FRANÇOISE DALPRÁ DARIVA

ADVANCING MOLECULAR BREEDING FROM MARKER-ASSISTED SELECTION TO GENOMIC PREDICTION IN TOMATO

Thesis submitted to the Plant Science Graduate Program of Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

Adviser: Carlos Nick Gomes

Co-advisers: Fernando França da Cunha
Edgard Augusto de Toledo Picoli

VIÇOSA – MINAS GERAIS
2023

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
D218a
2023

Dariva, Françoise Dalprá, 1992-
Advancing molecular breeding from marker-assisted selected to genomic prediction in tomato / Françoise Dalprá Dariva. – Viçosa, MG, 2023.
1 tese eletrônica (119 f.): il. (algumas color.).

Texto em inglês.
Orientador: Carlos Nick Gomes.
Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Agronomia, 2023.
Inclui bibliografia.
DOI: https://doi.org/10.47328/ufvbbt.2023.474
Modo de acesso: World Wide Web.

1. Tomate - Melhoramento genético. 2. Mapeamento cromossômico. 3. Marcadores genéticos. I. Gomes, Carlos Nick, 1979-. II. Universidade Federal de Viçosa. Departamento de Agronomia. Programa de Pós-Graduação em Fitotecnia. III. Título.

CDD 22. ed. 631.52

Bibliotecário(a) responsável: Euzébio Luiz Pinto CRB-6/3317

FRANÇOISE DALPRÁ DARIVA

ADVANCING MOLECULAR BREEDING FROM MARKER-ASSISTED SELECTION TO GENOMIC PREDICTION IN TOMATO

Thesis submitted to the Plant Science Graduate Program of Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

APPROVED: May 19, 2023.

Assent:

gov.br Documento assinado digitalmente
FRANCOISE DALPRA DARIVA
Data: 22/11/2023 13:38:36-0300
Verifique em https://validar.iti.gov.br

Françoise Dalprá Dariva
Author

gov.br Documento assinado digitalmente
CARLOS NICK GOMES
Data: 12/12/2023 12:22:25-0300
Verifique em https://validar.iti.gov.br

Carlos Nick Gomes
Adviser

To the dreamers.

ACKNOWLEDGMENTS

A BIG thanks

To God, for Him is my refuge and my fortress.

To my mother, Irani Dalprá Dariva (*in memorian*), that always believed in the power of Education, and most importantly, that always believed in me (I bet she is very proud of me now).

To my grandfather Antônio, my father Luiz Antônio, my sisters Fernanda, Fabrícia, and Felícitta, and my brother Filiphi, for all the love and support.

To Universidade Federal de Viçosa and The Ohio State University for the opportunity to pursue my education.

To Conselho de Aperfeiçoamento Pessoal de Nível Superior (CAPES) for granting my scholarship.

To my adviser Dr. Carlos Nick for all these years of mentoring and friendship.

To my co-advisers Dr. Fernando França da Cunha and Dr. Edgard Augusto de Toledo Picoli for all the help during my projects.

To Dr. David Francis for all the mentoring and support during my time as a Visiting Researcher in the US. Thank you so much for the opportunity to be part of your team, Dr. Francis!

To all my lab mates, interns, and farm/lab members for all the help during field trials.

To my boyfriend Nicolas for making me smile.

To all my friends for making this journey so joyful.

To everybody that somehow made this dream come true.

with love,

Fran

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Abbe Faria: *"In return for your help, I offer you something priceless."*
Edmond: *"My freedom?"*
Abbe Faria: *"No, freedom can be taken away, as you well know. I offer you my knowledge."*

(The Count of Monte Cristo)

ABSTRACT

DARIVA, Françoise Dalprá, D.Sc., Universidade Federal de Viçosa, May, 2023. **Advancing molecular breeding from marker-assisted selection to genomic prediction in tomato.** Adviser: Carlos Nick Gomes. Co-advisers: Fernando França da Cunha, and Edgard Augusto de Toledo Picoli.

This dissertation demonstrates how quantitative trait loci (QTL) mapping, marker-assisted selection (MAS), and genomic selection (GS) can be used to facilitate breeding for important agronomic traits in tomatoes. The first objective was to investigate the effect of deficit irrigation on yield and fruit quality attributes of four tomato introgression lines (IL) and hopefully locate QTLs for improved performance. Our results revealed losses in yield and fruit weight, but gains in soluble solids content, fruit redness, fruit firmness, and lycopene content as plants were subjected to deficit irrigation. QTLs for improved performance were detected for yield on tomato chromosome (Chr) 3, and lycopene content on Chr 2, and 3 under optimum irrigation, and for fruit firmness on Chr 3, and lycopene content on Chr 2, 3, and 7 under deficit irrigation. The second objective was to study the genetics of fruit puffiness, a physiological disorder that affects fruit quality and factory yield of tomatoes, and provide a solution to plant breeders that face this problem in their breeding populations. An advanced recombinant inbred line (RIL) and three-derived processing tomato populations were used for mapping and validation purposes, respectively. A dominant QTL for increased fruit puffiness was mapped on Chr 1 explaining from 5 to 22.5% of the total phenotypic variation. Missing heritability issues suggest polygenic control of fruit puffiness in tomatoes. A GS model developed from the mapping set predicted fruit puffiness in the validation set with an accuracy of $r = 0.52$ ($p = 2.36e^{-12}$). MAS using the markers solcap_snp_sl_20440 and solcap_snp_sl_18619 associated with the QTL on Chr 1 was as effective as GS. The third objective was to investigate genomic prediction accuracy of yield-related traits in tomato hybrids. First, we imputed a total of 22,681 tomato hybrids using SNP information for 303 tomato parents. Seven GS models using three-related populations and all their possible combinations were then developed. Fifty hybrids were actually created for further use in field validation trials. With correlation coefficients as high as 0.42 for yield and 0.58 for fruit weight, genomic prediction of tomato hybrids showed to be very accurate. Fruit weight predictions were better than yield predictions. Increase in training population size improved yield predictions of

hybrids. For fruit weight, better predictions were obtained from the model in which training lines were more genetically related to selected hybrids. Overall, these results suggest that GS may help breeders to choose which hybrids they should invest in. This dissertation provides useful information about QTL discovery and the role of MAS and GS tools in applied tomato breeding.

Keywords: QTL mapping. Marker-assisted selection. Genomic selection. Molecular breeding. Processing tomato.

RESUMO

DARIVA, Françoise Dalprá, D.Sc., Universidade Federal de Viçosa, maio de 2023. **Avanços no melhoramento molecular do tomateiro: da seleção assistida por marcadores à predição genômica.** Orientador: Carlos Nick Gomes. Coorientadores: Fernando França da Cunha e Edgard Augusto de Toledo Picoli.

Esta tese busca demonstrar como mapeamento de locos para caracteres quantitativos (QTL), seleção assistida por marcadores (SAM) e seleção genômica (SG) podem ser empregados de modo a facilitar o melhoramento para caracteres de importância econômica no tomateiro. O primeiro objetivo foi avaliar o efeito do déficit de irrigação na produção e qualidade de fruto de quatro linhagens de introgressão de tomateiro (ILs) e, com sorte, localizar QTLs para melhor desempenho. Nossos resultados revelaram perdas de produção e peso de frutos, porém ganhos em sólidos solúveis totais, na coloração e no teor de licopeno quando as plantas foram cultivadas sob déficit de irrigação. QTLs para aumento de desempenho foram observados para produção no cromossomo (Cr) 3, e teor de licopeno nos Cr 2 e 3 sob regime ótimo de irrigação, e para firmeza dos frutos no Cr 3, e teor de licopeno nos Cr 2, 3 e 7 sob regime de déficit. O segundo objetivo foi estudar a base genética da formação de frutos ocos em tomate, uma desordem fisiológica que afeta a qualidade e o rendimento industrial, bem como fornecer uma solução aos melhoristas que enfrentam esse problema nas suas populações de melhoramento. Uma população avançada de linhagens de recombinação (RIL), e três populações subsequentes oriundas dessa RIL foram utilizadas para mapeamento e validação de QTLs, respectivamente. Um QTL dominante para aumento de frutos ocos foi mapeado no Cr 1 o qual explicou de 5 a 22.5% da variação fenotípica total. Alta herdabilidade escondida sugere controle poligênico dessa característica em tomate. O modelo de SG desenvolvido a partir da população de mapeamento predisse a porcentagem de frutos ocos na população de validação com uma precisão de $r = 0.52$ ($p = 2.36e^{-12}$). SAM para os marcadores solcap_snp_sl_20440 e solcap_snp_sl_18619 associados com o QTL no Cr 1 foi tão eficiente quanto SG. O terceiro objetivo foi investigar a aplicabilidade da predição genômica para características de produção em híbridos de tomateiro. O primeiro passo foi imputar um total de 22.681 híbridos de tomate usando informações de marcadores SNP de 303 pais. Sete modelos de SG foram criados usando três populações aparentadas e todas a combinações possíveis entre elas. Cinquenta híbridos foram produzidos para uso nos ensaios de validação. Com coeficientes de

correlação de até 0,42 para produção e 0,58 para peso médio dos frutos, a predição genômica de híbridos de tomateiro mostrou-se bastante precisa. As previsões de peso médio dos frutos foram melhores do que as de produção. Aumentar o tamanho da população de treinamento melhorou a predição da produção. Para peso dos frutos, melhores predições foram obtidas para o modelo no qual as linhagens usadas no treinamento eram mais aparentadas dos híbridos selecionados. No geral, nossos resultados sugerem que modelos de SG podem ajudar melhoristas a escolher quais híbridos de tomate eles devem investir. Esta tese de doutorado fornece informação útil sobre descoberta de QTLs para caracteres de interesse e sobre o papel da SAM e da SG no melhoramento genético do tomateiro.

Palavras-chave: Mapeamento de QTL. Seleção assistida por marcadores. Seleção genômica. Melhoramento molecular. Tomate indústria.

LIST OF ILLUSTRATIONS

CHAPTER II

CHAPTER III

CHAPTER IV

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

QTL	Quantitative Trait Loci
MAS	Marker-assisted Selection
GS	Genomic Selection
IL	Introgression Line
OI	optimum irrigation
DI	deficit irrigation
ASW	available soil water
PY	total plant yield
FW	mean fruit weight
TSS	total soluble solids
TA	titratable acidity
RIL	recombinant inbred line
Chr	chromosome
SNP	Single Nucleotide Polymorphism
IM-BI	interval mapping for a binary model
IM-NP	interval mapping for a non-parametric model
CIM	composite interval mapping
LOD	logarithm of odds
LSI	LOD support interval
PVE	percentage of variance explained
GEBV	genomic estimated breeding value
MAF	minor allele frequency
GS	genomic selection
GEBV	genomic estimated breeding value
BLUP	best linear unbiased predictor
MLE	maximum likelihood estimator
LRT	likelihood ratio tests
BLUE	best linear unbiased estimate
PCA	principal component analysis

GxE genotype-by-environment

SNP Single Nucleotide Polymorphism

SL *Solanum lycopersicum*

SP *Solanum pimpinellifolium*

SUMMARY

INTRODUCTION

Breeding a variety is a cyclic process that involves a series of crossing, testing, and selection (Ceccarelli, 2015). First, plant materials with desired agricultural traits are crossed to create genetic variability. The progeny is then tested and the best genetic combinations are advanced. Lines or hybrid materials that reach the final stages of a breeding cycle are released as cultivars and/or used as parents so that the whole process can start again. These newly released cultivars should perform well under the challenging environmental conditions often seen in the field. In tomatoes, weather events including drought, flooding, excess radiation, and temperature fluctuations have a direct impact on yield components (fruit number and size) and fruit quality attributes (soluble solids, lycopene content, titratable acidity, pH, color, viscosity, firmness) (Cui et al., 2020; Ezin et al., 2010; Gent, 2007; Li et al., 2015; Lokesha et al., 2019; Mingchi et al., 2010; Petrović et al., 2019). Environmental constraints are also associated with disease epidemics and physiological disorders (blossom-end rot, cracking, cat-facing, fruit puffiness, etc.) that affect fruit set, weight, appearance, processing ability, and ultimately, sales (Olle and Williams, 2017; Shenge et al., 2018; Singh et al., 2015). In traditional tomato breeding, hybrid progenies, segregating populations as well as advanced inbred lines are tested in multi-location-year trials and selection is carried out based on phenotypic data. Because progenies showing disease and fruit disorder problems are discarded during the course of the breeding cycle, selection for biotic and abiotic stress tolerance can be achieved indirectly if environmental conditions cooperate.

The success of a tomato breeding program is measured by its ability to release elite cultivars in a short time and at a low cost. Time to release a cultivar is shortened by advancing as many generations per year as possible while costs are reduced by prioritizing phenotyping of promising germplasm. Although traditional tomato breeding efforts have been successful (Foolad, 2007), cultivar release often takes years when breeding decisions rely solely on phenotypic data from field trials. In general, 10 to 12 years are required for plant varieties to be launched in the market under traditional breeding circumstances (Hussain et al., 2015). Luckily, recent molecular breeding tools that involve genotype characterization such as marker-assisted selection (MAS), genomic selection (GS), and more recently, large-scale genomic prediction, have been developed to increase speed, selection accuracy, and improve resource allocation in breeding programs (Borrenpohl et al., 2020; Varshney et al., 2017). Marker-aided breeding relies on the association between genotypes and

phenotypes to select individuals based on their genotypic profile only. Marker-assisted selection differs from genomic selection in terms of genome coverage. While MAS targets specific regions or major-effect quantitative trait loci (QTLs) in the genome, GS uses comprehensive marker coverage to capture the effect of several small-effect genes involved in trait expression (Arruda et al., 2016). GS during early generations is more cost-effective than phenotypic selection (PS) (Borrenpohl et al., 2020). Moreover, GS provides higher gains from selection compared to PS approaches (Borrenpohl et al., 2020).

MAS is widely used in tomato breeding especially when it comes to introgression of disease-resistance genes. As for abiotic stress tolerance, less information in terms of QTL mapping is known compared to that of biotic stress tolerance. Furthermore, the use of MAS for abiotic stress tolerance in tomato breeding programs has not been documented so far (Foolad and Panthee, 2012). In contrast to MAS, first GS studies in tomato breeding have just been published (Cappetta et al., 2020). Therefore, the objectives of this study are 1) to describe how molecular tools have been used to facilitate breeding decisions for the tomato crop; 2) to assess the effect of deficit irrigation on yield and fruit quality components of selected processing tomato lines as well as map QTLs for improved performance under ideal and low water supply; 3) to map and validate QTLs for fruit puffiness, a physiological disorder that affects fresh consumption and factory processing of tomato fruits as well as suggest effective selection strategies for breeding of puffiness-free varieties; and 4) to investigate the prediction accuracy of GS models for large scale genomic prediction of yield components in processing tomato hybrids.

References

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

Borrenpohl, D., Huang, M., Olson, E., Sneller, C., 2020. The value of early-stage phenotyping for wheat breeding in the age of genomic selection. Theor. and Appl. Genet. 133, 2499–2520. https://doi.org/10.1007/s00122-020-03613-0

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, 1–14. https://doi.org/10.3390/plants9091236

Ceccarelli, S. (2015), Efficiency of Plant Breeding. Crop Sci. 55, 87-97. https://doi.org/10.2135/cropsci2014.02.0158

Cui, J., Shao, G., Lu, J., Keabetswe, L., Hoogenboom, G., 2020. Yield, quality and drought sensitivity of tomato to water deficit during different growth stages. Sci. Agric. 77. https://doi.org/10.1590/1678-992x-2018-0390

Ezin, V., De La Pena, R., Ahanchede, A., 2010. Flooding tolerance of tomato genotypes during vegetative and reproductive stages. Braz. J. Plant. Physiol. 22, 131-142. https://doi.org/10.1590/S1677-04202010000200007

Foolad, M.R., 2007. Genome mapping and molecular breeding of tomato. Int. J. Plant Genomics, 2007, 1-52. https://doi.org/10.1155/2007/64358

Foolad, M.R., Panthee, D.R., 2012. Marker-Assisted Selection in Tomato Breeding. CRC Crit Rev Plant Sci 31, 93–123. https://doi.org/10.1080/07352689.2011.616057

Gent, M.P.N., 2007. Effect of degree and duration of shade on quality of greenhouse tomato, HortScience, 42, 514-520. https://doi.org/10.21273/HORTSCI.42.3.514

Hussain, B., Raza, Q., Atif, R.M., Ahmad, M.Q., 2022. New breeding techniques (NBTs) and biotechnology for boosting rice grain yield to feed 5 billion in 2050. In: Sarwar, N., Atique-ur-Rehman, Ahmad, S., Hasanuzzaman, M. (eds) Modern techniques of tice crop production. Springer, Singapore. https://doi.org/10.1007/978-981-16-4955-4_33

Li, Y., Jia, Z., Niu, W., Wang, J., Zhang, M., 2015. Effect of post-infiltration soil aeration at different growth stages on growth and fruit quality of drip-irrigated potted tomato plants (*Solanum lycopersicum*). PLoS One 10. https://doi.org/10.1371/journal.pone.0143322

Lokesha, A.N., Shivashankara, K.S., Laxman, R.H., Geetha, G.A., Shankar, A.G., 2019. Effect of High Temperature on fruit quality parameters of contrasting tomato genotypes. Int. J. Curr. Microbiol. Appl. Sci 8, 1019–1029. https://doi.org/10.20546/ijcmas.2019.803.124

Mingchi, L., Xiangli, L., Jing, H., Lihong, G., 2010. Effect of simulated drought stress on plant growth, yield and fruit properties of tomato. Acta Hortic. 856, 193-202. 10.17660/ActaHortic.2010.856.26

Olle, M., Williams, I.H., 2017. Physiological disorders in tomato and some methods to avoid them. J. Hortic. Sci. Biotech., 92, 223-230. https://doi.org/10.1080/14620316.2016.1255569

Petrović, I., Savić, S., Jovanović, Z., Stikić, R., Brunel, B., Sérino, S., Bertin, N., 2019. Fruit quality of cherry and large fruited tomato genotypes as influenced by water deficit. Zemdirbyste 106, 123–128. https://doi.org/10.13080/z-a.2019.106.016

Shenge KC, Jimoh RO, Akpa AD, Chindo PS, Ajene IJ, 2018. Seasonal variations in incidence and severity of bacterial speck of tomato (*Solanum lycopersicum* L.) under rainfed and irrigated conditions in Samaru Zaria, Nigeria. Nigerian Journal of Plant Protection 31, 29-38.

Singh, R.K., Rai, N., Singh, M., Saha, S., Singh, S.N., 2015. Detection of tomato leaf curl virus resistance and inheritance in tomato (Solanum lycopersicum L.). J. Agr. Sci. 153, 78–89. https://doi.org/10.1017/S0021859613000932

Varshney, R.K., Roorkiwal, M., Sorrells, M.E., 2017. Genomic selection for crop improvement: An introduction. In: Varshney, R., Roorkiwal, M., Sorrells, M. (eds) Genomic selection for crop improvement. Springer, Cham. https://doi.org/10.1007/978-3-319-63170-7_1

1. CHAPTER I – BACKGROUND AND SIGNIFICANCE

1.1. Shifting phenotype-oriented to marker-aided breeding in tomatoes

Tomato breeding is a relatively young field with first systematic efforts to improve horticultural traits of tomato cultivars dating back to 1930s (Foolad and Panthee, 2012). Yield and disease resistance aside, desired traits in tomato cultivars vary depending on whether fruits are grown for either processing or fresh consumption. Fresh market varieties, for example, are expected to produce large round fruits that are firm, tasty, intense red, long-lasting, and uniform in size while for processing tomato varieties having jointless pedicel, high soluble solids content, improved viscosity, firmness, and adequate acidity is more important (Atanassova et al., 2007; Foolad, 2007). Until the 1980s, selection of superior tomato families for economically-important traits relied solely on phenotypic records from field trials. In fact, phenotype-based selection alone has provided great results in terms of cultivar improvement. A 7.2-fold increase in yield of processing tomato cultivars (from 10.1 to 72.4 t/ha) was achieved between 1920s and 1990s (Foolad, 2007). Marker-aided breeding in tomatoes was first reported for nematode resistance in 1980 (Foolad and Panthee, 2012; Medina-Filho and Stevens, 1980) which means that all this progress was accomplished before marker-based selection tools became popular in tomato breeding programs. However, there are situations where phenotypic selection may not be the best option. Phenotypic selection requires the availability of screening environments. In case of breeding for disease-resistant varieties, for example, screening environments should contain enough inoculum pressure and provide ideal conditions for pathogen infection and disease development otherwise phenotypic selection will not work. Phenotype-based selection is also time-consuming because requires growing plants for evaluation and often means that only one generation can be advanced per year (Guo et al., 2019).

Since their emergence in the 1970s, molecular markers have caught the attention of breeding initiatives due to their potential benefits for crop improvement especially during genotype selection stages (Schlotterer, 2004). By definition, molecular markers are tools for detecting variation at DNA level (Toppa and Jadoski, 2013). Selection decisions based on molecular marker information outperform phenotype-based selection for low-heritability traits (Dekkers, 2007; Heffner et al., 2009) and are more accurate and cost-effective during early breeding stages

(Borrenpohl et al., 2020). Among all types of molecular markers, Single Nucleotide Polymorphism (SNP) markers, the ones used in this study, are preferred because of their high abundance in the genome and adequacy for high-throughput genotyping automation (Guo et al., 2019). SNP markers are capable of detecting variation with a single-base pair resolution. In tomatoes, an array containing 7,720 SNPs has already been developed (Sim et al., 2012) and optimized (Gill et al., 2019) for mapping and marker-oriented selection purposes. Marker-assisted selection (MAS) and genomic selection (GS) are the two most important marker-oriented selection techniques employed in tomato breeding today. MAS differs from GS by its small genome coverage (selection through MAS is restricted to one or few loci of large effect while GS also captures the effect of several other small-effect genes that are distributed across the genome), and the prior need for gene identification and validation (Collard and Mackill, 2008; Moniruzzaman et al., 2014). The use of marker-based selection tools in tomato breeding is further discussed in detail.

1.2. Quantitative trait loci mapping (QTL) and marker-assisted selection (MAS) in tomato breeding

MAS has already become a routine practice in private and public tomato breeding programs. This technology uses molecular markers tagged to genes of interest to select individuals based on a desirable DNA band pattern (Collard and Mackill, 2008). Therefore, QTL mapping and validation studies involving construction of linkage maps are required before the actual implementation of MAS in breeding programs. The three main uses for MAS are: (1) confirmation of hybrid purity, (2) germplasm screening (especially important for early-generation progeny), and (3) marker-assisted backcrossing (Foolad and Panthee, 2012). It is important to point out that MAS is only recommended after reliable marker-trait associations have been established. In tomato breeding, MAS has been used to incorporate resistance to late blight (Akbar et al., 2016; Wang et al., 2016), fusarium wilt (Akbar et al., 2016; El Mohtar et al., 2007; Lim et al., 2006; Orchard et al., 2023), tomato yellow leaf curl virus (Nevame et al., 2018; Orchard et al., 2023; Prasanna et al., 2015), bacterial spot (Orchard et al., 2023; Yang and Francis, 2005), bacterial speck (Sun et al., 2011; Yang and Francis, 2005), tomato spotted wilt virus (Robbins et al., 2010; Shi et al., 2011), gray leaf spot (Orchard et al., 2023; Su et al., 2019), root-knot nematode (Devran and Halil Elekçioğlu, 2004; Zhang et al., 2007), to improve fruit quality by increasing carotenoid content

(Kinkade and Foolad, 2013; Park et al., 2009; Zhang and Stommel, 2001) and extended shelf life (Nguyen and Sim, 2017; Osei et al., 2019), and to modify plant morphological traits such as jointed pedicels (Lee et al., 2018). The number of studies involving QTL mapping and MAS for abiotic stress tolerance and abiotic stress-related disorders in tomato fruits is low in comparison to disease resistance efforts. So far, QTLs have been mapped for heat (Wen et al., 2019), water deficit (Diouf et al., 2018; Foolad et al., 2003), and salinity stress tolerance (Diouf et al., 2018) as well as problematic fruit disorders like blossom-end rot (Topcu et al., 2021), and cracking (Capel et al., 2017). Although some of these works mention the possibility of MAS for introgression and/or screening for desirable alleles, the real feasibility of MAS for introgression of environment-sensitive traits has yet to be proved. Normally, MAS works well for qualitative traits controlled by one or a few major effect genes but it has its efficacy compromised when only genes of moderate or small effect are involved in trait expression (Jannink et al., 2010). The percentage of phenotypic variation explained by a single QTL in the studies mentioned above has not exceeded 20%, which indicates that proving the efficacy of MAS for quantitative traits is essential prior to its utilization in tomato breeding.

1.3. Genomic selection (GS) in tomato breeding

GS is a recent marker-based tool proposed to overcome the limitations of MAS for quantitative trait selection. In practical terms, GS builds predictive models that are used to select individuals who have already been genotyped but not yet phenotyped (Meuwissen et al., 2001). GS relies on genome saturation with molecular markers in an attempt to capture most of the phenotypic variation present in the population. Selections are based on predicted breeding merit of individuals known as genomic estimated breeding values (GEBVs). Three distinct populations are required for GS implementation in breeding programs: a training population, consisting of individuals whose phenotype and genotype are known and are used to build the GS model; a testing population, in which only genotypic data of individuals is available and is used to assess prediction abilities of models; and a breeding population, that contains individuals that have only been genotyped and are directly involved in breeding (Desta and Ortiz, 2014). Phenotypes of individuals from the training population are regressed on markers in order to estimate their effects. The marker effect vector is then multiplied with the genotype matrix to predict GEBVs of the breeding population.

The higher the GEBV of an individual, the better its phenotypic performance if prediction abilities of models are high. Up to now, the use of genomic selection models in tomato breeding has been reported by Duangjit et al. (2016), Liabeuf et al. (2018), and Yamamoto et al. (2017). In tomatoes, prediction accuracies of GS models are maximized by increasing marker set, marker density, size of the training set, and genetic relatedness between training and testing lines (Duangjit et al., 2016). Furthermore, the higher the trait heritability, the better the prediction (Duangjit et al., 2016). Field validation of GS models is still required since almost all information regarding accuracy of GS models in tomatoes is based on cross-validation results.

1.4. Prospects for genomic hybrid breeding in tomatoes

Pioneer studies in rice have set the ground for the application of large-scale genomic prediction of hybrids in breeding (Cui et al., 2020; Wang et al., 2017; Xu et al., 2014). Genomic hybrid breeding uses GS models to predict unobserved phenotypes of theoretical hybrids. Theoretical hybrids are those imputed by combining genotypic data of parents. Cost-benefit is high because the number of parents required for genotyping is insignificant compared to the number of hybrids predicted. Predicted phenotypes of top-ranked hybrids are statistically different from that of bottom-ranked hybrids and prediction coefficients of GS models are moderate to high especially for high heritability traits (Xu et al., 2014). Hybrid prediction in tomatoes has been demonstrated for bacterial spot resistance (Liabeuf et al., 2018), suggesting a role for large-scale prediction for other traits.

1.5. References

Akbar, K., Abbasi, F.M., Sajid, M., Ahmad, M., Khan, Z.U., Aziz-Ud-Din, Ali, H., 2016. Marker-assisted selection and pyramiding of I1 and Ph3 genes for multiple disease resistance in tomato through PCR analysis. International Journal of Biosciences 9, 108–113. https://doi.org/10.12692/ijb/9.3.108-113

Atanassova, B., Stoeva-Popova, P., & Balacheva, E., 2006. Cumulating useful traits in processing tomato. In X International Symposium on the Processing Tomato 758, 27-36.

Borrenpohl, D., Huang, M., Olson, E., Sneller, C., 2020. The value of early-stage phenotyping for wheat breeding in the age of genomic selection. Theor Appl Genet 133, 2499–2520. https://doi.org/10.1007/s00122-020-03613-0

Capel, C., Yuste-Lisbona, F.J., López-Casado, G., Angosto, T., Cuartero, J., Lozano, R., Capel, J., 2017. Multi-environment QTL mapping reveals genetic architecture of fruit cracking in a tomato RIL *Solanum lycopersicum* × *S. pimpinellifolium* population. Theor Appl Genet 130, 213–222. https://doi.org/10.1007/s00122-016-2809-9

Collard, B. C., Mackill, D. J., 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences 363, 557-572. https://doi.org/10.1098/rstb.2007.2170

Cui, Y., Li, R., Li, G., Zhang, F., Zhu, T., Zhang, Q., Ali, J., Li, Z., Xu, S., 2020. Hybrid breeding of rice via genomic selection. Plant Biotechnol. J. 18, 57–67. https://doi.org/10.1111/pbi.13170

Dekkers, J. C. M., 2007. Prediction of response to marker-assisted and genomic selection using selection index theory. J Anim Breed Genet 124, 331-341. https://doi.org/10.1111/j.1439-0388.2007.00701.x

Desta, Z.A., Ortiz, R., 2014. Genomic selection: Genome-wide prediction in plant improvement. Trends Plant Sci 19, 592-601. https://doi.org/10.1016/j.tplants.2014.05.006

Devran, Z., Halil Elekçioğlu, İ. H., 2004. The screening of F_2 plants for the root-knot nematode resistance gene, Mi by PCR in tomato. Turk J Agric For 28, 253-257. https://journals.tubitak.gov.tr/agriculture/vol28/iss4/6

Diouf, I.A., Derivot, L., Bitton, F., Pascual, L., Causse, M., 2018. Water deficit and salinity stress reveal many specific QTL for plant growth and fruit quality traits in tomato. Front Plant Sci 9, 2019. https://doi.org/10.3389/fpls.2018.00279

Duangjit, J., Causse, M., Sauvage, C., 2016. Efficiency of genomic selection for tomato fruit quality. Mol Breeding 36, 1-16. https://doi.org/10.1007/s11032-016-0453-3

El Mohtar, C.A., Atamian, H.S., Dagher, R.B., Abou-Jawdah, Y., Salus, M.S., Maxwell, D.P., 2007. Marker-assisted selection of tomato genotypes with the I-2 gene for resistance to *Fusarium oxysporum* f. sp. *lycopersici* race 2. Plant Dis 91, 758–762. https://doi.org/10.1094/PDIS-91-6-0758

Foolad, M.R., 2007. Genome mapping and molecular breeding of tomato. Int J Plant Genomics 2007. https://doi.org/10.1155/2007/64358

Foolad, M.R., Panthee, D.R., 2012. Marker-assisted selection in tomato breeding. CRC Crit Rev Plant Sci 31, 93-123. https://doi.org/10.1080/07352689.2011.616057

Foolad, M.R., Zhang, L.P., Subbiah, P., 2003. Genetics of drought tolerance during seed germination in tomato: Inheritance and QTL mapping. Genome 46, 536–545. https://doi.org/10.1139/g03-035

Gill, U., Scott, J.W., Shekasteband, R., Ogundiwin, E., Schuit, C., Francis, D.M., Sim, S.C., Smith, H., Hutton, S.F., 2019. Ty-6, a major begomovirus resistance gene on chromosome 10, is effective against *Tomato yellow leaf curl virus* and *Tomato mottle virus*. Theor Appl Genet 132, 1543–1554. https://doi.org/10.1007/s00122-019-03298-0

Guo, Z., Wang, H., Tao, J., Ren, Y., Xu, C., Wu, K., Zou, C., Zhang, J., Xu, Y., 2019. Development of multiple SNP marker panels affordable to breeders through genotyping by target sequencing (GBTS) in maize. Mol Breeding 39, 1-12. https://doi.org/10.1007/s11032-019-0940-4

Heffner, E.L., Sorrells, M.E., Jannink, J.L., 2009. Genomic selection for crop improvement. Crop Sci 49, 1-12. https://doi.org/10.2135/cropsci2008.08.0512

Jannink, J.L., Lorenz, A.J., Iwata, H., 2010. Genomic selection in plant breeding: From theory to practice. Brief Funct Genomic Proteomic 9, 166–177. https://doi.org/10.1093/bfgp/elq001

Kinkade, M.P., Foolad, M.R., 2013. Validation and fine mapping of lyc12.1, a QTL for increased tomato fruit lycopene content. Theor Appl Genet 126, 2163–2175. https://doi.org/10.1007/s00122-013-2126-5

Lee, T.G., Shekasteband, R., Menda, N., Mueller, L.A., Hutton, S.F., 2018. Molecular markers to select for the j-2–mediated jointless pedicel in tomato. HortScience 53, 153–158. https://doi.org/10.21273/HORTSCI12628-17

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 108, 392–401. https://doi.org/10.1094/PHYTO-12-16-0431-R

Lim, G.T.T., Wang, G.P., Hemming, M.N., Basuki, S., McGrath, D.J., Carroll, B.J., Jones, D.A., 2006. Mapping the I-3 gene for resistance to Fusarium wilt in tomato: Application of an I-3 marker in tomato improvement and progress towards the cloning of I-3. Australasian Plant Pathology 35, 671–680. https://doi.org/10.1071/AP06073

Medina-Filho, H. and Stevens, M. 1980. Tomato breeding for nematode resistance: survey of resistant varieties for horticultural characteristics and genotype of acid phosphates. Acta Hort 100, 383–391. https://doi.org/10.17660/ActaHortic.1980.100.41/

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

Moniruzzaman, M., Khatun, R., Mintoo, A.A., 2014. Application of marker-assisted selection for livestock improvement in Bangladesh. Bangladesh Veterinarian 31, 1-11. https://doi.org/10.3329/bvet.v31i1.22837

Nevame, A.Y.M., Xia, L., Nchongboh, C.G., Hasan, M.M., Alam, M.A., Yongbo, L., Wenting, Z., Yafei, H., Emon, R.M., Ismail, M.R., Efisue, A., Gang, S., Wenhu, L., Longting, S., 2018. Development of a new molecular marker for the resistance to tomato *Yellow leaf curl virus*. Biomed Res Int 2018. https://doi.org/10.1155/2018/8120281

Nguyen, T.T., Sim, S.C., 2017. Development of a gene-based marker for the non-ripening (nor) gene in cultivated tomato. Korean Journal of Horticultural Science and Technology 35, 618–627. https://doi.org/10.12972/kjhst.20170066

Orchard, C.J., Kressin, J., Chompookam, S., Chuapong, J., Onmanee, N., Van Leeuwen, K., Struss, D., Balatero, C., Francis, D.M., 2023. Marker-assisted selection to combine alleles for four disease resistance genes of tomato collocated on Chromosome 11. HortScience 58, 495–501. https://doi.org/10.21273/HORTSCI16982-22

Osei, M.K., Danquah, E., Danquah, A., Massoudi, M., Maxwell, D., Adu-Dapaah, H., Blay, E., 2019. Validation of SNP marker linked to alc gene for long shelf life of tomato. J Crop Improv 33, 669–682. https://doi.org/10.1080/15427528.2019.1657216

Park, Y., Lee, Y., Kang, J., Choi, Y., Son, B., 2009. A Gene-based dCAPS marker for selecting old-gold-crimson (og_c) fruit color mutation in tomato. J Life Sci 19, 152–155.

