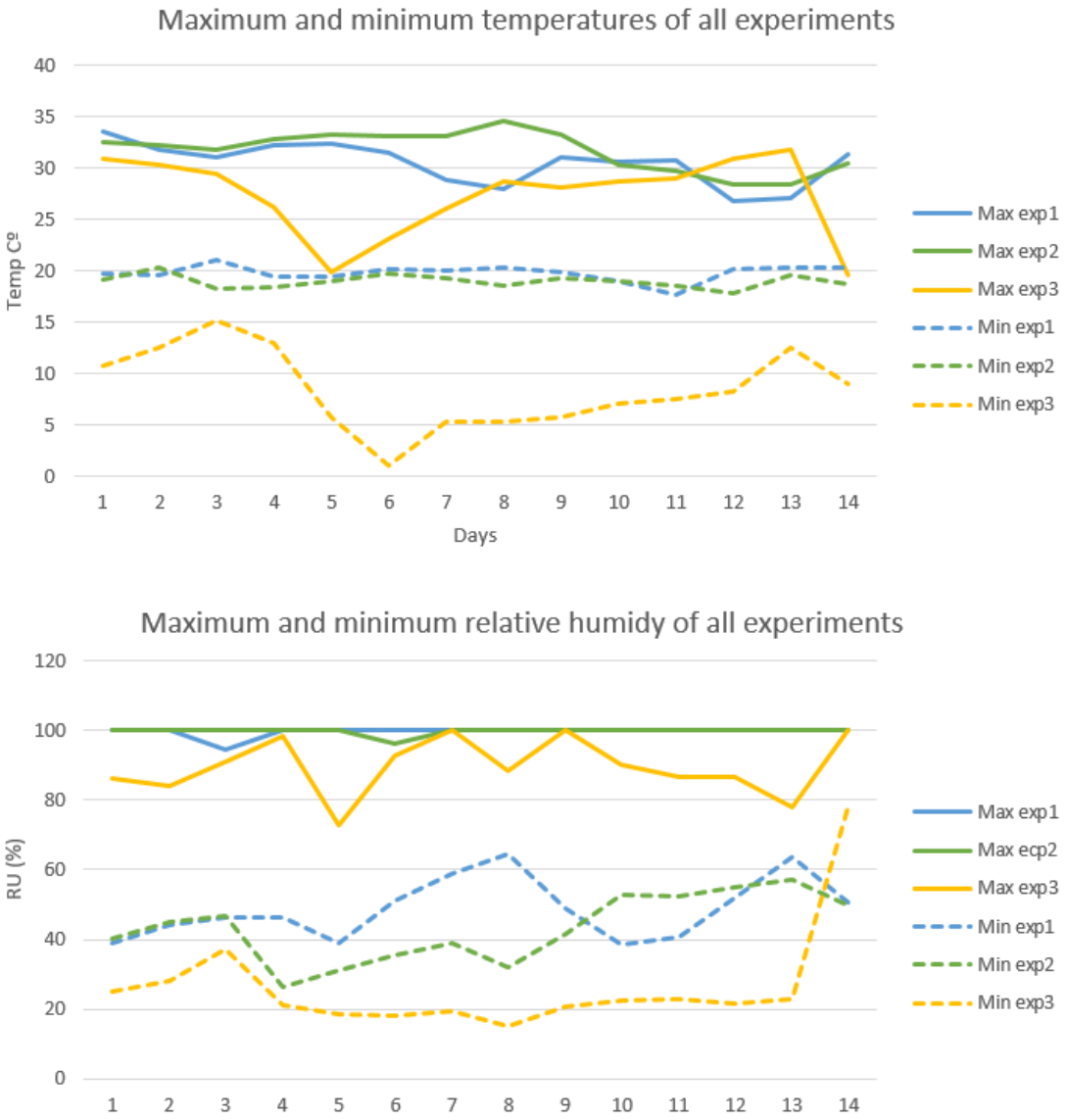
282	CGT-57	Angela Hiper	Absent	Present	Present
283	CGT-58	Europeel	Absent	Present	Present
284	CGT-62	Missouri	Present	Present	Present
285	CGT-64	Príncipe Gigante Ag. 590	Present	Present	Present
286	CGT-67	Yoshimatsu-4 (INPA)	Present	Present	Present
287	CGT-68	Hoffit	Present	Present	Present
288	CGT-72	Rimone	Present	Present	Present
289	CGT-73	Licapal	Present	Present	Present
291	CGT-77	Romitel	Absent	Present	Present
292	CGT-84	Santo Antonio	Present	Present	Present
293	CGT-85	Motelle	Absent	Present	Present
294	CGT-88	Tropicana	Absent	Present	Present
295	CGT-90	Rotam-4	Present	Present	Present
296	CGT-91	Rodade	Present	Absent	Present
297	CGT-97	Olho Roxo Melhorado	Present	Present	Present
298	CGT-98	P213 - PA (Tropicana)	Present	Present	Present
299		Tomate Ferraz Ipa-8 (Hortivale)	Present	Present	Present
326	GT0048	PAKMOR (VF)	Absent	Present	Present
336	GT0077	HEINZ 1350	Present	Present	Present
337	GT0085	SUPERMARKET 87170	Present	Present	Present
343	GT0101	MECANO	Absent	Present	Present
348	GT0118	TROPIC	Present	Present	Present
359	GT0181	"VERDINHO"	Present	Present	Present

Supplementary Figure 1. a) Hydroponic pools used for rootstock evaluations. The pools were built using concrete blocks covered with a layer of non-woven geotextile and a second layer of black plastic. Two recirculating pumps coupled to hoses that ran the entire pool length were used to keep the solution homogeneous; b) The seedlings were suspended over the solution in Styrofoam plates with sponge sheets; c) The seedlings were spaced eight centimeters apart, with three plants per plot, and two treatments per line; d) Seedlings after fourteen days

Supplementary Table 2. Modified Hoagland and Arnon solution used at 50% power in this research. The

Macronutrients	grams.1000 L^{-1}
$Ca(NO_3)_2.4H_2O$	118
KNO_3	50,5
$MgSO_4.7H_2O$	49,2
KH_2PO_4	13,6
Micronutrients	
H_3BO_3	7,2
$MnCl_2.4H_2O$	4,52
$ZnSO_4.7H_2O$	0,55
$CuSO_4.5H_2O$	0,2
$NaMoO_4$	0,225
Fe-EDTA	
EDTA-Na_2	33,2
NaOH	3,65
$FeSO_4.7H_2O$	25

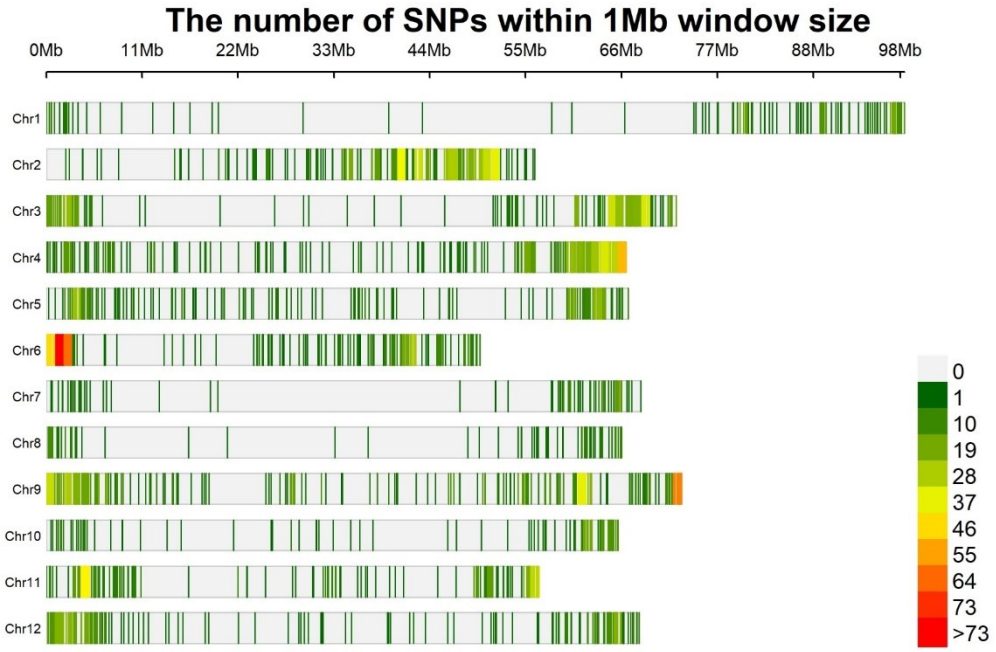
quantities below were used for each 1000 liters of nutrient solution



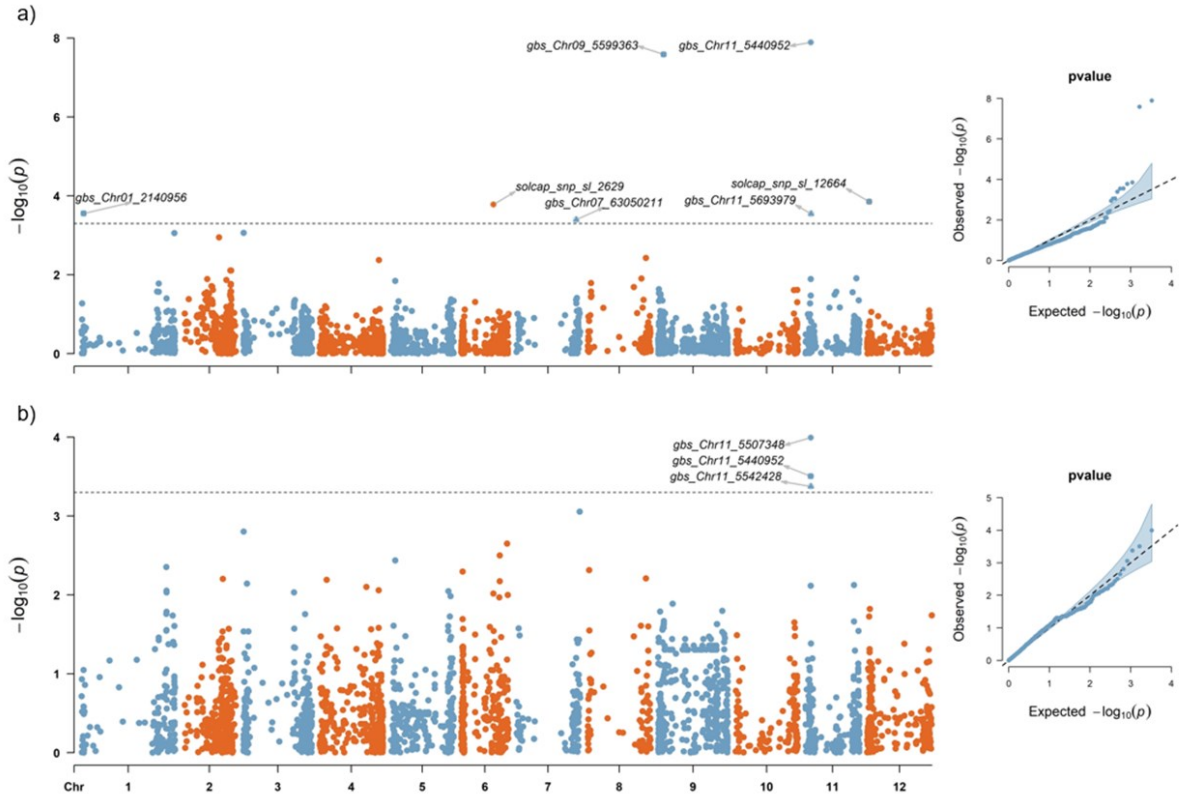
Supplementary Figure 2. Temperature and humidy data for all trials

Supplementary Table 3. Number of markers per chromosome

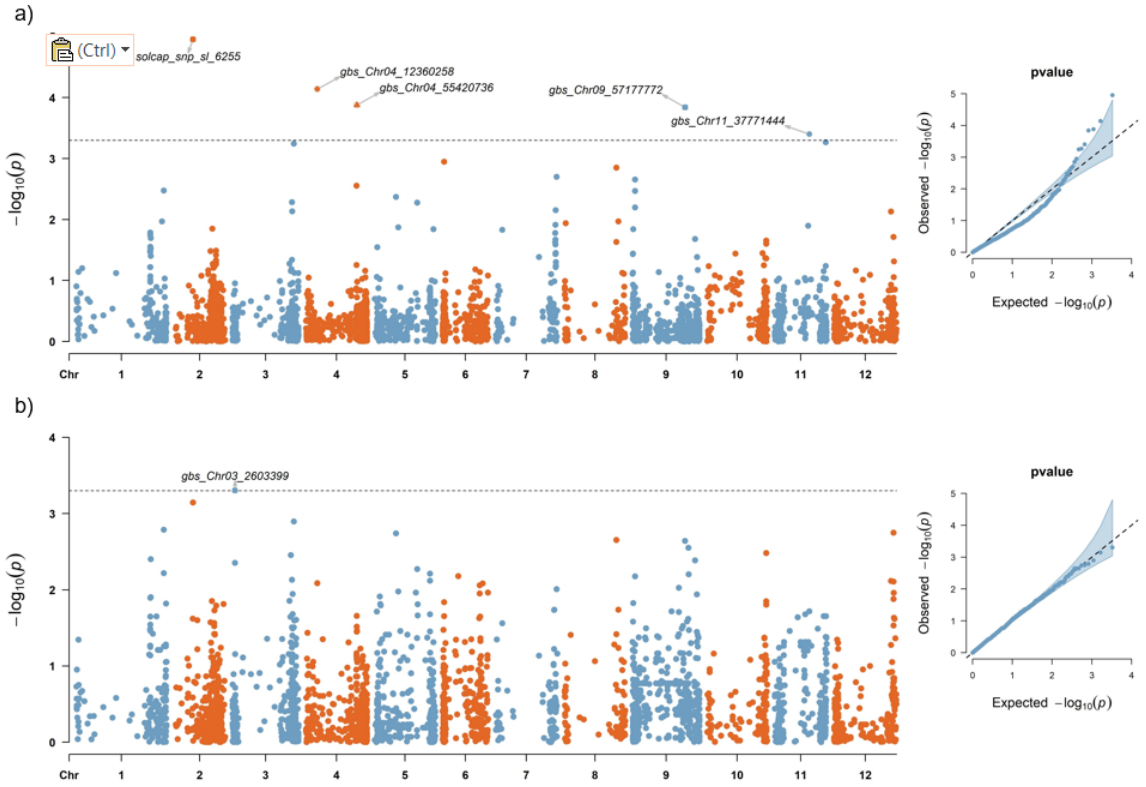
Chromosome	Number of markers
1	150
2	431
3	309
4	417
5	272
6	363
7	90
8	86
9	556
10	123
11	254
12	254
Total	3305
Média	275,4



