GENETICS AND BREEDING OF EARLY BLIGHT AND BACTERIAL SPOT RESISTANT TOMATOES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Taylor Anderson

December 2020

GENETICS AND BREEDING OF EARLY BLIGHT AND BACTERIAL SPOT RESISTANT TOMATOES

Taylor Anderson, Ph. D.

Cornell University 2020

Chapter I. We employed a rapid and low-cost trait introgression methodology to transfer broad-spectrum genetic resistance to bacterial spot (*Xanthamonas spp.*) across tomato (*Solanum lycopersicum* L.) market classes. High-density genotyping of finished lines confirmed resistance introgression and background genome recovery but uncovered cryptic introgressions lurking in the background genome. Near-isogenic fresh market breeding lines were evaluated for horticultural performance and multi-race bacterial spot resistance, demonstrating both race non-specific and race-specific resistance. Digital fruit image analysis revealed subtle changes in tomato fruit quality characteristics and yield trials found differences in maturity that were associated with bacterial spot resistance haplotypes.

Chapter II. We investigated the genetic determinants of early blight (*Alternaria linariae*) resistance in modern tomato breeding lines. Quantitative trait loci (QTL) associated with early blight resistance were detected in cross-market breeding populations and offered substantial protection against stem lesions (collar rot) and moderate protection against defoliation. Subsequent field trials validated the three most promising QTL, *EB-1.2*, *EB-5*, and *EB-9*. Resistance effects for *EB-5* and *EB-9* were consistent across breeding populations and environments, while *EB-1.2*'s effect was population-specific. We developed near-isogenic fresh market tomato lines and found their resistance to be largely mediated by *EB-5* and *EB-9*, together capturing 49.0% and 68.7% of the defoliation and stem lesion variance, respectively.

Chapter III. Whole-genome resequencing of modern tomato breeding lines paired with efficient local ancestry inference revealed cryptic early blight resistance introgressions that were traced to specific donors. Early blight stem lesion and foliar resistance from *EB-9* was traced to the vintage tomato *Devon Surprise*, which is probably derived from *Ailsa Craig*. Foliar resistance from *EB-5* was traced to *Hawaii 7998*. Definition of the shared ancestral haplotypes enabled fine mapping of the resistance loci and the identification of candidate variants possibly underlying resistance. Co-analysis with resequencing data for 769 accessions predicted *EB-9* resistance in several vintage and cherry tomatoes. Foliar *EB-5* resistance was rare among sequenced tomatoes and was not detected with high confidence in any accession. Similarly, we found little evidence of introgression from *Solanum habrochaites* PI 126445 in modern tomatoes, despite being a commonly cited source of resistance.

BIOGRAPHICAL SKETCH

Taylor Anderson grew up amongst the picturesque vineyards of California's Sonoma Valley. The first job he enjoyed was working in a winery, where he made, sold, and consumed wine with his friends and colleagues. These early experiences in winemaking led him to the Viticulture and Enology programs at the Santa Rosa Junior College and the University of California, Davis. There, his coursework introduced him to the fascinating histories of plant domestication and breeding, and he followed that passion into courses on plant breeding and genetics. While working as an Undergraduate Research Assistant, he had the opportunity to work directly with plant breeders and decided to become one himself. Taylor received his PhD in Plant Breeding from Cornell University, where he was fortunate to receive grants and fellowships in support of his professional development. He considers himself lucky to be the last graduate student of Dr. Martha Mutschler, under whom he studied the genetics of disease resistance and horticultural quality in tomato. As of this date, he has accepted a position as a breeder of hot peppers, which is fitting, since he puts hot sauce on literally everything 🌶️.

I dedicate this document to my wife, Abby, who has sacrificed so much in order for me to pursue my passion

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my committee chair and primary mentor, Dr. Martha Mutschler-Chu, who has provided critical support and guidance throughout my PhD. From you, I have learned how to be a plant breeder and scientist and how to approach my personal and professional goals with focus and intentionality. I admire your grit, determination, and passion, and I am immensely grateful to you for this opportunity.

Thank you to Dr. David Francis of The Ohio State University, who I consider an unofficial co-chair to my education. You have been an invaluable mentor, friend, and collaborator, reminding me to apply the law of parsimony throughout my work and offering clarifying guidance in moments of scientific complexity and personal confusion. Thank you for all your time and effort spent mentoring me despite no obligation to do so.

I would like to thank the members of my graduate committee, Dr. Christine Smart, Dr. Michael Gore, and Dr. Margaret Smith. You have all been incredibly generous with your time, knowledge, and support throughout my education. Collaboration with you and the members of your programs has significantly enriched my professional development. In particular, I would like to point out Nicholas S. Kaczmar of the Gore lab, whose collaboration enabled me to gain experience working with high-throughput phenotype data, and Martha Sudermann of the Smart lab, whose close working partnership has taught me much and has contributed significantly to this work. I also want to thank you all for your feedback on publications and this dissertation, and for all the time spent listening to me monologue in committee meetings.

Thank you to my graduate student collaborators Martha Amali Sudermann, in Dr. Chris Smart's program, and Eduardo Bernal, a member of Dr. David Francis' program. Martha Sudermann has been a close collaborator on early blight research in tomato and has made significant contributions throughout the third chapter of this dissertation, from helping to align genetic sequences, to data analysis, to drafting the chapter. Eduardo Bernal has made significant contributions to the first chapter, performing the majority of pathogen ratings,

answering questions from me throughout the project, and helping to review and edit the manuscript.

Thank you to my funding sources. Two years of my professional development were supported by the United States Department of Agriculture – National Institute of Food and Agriculture Predoctoral Fellowship No. 12680140. My research was also supported in part by two-year Sustainable Agriculture Research and Education (SARE) Graduate Student Grant GNE17-143. This research was funded in part by a Schmittau-Novak Grant from the School of Integrative Plant Science, Cornell University, as well as New York Specialty Crops Grant SCG17-001. Thank you to Michael Mazourek for employing me as a Technician just prior to the commencement of my graduate program and for allowing me to assist in teaching his undergraduate course in Plant Genetics, which supported me for one semester. Thank you to the Cornell SIPS Plant Breeding Section for providing summer research stipend support and for building and maintaining an enriched social and professional environment.

Thank you to the current and past members of the Mutschler group, Darlene DeJong, Stella Zitter, Michael Glos, Julie Bojanowski, Richard Gaisser, and Enosh Tzin. Darlene taught me how to perform molecular biology laboratory techniques in a highly efficient manner, has been vital to all operations of our laboratory program, and has contributed significantly to my work. She has also been a friend and has even allowed my wife and I to live in her nicely furnished basement for the next couple of months! Stella Zitter offered guidance and mentorship on methods in plant pathology and performed pilot experiments to optimize the early blight protocols. Michael and Julie have made so much of this work possible as the leaders of our greenhouse and field operations - thank you for all your hard work, for your forward thinking and planning, for being reliable, friendly and knowledgeable, and for interpreting my often late, vague, and sometimes cryptic instructional emails!

Thank you to Diane Ullman and Sulley Ben-Mahmoud of the University of California Davis for giving me the opportunity to obtain my first work experience as a scientist as an

Undergraduate Researcher, for mentoring me, and for collaborating with me on projects and publications in the earlier years of my graduate program.

Thank you to Dr. Randy Gardner, North Carolina State University, for supplying seed of critical breeding lines and for reviewing and providing critical feedback on the manuscript. Thank you to Dr. Esther van der Knaap, University of Georgia, for sharing seed of tomato accessions.

Thank you to my wife, Abigail Lawton, and my parents, Misty and Tony Anderson, for being my systems of support and encouragement throughout my higher education. At times, my graduate training was all-consuming. Thank you for sticking with me through it all as I selfishly pursued my dream of becoming a plant breeder and scientist.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

LIST OF TABLES

CHAPTER 1

CHAPTER 2

CHAPTER 3

PREFACE

This work in three chapters encompasses our efforts to breed elite fresh market tomatoes with combined genetic resistances to three damaging tomato diseases of the Northeast United States: early blight (*Alternaria linariae*), bacterial speck (*Pseudomonas syringae* pv. *tomato*), and bacterial spot (*Xanthamonas spp*). Early blight is a widespread and problematic disease affecting tomatoes. Infections are worst in humid growing regions such as the Northeast and Mid-Atlantic where regular sprays over the course of the season are needed to limit the spread of early blight disease. Bacterial speck has become one of the most commonly reported diseases of tomato in the Northeast in recent years, and bacterial spot is increasingly detected in the Northeast as summers become warmer and wetter (Pers. comm. Chris Smart). Chemical control has limited efficacy against either of these bacterial diseases. Genetic disease resistance presents an opportunity to improve the sustainability of Northeastern tomato production by offering a companion strategy to chemical control to limit tomato losses from plant disease and to help reduce the need for protective sprays.

Effective genetic resistances to bacterial speck and bacterial spot were previously mapped in the tomato genome by other researchers (see Chapter 1). Their work enabled us to use contemporary molecular breeding tools to pyramid these resistances into our best-performing fungal and oomycete resistant fresh market inbreds. Specifically, we used low-cost marker-assisted backcrossing with background genome selection to efficiently transfer bacterial resistances from processing tomatoes into our commercially relevant fresh market background. The resulting near-isogenic breeding lines facilitated studies of the individual and combinatorial effects of the resistance introgressions on disease resistance and fresh market horticultural type. Background selection then became something of a theme throughout this

work, being used repeatedly to generate novel breeding materials that became the focus of study in this dissertation and the basis for our recent inbred releases.

While molecular markers for bacterial speck and spot resistance were available at the start of this work, there were no such markers for early blight resistance. To perform genotypic selection for early blight resistance, we first had to map resistance loci in the tomato genome (see Chapter 2). To our surprise, we found loci associated with early blight resistance in both the Cornell University fresh market and The Ohio State processing backgrounds, and we used this information to breed new tomatoes with enhanced resistance in both market classes. Our breeding efforts afforded multiple opportunities to validate the resistance locus effects and markers in different growing environments. Ultimately, we were intrigued to find that much of the early blight resistance in our breeding populations could be explained by just two or three resistance loci, and we felt these valuable loci merited further study and finer mapping.

Traditional fine-mapping of quantitative trait loci is time consuming and expensive, prompting us to pursue an alternative fine-mapping-by-sequencing approach in Chapter 3. To do this, we researched the 80+ year history of early blight breeding efforts and traced our resistance QTL back to several landmark tomato breeding lines that we then sequenced. By comparing the genome sequences of these landmark lines, we could determine their shared ancestral resistance introgressions and use the underlying sequence to fine-map the resistance loci, to confirm the putative ancestry of the QTL, to predict the presence or absence of early blight resistance in hundreds of re-sequenced accessions, and to report dozens of molecular markers that breeders can use to select for these valuable resistance loci. We created a novel introgression detection and delineation method to facilitate this work, which is both highly scalable and flexible. The scripts needed to adapt this analysis to your own fine-mapping-by-sequencing efforts can be found in the supplemental materials for Chapter 3.