Prasanna, H.C., Kashyap, S.P., Krishna, R., Sinha, D.P., Reddy, S., Malathi, V.G., 2015. Marker-assisted selection of Ty-2 and Ty-3 carrying tomato lines and their implications in breeding tomato leaf curl disease-resistant hybrids. Euphytica 204, 407–418. https://doi.org/10.1007/s10681-015-1357-8

Robbins, M.D., Masud, M.A., Panthee, D.R., Gardner, R.G., Francis, D.M., Stevens, M.R., 2010. Marker-assisted selection for coupling phase resistance to *Tomato spotted wilt virus* and

Phytophthora infestans (late blight) in tomato, HortScience 45, 1424-1428. https://doi.org/10.21273/HORTSCI.45.10.1424

Schlotterer, C., 2004. The evolution of molecular markers— just a matter of fashion? Nat Rev Genet *5*, 63-69. https://doi.org/10.1038/nrg1249

Shi, A., Vierling, R., Grazzini, R., Chen, P., Caton, H., Panthee, D., 2011. Identification of molecular markers for Sw-5 gene of tomato spotted wilt virus resistance. American Journal of Biotechnology and Molecular Sciences 1, 8–16. https://doi.org/10.5251/ajbms.2011.1.1.8.16

Sim, S.C., Durstewitz, G., Plieske, J., Wieseke, R., Ganal, M.W., van Deynze, A., Hamilton, J.P., Buell, C.R., Causse, M., Wijeratne, S., Francis, D.M., 2012. Development of a large snp genotyping array and generation of high-density genetic maps in tomato. PLoS One 7. https://doi.org/10.1371/journal.pone.0040563

Su, X., Zhu, G., Huang, Z., Wang, X., Guo, Y., Li, B., Du, Y., Yang, W., Gao, J., 2019. Fine mapping and molecular marker development of the Sm gene conferring resistance to gray leaf spot (*Stemphylium* spp.) in tomato. Theor Appl Genet 132, 871–882. https://doi.org/10.1007/s00122-018-3242-z

Sun, W., Zhao, W., Wang, Y., Pei, C., Yang, W., 2011. Natural variation of pto and fen genes and marker-assisted selection for resistance to bacterial speck in tomato. Agric Sci China 10, 827–837. https://doi.org/10.1016/S1671-2927(11)60068-0

Topcu, Y., Sapkota, M., Illa-Berenguer, E., Nambeesan, S.U., van der Knaap, E., 2021. Identification of blossom-end rot loci using joint QTL-seq and linkage-based QTL mapping in tomato. Theor Appl Genet 134, 2931–2945. https://doi.org/10.1007/s00122-021-03869-0

Toppa, E.V.B., Jadoski, C.J., 2013. O uso dos marcadores moleculares no melhoramento genético de plantas. Scientia Agraria Paranaensis 12, 1–5. https://doi.org/10.18188/1983-1471/sap.v12n1p1-5

Wang, X., Li, L., Yang, Z., Zheng, X., Yu, S., Xu, C., Hu, Z., 2017. Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II. Heredity 118, 302–310. https://doi.org/10.1038/hdy.2016.87

Wang, Y.Y., Chen, C.H., Hoffmann, A., Hsu, Y.C., Lu, S.F., Wang, J.F., Hanson, P., 2016. Evaluation of the Ph-3 gene-specific marker developed for marker-assisted selection of late blight-resistant tomato. Plant Breeding 135, 636–642. https://doi.org/10.1111/pbr.12395

Wen, J., Jiang, F., Weng, Y., Sun, M., Shi, X., Zhou, Y., Yu, L., Wu, Z., 2019. Identification of heat-tolerance QTLs and high-temperature stress-responsive genes through conventional QTL mapping, QTL-seq and RNA-seq in tomato. BMC Plant Biol 19, 1-17. https://doi.org/10.1186/s12870-019-2008-3

Xu, S., Zhu, D., Zhang, Q., 2014. Predicting hybrid performance in rice using genomic best linear unbiased prediction. Proc. Natl. Acad. Sci. 111, 12456–12461. https://doi.org/10.1073/pnas.1413750111

Yamamoto, E., Matsunaga, H., Onogi, A., Ohyama, 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 (Edinb) 118, 202–209. https://doi.org/10.1038/hdy.2016.84

Yang, W., Francis, D.M., 2005. Marker-assisted selection for combining resistance to bacterial spot and bacterial speck in tomato, J Amer Soc Hort Sci 130, 716-721. https://doi.org/10.21273/JASHS.130.5.716

Zhang, L.H., Mojtahedi, H., Kuang, H., Baker, B., Brown, C.R., 2007. Marker-assisted selection of Columbia root-knot nematode resistance introgressed from *Solanum bulbocastanum*. Crop Sci 47, 2021–2026. https://doi.org/10.2135/cropsci2007.01.0003

Zhang, Y., Stommel, J.R., 2001. Development of SCAR and CAPS markers linked to the beta gene in tomato. Crop Sci, 41(5), 1602-1608. https://doi.org/10.2135/cropsci2001.4151602x

2. CHAPTER II - YIELD AND FRUIT QUALITY ATTRIBUTES OF SELECTED TOMATO INTROGRESSION LINES SUBJECTED TO LONG-TERM DEFICIT IRRIGATION

Françoise Dalprá Dariva[1], Herika Paula Pessoa[1], Mariane Gonçalves Ferreira Copati[1], Gabriella Queiroz de Almeida[1], Manoel Nelson de Castro Filho[1], Edgard Augusto de Toledo Picoli[2], Fernando França da Cunha[3], Carlos Nick[1]

[1]Departamento de Agronomia, Programa de Pós-graduação em Fitotecnia, Universidade Federal de Viçosa, Av. P.H. Rolfs, s/n, Campus Universitário, Viçosa, MG 36570-900, Brazil
[2]Departamento de Biologia Vegetal, Programa de Pós-graduação em Fitotecnia, Universidade Federal de Viçosa, Av. P.H. Rolfs, s/n, Campus Universitário, Viçosa, MG 36570-900, Brazil
[3]Departamento de Engenharia Agrícola, Programa de Pós-graduação em Engenharia Agrícola, Universidade Federal de Viçosa, Av. P.H. Rolfs, s/n, Campus Universitário, 36570-900, Viçosa, MG 36570-900, Brazil

Abstract

Breeding for high-yielding genotypes with enhanced fruit quality under challenging environmental conditions such as reduced water supply will be the next target of tomato breeders in the nearby future. In this study, we assessed yield and a broad set of fruit quality attributes in four *Solanum pennellii* introgression lines, previously selected according to their drought resistance at seed level (IL 3-5 and IL 10-1, considered drought-resistant, and IL 7-1 and IL 2-5, considered drought-sensitive), under two distinct irrigation regimes, a deficit irrigation (DI) and an optimum irrigation (OI) treatment where plants were kept at 50 and 100% available soil water throughout the whole plant flowering and fruiting stages, respectively. Water deficit decreased plant yield over 66% and mean fruit weight over 53%. Fruit quality, on the other hand, was improved as indicated by the increase in total soluble solids, fruit redness, fruit firmness, and lycopene content. Poor fruit quality in terms of fruit color and taste was observed for IL 10-1 under OI and DI, respectively. IL 3-5, on the opposite, displayed interesting high fruit firmness under DI. Quantitative trait loci (QTLs) for enhanced lycopene content were detected for IL 2- 5 and IL 3-5 under OI and IL 2-5, IL 3-5, and IL 7-1 under DI. The presence of QTLs for reduced plant yield under DI in all studied ILs (IL 2-5, IL 3-5, IL 7-1, and IL 10-1) indicates no association between drought resistance at seed germination and initial development and plant productive stages. In sum, our study provides insights into the potential use of IL 2-5, IL 3-5, IL 7-1, and IL 10-1 for tomato improvement under both optimum and deficit irrigation conditions.

Keywords: tomato breeding; *Solanum pennellii;* water deficit; lycopene; fruit sensory attributes

Abbreviations: OI, optimum irrigation; DI, deficit irrigation; ASW, available soil water; PY, total plant yield; FW, mean fruit weight; TSS, total soluble solids; TA, titratable acidity.

2.1. Introduction

Water scarcity is a pressing issue threatening food production in many regions of the world. Breeding for drought-resistant materials is considered a low-cost strategy to ensure crop yield when water availability is limited or unpredictable (Bernier et al., 2008), and hence the center of several initiatives. Those breeding initiatives involve finding genetic materials carrying drought-resistance genes, transferring such genes into elite germplasm, and evaluating the impact of this resistance on crop yield and quality (Schauer et al., 2005; Petrovi´c et al., 2019). In tomatoes, related wild species such as *Solanum peruvianum, S. chilense* (Tapia et al., 2016), and *S. pennellii* (Bolger et al., 2014) are seen as important sources of drought-resistance genes since little genetic variation is found for this trait within the *S. lycopersicum* species (Foolad, 2007).

Studies involving drought resistance are, however, complex due to the quantitative nature of the character (Schneider et al., 1997). Introgression line (IL) populations have proven to be a key resource on such studies as they allow the identification of quantitative trait loci (QTLs) (Liu et al., 2003) without the need for molecular analyses that are costly and time-consuming. Eshed and Zamir (1995) developed an Introgression line population derived from an initial cross between the drought-resistant tomato wild relative *S. pennellii* (accession LA 716) and the processing tomato variety M82. Each IL is genetically identical to the variety M82 except for a defined genomic segment donated by the wild species so that any difference found between an IL and the variety M82 is solely attributed to the *S. pennellii'*s introgressed segment it contains. Currently, this population comprises 76 nearly isogenic lines covering the whole *S. pennellii* genome, which allows a complete study of the wild species' reservoir of genes (Lippman et al., 2007).

Despite being of great importance, genetic resistance to drought conditions is not the only aspect breeders should look into when improving tomato varieties. Drought-resistant materials must also exhibit other important traits, including satisfactory agronomic performance and good fruit quality. According to Causse et al. (2003), tomato fruit quality refers to a set of sensory (size, color, firmness, taste, aroma) and nutritional attributes (minerals, vitamins, and antioxidant

compounds, such as carotenoids and lycopene) that have an impact on consumers' purchase decision and satisfaction. Although neglected for many years as breeders targeted other traits, such as yield, disease resistance, and fruit shelf life, improved fruit quality is now an important requisite of the breeding process (Causse et al., 2010b).

It is widely known that water deficit affects yield and fruit quality attributes of tomato plants. The impact of water deficit on yield and fruit quality, however, depends on the genotype, the duration, and intensity of the stress, and the developmental stage which plants and fruits are in (Ripoll et al., 2016; Petrovi´c et al., 2019). Numerous studies assessing the whole *S. pennellii* IL population have already been performed in order to uncover QTLs for yield (Eshed and Zamir, 1994; Schauer et al., 2006) and fruit quality attributes, such as total soluble solids (Eshed and Zamir, 1994; Fridman et al., 2004; Schauer et al., 2006), flavor compounds (Tieman et al., 2006), fruit color (Liu et al., 2003), red fruit firmness (Yang et al., 2016), antioxidant molecules (Rousseaux et al., 2005), and metabolites (Schauer et al., 2006). However, with the exception of the study carried out by Liu et al. (2003), none of them investigated whether the QTLs for enhanced yield and fruit quality attributes found under optimum watering conditions are the same as those found under water deficit conditions. Moreover, most of these studies were able to assess only a limited number of fruit quality parameters, possibly due to the large size of the IL population.

Assuming the fact that to be considered drought resistance, a tomato variety should display a level of drought resistance on all developmental stages of plant growth, screening for drought resistance should start on germination and early seedling growth stages. Only genotypes with a fair level of drought resistance on germination and initial growth stages will then be tested for drought resistance on the next phases of plant development, including vegetative growth, flowering, fruit set and ripening stages. In this study, we investigated the effect of long-term deficit irrigation on plant yield and a wide range of fruit quality attributes of four selected tomato ILs (IL 2-5, IL 3-5, IL 7-1, and IL 10-1) from the *S. pennellii* IL population. The ILs were selected based on their level of resistance to drought on germination and early seedling growth stages (data not published), in which IL 3-5 and IL 10-1 displayed the highest level of drought resistance and IL 7-1 and IL 2-5 the lowest. Our results will provide information on the way and magnitude the long-term deficit irrigation affected yield and fruit quality attributes of the selected ILs, and whether the drought-resistant genotypes, IL 3-5 and IL 10-1, are able to sustain higher levels of yield and fruit quality in comparison to the susceptible genotypes IL 2-5 and IL 7-1, and therefore should move on the

next phases of our tomato breeding program. Moreover, since we included the processing tomato variety M82 on the trial, we will also be able to observe the presence of QTLs for yield and fruit quality attributes on both optimum and deficit irrigation treatments. A correlation analysis will also be performed to provide us information on the relationship between the studied traits.

2.2. Materials and methods

2.2.1. Plant materials and greenhouse experiment

The introgression lines (ILs) IL 3-5 and IL 10-1, considered drought-resistant, and IL 7-1 and IL 2-5, considered drought-sensitive in a previous study (data not published), and the processing tomato variety M82 were grown under two distinct irrigation treatments (50 and 100% available soil water) throughout the entire growing season.

The greenhouse experiment was conducted in the Research and Extension Farm Unit *Horta Velha* at Universidade Federal de Viçosa, Viçosa, MG, Brazil (20° 45’ 14’’ S; 42° 52’ 53’’ W; 648,74 m of altitude), from January to June 2018.

Seeds were sown in polystyrene trays of 128 cells each containing Tropstrato substrate (Vida Verde, BRA). Transplanting occurred when plants reached the 3-4 true leaf stage. Plants were grown into 15L-pots (1 plant per pot) inside a greenhouse to ensure that they had only access to the exact amount of water we provided. Pots contained a mixture of soil, sand, and composted cow manure (3:1:1), and were spaced 0.5 x 1.0 m apart. Soil texture was classified as sandy clay (sand: 52.9 %, silt: 4.5 %, clay: 42.6%). Fertilizations, as well as fungicide/insecticide applications, were based on crop recommendations. Plants were tied up to bamboo stakes placed inside each pot to prevent them from falling.

The experiment was arranged in a 2 (irrigation treatments) x 5 (genotypes) factorial scheme in a randomized block design with three replications. Each replication consisted of three plants arranged side by side. All variables were measured for the three plants per repetition separately and the mean values were used on the statistical analysis.

2.2.2. Irrigation management and calculation

The five genotypes were grown under two distinct irrigation treatments: a deficit irrigation treatment (DI), where soil water content was kept at 50% available soil water (ASW), and an optimum irrigation treatment (OI), where soil water content was kept at 100% ASW.

A soil sample was taken to estimate the soil water retention curve parameters based on the Van Genuchten (1980) equation using SWRC Fit software (Seki, 2007).

All pots were filled with the same dry soil weight (W_{ds}). Water weight (W_{water}) at DI and OI treatments in the first irrigation was determined by multiplying W_{ds} with the soil water content (kg/kg) at soil water potentials of -130 and -33 kPa, respectively. To estimate water tension values corresponding to 50% (-130 kPa) and 100% ASW (-33 kPa), we assumed that field capacity and wilting point were reached at matric soil water potentials of -33 and -1,500 kPa, respectively (Bernardo et al., 2006).

Soil water content was then monitored by weighing each pot daily. On the following irrigations, W_{water} applied was determined as total pot weight of reference ($W_{totalpot}$), which varied depending on the irrigation treatment applied, minus total pot weight measured on the day. Total pot weight of reference was calculated as follows: $W_{totalpot} = W_{container} + W_{ds} + W_{water} + W_{plant} + W_{bs}$, where: $W_{container}$= container weight, W_{ds} = dry soil weight, W_{water} = water weight, W_{plant} = plant weight, and W_{bs} = bamboo stake weight. W_{plant} was determined by weighing same-age spare plants grown within the experiment. Every ten days, one spare plant was harvested and weighed to adjust W_{plant}. Water was provided once a day early in the morning.

2.2.3. Measurements

2.2.3.1. Yield

Tomato fruits were harvested when fully ripened during the course of the experiment. Yield parameters consisted of total plant yield (Kg plant^{-1}) and mean fruit weight (g). For total plant yield, we measured the fresh weight of all fruits on each plant. Mean fruit weight consisted of the mean values of the 3 to 4 fully-ripened fruits picked randomly to perform the fruit quality assessment.

2.2.3.2. Fruit quality attributes

Fruit quality measurements were performed on 3 to 4 fully-ripened fruits per plant picked randomly. All the 3-4 fruits per plant were homogeneous in size and color and harvested on the same day. Fruit quality assessment began right after harvest.

The color numeric components L*, a*, and b*, from the L*a*b* CIELAB color space (Commission Internationale de l'Eclairage, 1978), were measured on two different spots of the fruit skin (180° apart from one another) of each selected fruit using a colorimeter (model CR-10, Konica Minolta, China). L* component indicates the lightness and darkness of color and ranges from black (0) to white (100). a* and b* components indicate color directions: +a* is the red direction; and -a* is the green direction; +b* is the yellow direction; and -b* is the blue di- rection. Chroma (saturation or vividness of color) and hue (tint of color) were calculated using the formulas $(a^{*2} + b^{*2})^{1/2}$ and $\tan^{-1}(b^*/a^*)$, respectively. The a*/b* ratio was used to express redness intensity, which has also been pointed as positively correlated with lycopene content (Toor et al., 2006). The higher a*/b* values, the redder the tomato fruit (López Camelo and Gómez, 2004). Fruit firmness, described as the mean maximum penetration force required for pericarp rupture and expressed in Newtons (N), was measured destructively using a digital penetrometer (model PDF-200, Soilcontrol, USA) fitted with a cylindrical stainless-steel probe with a round tip (Ø 8mm). Two measurements, located 180° apart from one another, were taken on the equatorial region of each fruit.

After color and firmness measurements, all 3-to-4 selected fruits were macerated together in a blender to produce the tomato juice used to determine total acidity (pH), total soluble solids (TSS), titratable acidity (TA), and lycopene content. pH values of the juice sample were determined instantly using a benchtop pH meter (model pH 21, Hanna Instruments, Italy). TSS, expressed as °Brix, was determined using a digital refractometer (model HI 96801, Hanna Instruments, Italy). For TA measurements, about 5 g of tomato juice were transferred to 100 mL- volumetric flasks filled with distilled water up to full capacity. An aliquot of 10 mL from this solution was then titrated with a NaOH solution (0.005 mol L^{-1}) using phenolphthalein 1% as indicator. TA values were expressed as % citric acid and obtained by the formula: TA (% citric acid) = [(mL NaOH used) x (0.005 N NaOH) x (0.064 mEq acid citric factor) x (correction-factor) x 100] / grams of juice sample. TSS/TA ratio was calculated as an indicator of flavor as described by Kader et al. (1978).

Lycopene content was determined by light spectrophotometry (model SP1105, Bel Photonics). Pigment extraction consisted of macerating 5 g of tomato juice with 15 mL of acetone for about 5 min, followed by vacuum-filtering. This procedure was repeated until samples became visibly colorless (about 3 to 4 times). After that, the pigment solution was transferred to a separatory funnel containing 25 mL of petroleum ether and washed with distilled water until total removal of the acetone. Lycopene content of the juice sample was determined as detailed by Rodriguez-Amaya (2001). According to this procedure, absorbance values are firstly adjusted to fit within a 0.2 to 0.8 range interval. In order to do that, 0.5 mL aliquots of the pigment solution were transferred to 5 or 10 mL-volumetric flasks according to the pigment concentration of each sample and diluted to volume with petroleum ether. Absorbance readings of the diluted samples were taken at 470 nm. Lycopene content, expressed as $\mu\ g^{-1}$, was then calculated by the formula: lycopene content ($\mu\ g^{-1}$) = [(Absorbance reading) x 25 mL x 10000 x (sample dilution)] / [(grams of juice sample) x 3450].

2.2.4. Statistical analysis

A two-way analysis of variance (ANOVA) was used to test the effect of the genotypes and the two irrigation treatments and their interactions on each studied trait. When differences were significant by the F test, means were compared using Tukey's test at a significance level of 0.05. To assess the presence of QTLs for yield and fruit quality attributes, we performed two separate analyses of variance (one for each irrigation treatment). When differences were found by the F test, mean values for each trait of the selected ILs were compared to the mean values of a common control (M82) (Dunnett's test at a significance level of 0.05). Pearson's linear correlation coefficients were also estimated and tested (t-test at 0.05 significance level) to examine the relationships between the studied traits on each irrigation treatment separately. Statistical analyses were performed using the Genes (Cruz, 2013) and R (version 3.6.0) software.

2.3. Results

2.3.1. Effect of genotype and irrigation treatments on yield and fruit quality attributes

Table 1 shows the significance of effects from the two-way analyses of variance performed for each yield and fruit quality attributes studied. Genotype x irrigation treatment interaction was significant for total plant yield (PY), mean fruit weight (FW), fruit firmness, total soluble solids (TSS), and lycopene content ($p<0.05$).

Table 1. Significance of effects from the two-way analysis of variance.

		Significance of effects													
Effects	**DF**	**PY**	**FW**	**L***	**a***	**b***	**a*/b***	**Chr**	**Hue**	**pH**	**F**	**TSS**	**TA**	**TSS/TA**	**Lyc**
Block	2														
Irrigation	1	**	**	ns	*	ns	*	ns	*	**	**	**	**	*	*
Genotype	4	**	**	**	**	**	*	**	*	**	**	ns	**	*	**
I x G	4	**	**	ns	ns	ns	ns	ns	ns	ns	**	**	ns	ns	**
Residue	18														

DF = Degrees of freedom; PY = total plant yield; FW = mean fruit weight; Chr = chroma; F = fruit firmness; TSS = total soluble solids; TA = titratable acidity; Lyc = lycopene. ns = not significant; ** and * = significant at 0.01 and 0.05 probability by the F test.

Yield parameters were drastically affected by the deficit irrigation treatment (DI) imposed, as shown in Figure 1. Reductions greater than 66% and 53% were observed for PY and FW, respectively. Highest PY was found for IL 3-5 (2.65 kg plant^{-1}) and lowest for IL 10-1 (1.58 kg plant^{-1}) and IL 2-5 (1.95 kg plant^{-1}) at optimum irrigation conditions (OI) (Figure 1a). At DI, on the other hand, no statistical differences for PY were found between genotypes (0.53 kg plant^{-1} on average) (Figure 1a). Mean FW of M82 (55g) was statistically identical to that of IL 3-5 (48.6g) and IL 7-1 (52g) under OI. Even though the FW of M82 under DI was almost half of that found at OI (26g), this genotype still showed better performance for this variable at DI conditions compared to the others (Figure 1b). IL 2-5 and IL 10-1 showed to produce small, low-weight fruits even at OI conditions (mean FW of IL 2-5 and IL10-1 were 24g and 22.8g at OI, and 8.6g and 8.8g at DI, respectively).

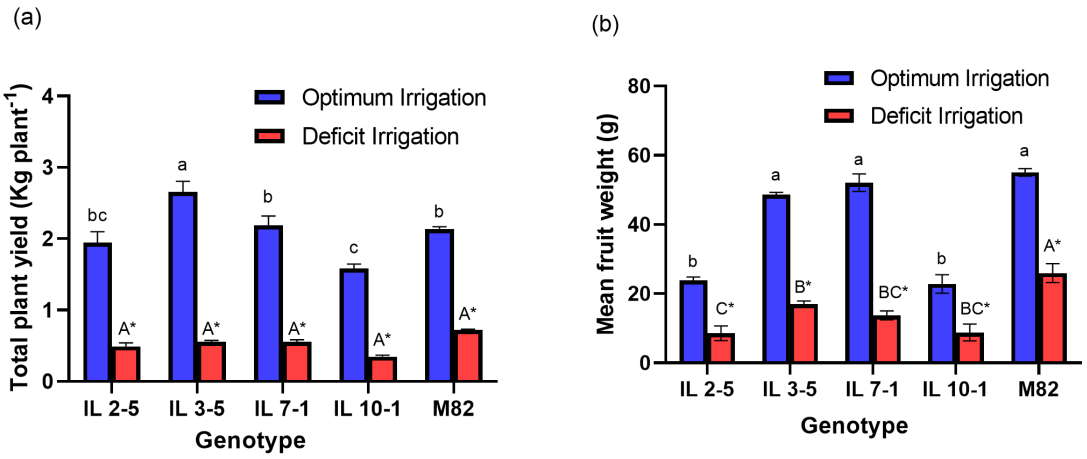


Figure 1. Total plant yield (PY) (a) and mean fruit weight (FW) (b) of five tomato genotypes (the ILs IL 3-5 and IL 10-1, considered drought resistant, IL 2-5 and IL 7-1, considered drought sensitive on a previous study, and the processing tomato variety M82) kept at two different irrigation treatments (50 and 100% ASW) throughout the entire growing season. Same lower- and upper-case letters indicate that the genotypes did not differ by the Tukey test ($p>0.05$), in the optimum irrigation (OI) (100% ASW) and in the deficit irrigation treatments (DI) (50% ASW), respectively. The asterisk (*) indicates statistical difference between OI and DI treatments for the same genotype. Data are expressed as mean ± standard error.

Genotype x irrigation treatment interaction was not significant for all color parameters (Table 1). By analyzing the effect of the irrigation treatment alone, we observed that the water deficit did not affect the variables L*, b*, and chroma ($p>0.05$). Still, it significantly affected a*, hue, and a*/b* ($p<0.05$) (Table 2). a* and a*/b* increased due to the water deficit while hue decreased. Higher values for a*/b* and lower values for hue in the DI than the OI treatment indicate that plants under water deficit produce fruits with a more intense red color. Except for IL 10-1, no statistical differences in fruit color were observed between the genotypes studied (Table 2). Overall, color parameters tended to be lower in IL 10-1 fruits compared to the others.

Table 2. Numeric fruit color parameters L*, a*, b*, a*/b*, Chroma, and Hue, using the "L*a*b*" CIELAB color space (Commission Internationale de l'Eclairage, 1978) of five tomato genotypes (the ILs IL 3-5 and IL 10-1, considered drought-resistant, IL 2-5 and IL 7-1, considered drought-sensitive on a previous study, and the processing tomato variety M82) kept at two different irrigation treatments (50 and 100% ASW) throughout the entire growing season. When genotype x irrigation treatment interaction was not significant by the F test ($p>0.05$), we studied both effects separately. Same lower-case letters indicate that the genotypes did not differ by the F or the Tukey's

test (p>0.05). Same upper-case letters indicate that there was no statistical difference between optimum (100% ASW) and deficit irrigation (50% ASW) treatments for the trait (p>0.05). Data are expressed as means ± standard error.