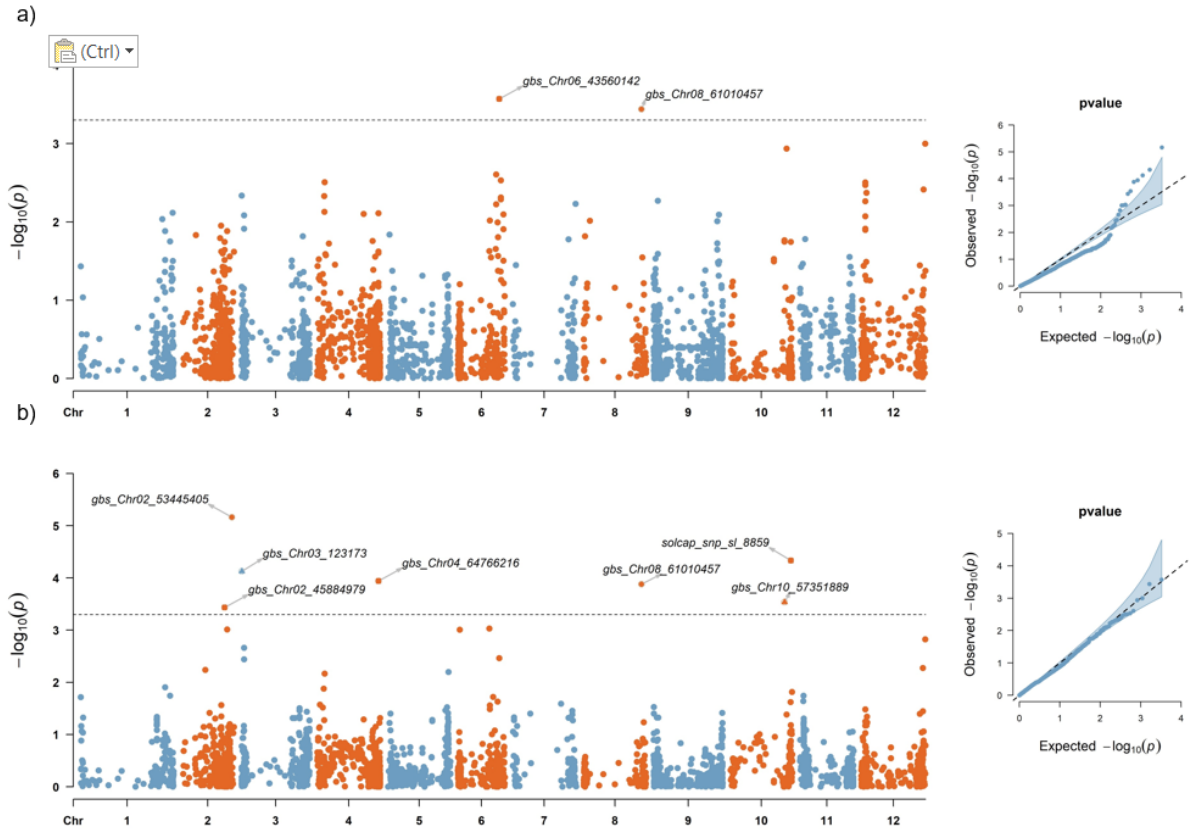
Supplementary Figure 3. Genome coverage of 3.305 SNPs markers used for GWAS analysis



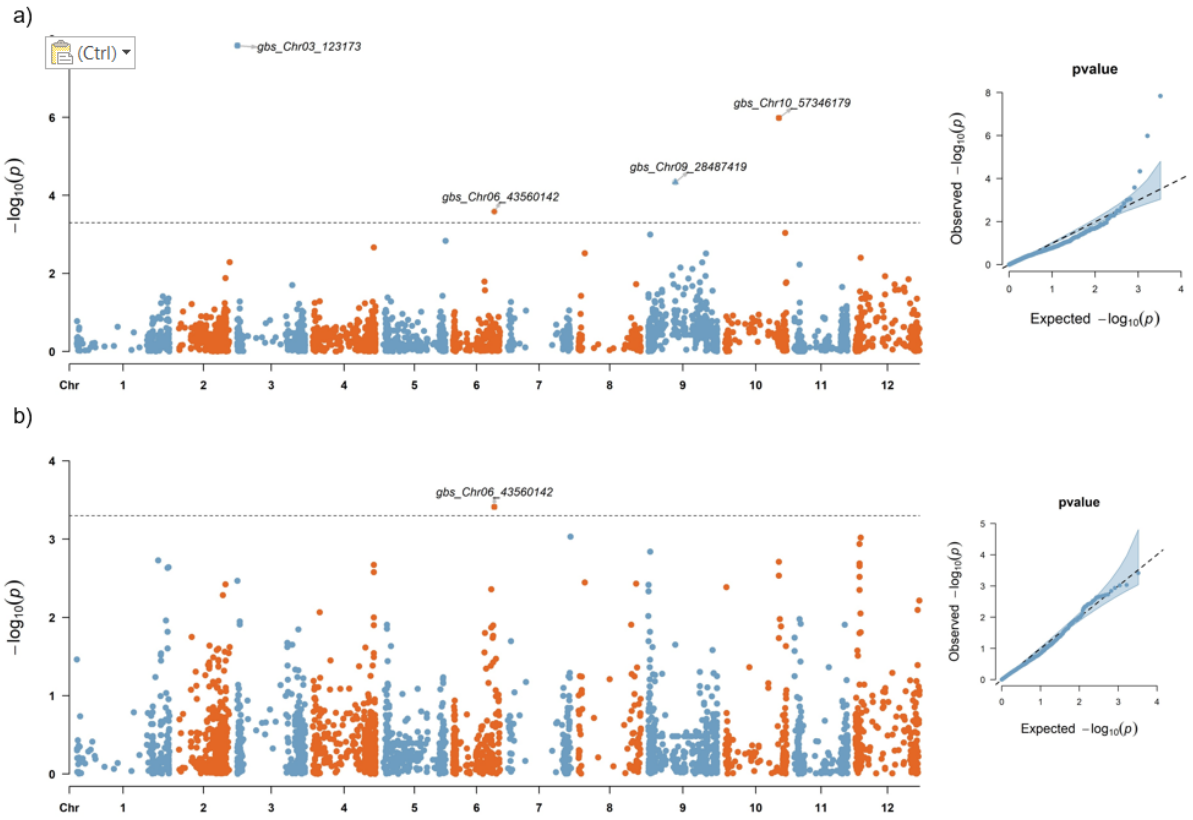
Supplementary Figure 4. Manhattan and QQ plots for shoot height. (a) FarmCPU model; (b) MLMM model. The grey dotted line on Manhattan plots represent the defined cut ($p > 0.0005$)



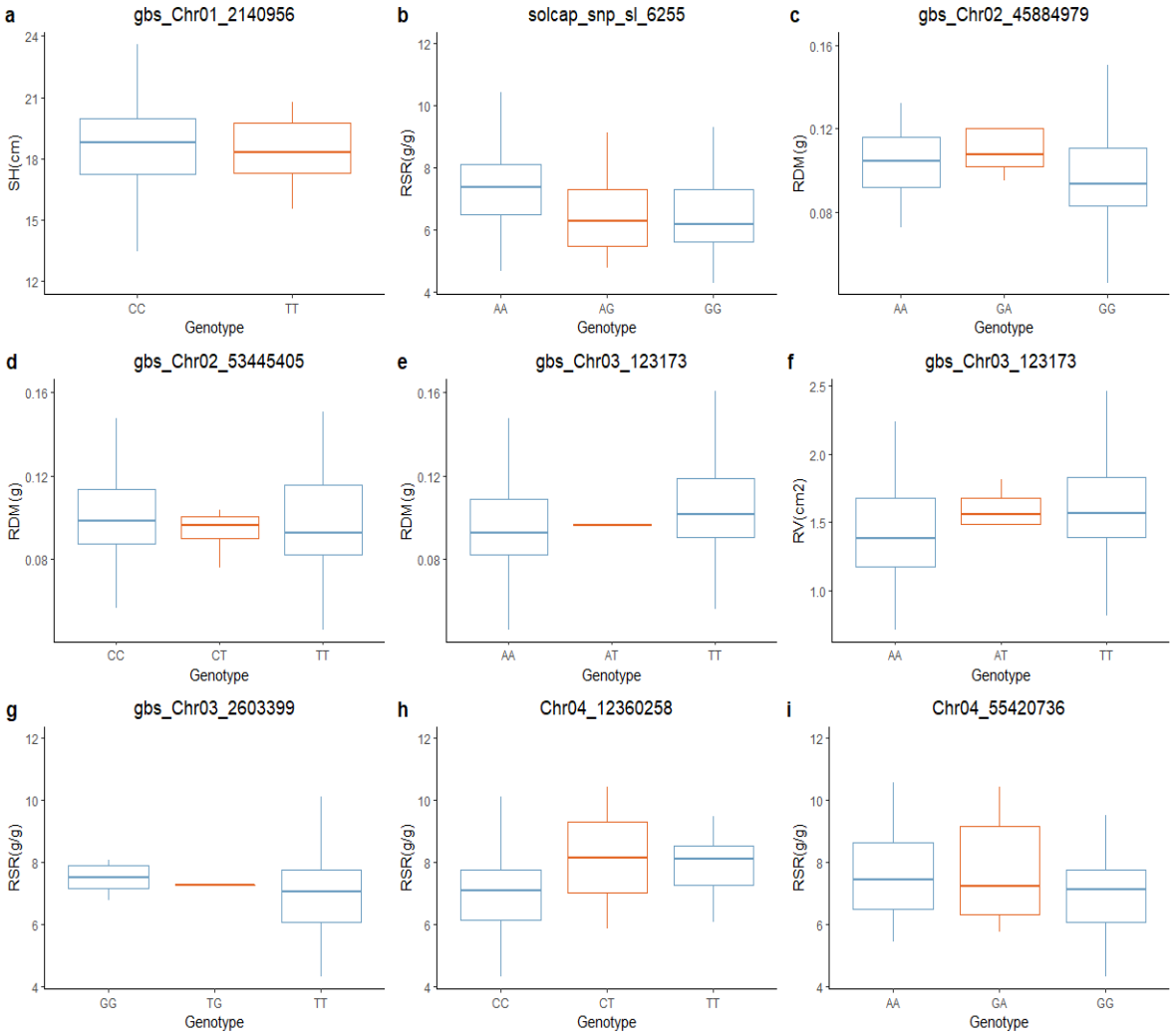
Supplementary Figure 5. Manhattan and QQ plots for ratio of shoot/root. (a) FarmCPU model; (b) MLMM model. The grey dotted line on Manhattan plots represent the defined cut ($p > 0.0005$)



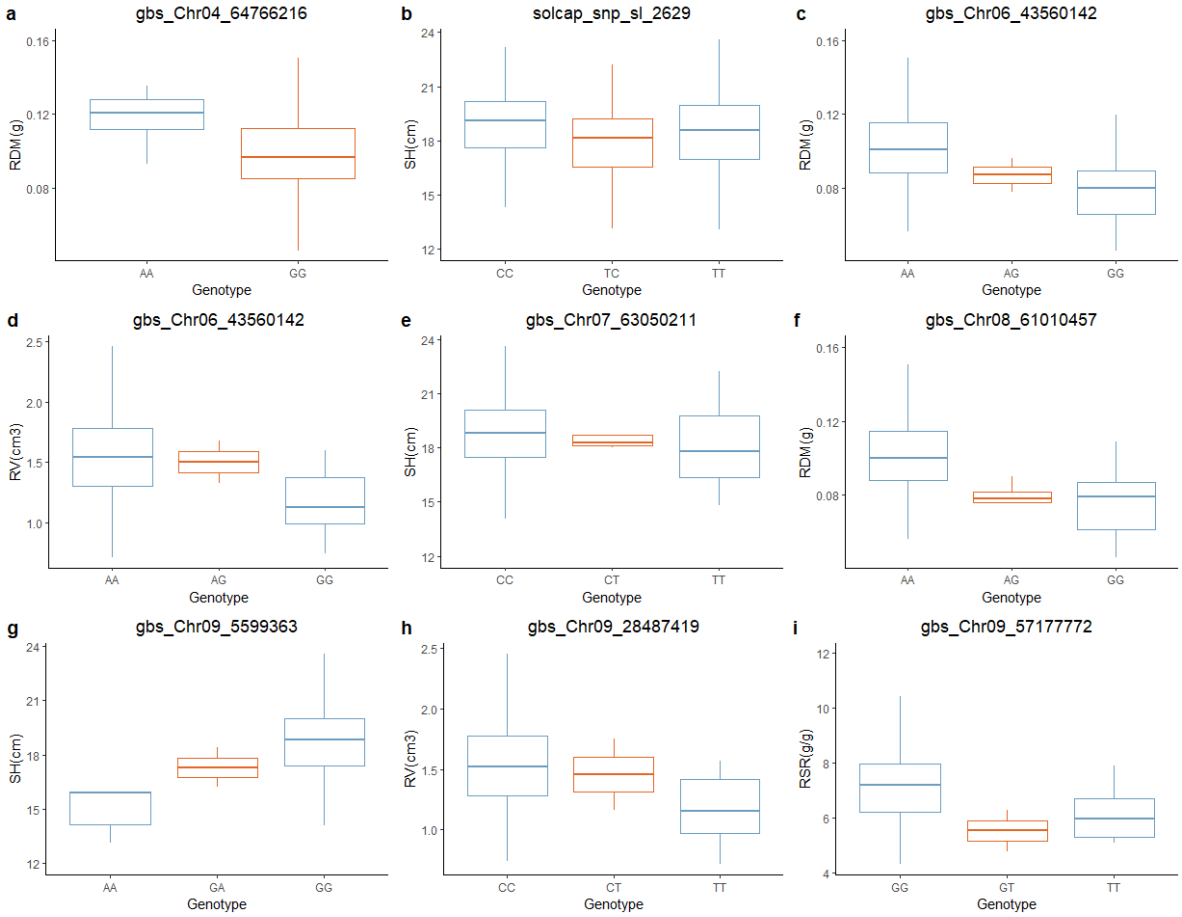
Supplementary Figure 6. Manhattan and QQ plots for root dry mass. (a) FarmCPU model; (b) MLMM model. The grey dotted line on Manhattan plots represent the defined cut ($p > 0.0005$)