CHAPTER 1

DEVELOPMENT OF TOMATO LINES WITH RESISTANCE TO BACTERIAL SPOT AND BACTERIAL SPECK USING LOW-COST MARKER-ASSISTED BACKCROSSING

Anderson TA, Bernal E, De Jong DM, Francis DM, and MA Mutschler

Affiliations:

Anderson, TA, De Jong, DM and MA Mutschler, Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, 240 Emerson Hall, Ithaca, NY 14853

Bernal E, Francis DM, Department of Horticulture and Crop Science, The Ohio State University, 1680 Madison Avenue, Wooster, OH 44691

Acknowledgements:

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by TAA, EB, DD, and DMF. The first draft of the manuscript was written by TAA and all authors edited, revised, and approved the final version of the manuscript. Thanks to Dr. Michael Gore, Dr. Christine Smart, and Dr. Margaret Smith of Cornell University for critically reviewing the manuscript.

Abstract

We used marker-assisted backcrossing with background genome selection to rapidly transfer disease resistances to bacterial speck (*Pseudomonas syringae* pv. *tomato*) and bacterial spot (*Xanthamonas spp.*) across tomato market classes. Resistance haplotypes on chromosomes 5 (*Pto*|*Rx-3*) and 11 (*QTL-11*|*Xv-3*) were introgressed from two processing tomato inbreds into a determine fresh market background. To minimize costs, we made breeding selections with as few as two or three markers per chromosome. *Post-hoc* denser genotyping of breeding selections revealed cryptic donor introgressions lurking in their background genomes, but these introgressions were small and did not undercut the effectiveness of background selection. For example, dense genotyping of a BC_2F_2 selection showed its recurrent parent genome content was similar to a typical BC_4. All fresh market lines demonstrated strong multi-race bacterial spot disease resistance, but those with added major gene resistances showed less disease from races inducing hypersensitive response (HR). A broad-spectrum *QTL-11* from 'Hawaii 7998' presented no additional resistance to bacterial spot races not inducing HR. Three horticultural trials found modest evidence that resistance introgressions were associated with differences in fresh market tomato quality metrics including fruit area, color, shape, uniformity, and pericarp thickness. Addition of *QTL-11*|*Xv-3* consistently increased early season and total marketable yields, suggesting greater potential for trait improvement through cross-market breeding. Our work offers practical insights for breeders wishing to exploit these bacterial resistances and presents a case study in the use of genotypic selection with sparse genetic data to efficiently access beneficial alleles across market classes.

CHAPTER 1 SUPPLEMENTAL MATERIALS

Supplement 1 Analysis of variance (ANOVA) Tables S1-S4 for disease resistance and horticultural performance experiments

Introduction

Tomato (*Solanum lycopersicum* L.) yield losses from bacterial disease are increasingly problematic for growers in the Northern and Eastern United States, as growing conditions have grown warmer and wetter (DeGaetano 2009; Qian et al 2012), as bacterial races have proliferated (Stall et al. 2009), and as chemical control has proven increasingly ineffective (Liao et al. 2019; Louws et al. 2001; Marco and Stall 1983). Common bacterial diseases of tomato include bacterial spot and bacterial speck, which cause lesions on tomato foliage and fruit that result in defoliation and losses in marketable fruits. Fortunately, there exist readily deployable genetic resistances to both diseases in tomato.

Bacterial speck is caused by *Pseudomonas syringae* pv. *tomato*. The pathogen is widespread and persistent, with losses to marketable yield exceeding 25% under cool, wet conditions (Gitaitis et al. 1985; Jardine et al. 1988; Jones 1991; Schneider et al. 1977; Young and Wilkie 1986). Two races of bacterial speck, races 0-1, have been described. Since the detection of race 1 in 1986, it has become the predominant race worldwide (Arredondo and Davis 2000; Cai et al. 2011; Kunkeaw et al. 2010; Lawton and MacNeill 1986). The *Pto* locus from *S. pimpinellifolium* PI 370093 works in concert with *Prf* to confer *AvrPto* and *AvrPtoB*-mediated hypersensitive response (HR) to bacterial speck race 0 (Ronald et al. 1992). Race 1 isolates are, by definition, pathogenic on tomatoes expressing *Pto*/*Prf*. However, recent T1 isolates from Eastern and Northeastern tomato fields were either race 0 or exhibited virulence intermediate to typical race 0 and race 1 due to host recognition of the *AvrPto* effector (Kraus et al. 2017; Veluchamy et al. 2014). Thus, there is a benefit to deploying *Pto* in the Northeastern fresh market class. The *Ptr1* locus from *Solanum lycopersicoides* grants HR to race 1 but has yet to be introgressed into elite genetic backgrounds (Mazo-Molina et al. 2019).

Bacterial spot is caused by a species complex that includes *Xanthomonas perforans, X. vesicatoria, X. euvesicatoria* and *X. hortorum* pv. *gardneri* (previously *X. gardneri*) (Morinière et al. 2020, Potnis et al. 2015; Stall et al. 2009; Timilsina et al. 2020). At least four bacterial spot races (T1, T2, T3, and T4) have been isolated from field-grown tomatoes. Race T1 is considered *X. euvesicatoria*, while races T3 and T4 are *X. perforans* (Scott et al. 1995; Scott et al. 2001). Race T2 is either *X. vesicatoria* or *X. hortorum* pv. *gardneri*, as there is no T2 differential germplasm (Bernal et al. 2020; Sharlach et al. 2013). Shifts in bacterial spot populations have been observed in recent decades throughout the United States. Race T1 predominated prior to 1990 in Midwestern and Southeastern states (Jones et al., 1986; Scott et al. 1997). In Florida, T1 was displaced by race T3 in the 1990s and was recently displaced again by race T4 (Timilsina et al. 2015). In North Carolina, race T4 is most common (Adhikari et al. 2019). Midwestern states tend to have mixed populations. In 2012, *X. hortorum* pv. *gardneri* (race unknown) predominated in Ohio and Michigan, followed by *X. perforans* race T3 and *X. evesicatoria* T1 (Ma et al. 2015). Recent surveys from 2017 to 2019 in Indiana and Ohio have shown mixed populations of *X. perforans* T4 and *X. hortorum* pv. *gardneri* (Bernal et al. *in* prep). In New York state, only *X. hortorum* pv. *gardneri* has been detected (Pers. comm. Chris Smart). Therefore, there is a need to monitor complex and evolving populations across diverse environments.

Genetic resistance to bacterial spot in tomato is gained through major gene resistances and quantitative trait loci (QTL). The *Rx-3* locus from 'Hawaii 7998' confers *avrRxv*-mediated HR to bacterial spot race T1 (Scott et al. 1995; Whalen et al. 1993; Yang et al. 2005; Yang and Francis 2005; Yu et al. 1995). The *Xv3* (syn. *Rx-4*) locus from 'Hawaii 7981' or PI 128216 offers *avrXv3*-mediated HR to race T3 (Pei et al. 2012; Robbins et al. 2009; Scott et al. 1996; Scott et al. 2001; Wang et al. 2011). Similarly, *RXopJ4* (Syn. *Xv4*) from *S. pennellii* LA0716 confers HR to race T4. Multi-race resistance is conferred through QTL, including the

centromeric *QTL-11* from 'Hawaii 7998' (Sim et al. 2015), PI 114490 (Hutton et al. 2010; Yang et al 2005), and LA2553 (Liabeuf 2016). A recent study of processing tomatoes near-isogenic for *QTL-11* from these three sources found that the *QTL-11* from 'Hawaii 7998' in combination with *Rx-4/Xv3* gave the greatest multi-race *Xanthamonas* field resistance (Bernal et al. 2020).

In response to the ephemerality of pathogenic bacterial speck and bacterial spot races, a strategy for deploying a combination of major gene resistances and less-penetrant, broader-spectrum resistance QTL was proposed (Poland et al. 2009; Sim et al. 2015). Towards this goal, processing tomato breeding lines were created with coupling phase genetic linkages that simplify the inheritance of bacterial disease resistance. One of these coupling linkages was between *Pto* and *Rx-3* on chromosome 5, hereto referred to as *Pto*|*Rx-3* (Yang and Francis 2005). The second was between *QTL-11* and *Xv3* on chromosome 11, hereto referred to as *QTL-11*|*Xv3* (Bernal et al. 2020; Sim et al. 2015). We used marker-assisted backcrossing with background genome selection (Hospital et al. 2009) to rapidly transfer these two coupling-phase resistance linkages into the fresh market tomato market class. Here, we present a novel investigation into the effectiveness of this backcrossing strategy that is now relatively common in the tomato breeding literature. Specifically, we determined whether background selection based on sparse genotypic data could generate truly near-isogenic tomato lines. Finally, we determined the resistance effect of these introgressions against four *Xanthamonas* isolates causing bacterial spot and investigated whether these introgressions presented tradeoffs with horticultural performance in the fresh market tomato background.

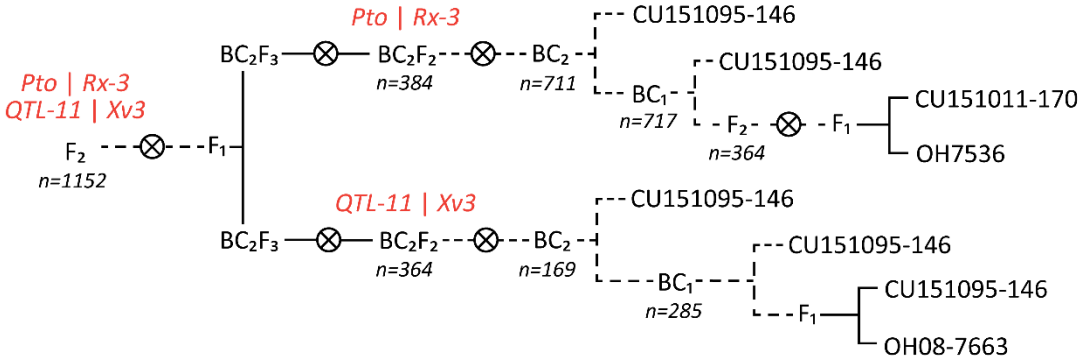
Materials and Methods

Starting plant materials. Tomatoes from the Cornell fresh market and The Ohio state processing tomato breeding programs were used as parents in a marker-assisted backcross gene pyramiding scheme. The processing tomato inbred 'Ohio 7536' (OH7536), whose creation is described in Yang and Francis (2005), has a coupling phase linkage between *Pto* from *S. pimpinellifolium* PI 370093 and *Rx-3* from 'Hawaii 7998' (*Pto*|*Rx-3*). The other resistance donor was the processing tomato 'Ohio 08-7663' (OH08-7663), which has a coupling linkage between *Xv4* from PI 128216 and *QTL-11* from 'Hawaii 7998' (*QTL-11*|*Xv3*) described in Sim et al. (2015). The recurrent background was the Northeast-adapted fresh market tomato Cornell 151095-146 (CU151095-146) and the closely related (>95% IBD) breeding line CU151011-170.