GENOTYPE	L*		
	Optimum Irrigation	Deficit Irrigation	Genotype overall mean
IL 2-5	32.00 ± 0.77	32.44 ± 0.31	32.22 a
IL 3-5	31.44 ± 0.86	33.09 ± 0.31	32.26 a
IL 7-1	32.50 ± 0.63	31.10 ± 0.60	31.80 a
IL 10-1	30.40 ± 0.59	29.30 ± 0.86	29.85 b
M82	33.15 ± 0.60	32.01 ± 0.32	32.58 a
Irrigation treatment overall mean	31.9 A	31.59 A	

GENOTYPE	a*		
	Optimum Irrigation	Deficit Irrigation	Genotype overall mean
IL 2-5	43.60 ± 0.90	46.95 ± 2.51	45.27 a
IL 3-5	45.82 ± 1.09	49.13 ± 1.28	47.48 a
IL 7-1	46.36 ± 0.62	46.73 ± 3.30	46.55 a
IL 10-1	35.50 ± 1.25	42.41 ± 1.03	38.96 b
M82	43.36 ± 0.26	45.14 ± 3.29	44.25 ab
Irrigation treatment overall mean	42.93 B	46.07 A	

GENOTYPE	b*		
	Optimum Irrigation	Deficit Irrigation	Genotype overall mean
IL 2-5	40.37 ± 0.63	40.70 ± 0.81	40.54 a
IL 3-5	37.65 ± 1.39	39.67 ± 0.90	38.66 ab
IL 7-1	41.35 ± 0.82	38.37 ± 0.36	39.86 a
IL 10-1	35.78 ± 2.12	36.14 ± 1.49	35.96 b
M82	39.68 ± 1.21	39.86 ± 1.61	39.77 a
Irrigation treatment overall mean	38.97 A	38.95 A	

GENOTYPE	a*/b*		
	Optimum Irrigation	Deficit Irrigation	Genotype overall mean
IL 2-5	1.08 ± 0.01	1.15 ± 0.05	1.12 ab
IL 3-5	1.22 ± 0.02	1.24 ± 0.05	1.23 a
IL 7-1	1.12 ± 0.04	1.22 ± 0.08	1.17 ab
IL 10-1	0.99 ± 0.03	1.18 ± 0.05	1.09 b
M82	1.09 ± 0.03	1.13 ± 0.07	1.11 ab

Irrigation treatment overall mean	1.10 B	1.18 A	

GENOTYPE	**Chroma**		
	Optimum Irrigation	**Deficit Irrigation**	**Genotype overall mean**
IL 2-5	59.42 ± 1.09	62.16 ± 2.29	60.79 a
IL 3-5	59.31 ± 1.71	63.17 ± 1.01	61.24 a
IL 7-1	62.14 ± 0.10	60.52 ± 2.74	61.33 a
IL 10-1	50.42 ± 2.30	55.75 ± 1.35	53.08 b
M82	58.79 ± 0.91	60.27 ± 3.18	59.53 a
Irrigation treatment overall mean	58.02 A	60.38 A	

GENOTYPE	**Hue**		
	Optimum Irrigation	**Deficit Irrigation**	**Genotype overall mean**
IL 2-5	42.80 ± 0.16	41.00 ± 1.20	41.90 ab
IL 3-5	39.39 ± 0.41	38.93 ± 1.09	39.16 a
IL 7-1	41.73 ± 0.94	39.55 ± 1.81	40.64 ab
IL 10-1	45.15 ± 1.01	40.41 ± 1.24	42.78 b
M82	42.43 ± 0.82	41.57 ± 1.77	42.00 ab
Irrigation treatment overall mean	42.30 A	40.29 B	

Table 3 shows means of fruit quality attributes, including fruit firmness, total acidity, total soluble solids (TSS), titratable acidity (TA), TSS:TA ratio, and lycopene content of the five tomato genotypes analyzed in two different irrigation conditions. For IL 2-5, IL 3-5, and IL 7-1, fruit firmness was higher at DI than at OI conditions (Table 3). However, for IL 10-1 and M82, no differences in fruit firmness regarding the irrigation treatment applied were found. At OI conditions, lowest fruit firmness was verified for IL 10-1, not differing from that of IL 2-5 and IL 3-5. At DI conditions, lowest fruit firmness was verified for M82, not differing from that of IL 2-5 and IL 7-1. In terms of total acidity, we observed a decrease in pH values when fruits were formed under DI conditions (Table 3). Fruit pH at DI (3.9 on average) was statistically lower than fruit pH at OI (4.1 on average). Lowest pH values were found for IL 10-1, not differing from IL 2-5 and IL 7-1. Opposite results were found for titratable acidity, in which fruit TA at DI (0.92% citric acid on average) was statistically higher than fruit TA at OI (0.46% citric acid on average). Highest TA

values were found for IL10-1 (0.85% citric acid), IL 2-5 (0.70% citric acid) and M82 (0.69% citric acid). IL10-1 differed from IL 3-5 and IL 7-1 in terms of fruit TA (Table 3). Although genotype x irrigation treatment interaction was significant, TSS content was significantly higher at DI conditions in all genotypes. TSS content was, on average, 73% higher in fruits subjected to DI than in fruits subjected to OI conditions. With the exception of IL 7-1 and IL10-1, which were statistically different in terms of TSS content in fruits, all genotypes showed similar TSS contents under OI conditions. Highest TSS values were observed for IL10-1 at OI, and they were statistically identical to those TSS values observed for IL 2-5, IL 3-5, and M82. Under DI conditions, on the other hand, no differences for TSS between genotypes were observed. TSS:TA ratio was lower at DI than it was at OI conditions (9.98 in average at DI and 8.66 at OI). IL 10-1 fruits had their flavor compromised, especially at DI conditions, as TSS/TA values were considerably lower (6.71). Unlike the other fruit quality attributes assessed in this study, lycopene content did not follow a predictable pattern. For IL 2-5, IL 10-1, and M82, for example, lycopene content was the same independently from the irrigation conditions fruits were formed in, while for IL 3-5 and IL 7-1, fruit lycopene content was different depending on the irrigation treatment. Moreover, the DI acted increasing lycopene content in IL 7-1 fruits while decreasing lycopene content in IL 3-5 fruits. Surprisingly high lycopene content was found for IL 2-5 (130.35 μg^{-1} fruit fresh weight in average) and IL 3-5 (153.55 μg^{-1} fruit fresh weight) at OI conditions and IL 2-5 (121.53 μg^{-1} fruit fresh weight) and IL 7-1 (133.67 μg^{-1} fruit fresh weight) at DI. Such lycopene values were about twice and three times as those observed for the parental line M82 at OI and DI conditions, respectively (62.34 μg^{-1} fruit fresh weight in average at OI and 40.09 μg^{-1} fruit fresh weight at DI).

Table 3. Fruit quality attributes firmness, total acidity, total soluble solids (TSS), titratable acidity (TA), TSS / TA ratio, and lycopene content, of five tomato genotypes (the ILs IL 3-5 and IL 10-1, considered drought-resistant, IL 2-5 and IL 7-1, considered drought-sensitive on a previous study, and the processing tomato variety M82) kept at two different irrigation treatments (50 and 100% ASW) throughout the entire growing season. When genotype x irrigation treatment interaction was not significant by the F test ($p>0.05$), we studied both effects separately. Same lower-case letters indicate that the genotypes did not differ by the F or the Tukey's test ($p>0.05$). Same upper-case letters indicate that there was no statistical difference between optimum (100% ASW) and deficit irrigation (50% ASW) treatments for the trait ($p>0.05$). Data are expressed as means ± standard error.

GENOTYPE	**Fruit firmness (N)**		
	Optimum Irrigation	**Deficit Irrigation**	**Genotype overall mean**
IL 2-5	17.67 ± 1.36 Bab	26.29 ± 1.73 Aab	21.98

IL 3-5	18.31 ± 1.08 Bab	30.53 ± 0.48 Aa	24.42
IL 7-1	19.83 ± 1.19 Ba	25.62 ± 1.94 Aab	22.73
IL 10-1	13.52 ± 1.32 Ab	16.79 ± 0.48 Ac	15.15
M82	20.93 ± 0.06 Aa	22.78 ± 1.44 Ab	21.85
Irrigation treatment overall mean	18.05	24.4	

GENOTYPE	**Total acidity (pH)**		
	Optimum Irrigation	**Deficit Irrigation**	**Genotype overall mean**
IL 2-5	4.09 ± 0.11	3.78 ± 0.03	3.94 ab
IL 3-5	4.22 ± 0.02	3.99 ± 0.02	4.11 a
IL 7-1	4.26 ± 0.04	3.94 ± 0.05	4.10 ab
IL 10-1	4.01 ± 0.09	3.81 ± 0.07	3.91 b
M82	4.22 ± 0.03	4.03 ± 0.06	4.12 a
Irrigation treatment overall mean	4.16 A	3.91 B	

GENOTYPE	**Total soluble solids (°Brix)**		
	Optimum Irrigation	**Deficit Irrigation**	**Genotype overall mean**
IL 2-5	4.61 ± 0.17 Bab	7.89 ± 0.49 Aa	6.25
IL 3-5	4.22 ± 0.26 Bab	8.39 ± 0.21 Aa	6.30
IL 7-1	3.80 ± 0.21 Bb	8.31 ± 0.15 Aa	6.05
IL 10-1	5.38 ± 0.63 Ba	6.96 ± 0.14 Aa	6.17
M82	4.50 ± 0.37 Bab	7.57 ± 0.25 Aa	6.04
Irrigation treatment overall mean	4.50	7.82	

GENOTYPE	**Titratable acidity (% citric acid)**		
	Optimum Irrigation	**Deficit Irrigation**	**Genotype overall mean**
IL 2-5	0.43 ± 0.04	0.96 ± 0.07	0.70 ab
IL 3-5	0.38 ± 0.02	0.85 ± 0.07	0.61 b
IL 7-1	0.38 ± 0.02	0.84 ± 0.03	0.61 b
IL 10-1	0.63 ± 0.04	1.06 ± 0.12	0.85 a
M82	0.47 ± 0.03	0.92 ± 0.06	0.69 ab
Irrigation treatment overall mean	0.46 B	0.92 A	

GENOTYPE	**TSS/TA**		
	Optimum Irrigation	**Deficit Irrigation**	**Genotype overall mean**
IL 2-5	10.74 ± 0.58	8.33 ± 0.88	9.54 ab
IL 3-5	11.11 ± 0.86	10.00 ± 0.65	10.56 a

IL 7-1	9.86 ± 0.33	9.98 ± 0.58	9.92 a
IL 10-1	8.52 ± 0.93	6.71 ± 0.62	7.62 b
M82	9.69 ± 1.21	8.27 ± 0.30	8.98 ab
Irrigation treatment overall mean	9.99 A	8.66 B	

GENOTYPE	**Lycopene (μ g^{-1} fruit fresh weight)**		
	Optimum Irrigation	**Deficit Irrigation**	**Genotype overall mean**
IL 2-5	130.35 ± 10.72 Aab	121.53 ± 5.89 Aa	125.94
IL 3-5	153.55 ± 8.41 Aa	73.89 ± 11.91 Bb	113.72
IL 7-1	97.85 ± 9.15 Bbc	133.67 ± 17.13 Aa	115.76
IL 10-1	45.27 ± 3.50 Ad	53.30 ± 3.99 Ab	49.29
M82	62.34 ± 9.64 Acd	40.09 ± 2.12 Ab	51.21
Irrigation treatment overall mean	97.87	84.5	

2.3.2. QTL analyses

As each IL is genetically identical to M82 except for a single homozygous defined genomic region, we assessed the presence of QTLs for yield and fruit quality attributes by comparing the mean values observed for each parameter of the selected ILs with the mean values of the variety M82 (Dunnett's test at 0.05 significance level) on each irrigation treatment separately. Such details are shown in Figures 2, 3, and 4. Asterisks indicate statistical differences between the M82 and the IL and, therefore, the presence of a QTL.

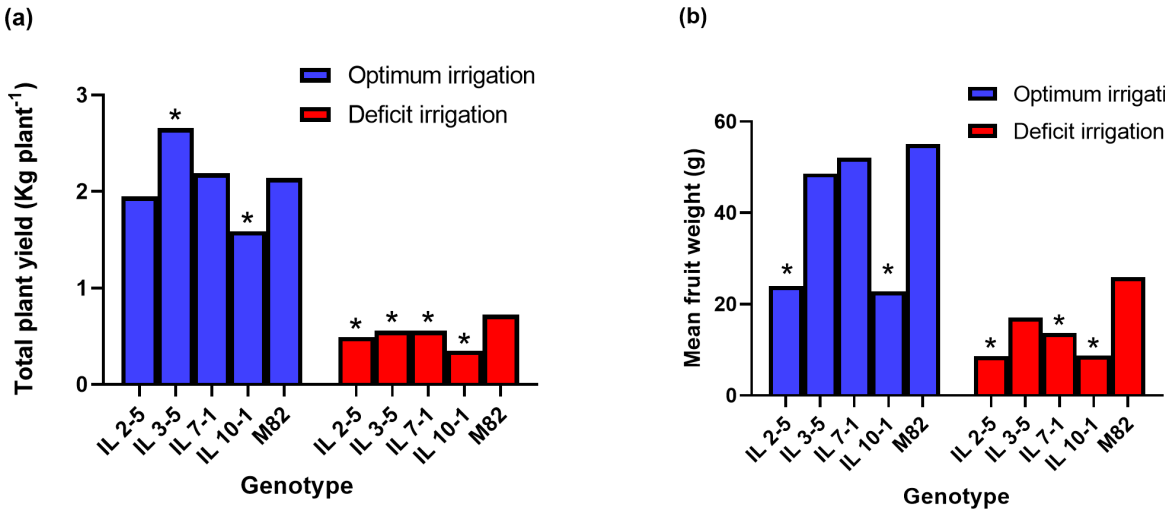


Figure 2. Mean values of total plant yield (PY) (a) and mean fruit weight (FW) (b) of four introgression lines (IL 3-5 and IL 10-1, considered drought resistant, IL 2-5 and IL 7-1 considered drought sensitive) compared to the mean values of a common control (variety M82) on each irrigation treatment separately. Asterisks (*) mean that the IL differed from M82 by the Dunnett's test ($p<0.05$).

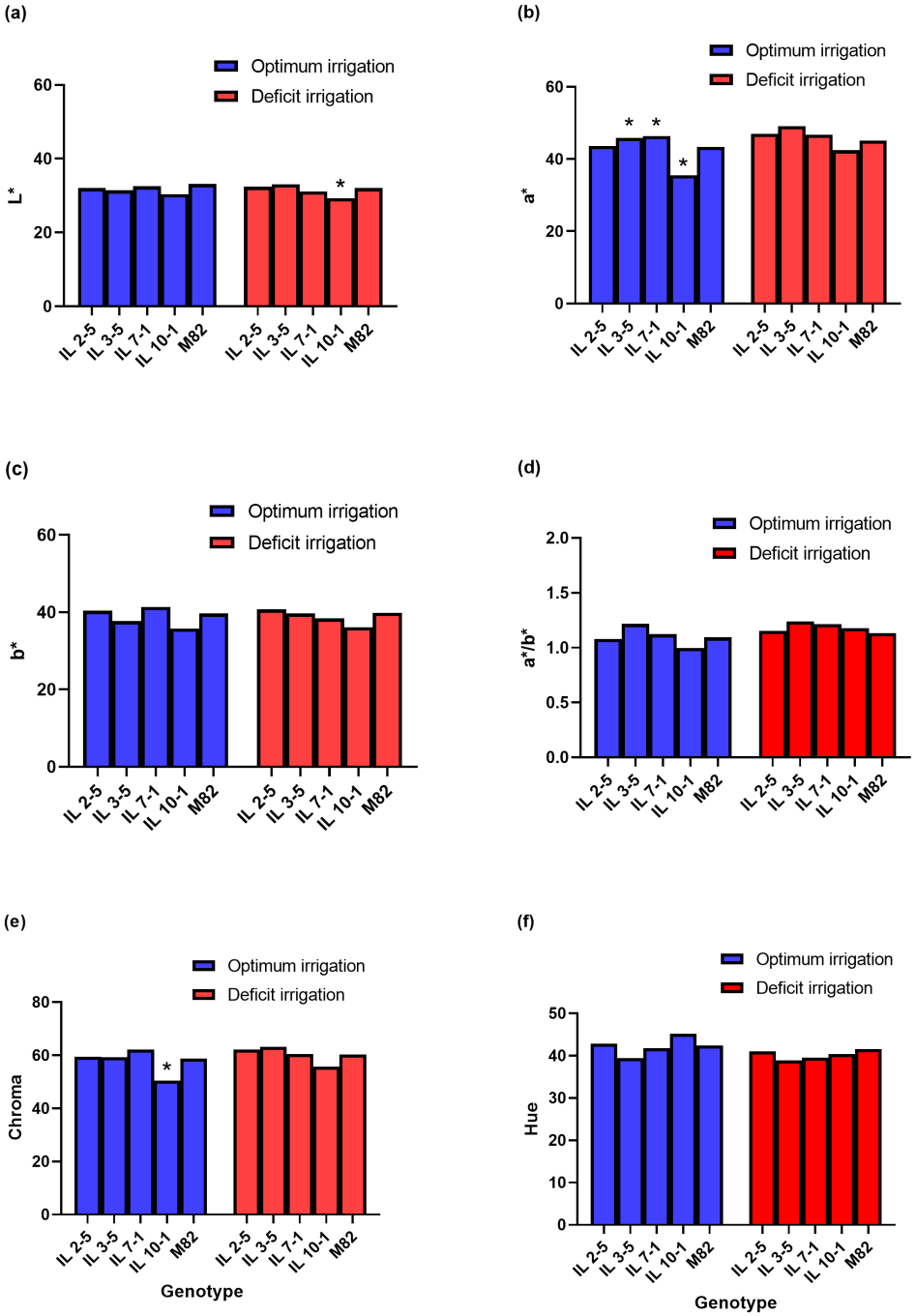


Figure 3. Mean values of numeric fruit color parameters L* (a), a* (b), b* (c), a*/b* (d), Chroma (e), and Hue (f), using the "L*a*b*" CIELAB color space (Commission Internationale de l'Eclairage, 1978) of four introgression lines (IL 3-5 and IL 10-1, considered drought resistant, IL 2-5 and IL 7-1 considered drought sensitive) compared to the mean values of a common control (variety M82) on each irrigation treatment separately. Asterisks (*) mean that the IL differed from M82 by the Dunnett's test ($p<0.05$).

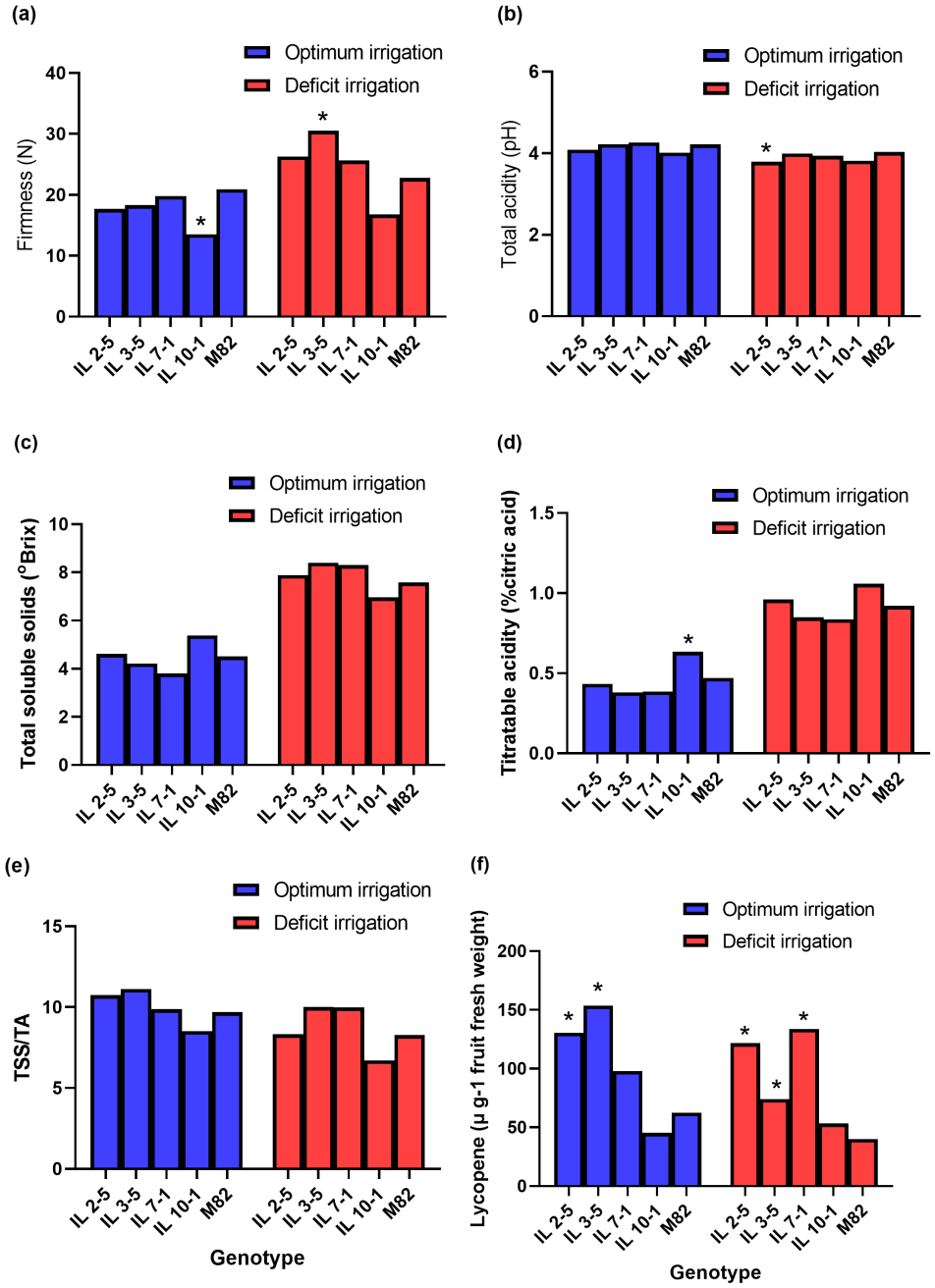


Figure 4. Mean values of fruit quality attributes firmness (a), total acidity (b), total soluble solids (TSS) (c), titratable acidity (TA) (d), TSS/TA (e), and lycopene content (f) of four introgression lines (IL 3-5 and IL 10-1, considered drought resistant, IL 2-5 and IL 7-1 considered drought sensitive) compared to the mean values of a common control (variety M82) on each irrigation treatment separately. Asterisks (*) mean that the IL differed from M82 by the Dunnett's test ($p<0.05$).

This study identified 12 QTLs for yield and fruit quality attributes at OI conditions and 13 QTLs at DI conditions. IL 3-5 contained a QTL for enhanced PY at OI conditions, while IL 10-1 contained a QTL for reduced PY (Figure 2a). Under DI conditions, all ILs contained QTLs for reduced plant yield. Similar results were also identified for mean fruit weight (Figure 2b). IL 10-1 also contained a QTL for reduced FW at OI conditions together with IL 2-5. Under DI conditions, all ILs showed QTLs for reduced FW except IL 3-5. QTLs for color parameters were identified for IL 10-1 (a QTL for decreased L* under DI, and decreased a* and chroma under OI), and IL 3-5 and IL 7-1 (both showed a QTL for enhanced a* under OI) (Figure 3). As for fruit firmness, IL 10-1 showed a QTL for decreased fruit firmness under OI, and IL 3-5 showed a QTL for increased fruit firmness under DI (Figure 4a). A QTL for decreased total acidity was found for IL 2-5 under DI, and a QTL for increased titratable acidity was found for IL 10-1 under OI. We did not identify QTLs for the variables b*, a*/b*, hue, TSS, and TSS/TA in neither irrigation condition. As expected, we found QTLs for enhanced lycopene content for IL 2-5 and IL 3-5 under OI and IL 2-5, IL 3-5, and IL 7-1 under DI (Figure 4f).

2.3.3. Correlation between traits

Our study on the association between yield and fruit quality attributes pointed out a total of 47 linear correlations under OI, in which 29 were positive and 18 negatives, and a total of 32 linear correlations under DI, in which 26 were positive and six were negative ($p<0.05$ by the t-test) (Figures 5 and 6). Overall, color parameters seemed to be strongly correlated with one another no matter the irrigation condition. Such results were already expected as some color variables (i.e., hue and chroma) are calculated as a function of others (i.e., a* and b*). Mean fruit weight showed a modest positive correlation ($r = 0.73$) with fruit firmness under OI, although such association was not maintained under DI. A similar pattern was observed for TSS, which correlated ($p<0.05$) with FW as well as PY under OI ($r = -0.54$ for FW and $r = -0.56$ for PY), but it did not correlate with any of them under DI ($p>0.05$). Total acidity, on the other hand, positively correlated with FW at both irrigation conditions ($r = 0.59$ under OI and $r = 0.73$ under DI). Lycopene content showed positive linear association with the color parameters a* ($r = 0.63$) and a*/b* ($r = 0.64$) under OI but surprisingly it did not happen under DI ($p>0.05$).

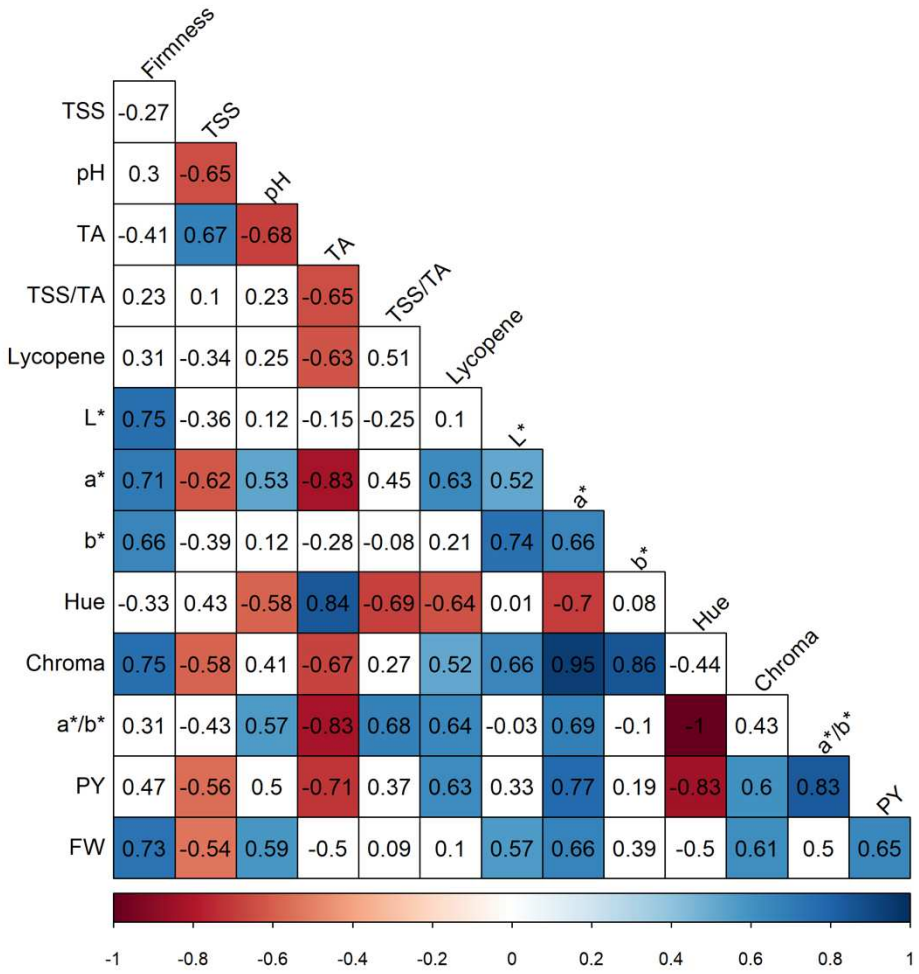


Figure 5. Pearson's correlation matrix between all studied traits at optimum irrigation conditions. Correlations were validated by the t test at 0.05 significance level. Blank squares mean that correlation was not significant ($p>0.05$).

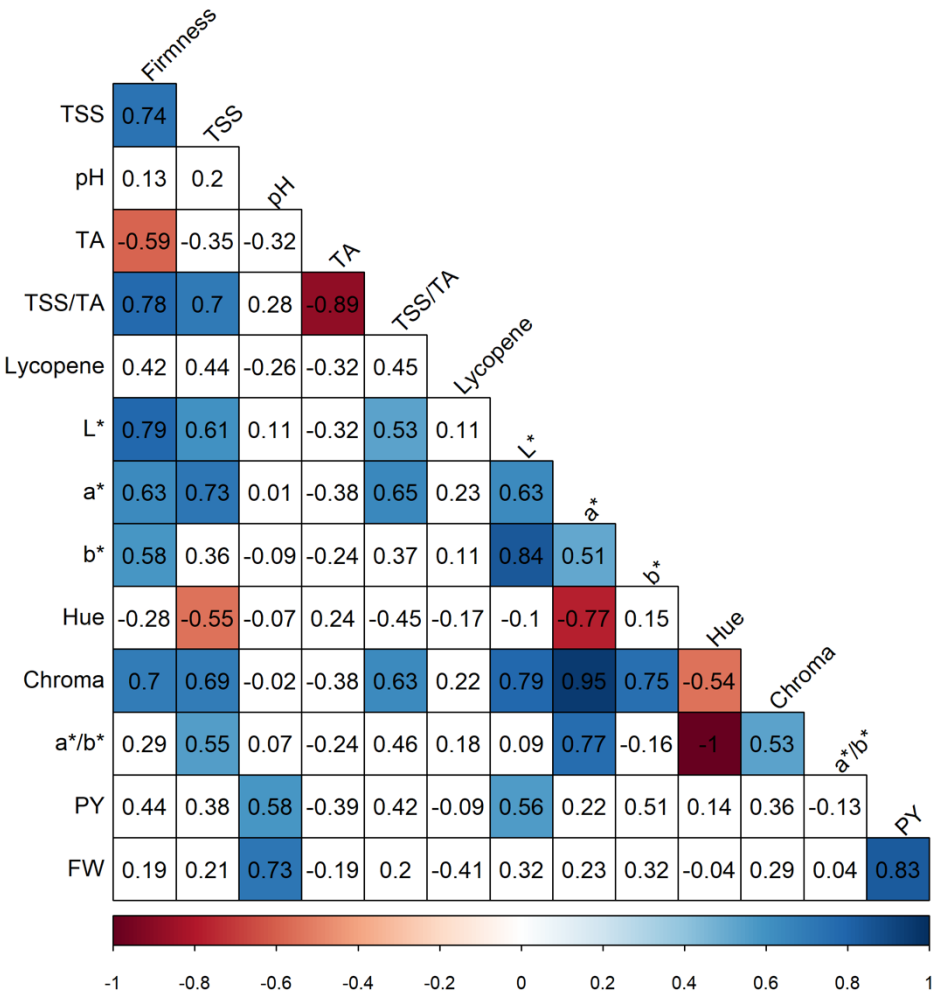


Figure 6. Pearson's correlation matrix between all studied traits at deficit irrigation conditions. Correlations were validated by the t test at 0.05 significance level. Blank squares mean that correlation was not significant ($p>0.05$).

2.4. Discussion

Plant breeding efforts consist of anticipating farmers' and consumers' needs and expectations for desirable genetic materials and plant products in the long run. If we consider future projections of water scarcity around the globe (Dai 2013), resistance to environmental constraints such as drought will be desirable for tomato growers in the following years as lack of water reduces

yield and fruit size and, therefore, profit (Petrovíc et al., 2019). However, for tomatoes, high yield is not the only guarantee of profit since consumers are very demanding in terms of fruit appearance (fruit shape, size, and firmness), taste, and flavor intensity (Causse et al., 2010a). Stevens et al. (1979) suggested breeding for cultivars with high levels of soluble solids and acids as they are the main components of taste. Recently, breeding programs have also been giving special attention to the nutritional quality of tomatoes (Bai and Lindhout, 2007).

Water is the main component of herbaceous plants accounting for 90% of their total fresh weight (Kramer and Boyer, 1995). A mature tomato fruit is composed of 93% to 95% water (Gameiro et al., 2007). It is within water that all chemical and biochemical processes of cells, including photosynthesis, takes place. Water deficit decreases turgor pressure to a minimum, stopping cell division and enlargement as well as stem and leaf elongation. Lack of water also inhibits photosynthesis, closes stomata, decreases respiration rates and other enzyme-mediated processes (Kramer; Boyer, 1995). Thus, it is not surprising that providing a reduced water supply to tomato plants would reduce their growth and yield.