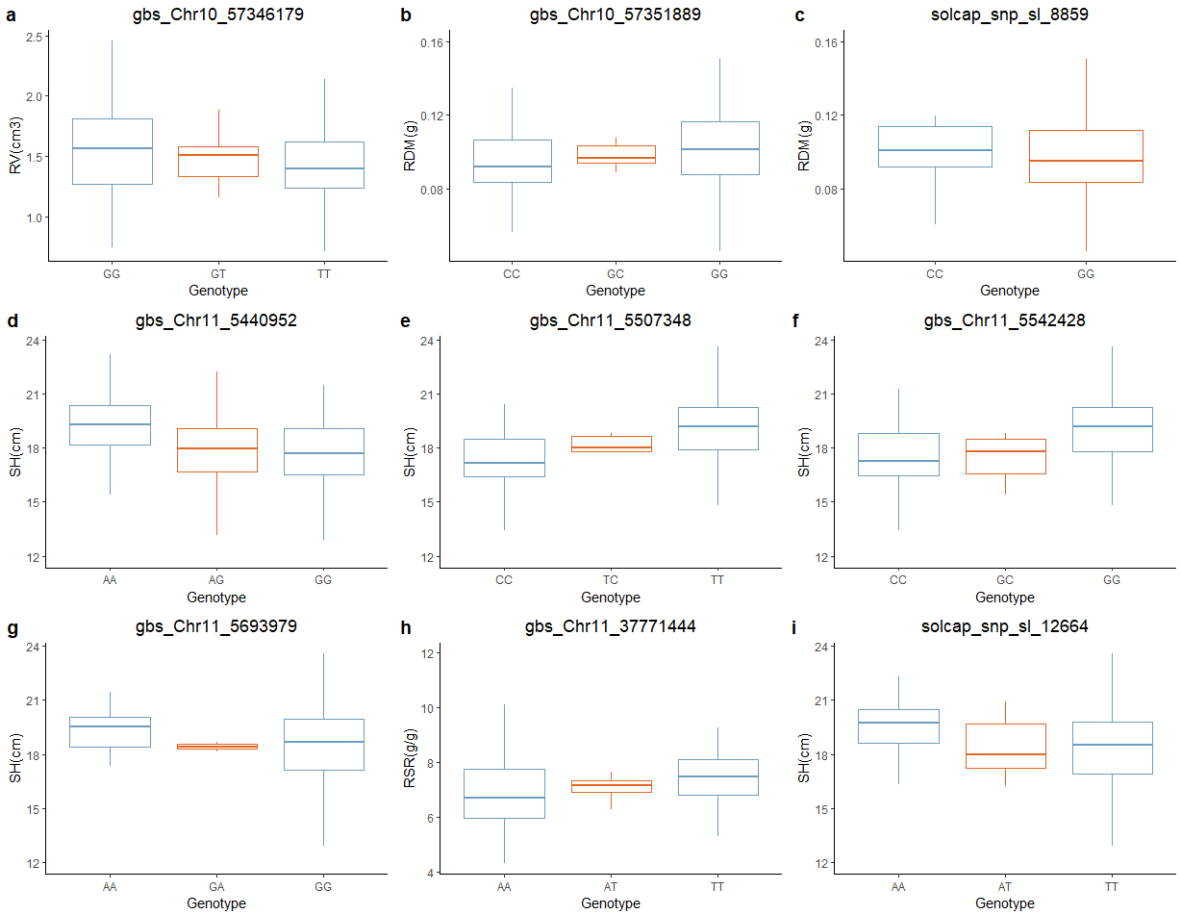
Supplementary Figure 7. Manhattan and QQ plots for root volume. (a) FarmCPU model; (b) MLMM model. The grey dotted line on Manhattan plots represent the defined cut ($p > 0.0005$)



Supplementary Figure 8. Boxplot visualization of the effect of allele substitution for detected QTNs of shoot height (SH), ratio of shoot/root (RSR), root dry mass (RDM), and root volume (RV)



Supplementary Figure 9. Boxplot visualization of the effect of allele substitution for detected QTNs of shoot height (SH), ratio of shoot/root (RSR), root dry mass (RDM), and root volume (RV)



Supplementary Figure 10. Boxplot visualization of the effect of allele substitution for detected QTNs of shoot height (SH), ratio of shoot/root (RSR), root dry mass (RDM), and root volume (RV)

3. IMPROVEMENT OF GENOMIC SELECTION MODELS FOR YIELD IN PROCESSING TOMATO THROUGH INCORPORATION OF LINKAGE AND GENE EFFECTS

Abstract

For specialty crops such as tomato, there is a lack of genetic information regarding important traits such as yield. This situation is due to many factors including the lack of a tradition for collecting quantitative data, limited resources, and small population sizes in breeding programs. Although genomic selection is now a widely used tool in plant breeding, its application to improving horticultural crops has lagged behind those in staple grain crops. Here, we use genomic analysis strategies in several processing tomato populations to search for yield-related quantitative trait loci (QTL) and incorporate this information into prediction and selection models. We first performed genome-wide association studies in a nested recombinant inbred line (RIL) population and composite interval mapping in a bi-parental RIL to identify genomic regions associated with yield and fruit weight. A yield-associated QTL on chromosome five was identified which explained 29% of the variation. Validation in subsequent generations suggested the QTL had dominant gene action. The QTL is associated with an introgression from Hawaii 7998 and contains a novel allele of the *self-pruning* gene (*Sp*5), a potential candidate gene, associated with both yield and plant architecture. We tested genomic selection models for yield and fruit size based on inbred line populations through cross-validation and the ability to predict an independent set of hybrids. Positional information on yield and fruit weight QLTs and estimates of gene action were added to these models and further tested. Inclusion of positional information and dominance for the QTL on chromosome 5 improved prediction accuracies in genomic selection models.

Keywords: *Solanum lycopersicum*, genomic prediction, large effect QTL, complex traits, yield genetics

3.1. Introduction

Crop yield is thought to be a complex trait influenced by many genetic and environmental factors. Despite the importance of yield, a description of the underlying genetic control is lacking for many crop plants. This paucity of information is ascribed to the complex genetic nature of yield and the underlying belief that quantitative agricultural traits are controlled by many genes of small effect, referred to as the "infinitesimal model" (Barton et al., 2017). However, studies of complex quantitative traits often demonstrate kurtosis suggesting that they are conditioned by both genes of small effect and genes of moderate to large effect (for example: Edwards et al., 1987 and Hayes and Goddered, 2001). In some crops such as tomato (*Solanum lycopersicum* L.), there is a lack of studies investigating yield in target populations that are relevant to the community engaged in crop improvement. Although several quantitative trait loci (QTLs) have been reported over the 12 tomato chromosomes, these studies are conducted in wide crosses and there is limited concordance across studies (Foolad, 2007; Hernández-Bautista et al., 2015).

An experimental limitation to genetic studies of yield in breeding relevant populations has been a lack of polymorphic markers. Despite a long history of molecular breeding in tomato, an emphasis on wide crosses led to a dearth of tools for use in cultivated populations. This situation changed shortly after the emergence of the first reference genome for tomato (The Tomato Genome Consortium, 2012). The availability of next-generation sequencing platforms (Patel et al., 2015), a reduction in genotyping costs (Wetterstrand, 2022), and the emergence of

highly parallel genotyping platforms allowed the applied research community to develop markers appropriate for dissecting complex traits and selection within breeding populations (Sim et al., 2012b, Sim et al., 2015).

Establishing linkage between protein or DNA sequence polymorphism and variation quantitative trait loci (QTL) opens several strategies for marker-assisted selection (MAS). In tomato, MAS for selection of major resistance loci has a history dating back more than fifty years (Rick and Fobes, 1974; Bolkan et al., 1983; Williamson and Colwell, 1991). More recently, the increased availability of markers polymorphic in breeding populations has facilitated forward selection coupled to background-genome selection (Orchard et al., 2021; Bernal et al., 2020). Finally, approaches to selection that use sequence polymorphism to develop predictive models, genomic selection, have demonstrated potential. Genomic selection (GS), originally proposed by Meuwissen et al. (2001), uses all available markers to estimate breeding values and predict performance. This approach is based on the infinitesimal model and the simplifying assumption that trait value is a result of the linear combination of additive genetic and nongenetic sources of variation (Fisher, 1919). GS is becoming a powerful tool for implementation in modern breeding programs (Bassi et al., 2016) and has shown promise in tomato for plant disease resistance (Liabeuf et al., 2018), fruit size (Hernández-Bautista et al., 2016), and fruit quality (Duangjit et al., 2016, Yamamoto et al., 2017). Predictive models that incorporate knowledge of linkage and non-additive gene effects have shown promise (Liabeuf et al., 2018).

In this study, we present evidence for a yield-related QTL on chromosome five of tomato. This QTL is inherited from inbred line OH987034 as part of an introgression from Hawaii 7998 and includes a novel allele of the *self-pruning* gene (*Sp*5). We investigated the effects of an allele substitution at this QTL, and demonstrate dominant gene action. We also investigated evidence for QTL fruit size, another component of yield. We aimed to test whether accounting for dominance effects and modeling yield-related QTL as a fixed effect could increase prediction of hybrid performance. We demonstrate that incorporating knowledge of linkage and gene action in a genomic prediction framework improves GS model performance.

3.2. Materials and Methods

Plant materials and field/greenhouse evaluations

This work consisted of three stages and different populations were used in each (Figure 1). First, we performed QTL mapping using a nested recombinant inbred line (RIL) population to identify regions associated with yield-related traits and defined chromosomal intervals associated with these. We validated QTL using subsequent generations and estimated the effects

of allele substitutions for SNPs within intervals using an independent testcross population. Finally, we developed genomic selection (GS) models using three training populations and tested prediction models through cross-validation and F1 prediction in an independent population of hybrids.

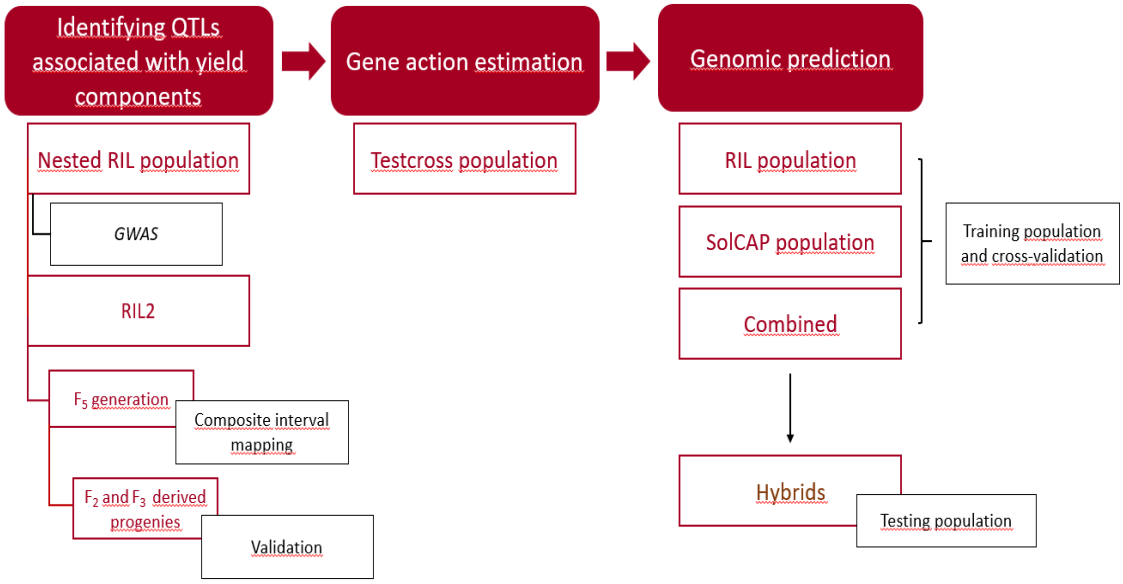


Figure 1. Graphic representation of the research stages. The nested RIL population was first analyzed using GWAS to search for regions associated with yield components. A bi-parental subset, RIL2 (OH7814 x OH987034), was then used for composite interval mapping to more precisely define the regions of the genome associated with yield components. RIL2-derived progenies were also studied to validate these QTLs and demonstrated an association with the candidate gene SP5. Gene effects were estimated for intervals associated with yield components using an independent hybrid testcross population. Three populations were then used to develop and test genomic prediction models that accounted for these QTLs and estimated gene effects. Finally, the accuracy of the models was tested against experimental data for an independent set of hybrids.

The training populations consisted of the RIL population formed from three crosses involving OH2641 x OH987034, OH7814 x OH987034, and OH2641 x OH981136 with 95, 119, and 68 fifth generation self-fertilized (F_5) progenies, respectively. A subset of the RIL population, RIL2, derived from the OH7814 x OH987034 cross, was used for the genetic mapping and QTL analysis of yield-related traits using interval mapping. OH7814 is an early-season inbred (Berry and Gould, 1983) and OH987034 is a high-yielding bacterial spot resistant inbred (Francis and Miller, 2005). The second training population was the Solanaceae Coordinated Agriculture Project (SolCAP) inbred line population which represents parent germplasm for the North American processing tomato industry as described previously (Merk et al, 2012 and Sim et al., 2012b). The 143 SolCAP inbred lines were used as a training population

for genomic prediction and selected inbreds were used as test cross parents to estimate gene action and the effect of allele substitutions for markers associated with yield related traits. SolCAP inbreds were also used to test hybrid prediction. The third training population was the combined SolCAP and RIL populations.

Both SolCAP and RIL inbred lines were evaluated in the field using an augmented experimental design (Federer et al., 1975; Lin and Poushinsky, 1983; Lin et al., 1983). We evaluated the RIL and RIL2 sub-population in an augmented design, planted at the North-Central Agriculture Research Station in Fremont, Ohio, and the Horticulture Unit 1 Research Farm in Wooster, Ohio over the course of two years (2011 and 2012) representing four environments. Row, Column, and Quadrant were variables with over-replicated checks used to account for spatial variation in three dimensions within each environment. The best linear unbiased predictor (BLUP) for each genotype and environment was estimated based on the model (Y = Genotype + Row + Column + Quadrant + Error), where all model factors were random. Analysis was performed with R software 4.1.1 (R Core Team, 2022), and functions "lmer" and "ranef" from the lme4 package (Bates et al., 2015). BLUPs across the four environments were subsequently averaged.

We used an independent testcross population to estimate gene action and the effect of allele substitutions for markers associated with yield-related traits. This population consisted of fifty-seven hybrids and was formed by crossing selections from the RIL population with inbred lines 2K1-1439, OH05-8127, OH05-8157, OH08-7460, OH8556, selected from the SolCAP population (Merk et al., 2012; Sim et al., 2012b). The testcross hybrids were evaluated in Wooster and Fremont in 2017, using a randomized complete block design with two blocks at each location. The experimental design was treated as fully random and BLUPs were averaged across the two locations.