Near-isogenic line development. The details of our backcross pyramiding scheme and the population sizes used for genotypic selection can be found in Fig. 1. We separately introgressed the *Pto*|*Rx-3* and *QTL-11*|*Xv3* haplotypes into the fresh market background. Two backcrosses to CU151095-146 and a self-pollination step generated BC_2F_2 mono-introgression lines homozygous for the resistance haplotypes. These near-isogenic lines (NIL) were intermated and self-pollinated to create di-introgression NILs with the full complement of resistance genes. Plants were grown to maturity in a greenhouse with inline fertigation and 14-hour days held between 22 and 28°C. Seedlings were grown in 4-inch pots and selections were transplanted into 5-gallon plastic pots (Griffin Horticultural, MA).

Marker-assisted selection. We used single-nucleotide polymorphisms (SNPs) from the SolCAP panel (Sim et al. 2012) to select BC_1 or BC_2 plants heterozygous for the resistance loci (foreground selection), at or above the 75th percentile for identity by descent (IBD) to the

Fig. 1 Introgression lines resistant to bacterial speck and spot were created by separately introgressing the coupling phase resistance linkages *Pto*|*Rx-3* (OH7536) and *QTL-11*|*Xv3* (OH08-7663) into the REC-P background of CU151095-146 and its closely related sibling CU151011-170. Near-isogenic mono-introgression lines generated at the BC_2F_2 generation were intermated to create an F_2 population segregating for bacterial resistance loci in a uniform genetic background. Dashed lines between generations indicate genotypic selection, with genotyped population sizes shown below



recurrent parent genetic background, and with the greatest consolidation of heterozygous haplotypes on the fewest chromosomes (background selection). Foreground selection was done prior to background genotyping using the CAPS/dCAPS markers solcap_snp_sl_50870 and solcap_snp_sl_194 to broadly flank the *Pto*|*Rx-3* haplotype (chromosome 5 positions SL3.0ch05:6.3 - 64.7 Mbp, SL4.0ch05:6.3 - 63.2 Mbp) and solcap_snp_sl_4524 and solcap_snp_sl_2677 to select for the *QTL-11*|*Xv3* haplotype (chromosome 11 positions SL3.0ch11:8.8 - 55.2 Mbp, SL4.0ch11:8.9 - 52.9 Mbp). Protocol details, including primer sequences, are described in Chapter 2 and Supplement 3.

Early BC_1 and F_2 generations required genome-wide marker coverage. We aimed to use four SolCAP SNPs per chromosome for background selection. We will use "breeding marker density", or BMD, to refer to this relatively low density of markers used for breeding and genotypic selection. BMD markers were chosen based on prior polymorphic information content (Blanca et al 2015) and even spacing on the genetic distance map, with physical distance used as a backup criterion. However, some chromosomes lacked four well-spaced polymorphic markers, so fewer markers were used on those chromosomes. Markers in non-segregating regions of the genome were excluded in subsequent generations to reduce costs. Markers for background selection were typed using Kompetitive allele specific PCR (LGC, Teddington UK) or amplicon sequencing (Agriplex Genomics, Cleveland OH). Plants from the intermated F_2 population were genotypically selected and grouped based on homozygosity for resistance-linked CAPs/dCAPs markers prior to transplanting in the field.

Re-genotyping of selections. A BC_1 plant from the *Pto*|*Rx-3* pedigree and a BC_2F_2 plant from the *QTL-11*|*Xv3* pedigree were genotyped at higher depth using genotyping-by-sequencing (GBS). A multiplexed, barcoded sequencing library of the *QTL-11*|*Xv3* selection was prepared with the PstI/MspI enzyme combination and sequenced to approximately 0.4x depth

on a NextSeq500 sequencer (Poland et al. 2012). The pooled library for the BC_1 *Pto*|*Rx-3* tomato was similarly prepared, except the ApeKI enzyme was used, and sequencing was to 0.7x depth. SNP variant identification and filtration was done with the GBSv2 pipeline in Tassel 5.0 using filtering thresholds and error correction to reduce spurious over-calling of heterozygotes, as described in Chapter 2 (Bradbury et al. 2007; Elshire et al. 2011; Furuta et al. 2017). The same BC_2F_2 selection with *QTL-11*|*Xv3* was genotyped by amplicon sequencing for an optimized panel of SolCAP SNPs (Agriplex Genomics, OH).

Inoculated disease trials. Inoculated field trials were done in the summer of 2018 and 2019 using infiltration protocols described in Bernal et al. 2020. In 2018 trials were independently inoculated with *Xanthomonas* strains (*X. euvesicatoria* [race T1 *Xcv1*10c and *Xcv*767], *X. perforans* [race T3 *Xcv*761], or *X. hortorum pv. gardneri* [SM775-12 and SM605-11]. In 2019, trials were inoculated with either *X. euvesicatoria* T1 or *X. perforans* [race T4 SM1828-17]. Replicated randomized complete block designs for 2018 had four blocks for the *X. perforans* T3 trial and two blocks for the *X. euvesicatoria* T1 and *X. hortorum* pv. *gardneri* trials. The 2018 design included NILs, parental controls, and the susceptible controls OH88119 and OH8556. The 2018 *X. euvesicatoria* and *X. hortorum* pv. *gardneri* trials were located at the Ohio State University's research farms in Wooster, OH, and the *X. perforans* T3 trial was conducted at the Ohio State University's North Central Agricultural Research Station (NCARS) in Fremont, OH. In 2019, *X. euvesicatoria* T1 and *X. perforans* race T4 trials were conducted in Wooster, Ohio and NCARS, respectively. The 2019 experimental designs included NILs, hybrids, parental controls, and OH88119. Plants were visually scored for disease at the mature green (early) and ripe fruit (late) timepoints using the Horsfall-Barratt scale (Horsfall and Barratt 1945).

Horticultural trials. NILs from the *QTL-11*|*Xv3* introgression program were evaluated for horticultural phenotypes in a 2018 disease-free field trial at the Thompson Vegetable

Research Farm in Freeville, NY. Mono-introgression NILs from the *QTL-11*|*Xv3* and *Pto*|*Rx-3* introgression programs and the intermated F_2 generation (di-introgression lines) were evaluated separately in 2019. Experiments were randomized complete blocks with four replications of experimental families, parental controls, and commercial benchmarks. Field trials were managed as described in Chapter 2, except that plants were trellised by the Florida Weave method (http://go.ncsu.edu/readext?681800) in 2019.

A preliminary yield trial was harvested twice, two weeks apart, in 2018. An exhaustive harvest trial was done in 2019, with weekly harvests of fruits at the red-ripe stage for 6 weeks, until >90% of fruit were removed. Fruits from each plot were sorted for marketability, and the resulting groups were counted and weighed. Six to twelve fruits per plot were chosen for imaging by sorting the fruits from a plot by size and selecting a representative subset. Latitudinal cross-sections were imaged on an Epson Perfection V550 flatbed scanner and analyzed using Tomato Fruit Analyzer software (Brewer et al. 2006; Darrigues et al. 2008; Gonzalo et al. 2009) on a PC running Windows 10.

Data analysis. Trial data from each field experiment were modeled independently using Analysis of Variance (ANOVA) in R version 3.6.0 (R Core Team 2017) for Windows 10. Disease ratings at the maximal timepoint were modeled for each isolate as a function of the fixed effects for introgression and block. Horsfall disease data were analyzed as a linear dependent variable following the convention of Bernal et al. 2020. TFA ontology phenotypes were independently modeled as a function of introgression, harvest date, introgression x harvest date interaction, and block. We considered *early yield* harvest data to be the first of two harvests in 2018, or the sum of the first 3 of 6 harvests in 2019. The sum of harvest data across all harvest dates constituted *total yield*. The *percent marketable yield* (the percentage of all fruit that passed quality grading) and the *Average Marketable Fruit Weight* (the average

weight of an individual marketable tomato) were weighted by the proportion of total yield on each harvest date. These harvest phenotypes were modeled as a function of the fixed effects for introgression and block. We used Type III sums of squares because all data except for the 2019 F_2 dataset were unbalanced. Post-hoc tests were protected lsmeans as calculated from emmeans with compact letter groupings by multcompView (Graves et al. 2019). Model residuals were inspected to ensure normality, homoscedasticity, and the absence of outliers. To calculate estimated marginal means for tomato fruit color and shape variables across all three experiments, we fit a mixed linear model with a fixed introgression effect and random intercept estimates for family, family x experiment, harvest date, and family x harvest date with lme4 (Bates et al. 2015) and emmeans (Length 2017). Principal components were calculated from marginal means using bpca (Faria and Demétrio 2019).