Similar to what we observed here, several other studies with tomatoes also reported yield losses and differences in fruit quality due to water limitation (Petrovíc et al., 2019; Chen et al., 2013; Medyouni et al., 2021). The magnitude of yield loss, however, depends on the degree and duration of the stress and the developmental stage plants subjected to (Chen et al., 2013; Ghannem et al., 2020). Ghannem et al. (2020) observed that moderate water stress during the vegetative phase did not affect final yield of greenhouse-grown potted tomatoes. As for Keabetswe et al. (2019), moderate water stress decreased yield when applied during flowering and fruiting stages. The water stress applied to the plants in this study can be considered severe as over 66% and 53% reduction in plant yield and mean fruit weight were observed, respectively. These high yield losses can be attributed not only to the degree of the stress but also to the fact that plants spent their whole flowering and fruiting stages under deficit irrigation conditions. By choosing to keep our stressed plants under 50% ASW, we intended to apply a deficit irrigation high enough to avoid selecting a genotype as drought-resistant when in fact it is sensitive but not too high that we could not discriminate genotypes according to their drought resistance level. Although there were no differences in plant yield between genotypes under DI in the two-way analysis of variance, our QTL analysis for the DI condition separately clearly differentiated the selected ILs from the parental line M82 indicating that the water stress intensity chosen served perfectly to our purpose.

Our QTL analysis also indicates that IL 3-5 and IL10-1, previously selected as drought-resistant at seed level, cannot be considered drought-resistant at plant productive stages. A QTL for enhanced fruit yield in IL 3-5 at OI conditions suggests this genotype as an important genetic resource for yield improvement under well-watered conditions.

Despite losses in yield and fruit size, water deficit seems to improve fruit quality of tomatoes (Keabetswe et al., 2019). Increases in total soluble solids, fruit redness, fruit firmness, and lycopene content are some of the benefits of water deficit (Zegbe-Domínguez et al., 2003; Favati et al., 2009; Chen et al., 2013; Romero and Rose, 2019) exactly as we observed in this study. Water accumulation in tomato fruits is ex- expected to reduce with less water being provided to plants resulting in higher concentrations of sugars, acids, and pigments (Petrovíc et al., 2019). Notice that we also observed an increase in acid concentrations due to water deficit, indicated by lower fruit pH and higher TA values in the DI treatment. Moreover, water deficit is known for triggering the synthesis of antioxidant compounds such as Vitamin C and carotenoids (Dorais et al., 2008; Fanciullino et al., 2014) in order to overcome the dangerous effects of reactive oxygen species (ROS) created during stress, which may also help us to understand why some stressed fruits had higher lycopene content than non-stressed fruits. As lycopene quenching rate towards oxygen singlet is twice as higher as that of β-carotene and ten times higher than that of α-tocopherol (Kong et al., 2010), it is more advantageous for plants to invest in lycopene synthesis compared to other carotenoids during stress conditions.

Fruit color is a crucial factor ensuring tomato marketing for fresh consumption. For Sacks and Francis (2001), breeding efforts for tomato color should be made towards achieving fruits that are red, dark, and saturated in color, which means fruits with decreased L* and hue, and increased chroma. The higher the a* and a*/b* parameters the redder the fruit (López Camelo and Gómez, 2004). The lower the hue angle the redder the fruit (Zegbe-Domínguez et al., 2003). Higher a* and a*/b* values and lower hue angles recorded for the DI condition in comparison to the OI condition indicate that water deficit improves fruit color, which is consistent with the findings of Chen et al. (2013), Favati et al. (2009) and Zegbe-Domínguez et al. (2003). Lycopene is the pigment responsible for the red color of tomato fruits (Kong et al., 2010), and hence positive linear associations between a*, a*/b*, and lycopene as those observed in this study are commonly reported (Arias et al., 2000; Dumas et al., 2003; Brandt et al., 2006; Favati et al., 2009). Significant changes in a* are recorded during chlorophyll degradation and lycopene synthesis (López Camelo

and Gómez, 2004). The presence of QTLs for reduced a* and chroma reveals poor fruit color for IL 10-1 under OI.

Fruit firmness is considered an essential attribute not only to the fresh market but also to the processing industry. According to Ferreira (2010), consumers prefer fruits that are firm to the touch and do not deform easily. Firmer tomatoes also tolerate long-distance transportation to the processing factories without pericarp rupture. Digital penetrometers are often used to assess tomato fruit firmness (Barrett et al., 1998; Yang et al., 2016). Yang et al. (2016) assessed red fruit firmness of the whole *S. pennellii* IL population under well-watered conditions at two consecutive-year trials. These authors identified 25 QTLs, 9 for increased and 16 for decreased red fruit firmness. Despite the fact that we used different methods to quantify fruit firmness (we measured fruit firmness through puncturing by a penetrometer while they measured through flat-plate compression), our findings were remarkably consistent (IL 2-5, IL 3-5, and IL 7-1 did not contain QTLs for red fruit firmness whereas IL 10-1 contains a QTL for decreased red fruit firmness under OI). The presence of a QTL for enhanced red fruit firmness in IL 3-5 only at DI conditions reveals different gene expression as plants are cultivated under different irrigation treatments. Fruit softening is a natural-occurring process during ripening considered a result of polysaccharide degradation in the cell wall and middle lamella (Uluisik et al., 2016). Fruits with lower cuticle water permeability tend to stay firm for long periods as tissue collapse due to fruit desiccation does not occur. Water deficit is known for substantially decreasing fruit transpirational water loss by increasing cuticle load and thickness, making fruits firmer (Romero and Rose, 2019). Romero and Rose (2019) showed that water deficit stress increases the expression of cuticle-related genes in a genotype-specific way, explaining why only IL 2-5, IL 3-5, and IL 7-1 fruits had their firmness improved due to water deficit.

Increasing TSS content in fruits is often the main target of breeding programs as it determines the efficiency of tomato processing. The higher the TSS content, the lower the energy required to evaporate water from fruits in the preparation of tomato paste and concentrated juice (Bennett, 2012). TSS and acid contents are also important for fresh consumption as they are major components of overall flavor intensity (Causse et al., 2002). Grierson and Kader (1986) suggests sugar:acid ratio as a good indicator of tomato flavor. TSS content of today's top tomato hybrids ranges from 3.6 to 5.0 ℃B, values that are well below ideal (Nick and Silva, 2016). Under DI, TSS content significantly increased due to water deficit, but it happened at the very expense of fresh

fruit weight, which is not interesting for breeding purposes. The inverse relationship between TSS and yield, as we found here, is often reported in research papers and remains a big challenge for tomato breeders (Grandillo et al., 1999). The absence of QTLs in both irrigation treatments suggests that these ILs have no use for TSS improvement. By studying TSS within the *S. pennellii* ILs, Eshed and Zamir (1994) reported QTLs for enhanced TSS content on chromosomes 1, 5, and 7.

For tomato fruits, ideal ranges are 3.7 - 4.5 for pH (Sulieman et al., 2011; Tudor-Radu et al., 2016), 0.4 – 0.6% citric acid for TA, and 8 – 12 for SST/TA (Divéky-Ertsey et al., 2012). SST/TA above 12, tomatoes are sour while SST/TA below 8, tomatoes are tasteless (Divéky-Ertsey et al., 2012). All pH values recorded in this study belonged to the ideal interval, whereas TA values were higher than those recommended under DI. Although TA was significantly higher at DI conditions and yet outside of the ideal range (0.92% citric acid), it had little influence in taste as all SST/TA values were between 8 and 12, except for IL 10-1 that had its taste compromised under DI (SST/TA = 6.71 in average).

Outstanding lycopene content values observed for IL 2-5 and IL 3-5 under OI and IL 2-5, IL 3-5, and IL 7-1 under DI compared to M82 suggest the presence of lycopene-enhancing genes in these materials. The high antioxidant properties of lycopene, which have also been associated with reduced risk of some types of cancer (Van Breemen and Pajkovic, 2008), together with its high consumption through tomato products, awakens the breeding sector's attention for lycopene-enriched fruits. Major QTLs for lycopene synthesis in the tomato species have already been reported in chromosomes 7 and 12 (Ashrafi et al., 2012). Differently from what we observed in this study, Rousseaux et al. (2005) detected only four QTLs for lycopene content within the *S. pennellii* IL population (IL 3-2, IL 6-2, IL 12-2 and IL 12-3), all of them responsible for decreasing lycopene concentration on tomato fruits. Such inconsistency of results may be explained by the fact that plants were grown under different environmental conditions, and lycopene synthesis is strongly affected by temperature and direct radiation on fruits (Brandt et al., 2006). According to Brandt et al. (2006), lycopene synthesis is inhibited over 30°C, and radiations of about 2990 mmol m^{-2} s^{-1} for 1.5 to 4 hours are harmful. Even Rousseaux et al. (2005) mentioned differences in QTL expression over the years. Yet, the greenhouse QTL assessment seemed to be more precise as variations due to environmental causes were much lower compared to field assessments (Rousseaux et al., 2005).

2.5. Conclusion

The deficit irrigation applied to plants in this study reduced yield and mean fruit weight of all genotypes compared to optimum irrigation conditions while positively affecting fruit quality attributes. Overall, the plants under severe water deficit during flowering and fruiting stages produced redder, firmer fruits with higher concentrations of total soluble solids and lycopene. IL 10-1 showed poor fruit quality in terms of fruit color and taste under optimum irrigation and deficit irrigation conditions, respectively. Common associations such as the positive correlations between lycopene content and chromaticity parameters a* and a*/b* and the negative correlation between total soluble solids and plant yield at well-watered conditions were reinforced. Interestingly, such relationships were not maintained under long-term deficit irrigation conditions. The presence of QTLs for enhanced-lycopene content in IL 2-5 and IL 3-5 under OI and IL 2-5, IL 3-5, and IL 7-1 under DI suggests these ILs as important gene bank for lycopene improvement. Yet, we detected a QTL for increased fruit firmness in IL 3-5 under deficit irrigation conditions. As QTLs for reduced plant yield under DI were found for all the selected ILs, none can be considered drought-resistant at plant productive stages.

2.6. References

Arias, R., Lee, T.-C., Logendra, L., Janes, H., 2000. Correlation of Lycopene Measured by HPLC with the L*, a*, b* Color Readings of a Hydroponic Tomato and the Relationship of Maturity with Color and Lycopene Content. J. Agric. Food Chem. 48, 1697–1702. https://doi.org/10.1021/jf990974e.

Ashrafi, H., Kinkade, M.P., Merk, H.L., Foolad, M.R., 2012. Identification of novel quantitative trait loci for increased lycopene content and other fruit quality traits in a tomato recombinant inbred line population. Mol. Breed. 30, 549–567. https://doi. org/10.1007/s11032-011-9643-1.

Bai, Y., Lindhout, P., 2007. Domestication and breeding of tomatoes: What have we gained and what can we gain in the future? Ann. Bot. 100, 1085–1094. https://doi. org/10.1093/aob/mcm150.

Barrett, D.M., Garcia, E., Wayne, J.E., 1998. Textural modification of processing tomatoes. Crit. Rev. Food Sci. Nutr. 38, 173–258. https://doi.org/10.1080/ 10408699891274192.

Bennett, A.B., 2012. Taste: Unraveling tomato flavor. Curr. Biol. 22, R443–R444. https://doi.org/10.1016/j.cub.2012.04.017.

Bernardo, S., Soares, A.A., Mantovani, E.C., 2006. Manual de irrigação, eighth ed. UFV, Viçosa.

Bernier, J., Atlin, G.N., Serraj, R., Kumar, A., Spaner, D., 2008. Breeding upland rice for drought resistance. Journal of the science of food and agriculture 88, 927–939.

Bolger, A., Scossa, F., Bolger, M.E., Lanz, C., Maumus, F., Tohge, T., Quesneville, H., Alseekh, S., Sørensen, I., Lichtenstein, G., Fich, E.A., Conte, M., Keller, H., Schneeberger, K., Schwacke, R., Ofner, I., Vrebalov, J., Xu, Y., Osorio, S., Aflitos, S. A., Schijlen, E., Jim´enez-Gom´ez, J.M., Ryngajllo, M., Kimura, S., Kumar, R., Koenig, D., Headland, L.R., Maloof, J.N., Sinha, N., Van Ham, R.C.H.J., Lankhorst, R. K., Mao, L., Vogel, A., Arsova, B., Panstruga, R., Fei, Z., Rose, J.K.C., Zamir, D., Carrari, F., Giovannoni, J.J., Weigel, D., Usadel, B., Fernie, A.R., 2014. The genome of the stress-tolerant wild tomato species Solanum pennellii. Nat. Genet. 46, 1034–1038. https://doi.org/10.1038/ng.3046.

Brandt, S., Pék, Z., Barna, É., Lugasi, A., Helyes, L., 2006. Lycopene content and colour of ripening tomatoes as affected by environmental conditions. J. Sci. Food Agric. 86, 568–572. https://doi.org/10.1002/jsfa.2390.

Causse, M., Buret, M., Robini, K., Verschave, P., 2003. Inheritance of nutritional and sensory quality traits in fresh market tomato and relation to consumer preferences. J. Food Sci. 68, 2342–2350 https://doi.org/Article.

Causse, M., Friguet, C., Coiret, C., LéPicier, M., Navez, B., Lee, M., Holthuysen, N., Sinesio, F., Moneta, E., Grandillo, S., 2010a. Consumer Preferences for Fresh Tomato at the European Scale: A Common Segmentation on Taste and Firmness. J. Food Sci. 75, 531–541. https://doi.org/10.1111/j.1750-3841.2010.01841.x.

Causse, M., Friguet, C., Coiret, C., LéPicier, M., Navez, B., Lee, M., Holthuysen, N., Sinesio, F., Moneta, E., Grandillo, S., 2010b. Consumer Preferences for Fresh Tomato at the European Scale: A Common Segmentation on Taste and Firmness. J. Food Sci. 75 https://doi.org/10.1111/j.1750-3841.2010.01841.x.

Causse, M., Saliba-Colombani, V., Lecomte, L., Duffé, P., Rousselle, P., Buret, M., 2002. QTL analysis of fruit quality in fresh market tomato: a few chromosome regions control the variation of sensory and instrumental traits. J. Exp. Bot. 53, 2089–2098. https://doi.org/10.1093/jxb/erf058.

Chen, J., Kang, S., Du, T., Qiu, R., Guo, P., Chen, R., 2013. Quantitative response of greenhouse tomato yield and quality to water deficit at different growth stages. Agric. Water Manag. 129, 152–162. https://doi.org/10.1016/j.agwat.2013.07.011.

Cruz, C.D., 2013. GENES - Software para análise de dados em estatística experimental e em genética quantitativa. Acta Sci. - Agron. 35, 271–276. https://doi.org/10.4025/ actasciagron.v35i3.21251.

Dai, A., 2013. Increasing drought under global warming in observations and models. Nat. Clim. Chang. 3, 52–58. https://doi.org/10.1038/nclimate1633.

de l'Eclairage, C.I., 1978. Recommendations on uniform color spaces, color-difference equations, psychometric color terms. Paris CIE.

Diveky-Ertsey, A., Csambalik, L., Kókai, Z., Stefanovits-Bányai, É., Pap, Z., Krisztiánne Kis, M., Sipos, L., 2012. Antioxidant, polyphenol and sensory analysis of cherry tomato varieties and landraces. Int. J. Hortic. Sci. 18, 75–80. https://doi.org/ 10.31421/ijhs/18/1/997.

Dorais, M., Ehret, D.L., Papadopoulos, A.P., 2008. Tomato (Solanum lycopersicum) health components: From the seed to the consumer. Phytochem. Rev. 7, 231–250. https://doi.org/10.1007/s11101-007-9085-x.

Dumas, Y., Dadomo, M., Di Lucca, G., Grolier, P., 2003. Effects of environmental factors and agricultural techniques on antioxidant content of tomatoes. J. Sci. Food Agric. 83, 369–382. https://doi.org/10.1002/jsfa.1370.

Eshed, Y., Zamir, D., 1995. An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141, 1147–1162.

Eshed, Y., Zamir, D., 1994. Introgressions from Lycopersicon pennellii can improve the soluble-solids yield of tomato hybrids. Theor. Appl. Genet. 88, 891–897. https://doi. org/10.1007/BF01254002.

Fanciullino, A.L., Bidel, L.P.R., Urban, L., 2014. Carotenoid responses to environmental stimuli: Integrating redox and carbon controls into a fruit model. Plant, Cell Environ 37, 273–289. https://doi.org/10.1111/pce.12153.

Favati, F., Lovelli, S., Galgano, F., Miccolis, V., Di Tommaso, T., Candido, V., 2009. Processing tomato quality as affected by irrigation scheduling. Sci. Hortic. (Amsterdam). 122, 562–571. https://doi.org/10.1016/j.scienta.2009.06.026.

Ferreira, S.M.R., Quadros, D.A.De, Karkle, E.N.L., Lima, J.J.De, Tullio, L.T., Freitas, R.J. S.De, 2010. Qualidade pós-colheita do tomate de mesa convencional e orgˆanico. Ciˆencia e Tecnol. Aliment. 30, 858–869. https://doi.org/10.1590/S0101- 20612010000400004.

Foolad, M.R., 2007. Genome mapping and molecular breeding of tomato. Int. J. Plant Genomics 2007. https://doi.org/10.1155/2007/64358.

Fridman, E., Carrari, F., Liu, Y.S., Fernie, A.R., Zamir, D., 2004. Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science 305, 1786–1789. https://doi.org/10.1126/science.1101666.

Gameiro, A.H., Filho, Caixeta, J., V, Rocco, C.D., Rangel, R., 2007. Processamento Industrial No Estado De Goia´s. Informações Econômicas 37, 7–16.

Ghannem, A., Ben Aissa, I., Majdoub, R., 2020. Effects of regulated deficit irrigation applied at different growth stages of greenhouse grown tomato on substrate moisture, yield, fruit quality, and physiological traits. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-020-10407-w.

Grandillo, S., Zamir, D., Tanksley, S.D., 1999. Genetic improvement of processing tomatoes: A 20 years perspective. Euphytica 110, 85–97. https://doi.org/10.1023/ A:1003760015485.

Grierson, D., Kader, A.A., 1986. Fruit ripening and quality. In: Atherton, J.G., Rudich, J. (Eds.), The Tomato Crop. Chapman and Hall, London, pp. 241–280.

Kader, A.A., Morris, L.L., Stevens, M.A., Albright-Holton, M., 1978. Composition and Flavor Quality of Fresh Market Tomatoes as Influenced by Some Postharvest Handling Procedures. J. Am. Soc. Hortic. Sci.

Keabetswe, L., Shao, G.C., Cui, J., Lu, J., Stimela, T., 2019. A combination of biochar and regulated deficit irrigation improves tomato fruit quality: A comprehensive quality analysis. Folia Hortic 31, 181–193. https://doi.org/10.2478/fhort-2019-0013.

Kong, K.W., Khoo, H.E., Prasad, K.N., Ismail, A., Tan, C.P., Rajab, N.F., 2010. Revealing the power of the natural red pigment lycopene. Molecules 15, 959–987. https://doi. org/10.3390/molecules15020959.

Kramer, P.J., Boyer, J.S., 1995. Functions and properties of water. In: Kramer, P.J., Boyer, J.S. (Eds.), Water relations of plants and soils. Academic Press, California, pp. 20–21.

Lippman, Z.B., Semel, Y., Zamir, D., 2007. An integrated view of quantitative trait variation using tomato interspecific introgression lines. Curr. Opin. Genet. Dev. 17, 545–552. https://doi.org/10.1016/j.gde.2007.07.007.

Liu, Y.-S., Gur, A., Ronen, G., Causse, M., Damidaux, R., Buret, M., Hirschberg, J., Zamir, D., 2003. There is more to tomato fruit colour than candidate carotenoid genes. Plant Biotechnol. J. 1, 195–207. https://doi.org/10.1046/j.1467- 7652.2003.00018.x.

López Camelo, A.F., Gómez, P.A., 2004. Comparison of color indexes for tomato ripening. Hortic. Bras. 22, 534–537. https://doi.org/10.1590/s0102- 05362004000300006.

Medyouni, I., Zouaoui, R., Rubio, E., Serino, S., Ahmed, H.Ben, Bertin, N., 2021. Effects of water deficit on leaves and fruit quality during the development period in tomato plant. Food Sci. Nutr. 1–12 https://doi.org/10.1002/fsn3.2160.

Nick, C., Silva, D.J.H., 2016. Melhoramento de Tomate. In: Nick, C., Bor´em, A. (Eds.), Melhoramento de Hortaliças. UFV, Viçosa, pp. 396–431.

Petrovíc, I., Savíc, S., Jovanovíc, Z., Stikíc, R., Brunel, B., Sérino, S., Bertin, N., 2019. Fruit quality of cherry and large fruited tomato genotypes as influenced by water deficit. Zemdirbyste-Agriculture 106, 123–128. https://doi.org/10.13080/z- a.2019.106.016.

Ripoll, J., Urban, L., Brunel, B., Bertin, N., 2016. Water deficit effects on tomato quality depend on fruit developmental stage and genotype. J. Plant Physiol. 190, 26–35. https://doi.org/10.1016/j.jplph.2015.10.006.

Rodriguez-Amaya, D., 2001. A Guide to Carotenoids Analysis in Food. International Life Sciences Institute Press, Washington.

Romero, P., Rose, J.K.C., 2019. A relationship between tomato fruit softening, cuticle properties and water availability. Food Chem 295, 300–310. https://doi.org/ 10.1016/j.foodchem.2019.05.118.

Rousseaux, M.C., Jones, C.M., Adams, D., Chetelat, R., Bennett, A., Powell, A., 2005. QTL analysis of fruit antioxidants in tomato using Lycopersicon pennellii introgression lines. Theor. Appl. Genet. 111, 1396–1408. https://doi.org/10.1007/s00122-005- 0071-7.

Sacks, E.J., Francis, D.M., 2001. Genetic and environmental variation for tomato flesh color in a population of modern breeding lines. J. Am. Soc. Hortic. Sci. 126, 221–226. https://doi.org/10.21273/jashs.126.2.221.

Schauer, N., Semel, Y., Roessner, U., Gur, A., Balbo, I., Carrari, F., Pleban, T., Perez- Melis, A., Bruedigam, C., Kopka, J., Willmitzer, L., Zamir, D., Fernie, A.R., 2006. Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nat. Biotechnol. 24, 447–454. https://doi.org/ 10.1038/nbt1192.

Schauer, N., Zamir, D., Fernie, A.R., 2005. Metabolic profiling of leaves and fruit of wild species tomato: A survey of the Solanum lycopersicum complex. J. Exp. Bot. 56, 297–307. https://doi.org/10.1093/jxb/eri057.

Schneider, K.A., Brothers, M.E., Kelly, J.D., 1997. Marker-assisted selection to improve drought resistance in common bean. Crop Sci 37, 51–60. https://doi.org/10.2135/ cropsci1997.0011183X003700010008x.

Seki, K., 2007. SWRC fit – a nonlinear fitting program with a water retention curve for soils having unimodal and bimodal pore structure. Hydrol. Earth Syst. Sci. Discuss 4, 407–437.

Stevens, M., Kader, A.a., Albright-Holton, M., 1979. Potential for increasing tomato flavor via increased sugar and acid content. J. Amer. Soc. Hort. Sci.

Sulieman, A.M.E., Awn, K.M.A., Yousif, M.T., 2011. Suitability of Some Tomato (Lycopersicon esculentum Mill.) genotypes for Paste Production 12, 45–51.

Tapia, G., Méndez, J., Inostroza, L., 2016. Different combinations of morpho-physiological traits are responsible for tolerance to drought in wild tomatoes Solanum chilense and Solanum peruvianum. Plant Biol 18, 406–416. https://doi. org/10.1111/plb.12409.

Tieman, D.M., Zeigler, M., Schmelz, E.A., Taylor, M.G., Bliss, P., Kirst, M., Klee, H.J., 2006. Identification of loci affecting flavour volatile emissions in tomato fruits. J. Exp. Bot. 57, 887–896. https://doi.org/10.1093/jxb/erj074.

Toor, R.K., Savage, G.P., Lister, C.E., 2006. Seasonal variations in the antioxidant composition of greenhouse grown tomatoes. J. Food Compos. Anal. 19, 1–10. https://doi.org/10.1016/j.jfca.2004.11.008.

Tudor-Radu, M., Vîjan, L.E., Tudor-Radu, C.M., Tita, I., Sima, R., Mitrea, R., 2016. Assessment of Ascorbic Acid, Polyphenols, Flavonoids, Anthocyanins and Carotenoids Content in Tomato Fruits. Not. Bot. Horti Agrobot. Cluj-Napoca 44, 477–483. https://doi.org/10.15835/nbha44210332.

Uluisik, S., Chapman, N.H., Smith, R., Poole, M., Adams, G., Gillis, R.B., Besong, T.M.D., Sheldon, J., Stiegelmeyer, S., Perez, L., Samsulrizal, N., Wang, D., Fisk, I.D., Yang, N., Baxter, C., Rickett, D., Fray, R., Blanco-Ulate, B., Powell, A.L.T., Harding, S.E., Craigon, J., Rose, J.K.C., Fich, E.A., Sun, L., Domozych, D.S., Fraser, P. D., Tucker, G.A., Grierson, D., Seymour, G.B., 2016. Genetic improvement of tomato by targeted control of fruit softening. Nat. Biotechnol. 34, 950–952. https://doi.org/ 10.1038/nbt.3602.

Van Breemen, R.B., Pajkovic, N., 2008. Multitargeted therapy of cancer by lycopene. Cancer Lett 269, 339–351. https://doi.org/10.1016/j.canlet.2008.05.016.

Van Genuchten, M.T., 1980. A Closed-form Equation for Predicting the Hydraulic of Unsaturated Soils. Soil Science Society of America Journal 44 (5), 892.

Yang, S., Yu, Q., Wang, B., Yang, T., Li, N., Tang, Y., Aisimutuola, P., Wang, Q., Xu, J., Gao, J., 2016. Identification of QTLs for red fruit firmness using the wild tomato species Solanum pennellii LA716 introgression lines. Plant Breed 135, 728–734. https://doi.org/10.1111/pbr.12423.

Zegbe-Domínguez, J.A., Behboudian, M.H., Lang, A., Clothier, B.E., 2003. Deficit irrigation and partial rootzone drying maintain fruit dry mass and enhance fruit quality in “Petopride” processing tomato (Lycopersicon esculentum, Mill.). Sci. Hortic. (Amsterdam). 98, 505–510. https://doi.org/10.1016/S0304-4238(03) 00036-0.

3. CHAPTER III - IDENTIFICATION AND VALIDATION OF QUANTITATIVE TRAIT LOCI ASSOCIATED WITH FRUIT PUFFINESS IN A PROCESSING TOMATO INBRED LINE POPULATION

Françoise Dalprá Dariva[1,2], Su Subode[1], Jin Cho[1], Carlos Nick[2], David Francis[1]

[1]Department of Horticulture and Crop Science, The Ohio State University, 1680 Madison Ave, Wooster 44691, OH, United States.
[2]Departamento de Agronomia, Programa de Pós-graduação em Fitotecnia, Universidade Federal de Viçosa, Av. P.H. Rolfs, s/n, Campus Universitário, 36570-900, Viçosa, MG 36570-900, Brazil.

Abstract

Physiological disorders impact yield and quality of marketable fruit in tomato. Puffy fruit caused by cavities inside the locule can be problematic for processing and fresh-market quality. In this paper, we used a recombinant inbred line (RIL) and three derived processing tomato populations to map and validate quantitative trait loci (QTLs) for fruit puffiness across environments. Binary interval mapping was used for mapping of fruit puffiness incidence and non-parametric interval mapping and parametric composite interval mapping were used for mapping severity. Marker-trait regressions were carried out to validate putative QTLs in subsequent crosses. Results from different mapping approaches were consistent and pointed to QTLs on Chromosome (Chr) 1, 2, and 4. Only the QTL on Chr 1 was validated in subsequent progeny. QTL on Chr 1 explained, at most, 22.5% of the total variance in percentage of puffy fruit, suggesting the potential involvement of other loci below the detection threshold. A significant interaction between loci on Chr 2 and 4 increased percentage of puffy fruit an additional 15%. The allele responsible for puffy fruits in the populations was donated by parent FG02-188 and was dominant towards increased incidence and severity. Markers solcap_snp_sl_20440 and solcap_snp_sl_18619 are strongly associated with the QTL interval on Chr 1 and are suggested for marker-assisted selection (MAS). MAS for QTL on Chr 1 was as efficient as genomic selection (GS) despite the potential contribution of other loci. Selection for puffiness-resistant lines by either MAS or GS at early breeding stages is recommended for better allocation of phenotyping resources.

Keywords: puffy fruit; marker-assisted selection; genomic selection; tomato breeding

Abbreviations: RIL, recombinant inbred line; QTL, quantitative trait loci; Chr, chromosome; MAS, marker-assisted selection; GS, genomic selection; SNP, Single Nucleotide Polymorphism; IM-BI, interval mapping for a binary model; IM-NP, interval mapping for a non-parametric model; CIM, composite interval mapping; LOD, logarithm of odds; LSI, LOD support interval; PVE, percentage of variance explained; GEBV, genomic estimated breeding value; MAF, minor allele frequency.

3.1. Introduction

Tomato is a major vegetable crop cultivated around the globe with more than 189 million tons of fruit harvested in 2021 (FAOSTAT, 2023). Physiological disorders impact yield and quality of marketable fruit, and are of growing concern due to increased severity of symptoms due to heat, drought, and excess moisture (Masarirambi et al., 2009). Physiological disorders are defined as defects, usually of fruit, and are exacerbated by exposure to abiotic stress conditions including nutrient imbalances, temperature fluctuation, and water extremes. Cat facing, blossom-end rot, yellow shoulder, cracking, and puffiness are examples of physiological disorders of tomato. Genetic predisposition, plant nutrition, and production practices including pruning and training contribute to the occurrence of physiological disorders in tomato fields (Peet, 2009).

Puffiness is defined as the presence of open cavities between outer walls and locular contents in fruit that would normally be filled with gel, seeds, or placental tissue (Gatahi, 2020). Because puffy fruit tend to be soft, ribbed, spongy, and gel-free, they are disliked by consumers of fresh tomato (Gangadhara et al., 2021). Grading standards for fresh-market tomatoes set limits for puffiness (USDA, 2005). Puffy fruit are also considered undesirable for the processing industry (Doganlar et al., 2002) with anectodical reports of factory yield loss due to fruit floating during water aided sorting and exploding during heating in the peeling process.

Puffiness occurs due to either an overgrowth of the pericarp or poor development of the locule tissue (Nawata et al., 1985). Hence, any factor that impairs pollination, fertilization, or seed set promotes fruit puffiness (Olson, 2004). Environmental conditions associated with fruit puffiness include high and low temperatures, excessive rain, low light intensity, and high nitrogen combined with low potassium levels (Nawata et al., 1985; Olson, 2004; Shabtai et al., 2007). Maintaining ideal temperatures for plant growth can minimize the impact of this disorder

(Gangadhara et al., 2021) which can only be achieved in indoor tomato production. Alternatively, breeding for tomato varieties which are less prone to puffiness may provide a solution for field-grown tomato. A genetic component is associated to fruit puffiness in tomatoes (Masarirambi et al., 2009) and defining the genetic predisposition or resistance could provide a means to minimize this disorder in diverse tomato production systems.

Efforts to identify quantitative trait loci (QTLs) in tomato for physiological disorders such as cracking (Capel et al., 2017) and susceptibility to blossom-end rot (Topcu et al., 2021) have been successful. Mapping QTL provides information relative to the location of loci, gene action, and the effect of an allele substitution which may facilitate selection decisions (Asíns, 2002). QTL mapping is also an important step in prospecting or mining candidate genes (Pessoa et al., 2021). Here we describe a multi-generation analysis of the genetics behind fruit puffiness within a processing tomato population. Three different statistical approaches were used to map QTLs for both incidence and severity of fruit puffiness across two environments. QTL interactions increasing fruit puffiness severity were detected. Single Nucleotide Polymorphism (SNP) markers for marker-assisted selection (MAS) to reduce susceptibility were identified. Strategies for selection based on MAS and genomic selection (GS) were compared under different intensities with results showing promise to minimize puffiness in breeding populations.