Finally, a set of 76 F_1 hybrids was used to test the prediction ability of our models (stage three, Figure 1). These hybrids were developed by crossing 32 SolCAP lines in various combinations and were evaluated in a randomized complete block design with two replications per environment in Wooster and Fremont over two years. BLUPs were estimated in each environment and were subsequently averaged. All populations were evaluated for total yield (kg.plant^{-1} and ripe kg.plant^{-1}) and fruit weight (grams.fruit^{-1}).

DNA extraction and molecular marker genotyping

Genomic DNA was isolated from fresh, young leaf tissue using a CTAB method scaled to 96-well format (Sim et al., 2015). Single-nucleotide polymorphisms (SNPs) were genotyped on

two platforms. The "SolCAP" Infinium array (Illumina Inc., San Diego, CA) developed for tomato (Sim et al., 2012a) contained 7,700 SNPs, 3,255 of which were polymorphic within the SolCAP inbred population. Previous data suggested that recombination in tomato was limited and the 3,255 SNPs would produce redundant genome coverage (Sim et al., 2012a). SNPs for the RIL population were therefore selected to cover the genome based on polymorphic information content, recombination, and physical position (Sim et al., 2012b). A 384-marker panel was selected as an optimized SNP set designed for use in a Competitive Allele Specific Polymorphism (KASP) assay (Semagn et al., 2014). Genotyping of the progenies was performed as a service by LGC Genomics (LGC Group, Middlesex, UK). All genotypes were coded relative to OH8245 as a genotype reference. Test-cross and hybrid genotypes were inferred by combining data for the two parent genotypes.

Genome-wide association studies (GWAS), genetic mapping, QTL analysis, and single-marker analysis

GWAS analysis was conducted using GAPIT (Lipka et al., 2012) by fitting the fixed and random circulating probability unification (FarmCPU) model in the RIL population (Liu et al., 2016). The FarmCPU model uses only the associated genetic markers as pseudo Quantitative Trait Nucleotides (QTNs) to derive kinship correction, controlling false positives as well as the MLM model with reductions in both false negatives and computing times (Liu et al., 2016). The first three principal component analysis scores were used as fixed effects to account for population structure.

A set of 260 SNP markers from the 384 marker panel were polymorphic in the RIL2 population (OH7814 x OH987034). Of these, 196 markers passed quality filter for segregation distortion, missing data, and independence and were used to construct a genetic map using qtl package 1.50 (Broman et al. 2003). Map order was adjusted using the "Reorder" function in the qtl package and quality was assessed based on the logarithm of odds (LOD), chromosome lengths, and heatmaps. The Kosambi map function was applied for calculating the map distance (cM) between adjacent markers (Kosambi, 1944). Finally, map quality was assessed using linear regression to compare genetic and physical position relative to the SL.4.0 tomato genome (Hosmani et al., 2019).

QTL analysis was performed by composite interval mapping (CIM) within the qtl package, using a 1 cM scan to detect putative QTL. We included a single cofactor with a 10 cM window size. The 95% significance LOD threshold was determined using 1000 permutations (Churchill and Doerge, 1994). A linear regression of phenotypes on markers was used to infer the

proportion of variance explained by the QTL. A LOD-defined interval (CIM interval) was used to detect SNPs highly associated with the putative QTLs. We also performed a single marker analysis aiming to define a wider region of association around the peaks (SMA interval), using the simplified model (Y= m + e) and considering as associated markers those that had a P-value < 0.01. The markers included in these intervals were tested for linkage disequilibrium, and when dependency between two markers was detected we eliminated one of those. This step was necessary because modeling such markers as fixed covariates inside the mixed model approach in the third research stage does not allow dependency between factors.

Gene effect estimation. Action and allele substitution

We estimated the gene effects for all markers detected in the previous stage in the Testcross population. The linear regression model (Y = m + e) was used to obtain each locus genotype value and infer the gene action effect. These estimations were used later for modeling different degrees of dominance in genomic prediction. A phase adjustment was made when necessary to match the Testcross QTL signal by multiplying genotypes values by -1 when the beneficial allele was coded with a negative sign so that the beneficial genotype value was always positive.

Candidate gene identification and association analysis of Hawaii 7998 SP5 allele with plant architecture and yield components

Sequence for *SP5G* was retrieved from the National Center for Biotechnology Information (NCBI) using GeneBank identification AY186736.1 (Carmel-Goren et al., 2003). The Basic Local Alignment Search Technique (BLAST) tool at the Sol Genomics Network (https://solgenomics.net/) was used to identify the International Tomato Annotation Group (ITAG) gene identification for *SP5G*.

Two strategies were used to identify polymorphisms in *SP5G*. First, genome sequences provided by the Tomato Genome Consortium (The Tomato Genome Consortium, 2012) for the red-fruited tomato varieties Heinz 1706 (Hosmani et al., 2019), LA1589 (The Tomato Genome Consortium, 2012), and Hawaii 7998 (Anderson, 2020) were searched using stand-alone BLAST. Sequences matching AY186736.1 were retrieved for each accession. Second, a PCR amplification strategy was used for sequencing to confirm polymorphisms identified in the bioinformatic pipeline. We developed eleven pairs of tiled primers each amplifying approximately 600 bp spanning the structural gene. Primer pairs (5' to 3') included: SPOH-1 F: TCCCAATAAACAAAGGAAAAAG R: TTCATTGCTACTTAAGTGGTTTCTC; SPOH-2 F:

CGAGAAAGGATTTAATTTCTCAAA R: TGGATACATGAGCCATGACAA; SPOH-3 F: AATATGACGCATAAATCATTCCA R: GGCCAAACGTTTCTTTACCA; SPOH-4 F: TGTCAAAATTACTTGATTCTCCAC R: AATCAATGGTTGACAAGTCGTG; SPOH-5 F: TGAAGATGCGACTTTGTTTGA R: ACTTGTGAAGGCCTCAAGGA; SPOH-6 F: AAAATATCATCCATCCATCTCATGT R: GCCTTTGCACATATTGTCCTT; SPOH-7 F: GGTTGACATTGATGGAGACG R: GCATCAGGATCCACCATAATC; SPOH-8 F: GTAACATGTATGTTTAATATTGCAAGC R: TGTAAGGGGATTAAGCAATCG; SPOH-9 F: TCGTCGATAGCGAAGACAAA; R: CGCGTATTTTCCTTGTTTCAA; SPOH-10 F: TTACGCAGTGACGAAGCAAG R: CAGCAACAGGCAAACCTAGA; SPOH-11 F: CGTTATACAGGCAATGAAGTCG R: TCTCATAATTTCAATTTTCTTCGAT. Sequencing, quality control, and alignment followed procedures described previously for the Beta locus (Orchard et al., 2021).

For confirmation of QTL discovery from the GWAS and CIM analysis, F3 individuals derived from OH7814 x OH987034 and OH2641 x OH987034 F2 progeny were selected for evaluation. The F2 seeds used to select F3 individuals and subsequent F4 families were from the same crosses used to develop the nested RIL but were independent of the RIL lines advanced. Selected F3 individuals were evaluated in a greenhouse trial for fourteen traits related to plant architecture: plant height (distance in centimeters from the first node to the first flower cluster), number of nodes (before the first flower cluster), internode length (average internode length in centimeters), flowering time (number of days before the first flower was observed), number of flower clusters, number of flowers (total number of flowers on all flower clusters), number of flowers on the first and second clusters, first and second flower cluster position (main or lateral stem), first and second clusters type A (single or double), and first and second clusters type B (bifurcated or not). F_4 selections derived from five heterozygous F_3 plants were generated based on polymorphism detected by SP5OH-9 primers (F: TCGTCGATAGCGAAGACAAA; R: CGCGTATTTTCCTTGTTTCAA). From each of these five families, we selected three possible genotypes (homozygous Hawaii 7998, heterozygous, and homozygous for the common allele found in processing tomato). A trial containing a total of 30 plots (5 families x 3 genotypes x 2 blocks) was evaluated in the field as a randomized complete block design. Plots consisted of 20 plants each based on the SP5G allele. We evaluated production by measuring ripe, green, cull, and total fruit for the central five plants based on yield (kg.plant-1), marketable fruit (kg.plant-1), fruit weight (grams.fruit-1), and plot width (centimeters). Analysis of F3 individuals was based on the linear regression model ($Y = m + e$) where m was the marker effect. For binary traits, a Kruskal–Wallis test was applied using the function kruskal.test in R (R Core team, 2022). For F_4

families the model was (Y = f + m + e) where f was a variable for family which traces back to specific F3 individual, which was self-pollinated to produce F4 families, separated by genotype as described above.

Genomic prediction models

Genomic prediction models were developed and tested on three populations: SolCAP inbreds, the nested RIL, and the two populations combined. A total of 268 markers passed quality control based on <5% missing data, a minimum 5% minor allele frequency (MAF), and were used for prediction. Two types of models were evaluated: 1) a completely random model, with the assumption that all markers had a small effect and a common variance estimated using ridge regression; and, 2) a mixed model, in which markers linked to a QTL were considered as fixed effect covariates, and markers out of the interval were considered as random effects. Within the mixed model, we also tested two intervals of QTL-linked markers, a narrow interval detected by the CIM threshold, and a wider interval detected through single marker analysis (SMA).

Genomic selection (GS) models were developed using the rrBLUP package in R (Endelman, 2011). We estimated the marker effects and genomic estimated breeding value (GEBV) using three training populations, as described above. We assessed the model's prediction accuracy through leave-one-out cross-validation. The Pearson coefficient of correlation between estimated GEBV and the phenotypic BLUP (rg), accuracy, was used to measure prediction performance.

To access the potential of these models to predict further generations, we calculated GEBVs for hybrids using the estimated marker effects and evaluated the prediction accuracy based on the correlation between their GEBVs and phenotypic BLUPs (rg). We adjusted the original GS models to account for QTL dominance effects using three different scenarios: a) Additivity, where the genotype scoring matrix was the standard -1, 0, and 1; b) Dominance, in which we re-scored linked heterozygous SNPs according to the gene action estimate to reflect the beneficial allele (-1 or 1); c) Estimated degree of dominance, where we used Testcross gene effect estimations in the marker matrix. For this final approach, we normalized the gene effect estimation to a range of two to be consistent with the original range of marker scoring.

We also performed all previously described analyses in the SolCAP population with the full marker data set of 3,255 markers. Inside the QTL intervals, we retained only markers that were present in the reduced marker set. This step was necessary to allow faithful comparisons between marker sets.

3.3. Results

Identifying QTLs associated with yield components

GWAS analysis of the nested RIL population revealed a peak on chromosome five associated with kg production per plant, represented by marker solcap_snp_sl_12285 (Figure 2). We also detected four peaks related to fruit weight, one peak each on chromosome one (solcap_snp_sl_13762) and five (solcap_snp_sl_23786), and two peaks on chromosome three (solcap_snp_sl_62180 and solcap_snp_sl_20723).

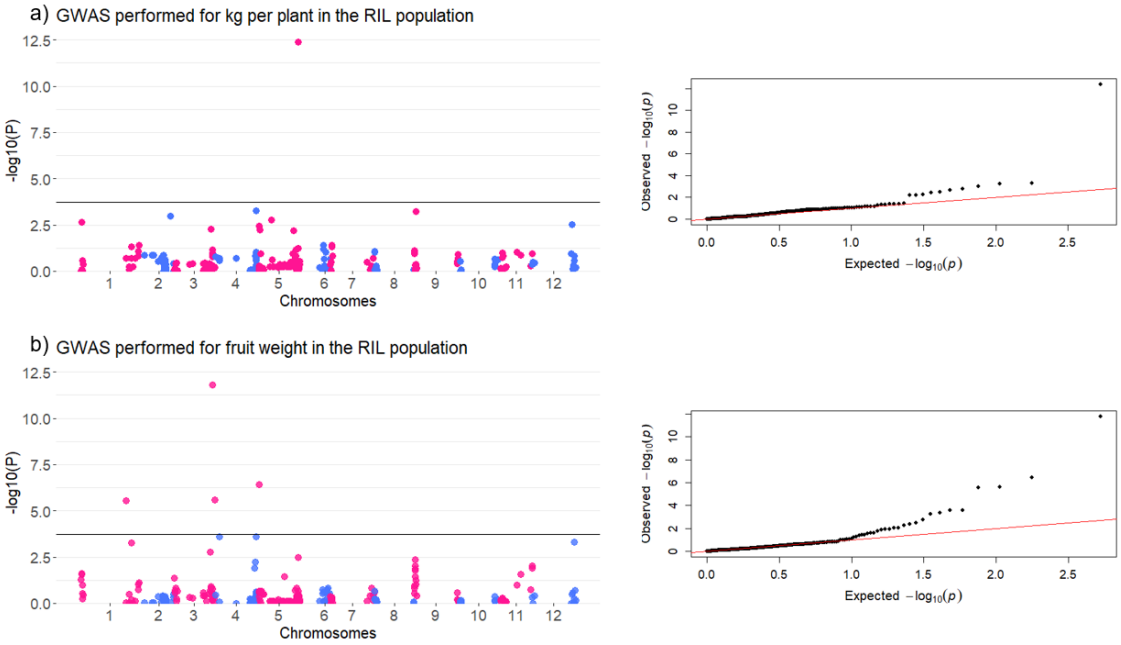


Figure 2. Manhattan and QQ plots resulted from FarmCPU model for kg per plant (a), and fruit weight (b) of RIL population. The black lines represents Bonferroni correction threshold.

We constructed a linkage map of the OH7814 x OH987034 RIL population (n = 119) using 196 SNP markers (Supplementary Table S1) and obtained 128 genetic bins. The map and map quality are summarized in Table 1. For almost all the markers the reference order resulted in a higher LOD score and/or a shorter distance between adjacent markers. There was no information for marker solcap_snp_19661 in the SL.4.0 tomato genome reference. This marker was reordered on chromosome 3, resulting in the best fit. We split chromosome 9 into two linkage groups because a recombination fraction greater than 0.5 was found for adjacent markers. The map covered a total genetic distance of 588 cM, the average distance between markers was around 3 cM, and the maximum gap was 47.4 cM on chromosome 8 (Table 1). For each chromosome, the number of markers ranged from 5 to 65, and the length ranged between 13.1 and 81.5 cM.