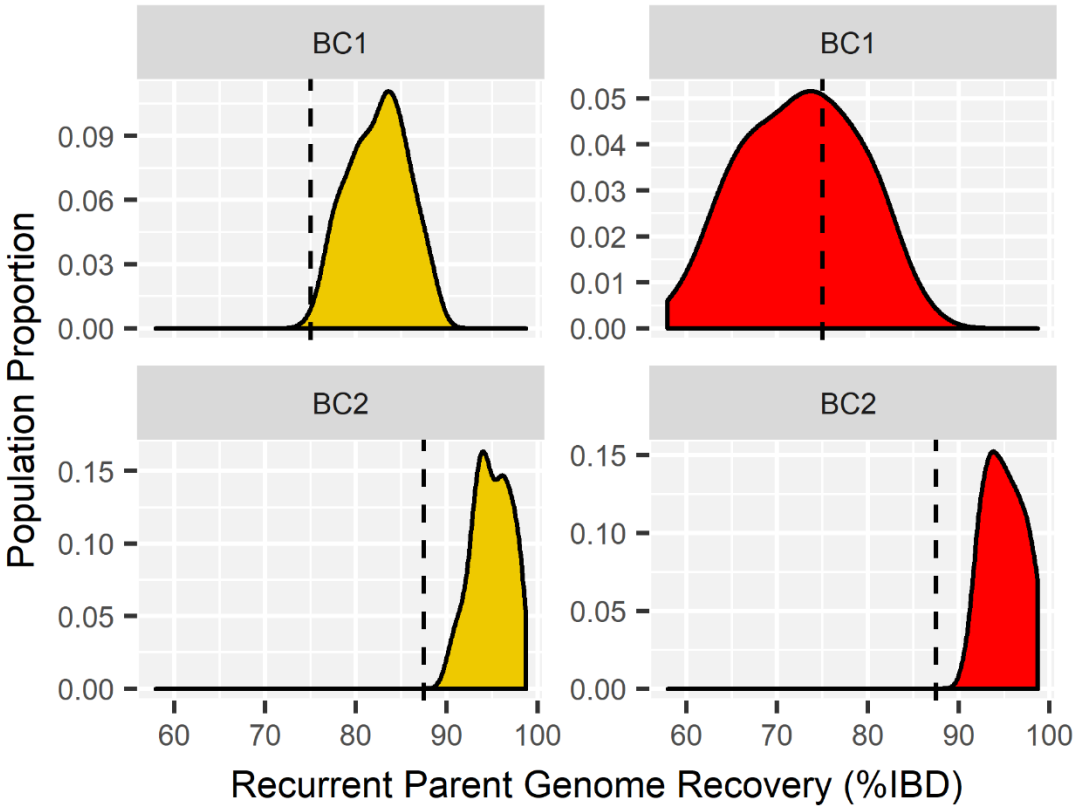
Results

Marker-assisted selection enabled rapid recovery of the recurrent parent genome

Background genome selection allowed us to generate near-isogenic lines in as few as four generations, including two backcross generations (Fig. 1). We measured the relative efficiency of our introgression program by comparing the marker-estimated percentage of the recurrent parent genetic background (% IBD) in our selected tomatoes against the expected % IBD for a given backcross generation without background selection. The % IBD distributions for the BC_1 and BC_2 generations from our parallel introgression programs are shown in Fig. 2. The selected top ~1% of BC_1 (n=3) plants in the *QTL-11*|*Xv3* introgression program averaged 85.0 ± 1.4% IBD, similar to the expected value of 87.5% IBD for an unselected BC_2 population. After backcrossing these selections for a second time, the top ~2% of their BC_2 progeny (n=4) averaged 98.7 ± 0.0% IBD[1], similar to an unselected BC_4 (96.9% IBD) or BC_5 (98.4% IBD). All of our BC_2 selections were heterozygous for one off-target marker in the background genome, which we subsequently selected against in the BC_2F_2 population. For the *pto*|*Rx-3* introgression program, which began by backcrossing an advanced F_2 selection rather than an F_1, we observed a BC_1 distribution for the entire population (n=717) centered around 82.5 ± 3.2% IBD, while the selected top ~1% of plants (n=7) averaged 87.8 ± 0.6% IBD, equivalent to a BC_2. After backcrossing these selections once more, the top ~1% of BC_2 plants (n=4) averaged 98.0 ± 0.1% IBD, between the expected %IBD for a BC_4 or BC_5. These BC_2 individuals were still heterozygous for one or two markers in the background genome that were subsequently selected against in the BC_2F_2 generation. Overall, background genome selection yielded

[1] All selections had the same %IBD by our markers

Fig. 2 Recovery of recurrent parent genome in BC_1 and BC_2 generations of two mono-introgression programs. **Left panel**: breeding progress in the *pto*|*Rx-3* introgression program. **Right panel**: breeding progress during the *QTL-11*|*Xv3* introgression program. Dashed vertical lines show the expected % recovery in theoretical populations without BGS. In the case of the *pto*|*Rx-3* introgression program, an F_2 selection 60% IBD was backcrossed, resulting in upward skew



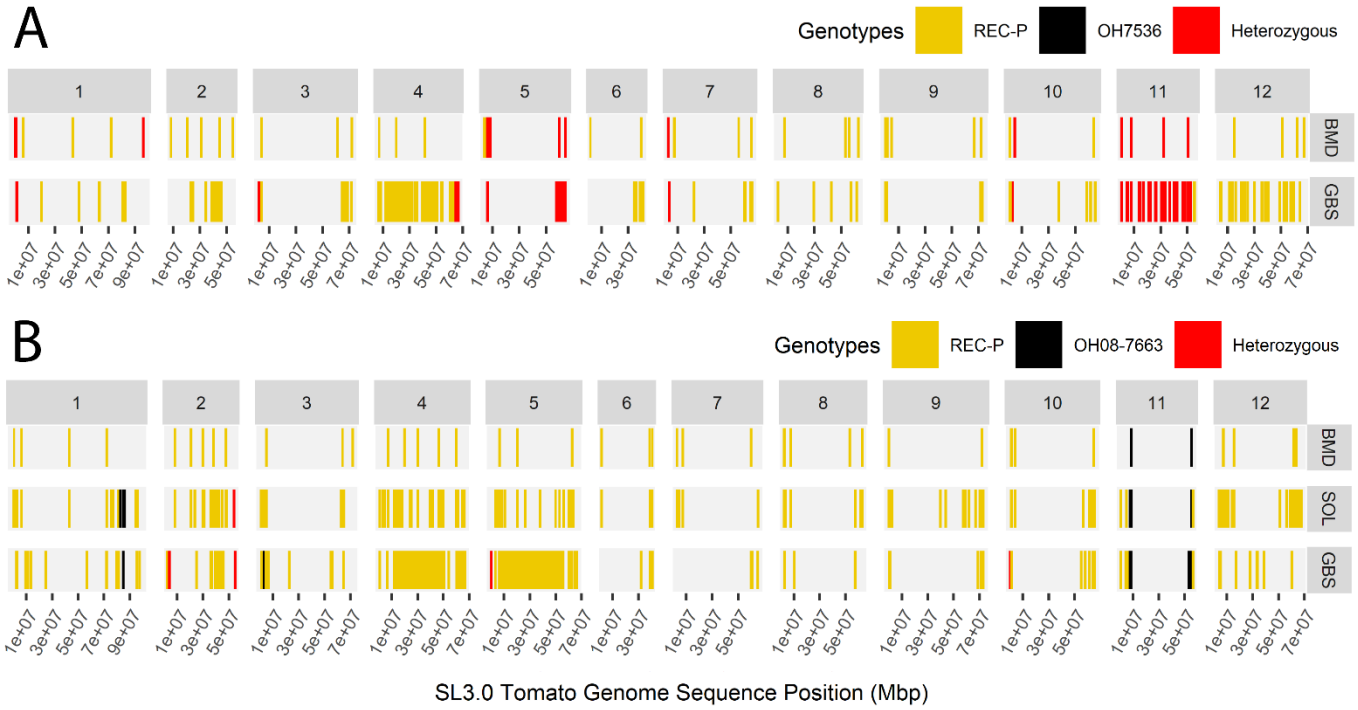
recurrent genome recovery advantages at a rate of at least two times the number of backcrosses done.

Genome comparisons reveal cryptic introgressions

To determine whether using low-density molecular markers for background selection resulted in cryptic and unwanted donor introgressions in our breeding selections, a breeding selection from each of the parallel introgression programs was re-genotyped to a relatively high-density, enabling comparisons to BMD genotypes (Fig. 3). A BC_1 individual from the *Pto*|*Rx-3* introgression program was re-genotyped using GBS, yielding 725 high-quality SNP markers across the SL3.0 tomato genome sequence (Tomato Genome Consortium 2012). Similarly, GBS genotyping of a BC_2F_2 selection from the *QTL-11*|*Xv3* program yielded 420 genome-wide SNPs. The same *QTL-11*|*Xv3* BC_2F_2 selection was also genotyped at sites from an optimized SolCAP marker panel (SOL), yielding 195 polymorphic markers. While GBS identified more polymorphic loci than the SOL panel, many of the additional SNPs were clustered in genomic regions of high polymorphism, such as on chromosomes 4 and 5. In contrast with the SOL panel, GBS failed to detect any variants on the upper arms of chromosomes 6 and 7. GBS genotypes offered moderately better coverage of centromeric and pericentromeric chromosomal regions compared to the SNPs in the SOL panel.

While markers at BMD correctly predicted much of the genomes of the breeding selections, higher-density genotyping revealed cryptic introgressions in both pedigrees that were not observable at BMD. In the case of the BC_1 selection from the *Pto*|*Rx-3* program, GBS genotypes confirmed heterozygosity for the chromosome 5 *Pto*|*Rx-3* resistance haplotype, but also uncovered undesired heterozygosity on chromosomes 3 (<4.5 Mbp) and 4 (<0.5 Mbp) that were not visible at BMD (Fig. 3). In total, the heterozygous portions of the background genome

Fig. 3 Genome compositions of two tomato lines from the breeding program as determined using different marker types and densities. Genotypes were either homozygous for the recurrent fresh market parent alleles (REC-P), heterozygous, or homozygous for either of the processing tomato donor parent alleles (OH7536 and OH08-7663). **Panel A:** A BC_1 plant from the *pto*|*Rx-3* introgression program was genotypically selected using a small number of markers (n = 30[2]), referred to as breeding marker density (BMD), and then re-genotyped at higher marker density (n = 420) using GBS. **Panel B:** A BC_2F_2 selection from the *QTL-11*|*Xv3* introgression program was similarly selected using BMD marker density (n = 41) and subsequently genotyped to higher SNP density for the SolCAP SNP panel (SOL) (n = 195) and by GBS (n = 725)



[2] More than 30 SNPs are shown in Panel A – BMD because homozygous markers typed at the F_2 generation are known to be heterozygous after a backcross, and these markers are plotted