3.2. Materials and Methods

3.2.1. Mapping populations

The initial population used to quantify the occurrence of fruit puffiness and map quantitative trait loci (QTLs) was derived from a cross between Ohio 2K9-5533-1 and FG02-188. Ohio 2K9-5533-1 is an inbred line with 50% fresh-market Roma genetic background and 50% processing tomato background. Ohio 2K9-5533-1 possesses resistance to tomato spotted wilt virus (*Sw5*) and *Phytophthora infestans* (*Ph3*) (Robbins et al., 2010). FG02-188 is a promising tomato parent for commercial hybrids with resistance to bacterial canker (*Clavibacter michiganensis* subsp. *michiganensis*). The population consisted of 159 F_5 recombinant inbred lines (RILs).

Three populations derived from crosses to two commercial lines (UGP02 and UGP01), and a bacterial spot resistant breeding line (OH813A) and three selections from the RIL population

(2K19-8107-1, 2K19-8052-3, and 2K19-8022) were developed for QTL validation. The parent OH813A was previously described (Bernal and Francis, 2021). The hybrids, 2K20-8312 (2K19-8107-1 X UGP02), 2K20-8322 (2K19-8052-3 X UGP01), and 2K20-8357 (OH813A X 2K19-8022) were self-pollinated and F_2 progeny were advanced to produce 161 F_3 generation families (53 lines from 2K20-8312 and 2K20-8322 each, and 55 lines from 2K20-8357) for evaluation.

3.2.2. Field trials

Tomato seedlings were grown inside a glass-covered greenhouse with environmental conditions set to 24-26°C daytime and 21-24°C nighttime temperatures, 12h photoperiod, and minimum light intensity of 190 μmol m^{-2} s^{-1} in the photosynthetic active range was maintained using supplemental high pressure sodium lighting. Seeds were sown in 288-cell flats filled with commercial growing medium (Promix FLX, Premier Horticulture Inc., Canada). Water was applied daily as needed. Fertilizer applications started when seedlings were 2 weeks old and were performed once a week by adding a 220-ppm solution of Jack's professional 20-20-20 (Scotts Sierra Horticultural Products, Co., Marysville, OH) to the irrigation water. Field transplanting was carried out when seedlings were 4-6 weeks old.

Field trials to assess fruit puffiness were established in Wooster, OH, where 154 tomato RILs were assessed, and Fremont, OH, where 158 RILs were assessed, during the 2019 summer season. Plots were arranged according to an augmented experimental design. Three commercial tomato checks (PS696, UG16112, and H3402) were included in the trials to estimate experimental error. These checks were repeated eighteen times each throughout the tomato field in a way that we could account for spatial variation across rows, columns, and quadrants. Each tomato plot was 6m long and contained 20 plants spaced 30cm apart. Rows were spaced 1.5m apart. Plant care followed standard production practices used in the Midwest United States (Precheur, 2005).

3.2.3. Fruit puffiness measurements

We measured fruit puffiness in nine randomly harvested fruit per plot. Fruit were cut longitudinally, scanned according to protocols outlined previously (Rodríguez et al., 2011, 2010), and the images used to manually score fruit puffiness both as percentage ([number of puffy

fruit/9]*100) and incidence of puffy fruit. For incidence, RILs were scored as 1, indicating presence of puffy fruit, or 0 indicating absence.

3.2.4. DNA isolation and genotyping

DNA was extracted from young leaves using the modified cetyltrimethylammonium bromide method described in Sim et al. (2015). DNA samples were then genotyped using a 384-SNP panel optimized based on recombination, genome coverage, and polymorphism rates (Gill et al., 2019; Sim et al., 2012a, 2012b). The amplicon-based genotyping by sequencing PlexSeq^{TM} platform was used for SNP genotyping (AgriPlex Genomics, Cleveland, Ohio, USA). SNP calls were phased as "A" and "B" in the RIL population, and "A", "B", and "H", in the validation population, in which "A" and "B" designate homozygotes for FG02-188 and Ohio 2K9-5533-1 alleles, respectively. Heterozygotes were designated "H" in the validation population and set to missing data (NA) in the RIL.

3.2.5. Genetic linkage map construction

Data quality control consisted of removing monomorphic markers, markers and individuals with more than 50% missing genotypic data, and markers with segregation distortion detected at $p<0.0001$. We removed 15 markers with failed SNP calls for the reference parent FG02-188, 198 markers that were monomorphic in the RIL population, 3 markers with failed SNP calls for more than 50% of the progeny, and 32 markers with strong segregation distortion ($p<0.0001$). No individuals were removed. The 136 SNP markers left were then assigned to the 12 tomato chromosomes according to their physical position in the Tomato Genome version SL4.0 (Hosmani et al., 2019).

A genetic map was formed using the R/qtl package (Broman et al., 2003) in the R statistical software (R Core Team 2021). The function est.rf()was used to estimate recombination frequencies between markers. Marker order was estimated through the orderMarkers() function with the map.function = "kosambi" argument. The Kosambi mapping procedure converts recombination frequencies into genetic distances (cM) and has the advantage of correcting for multiple crossovers (Kosambi, 1944). Chromosomes were split into two linkage groups if recombination fractions

between markers within a chromosome were greater than 0.5 and the LOD scores for the test of recombination fractions were less than 2. The ripple() (window = 4 and method = "likelihood") and switch.order() functions were used to compare the initial marker order with alternative orders including the marker order in the reference genome (Tomato Genome version SL4.0) based on logarithm of odds (LOD) scores and chromosome length. Higher LOD scores and lower chromosome lengths were observed mostly when markers were arranged in the same order as in the reference genome and were therefore preferred. Quality of the final map was assessed based on genome coverage, by linear regression, and by rank correlation analyses between marker position in the genetic linkage map and the physical position in physical map version SL4.0 (Hosmani et al., 2019) (Table 1).

3.2.6. Phenotypic data analysis

To investigate spatial variation in the field, the effects of genotype, rows, columns, and quadrants on the response variable percentage of puffy fruit, we fit a linear regression model via maximum likelihood estimators using the lmer() function in the lme4 package in R (Bates et al. 2015; R Core Team, 2021). The full experimental model was $y = u + GEN + ROW + COL + QUAD + e$, where, y was the vector of phenotypic observations, in this case, percentage of puffy fruit; u was the model's overall mean; GEN was the effect of genotype; ROW was the effect accounting for variation across rows consisting of a single tier of plots; COL was the effect accounting for variation across columns consisting of three beds; $QUAD$ was the effect accounting for variation across quadrants; and e was the residual error; with all these effects considered random. The variables ROW, COL and $QUAD$ describe a three-dimensional grid with over-replicated checks used to estimate variation. We used the anova(model 1, model 2) function and syntax to compare the full model (model 1) against a second model (model 2) with either GEN, COL, ROW, or $QUAD$ dropped to test the significance of the effect in question. P-values of the chi-square statistics equal to or lower than 0.05, obtained from maximum likelihood ratio test comparisons, indicated whether the dropped effect was significant. Additionally, correlation analyses were used to investigate the relationship between percentage and incidence of puffy fruit across Wooster and Fremont environments.

3.2.7. QTL mapping

QTL mapping for fruit puffiness within the RIL population was performed for percentage and incidence of puffy fruit in the two locations (Wooster and Fremont) separately and combined, using three different statistical approaches. Genotype probabilities were estimated using the calc.genoprob() function in R/qtl with a step size of 1cM. QTLs for incidence of puffy fruit were detected through Interval Mapping using the scanone() function and the model = "binary" argument (IM-BI) in R/qtl (Broman et al., 2003). For percentage of puffy fruit, QTLs were detected through Interval Mapping using the scanone() function and a non-parametric model (model = "np" argument) (IM-NP) since residuals were not normally distributed according to the Shapiro-Wilk test ($p<0.05$). Composite Interval Mapping (CIM) using the cim() function with one marker covariate and Haley-Knott regression (Haley and Knott, 1992) as the solution-generating algorithm was also used to detect QTLs for percentage of puffy fruit. The non-parametric model and CIM were compared to verify CIM sensitivity to the violation of normality assumptions especially because this method has recently been widely implemented as seen in Adhikari and Missaoui (2019), Fenstemaker et al. (2022), and Lin et al. (2018). LOD significance thresholds to declare a QTL were determined by permutation tests ($\alpha = 0.05$, n = 1000; Churchill and Doerge, 1994). We defined QTL location based on 1-LOD support interval (LSI) from the LOD peak (maximum LOD score in a chromosome that is above the LOD cutoff at 0.05 significance level). Percentage of variance explained (PVE) and additive effects of QTLs for percentage of puffy fruit were based on the nearest marker to the QTL peak position and estimated using the makeqtl() and fitqtl() functions of R/qtl (Broman et al., 2003). PVE was given by the formula $1–10^{-2\ LOD/n}$, where n is the sample size and LOD is the LOD score. Additive effects were equal to half the difference between the phenotype averages for the two homozygotes. A univariate binary logistic regression was fit for fruit puffiness incidence data using the glm() function and family=binomial argument in the R Core package. Odds ratios were expressed as e^{β_1}, where e is the natural base and equals to 2.718 and β_1 is the regression coefficient for the homozygote A.

3.2.8. Interaction between QTLs

Interaction between QTLs were investigated using a linear model ANOVA for all nearest marker combinations using the "lm" function in the R core package (R Core Team, 2021). The linear model used was $y = u + M_x + M_y + M_x{:}M_y + e$, where, y is the vector of phenotypic observations, in this case, percentage of puffy fruit; u is the model's overall mean; M_x is the effect of the nearest marker to the QTL peak position on chromosome x (with x being any chromosome with a QTL for percentage of puffy fruit); M_y is the effect of the nearest marker to the QTL peak position on chromosome y (with $y \neq x$, and equal to any chromosome containing a QTL for percentage of puffy fruit) ; $M_x{:}M_y$ is the interaction effect between M_x and M_y; and e is the random error; with all effects considered fixed except the random error. Tukey's Honest Significant Difference test at 0.05 significance was used to compare the means for percentage of puffy fruit between all genotype combinations, in this case, $A_{M_x}A_{M_y}$, $B_{M_x}A_{M_y}$ $A_{M_x}B_{M_y}$, and $B_{M_x}B_{M_y}$, in which A_{M_x} and A_{M_y} are homozygotes for the parent FG02-188 allele at marker M_x and M_y, and B_{M_x} and B_{M_y} are homozygotes for the alternative allele at marker M_x and M_y, respectively. Mean comparison tests were conducted using the HSD.test() function of the agricolae package in R (de Mendiburu, 2021; R Core Team, 2021).

3.2.9. QTL validation and QTL action

Validation was conducted using F_3 families grown in Fremont during the 2022 summer season. Seedling growth, plot size, plant care, and fruit puffiness measurements followed the same procedure described for the RIL population. Measurement of fruit puffiness and incidence of puffiness and genotyping were also as described for the RIL. One-hundred and ninety-eight out of the 384 SNP markers were polymorphic for the validation families. We performed QTL validation by fitting linear regression models between the variable percentage of puffy fruit and each marker potentially-linked to a QTL for fruit puffiness according to interval mapping results (or within the 1-LSI QTL interval) using the lm() and anova() functions of the R core package (R Core Team 2021). We considered cross as a covariate in the regression analysis so that the model was denotated as $y = u + G + C + e$, where y was the percentage of puffy fruit, G and C were the genotype and cross effects, respectively, and e was the random error. Sub-populations were also analyzed separately using the same model with the cross effect dropped. A QTL was considered

validated when the effect of a marker previously linked to a QTL was significant by the linear ANOVA's F-test at 0.05 significance level with a Bonferroni correction applied.

The presence of heterozygotes in the validation population allowed us to investigate QTL action through mean comparison tests between genotype classes. A Tukey's Honest Significant Difference test at 0.05 probability level was conducted to compare percentage of puffy fruit averages between homozygotes A, homozygotes B, and heterozygotes for validated markers using the HSD.test() function of the agricolae package in R (de Mendiburu, 2021; R Core Team, 2021).

3.2.10. Comparison of selection strategies for decreased fruit puffiness

To investigate whether MAS for validated QTLs would result in decreased fruit puffiness, we compared percentage of puffy fruit averages of randomly selected families from the validation population with that of families selected through MAS, GS, and the combination of GS and MAS (GS+MAS), under selection intensities of k=1.66 (10%) and k=0.80 (20%). The MAS strategy consisted of randomly selecting families that were homozygous for the allele B of solcap_snp_sl_20440 and solcap_snp_sl_18619. These are the two markers most likely to be linked to the QTL on chromosome 1 (Table 3). The GS strategy consisted of selecting families with lowest genomic estimated breeding values (GEBVs) for percentage of puffy fruit. For the GS+MAS strategy, instead of randomly selecting families that were homozygous for the allele B of solcap_snp_sl_20440 and solcap_snp_sl_18619 markers, we selected those with lowest GEBVs for percentage of puffy fruit. k=1.66 and k=0.80, in this case, mean that 16 and 32 out of 162 families from the validation population were selected, respectively. Means of selected families were compared by the Tukey's Honest Significant Difference test at α=0.05 significance level (de Mendiburu, 2021; R Core Team, 2022).

The GS model used to predict GEBVs for percentage of fruit in the validation population was developed using genotypic and phenotypic data from the mapping population *Cmm*-RIL. Model development included genotypes for 163 SNPs and percentage of puffy fruit averages for 149 *Cmm*-RIL lines across two locations. SNP calls were phased as '-1,0,1', referring to homozygotes for the allele of Ohio 2K9-5533-1, heterozygotes, and homozygotes for the allele of FG02-188, respectively. Such change is a requirement of the rrBLUP package used to run the GS model. Call rate >80% and MAF <0.05 were the criteria adopted for quality control of SNPs. The

few NAs left were replaced by the most frequent mode. Effects of markers were estimated by the mixed.solve() function of rrBLUP (Endelman, 2011) in R (R Core Team, 2022). mixed.solve() fitted the ridge regression model $y = X\beta + Zu + \varepsilon$ to the data, where y is the corrected phenotypic data for each line, X is the design matrix for the fixed effects β (which was not provided here), Z is the genotype matrix for the random marker effects u with $u \sim N(0, K\sigma_u^2)$, where K is a positive semidefinite matrix (an identity matrix in this case), and ε is the error vector (Endelman, 2011). rrBLUP assumes that all markers have some effect on the trait and share a common variance (Baral et al., 2020). Fruit puffiness GEBVs of validation families were estimated by simply multiplying the vector of marker effects with the genotype matrix. GS model quality and predictive ability was checked by leave-one-out cross-validation and linear correlation between predicted GEBVs and measured percentage of puffy fruit averages for the two locations.

3.3. Results

3.3.1. Fruit puffiness assessment

Symptoms of fruit puffiness are shown in Fig.1. Sixty-six percent of the RILs (101 out of 154) had no puffy fruit in the Wooster field trial while 59% lacked puffy fruit in the Fremont trial (93 out of 158). Sixty-six RILs had no puffy fruit in both locations. There was a significant correlation between locations ($r = 0.43$, $p<0.001$) suggesting that RILs with high percentage of puffy fruit in Wooster tend to have a high percentage of puffy fruit in Fremont and vice-versa. Based on incidence, RILs with puffy fruit in Wooster also tended to have puffy fruits in Fremont ($r = 0.21$, $p = 0.01$).

Figure 1. Longitudinal section of a normal tomato fruit on the left and a puffy tomato fruit on the right.

Genetic causes were associated with variation for fruit puffiness within the RIL population based on pairwise comparison of models ($p \leq 0.0026$). Row, column, and quadrant effects, on the other hand, were not significant (p-values ranging from 0.18 to 1.0), and therefore, no correction of the original data for spatial variation was needed for the QTL analysis.

Fruit puffiness was more severe in the validation populations than in the mapping population. Validation families had an average percentage puffy fruit of 26.7% compared to 7.1% and 10% observed for the RILs in Wooster and Fremont field trials, respectively. Variation in percentage of puffy fruit was also observed among validation sub-populations. Average percentage of puffy fruit for 20K20-8312, 2K20-8322, and 2K20-8357 were 42.8%, 27.9%, and 10.1%, respectively.

3.3.2. Linkage map quality

The quality of the linkage map was assessed based on genome coverage and correlation with the physical map. Table 1 summarizes marker distribution across the 12 tomato chromosomes. Chromosome (Chr) 4 was split into two because the markers solcap_snp_sl_4042, solcap_snp_sl_47843, and solcap_snp_sl_4139 showed recombination fractions greater than 0.5 and test LOD score lower than 2 relative to other Chr 4 SNPs. The number of markers per chromosome ranged from 3, on Chr 11, to 30 on Chr 4 (27 on 4a and 3 on 4b). Lowest and highest chromosome length estimates ranged from 11.1 cM for Chr 11 to 93.3 cM for Chr 6. The average

distance between markers was 4.9 cM and the largest distance between markers was 51.4 cM, observed on Chr 9. Marker physical position in the tomato Sl4.0 physical map (Hosmani et al., 2019)agreed with the estimated genetic position in the linkage map (Table 1). All rank correlation coefficients were equal to 1.00 (Table 1). The R^2 for linear regressions were high for all linkage groups except in 11 and 12 and ranged from 0.48 to 0.99 (Table 1). Poor fit of the data on groups 11 and 12 is likely due to the small number of markers for these linkage groups (3 on Chr 11, and 5 on Chr 12).

Table 1. Genetic linkage map for the Cmm-RIL population.

					Genetic map vs physical map (Sl4.0) correlation		
Chr	**Number of markers**	**Chr length (cM)**	**Average distance between markers (cM)**	**Largest distance between markers (cM)**	**[a]P-value**	**[b]R^2**	**[c]ρ**
1	12	54.2	4.9	10.6	0.0001	0.7581	1.0000
2	14	81.2	6.2	34.3	0.0000	0.7522	1.0000
3	6	41	8.2	28.7	0.0006	0.9509	1.0000
4a	27	37.9	1.5	13.6	0.0000	0.4790	1.0000
4b	3	0.5	0.2	0.5	0.2305	0.7491	1.0000
5	16	35	2.3	9.9	0.0000	0.7728	1.0000
6	9	93.3	11.7	46.8	0.0001	0.8960	1.0000
7	11	47.7	4.8	17.2	0.0002	0.7813	1.0000
8	7	55.1	9.2	33	0.0061	0.7671	1.0000
9	13	62.6	5.2	51.4	0.0000	0.9945	1.0000
10	10	63.6	7.1	31	0.0012	0.7198	1.0000
11	3	11.1	5.6	9.1	0.5260	-0.0803	1.0000
12	5	14.9	3.7	8.2	0.1730	0.3516	1.0000
Overall	136	598	4.9	51.4			

Chr = chromosome. [a,b]P-values and Determination coefficients (R^2) for the linear regression ANOVA and [c]Spearman's rank correlation coefficients between marker position in the genetic linkage map and in the physical map (Tomato Genome version SL4.0 published by Hosmani et al., 2019).

3.3.3. QTL identification

QTLs associated with fruit puffiness were detected in Fremont and in the combined data but not in Wooster. The number of QTLs observed among mapping approaches were consistent, with differences reflecting the environment in which data were collected and the confidence interval rather than the chromosome location, method of scoring the trial, or statistical approach used. For percentage of puffy fruit, the non-parametric model (IM-NP) detected a QTL located between 11 and 38 cM on Chr 1, and a QTL between 7 and 38 cM on Chr 2 in Fremont, and a QTL between 14 and 38 cM on Chr 1, and another QTL between 22 and 37 cM on Chr 2 using data for both locations (Fig. 2). CIM detected QTLs on Chr 1, 2, and 4 between 9 and 30 cM; 28 and 37 cM; and 0 and 5 cM, respectively, in Fremont, and on Chr 2 between 30 and 36 cM in both locations (Fig. 2). As for incidence of puffy fruit, the binary model (IM-BI) detected a QTL on Chr 1 between 11 and 38 cM in Fremont, and on Chr 1 and 2 between 16 and 39 cM, and 22 and 39, respectively, in both locations (Fig. 2). The one-LOD support intervals (LSI) estimated on Chr 1 and 2 using different mapping approaches overlap (Fig. 2) which suggests that different mapping approaches are all detecting the same QTL. The SNP markers solcap_snp_sl_18619, and solcap_snp_sl_4963 on Chr 1, solcap_snp_sl_13625 and solcap_snp_sl_23850 on Chr 2, and solcap_snp_sl_21372 on Chr 4 appear within 1-LSI QTL intervals for all mapping approaches (Fig. 2) and are therefore more likely to be linked to these respective QTLs. In most cases, B was the beneficial allele (Table 2) suggesting that FG02-188 was the parent responsible for increased fruit puffiness in the RIL population. The allele from the parent Ohio 2K9-5533-1 on chromosome 1 decreased percentage of puffy fruit by 3.1 to 4.4%, and on Chr 2 by 4.1 to 5%. On Chr 4 the allele from the parent Ohio 2K9-5533-1 increased percentage of puffy fruit by 4.7%. The QTL on Chr 1 explained between 5% to 6.4% of the total variation in percentage of puffy fruit (Table 2). On Chr 2, the QTL explained from 8.3% to 9.5% of the variation in percentage of puffy fruit (Table 2). The QTL on Chr 4 explained 7.1% of the variation in percentage of puffy fruit in Fremont (Table 2). The odds of having puffy fruit for homozygotes A on Chr 1 are 3.4 to 3.6 times the odds for homozygotes B. The odds of having puffy fruit for homozygotes A on Chr 2 were 3.9 times the odds for homozygotes B.

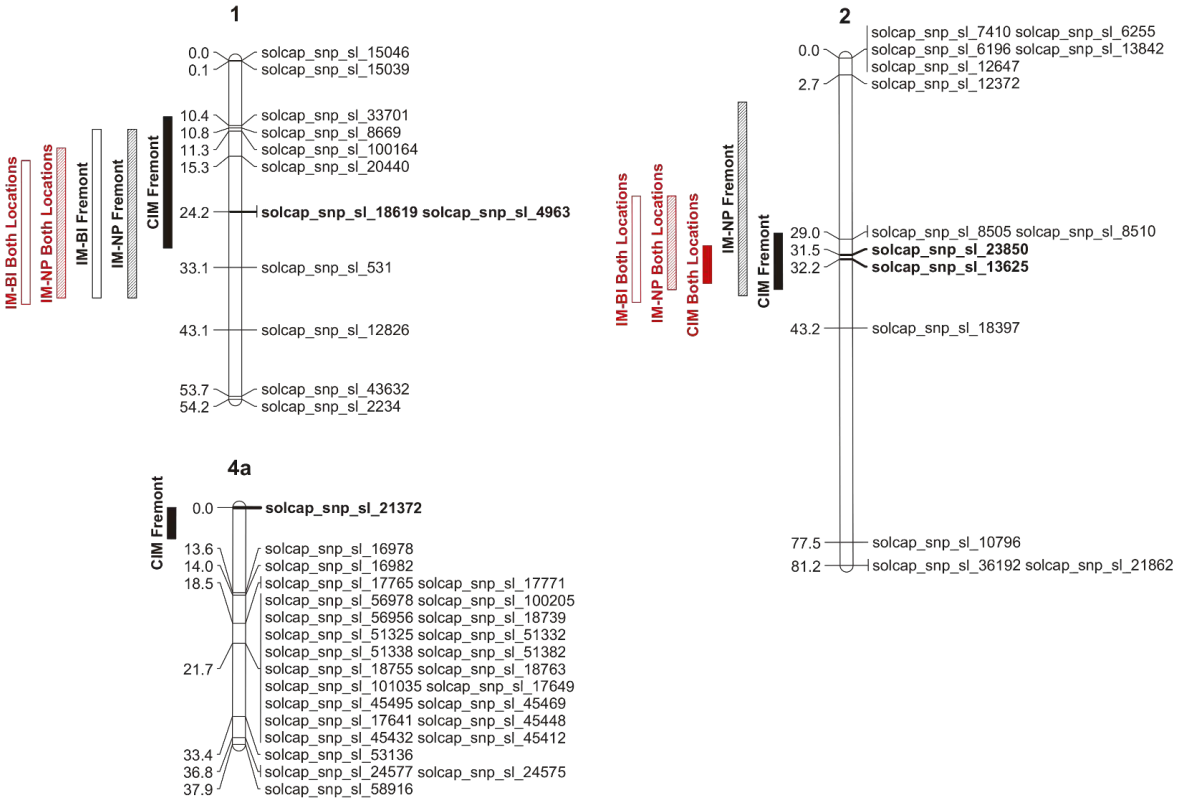


Figure 2. Schematic representations of tomato chromosomes 1, 2 and 4a, based on genetic linkage map results, showing 1-LOD support intervals (LSI) for all detected QTLs using different mapping approaches: IM-BI = Interval Mapping for incidence of puffy fruit using a binary model; IM-NP = Interval Mapping for percentage of puffy fruit using a non-parametric model; CIM = Composite Interval Mapping for percentage of puffy fruit using Haley-Knott regression as the solution-generating algorithm. Markers in bold appear within 1-LSI QTL intervals of all mapping approaches. Chromosome representations were created using the MapChart software (Voorrips, 2002).

Table 2. QTL mapping results for the Cmm-RIL population.

Location	Variable	Mapping method	Chr	QTL Peak Position (cM)	LOD Peak	[a]LOD cutoff	[b]Nearest marker	[c]PVE	[d]Additive effect	[e]Odds ratio	[f]Beneficial allele	[g]Physical Position of Nearest Marker (bp)
Fremont	Percentage of puffy fruit	CIM	1	17	2.69	2.46	solcap_snp_sl_20440	6.40	-4.40	-	B	3637274
			2	32	3.06	2.46	solcap_snp_sl_13625	8.30	-5.00	-	B	40480064
			4	0	2.83	2.46	solcap_snp_sl_21372	7.10	4.70	-	A	2951634
	Percentage of puffy fruit	IM-NP	1	20	2.86	2.40	solcap_snp_sl_18619	5.00	-3.90	-	B	68611189
			2	32	2.63	2.40	solcap_snp_sl_13625	8.30	-5.00	-	B	40480064
	Incidence of puffy fruit	IM-BI	1	21	2.67	2.48	solcap_snp_sl_18619	-	-	3.40	B	68611189
Both Locations	Percentage of puffy fruit	CIM	2	32	3.59	2.40	solcap_snp_sl_13625	9.50	-4.10	-	B	40480064
	Percentage of puffy fruit	IM-NP	1	29	3.50	2.46	solcap_snp_sl_531	5.40	-3.10	-	B	73097261
			2	32	3.18	2.46	solcap_snp_sl_13625	9.50	-4.10	-	B	40480064
	Incidence of puffy fruit	IM-BI	1	28	3.25	2.49	solcap_snp_sl_4963	-	-	3.60	B	70083502
			2	32	2.55	2.49	solcap_snp_sl_13625	-	-	3.90	B	40480064

CIM = Composite Interval Mapping; IM-NP = Interval Mapping using a non-parametric model; IM-BI = Interval mapping using a binary model; Chr = chromosome. aLOD cuttoffs were defined through permutation tests ($\alpha = 0.05$, n = 1000; Churchill and Doerge, 1994). bNearest marker refers to the nearest marker to the QTL Peak Position (cM). cPVE is the percentage of variance explained by a QTL based on the nearest marker to the QTL Peak Position (cM) estimated by the formula 1–10-2 LOD/n, where n is the sample size and LOD is the LOD score. dThe additive effect corresponds to half the difference between the phenotype averages for the two homozygotes. eOdds ratios estimated by e^{β_1}, where e is the natural base and equals to 2.718, and β_1 is the regression coefficient of the logistic regression for the homozygote A. The odds ratio for incidence of puffy fruit in Fremont means that the odds of having puffy fruit for homozygotes A on chromosome 1 are 3.4 to 3.6 times the odds for homozygotes B. fAllele responsible for decreased fruit puffiness in the RIL population in which "A" is the allele donated by the parent FG02-188 and "B" is the allele donated by the parent Ohio 2K9-5533-1. gPhysical position of nearest marker (bp) in the Tomato Genome version SL4.0(Hosmani et al., 2019).

3.3.4. QTL interaction

An interaction between QTLs on Chr 2 and 4 may increase the percentage of puffy fruit. A significant interaction was observed between solcap_snp_sl_13625, the nearest marker to the QTL peak position on Chr 2, and solcap_snp_sl_21372, the nearest marker to the QTL peak position on Chr 4, for the Fremont data and combined data ($p<0.05$). Although the interaction was not significant for the Wooster data ($p>0.10$), genotype mean values were numerically consistent with an interaction. RILs with the marker combination, $A_{M_{13625}}B_{M_{21372}}$, homozygous for the allele A of solcap_snp_sl_13625, and homozygous for the allele B of solcap_snp_sl_21372 had significantly higher percentage of puffy fruit in Fremont and in the data combined from both locations compared to other allele combinations (Fig. 3). Although not statistically different for Wooster data, percentage of puffy fruit in $A_{M_{13625}}B_{M_{21372}}$-RILs (12.3%) was on average higher than that of $A_{M_{13625}}A_{M_{21372}}$ (7.5%), $B_{M_{13625}}A_{M_{21372}}$ (2.8%), and $B_{M_{13625}}B_{M_{21372}}$ (4.4%) (Fig. 3).

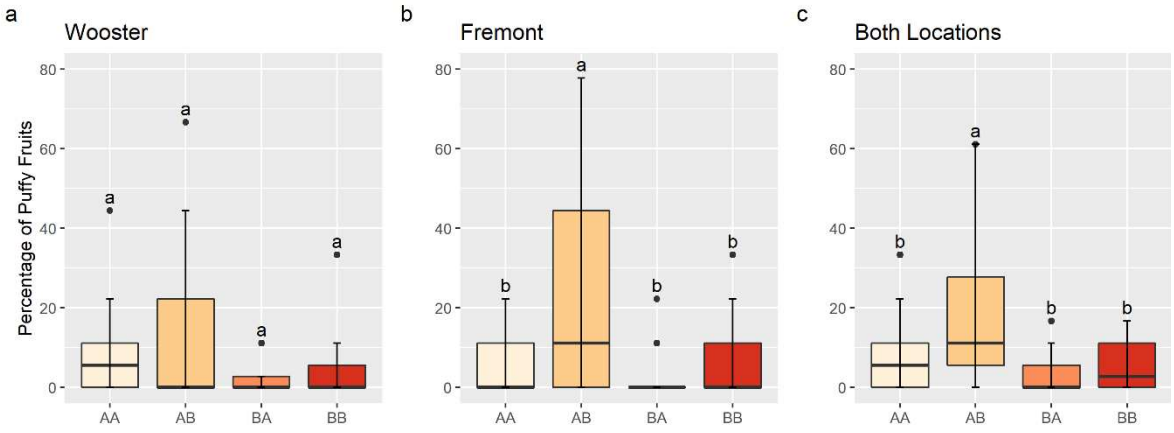


Figure 3. Boxplots for percentage of puffy fruit by chromosome 2 and 4 genotypes for Wooster (a), Fremont (b), and data for Both Locations combined (c). AA indicates a homozygote for the allele A of solcap_snp_sl_13625 (chromosome 2) and solcap_snp_sl_21372 (chromosome 4); AB indicates a homozygote for the allele A of solcap_snp_sl_13625 and a homozygote for the allele B of solcap_snp_sl_21372; BA indicates a homozygote for the allele B of solcap_snp_sl_13625 and a homozygote for the allele A of solcap_snp_sl_21372; and BB indicates a homozygote for the allele B of solcap_snp_sl_13625 and solcap_snp_sl_21372. Different letters indicate statistically significant differences among genotypes by the Tukey's Honest Significant Difference Test ($p<0.05$).