Table 1. Genetic map quality for F5 recombinant inbred lines population OH7814 x OH987034

Chromosome	Number of markers	Length	Average distance between markers	Largest distance between markers	Percentage of reference genome covered[a]	Linear regression between genetic map vs physical map (SL4.0)	
						Adjusted r^2	p-value
		cM					
1	13	81.5	7.4	18.5	94.75	0.67	3E-04
2	28	62.2	2.8	21.6	35.28	0.99	2.2E-16
3	9	49.8	8.3	20.0	83.23	0.91	1E-04
4	11	13.1	2.2	8.7	75.99	0.79	1.3E-04
5	64	53.4	1.9	9.0	94.01	0.64	8.5E-16
6	5	50.7	12.7	30.4	24.54	0.99	4.9E-05
7	10	59.8	8.5	22.6	93.25	0.75	7E-04
8	12	69.4	9.9	47.4	97.56	0.65	1E-03
9a	8	21.4	3.1	17.0	6.04	0.98	1.2E-08
9b	2	0	-	-	-	-	-
10	8	19.7	6.6	15.5	86.92	0.48	3.3E-02
11	15	28.4	3.6	10.7	84.80	0.85	1.9E-07
12	11	78.9	9.9	23.9	99.18	0.80	1.1E-03
Total	196	588.2	-	-	-	-	-

[a]Percentage calculated based on the region covered between the first and last marker of the genetic map and the region covered between the first and last marker of the reference genome.

A region around 41 cM on chromosome 5 behaves as a putative QTL for kg production per plant and ripe kg per plant (Figure 2), and explained 0.29 of the phenotypic variation. We detect this peak for most or all the individual trials, suggesting this QTL is robust and can be detected despite variation in environmental conditions (Table 2). Two QTLs were detected for fruit weight, one on chromosome 2 and another on 5, which explains 0.11 and 0.18 of the variation, respectively. The chromosome 5 peak was stable in three environments. Although QTL for fruit weight on chromosome two was only detected in one of the trials, its appearance in the combined data analysis may suggest that it has a consistent small effect across all the trials.

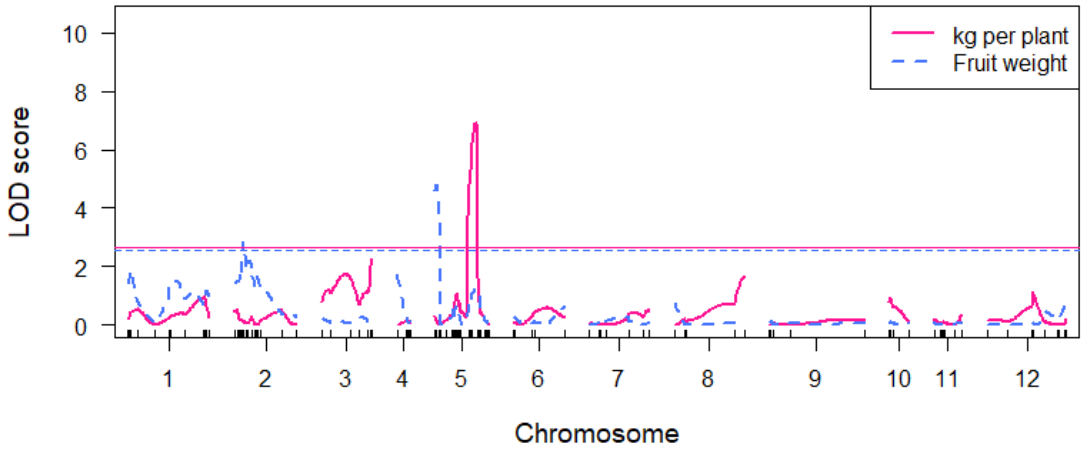


Figure 3. Composite interval mapping (CIM) analysis of kilogram production per plant, fruit weight, in the OH7814 x OH987034 recombinant inbreed line population. The horizontal lines represent the resampled LOD significance cutoff (α=0.05, N=1000 permutations)

Table 2. Quantitative trait loci detected by composite interval mapping in the OH7814 x OH987034 recombinant inbred line population

Trait	Chromosome	Peak position (cM)	LOD	Threshold LOD[a]	Stability[b]	Nearest marker
kg per plant	5	41	6.93	2.65	3/4	solcap_snp_sl_222
Ripe kg per plant	5	40	2.65	2.66	1/4	solcap_snp_sl_222
Fruit weight	2	8.98	2.85	2.54	1/4	Le001778_68_solcap_snp_sl_33474
	5	1.46	4.81	2.54	3/4	solcap_snp_sl_23786

[a]Calculated based resampling (α=0.05, N=1000 permutations). [b]Number of times QTL appears in individual trial/total trials.

Interval of QTL definition and gene effect estimation

We considered the peak detected on chromosome five for kg.plant-1 in both analysis, GWAS and CIM, to define a yield QTL. Three markers were detected in that region for the narrow interval defined by CIM, solcap_snp_sl_12285, solcap_snp_sl_231, and solcap_snp_sl_37588. The marker solcap_snp_sl_231 did not pass linkage disequilibrium criteria for consideration as a fixed effect in GS models and was removed from further analyses. The two remaining markers represent the region between 62.55 and 63.43 Mbp on chromosome 5. A wider interval that included seven additional markers (solcap_snp_sl_12232, solcap_snp_sl_12233, solcap_snp_sl_12244, solcap_snp_sl_22642, solcap_snp_sl_37763,

solcap_snp_sl_345, and solcap_snp_sl_356) was defined through SMA around the QTL peak, and accounts for the region between 62.18 and 64.23 Mpb.

To search regions associated with fruit weight, we only considered the peak found on chromosome 5, as it was the only region detected for both GWAS and CIM. The independent markers found in this region for the narrow interval were solcap_snp_sl_23832 and solcap_snp_sl_2378, representing the region between 3.48 and 3.72 Mpb. For the wider interval, the marker solcap_snp_sl_23722 was added and increased the region of interest to between 3.48 and 4.11 Mpb.

Estimations of the effect of an allele substitution for all markers related to the yield QTL on chromosome 5 indicated dominant to overdominant gene action (Supplementary Table S2). For the fruit weight QTL on chromosome 5 analysis did not detect a consistent or clear pattern of gene action, marker solcap_snp_sl_23832 showed dominance for larger fruit, while markers solcap_snp_sl_23786 and solcap_snp_sl_23722 suggested overdominance for small fruit (Supplementary Table S3).

The chromosome 5 QTL is found in an introgression inherited from Hawaii 7998 which contains a novel allele of SP5

The chromosome 5 QTL region, narrowly defined as 62.55 to 63.43 Mb, contains a paralog of *self-pruning* (*sp*) gene, SP5G (GeneBank ID AY186736.1, Carmel-Goren et al., 2003) which has the ITAG annotation *Solyc05g053850.2* and maps to a physical location between 63,889,954 and 63,891,979 bp relative to the H1706 reference. Sequencing identified a 3.6 Kb contig of *SP5G* including SPOH-3F through SPOH-11 F with a 198 bp insertion polymorphism distinguishing OH987034 from the common processing tomato allele shared by OH2461, OH7814, and H1706 with a 3.4 Kb contig. The OH987034 allele is shared by Hawaii 7998, which is the source of the chromosome 5 introgression (Anderson et al., 2021). The primer pairs SPOH-9 and SPOH-10 flank the indel with the larger allele found in the Hawaii 7998 derived introgression. A SolCAP SNP marker, solcap_snp_sl_101092, which was not in our panel, lies in the first intron of *Solyc05g053850.2*.

Of the 14 traits evaluated for plant architecture in independent F_3 progeny from OH7814 x OH987034 and OH2641 x OH987034, 7 showed differences associated with the *SP5G* genotype classes (Table 3). Plants with the Hawaii 7998 allele were taller and later than plants with the common processing allele. They had fewer flowers and flower clusters, but produced more axillary branches. Variation in inflorescence position, type, and internode length were not detected.

Table 3. Association of *SP5* alleles with traits related to plant architecture

Trait	Genotypic means of markers classes[a]			R^2	p-value	$LSD_{0.05}$
	PP	HH	PH			
Height	30.39	38.06	33.8	0.44	<.0001	2.14
Number of nodes	9.85	11.16	10.3	0.27	<.0001	0.54
Flowering time	7.76	15.43	11.83	0.71	<.0001	1.23
Number of flower clusters (FC)	3.61	2.06	2.93	0.48	<.0001	0.4
Number of flowers	30.53	17.79	25.61	0.56	<.0001	2.78
Number of flowers on the first FC	12.59	10.94	10.65	0.07	0.0128	1.44
Number of flowers on the second FC	8.08	6.78	7.64	0.1	0.022	0.97

[a]PP=Homozygous for Processing allele, HH= Ha7998 homozygote, and PH=heterozygote

The Hawaii 7998 allele of *SP5G* was significantly associated with yield evaluated in F_4 families as measured by the weight of ripe, green, and total fruit harvested (Table 4). There was evidence of dominant gene action for total yield, as the heterozygous genotypes approached homozygous values. There was no difference in fruit size. In these families, *SP5G* association explained 0.71 to 0.1 proportion of the variation (R^2) for plant architectural traits (Table 3) and 0.26 to 0.1 proportion of the variation for yield (Table 4). The Hawaii 7998 allele was associated with later flowering and larger plants with more prolific branching, though flower number per inflorescence was reduced. The effect of an allele substation was approximately 6 T/A of ripe fruit and 9 T/A of total fruit with gene action additive to dominant in these early generation families. The effect of family was non-significant.

Table 4. Association of *SP5* alleles with yield components

Trait	Genotypic means of markers classes[a]			R^2	p-value	$LSD_{0.05}$
	PP	HH	PH			
Ripe fruit weight	30.03	36.75	32.88	0.10	0.0316	4.93
Green fruit weight	5.11	10.4	8.61	0.26	0.0079	3.23
Cull fruit weight	4.35	2.33	4.21	-	ns	-
Total fruit weight	39.49	49.46	45.7	0.11	0.0214	6.93
Fruit size	6.15	6.4	6.32	-	ns	-

[1]PP=Homozygous for Processing allele, HH= Ha7998 homozygote, and PH=heterozygote

Genomic prediction analysis

To account for missing heritability, variation which was not accounted for in the QTL analysis, genomic prediction was explored. Prediction accuracies for yield based on cross-validation ranged from a high of 0.39 for the Nested RIL, 0.38 for the SolCAP population, to 0.30 for the combined population using the limited marker set and a completely random model (Table 5). Accuracies for the SolCAP population were 0.39 with the set of 3,255 markers. When treating markers associated with the chromosome 5 yield QTL based on the narrow interval defined by CIM as fixed effects, the accuracies increased in the Nested RIL and combined populations but not the SolCAP population for either marker set. Using the wider interval defined by SMA, accuracies increased in all training populations with both marker sets (Table 5). In contrast, accuracies for fruit size prediction in cross validation were higher, ranging from 0.35 for the SolCAP training population to 0.58 for the nested RIL. Adding markers increased the prediction in the SolCAP population to 0.43, but modeling linked markers as fixed effects did not increase accuracy over the completely random model (Table 5).

For prediction of hybrid performance, completely random models were predictive of yield only for the SolCAP model defined by the complete marker set as noted previously (Orchard, 2022). When markers linked to the yield-related QTL were treated as fixed effects, the accuracy of the additive SolCAP (3255) model increased from 0.23 to 0.31 when modeling the narrow interval. Modeling the larger interval increased accuracy for the SolCAP (3255) model from 0.23 to 0.29 and the model with 268 markers became significant ($P = 0.04$) with an accuracy of 0.23 (Table 6). We also investigated modeling the QTL-linked markers based on dominant gene action detected in the test cross population (Supplementary Table S2). Accounting for complete dominance or estimates based on the test-cross population using the window identified by CIM led to the same result, with accuracies of 0.35-0.36 for the SolCAP (3255) model and 0.25 for the model based on fewer markers (Table 6). Expanding the interval resulted in a marginal decrease in accuracies, though significant correlations were observed for both SolCAP models. The method of accounting for dominance had a marginal effect on significance and accuracies.

Accuracies of hybrid performance were higher for fruit size (Table 7). Highly significant correlations between predicted and observed values were detected for all training models with accuracies ranging from 0.64-0.76 for completely random models with additive gene action. Accuracies of random models did not increase when modeling linked markers as fixed effects. Incorporating dominance decreased accuracies marginally for all models, though correlations were still significant. Accuracies were marginally higher when the wider interval was modeled for

this trait (Table 7). The decrease in accuracy was greater for the mixed models when using the Nested RIL and Combined populations as training populations. Increasing marker number had a marginal effect in the SolCAP training sets (Table 7).

Table 5. Cross-validation prediction accuracies for yield and fruit size using random and mixed models

	Yield					
Population[a]	**Random model**[b]		**Mixed model**[c]			
			Narrow interval (CIM)		**Wide interval (SMA)**	
	r	p-value	r	p-value	r	p-value
SolCAP (268)	0.38	3.04E-06	0.35	2.73E-05	0.41	4.22E-07
Nested RIL	0.39	4.45E-11	0.44	9.05E-14	0.43	7.04E-13
Combined	0.30	6.27E-10	0.32	4.89E-11	0.31	2.48E-10
SolCAP(3255)	0.39	1.33E-06	0.36	1.28E-05	0.42	1.92E-07
	Fruit size					
Population	**Random model**		**Mixed model**			
			Narrow interval (CIM)		**Wide interval (SMA)**	
	r	p-value	r	p-value	r	p-value
SolCAP (268)	0.35	3.46E-05	0.32	1.42E-04	0.33	1.02E-04
Nested RIL	0.58	1.09E-24	0.54	3.41E-21	0.53	2.04E-20
Combined	0.40	5.05E-17	0.41	5.11E-17	0.40	2.00E-16
SolCAP(3255)	0.43	9.24E-08	0.36	1.85E-05	0.37	7.76E-06

[a]SolCAP. Nested RIL. and combined population were evaluated using 268 markers. SolCAP(3255) corresponds to this population evaluation with full genotypic information. [b]RR-BLUP conventional model. with all markers treated as random. [c]A mixed model including QTL-linked markers as fixed covariates and the remaining markers as random. Two intervals of markers were tested. a narrow one detected through composite interval mapping (CIM) and a wider one detected with single markers analysis.