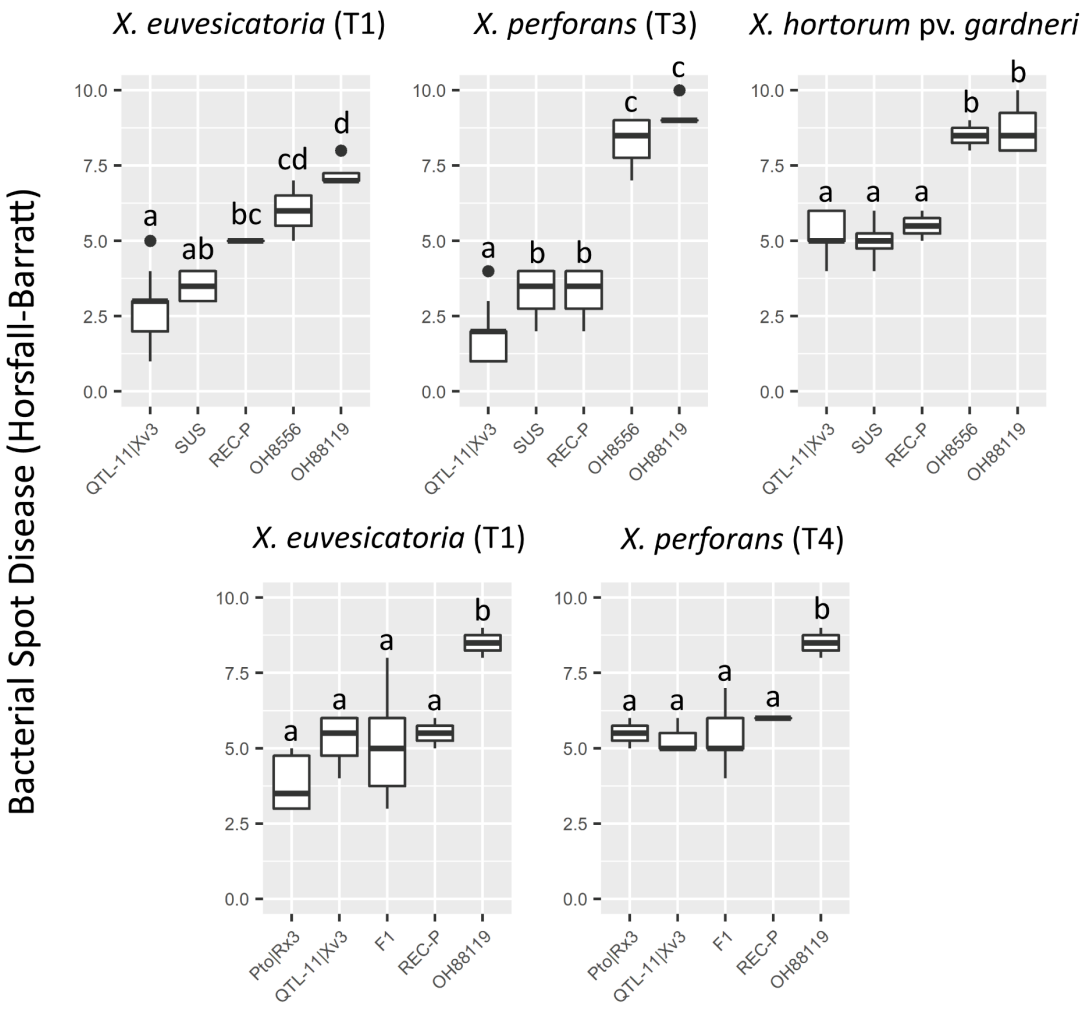
for this selection spanned an estimated physical distance of 65.0 - 138.1 Mbp[3]. Using this high-density marker information, we calculated a revised 90.9 – 95.7% IBD for the genetic background of our BC_1 selection, making it approximately equivalent to a traditional BC_2 or BC_3 obtained without background selection. Genotyping at SOL sites for the BC_2F_2 *QTL-11*|*Xv3* selection confirmed successful introgression of the chromosome 11 *QTL-11*|*Xv3* resistance haplotype and revealed a homozygous cryptic introgression on chromosome 1 and a stretch of residual heterozygosity on chromosome 2 that was previously undetected. GBS of this same entry revealed another homozygous introgression on chromosome 3, and three more putative regions of heterozygosity on chromosomes 2, 5, and 10 that were undetectable at BMD and SOL marker density. In total, these off-target introgressions spanned an estimated physical distance of 21.5 - 39.3 Mbp of the sampled BC_2F_2 genome. A majority (57.2 – 98.0 %) of this off-target distance was heterozygous. Using this high-density GBS marker information, we calculated a revised 96.4 – 98.6% IBD for the genetic background of our BC_2F_2 selection, making the selection approximately equivalent to a BC_4 without background selection. Thus, high density genotyping did not undercut the expectation of accelerated background genome recovery in either the BC_1 or $BC_2 2F_2$ selections.

Resistance effect of introgressions on bacterial disease

Every fresh market breeding line we tested was more resistant to bacterial spot disease than the susceptible processing tomato control OH88119, regardless of the *Xanthamonas* isolate ($p < .05$; Fig. 4). However, there were differences in bacterial spot disease in 2018 between NILs homozygous for the newly introgressed *QTL-11*|*Xv3* haplotype, their sibling NILs selected for the absence of the resistance haplotype, and the recurrent parent line

[3] A range of reflects that fact that we do not know the recombination site in the interval between two markers

Fig. 4 Incidence of bacterial spot disease caused by five different pathogen isolates revealing the resistance effect of added coupling phase resistance linkages in the fresh market tomato background. Inoculated field experiments in 2018 were against *X. euvesicatoria* T1, *X. perforans* T3, and *X. gardneri* (top panel) and included the *QTL-11*|*Xv3* mono-introgression lines, near-isogenic siblings selected for the absence of the *QTL-11*|*Xv3* resistance haplotype, the recurrent parent fresh market tomato line (REC-P), and two susceptible processing tomato controls, OH8556 and OH88119. Inoculated trials in 2019 (bottom panel) were against *X. euvesicatora* T1 and *X. perforans* T4, and included the (-)*QTL-11*|*Xv3* and *Pto*|*Rx-3* mono-introgression lines, the F_1 hybrid between these two mono-introgression lines, the recurrent parent line and the susceptible control OH88119. Field trials in both years were conducted in Wooster, OH



(Fig. 4; Table S1, Supplement 1). The added resistance from *QTL-11*|*Xv3* was most apparent against race T3 of *X. perforans*; lines with this haplotype were more resistant than the susceptible NILs (t_{56} = 4.10, p < .001) and the recurrent parent (t_{56} = 3.03, p = .004). Since the *Rx-3* locus was not present in the *QTL-11*|*Xv3* lines, any resistance to *X. euvesicatoria* race T1 beyond that of the recurrent parent would presumably have been due to partial resistance from *QTL-11*. Our data suggested enhanced resistance in NILs with *QTL-11*|*Xv3* against *X. euvesicatoria* T1 in the contrast with the recurrent parent (t_{26} = 3.56, p = .002), but the result was only marginally statistically different in the contrast with sibling NILs lacking the *QTL-11*|*Xv3* introgression (t_{26} = 1.81, p = .083). We observed little evidence of increased resistance to *X. hortorum* pv. *gardneri* (no defined race) in *QTL-11*|*Xv3* lines compared to either the recurrent parent (t_{26} = 0.50, p = .630) or to the susceptible NILs (t_{26} = 0.66, p = .514).

Inoculated field trials in 2019 enabled tests of the recently completed *Pto*|*Rx-3* mono-introgression lines alongside *QTL-11*|*Xv3* lines and F_1 hybrids that were heterozygous for the *Pto*|*Rx-3* and *QTL-11*|*Xv3* resistance haplotypes (Table S1, Supplement 1). These trials were smaller than those in 2018, resulting in lower statistical power for resolving differences in disease resistance relative to the recurrent parent. There was little evidence that lines homozygous for either the *Pto*|*Rx-3* or *QTL-11*|*Xv3* resistance introgressions were more resistant to *X. perforans* race T4 compared to the recurrent parent. In contrast, we can be modestly confident that lines with the *Pto*|*Rx-3* introgression had greater resistance to *X. euvesicatoria* T1 than the recurrent parent (t_{20} = 1.48, p = .154). There was little evidence for added resistance to either bacterial spot isolate when both haplotypes were heterozygous in the F_1. Thus, it appears that the *Rx-3* hypersensitive resistance effect against *X. euvesicatoria* T1 was at least partially recessive in our fresh market tomato background.

Impact of introgressions on horticultural type

We conducted three separate uninoculated field trials over the course of two years to investigate the potential impact of the resistance haplotypes on fresh market horticultural type. The first trial in 2018 (Table 1, experiment A; Table S2, Supplement 1) investigated the *QTL-11*|*Xv3* NILs in comparisons to each other and to the recurrent parent. A second horticultural trial, conducted in 2019, added the newly created lines homozygous for the *Pto*|*Rx-3* introgression and the F_1 hybrid between the *QTL-11*|*Xv3* and *Pto*|*Rx-3* mono-introgression lines (Table 1, experiment B; Table S3, Supplement 1). A third trial, also conducted in 2019, investigated the individual and combined effects of the *QTL-11*|*Xv3* and *Pto*|*Rx-3* resistance introgressions in inbreds (Table 1, experiment C; Table S4, Supplement 1).

Analysis of harvest data from the three separate field experiments revealed differences in the horticultural performance of our introgression lines and hybrids that are likely due to the introgressions carrying the *QTL-11*|*Xv3* and *Pto*|*Rx-3* resistances. Lines homozygous for the *QTL-11*|*Xv3* introgression consistently exhibited the greatest early-season marketable yields relative to the recurrent parent, with statistically significant effects ($p < .05$) in experiments A and B. Lines homozygous for the *Pto*|*Rx-3* introgression had the lowest early season yields in both experiments and the result was marginally different from the recurrent parent in experiment B ($t_{45} = 1.60$, $p = .116$). While we observed similar trends in early-season marketable yield for lines with *QTL-11*|*Xv3* and *Pto*|*Rx-3* in experiment C, this comparatively small experiment lacked the statistical power needed to confidently differentiate the introgression lines for this trait. Homozygosity for the *QTL-11*|*Xv3* introgression also increased the total marketable yield relative to the recurrent fresh market parent in experiments A and B. While there appeared to be a tradeoff between total marketable yield and fruit weight for the *QTL-11*|*Xv3* lines in 2018 (experiment A), we did not observe this tradeoff in the following

year's horticultural trial (experiment B). Lines with the *Pto*|*Rx-3* introgression had higher percentages of marketable fruit relative to the recurrent parent and to the susceptible NILs in experiments B and C, respectively. Finally, the F_1 combination heterozygous for the *QTL-11*|*Xv3* and *Pto*|*Rx-3* introgressions had the greatest total marketable yield of any entry, had similarly high percentages of marketable fruits to the *Pto*|*Rx-3* parent, and had mid-parent early yields (experiment B).