3.3.5. QTL validation

Significant marker-trait associations detected in the validation population confirms only QTL on Chr 1 (Table 3). Linear regressions for percentage of puffy fruit on the markers solcap_snp_sl_33701, solcap_snp_sl_8669, solcap_snp_sl_100164, solcap_snp_sl_20440, solcap_snp_sl_18619, solcap_snp_sl_4963, and solcap_snp_sl_531 located within 1-LSI on Chr 1 (Fig. 2) were highly significant ($p<0.0001$) in the validation population. Linear regressions for percentage of puffy fruit on markers within 1-LSI QTL interval on Chr 2 (p-values ranging from 0.76 to 0.80) and 4 ($p = 0.94$), on the other hand, were not significant.

Because cross effect was significant for all marker-trait regressions, we also analyzed marker-trait relationships for the three sub-populations separately. Potential QTLs on Chr 2 and 4 were still dismissed (Table 3). All significant marker-trait associations in the validation population, except those on the markers solcap_snp_sl_20440 and solcap_snp_sl_18619, were considered not significant in at least one sub-population (Table 3). Since regressions upon solcap_snp_sl_20440 and solcap_snp_sl_18619 had the highest significance values ($p = 3.02e^{-11}$) in the validation population compared to the other markers and there is no evidence to deny association between these markers and the trait due to monomorphism within sub-populations, they could be useful tools for marker-assisted selection. Interestingly, solcap_snp_sl_20440 and solcap_snp_sl_18619 were monomorphic for the beneficial allele B in the population 2K20-8357 with lowest percentage of puffy fruit (10.1%) and monomorphic for the allele A in the population 2K20-8312 with highest percentage of puffy fruit (42.76%). solcap_snp_sl_20440 and solcap_snp_sl_18619 explained 22.5% of the total variation in percentage of puffy fruit observed ($R^2 = 0.225$).

The validation population was not appropriate to confirm the existence of an interaction between Chr 2 and 4 because genotype classes in it were not fairly represented. There were 24 $A_{M_{13625}}A_{M_{21372}}$ and 33 $A_{M_{13625}}B_{M_{21372}}$ lines against only 2 $B_{M_{13625}}A_{M_{21372}}$, and 5 $B_{M_{13625}}B_{M_{21372}}$ lines in this population.

Table 3. P-values of ANOVA's F-tests from marker-trait regressions for the validation populations studied separately and combined. Selected markers are those identified by interval mapping as potentially linked to QTLs for percentage of puffy fruit in the mapping population.

		P-values of ANOVA F tests from marker-trait regressions			
Chr	Marker	Validation population	2K20-8312 sub-population	2K20-8322 sub-population	2K20-8357 sub-population
1	solcap_snp_sl_33701	$5.59e^{-07}$	0.83	monomorphic B	monomorphic B
1	solcap_snp_sl_8669	$1.04e^{-04}$	0.71	0.33	monomorphic B
1	solcap_snp_sl_100164	$1.32e^{-09}$	monomorphic A	0.31	monomorphic B
1	solcap_snp_sl_20440	$3.02e^{-11}$	monomorphic A	monomorphic A	monomorphic B
1	solcap_snp_sl_18619	$3.02e^{-11}$	monomorphic A	monomorphic A	monomorphic B
1	solcap_snp_sl_4963	$4.54e^{-07}$	0.85	0.45	monomorphic B
1	solcap_snp_sl_531	$5.05e^{-09}$	0.75	0.04	monomorphic B
2	solcap_snp_sl_8505	0.80	monomorphic A	0.92	monomorphic A
2	solcap_snp_sl_8510	0.80	monomorphic A	0.92	monomorphic A
2	solcap_snp_sl_23850	0.68	monomorphic A	0.25	monomorphic A
2	solcap_snp_sl_13625	0.76	monomorphic A	0.59	monomorphic A
4	solcap_snp_sl_21372	0.94	0.44	0.94	0.36

Monomorphic A and B = monomorphic for the allele "A" donated by the parent FG02-188 and for the allele "B" donated by the parent Ohio 2K9-5533-1.

3.3.6. QTL action

The QTL on Chr 1 exhibits complete dominance for increased fruit puffiness. Fig. 5 contains mean comparisons for percentage of puffy fruit by genotype for the markers solcap_snp_sl_33701, solcap_snp_sl_8669, solcap_snp_sl_4963, and solcap_snp_sl_531, located on Chr 1. Linear regressions for percentage of puffy fruit on these markers were significant in the field validation (Table 3). Average percentage of puffy fruit between homozygotes A and heterozygotes did not differ but it was significantly lower for homozygotes B (Fig. 5). No conclusions could be drawn based on the markers solcap_snp_sl_20440 and solcap_snp_sl_18619 because they were either homozygous A or homozygous B for all families.

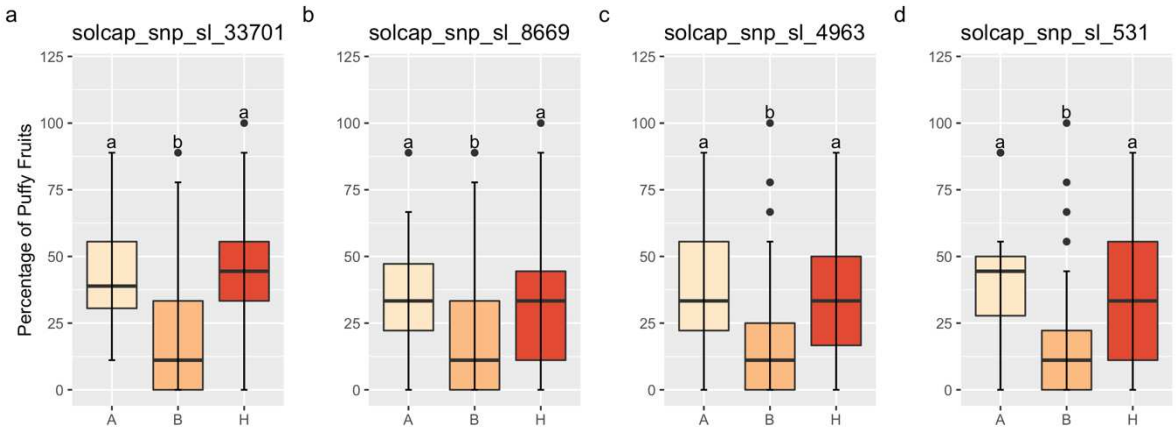


Figure 4. Boxplots for percentage of puffy fruit by genotype for the validation populations. "H" means heterozygotes, and "A" and "B" indicates homozygotes for the FG02-188 and Ohio 2K9-5533-1 alleles, respectively. The markers solcap_snp_sl_33701, solcap_snp_sl_8669, solcap_snp_sl_4963, and solcap_snp_sl_531, located on chromosome 1, are validated. Same lower-case letters mean that genotype classes did not differ by the Tukey's Honest Significant Difference Test ($p<0.05$). Similar means between heterozygotes and homozygotes for the allele A in the field validation data suggest a major dominant QTL for higher fruit puffiness.

3.3.7. Marker-assisted selection for QTL on Chromosome 1

Based on our results, QTLs explained only a small portion of the variation for percentage of puffy fruit in the whole data set (22.5% at most) which suggests that we are dealing with a quantitative trait. In this case, we expect GS to be a better strategy since it considers both large and small-effect genes for selection. In order to investigate how relevant QTL on Chr 1 is and if we can benefit from marker-assisted selection (MAS) on it, we compared percentage of puffy fruit of randomly selected families to percentage of puffy fruit of families selected either by MAS, GS, or the combination of MAS and GS (GS+MAS) under two different selection intensities (Figure 5). Tomato families selected by either MAS, GS, or GS+MAS had lower percentage of puffy fruit compared to randomly selected families regardless the selection intensity tested (Figure 5). MAS for the markers solcap_snp_sl_20440 and solcap_snp_sl_18619 was as good as GS and GS+MAS. Although there is no statistical difference between MAS or GS, using GS instead of MAS prevents selection of outliers with high percentage of puffy fruit as suggested in Figure 5b. The GS model had high predictability results as demonstrated by both cross-validation ($r=0.27$, $p=7.23e^{-4}$) and field-validation ($r=0.52$, $p=2.36e^{-12}$) coefficients.

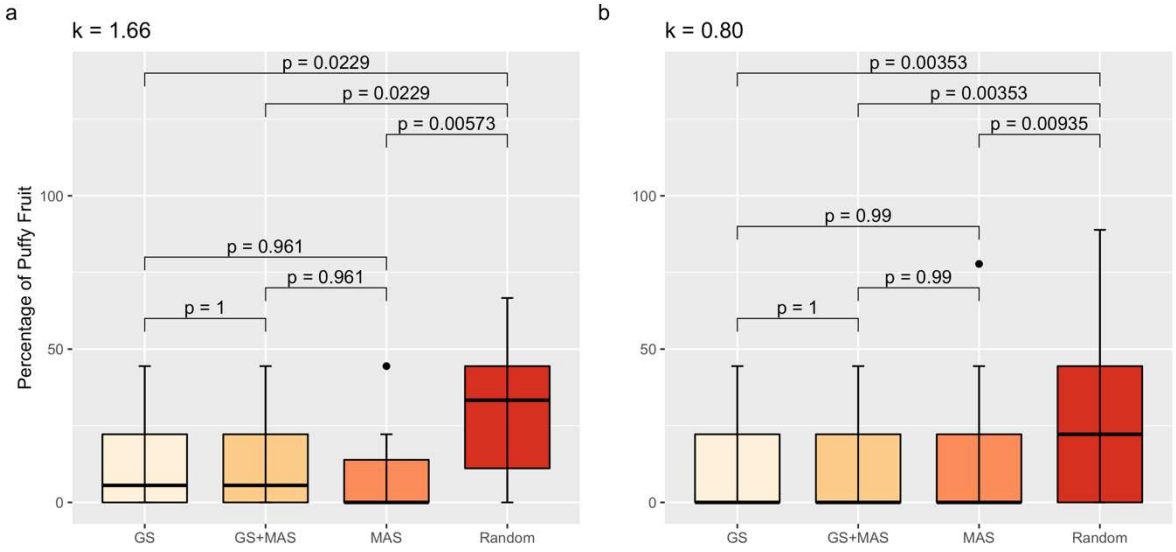


Figure 5. Percentage of puffy fruit averages by selection strategy group for k=1.66 (a) and k=0.80 (b). GS = genomic selection; MAS = marker-assisted selection for the markers solcap_snp_sl_20440 and solcap_snp_sl_18619.

3.4. Discussion

Fruit puffiness has an impact on tomato fruit quality, affecting both processing and fresh market classes. In this study, we presented evidence for putative QTLs for fruit puffiness on tomato chromosomes 1, 2, and 4 and a potential interaction increasing fruit puffiness up to 15% between loci on Chr 2 and 4. Validation efforts in subsequent processing tomato populations confirmed only QTL on Chr 1 which explained 22.5% of the total variation for percentage of puffy fruit in the validation set. QTLs for fruit puffiness were previously reported (Bernacchi et al., 1998; Doganlar et al., 2002; Frary et al., 2004; Fulton et al., 2000, 1997; Tanksley et al., 1996) but none of them were located on Chr 1. Such differences in QTL number and location are common when fruit puffiness alleles come from different tomato germplasm. Fruit puffiness alleles in these previous studies were donated by wild tomato parents while FG02-188, a cultivated elite line, was the source of fruit puffiness alleles here. In addition, previously described QTLs explained a small proportion of the variation also suggesting that fruit puffiness might be a polygenic trait with many loci of small effect and highly influenced by environmental conditions. A comparison between average phenotypes of homozygous and heterozygous lines for makers associated with QTL on Chr 1

suggests a dominance effect for the FG02-188 allele towards increased fruit puffiness. MAS for decreased fruit puffiness on Chr 1 must therefore target only individuals that are homozygous for the allele of the Ohio 2K9-5533-1 parent.

Different data scoring types and data distribution issues required more than one statistical approach for QTL mapping. We therefore mapped QTLs for incidence of puffy fruit through interval mapping using a binary method (IM-BI), and for percentage of puffy fruit using non-parametric interval mapping (IM-NP) and parametric composite interval mapping (CIM). Single QTL analysis (one for each location) and a joint QTL analysis (using average data for the two locations) were also used to determine whether QTLs for fruit puffiness were generally expressed or expressed only in a particular location. Results from different mapping approaches pointed to QTLs, especially on Chr 1 and 2. Because residuals for percentage of puffy fruit in this study were not normally distributed, we opted to map QTLs using both parametric and non-parametric methods. Differences in QTL detection were noted for the parametric CIM method which detected a QTL on Chr 4 and failed to detect the QTL on Chr 1 in Fremont. CIM usually eliminates ghost QTLs (Doerge, 2002) and has more mapping power and precision than standard interval mapping (Zou, 2009). However, CIM should be used with caution when departure from data normality is substantial as in our dataset (Rebaï, 1997). QTLs for fruit puffiness in the RIL population were detected in the joint analyses for both locations and in Fremont suggesting that selection against puffy lines based on phenotype is location-dependent and is more likely to succeed if performed in the Fremont location.

QTL discovery on Chr 1 opens up the possibility for MAS of puffiness-resistant lines but because PVE was low other selection strategies like GS should be considered. GS is often referred to be better suited for complex traits because it uses genome-wide marker coverage to capture the effect of as many small-effect trait-related genes across the genome as possible in an attempt to explain the total genetic variation (Heffner et al., 2009). GS for decreased fruit puffiness in the validation population exhibited promising results with a significant correlation (accuracy) of $r = 0.52$ ($p=2.36e^{-12}$) between predicted GEBVs and measured percentage of puffy fruit in the validation set. However, the percentage of puffy fruit for validation lines selected through either MAS or GS were statistically the same (Figure 5). In addition, combining GS and MAS did not improve selection over using MAS and GS strategies alone. The similarity between selection strategies is due to the fact that all 16 or 32 lines ($k=1.66$ or $k=0.80$) selected through GS or

GS+MAS belonged to the sub-population with lowest percentage of puffy fruit (2K20-8357) on which all individuals were homozygous for the allele B of solcap_snp_sl_20440 and solcap_snp_sl_18619 markers. These are the same lines that would be selected through MAS. In fact, markers associated with QTL on Chr 1 had great impact on GEBV predictions as suggested by their large effects in the GS model.

3.5. Conclusions

Alleles identified in an elite tomato line were found to be the cause of elevated fruit puffiness in advanced tomato breeding populations. Evidence pointed to the discovery of putative QTLs on chromosomes 1, 2, and 4 and a potential interaction between loci on chromosomes 2 and 4 increasing percentage of puffy fruit up to 15%. From these putative QTLs, only that on Chr 1 was reproducible across populations and mapping approaches. QTL on Chr 1 is dominant towards increased fruit puffiness and accounts for 5 to 22.5% of the total variation in percentage of puffy fruit. Approaches using either MAS or GS are effective for selecting against puffy lines compared to when no selection strategy is used at all. Missing heritability issues and effectiveness of MAS suggest that fruit puffiness in our breeding populations is controlled by a major QTL on Chr 1 and several other small-effect QTLs located throughout the genome.

3.6. References

Adhikari, L., Missaoui, A.M., 2019. Quantitative trait loci mapping of leaf rust resistance in tetraploid alfalfa. Physiol Mol. Plant Pathol. 106, 238–245. https://doi.org/10.1016/j.pmpp.2019.02.006

Asíns, M.J., 2002. Present and future of quantitative trait locus analysis in plant breeding. Plant Breeding 121, 281-291. https://doi.org/10.1046/j.1439-0523.2002.730285.x

Baral, K., Coulman, B., Biligetu, B., Fu, Y.B., 2020. Advancing crested wheatgrass [Agropyron cristatum (L.) Gaertn.] breeding through genotyping-by-sequencing and genomic selection. PLoS One. 15, e0239609. https://doi.org/10.1371/journal.pone.0239609

Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw., 67, 1-48. https://doi.org/10.18637/jss.v067.i01.

Bernacchi, D., Beck-Bunn, T., Eshed, Y., Lopez, J., Petiard, V., Uhlig, J., Zamir, D., Tanksley, S., 1998. Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from Lycopersicon hirsutum. Theor. Appl. Genet. 97, 381-397. https://doi.org/10.1007/s001220050908

Bernal, E., & Francis, D. M., 2021. Processing tomato germplasm with improved resistance to bacterial spot. HortScience, 56, 519-520. https://doi.org/10.21273/HORTSCI15616-20

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

Capel, C., Yuste-Lisbona, F.J., López-Casado, G., Angosto, T., Cuartero, J., Lozano, R., Capel, J., 2017. Multi-environment QTL mapping reveals genetic architecture of fruit cracking in a tomato RIL *Solanum lycopersicum* × *S. pimpinellifolium* population. Theor. Appl. Genet. 130, 213–222. https://doi.org/10.1007/s00122-016-2809-9

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

de Mendiburu, F., 2021. agricolae: Statistical Procedures for Agricultural. Research. R package version 1.3-5. https://CRAN.R-project.org/package=agricolae

Doerge, R. W. (2002). Mapping and analysis of quantitative trait loci in experimental populations. Nature Reviews Genetics, 3, 43-52. https://doi.org/10.1038/nrg703

Doganlar, S., Frary, A., Ku, H.M., Tanksley, S.D., 2002. Mapping quantitative trait loci in inbred backcross lines of Lycopersicon pimpinellifolium (LA1589). Genome 45, 1189–1202. https://doi.org/10.1139/g02-091

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

FAOSTAT, 2023. Food and Agriculture Organization of the United Nations. Food and agriculture data. https://www.fao.org/faostat (accessed 26 January 2023).

Fenstemaker, S., Sim, L., Cooperstone, J., Francis, D., 2022. *Solanum galapagense*-derived purple tomato fruit color is conferred by novel alleles of the anthocyanin fruit and atroviolacium loci. Plant Direct 6, e394. https://doi.org/10.1002/pld3.394

Frary, A., Fulton, T.M., Zamir, D., Tanksley, S.D., 2004. Advanced backcross QTL analysis of a *Lycopersicon esculentum* x *L. pennellii* cross and identification of possible orthologs in the Solanaceae. Theor. Appl. Genet. 108, 485–496. https://doi.org/10.1007/s00122-003-1422-x

Fulton, T.M., Beck-Bunn, T., Emmatty, D., Eshed, Y., Lopez, J., Petiard, V., Uhlig, J., Zamir, D., Tanksley, S.D., 1997. QTL analysis of an advanced backcross of Lycopersicon peruvianum to the cultivated tomato and comparisons with QTLs found in other wild species. Theor. Appl. Genet. 95, 881-894. https://doi.org/10.1007/s001220050639

Fulton, T.M., Grandillo, S., Beck-Bunn, T., Fridman, E., Frampton, A., Lopez, J., Petiard, V., Uhlig, J., Zamir, D., Tanksley, S.D., 2000. Theor. Appl. Genet. 100, 1025–1042. https://doi.org/10.1007/s001220051384

Gangadhara, K., Selvakumar, R., Singh, P.K., Manjunathagowda, D.C., Kumar, J., 2021. Tomato: Physiological Disorders and Their Management. Biotica Research Today 3, 944-948. https://biospub.com/index.php/biorestoday/article/view/1181

Gatahi, D.M., 2020. Challenges and Opportunities in Tomato Production Chain and Sustainable Standards Introduction. Int. J. Hortic. Sci. Technol. 7, 235–262. https://doi.org/10.22059/ijhst.2020.300818.361

Gill, U., Scott, J.W., Shekasteband, R., Ogundiwin, E., Schuit, C., Francis, D.M., Sim, S.C., Smith, H., Hutton, S.F., 2019. Ty-6, a major begomovirus resistance gene on chromosome 10, is effective against Tomato yellow leaf curl virus and Tomato mottle virus. Theor. Appl. Genet. 132, 1543–1554. https://doi.org/10.1007/s00122-019-03298-0

Haley, C.S., Knott, S.A., 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69, 315-324. https://doi.org/10.1038/hdy.1992.131

Heffner, E.L., Sorrells, M.E., Jannink, J.L., 2009. Genomic selection for crop improvement. Crop Sci. 49, 1-12. https://doi.org/10.2135/cropsci2008.08.0512

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

Kosambi, D.D., 1944. The estimation of the map distance from recombination values. Annals of Eugenics. 12, 172-175. https://doi.org/10.1007/978-81-322-3676-4_16

Lin, F., Wani, S.H., Collins, P.J., Wen, Z., Gu, C., Chilvers, M.I., Wang, D., 2018. Mapping quantitative trait loci for tolerance to *Pythium irregulare* in soybean (*Glycine max* L.). G3: Genes, Genom. Genet. 8, 3155–3161. https://doi.org/10.1534/g3.118.200368

Masarirambi, M. T., Mhazo, N. O. R. M. A. N., Oseni, T. O., & Shongwe, V. D., 2009. Common physiological disorders of tomato (*Lycopersicon esculentum*) fruit found in Swaziland. J. Agric. Soc. Sci, *5*, 123-127.

Nawata, E., Inden, H., Asahira, T., 1985. Effects of ccc on the occurrence of tomato puffy fruits and the endogenous cytokinin activities. Sci. Hortic. 26, 119-127. https://doi.org/10.1016/0304-4238(85)90004-4

Olson, S. M., 2004. Physiological, Nutritional, and Other Disorders of Tomato Fruit: HS-954/HS200, 2/2004. EDIS, 2004(3).

Peet, M.M., 2009. Physiological disorders in tomato fruit development. Acta Hortic. 821, 151-160. https://doi.org/10.17660/ActaHortic.2009.821.16

Pessoa, H.P., da Silva, L.J., Chagas, R.R., Ferreira Copati, M.G., Dariva, F.D., Dias, F. de O., Gomes, C.N., 2021. Uncovering tomato candidate genes for seed germination performance and seedling vigour using the *Solanum pennellii* introgression line population. Sci. Hortic. 284, 110134. https://doi.org/10.1016/j.scienta.2021.110134

Precheur, R. J., 2005. Ohio Vegetable Production Guide. The Ohio State University. Ext. Bull. 672.

R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rebaï, A., 1997. Comparison of methods for regression interval mapping in QTL analysis with non-normal traits. Genet. Res. 69, 69–74. https://doi.org/10.1017/S0016672396002558

Robbins, M. D., Masud, M. A., Panthee, D. R., Gardner, R. G., Francis, D. M., & Stevens, M. R., 2010. Marker-assisted selection for coupling phase resistance to tomato spotted wilt virus and *Phytophthora infestans* (late blight) in tomato. HortScience, 45, 1424-1428. https://doi.org/10.21273/HORTSCI.45.10.1424

Rodríguez, G.R., Moyseenko, J.B., Robbins, M.D., Morejón, N.H., Francis, D.M., van der Knaap, E., 2010. Tomato analyzer: A useful software application to collect accurate and detailed morphological and colorimetric data from two-dimensional objects. J. Vis. Exp. 37, e1856. https://doi.org/10.3791/1856

Rodríguez, G.R., Muños, S., Anderson, C., Sim, S.C., Michel, A., Causse, M., McSpadden 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 Physiol. 156, 275–285. https://doi.org/10.1104/pp.110.167577

Shabtai, S., Salts, Y., Kaluzky, G., Barg, R., 2007. Improved yielding and reduced puffiness under extreme temperatures induced by fruit-specific expression of rolB in processing tomatoes. Theor. Appl. Genet. 114, 1203–1209. https://doi.org/10.1007/s00122-007-0511-7

Sim, S.C., Durstewitz, G., Plieske, J., Wieseke, R., Ganal, M.W., van Deynze, A., Hamilton, J.P., Buell, C.R., Causse, M., Wijeratne, S., Francis, D.M., 2012a. Development of a large snp genotyping array and generation of high-density genetic maps in tomato. PLoS One 7, e40563. https://doi.org/10.1371/journal.pone.0040563

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, 1437–1445. https://doi.org/10.1094/PHYTO-02-15-0051-R

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., Mutschler, 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, e45520. https://doi.org/10.1371/journal.pone.0045520

Tanksley, S.D., Grandillo, S., Fulton Zamir, T.D., Eshed, Y., Petiard -J Lopez Beck-Bunn, V.T., Zamir, D., Eshed, T., Lopez Hispareco, J.S., 1996. Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*, Theor. Appl. Genet. 92, 213-224. https://doi.org/10.1007/BF00223378

Topcu, Y., Sapkota, M., Illa-Berenguer, E., Nambeesan, S.U., van der Knaap, E., 2021. Identification of blossom-end rot loci using joint QTL-seq and linkage-based QTL mapping in tomato. Theor. Appl. Genet. 134, 2931–2945. https://doi.org/10.1007/s00122-021-03869-0

USDA, 2005. Tomatoes. Shipping Point and Market Inspection Instructions. United States Department of Agriculture: Washington, DC, USA.

Voorrips, R. E., 2002. MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs, *J. Hered.*, 93, 77-78. https://doi.org/10.1093/jhered/93.1.77

Zou, F., 2009. QTL Mapping in Intercross and Backcross Populations, in: DiPetrillo, K. (Eds.) Cardiovascular Genomics. Methods in Molecular Biology, 573. Humana Press, Totowa, NJ, pp.157-173. https://doi.org/10.1007/978-1-60761-247-6_9

4. CHAPTER IV - LARGE-SCALE GENOMIC PREDICTION OF TOMATO HYBRIDS

Françoise Dalprá Dariva[1,2], Su Subode[1], Caleb Orchard[1], Jin Cho[1], Carlos Nick[2], David Francis[1]

[1]Department of Horticulture and Crop Science, The Ohio State University, 1680 Madison Ave, Wooster 44691, OH, United States.
[2]Departamento de Agronomia, Programa de Pós-graduação em Fitotecnia, Universidade Federal de Viçosa, Av. P.H. Rolfs, s/n, Campus Universitário, 36570-900, Viçosa, MG 36570-900, Brazil.

Abstract

Heterosis for yield in tomatoes is unpredictable and heterotic groups have not been defined. As a result, tomato hybrids are created by combining elite lines in a trial-and-error fashion. This lack of criteria guiding selection of hybrid combinations results in tomato breeding programs investing in combinations based on the phenotypic performance of parents. As a consequence, thousands of potential candidate hybrids are left behind. This study demonstrates that large-scale prediction of hybrid performance can be accomplished through genomic selection (GS) models. We predicted the performance of over 22,000 tomato hybrids using seven GS models developed from inbred and partially inbred training populations. Prediction abilities (r) for field validation were as high as 0.42 ($p=2.87e^{-03}$) for yield and 0.58 ($p=1.12e^{-05}$) for fruit weight with variation depending on the GS model and the location of validation trials. The larger the training population size the higher the predictability, though we obtained significant yield predictions for models with a training population size of as few as 69 individuals and a modest marker set of 248 SNPs. These findings are encouraging for the use of GS for hybrid prediction in the context of vegetable breeding programs which often operate with small effective population sizes, few markers, and phenotyping budget constraints.

Keywords: genomic selection models, hybrid breeding, processing tomato

Abbreviations: GS, genomic selection; GEBV, genomic estimated breeding value; BLUPs, best linear unbiased predictors; MLE, maximum likelihood estimators; LRT, likelihood ratio tests; BLUEs, best linear unbiased estimates; PCA, principal component analysis; GxE, genotype-by-

environment; SNP, Single Nucleotide Polymorphism; SL, *Solanum lycopersicum*; SP, *Solanum pimpinellifolium*;

4.1. Introduction

Tomato is a high-value crop cultivated in most countries. Contemporary tomato varieties are predominantly F1 hybrids (Lammerts Van Bueren et al., 2011). The primary reason for commercializing tomato hybrids is protection of intellectual property in the form of parents and guaranteed seed sales. Heterosis, also referred to as hybrid vigor, describes the superior performance of a hybrid compared to the parents and is especially relevant in cross-pollinated species that suffer from inbreeding depression (Liu et al., 2020). Although several publications report evidence for heterosis for several traits including early maturity, fruit weight, total soluble solids, lycopene content, and even yield (Avdikos et al., 2021; Kumar et al., 2019; Liu et al., 2021; Soresa et al., 2020), the evidence heterosis for commercial traits is unpredictable (Soresa et al., 2020), and of low magnitude or nonexistent under stress-free conditions (Duvick, 1999a). In contrast, for hybrid maize heterosis alone accounts for a 15% increase in yield compared to the best open-pollinated varieties (Duvick, 1999b).

In addition to germplasm protection, hybrid tomatoes have advantages for pyramiding resistance genes. Breeders can combine complementary lines to achieve a hybrid with a comprehensive set of resistances or other traits with dominant gene action. Advantages of combining loci with additive action includes the long shelf-life genes “rin” and “nor” which are only desirable in the heterozygous form because fruits from homozygous plants for the nor allele (nor/nor) or rin allele (rin/rin) do not fully ripen while heterozygotes possess desirable long shelf-life properties (Garg and Cheema, 2008; Ito et al., 2021). From a financial standpoint, tomato hybrids are affordable to both seed producers and buyers due to the high value of the crop, low seed requirement per unit area, and high seed set per single pollination (Duvick, 1999a; Soresa et al., 2020).

A major challenge for breeding crops such as tomato, where hybrids dominate the commercial sector, is designing new crosses. In corn parents are chosen from distinct heterotic groups and tester lines can be used to make group assignments. The absence of defined heterotic groups in tomato causes hybrid design to rely on a largely random process of crossing superior

inbreds. Although diallel and testcross methodologies may shed some light on the potential of specific tomato parents for crossing and promising hybrid combinations (Aisyah et al., 2016; Figueiredo et al., 2015; Liu et al., 2021; Vekariya et al., 2019), such methods are limited by phenotyping capacity. In the absence of testers for heterosis, genomic selection (GS) models may offer an approach to predict hybrid performance. An advantage of GS is the potential to predict as many hybrids as desired without the need for either large-scale phenotyping or genotyping. GS models are based on a training population for which both phenotypic and genotypic information are available to build a predictive model. Genomic Estimated Breeding Values (GEBVs) are then estimated for individuals for which only genotypic data are available (Meuwissen et al., 2001). Genotypes of hybrids may be imputed based on data for inbreds or partially inbred parents. Selection or choice of which hybrids to create could then be based on GEBVs rather than phenotypic data. The predictive potential of GS is documented based on both cross-validation and field validation results (Duangjit et al., 2016; Duhnen et al., 2017; Lozada et al., 2019; Zhao et al., 2012). GS has also the advantage of being cheaper and faster than phenotypic selection at early-field testing stages (Borrenpohl et al., 2020).

The objective of this study was to assess accuracy of genomic selection (GS) models in predicting plant yield and fruit weight of tomato hybrids. We used genomic estimated breeding values (GEBVs) to predict performance of 22,681 possible crosses. We then applied selection criteria to identify promising hybrids for field validation trials. Assessment of GS was based on cross-validation and field validation coefficients, and on comparing hybrids selected for performance to randomly selected hybrids.

4.2. Materials and Methods

4.2.1. Plant Material

Three populations were used to develop and train genomic selection (GS) models. The SolCAP population consisted of 140 processing tomato inbred lines considered base breeding germplasm for the North American market. Phenotypic analysis and genotyping of this population were previously described (Merk et al., 2012; Sim et al., 2012a). From the 140 lines used in this study, twenty-four are derived from tomato germplasm adapted to Oregon or California's arid

environmental conditions. Ninety-two lines originated from breeding programs located in the midwestern United States and Ontario, Canada (Great Lakes region) and are considered humid adapted. For the remaining twenty-four lines, no information regarding pedigree or environmental adaptation was available (Merk et al., 2012).