Table 6. Experimental prediction accuracies for hybrid yield

Narrow interval (CIM)[b]						
	Random model					
Population[a]	**Additive[c]**		**Dominance[d]**		**Estimated dominance[e]**	
	r	p-value	r	p-value	r	p-value
SolCAP (268)	0.13	0.25	0.14	0.24	0.13	0.24
Nested RIL	0.09	0.43	0.12	0.30	0.12	0.31
Combined	0.19	0.11	0.21	0.07	0.21	0.07
SolCAP (3255)	0.23	0.04	0.23	0.04	0.23	0.04
	Mixed model[3]					
Population	**Additive**		**Regular dominance**		**Estimated dominance**	
	r	p-value	r	p-value	r	p-value
SolCAP (268)	0.18	0.11	0.25	0.03	0.25	0.03
Nested RIL	0.03	0.76	0.07	0.53	0.07	0.53
Combined	0.13	0.26	0.19	0.09	0.21	0.07
SolCAP (3255)	0.31	6.12E-3	0.35	1.65E-3	0.36	1.68E-3
Wide interval (SMA)[b]						
	Random model					
Population	**Additive**		**Dominance**		**Estimated dominance**	
	r	p-value	r	p-value	r	p-value
SolCAP (268)	0.13	0.25	0.15	0.18	0.15	0.19
Nested RIL	0.09	0.43	0.17	0.13	0.16	0.16
Combined	0.19	0.11	0.25	0.02	0.25	0.02
SolCAP (3255)	0.23	0.04	0.23	0.04	0.23	0.04
	Mixed model					
Population	**Additive**		**Regular dominance**		**Estimated dominance**	
	r	p-value	r	p-value	r	p-value
SolCAP (268)	0.23	0.04	0.25	0.02	0.21	0.06
Nested RIL	0.03	0.78	0.02	0.82	0.01	0.90
Combined	0.14	0.21	0.20	0.07	0.19	0.09
SolCAP (3255)	0.29	0.01	0.29	9.96E-3	0.27	0.01

[a]SolCAP, Nested RIL, and combined populations were evaluated using 268 markers. SolCAP(3255) corresponds to the SolCAP population evaluated with 3,255 polymorphic markers from the Illumina Tomato array. [b]The adjustments were made in two intervals, the narrow one included 2 markers detected through composite interval mapping (CIM), and the wider interval included 9 markers detected through single markers analysis (SMA). [c]Assumes additive gene action, with markers scored -1, 0, 1. [d]Heterozygous loci were recoded according to the gene action estimate to reflect the beneficial allele (-1 or 1). [e]Heterozygous loci were recoded based on the estimated effect of an allele substitution.

Table 7. Experimental prediction accuracies for hybrid fruit size

Narrow interval (CIM)[b]						
	Random model					
Population[a]	**Additive[c]**		**Regular dominance[d]**		**Estimated dominance[e]**	
	r	p-value	r	p-value	r	p-value
SolCAP (268)	0.74	4.60E-14	0.73	4.69E-14	0.73	1.48E-09
Nested RIL	0.64	5.55E-10	0.64	2.14E-10	0.64	8.83E-13
Combined	0.69	9.50E-12	0.68	5.28E-11	0.68	1.76E-11
SolCAP (3255)	0.76	3.89E-15	0.76	3.82E-15	0.76	5.80E-15
	Mixed model					
Population	**Additive**		**Regular dominance**		**Estimated dominance**	
	r	p-value	r	p-value	r	p-value
SolCAP (268)	0.72	3.23E-13	0.69	3.21E-12	0.69	7.41E-12
Nested RIL	0.39	4.74E-04	0.29	1.13E-02	0.31	7.10E-03
Combined	0.62	2.58E-09	0.59	2.36E-08	0.58	4.26E-08
SolCAP (3255)	0.75	6.62E-15	0.73	3.44E-14	0.73	5.50E-14
Wide interval (SMA)[b]						
	Random model					
Population	**Additive**		**Regular dominance**		**Estimated dominance**	
	r	p-value	r	p-value	r	p-value
SolCAP (268)	0.74	4.60E-14	0.74	1.47E-13	0.74	3.60E-14
Nested RIL	0.64	5.55E-10	0.62	2.28E-09	0.62	2.71E-09
Combined	0.69	9.50E-12	0.68	1.E-11	0.68	2.46E-11
SolCAP (3255)	0.76	3.89E-15	0.75	5.06E-15	0.75	4.54E-15
	Mixed model					
Population	**Additive**		**Regular dominance**		**Estimated dominance**	
	r	p-value	r	p-value	r	p-value
SolCAP (268)	0.74	2.75E-14	0.69	8.65E-12	0.68	1.60E-11
Nested RIL	0.35	1.77E-03	0.50	4.78E-06	0.50	5.81E-06
Combined	0.66	1.17E-10	0.63	1.13E-09	0.63	8.71E-10
SolCAP (3255)	0.77	7.84E-16	0.72	1.63E-13	0.72	2.28E-13

[a]SolCAP, Nested RIL, and combined populations were evaluated using 268 markers. SolCAP(3255) corresponds to the SolCAP population evaluated with 3,255 polymorphic markers from the Illumina Tomato array. [b]The adjustments were made in two intervals, the narrow one included 2 markers detected through composite interval mapping (CIM), and the wider interval included 9 markers detected through single markers analysis (SMA). [c]Assumes additive gene action, with markers scored -1, 0, 1. [d]Heterozygous loci were recoded according to the gene action estimate to reflect the beneficial allele (-1 or 1). [e]Heterozygous loci were recoded based on the estimated effect of an allele substitution.

3.4. Discussion

We explore the genetic basis of yield in processing tomato and identified a QTL on chromosome 5 which explains 29 percent of the variation for total fruit produced. Identifying a major yield-related QTL seems to contradict the infinitesimal model, though a review of the literature supports the existence of QTL for yield and yield-related traits. The majority of QTL analyses carried out on tomatoes used interspecific populations, which are often genetically distant from materials readily used in breeding programs. QTLs that are identified in wide crosses are often specific to those crosses and cannot be extrapolated. Although several yield-related QTLs have been identified in tomatoes, there is limited concordance across studies (Foolad, 2007; Hernández-Bautista et al., 2015). Regarding fruit size, most QTLs were identified in chromosomes 1, 2, 3, 4, 6, 7, and 11, with many independent groups finding the same regions for different populations (Foolad, 2007). Several fruit size QTL identified in wide crosses have now been cloned and the beneficial alleles tend to be fixed in breeding programs suggesting that these alleles are domestication related and contribute little to variation in contemporary breeding populations (Rodríguez et al., 2011). More recently studies have begun to investigate the genetic basis of yield in breeding populations. QTL mapping in an F2 population derived from a cross between two commercial F1 hybrids of tomato identified 13 QTLs for plant growth, yield, and fruit (Ohyama et al., 2017). Although none of these QTL overlapped with those we identified here, this work serves to illustrate that exceptions to the infinitesimal model exist.

The chromosome 5 QTL we identified, located between 62.55 to 63.43 Mb, contains a paralog of *self-pruning* (*sp*) *SP5G* (GeneBank ID AY186736.1, Carmel-Goren et al., 2003), which has the ITAG annotation *Solyc05g053850.2* and maps to the physical location at approximately 63.89 bp relative on the tomato reference. Self-pruning belongs to a family of genes which includes at least five other loci that are hypothesized to be involved in plant growth and development due to their sequence homology (Carmel-Goren et al., 2003). The discovery of a tomato line with a mutation in the *SP* gene on chromosome 6 (Yeager, 1927) is considered a milestone for tomato cultivation. Determinate plants are homozygous for the recessive allele (sp), and display a progressive decrease in the number of leaves separating each inflorescence until growth is ended by the production of two successive inflorescences. This growth pattern resulted in a compact plant with synchronized flowering and fruit ripening, and finally allowed mechanical harvesting (MacArthur, 1932; Pnueli et al., 1998).

The region we defined overlaps with the location of *obscuravenosa* (*obv*), a gene which determines whether chloroplasts are found in cells around the leaf veins causing them to appear dark (Jones et al., 2007). *Obscuravenosa* was mapped to 63.05–64.01 Mb and the candidate gene,

Solyc05g054030, was identified as a C_2H_2 zinc finger transcription factor (Lu et al., 2021). It has been hypothesized that the prevalence of obv in processing varieties resulted from its linkage to a QTL, possibly an allele of *SP5G*, which affected compact plant habit associated with mechanically harvested tomatoes (Jones et al., 2007). Although flower number per inflorescence was lower in the greenhouse, the increased branching could explain higher yield in the field. Alleles of *SP5G* were shown to have flower-repressing activity (Cao et al., 2016), consistent with our observation. Variation in this locus may also be responsible for the loss of day-length-sensitive flowering and may have played an important role in the expansion of cultivated tomato to beyond its center of origin (Soyk et al., 2017). The Hawaii 7998 introgression on chromosome 5 with the yield QTL also contains resistances *EB-5* and *Rx-3*, which confer resistance to early blight (*Alternaria linariae*) (Anderson, 2020) and bacterial spot race T1 (*Xanthomonas* sp.) (Yang et al. 2005; Sim et al., 2015), respectively. This introgression was found to be rare based on a genome comparison of 770 sequenced accessions of tomato (Anderson, 2020). This introgression is still therefore rare in breeding populations and could be fixed through selection.

The discovery of a major QTL affecting processing tomato yield has implications for breeding programs. One approach is to fix the introgression through MAS. Alternatively, a GS approach which uses whole-genome coverage to incorporate small effect QTL (Meuwissen et al, 2001). However, traditional GS approaches assume additive gene action. Given our knowledge of genome position and gene action, we explored GS models that incorporate QTL information by modeling linked markers as fixed covariates. Additionally, we modeled dominant gene action. The mixed models accounting for the yield QTL on chromosome 5 improved hybrid prediction accuracy from 0.23 to 0.36 when dominance was considered and the full marker set was used for prediction. Also, incorporating linkage and dominance in mixed models improved accuracy from 0.13 to 0.25 with the smaller marker set. Defining the QTL based on the narrow CIM interval performed better than the larger interval based on SMA. Hybrids were developed using lines from the SolCAP population, and models based on this training set were superior.

For fruit weight, random models showed high hybrid prediction accuracies (0.74 - 0.76). Modeling different intervals and dominance failed to increase prediction accuracy. This failure held based on marker number and training population. For this trait, high prediction accuracies were observed across marker sets and training populations with fully random models. The fruit weight QTL we detected explained up to 18% of the variation, though their estimates of gene action were inconclusive because adjacent markers provided inconsistence estimates across the QTL interval. We conclude that this QTL may be restricted to the RIL population and cannot be

extrapolated. In addition, dominance does not seem to play an important role in fruit weight. The infinitesimal model may be a better description of genetic architecture for this case.

The literature is divided regarding the value of incorporating linkage and gene action into predictive models. The inclusion of fixed covariates accounting for QTL signals in GS models has been shown to be effective in improving prediction accuracies in rice (Spindel et al., 2016), tomato (Liabeuf et al., 2018), wheat (Arruda et al., 2016; Sarinelli et al., 2019; Zaim et al., 2020), and maize (Bian and Holland, 2017). However, simulation studies including fixed effects for markers tagging GWAS peaks using wide data for maize and sorghum diversity panels showed no increase, or even a decrease, in prediction for most of the simulated genetic architectures (Rice and Lipka, 2019). The results reported here suggest there is practical value in incorporating information on linkage and gene action.

3.5. Conclusion

Chromosome 5 holds a dominant yield-related QTL in processing tomatoes. This QTL co-locates to an allele of *SP5* that affects plant architecture. Adding information on this QTL and its gene action improved GS models.

REFERENCES

Anderson, T. (2020). Genetics and Breeding of Early Blight and Bacterial Spot Resistant Tomatoes [Doctoral dissertation, Cornell University]. Cornell University Digital Repository. https://ecommons.cornell.edu/bitstream/handle/1813/103279

Anderson, T.A., Zitter, S. M., De Jong, D. M., Francis, D. M., & Mutschler, M. A. (2021). Cryptic introgressions contribute to transgressive segregation for early blight resistance in tomato. Theoretical and Applied Genetics, 134, 2561–2575. https://doi.org/10.1007/s00122-021-03842-x

Arruda, M. P., Lipka, A. E., Brown, P. J., Krill, A. M., Thurber, C., Brown- Guedira, G., Dong, Y., Foresmam, B. J., & Kolb, F. L. (2016). Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (Triticum aestivum L.). Molecular Breeding, 36(84), 1:11. http://doi.org/10.1007/s11032-016-0508-5

Barton, N. H., Etheridge, A. M., & Veber, A. (2017). The infinitesimal model: Definition, derivation, and implications. Theoretical Population Biology, 118, 50–73. htttp://doi.org/10.1016/j.tpb.2017.06.001

Bassi, F. M., Bentley, A. R., Charmet, G., & Ortiz, R. (2016). Breeding schemes for the implementation of genomic selection in wheat (Triticum spp.). Plant Science, 242, 23–36. http://doi.org/10.1016/j.plantsci.2015.08.021

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bernal, E., Liabeuf, D., & Francis, D.M. (2020). Evaluating quantitative trait locus resistance in tomato to multiple Xanthomonas spp. Plant Disease, 104(2), 423-429. https://doi.org/10.1094/PDIS-03-19-0669-RE

Berry, S. Z., & Gould, W. A. (1983). 'Ohio 7814'tomato. HortScience, 18, 494–496.

Bian, Y., Holland, J. (2017) Enhancing genomic prediction with genome-wide association studies in multiparental maize populations. Heredity, 118, 585–593. https://doi.org/10.1038/hdy.2017.4

Bolkan, H. A., Williamson, V.M, & Waters, C. M. (1983). Use of cellulose acetate electrophoresis as an alternative to starch gel electrophoresis for detecting root-knot nematode resistance in tomato. Plant Disease, 71, 1001-1003.