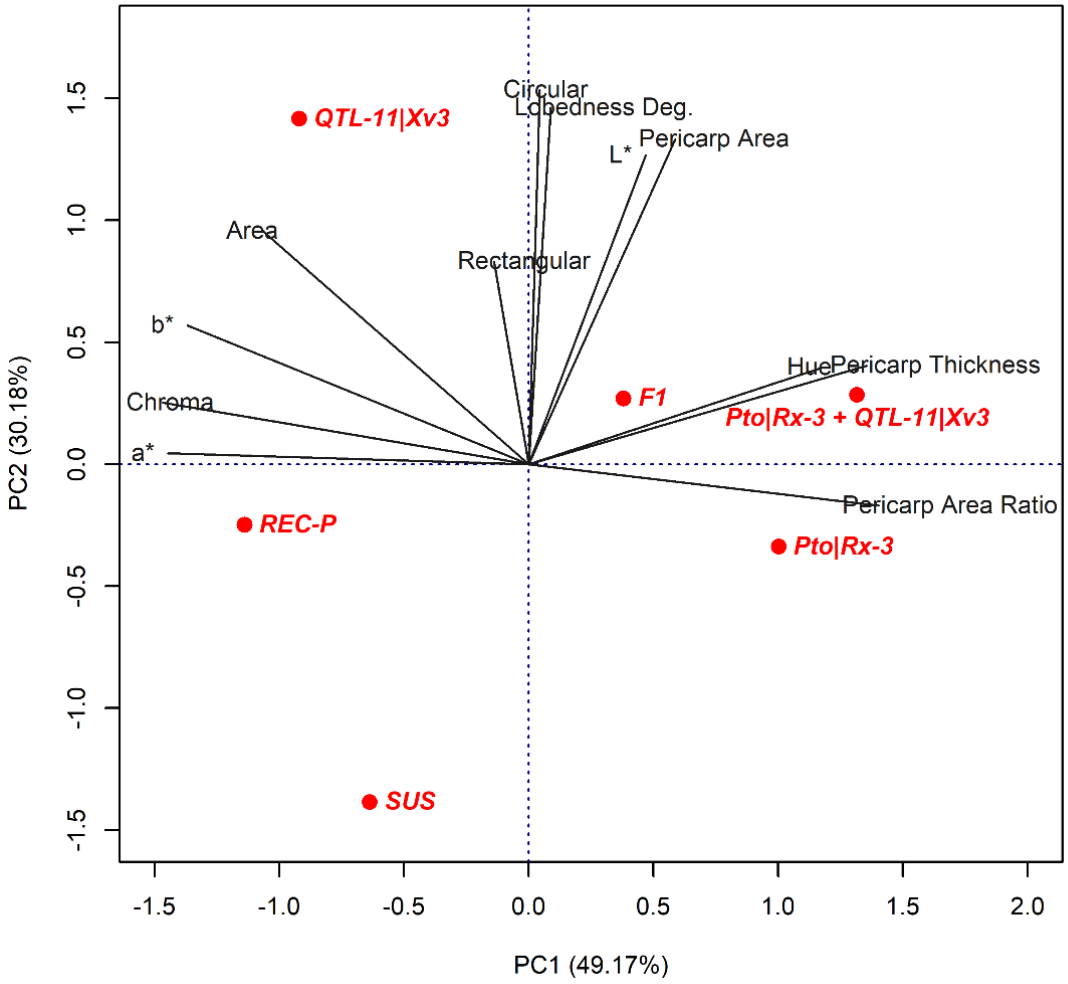
Multivariate trends in fruit color and shape variables were studied across all three experiments using principal component analysis on data from tomato fruit images. For context, we also present univariate analyses for an uncorrelated subset of the phenotypes in Table 1, revealing differences at the 95% confidence level across entries for cross-sectional fruit area (experiment B), the ratio of pericarp to total fruit area (experiment B), the degree of fruit shape irregularity (experiment A), and the blue-yellow fruit color metric *b** (experiment A). The first and second principal components of the PCA analysis explained 79.4% of the phenotypic variance for fruit shape and color-related ontology terms (Fig. 5). The primary sources of variation underlying PC1 were the fruit color metrics *a**, *Chroma*, and *Hue*, as well as *Tomato Pericarp Thickness* and *Tomato Pericarp Area Ratio*. PC2 was influenced by variation in the fruit shape uniformity measures *Circular* and *Lobedness Degree*, as well as *L** fruit color lightness. Projection of breeding lines onto this PCA space revealed that tomatoes from plants with the *Pto*|*Rx-3* introgression tended to have a smaller fruit area, thicker pericarps, duller Chroma, and a yellower Hue. In contrast, tomatoes from plants with the *QTL-11*|*Xv3* introgression were more likely to have larger and irregularly shaped fruits (larger values for *Circular* or *Lobedness Degree*) but were otherwise unchanged relative to the recurrent parent. Based on the centralized position of the F_1 in Fig. 5, changes in fruit color and shape were mid-

parent or partially dominant. Interestingly, the changes in imaged fruit area and pericarp thickness are not reflected in the table of average marketable fruit weights (Table 1).

Table 1 Effect of bacterial disease resistance introgressions on horticultural performance in three separate field experiments (labeled Experiments A-C, at left). Statistical effects were modeled independently for each experiment. **A**. The 2018 trial compared mono-introgression lines homozygous for the *QTL-11*|*Xv3* introgression (*QTL-11*|*Xv3*), a susceptible selection without the resistance introgression (SUS), and the recurrent parent line (REC-P). **B.** The 2019 trial compared lines homozygous for either the *Pto*|*Rx-3* or the *QTL-11*|*Xv3* introgressions, F_1 hybrids between these two types of lines that were heterozygous for both resistance introgressions (F_1), and the REC-P. **C**. A 2019 di-introgression trial of F_2 selections homozygous for either the *Pto*|*Rx-3* or *QTL-11*|*Xv3* resistance introgressions, homozygous for both resistance introgressions (*Pto*|*Rx-3* + *QTL-11*|*Xv3*), or homozygous for the susceptible recurrent parent genotypes at these loci (SUS). Means are bolded if the effect of introgression was significant at $p < .05$. Letter groupings show differences in entry means at the 95% confidence level

		Marketable Yield				Tomato Fruit Analyzer					
	Line/Hybrid Type within Experiment	Early (kg/plant)	Total (kg/plant)	% Marketable	Fruit Weight (g)	Area (cm^2)	Lobedness Degree	Pericarp Area Ratio	L*	a	b
A	*QTL-11*|*Xv3*	**1.73^{b}**	**4.93^{b}**	**47.3^{a}**	**162^{a}**	45.1	**1.38^{b}**	0.37	43.7	32.9	**27.1^{b}**
	SUS	**1.31^{ab}**	**5.34^{b}**	**55.4^{b}**	**155^{a}**	41.7	**1.23^{a}**	0.38	42.8	32.3	**26.3^{a}**
	REC-P	**0.98^{a}**	**3.60^{a}**	**54.0^{ab}**	**179^{b}**	43.6	**1.22^{a}**	0.36	43.4	32.7	**26.6^{ab}**
B	*Pto*|*Rx-3*	**0.75^{a}**	**5.92^{a}**	**79.6^{b}**	169	**37.0^{a}**	1.29	**0.42^{c}**	47.3	23.6	27.1
	QTL-11|*Xv3*	**1.69^{c}**	**6.92^{b}**	**67.7^{a}**	164	**44.6^{c}**	1.34	**0.38^{a}**	47.5	24.6	27.7
	F_1	**1.13^{b}**	**7.77^{c}**	**76.9^{b}**	169	**39.9^{b}**	1.34	**0.41^{b}**	46.6	24.6	27.2
	REC-P	**1.01^{ab}**	**5.32^{a}**	**67.9^{a}**	163	**40.9^{b}**	1.43	**0.38^{a}**	46.7	25.5	27.8
C	*Pto*|*Rx-3*	1.30	6.43	**64.6^{b}**	165	39.6	1.42	0.39	44.6	24.5	26.7
	QTL-11|*Xv3*	2.45	5.55	**50.3^{a}**	158	41.4	1.26	0.35	46.6	26.2	28.4
	Pto|*Rx-3 + QTL-11*|*Xv3*	2.42	6.91	**66.9^{b}**	162	40.1	1.31	0.39	47.0	23.8	27.1
	SUS	1.82	5.18	**44.1^{a}**	172	40.6	1.22	0.37	45.7	26.0	27.8

Fig. 5 PCA loadings biplot of fresh market breeding lines projected onto a PCA space defined by tomato fruit shape and color ontology measures. These data are extracted from images of latitudinally-cut tomato fruit. The contribution of the fruit color or shape variables to the first and second principal components are represented as named vectors. The recurrent parent (REC-P) and the susceptible genetic background controls that were selected for the recovery of the recurrent parent genome and the absence of the resistance haplotype (SUS) cluster in the bottom left corner of the PCA space. The effect of adding the *Pto|Rx-3* introgression can be visualized along the primary axis of variation, while the effect of adding the *QTL-11|Xv3* introgression is represented by the secondary axis of variation



Discussion

The pathogens underlying bacterial speck and bacterial spot disease have proven capable of rapid evolution and pathogenic race proliferation, making it challenging to create tomato varieties with durable, effective, and broad-spectrum resistance. One strategy to overcome this challenge is to pyramid highly penetrant resistance loci and broadly effective quantitative trait resistance loci. Genetic linkage can be leveraged to simplify this task by consolidating linked resistance loci into coupling phase linkages that increase the probability of co-inheritance and facilitate reductions in breeding population sizes. This approach was demonstrated in processing tomatoes, where the *QTL-11*|*Xv3* linkage in OH08-7663 gave a surprisingly high degree of resistance to the bacterial spot species X. *perforans* and *X. hortorum* pv. *gardneri* (Bernal et al. 2020). In this work, we sourced the same *QTL-11*|*Xv3* haplotype and a separate linkage of *Pto*|*Rx-3* from two processing tomato breeding lines, transferring these haplotypes into an elite fresh market tomato background. It was unknown how pyramiding these resistances from processing tomatoes together in a fresh market tomato might enhance the level of bacterial disease resistance or whether it would affect important fresh market horticultural traits such as fruit size and maturity. Notably, we also looked into the effectiveness of the now commonly used trait introgression methodology of background genome selection in tomato, specifically investigating a low-cost variant of this method that relies on sparse marker data for genotypic selection.

In practice, there is little crossing between cultivated market classes despite valuable genetic traits in each class. By framing the exchange of genetic material between market classes as a trait introgression problem, we were able to access bacterial resistance coupling linkages from processing tomato and rapidly recover fresh market type. Interestingly, we experienced little linkage drag in the resulting lines, despite incorporating large introgressions

that encompassed a majority of two chromosomes. This suggests there is greater potential to enhance horticultural performance and disease resistance through cross-market breeding populations. This hypothesis of cross-market breeding potential is further substantiated by our recent efforts to characterize early blight (*Alternaria linariae*) resistance in a fresh market breeding line. In that work, we serendipitously observed transgressive segregation for early blight resistance after intermating a resistant fresh market tomato and a largely susceptible OSU processing tomato, leveraging the novel resistance on chromosome 5 of processing tomato (*EB-5*) to bolster early blight resistance in the fresh market class (*see* Chapter 2). In a reciprocal backcross program, we also transferred early blight and late blight (*Phytophthora infestans*) resistance from chromosome 9 of fresh market tomato (*EB-9*|*Ph-3*) into the processing class, creating blight-resistant processing tomato lines that are now in testing.