The Second population consisted of a biparental recombinant inbred line (RIL) population using Ohio 2K9-5533-1 as the female parent and FG02-188 as the male parent. The inbred line Ohio 2K9-5533-1 carries tomato spotted wilt virus (Sw5) and Phytophthora infestans (Ph3) resistances in coupling phase (Robbins et al. 2010). The line 2K9-5533-1 retains 50% fresh-market Roma genetic background and is determinate. The male parent, FG02-188, has bacterial canker resistance believed to be derived from Bulgaria 12. FG02-188 has shown commercial potential as a parent for processing tomato hybrids. This RIL population consisted of 159 inbred lines.

Finally, a third population, which we named “next generation” (NG), consisted of 69 F4 families derived from crosses to lines from the RIL. The F4 families were derived from three hybrids 2K20-8312, 2K20-8322, 2K20-8357. The hybrid 2K20-8312 was derived from a cross between UGP02, a California-adapted line used as a parent for commercial hybrids, and 2K19-8107-1, a selection from the RIL population. 2K20-8322 was derived from the cross between UGP01, another California-adapted commercial parent, and 2K19-8052-3, another selection from the RIL population. 2K20-8357 was derived from the OH813A x 2K19-8022 cross, where OH813A is an advanced breeding line with bacterial spot resistance (Bernal et al., 2020; Bernal and Francis, 2021), and 2K19-8022 is a RIL selection. Hybrids were self-pollinated to produce F2 seed. Six hundred and fifty-eight F2 progenies were sown in the greenhouse and genotyped. One hundred and sixty-two selections were advanced to the F3. Of these selections, 69 were intended for yield evaluation. These selections were based on predictive GS models from a previous study (data not published), with 27 high GEBV, 27 low GEBV, and 15 random selections based on yield. The F3 families were then evaluated in the field in a two-location trial, and single-plant selections were advanced after family-based selection for performance. These sixty-nine families constituted the training set. The remaining 93 out of the 162 were advanced by single seed descent as potential parents. Genotypic data for the entire set of 162 were used for hybrid imputation, but only 69 were used for training.

New hybrids were created based on genomic selection models (see section 2.5 below). Parents of selected tomato hybrids were grown in a greenhouse during autumn/winter 2021 to

perform the crosses. F4 plants of NG families were used as females whereas SolCAP lines were used as males. Hybrid seeds were extracted from mature red fruits and treated with hydrochloric acid (0.25 M for 20 min) and trisodium phosphate (10% for 20 min) solutions. Seeds were washed with tap water after each treatment and dried at room temperature prior to heat treatment at 70°C for 72 hours. Seed were stored for future use at 4°C and 20% relative humidity. Seeding and transplanting for field evaluation were as described previously (Merk et al., 2012).

4.2.2. Genotyping

The RIL and NG populations were genotyped using 384 Single-nucleotide polymorphisms (SNPs) as markers. The set of 384 SNPs was optimized based on genome distribution from 7,720 SNPs (Sim et al. 2012a) and polymorphic information content (Sim et al. 2012b). The amplicon-based genotyping-by sequencing PlexSeqTM platform was used for SNP genotyping (AgriPlex Genomics, Cleveland, Ohio, USA). Raw genotyping data was returned as SNP calls which were then phased as “1”, “-1”, “0” with those homozygous for the OH8245 allele designated allele “1”, the alternate allele as “-1”, and heterozygotes as “0”.

Quality control for markers consisted of eliminating monomorphic markers and those with more than 20% missing values. Remaining missing genotypes were imputed as the common allele. Following these data quality control steps, 248 markers were retained for training GS models and for estimation of Genomic Estimated Breeding Values (GEBVs) in prediction populations.

4.2.3. Genomic Selection Model Development

In total, we created seven models for both yield and fruit weight based on individual populations and combinations: SolCAP, RIL, NG-sel, RIL+SolCAP, RIL+NG-sel, SolCAP+NG-sel, RIL+SolCAP+NG-sel using the leove-one-out cross-validation approach described in Liabeuf et al. (2018). The workflow for the development and validation of GS models is summarized in Figure 1.

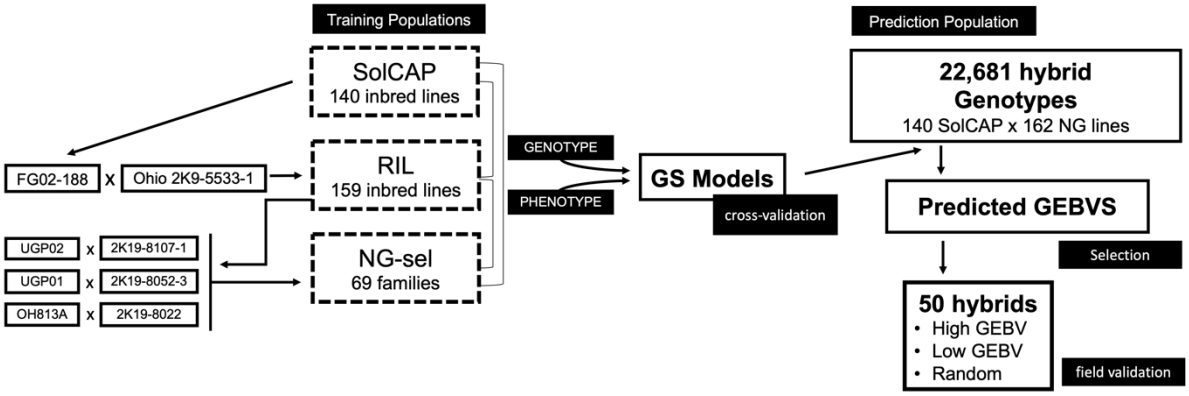


Figure 13. Flowchart summarizing GS model development and validation.

Yield and fruit weight phenotypic data used to train the GS models were obtained from field trials organized according to a randomized complete block design for the SolCAP population, as described previously (Merk et al., 2012). Evaluation of the RIL population was based on an augmented design assessed over two years at two locations per year. For this trial, within-field variation was adjusted based on row, column, and quadrant which contained over-replicated checks. The augmented design model was $y = u + GEN + ROW + COL + QUAD + e$, where y was the vector of phenotypic observations, u was the model's overall mean, GEN was the effect of genotype, ROW was the effect accounting for variation across rows consisting of a single tier of plots; COL was the effect accounting for variation across columns consisting of three beds; $QUAD$ was the effect accounting for variation across quadrants; and e was the residual error, with all these effects considered random. The lmer() function of the lme4 package in R (Bates et al. 2015; R Core Team, 2022) was used to fit those linear regression models to the phenotypic data through the maximum likelihood approach. Best unbiased linear predictors (BLUPs) for genotypes were extracted using the coef() function in the lme4 package. Adjusted data were then averaged across years to create a single value for each line. The F3 families of NG-sel were evaluated in randomized complete block designs with two replicates, in two locations (Wooster, and Fremont, OH).

Marker effects for each GS model were estimated using ridge regression with the mixed.solve() function in the rrBLUP package (Endelman, J. B. 2011). The mixed.solve() function uses maximum likelihood (ML) or restricted maximum likelihood (REML) solutions to fit the ridge regression model $y = X\beta + Zu + \varepsilon$, where y is the corrected phenotypic data for each line, X is the design matrix for the fixed effects β (which was not provided here), Z is the genotype matrix

for the random marker effects u with $u \sim N(0, K\sigma_u^2)$, where K is a positive semidefinite matrix (an identity matrix in this case), and ε is the error vector. Because we did not provide X values, mixed.solve() added an intercept term to the equation by default (Endelman, J. B. 2011). GEBVs of hybrids were predicted by multiplying the vector of marker effects with the hybrid genotype matrix. Because these GEBVs are centered to 0 we added back the constant obtained from the beta output of the mixed.solve() function so that predicted GEBVs for yield and fruit weight are expressed in tons per acre (t/a) and grams (g), respectively. The constant was the average of all betas obtained from each leave-one-out run of the model.

4.2.4. Hybrid Imputation

We imputed genotypic data for all possible hybrid combinations between the 140 tomato lines in the SolCAP population, the 162 tomato families in the NG population, and an additional inbred, line 2K17-2130, by combining genomic data for 248 SNPs used in the training set. A total of 22,681 hybrids were imputed. A hybrid between parents that were homozygous for the same allele were scored as that allele (either 1 or -1); when the two parents differed, the hybrid was designate “0" for heterozygous; in cases where SNPs were heterozygous in one or both parents, the hybrid was imputed as heterozygous (0). Missing data accounted for 0.0046% of all SNPs (345 out of 74,896 SNPs) and were imputed based on the common allele.

4.2.5. Hybrid selection for field validation

GS model accuracy for yield and fruit weight of hybrids was assessed by comparing predicted GEBVs to measured field phenotypes. In total, 50 tomato hybrids were selected and created for field evaluation. Selection criteria were based on GEBVs for yield and fruit weight. Sixteen hybrids with high-yield GEBV scores (high GEBV), 13 with low-yield GEBV scores (low GEBV), and 21 randomly picked hybrids (random GEBV) were based on the model with the highest cross-validation result (in this case, RIL+SolCAP model) were selected (see Table 1). During the process, we also imposed selection such that a single parent did not participate in more than four crosses. GEBV scores for high-yield selections were at least one standard deviation greater than the mean while GEBV scores for low-yield selections were at least one standard

deviation lower than the mean. We also imposed step-selection for fruit weight such that high-yield GEBV selections must have fruit size of 65 g (± 5 g). This fruit size restriction was imposed to meet the desired range for whole-peel tomato processors.

4.2.6. Field evaluation

Seedling production for field plantings was initiated in April 2022 under a controlled environment (24-26°C daytime and 21-24°C nighttime temperatures, 12h photoperiod, and minimum light intensity of 190 μmol m^{-2} s^{-1} in the photosynthetic active range) as described previously (Merk et al., 2012). Seeds of the 50 selected hybrids and two commercial hybrid checks, H3406 and H5108, were sown in 288-cell flats filled with commercial growing medium (Promix FLX, Premier Horticulture Inc., Canada). Water was provided to seedlings daily according to their needs. Weekly fertilizations started when seedlings were 2 weeks old and consisted of adding a 220-ppm solution of Jack's professional 20-20-20 (Scotts Sierra Horticultural Products, Co., Marysville, OH) to the irrigation water. Four-to-six-week-old seedlings were transplanted to the field using a three-row planter for Fremont trials and a two-row transplanter for Wooster trials.

Hybrids were evaluated in a 2-replicate randomized complete block design in two locations (Wooster and Fremont, OH, USA) during the 2022 Summer season. Tomato plots were 3m long each and contained 10 plants. In-row and between-row spacing were 0.3m x 1.5m, respectively. Overall plant care followed standard tomato production procedures recommended for the Midwest United States (Precheur, 2005). Yield-related traits were total yield (adjusted to t/acre) and fruit weight (grams). Total yield was measured based on five middle plants per plot using a destructive harvest. Fruit weight (g) was estimated by averaging 25 size-representative fruits per plot that were randomly picked.

4.2.7. Statistical analysis

4.2.7.1. Phenotypic data analysis

Field trials in Wooster and Fremont were analyzed combined using the statistical model $y = u + GEN + LOC + REP\%IN\%LOC + GEN:LOC + e$, where y was the vector of phenotypic

observations, u was the model's overall mean, GEN was the effect of genotype, LOC was the effect of location, $REP\%IN\%LOC$ was the effect of blocks nested within locations, $GEN:LOC$ was the effect of the genotype x location interaction, and e was the residual error.

To investigate the effects of genotype, location, and genotype x location interactions on the variables yield and fruit weight, we fitted linear mixed models through maximum likelihood estimators (MLE) and compared the goodness of fit between the complete model (model containing all the effects described above) and the reduced model (model without the effect under test) through likelihood ratio tests (LRT) using the lmer() function for random and mixed models and anova(complete_model, reduced_model) functions and syntax to compare models (Bates et al. 2015; R Core Team, 2021). All effects were considered random with the exception of genotype which we considered fixed. The fixef() function from the lme4 package was used to extract best linear unbiased estimates (BLUEs) for genotypes across locations.

Performance of hybrids was also assessed on each location independently using the statistical model $y = u + GEN + REP + e$, where y was the vector of phenotypic observations, u was the model's overall mean, GEN was the effect of genotype considered fixed, REP was the effect of blocks considered random, and e was the residual error. The ML approach was applied to single-location data and the fixef() function was used to extract genotype BLUEs for each location separately as detailed before. Genotype BLUEs were used to study linear associations between a single trait across the two locations using the cor.test() function in R (R Core Team, 2021).

4.2.7.2. Cross-validation of GS models

Cross-validation of models was performed using the leave-one-out method. In this method, we remove one individual from the training set and use phenotypic and genotypic information of the remaining individuals to run a GS model that will be used to predict the GEBV of the individual left out. This procedure is then repeated n times, where n is the number of individuals in the training set. Cross-validation coefficients (r) are linear correlations between observed phenotypes of individuals from the training set and their predicted GEBVs estimated one at a time on each run of the model.

4.2.7.3. Field validation of GS models

Field validation was performed across all GS models based on linear correlations (r) between predicted GEBVs and observed hybrid data from the field trials. The similarity between GS models was assessed by linear correlation analysis for predicted GEBVs using the cor.test() function in R (R Core Team, 2021).

4.2.7.4. Principal Component Analysis

To investigate genetic similarity of training populations, we applied principal component analysis (PCA) to genomic data of individuals from SolCAP, RIL, NG-sel, and the hybrid populations. The population x marker (N x K) matrix was converted to an NxN covariance matrix as detailed in Price (2006). The NxN matrix was created by multiplying the genotype (N) by SNP (K) matrix g_{nk}, with n genotypes set to rows and k SNPs set to columns, to its normalized transposed matrix as follows $g_{nk} \times {g_{nk}}^{-1}$. Matrix normalization was performed using the normalize() function of the 'som' (Yan, 2016) package. PCA was performed using the PCA() function of the 'factoextra' package in R (Kassambara and Mundt, 2008) in R.

4.2.7.5. Mean comparison of GEBV groups

Tukey's Honest Significant Difference test (p = 0.05) was used to compare average phenotypes of selection groups using the HSD.test() function embedded in the agricolae() package (de Mendiburu, F., 2021).

4.3. Results

We predicted yield and fruit weight performance for 22,681 hybrids, a number which would be unreasonable to create and evaluate in any breeding program, using GS models based on inbred and partially inbred-line training populations. Figure 2 and 3 shows yield and fruit weight distributions of hybrids predicted in this study using different training populations. Colored points represent the distribution of hybrids selected and created for subsequent field validation of models. The RIL+SolCAP model was used to guide the selection. Selections were made across the range

of predicted phenotypes with 16 selected as high yield, 13 as low yield, and 21 as random selections. To verify whether the other models would also be able to separate high-yield from low-yield classes, hybrids were reassigned to groups according to the same selection criteria adopted for RIL+SolCAP but using other models. High-yield hybrids were those with predicted yield at least one standard deviation greater than the training population mean while low-yield hybrids were those with predicted yield at least one standard deviation lower than the mean. Everything else was considered random. Overall predicted mean for yield using the RIL+SolCAP model was 36.38 t/a with a standard deviation of 1.78. Predicted yield for high-yield selections ranged from 38.37 to 41.10 t/a whereas predicted yield for low-yield selections ranged from 30.83 to 33.88 t/a. Predicted fruit weight of high-yield hybrids ranged from 62.66 to 67.04g. Predicted means and high/low thresholds for each model and trait as well as the number of hybrids per selection group after reassignment are detailed in Supplementary Tables 1 and 2 and can also be visualized in Figures 2 and 3.

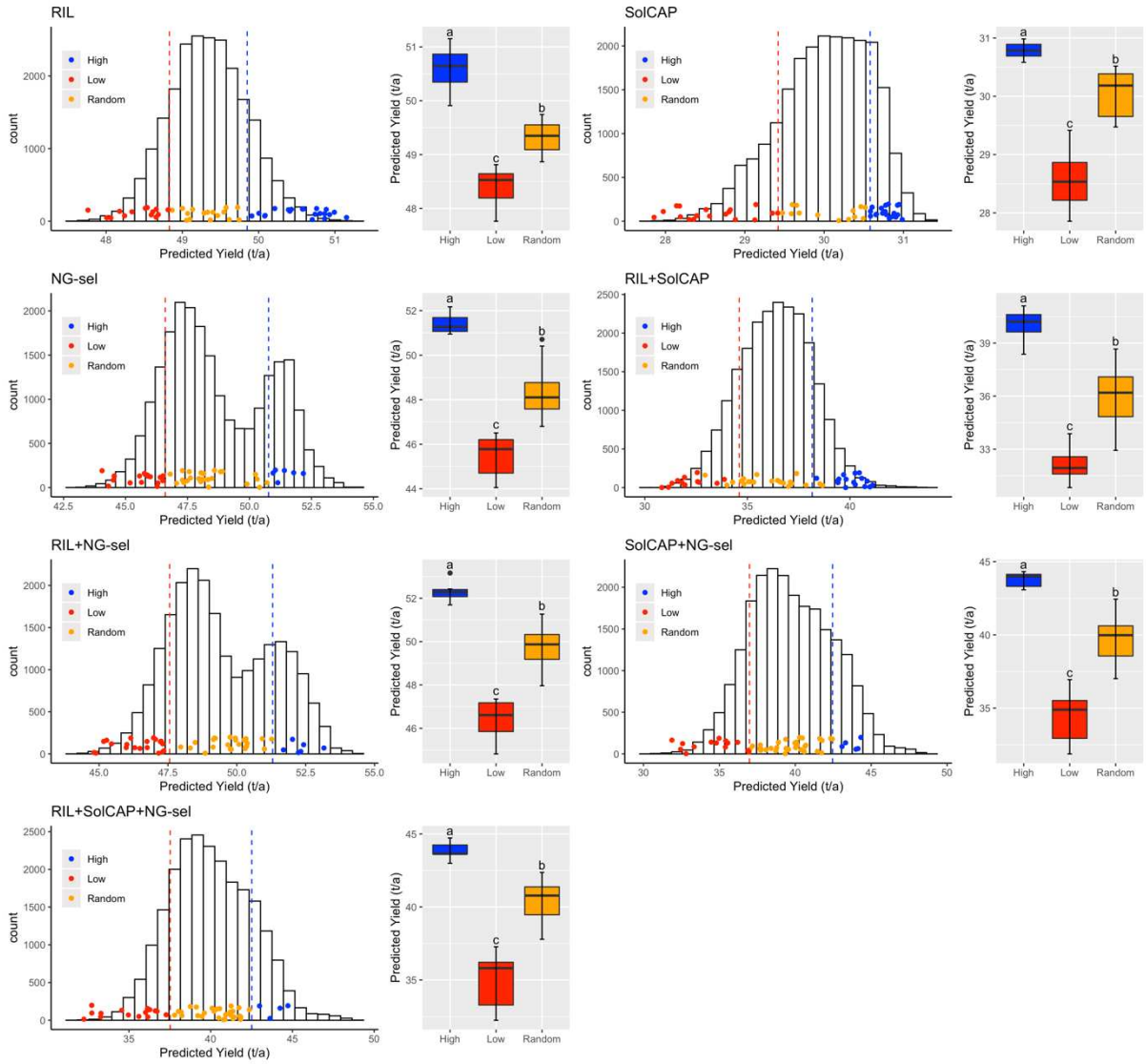


Figure 2. Yield GEBV distributions and mean comparisons for imputed hybrids predicted for each GS model. Red, yellow, and blue points represent low, random, and high-GEBV selections, respectively. Red and blue dashed lines represent low and high-GEBV selection thresholds, respectively. Low-case letters above box plot distributions indicate means that were statistically different according to Tukey's Honest Significant Difference test at a 0.05 probability level.

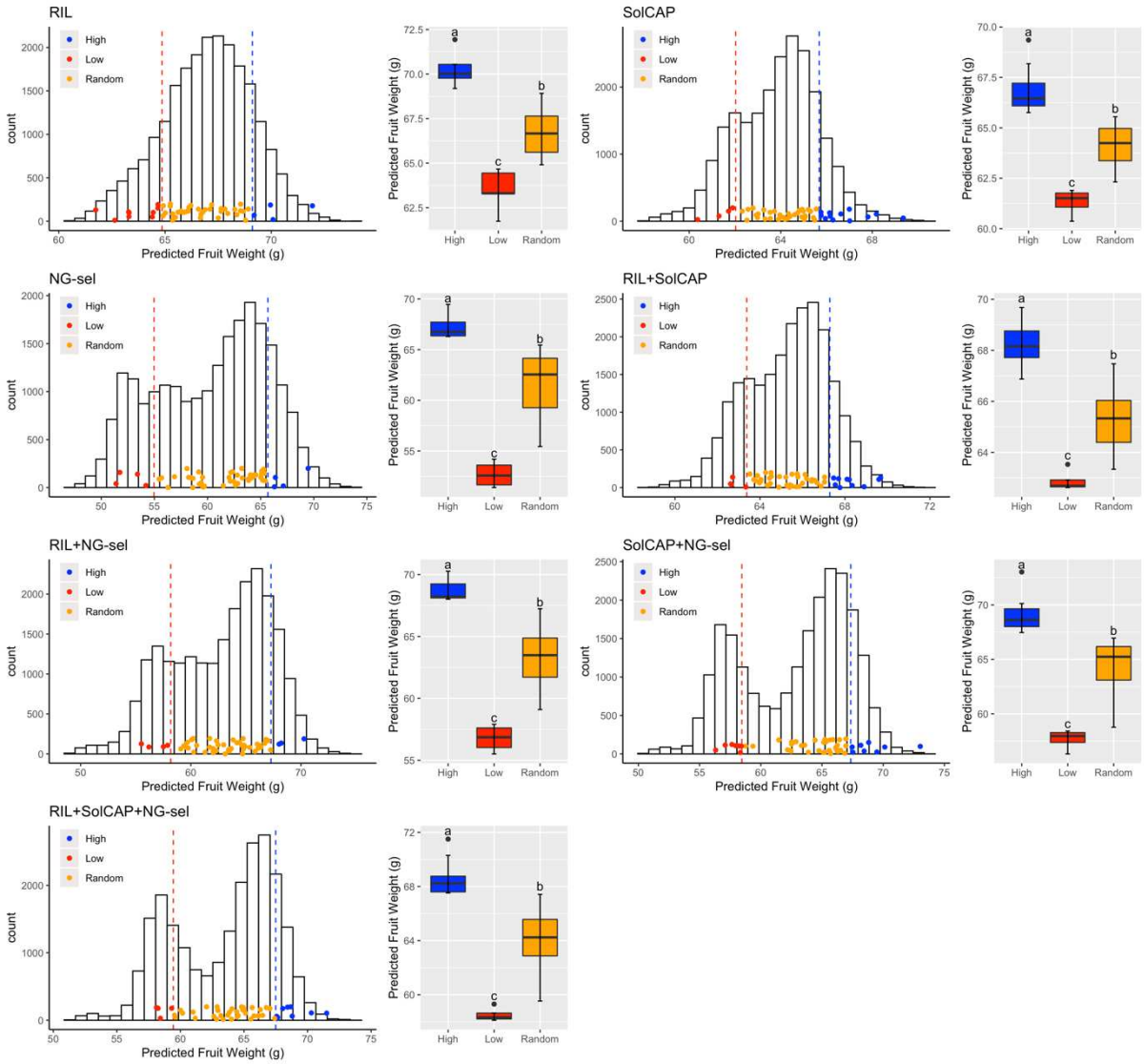


Figure 3. Fruit weight GEBV distributions for hybrids and GEBV mean comparisons for hybrid classes predicted for each GS model. Red, yellow, and blue points represent low, random, and high-GEBV selections, respectively. Red and blue dashed lines represent low and high-GEBV thresholds, respectively. Low-case letters above box plot distributions indicate means that were statistically different according to Tukey's Honest Significant Difference test at a 0.05 probability level.

Predicted phenotypes of hybrids varied in magnitude across models. The smallest predicted phenotype range was found for the SolCAP model where GEBVs of hybrids ranged from 27.81 to 31.38 t/a for yield. The ranges were 47.48 to 51.28 t/a for the RIL model, 42.77 to 54.22 t/a for the NG-sel model, 44.07 to 54.34 t/a for the RIL+NG-sel model, and 30.12 to 44 t/a for RIL+SolCAP, 30.36 to 48.73 t/a for SolCAP+NG-sel, 31.36 to 48.84 t/a for RIL+SolCAP+NG-sel models (Figure

2). Fruit weight GEBVs of hybrids ranged from 58.05 to 70.42 g for SolCAP, 60.61 to 73.84 g for RIL, 47.41 to 74.15 g for NG-sel, 49.08 to 74.61 g for RIL+NG-sel, 58.73 to 71.96 g for RIL+SolCAP, 50.41 to 73.92 g for SolCAP+NG-sel, 51.51 to 73.93 g for RIL+SolCAP+NG-sel models (Figure 3). Bi-modal distributions were observed when the NG-sel population was included in the training set, probably because of directional selection for yield in this sub-set of lines. SolCAP was the model that rendered lowest predicted phenotypes, with hybrids yielding as low as 27.81 t/a. In contrast, the highest predicted phenotypes were found for the RIL+NG-sel, with hybrids yielding as high as 54.34 t/a. The lowest fruit weight predicted value was 47.41 g found for the NG-sel model and highest was 74.61 g found for the RIL+NG-sel model. Mean comparison analyses separated GEBV predictions of high, low, and random groups for all GS models (Figures 2 and 3). Predicted yield of High-GEBV selections is from 0.8 t/a (SolCAP model) to 4 t/a (SolCAP+NG-sel model) greater than predicted yield mean of the entire hybrid population. For fruit weight, the difference between high-fruit weight selections and the overall predicted mean for the 22,681 hybrids varied from 3 g (SolCAP and RIL+SolCAP models) to 7 g (NG-sel model).

Prediction abilities for GS models based on inbred and partially-inbred line populations were first assessed by leave-one-out cross-validation (Table 1). Cross-validation was significant (p-values from $3.42e^{-124}$ to $8.29e^{-03}$) for all GS models. Accuracy (r) varied from 0.26 to 0.92 for yield and from 0.25 to 0.79 for fruit weight suggesting that GS models may be useful predictive tools (Table 1). The RIL+SolCAP was the model with highest cross-validation prediction ($r = 0.92$, $P= 3.42e^{-124}$) for yield (Table 1). Yield performance of the RIL+SolCAP model was then followed by RIL+SolCAP+NG-sel ($r= 0.83$, $P = 1.96e^{-95}$), and the model with the lowest accuracy was RIL ($r=0.26$, $P = 8.22e^{-04}$). For fruit size, NG-sel was the model with highest cross-validation accuracy ($r=0.79$, $P = 1.26e^{-15}$), and SolCAP was the model with the lowest ($r=0.25$, $P = 5.23e^{-11}$).

Table 1. Model accuracy and significance of leave-one-out cross-validation for individual and combined training populations.

Training population	Total Yield		Fruit Weight	
	r	p-value	r	p-value
RIL	0.26	$8.22e^{-04}$	0.44	$5.29e^{-09}$
SolCAP	0.40	$1.45e^{-06}$	0.25	$3.17e^{-03}$
NG-sel	0.32	$8.29e^{-03}$	0.79	$1.26e^{-15}$
RIL+SolCAP	0.92	$3.42e^{-124}$	0.37	$5.23e^{-11}$
RIL+NG-sel	0.33	$3.94e^{-07}$	0.63	$2.34e^{-26}$
SolCAP+NG-sel	0.78	$2.56e^{-43}$	0.67	$3.16e^{-28}$
RIL+SolCAP+NG-sel	0.83	$1.96e^{-95}$	0.57	$4.13e^{-32}$

Hybrid validation trials identified promising genotypes for both yield and for fruit weight. Sixteen hybrids had higher yields than the best-performing check H5108. Yield adjusted means of top-ranking hybrids such as FG21-439 across the two locations was as high as 52 t/acre, against 35.9 t/acre observed for H5108. Twenty-one hybrids had higher fruit weight (from 61.2 to 85.7g) than the best-performing check, in this case, H3406 (60.5 g). Variation in hybrid performance is shown in Supplementary Figures 1a and 1b. Hybrid evaluation locations experienced severe weather in 2022. The Fremont location received 47.4 cm of precipitation between May 26 and September 30, with 9.4 cm falling before July 1 and 15 cm falling between in a little over one-week between July 6 and July 14, 2022 (https://weather.cfaes.osu.edu/). Wooster experienced 49.7 cm during the same May-September time period, but 19.3 cm of precipitation before July and a single June 14 event accounting for 6.5 cm. Field slope in Wooster permitted more rain to run off. There was a lack of correlation for yield ($r = -0.025$; $p = 0.86$) between the two locations. In contrast, there was a significant correlation between locations for fruit weight ($r = 0.62$; $p = 1.9e^{-06}$). Average tomato yield in Fremont were 42.9 t/acre, significantly higher than in Wooster ($P = 2.2e^{-16}$), which was 26.2 t/acre on average (Supplementary Figure 1c). Fruit weight in Fremont (61.7 g) was similar to that in Wooster (59.4 g) (Supplementary Figure 1d). Due to the strong location effects, we decided to investigate correlations between predicted and observed phenotypes using combined and single-location data.

Model predictions for yield performance of hybrids were location-dependent. For observed hybrid yield across both locations, predicted GEBV and observed yield was significant ($p = 0.04$) only for the RIL+SolCAP+NG-sel training population with an accuracy of 0.3. All seven GS models tested predicted hybrid yield well in Wooster (r ranged from 0.32 to 0.42 with p between 0.02 and 0.002). In contrast, model predictions for Fremont yield data showed accuracies that ranged from -0.07 to 0.08 with $p>0.6$ (Table 2). Yield predictions for the Wooster trial were significant and moderately high even for the NG-sel model with the smallest training population size ($r = 0.32$; $p = 0.02$) (Table 2). However, hybrid prediction accuracies for yield tended to increase and become more significant as training populations were combined for GS modeling. Accuracies for RIL, SolCAP, and NG-sel predicted GEBVs and observed yield averages (BLUEs) for the combined data were 0.16 ($p = 0.28$), 0.13 ($P = 0.38$), and 0.14 ($p = 0.34$). Accuracies for RIL+SolCAP $r = 0.22$ ($p = 0.13$), SolCAP+NG-sel $r = 0.27$ ($p = 0.06$), and RIL+SolCAP+NG-sel $r = 0.30$ ($p = 0.04$). A similar pattern was also observed for single-location predictions (Table 2).

Table 2. Prediction ability of GS models for individual and combined-location validation trials.

Training population	Combined locations				Wooster location				Fremont location			
	Total Yield		Fruit Weight		Total Yield		Fruit Weight		Total Yield		Fruit Weight	
	r	p-value	r	p-value	r	p-value	r	p-value	r	p-value	r	p-value
RIL	0.16	0.28	-0.17	0.24	0.38	$6.80e^{-03}$	-0.05	0.71	-0.04	0.80	-0.26	0.07
SolCAP	0.13	0.38	0.58	$1.12e^{-05}$	0.32	0.02	0.58	$1.25e^{-05}$	-0.05	0.75	0.47	$6.45e^{-04}$
NG-sel	0.14	0.34	0.40	$4.36e^{-03}$	0.32	0.02	0.37	$1.00e^{-02}$	-0.05	0.72	0.34	0.02
RIL+SolCAP	0.22	0.13	0.46	$8.06e^{-04}$	0.42	$2.87e^{-03}$	0.48	$4.47e^{-04}$	0.01	0.96	0.35	0.01
RIL+NG-sel	0.15	0.31	0.19	0.19	0.37	0.01	0.23	$1.00e^{-01}$	-0.07	0.65	0.10	0.49
SolCAP+NG-sel	0.27	0.06	0.55	$3.43e^{-05}$	0.38	0.01	0.52	$1.08e^{-04}$	0.06	0.70	0.47	$5.25e^{-04}$
RIL+SolCAP+NG-sel	0.30	0.04	0.49	$2.70e^{-04}$	0.41	$2.85e^{-03}$	0.49	$3.01e^{-04}$	0.08	0.60	0.40	$4.08e^{-03}$

Unlike yield, fruit weight predictions were more dependent on the GS model used than the location where selected hybrids were tested. Using the data from combined locations, fruit weight of hybrids was predicted by all models except RIL and RIL+NG-sel (Table 2) with prediction accuracies ranging from 0.4 to 0.58 and significance ranging from $4.36e^{-03}$ to $1.12e^{-05}$ (Table 2). Wooster prediction coefficients ranged from 0.37 to 0.58 (with p-values ranging from $1.08e^{-04}$ to $1.25e^{-05}$) and were slightly better than for Fremont which ranged from 0.30 to 0.47 (with p-values ranging from $5.25e^{-04}$ to 0.03) (Table 2). SolCAP was the most accurate model for fruit weight, with r = 0.58 for the combined-location data and r = 0.58 and 0.47 for Wooster and Fremont data, respectively, followed by the SolCAP+NG-sel model with r = 0.55 for the combined-location data and r = 0.52 and 0.47 for Wooster and Fremont data, respectively. GS models predicted fruit weight of hybrids more accurately than they did for yield as suggested by higher prediction coefficients (Table 2).