Broman, K. W., Wu, H., Sen, Ś., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. Bioinformatics, 19, 889–890. https://rqtl.org/ 10.1093/bioinformatics/btg112

Cao, K., Cui, L., Zhou, X., Ye, L., Zou, Z., & Deng, S. (2016). Four tomato FLOWERING LOCUS T-like proteins act antagonistically to regulate floral initiation. Frontiers in Plant Science, 6, 1213. https://doi.org/10.3389/fpls.2015.01213

Carmel-Goren, L., Liu, Y. S., Lifschitz, E., & Zamir, D. (2003) The SELF PRUNING gene family in tomato. Plant Molecular Biology, 52, 1215–1222. https://doi.org/10.1023/B:PLAN.0000004333.96451.11

Churchill, G. A., & Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. Genetics, 138, 963–971. https://doi.org/10.1093/genetics/138.3.963

Duangjit, J., Causse, M., & Sauvage, C. (2016) Efficiency of genomic selection for tomato fruit quality. Molecular Breeding, 36, 29. https://doi.org/10.1007/s11032-016-0453-3

Edwards, M. D., Stuber, C. S., & Wendel, J. F. (1987). Molecular-marker-facilitated investigations of quantitative-trait loci in maize. i. numbers, genomic distribution and types of gene action. Genetics, 116, 113-125. https://doi.org/10.1093/genetics/116.1.113

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome, 4, 250–255. https://doi.org/10.3835/plantgenome2011.08.0024

Federer, W. T., R. C. Nair, & Raghavarao, D. (1975). Some augmented row-column designs. Biometrics, 31, 361–373. https://doi.org/10.2307/2529426

Fisher, R. A. (1919). The correlations between relatives on the supposition of Mendelian inheritance. Philosophical Transactions of the Royal Society of Edinburgh, 52, 399–433. http://doi.org/10.1017/S0080456800012163

Foolad, M. R. (2007). Genome mapping and molecular breeding of tomato. International Journal of Plant Genomics, 64358, 1-52. https//doi.org/10.1155/2007/64358

Francis, D. M., & Miller, S. (2005). Ohio 9834 and Ohio 9816: Processing tomato breeding lines with partial resistance to race T1 of bacterial spot. Hortscience, 40(5), 1566-1568. https://doi.org/10.21273/HORTSCI.40.5.1566

Hayes, B., & Goddered, M. E. (2001). The distribution of the effects of genes affecting quantitative traits in livestock. Genetics Selection Evolution, 33, 209–229 https://doi.org/10.1186/1297-9686-33-3-209

Hernández-Bautista, A., Lobato-Ortiz, R., Cruz-Izquierdo, S., García-Zavala, J. J., Chávez-Servia, J. L., Hernández-Leal, E., & Bonilla-Barrientos. (2015). Fruit size QTLs affect in a major proportion the yield in tomato. Chilean Jounal of Agricultural Research, 74(4), 402-409. http://dx.doi.org/10.4067/S0718-58392015000500004

Hernández-Bautista, A., Lobato-Ortiz, R., García-Zavala, J. J., Parra-Gómez, M. A., Cadeza-Espinosa, M., Canela-Doñan, D., Cruz-Izquierdo, S., & Chávez-Servia, J. (2016). Implications of genomic selection for obtaining F2:3 families of tomato. Scientia Horticulturae, 207, 7-13. https://doi.org/10.1016/j.scienta.2016.05.005

Hosmani, P. S., Flores-Gonzalez, M., Van de Geest, H., Maumus, F., Bakker, L. V., Schijlen, E., van Haarst, J., Cordewener, J., Sanchez-Perez, G., Peters, S., Fei, Z., Giovannoni, J. J., Mueller, L .A., & Saha, S. (2019). An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. BioRxiv, 767764. https://doi.org/10.1101/767764

Jones, C. M., Rick, C. M., Adams, D., Jernstedt, J., & Chetelat, R. T. (2007). Genealogy and fine mapping of obscuravenosa, a gene affecting the distribution of chloroplasts in leaf veins, and evidence of selection during breeding of tomatoes (Lycopersicon esculentum; Solanaceae). American Journal of Botany, 94(6), 935-47. http://doi: 10.3732/ajb.94.6.935. PMID:

Kosambi, D. D. (1944). The estimation of map distance from recombination values. Annals of Eugenics, 12, 172–175. https://doi.org/10.1111/j.1469-1809.1943.tb02321.x

Liabeuf, D., Sim, S-C., & Francis, D. M. (2018). Comparison of marker-based genomic estimated breeding values and phenotypic evaluation for selection of bacterial spot resistance in tomato. Phytopathology, 180, 392-401. https://doi.org/10.1094/PHYTO-12-16-0431-R

Lin, C. S. & Poushinsky, G. (1983). A modified augmented design for an early stage of plant selection involving a large number of test lines without replication. Biometrics, 39, 553–561. https://doi.org/10.2307/2531083

Lin, C. S., Poushinsky, G., & Jui, P. Y. (1983). Simulation study of three adjustment methods for the modified augmented design and comparison with the balanced lattice square design. Journal of Agricultural Science, 100, 527–534. https://doi.org/10.1017/S0021859600035279

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S., & Zhang, Z. (2012). GAPIT: genome association and prediction integrated tool. Bioinformatics, 28(18), 2397–2399. https://doi.org/10.1093/bioinformatics/bts444

Liu, X., Huang, M., Fan, B., Buckler, E. S., & Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genetics, 12(2), e1005767. https://doi.org/10.1371/journal.pgen.1005767

Lu, J., Pan, C., Li, X. Huang, Z., Shu, J., Wang, X., Lu, X., Pan, F., Hu, j., Zhang, H., Su, W., Zhang, M., Du, Y., Liu, L., Guo, Y., & Li, J. (2021). OBV (obscure vein), a C2H2 zinc finger transcription factor, positively regulates chloroplast development and bundle sheath extension formation in tomato (Solanum lycopersicum) leaf veins. Horticulture Research, 8, 230. https://doi.org/10.1038/s41438-021-00659-z

MacArthur, J. W. (1932). Inherited character in tomato. I-The self-pruning habit. Journal of Heredity, 23, 394–395. https://doi.org/10.1093/oxfordjournals.jhered.a103514

Merk, H. L., Yarnes, S. C., Van Deynze, A., Tong, N., Menda, N., Mueller, L. A., Mutschler, M. A., Loewen, S. A., Myers, J. R., & Francis, D. M. (2012). Trait diversity and potential for selection indices based on variation among regionally adapted processing tomato germplasm. Journal of American Society for Horticultural Science, 137(6), 427–437. https://doi.org/10.21273/JASHS.137.6.427

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. https://doi.org/10.1093/genetics/157.4.1819

Ohyama, A., Shirasawa, K., Matsunaga, H., Negoro, S., Miyatake, K., Yamaguchi, H., Nunome, T., Iwata, H., Fukuoka, H., & Hayashi, T. (2017). Bayesian QTL mapping using genome-wide SSR markers and segregating population derived from a cross of two commercial F1 hybrids of tomato. Theorical and Applied Genetics, 130, 1601-1616. http://doi.org/ 10.1007/s00122-017-2913-5

Orchard, C. (2022). Pre-breeding to combine genes for resistance and agronomic traits in processing and fresh-market tomato [Doctoral dissertation, The Ohio State University].

Orchard, C. J., Cooperstone, J. L., Gas-Pascual, E., Andrade, M. C., Abud, G, Schwartz, S. J., & Francis, D. M. (2021). Rapid identification and assessment of alleles in the promoter of the Cyc-B gene that modulate levels of β-carotene in ripe tomato fruit. Plant Genome,14(1), e20085. https://doi.org/10.1002/tpg2.20085.

Patel, D., Zander, M., Dalton-Morgan, J., & Batley, J. (2015). Advances in plant genotyping: where the future will take us. In Batley J (Ed.), Plant genotyping (pp. 1–11). Springer. https://doi.org/10.1007/978-1-4939-1966-6_1

Pnueli, L., Carmel-Goren, L., Hareven, D., Gutfinger, T., Alvarez, J., Ganal, M., Zamir, D., & Lifschitz, E. (1998) The SELF-PRUNING gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of CEN and TFL1. Development, 125, 1979–1989. http://doi.org/ 10.1242/dev.125.11.1979.

R Core Team. (2022). R: a language and environment for statistical computing. https://www.R-project.org

Rice, B., & Lipka, A. E. (2019). Evaluation of RR-BLUP genomic selection models that incorporate peak genome-wide association study signals in maize and sorghum. Plant Genome, 12(1). http://doi.org/10.3835/plantgenome2018.07.0052

Rick, C.M., & Fobes, J. (1974). Association of an allozyme with nematode resistance. Tomato Gen. Cooperative Rep, 24:25. Available at: https://tgc.ifas.ufl.edu/vol24/vol24.pdf (verified 11/10/2022)

Rodríguez, G. R., Muños, S., Anderson, C., Sim, S. C., Michel, A., Causse, M., Gardener, B. B., Francis, D., & van der Knaap, E. (2011). Distribution of SUN, OVATE, LC, and FAS in the tomato germplasm and the relationship to fruit shape diversity. Plant Physiology, 156(1), 275-285. https://doi.org/10.1104/pp.110.167577

Sarinelli, J. M., Murphy, J. P., Tyagi, P., Holland, J. B., Johnson, J. W., Mergoum, M., Mason, R. E., Babar, A., Harisson, S., Sutton, R., Griffey, C. A., & Brown-Guedira, G. (2019). Training population selection and use of fixed effects to optimize genomic predictions in a historical USA winter wheat panel. Theorical and Applied Genetics, 132, 1247–1261. http://doi.org/10.1007/s00122-019-03276-6

Semagn, K., Babu, R., Hearne, S., & Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement. Molecular Breeding, 33, 1-14. https://doi.org/10.1007/s11032-013-9917-x

Sim, S-C, Robbins, M. D., Wijeratne, S., Wang, H., Yang, W., & Francis, D. M. (2015). Association analysis for bacterial spot resistance in a directionally selected complex breeding population of tomato. Phytopathology, 105(11), 1437-1445. https://doi.org/10.1094/PHYTO-02-15-0051-R

Sim, S-C., Surstewitz, G., Plieske, J., Ganal, M. W., Van Deynze, A., Stoffel, K., Hamilton, J., Buell, C. R., Zarka, D., Douches, D. S., & Francis, D. M.. (2012a). Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. PLoS ONE, 7(7), e40563. https://doi.org/10.1371/journal.pone.0040563

Sim, S-C., Van Deynze, A., Stoffel, K., Douches, D. S., Zarka, D. Ganal, M. W., Chetelat, R. T, Hutton, S. F., Scott, J. W., Gardner, R. G., Panthee, D. R, Mutschelr, M., Myers, J. R., & Francis, D. M. (2012b). High-density SNP genotyping of tomato (Solanum lycopersicum L.) reveals patterns of genetic variation due to breeding. PLoS ONE 7(9): e45520. https://doi.org/10.1371/journal.pone.0045520

Soyk, S., Müller, N., Park, S. J., Schmalenbach, I., Jiang, K., Hayama, R., Zhang, L., Van Eck, J., Jiménez-Gómez, J. M., & Lippman, Z. B. (2017). Variation in the flowering gene SELF PRUNING 5G promotes day-neutrality and early yield in tomato. Nature Genetics, 49, 162–168. https://doi.org/10.1038/ng.3733

Spindel, J. E., Begum, H., Akdemir, D., Collard, B., Redona, E., Jannink, J-L, & McCouch, S. (2016). Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. Heredity, 116, 395–408. http://doi.org/10.1038/hdy.2015.113

The Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature, 485, 635–641. https://doi.org/10.1038/nature11119

Wetterstrand, K. A. (2022). DNA sequencing costs: data from the NHGRI large-scale genome sequencing program. National. Human Genome Research Institute. http://www.genome.gov/sequencingcosts

Williamson, V. M., & Colwell, G. (1991). Acid Phosphatase-1 from Nematode Resistant Tomato Isolation and Characterization of its Gene. Plant Physiology. 97(1), 139–146. https://doi.org/10.1104/pp.97.1.139

Yamamoto, E., Matsunaga, H., Onogi, A., Miyatake, K., Yamaguchi, H., Nunome, T., Iwata, H., & Fukuoka, H. (2017) Efficiency of genomic selection for breeding population design and phenotype prediction in tomato. Heredity, 118, 202–209. https://doi.org/10.1038/hdy.2016.84

Yang, W., Sacks, E. J., Ivey, M. L .L., Miller, S. A., & Francis, D. M. (2005). Resistance in Lycopersicon esculentum intraspecific crosses to race T1 strains of Xanthomonas campestris pv. vesicatoria causing bacterial spot of tomato. Phytopathology, 95(5):519-27. http://doi.org/10.1094/PHYTO-95-0519

Yeager, A.F. (1927). Determinate growth in tomato. Journal of Heredity, 18, 263–265. https://doi.org/10.1093/oxfordjournals.jhered.a102869

Zaim , M., Kabbaj, H., Kehel. Z., Gorjanc, G., Filali-Maltouf, A., Belkadi, B., Nachit, M. M., & Bassi, F. M. (2020). Combining QTL analysis and genomic predictions for four durum wheat populations under drought conditions. Frontiers in Genetics, 11, 316. https://doi.org/10.3389/fgene.2020.00316

Supplementary Material

Supplementary Table S2 Genetic map for RIL population derived from OH7814 x OH987034

Markers	Synonyms	Chromosome	Linkage group	Physical poistion Sl4.0	Genetic position (cM)
solcap_snp_sl_15013		1	1	2677968	0
solcap_snp_sl_60078		1	1	2903136	1,375979188
solcap_snp_sl_20440		1	1	3637274	3,546025206
solcap_snp_sl_30116		1	1	70056469	9,758336778
solcap_snp_sl_36902		1	1	73555274	27,43201958
solcap_snp_sl_9751		1	1	77152068	41,40000748
solcap_snp_sl_34568		1	1	77388343	41,40000753
solcap_snp_sl_2234		1	1	79025804	43,05319029
solcap_snp_sl_2440		1	1	82377024	57,60635
solcap_snp_sl_31775		1	1	86869941	76,1079079
solcap_snp_sl_14323		1	1	87223580	77,59078319
solcap_snp_sl_100168	CL009293.0681	1	1	87430827	78,54911421
solcap_snp_sl_4283		1	1	88836087	81,47642809
solcap_snp_sl_12372		2	2	33200863	0
solcap_snp_sl_8386		2	2	35201344	3,180948923
solcap_snp_sl_20325		2	2	35260118	3,696973825
solcap_snp_sl_8405		2	2	35340445	4,22419779
solcap_snp_sl_8439		2	2	35811913	5,3135172
solcap_snp_sl_14951		2	2	36144718	6,373840132
solcap_snp_sl_8464		2	2	36559238	7,444201024
solcap_snp_sl_33474	Le001778_68	2	2	36977097	8,98199211
solcap_snp_sl_100784	SGN.U568794_snp106	2	2	37602239	9,547000951
solcap_snp_sl_33636		2	2	37825099	11,94146744
solcap_snp_sl_36037		2	2	38037210	12,57106944
solcap_snp_sl_36017	CL015660.0224	2	2	38073580	12,57106949
solcap_snp_sl_66052		2	2	38303824	13,10479665
solcap_snp_sl_35968		2	2	38392687	13,1047967
solcap_snp_sl_		2	2	38412477	13,10479675