We used marker-assisted backcrossing with background selection to rapidly transfer the two resistances haplotypes into our elite genetic background, demonstrating a simple, quick, and affordable approach to moving valuable haplotypes between market classes. Simulations (Hospital and Mulsant 1992) and real-world studies in tomato (Bernal et al. 2020) suggested that background selection is effective with few markers per chromosome, minimizing genotyping costs and simplifying the genotypic selection methodology. Our implementation of background selection aimed to place at least four markers per chromosome. However, a lack of polymorphism between the donor and recurrent parents further reduced the number of informative markers. Despite the low density of markers, we confirmed that background selection had accelerated the recovery of the recurrent parent genetic background in our breeding selections. Traditional programs often require four to six backcrosses; for tomato, each generation takes approximately 5 months. Since we achieved

equivalent levels of inbreeding with two backcrosses, we estimate that background selection saved us between 10 and 20 months of NIL development.

The use of low-density markers failed to detect several cryptic introgressions in our breeding selections that were observable with greater marker densities. However, these introgressions were small. Furthermore, when we used the higher-density genotypic information to calculate the proportion of the selected genomes identical-by-descent to the recurrent parent, the resulting values were similar to those observed at BMD. Thus, low-density markers were able to correctly identify individuals with greater recovery of the recurrent genome, despite missing cryptic donor introgressions. The importance of these cryptic introgressions depends on the donor's adaptability, in our case, a relatively well-adapted tomato line from another market class. Our low-density markers were more accurate in the BC_1 generation, failing to detect several cryptic introgressions in a BC_2F_2 line that had experienced additional rounds of meiosis. Most discrepancies in the genomic construction across marker systems occurred for the larger, lower-numbered chromosomes that recombine frequently (Tanksley et al. 1992). Therefore, a greater density of markers should be used when working with an unadapted resistance donor, when it is necessary to monitor cryptic introgressions on larger chromosomes, and as the number of breeding generations increases.

While *P. syringae* pv. *tomato* race 1 now predominates worldwide, *Pto*-mediated resistance may still be valuable for the Eastern United States. Race 0 isolates were recently isolated from tomatoes in western North Carolina (Veluchamy et al. 2014). Recent isolates of *P. syringae* pv. *tomato* from field-grown tomatoes in New York state indicate a geographically uniform population that produces the *AvrPto* effector but does not accumulate the *AvrPtoB* protein (Kraus et al. 2017). Interestingly, greenhouse tests showed these populations were intermediately pathogenic relative to typical race 0 and race 1 isolates on a *Pto* cultivar,

presumably because of additional virulence factors suppressing host immunity. Few fresh market tomato cultivars possess *Pto*, while *Pto* is nearly universal among processing tomatoes (Kraus et al. 2017; Kunkeaw et al. 2009; Pedley and Martin 2003). Kraus and colleagues proposed that because the Northeast does not have a processing tomato industry, there may not have been strong selective pressure to eliminate, modify, or down-regulate the race 0 effectors. As a result, our fresh market lines with *Pto* resistance may help to reduce crop losses from bacterial speck in the Northeast. We eagerly anticipate the availability of race 1 bacterial speck resistance in a cultivated background, which will help to ensure protection from bacterial speck disease in the near future.

By transferring bacterial spot resistances into a modern large-fruited inbred background that was adapted to the Northeast, we could study multi-race resistance effects in a pathosystem more typical of a commercial breeding program, wherein breeding lines are rarely as susceptible as a known susceptible check variety. The addition of *Pto*|*Rx-3* and *QTL-11*|*Xv3* to our fresh market breeding lines offered modest increases in bacterial spot resistance that were only detectable when a major gene resistance was deployed against a matching pathogen race. For example, there was strong evidence that deployment of HR from *Xv3* (in combination with *QTL-11*) enhanced resistance to bacterial spot race T3 of *X. perforans* in our background. Similarly, our results suggested that addition of *Rx-3* in our breeding lines improved resistance to *X. euvesicatoria* race T1 in our low-powered trials.

The less-penetrant quantitative resistance effect of *QTL-11* was largely indistinguishable in our background. We hypothesized that *QTL-11*|*Xv3* would confer enhanced resistance to *X. hortorum* pv. *gardneri* and *X. perforans* race T4 (against which we did not deploy a matching major gene resistance), as was recently observed in processing and fresh market tomatoes (Bernal et al. 2020; Hutton et al. 2010). However, the *QTL-11*|*Xv3*

haplotype offered at most a very modest improvement in resistance to these isolates in our genetic background. Similarly, we failed to detect enhanced resistance to *X. euvesicatoria* race T1 in NILs homozygous for *QTL-11*|*Xv3* in either 2018 or 2019 tests. It is possible that *QTL-11*'s resistance effect was masked by the appreciable resistance already shown by our recurrent fresh market tomato. Thus, breeding programs with strong, existing quantitative bacterial spot resistance may not be justified in "locking up" the genomic region encompassed by the whole *QTL-11*|*Xv3* haplotype, as enhanced resistance may not materialize. Alternatively, the lack of tightly linked markers and the centromeric positioning of *QTL-11* raises the possibility that this resistance locus is more prevalent among cultivated tomatoes than is currently postulated and was already present in our recurrent parent. We have never intentionally selected for *QTL-11* resistance, nor knowingly crossed with a *QTL-11* donor until now. However, a centromeric chromosome 11 QTL has been identified in several accessions (Bernal et al. 2020; Scott et al 2015), including fresh market breeding lines from the University of Florida (Hutton et al. 2010).

Tests of introgression effects on horticultural performance allowed us to weigh the costs and benefits of the resistance linkages in our fresh market background. For example, our work found the *QTL-11*|*Xv3* haplotype increased *X. perforans* race T3 bacterial spot disease resistance and raised early season marketable yields without negatively impacting critical horticultural traits, especially fruit size. Although we observed some evidence for greater fruit irregularity in *QTL-11*|*Xv3* lines in one environment, this is to be expected for fruit with a larger cross-sectional area. The *Pto*|*Rx-3* haplotype appeared to improve resistance to *X. euvesicatoria* race T1 and resulted in a greater proportion of marketable fruit with thicker pericarps than the recurrent parent. However, lines with *Pto*|*Rx-3* also exhibited lower early season yields, which is a significant issue for tomato growers in short-season environments like the Northeast. The later maturity among the *Pto*|*Rx-3* introgression lines is also consistent

with our empirical observation that these lines had larger vine size (not shown), and our horticultural findings that these lines had less red fruit and had fewer unmarketable fruit defects (which are often due to rapid fruit growth). Later maturity can also confound the resistance effects of bacterial spot resistance loci (Yang et al 2005). Fortunately, the negative horticultural associations with the *Pto*|*Rx*-3 introgression were not fully dominant and can be ameliorated through strategic hybridization. We hope that our findings on the costs and benefits of two valuable coupling phase bacterial resistance linkages will empower tomato breeders to develop better hybridization plans that balance the objectives of elite horticultural quality and broad-spectrum bacterial resistance.

.

REFERENCES

Arredondo CR, Davis RM (2000) First report of Pseudomonas syringae pv. tomato race 1 on tomato in California. Plant Disease 84(3):371-371

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

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

Blanca J, Montero-Pau J, Sauvage C, Bauchet G, Illa E, Díez MaJ, Francis D, Causse M, van der Knaap, E, Cañizares Jn (2015) Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. BMC Genomics 16:257. https://doi.org/10.1186/s12864-015-1444-1

Brewer MT, Lang L, Fujimura K, Dujmovic N, Gray S, van der Knaap E (2006) Development of a controlled vocabulary and software application to analyze fruit shape variation in and other plant species. Plant Physiology 141:15-25. https://dx.doi.org/10.1104%2Fpp.106.077867

Cai R, Lewis J, Yan S, Liu H, Clarke CR, Campanile F, Almeida NF, Studholme DJ, Lindeberg M, Schneider D, Zaccardelli M, Setubal JC, Morales-Lizcano NP, Bernal A, Coaker G, Baker C, Bender CL, Leman S, Vinatzer BA (2011) The plant pathogen *Pseudomonas syringae* pv. *tomato* is genetically monomorphic and under strong selection to evade tomato immunity. PLoS Pathog. 7(8):e1002130. https://doi.org/10.1371/journal.ppat.1002130

Darrigues A, Hall J, van der Knaap E, Francis DM (2008) Tomato Analyzer-Color Test: a new tool for efficient digital phenotyping. J. Am. Soc. of Hort. Sci. 133:579-586. https://doi.org/10.21273/JASHS.133.4.579

DeGaetano AT (2009). Time-dependent changes in extreme-precipitation return-period amounts in the continental United States. Journal of Applied Meteorology and Climatology 48(10), 2086-2099. https://doi.org/10.1175/2009JAMC2179.1

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

Faria JC, Demétrio CGB (2019) bpca: Biplot of multivariate data based on Principal Components Analysis. ESALQ, USP, Brasil

Furuta T, Ashikari M, Jena KK, Doi K, Reuscher S (2017) Adapting genotyping-by-sequencing for rice F_2 populations. G3: Genes, Genomes, Genetics 7(3):881-893. https://doi.org/10.1534/g3.116.038190

Gitaitis RD, Jones JB, Jaworski CA, Phatak SC (1985) Incidence and development of *Pseudomonas syringae* pv. syringae on tomato transplants in Georgia. Plant Disease 69:32–35

Gonzalo MJ, Brewer MT, Anderson C, Sullivan D, Gray S, van der Knaap E (2009) Tomato Fruit Shape Analysis Using Morphometric and Morphology Attributes Implemented in Tomato Analyzer Software Program. J. Am. Soc. of Hort. Sci. 134:77-87. https://doi.org/10.21273/JASHS.134.1.77

Graves S, Piepho HP, Selzer L, Sundar DR (2019). multcompView: Visualizations of Paired Comparisons. R package version 0.1-8. https://CRAN.R-project.org/package=multcompView

Hospital F (2009) Challenges for effective marker-assisted selection in plants. Genetica 136:303-310. https://doi.org/10.1007/s10709-008-9307-1

Horsfall JG, Barratt R (1945) An improved grading system for measuring plant diseases. Phytopathology 35:655.