Despite differences in predictability of models, linear associations between GEBV predictions for all model combinations showed significant correlation. Tables 3 and 4 contain correlation coefficients between predicted yield and fruit weight among models, respectively. Correlations for predicted yield between single-training population models such as RIL and SolCAP (r = 0.05), RIL and NG-sel (r = 0.10), NG-sel and SolCAP (r = 0.37) models were relatively weak. On the other hand, correlations for predicted yield between single and combined training-population models such as RIL and RIL+SolCAP (r = 0.47), RIL and RIL+NG-sel (r = 0.25) SolCAP and RIL+SolCAP (r = 0.50), SolCAP and SolCAP+NG-sel (r = 0.45), NG-sel and RIL+NG-sel (r = 0.97), NG-sel and SolCAP+NG-sel (r = 0.82), were strong if the single-population was included in the combined-population model. Correlations among fruit weight predictions of models were higher than for yield but followed a similar pattern. The RIL+SolCAP+NG-sel model was identified as the most precise model due to its high accuracy in most cases. The highly significant correlations between RIL+SolCAP+NG-sel and SolCAP+NG-sel models for both yield (r = 0.95) and fruit weight (r = 0.97) suggest that the RIL population has little impact on the predictive ability of RIL+SolCAP+NG-sel.

Table 3. Similarities between GEBV predictions for yield among GS models. Numbers below the matrix diagonal are correlation coefficients (r) with their respective p-values above the diagonal.

	RIL	**SolCAP**	**NG-sel**	**RIL+SolCAP**	**RIL+NG-sel**	**SolCAP+ NG-sel**	**RIL+SolCAP+ NG-sel**
RIL	1	$8.1e^{-14}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$
SolCAP	0.05	1	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$
NG-sel	0.10	0.37	1	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$
RIL+SolCAP	0.47	0.50	0.09	1	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$
RIL+NG-sel	0.25	0.37	0.97	0.20	1	$< 2.2e^{-16}$	$< 2.2e^{-16}$
SolCAP+NG-sel	0.10	0.45	0.82	0.37	0.79	1	$< 2.2e^{-16}$
RIL+SolCAP+NG-sel	0.18	0.45	0.77	0.55	0.79	0.95	1

Table 4. Similarities between GEBV predictions for fruit weight among GS models. Numbers below the matrix diagonal are correlation coefficients (r) with their respective p-values above the diagonal.

	RIL	**SolCAP**	**NG-sel**	**RIL+SolCAP**	**RIL+NG-sel**	**SolCAP+ NG-sel**	**RIL+SolCAP+ NG-sel**
RIL	1	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$
SolCAP	0.31	1	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$
NG-sel	0.50	0.68	1	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$
RIL+SolCAP	0.55	0.88	0.76	1	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$
RIL+NG-sel	0.79	0.63	0.89	0.79	1	$< 2.2e^{-16}$	$< 2.2e^{-16}$
SolCAP+NG-sel	0.44	0.87	0.91	0.88	0.85	1	$< 2.2e^{-16}$
RIL+SolCAP+NG-sel	0.54	0.84	0.89	0.94	0.90	0.97	1

The weak correlation between RIL, NG-sel, and SolCAP-predicted GEBVs may be related to genetic differentiation among individuals from these training populations. A PCA graph based on marker covariances illustrates genetic differences with principal components 1 and 2 explaining 57.4% of the total variation in the data set (Figure 3). The first principal component explains over 45% of the observed variation and separates the training populations. Observations from this analysis are consistent with known attributes of the training populations. For example, the SolCAP population was characterized by high variation for PC1 and PC2, with clustering skewed toward positive values on PC1. The large ellipse for this population is consistent with higher genetic variability within this collection of inbreds. The RIL population, in contrast, clustered toward negative values of PC1 and was delimited by a much smaller ellipse consistent with a bi-parental population with ~25% pedigree contribution from fresh-market Roma tomato. The NG-sel population has the second largest ellipse, consistent with its six-parent origin. Hybrids clustered intermediate to the SolCAP and NG-sel populations.

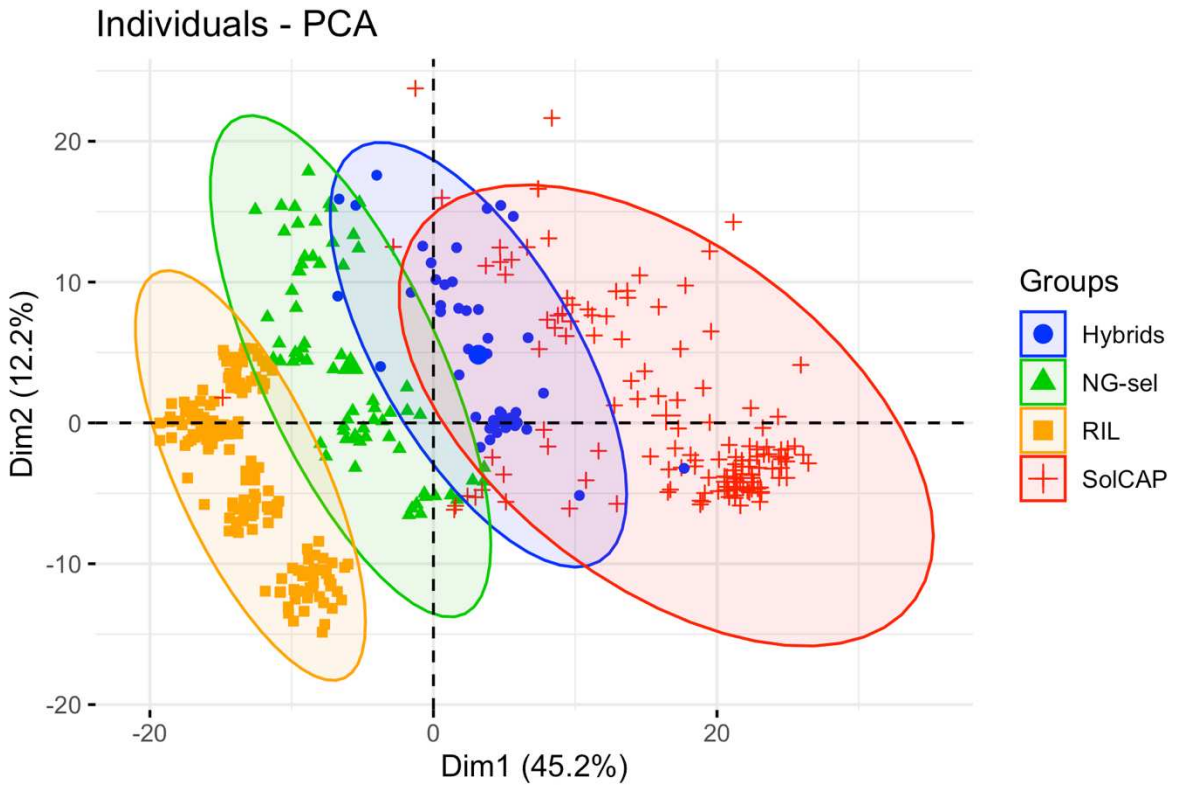


Figure 4. Principal component analysis (PCA) to visualize genetic relationships between individuals in training populations. Color coding in the PCA graph distinguishes the three training populations involved in GS modeling as well as the selected hybrids based on covariance of the genotype information from 248 SNPs. Ellipses represent concentration ellipses for each group.

Comparisons between selection groups were consistent with validation results. Figure 4 contains mean comparison tests for selection groups considering the best model (RIL+SolCAP+NG-sel). Average yield for high GEBV selections differed from low GEBV selections in Wooster ($p<0.05$) where the predictive ability of RIL+SolCAP+NG-sel was significant ($r = 0.41$; $p = 2.85e^{-03}$) (Figure 4b). In Fremont, where the predictive ability was not significant in correlation analysis, we also did not observe differences between selection groups ($p>0.05$) (Figure 4c). When the RIL+SolCAP+NG-sel model displayed moderate prediction, such as for yield in the combined locations (Figure 4a), we could not differentiate high from low yield-GEBV classes based on mean comparison tests and yet there was a tendency for high-GEBV to surpass low-GEBV selections. Differences in fruit weight between high and low-GEBV selections were highly significant in the combined data and across locations for the RIL+SolCAP+NG-sel

model (Figure 4d, e, and f). Fruit weight averages for the random fruit weight-GEBV group was lower than the high fruit weight-GEBV group suggesting that fruit weight improvement is more likely to be achieved through selection based on estimates of genetic potential than on random selection of hybrids.

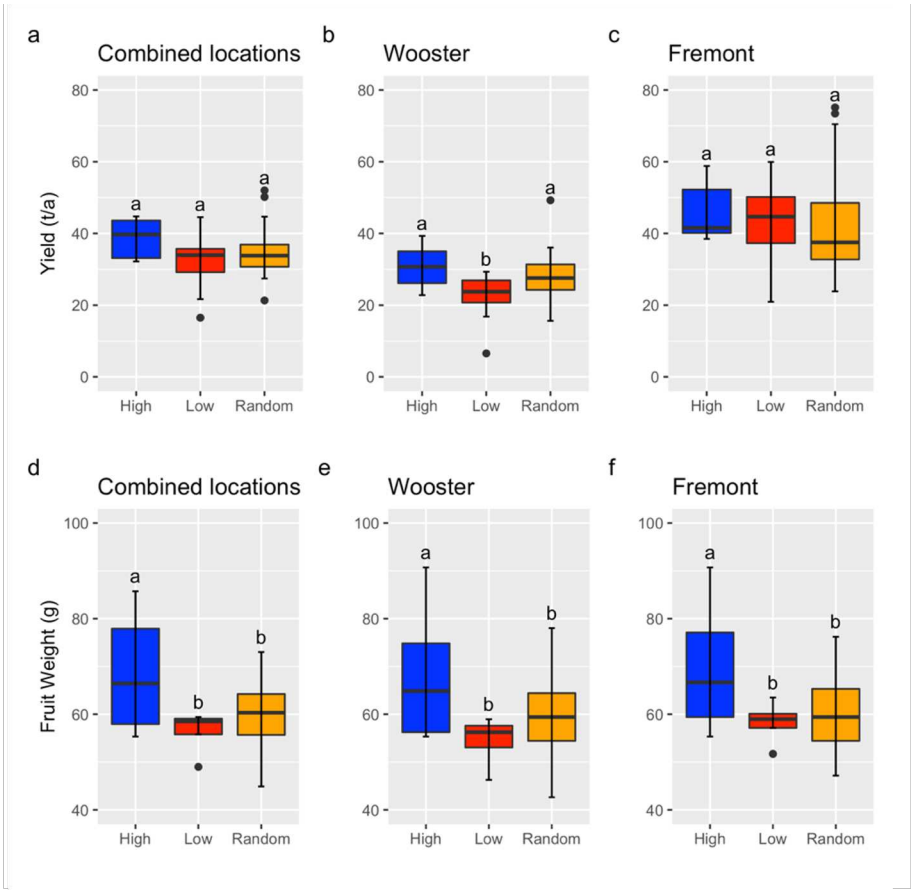


Figure 5. Mean comparisons for observed yield (a, b, c), fruit weight (d, e, f), and selection index (a, h, i) among selection groups for the RIL+SolCAP+NG-sel model considering data from Wooster, Fremont, and combined locations. High refers to high GEBV selections, Low to low GEBV selections, and Random to random GEBV selections for each trait. Different low-case letters indicates that means were statistically different according to the Tukey's Honest Significant Difference test at a 0.05 probability level. For more information about selection groups see Supplementary Tables 1 and 2.

4.4. Discussion

GS is an under-explored tool for horticultural crops such as tomato (Cappetta et al., 2020) despite positive evidence for applicability (Duangjit et al., 2016; Liabeuf et al., 2018; Yamamoto et al., 2017). Preliminary studies in rice have shown great promise of GS in predicting performance of thousands of hybrids that have not yet been created (Cui et al., 2020; Wang et al., 2017; Xu et al., 2014). Because the genotype of hybrids can be imputed based on the genotype of parents, GS can predict the performance of as many hybrids as there are combinations of genotyped parents with no hybrid genotyping needed. Then, only top-ranked hybrids would be selected for field testing with great potential for targeted hybridization and field evaluation within the scope of defined resources. Selection of hybrids could become less subjective and is more likely to succeed given the predictive ability of models. This paper demonstrates that GS models can be used for large-scale prediction of yield-related traits in tomato hybrids. Predictive abilities for field validation trials were optimistic with coefficients as high as 0.42 ($p=2.87e^{-03}$) for yield and 0.58 ($p=1.12e^{-05}$) for fruit weight. Results from mean comparison tests corroborated with these field validation coefficients as observed phenotype averages for high-GEBV selections were greater than for low-GEBV selections in most cases.

A significant genotype-by-environment (GxE) interaction seems to compromise predictability of the GS models and validation presented here. The lack of phenotypic correlation across locations suggested a strong GxE interaction for yield in the hybrid validation trial. Yield predictions that were fairly accurate in the Wooster location were lost in the Fremont and combined location analyses. For fruit weight, in which correlation across locations was relatively high, GS models worked for both single and combined location data. GxE interaction has been shown to affect prediction for traits such as yield and tiller number in hybrid rice validation trials (Xu et al. 2014). These traits had low heritability values ($H^2 = 0.38$ for yield and $H^2 = 0.50$ for tiller number). The fact that we observed significant predictions for the RIL+SolCAP+NG-sel model in the combined location analysis suggests that increasing the training population size may help to attenuate the GxE interaction problem. Increases in population size have been shown to avoid ascertainment bias providing more precise marker effect estimates (de Bem Oliveira et al., 2020). Alternative GS models that account for GxE effects may also be used to improve prediction accuracies (Jighly et al., 2021).

GS studies in tomato reported improvements in predictive ability of models for fruit quality traits with increase in marker number, training population size, and genetic relatedness between

training and testing lines (Duangjit et al. 2016). Duangjit et al. 2016 obtained cross-validation prediction averages of 0.81 for fruit weight when a mixture of *S. lycopersicum* (SL) and *S. pimpinellifolium* (SP) accessions were used for training. The GS model trained with SL accessions only, on the other hand, predicted fruit weight of other SL accessions well but failed to predict fruit weight of SP accessions. The best fruit weight predictions here were observed for GS models in which at least one set of parents was used for training also suggesting that genetic proximity between training and prediction sets has an impact on predictive ability of models. The inclusion of SolCAP parents in model training was imperative for fruit weight predictions to work. These observations suggest that maximizing fruit weight predictions of hybrids requires not only including parents in training but also making sure that the parents included are representative of the total genetic variation present in the prediction set. Because the RIL model predicted yield but did not predict fruit weight of hybrids in Wooster, we suspect that close genetic similarity between training lines and hybrids is more important for fruit weight than it is for yield. In contrast to fruit weight, yield predictions were maximized by increasing training population size.

4.5. Conclusions

Our results suggest that genomic prediction is a promising alternative to the random process of designing new tomato hybrids. Large-scale genomic prediction for yield-related traits in tomato was effective even for modest training population sizes (69 individuals) and low marker numbers (248 SNPs). Fruit weight predictions of models were more accurate than yield predictions as indicated by higher validation coefficients for fruit weight. GxE interaction affected predictions for environmentally-sensitive traits like yield. Including parents of hybrids in GS modeling and increasing training population sizes is recommended to maximize predictive abilities for fruit weight and yield models, respectively.

4.6. References

Aisyah, S.I., Wahyuni, S., Syukur, M., Witono, J.R., 2016. The estimation of combining ability and heterosis effect for yield and yield components in tomato (*Solanum lycopersicum* Mill.) at lowland. Ekin J. 2, 23-29.

https://dergipark.org.tr/en/pub/ekinjournal/issue/22787/243193

Avdikos, I.D., Nteve, G.M., Apostolopoulou, A., Tagiakas, R., Mylonas, I., Xynias, I.N., Papathanasiou, F., Kalaitzis, P., Mavromatis, A.G., 2021. Analysis of re-heterosis for yield and fruit quality in restructured hybrids, generated from crossings among tomato recombinant lines. Agronomy 11, 822. https://doi.org/10.3390/agronomy11050822

Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw., 67, 1-48. https://doi.org/10.18637/jss.v067.i01.

Bernal, E., Francis, D.M., 2021. Processing tomato germplasm with improved resistance to bacterial spot. HortScience 56, 519-520. https://doi.org/10.21273/HORTSCI15616-20

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

Borrenpohl, D., Huang, M., Olson, E., Sneller, C., 2020. The value of early-stage phenotyping for wheat breeding in the age of genomic selection. Theoretical and Applied Genetics 133, 2499–2520. https://doi.org/10.1007/s00122-020-03613-0

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

Cui, Y., Li, R., Li, G., Zhang, F., Zhu, T., Zhang, Q., Ali, J., Li, Z., Xu, S., 2020. Hybrid breeding of rice via genomic selection. Plant Biotechnol. J. 18, 57–67. https://doi.org/10.1111/pbi.13170

de Bem Oliveira, I., Amadeu, R.R., Ferrão, L.F.V., Muñoz, P.R., 2020. Optimizing whole-genomic prediction for autotetraploid blueberry breeding. Heredity 125, 437–448. https://doi.org/10.1038/s41437-020-00357-x

de Mendiburu, F., 2021. agricolae: Statistical Procedures for Agricultural. Research. R package version 1.3-5. https://CRAN.R-project.org/package=agricolae

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

Duhnen, A., Gras, A., Teyssèdre, S., Romestant, M., Claustres, B., Daydé, J., Mangin, B., 2017. Genomic selection for yield and seed protein content in soybean: A study of breeding program data and assessment of prediction accuracy. Crop Sci. 57, 1325–1337. https://doi.org/10.2135/cropsci2016.06.0496

Duvick, D.N., 1999a. Commercial strategies for exploitation of heterosis. *Genetics and exploitation of heterosis in Crops,* 295-304.

Duvick, D.N., 1999b. Heterosis: Feeding people and protecting natural resources. *Genetics and exploitation of heterosis in crops*, 19-29.

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

Figueiredo, A.S.T., de Resende, J.T.V., Faria, M.V., de Paula, J.T., Schwarz, K., Zanin, D.S., 2015. Combining ability and heterosis of relevant fruit traits of tomato genotypes for industrial processing. Crop Breeding and Applied Biotechnology 15, 154–161. https://doi.org/10.1590/1984-70332015v15n3a27

Garg, N., Cheema, D.S., 2008. Genotype environment interactions for shelf life and yield attributes in tomato hybrids heterozygous at rin, nor, or alc loci. J. Crop Improv. 22, 17–30. https://doi.org/10.1080/15427520802042697

Ito, Y., Nakamura, N., Kotake-Nara, E., 2021. Semi-dominant effects of a novel ripening inhibitor (rin) locus allele on tomato fruit ripening. PLoS One 16, e0249575. https://doi.org/10.1371/journal.pone.0249575

Jighly, A., Hayden, M., Daetwyler, H., 2021. Integrating genomic selection with a genotype plus genotype x environment (GGE) model improves prediction accuracy and computational efficiency. Plant Cell Environ. 44, 3459–3470. https://doi.org/10.1111/pce.14145

Kassambara, A., Mundt, F., 2020. _factoextra: Extract and Visualize the Results of Multivariate Data Analyses_. R package version 1.0.7, https://CRAN.R-project.org/package=factoextra.

Kumar, R., Srivastava, K., Kumar, V., Saroj, S.K., Sharma, S.K., Singh, R.K., 2019. Heterosis analysis in tomato (*Solanum lycopersicum* L.) for lycopene, TSS, titrable acidity, and Ascorbic acid. Electronic Journal of Plant Breeding 10, 1547–1553. https://doi.org/10.5958/0975-928X.2019.00198.4

Lammerts Van Bueren, E.T., Jones, S.S., Tamm, L., Murphy, K.M., Myers, J.R., Leifert, C., Messmer, M.M., 2011. The need to breed crop varieties suitable for organic farming, using wheat, tomato and broccoli as examples: A review. NJAS - Wageningen Journal of Life Sciences 58, 193–205. https://doi.org/10.1016/j.njas.2010.04.001

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 108, 392–401. https://doi.org/10.1094/PHYTO-12-16-0431-R

Liu, J., Li, M., Zhang, Q., Wei, X., Huang, X., 2020. Exploring the molecular basis of heterosis for plant breeding. J Integr Plant Biol 62, 287–298. https://doi.org/10.1111/jipb.12804

Liu, Z., Jiang, J., Ren, A., Xu, X., Zhang, H., Zhao, T., Jiang, X., Sun, Y., Li, J., Yang, H., 2021. Heterosis and combining ability analysis of fruit yield, early maturity, and quality in tomato. https://doi.org/10.3390/agronomy

Lozada, D.N., Mason, R.E., Sarinelli, J.M., Brown-Guedira, G., 2019. Accuracy of genomic selection for grain yield and agronomic traits in soft red winter wheat. BMC Genet. 20, 1-12. https://doi.org/10.1186/s12863-019-0785-1

Merk, H.L., Yarnes, S.C., Deynze, A. Van, 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, J. Amer. Soc. Hort. Sci. *137*, 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

Precheur, R. J., 2005. Ohio Vegetable Production Guide. The Ohio State University. Ext. Bull. 672.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., Reich, D., 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904–909. https://doi.org/10.1038/ng1847

R Core Team, 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Robbins, M.D., Masud, M.A., Panthee, D.R., Gardner, R.G., Francis, D.M., Stevens, M.R., 2010. Marker-assisted selection for coupling phase resistance to Tomato spotted wilt virus and *Phytophthora infestans* (late blight) in tomato, HortScience, *45*, 1424-1428. https://doi.org/10.21273/HORTSCI.45.10.1424

Sim, S.C., Durstewitz, G., Plieske, J., Wieseke, R., Ganal, M.W., van Deynze, A., Hamilton, J.P., Buell, C.R., Causse, M., Wijeratne, S., Francis, D.M., 2012a. Development of a large snp

genotyping array and generation of high-density genetic maps in tomato. PLoS One *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., Mutschler, 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, 1–18. https://doi.org/10.1371/journal.pone.0045520

Soresa, D.N., Nayagam, G., Bacha, N., Jaleta, Z., 2020. Heterosis in tomato (*Solanum lycopersicum* L.) for yield and yield component traits. Adv. Res. 21, 141–152. https://doi.org/10.9734/air/2020/v21i930242

Vekariya, T., Kulkarni, G., Vekaria, D., Dedaniya, A., Memon, J., 2019. Combining ability analysis for yield and its components in tomato (*Solanum lycopersicum* L.). Acta Scientific Agriculture 3, 185–191. https://doi.org/10.31080/asag.2019.03.0541

Wang, X., Li, L., Yang, Z., Zheng, X., Yu, S., Xu, C., Hu, Z., 2017. Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II. Heredity 118, 302–310. https://doi.org/10.1038/hdy.2016.87

Xu, S., Zhu, D., Zhang, Q., 2014. Predicting hybrid performance in rice using genomic best linear unbiased prediction. Proc. Natl. Acad. Sci. 111, 12456–12461. https://doi.org/10.1073/pnas.1413750111

Yamamoto, E., Matsunaga, H., Onogi, A., Ohyama, 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

Yan, J., 2016. _som: Self-Organizing Map_. R package version 0.3-5.1. https://CRAN.R-project.org/package=som

Zhao, Y., Gowda, M., Liu, W., Würschum, T., Maurer, H.P., Longin, F.H., Ranc, N., Reif, J.C., 2012. Accuracy of genomic selection in European maize elite breeding populations. Theoretical and Applied Genetics 124, 769–776. https://doi.org/10.1007/s00122-011-1745-y

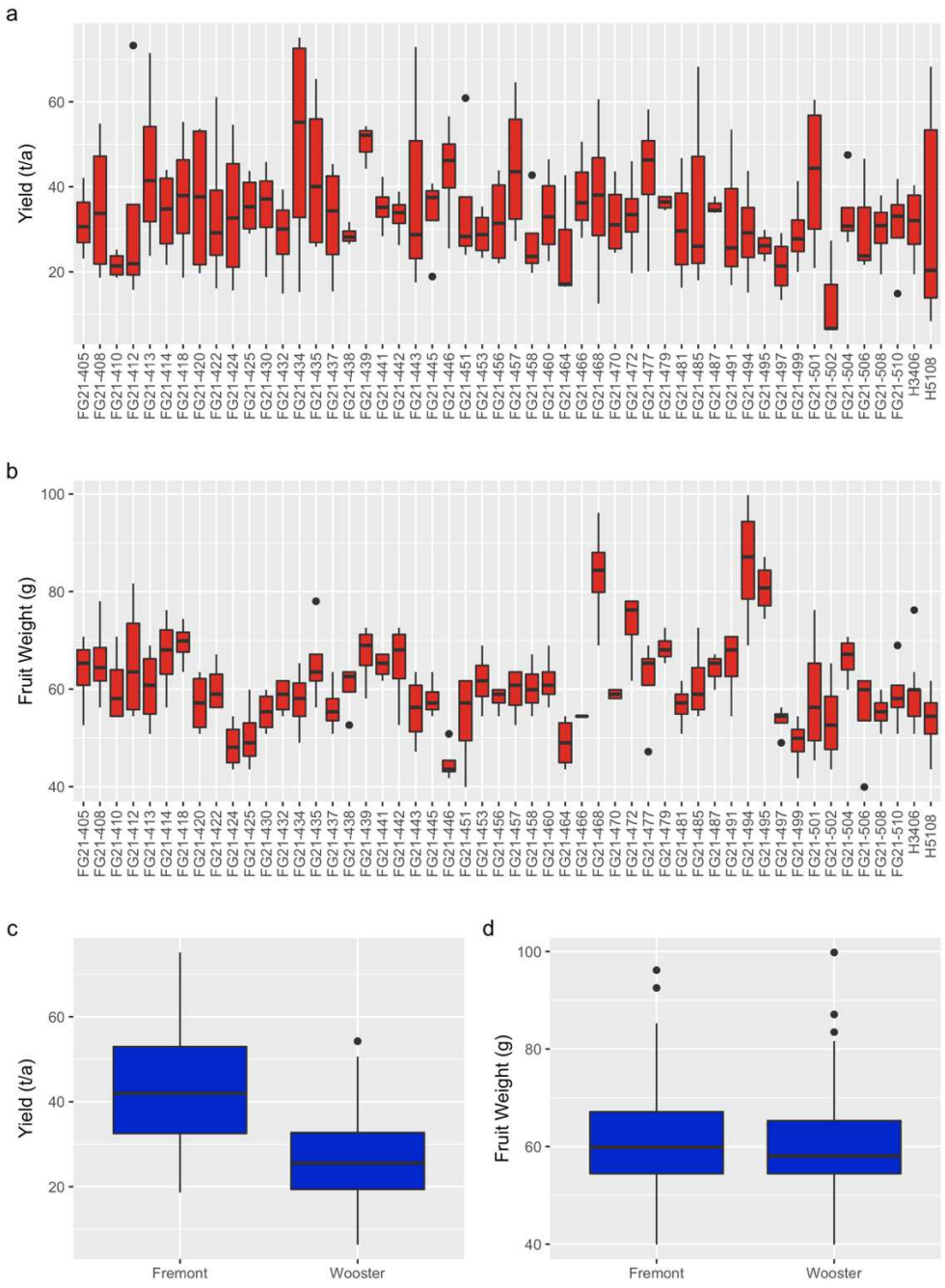
4.7. Supplementary materials

Supplementary Table 1. GEBV means (Mean) and standard deviations (SD) as well as high and low GEBV thresholds for each model.

Model	Yield (t/a)		Thresholds		Fruit Weight (g)		Thresholds	
	Mean	SD	High	Low	Mean	SD	High	Low
Cmm	49.34	0.51	49.85	48.83	66.98	2.12	69.11	64.86
SolCAP	30.00	0.58	30.58	29.42	63.86	1.83	65.68	62.03
NG-sel	48.68	2.08	50.77	46.60	60.32	5.36	65.68	54.96
RIL+SolCAP	36.38	1.78	38.16	34.60	65.33	1.96	67.29	63.38
RIL+NG-sel	49.43	1.87	51.30	47.56	62.73	4.55	67.28	58.17
SolCAP+NG-sel	39.72	2.74	42.46	36.98	62.90	4.45	67.35	58.45
RIL+SolCAP+NG-sel	40.01	2.49	42.51	37.52	63.46	4.02	67.49	59.44

Supplementary Table 2. Number of hybrids for each selection group after reassignment. High = high GEBV selections for yield; Low GEBV selections for yield; Random GEBV selections.

Model	Yield selection groups			Fruit Weight selection groups		
	High	Low	Random	High	Low	Random
Cmm	20	14	16	4	9	37
SolCAP	19	18	13	12	4	34
NG-sel	6	18	26	4	4	42
RIL+SolCAP	16	13	21	10	4	36
RIL+NG-sel	6	20	24	3	4	43
SolCAP+NG-sel	5	14	31	8	7	35
RIL+SolCAP+NG-sel	5	16	29	10	4	36



Supplementary Figure 1. Boxplots illustrating differences in yield and fruit size across hybrids (a, b) and locations (c, d).

CONCLUSIONS

This dissertation provides resources and suggests molecular-based strategies for tomato cultivar improvement under unpredictable weather patterns. The first study detected QTLs for increased yield (Chr 3) and lycopene content (Chr 2, and 3) under optimal irrigation and for fruit firmness (Chr 3) and lycopene content (Chr 2, 3, and 7) under deficit irrigation conditions within a *Solanum pennellii* IL population. Water deficit treatments reduced yield while positively affecting fruit quality of selected *S. pennellii* ILs. In the second study, we identified a major dominant QTL for increased fruit puffiness, an abiotic-stress-related disorder in tomato fruits. The QTL located on Chr 1 was reproducible across populations and mapping approaches. FG02-188, and advanced tomato line, donated the allele responsible for fruit puffiness in both mapping and validation populations. MAS for the markers solcap_snp_sl_20440 and solcap_snp_sl_18619 and GS are recommended for fast-tracking and selection of puffiness-resistant lines. The third study is the first attempt at genomic hybrid breeding in tomatoes. Phenotype prediction of over 22,000 tomato hybrids was possible through GS modeling. Model training with population and marker numbers as small as 69 individuals and 248 SNPs, respectively, provided significant yield predictions. GxE interaction was problematic for successful yield prediction but did not affect fruit weight. Fruit weight predictions were maximized when parents of hybrids were involved in training while yield predictions were higher the larger the training population size.