35955					
solcap_snp_sl_25485		2	2	39026930	16,79160362
solcap_snp_sl_25482		2	2	39098574	17,25379231
solcap_snp_sl_35798		2	2	39141599	17,66738526
solcap_snp_sl_25429		2	2	39367329	18,14336976
solcap_snp_sl_25418		2	2	39445254	18,14336981
solcap_snp_sl_25405		2	2	39606070	20,72890516
solcap_snp_sl_100811	SGN.U574837_snp399	2	2	39718860	20,72890526
solcap_snp_sl_13550		2	2	39904203	21,18695029
solcap_snp_sl_13581		2	2	40260152	23,14600218
solcap_snp_sl_100718	S_427	2	2	40326455	23,59645921
solcap_snp_sl_100015	X241_2F_264_241_2b_60_b	2	2	40783633	25,90988183
solcap_snp_sl_12841		2	2	47236494	47,52628332
solcap_snp_sl_58447		2	2	52079967	62,21594321
solcap_snp_sl_9703		3	3	2397009	0
solcap_snp_sl_23192		3	3	4290879	8,680862357
solcap_snp_sl_5722		3	3	49252201	28,70776118
solcap_snp_sl_21685		3	3	54291369	37,38732617
solcap_snp_sl_19661		3	3	NA	45,96125931
solcap_snp_sl_7940		3	3	56344706	48,30494745
solcap_snp_sl_7939		3	3	56363081	48,3049475
solcap_snp_sl_7919		3	3	56797388	49,83155803
solcap_snp_sl_15960		3	3	57003043	49,83155808
solcap_snp_sl_1698		4	4	7259196	0
solcap_snp_sl_1701		4	4	7346350	5,00E-08
solcap_snp_sl_6946		4	4	32440265	1,00E-07
solcap_snp_sl_53149		4	4	53798380	8,655333945
solcap_snp_sl_24135		4	4	53847108	8,655333995
solcap_snp_sl_58945		4	4	54252326	8,655334045
solcap_snp_sl_24606		4	4	54364143	9,700764772
solcap_snp_sl_24577		4	4	54444551	10,46590309

solcap_snp_sl_24575		4	4	54509967	11,37102051
solcap_snp_sl_24665		4	4	54693934	12,27645893
solcap_snp_sl_24649		4	4	54766370	13,08377728
solcap_snp_sl_23832		5	5	3484061	0
solcap_snp_sl_23786		5	5	3720272	1,460456297
solcap_snp_sl_23756		5	5	3871993	4,224369215
solcap_snp_sl_48900		5	5	3997782	4,92714058
solcap_snp_sl_23722		5	5	4107281	5,691270592
solcap_snp_sl_23712		5	5	4178073	5,691270642
solcap_snp_sl_29473		5	5	4880444	11,69775874
solcap_snp_sl_51007		5	5	7678832	17,05841924
solcap_snp_sl_51015		5	5	7858705	17,05841929
solcap_snp_sl_51043		5	5	7892027	17,05841934
solcap_snp_sl_13798		5	5	8218805	18,58745608
solcap_snp_sl_51094		5	5	8414773	19,03337044
solcap_snp_sl_51118		5	5	9871599	19,46325739
solcap_snp_sl_51134		5	5	10637662	20,76734653
solcap_snp_sl_51148		5	5	10883430	20,76734658
solcap_snp_sl_51607		5	5	20333768	20,76734663
solcap_snp_sl_51601		5	5	20507268	20,76734668
solcap_snp_sl_51573		5	5	22470726	20,76734673
solcap_snp_sl_51543		5	5	23946024	20,76734678
solcap_snp_sl_51538		5	5	24301659	20,76734683
solcap_snp_sl_67774		5	5	25595831	20,76734688
solcap_snp_sl_100313	CL015854.0378	5	5	25941163	20,76734693
solcap_snp_sl_51281		5	5	26425064	22,42775742
solcap_snp_sl_52271		5	5	31025943	22,42775761
solcap_snp_sl_55294		5	5	33094734	22,42775766
solcap_snp_sl_55302		5	5	35539041	22,42775771
solcap_snp_sl_55319		5	5	38042466	22,42775776
solcap_snp_sl_55326		5	5	39865339	22,42775781

solcap_snp_sl_55342		5	5	41350302	22,42775786
solcap_snp_sl_55348		5	5	41751153	22,42775791
solcap_snp_sl_38800		5	5	48268743	22,42775796
solcap_snp_sl_69405		5	5	51391717	22,42775806
solcap_snp_sl_69381		5	5	54381103	22,42775811
solcap_snp_sl_56076		5	5	56372679	22,42775816
solcap_snp_sl_100588	CL017527.0194	5	5	57799911	22,42775821
solcap_snp_sl_51459		5	5	58434534	23,40758667
solcap_snp_sl_41311		5	5	58700683	23,88742351
solcap_snp_sl_22560		5	5	58999871	23,88742356
solcap_snp_sl_22563		5	5	59204625	23,88742361
solcap_snp_sl_22565		5	5	59267541	23,88742366
solcap_snp_sl_22567		5	5	59268343	23,88742371
solcap_snp_sl_22572		5	5	59470695	23,88742376
solcap_snp_sl_16137		5	5	60026139	24,56240355
solcap_snp_sl_12232		5	5	62428315	33,60809612
solcap_snp_sl_12233		5	5	62443859	33,60809617
solcap_snp_sl_12244		5	5	62555822	34,07501402
solcap_snp_sl_22642		5	5	62611025	34,53670138
solcap_snp_sl_12285		5	5	62800576	36,06236565
solcap_snp_sl_222		5	5	63365854	43,00253532
solcap_snp_sl_231		5	5	63496001	43,00253537
solcap_snp_sl_37588		5	5	63665542	43,53863046
solcap_snp_sl_249		5	5	63701843	43,53863051
solcap_snp_sl_37612		5	5	63727587	44,08377588
solcap_snp_sl_258		5	5	63842577	44,63243203
solcap_snp_sl_280		5	5	63978111	45,75999253
solcap_snp_sl_37689		5	5	64061065	45,75999258
solcap_snp_sl_308		5	5	64293857	49,30917283
solcap_snp_sl_37763		5	5	64393173	49,82991963
solcap_snp_sl_345		5	5	64438672	49,82991968

solcap_snp_sl_356		5	5	64473025	51,44493518
solcap_snp_sl_37808		5	5	64555534	51,9570072
solcap_snp_sl_37812		5	5	64559529	51,95700725
solcap_snp_sl_37888	SGN.U580266_snp391	5	5	64891762	53,40434829
solcap_snp_sl_37896		5	5	64942752	53,40434834
solcap_snp_sl_25660		6	6	34651470	0
solcap_snp_sl_11340		6	6	34879976	0,465075695
solcap_snp_sl_27197		6	6	38211008	17,84208572
solcap_snp_sl_17024		6	6	39227663	20,32898817
solcap_snp_sl_24258		6	6	46260408	50,69711324
solcap_snp_sl_22109		7	7	1791908	0
solcap_snp_sl_68098		7	7	2747420	8,939741328
solcap_snp_sl_15789		7	7	2851544	8,939746328
solcap_snp_sl_15785		7	7	2859618	8,939751328
solcap_snp_sl_22071		7	7	3420679	10,88117622
solcap_snp_sl_22770		7	7	55611190	16,35864344
solcap_snp_sl_6372		7	7	61024967	39,00656636
solcap_snp_sl_6371		7	7	61252038	39,49633553
solcap_snp_sl_7025		7	7	63561726	53,92753817
solcap_snp_sl_37097		7	7	65087011	59,75009853
solcap_snp_sl_7232		8	8	49686	0
solcap_snp_sl_24383		8	8	153948	5,00E-08
solcap_snp_sl_14530		8	8	2588509	9,71852564
solcap_snp_sl_7386		8	8	2853687	10,16692743
solcap_snp_sl_15757		8	8	2862357	10,16692748
solcap_snp_sl_7388		8	8	2931501	11,15772352
solcap_snp_sl_5428		8	8	3283178	11,64723457
solcap_snp_sl_5431		8	8	3418113	12,14209928
solcap_snp_sl_5429		8	8	3370149	12,14209933
solcap_snp_sl_34763		8	8	59822463	59,55866057
solcap_snp_sl_21473		8	8	59906254	59,55866062

solcap_snp_sl_10246		8	8	62551945	69,36626256
solcap_snp_sl_17500		9	9a	489677	0
solcap_snp_sl_17481		9	9a	628839	0,762761417
solcap_snp_sl_28404		9	9a	650754	0,762761467
solcap_snp_sl_7775		9	9a	1241226	1,349612612
solcap_snp_sl_14653		9	9a	1423898	3,872422178
solcap_snp_sl_19983		9	9a	1531081	4,373949084
solcap_snp_sl_19982		9	9a	1532546	4,373949134
solcap_snp_sl_12501		9	9a	4449418	21,41040411
solcap_snp_sl_25721		9	9b	65902112	0
solcap_snp_sl_69576		9	9b	65918843	0
solcap_snp_sl_34373		10	10	3783034	0
solcap_snp_sl_25001		10	10	3932012	1,015566379
solcap_snp_sl_9598		10	10	4054610	1,015571379
solcap_snp_sl_9597		10	10	4118057	1,015576379
solcap_snp_sl_34365		10	10	4515296	1,015593344
solcap_snp_sl_51313		10	10	55081349	4,144395537
solcap_snp_sl_24001		10	10	55781326	4,144400544
solcap_snp_sl_100743	SGN.U317657_C2_At3g47930_snp417	10	10	60135477	19,68060582
solcap_snp_sl_21102		11	11	4530744	0
solcap_snp_sl_21115		11	11	4586610	1,00E-07
solcap_snp_sl_24970		11	11	6618352	6,588224515
solcap_snp_sl_24976		11	11	6688073	7,053126915
solcap_snp_sl_15284		11	11	7311473	7,053126965
solcap_snp_sl_55212		11	11	7723289	7,053127015
solcap_snp_sl_6901		11	11	8024452	7,675922225
solcap_snp_sl_737		11	11	9885158	8,968448505
solcap_snp_sl_732		11	11	10109098	8,968448555
solcap_snp_sl_26266		11	11	11831866	9,591734022
solcap_snp_sl_12406		11	11	12013193	9,591734072
solcap_snp_sl_676		11	11	13022974	10,1689834

solcap_snp_sl_19149		11	11	49787007	20,82129667
solcap_snp_sl_14075		11	11	50076293	20,82129672
solcap_snp_sl_100695	Le014880s_88	11	11	51621630	28,41795444
solcap_snp_sl_17703		12	12	107882	0
solcap_snp_sl_1572		12	12	4080943	20,01170496
solcap_snp_sl_22748		12	12	59842648	43,90511686
solcap_snp_sl_3112		12	12	60847191	45,36338537
solcap_snp_sl_19345		12	12	61458568	45,36338542
solcap_snp_sl_12856		12	12	62630479	56,63020933
solcap_snp_sl_31973		12	12	63926751	69,08018491
solcap_snp_sl_12646		12	12	64080252	69,08018496
solcap_snp_sl_25007		12	12	64291758	71,60059207
solcap_snp_sl_19393		12	12	65668574	77,03878095
solcap_snp_sl_6526		12	12	66315373	78,87276432

Marker	Genotype	Actual estimate	Normalized	BGA	Gene action	Detection interval[b]
solcap_snp_sl_12232	-1	-0,6241	-1,608090698	-1,008090698	Overdominace	SMA
	0	0,1521	0,391909302	0,991909302		
	1	-0,2235	-0,575882505	0,024117495		
solcap_snp_sl_12233	-1	-0,6241	-1,608090698	-1,008090698	Overdominace	SMA
	0	0,1521	0,391909302	0,991909302		
	1	-0,2235	-0,575882505	0,024117495		
solcap_snp_sl_12244	-1	-0,6967	-1,537120794	-0,997120794	Overdominace	SMA
	0	0,2098	0,462879206	1,002879206		
	1	-0,2235	-0,49310535	0,04689465		
solcap_snp_sl_22642	-1	-0,6967	-1,192366935	-0,992366935	Dominance	SMA
	0	0,1368	0,234126305	0,434126305		
	1	0,4719	0,807633065	1,007633065		
solcap_snp_sl_12285	-1	-0,6967	-1,192366935	-0,992366935	Dominance	SMA and CIM
	0	0,1368	0,234126305	0,434126305		
	1	0,4719	0,807633065	1,007633065		
solcap_snp_sl_231[a]	-1	0,2235	0,729081716	0,489081716	Overdominace	SMA and CIM
	0	-0,2355	-0,768227043	-1,008227043		
	1	0,3776	1,231772957	0,991772957		
solcap_snp_sl_37588[a]	-1	0,2235	0,729081716	0,479081716	Overdominace	SMA and CIM
	0	-0,2355	-0,768227043	-1,018227043		
	1	0,3776	1,231772957	0,981772957		
solcap_snp_sl_37763[a]	-1	0,27351	1,054801388	0,784801388	Overdominace	SMA
	0	-0,1907	-0,735441573	-1,005441573		
	1	0,3279	1,264558427	0,994558427		
solcap_snp_sl_345[a]	-1	0,2735	0,844135802	0,674135802	Overdominace	SMA
	0	-0,2704	-0,834567901	-1,004567901		
	1	0,3776	1,165432099	0,995432099		
solcap_snp_sl_356[a]	-1	0,2735	0,844135802	0,674135802	Overdominace	SMA
	0	-0,2704	-0,834567901	-1,004567901		
	1	0,3776	1,165432099	0,995432099		

Supplementary Table S3 Gene effects estimations on Testcross population, and normalized and best genotype adjusted (BGA) scoring for markers detected within the yield QTL intervals on chromosome 5

[a]Phase adjustment was applied to match the QTL signal by multiplying the gene effects by -1. For these markers, a negative number gives the best phenotypic value; [b]Single marker analysis (SMA) and composite interval mapping (CIM).

Supplementary Table S3 Gene effects estimations on Testcross population, and normalized and best genotype Adjusted (BGA) scoring for markers detected within the fruit size QTL interval on chromosome 5.

Marker	Genotype	Actual estimate	Normalized	BGA	Gene action	Interval[b]
solcap_snp_sl_23832	-1	-3,3159	-1,5014603	-0,9914603	Dominance	SMA and CIM
	0	0,7275	0,329416559	0,839416559		
	1	1,101	0,4985397	1,0085397		
solcap_snp_sl_23786	-1	0,1703	0,057580471	-0,452419529	Overdominance	SMA and CIM
	0	-1,4372	-0,485934542	-0,995934542		
	1	4,478	1,514065458	1,004065458		
solcap_snp_sl_23722[a]	-1	-1,7637	-1,121732494	-1,001732494	Overdominance	SMA
	0	1,3809	0,878267506	0,998267506		
	1	0,2629	0,167207276	0,287207276		

[a]Phase adjustment was applied to match the QTL signal by multiplying the gene effects by -1. For these markers, a negative number gives the best phenotypic value; [b]Single marker analysis (SMA) and composite interval mapping (CIM).