Hospital F, Chevalet C, Mulsant P (1992) Using markers in gene introgression breeding programs. Genetics 132(4):1199-1210

Hutton, SF, Scott JW, Yang W, Sim SC, Francis DM, Jones JB (2010) Identification of QTL associated with resistance to bacterial spot race T4 in tomato. Theor. Appl. Genet. 121:1275-1287. https://doi.org/10.1007/s00122-010-1387-5

Jardine DJ, Stephens CT, Fulbright DW (1988) Potential sources of initial inoculum for bacterial speck in early planted tomato crops in Michigan: debris and volunteers from previous crops. Plant Dis. 72:246–250

Jones JB. 1991. Bacterial speck. In Compendium of tomato diseases, ed. JB Jones, JP Jones, RE Stall, TA Zitter, pp. 26–27. St. Paul, MN: APS Press

Kraus CM, Mazo-Molina C, Smart CD, Martin GB (2017) *Pseudomonas syringae* pv. *tomato* Strains from New York Exhibit Virulence Attributes Intermediate Between Typical Race 0 and Race 1 Strains. Plant Disease 101(8):1442-1448. https://doi.org/10.1094/PDIS-03-17-0330-RE

Kunkeaw S, Tan S, Coaker G (2010) Molecular and evolutionary analyses of *Pseudomonas syringae* pv. *tomato* race 1. Molecular Plant-Microbe Interactions 23(4):415-424. https://doi.org/10.1094/MPMI-23-4-0415

Lawton MB, MacNeill BH (1986) Occurrence of race 1 of *Pseudomonas syringae* pv. *tomato* on field tomato in southwestern Ontario. J. Plant Pathol. 8:85-88

Length R (2017) emmeans: estimated marginal means, aka least-squares means. R package version 1.0. https://CRAN.R-project.org/package=emmeans

Liabeuf D (2016) Development of processing tomato lines resistant to *Xanthomonas gardneri*: From screening to breeding [dissertation]. Ohio State University, Columbus, OH

Liao YY, Strayer-Scherer AL, White J, Mukherjee A, De La Torre-Roche R, Ritchie L, Colee J, Vallad GE, Freeman JH, Jones JB, Paret ML (2019) Nano-magnesium oxide: A novel bactericide against copper-tolerant *Xanthomonas perforans* causing tomato bacterial spot. Phytopathology 109:52-62. https://doi.org/10.1094/PHYTO-05-18-0152-R

Louws FJ, Wilson M, Campbell HL, Cuppels DA, Jones JB, Shoemaker PB, Sahin F, Miller SA (2001) Field control of bacterial spot and bacterial speck of tomato using a plant activator. Plant Dis. 85:481-488. https://doi.org/10.1094/PDIS.2001.85.5.481

Marco G, Stall R (1983) Control of bacterial spot of pepper initiated by strains of *Xanthomonas campestris* pv. *vesicatoria* that differ in sensitivity to copper. Plant Dis. 67:779-781. https://doi.org/10.1094/PD-67-779

Mazo-Molina C, Mainiero S, Hind SR, Kraus CM, Vachev M, Maviane-Macia F, Lindeberg M, Saha S, Strickler SR, Feder A, Giovannoni JJ, Smart CD, Peeters N, Martin GB (2019) The *Ptr1* locus of *Solanum lycopersicoides* confers resistance to race 1 strains of *Pseudomonas syringae* pv. *tomato* and to *Ralstonia pseudosolanacearum* by recognizing the type III effectors AvrRpt2 and RipBN." Molecular Plant-Microbe Interactions 32(8):949-960

Morinière L, Burlet A, Rosenthal ER, Nesme X, Portier P, Bull CT, Lavire C, Fischer-Le Saux M, Bertolla F (2020) Clarifying the taxonomy of the causal agent of bacterial leaf spot of lettuce through a polyphasic approach reveals that *Xanthomonas cynarae* Trébaol *et al.* 2000 emend. Timilsina *et al.* 2019 is a later heterotypic synonym of *Xanthomonas hortorum* Vauterin *et al.* 1995. Systematic and Applied Microbiology 43(4) 126087. https://doi.org/10.1016/j.syapm.2020.126087

Pedley KF, Martin GB (2003) Molecular basis of *Pto*-mediated resistance to bacterial speck disease in tomato. Annu. Rev. Phytopathol. 41:215-243. https://doi.org/10.1146/annurev.phyto.41.121602.143032

Pei C, Wang H, Zhang J, Wang Y, Francis DM, Yang W (2012) Fine mapping and analysis of a candidate gene in tomato accession PI128216 conferring hypersensitive resistance to bacterial spot race T3. Theor. Appl. Genet. 124:533-542. https://doi.org/10.1007/s00122-011-1726-1

Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC, Nelson RJ (2009) Shades of gray: the world of quantitative disease resistance. Trends in Plant Sci. 14(1):1-29. https://doi.org/10.1016/j.tplants.2008.10.006

Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLOS ONE 7(2):e32253. https://doi.org/10.1371/journal.pone.0032253

Potnis N, Timilsina S, Strayer A, Shantharaj D, Barak JD, Paret ML, Vallad GE, Jones JB. 2015. Bacterial spot of tomato and pepper: Diverse *Xanthomonas* species with a wide variety of virulence factors posing a worldwide challenge. Mol. Plant Pathol. 16:907-920. https://doi.org/10.1111/mpp.12244

Qian B, Gameda S, Zhang X, De Jong R (2012) Changing growing season observed in Canada. Climatic Change 112, 339–353. https://doi.org/10.1007/s10584-011-0220-8

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Robbins MD, Darrigues A, Sim SC, Masud MA, Francis DM (2009) Characterization of hypersensitive resistance to bacterial spot race T3 (*Xanthomonas perforans*) from tomato accession PI 128216. Phytopathology 99:1037-1044. https://doi.org/10.1094/PHYTO-99-9-1037

Ronald PC, Salmeron JM, Carland FM, Staskawicz BJ (1992) The cloned avirulence gene avrPto induces disease resistance in tomato cultivars containing the Pto resistance gene. J. Bacteriol. 174:1604–11. https://dx.doi.org/10.1128%2Fjb.174.5.1604-1611.1992

Schneider RW, Grogan RG (1977) Bacterial speck of tomato: sources of inoculum and establishment of a resident population. Phytopathology 67:388–94

Scott JW, Hutton SF, Shekasteband R, Sim SC, Francis DM (2015) Identification of tomato bacterial spot race T1, T2, T3, T4, and *Xanthomonas gardneri* resistance QTLs derived from PI 114490 populations selected for race T4. Proceedings of the IV International Symposium on Tomato Diseases 1069:53-58

Scott JW, Jones JB, Somodi GC (2001) Inheritance of resistance in tomato to race T3 of the bacterial spot pathogen. J. Am. Soc. Hortic. Sci. 126:436-441. https://doi.org/10.21273/JASHS.126.4.436

Scott JW, Jones JB, Somodi GC, Stall RE (1995) Screening tomato accessions for resistance to *Xanthomonas campestris* pv. *vesicatoria*, race T3. HortScience 30(3):579-581. https://doi.org/10.21273/HORTSCI.30.3.579

Scott JW, Stall RE, Jones J, Somodi GC (1996) A single gene controls the hypersensitive response of Hawaii 7981 to race 3 (T3) of the bacterial spot pathogen. Tomato Genet. Coop. 46:23

Sharlach M, Dahlbeck D, Liu L, Chiu J, Jiménez-Gómez JM, Kimura S, Koenig D, Maloof JN, Sinha N, Minsavage GV, Jones JB, Stall RE, Staskawicz BJ (2013) Fine genetic mapping of *RXopJ4*, a bacterial spot disease resistance locus from *Solanum pennellii* LA716. Theoretical and Applied Genetics 126(3):601-9

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

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

Stall RE, Jones JB, Minsavage GV (2009) Durability of resistance in tomato and pepper to xanthomonads causing bacterial spot. Annual review of phytopathology 47:265-284. https://doi.org/10.1146/annurev-phyto-080508-081752

Tanksley, SD, Ganal MW, Prince JP, De Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB (1992) High density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141-1160

Timilsina S, Kara S, Jacques MA, Potnis N, Minsavage GV, Vallad GE, Jones JB, Fischer-Le Saux M (2020) Reclassification of *Xanthomonas gardneri* (ex Šutič 1957) Jones et al. 2006 as a later heterotypic synonym of *Xanthomonas cynarae* Trébaol et al. 2000 and description of *X. cynarae* pv. *cynarae* and *X. cynarae pv. gardneri* based on whole genome analyses. International journal of systematic and evolutionary microbiology 69(2):343-9. https://doi.org/10.1099/ijsem.0.003104

Timilsina S, Jibrin MO, Potnis N, Minsavage GV, Kebede M, Schwartz A, Bart R, Staskawicz B, Boyer C, Vallad GE, Pruvost O, Jones J, Goss EM (2015) Multilocus sequence analysis of Xanthomonads causing bacterial spot of tomato and pepper plants reveals strains generated by recombination among species and recent global spread of *Xanthomonas gardneri*. Applied and Environmental Microbiology 81(4):1520-1529. https://doi.org/10.1128/AEM.03000-14

Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485(7400):635–641. https://doi.org/10.1038/nature11119

Veluchamy S, Hind SR, Dunham DM, Martin GB, Panthee DR (2014) Natural variation for responsiveness to flg22, flgII-28, and csp22 and *Pseudomonas syringae* pv. *tomato* in heirloom tomatoes. PLoS One 9:e106119. https://doi.org/10.1371/journal.pone.0106119

Wang H, Hutton SF, Robbins MD, Sim SC, Scott JW, Yang W, Jones JB, Francis DM (2011) Molecular mapping of hypersensitive resistance from tomato 'Hawaii 7981' to *Xanthomonas perforans* race T3. Phytopathology 101:1217-1223. https://doi.org/10.1094/PHYTO-12-10-0345

Whalen MC, Wang JF, Carland FM, Heiskell ME, Dahlbeck D, Minsavage GV, Jones JB, Scott JW, Stall RE, Staskawicz BJ (1993) Avirulence gene avrRxv from *Xanthomonas campestris* pv. *vesicatoria* specifies resistance on tomato line Hawaii 7998. Molecular Plant Microbe Interactions 6(5):616-627. https://doi.org/10.1094/mpmi-6-616

Yang W, Francis DM (2005) Marker-assisted selection for combining resistance to bacterial spot and bacterial speck in tomato. J. Am. Soc. Hortic. Sci. 130:716-721. https://doi.org/10.21273/JASHS.130.5.716

Yang W, Miller S, Scott J, Jones J, Francis D (2005) Mining tomato genome sequence databases for molecular markers: Application to bacterial resistance and marker assisted selection. Acta Horticulturae 695:241-250. https://doi.org/10.17660/ActaHortic.2005.695.26

Young, Dye DW, Wilkie JP (1986) Bacterial Speck. *In* Vegetable Diseases and Their Control, ed. AF Sherf, pp. 610–14. New York: Wiley

Yu ZH, Wang JF, Stall RE, Vallejos CE (1995) Genomic localization of tomato genes that control a hypersensitive reaction to *Xanthomonas campestris* pv. *vesicatoria* (Doidge) dye. Genetics 141:675-682

CHAPTER 2

CRYPTIC INTROGRESSIONS CONTRIBUTE TO TRANSGRESSIVE SEGREGATION FOR EARLY BLIGHT RESISTANCE IN TOMATO

Anderson, TA, Zitter SM, De Jong DM, Francis DM, and MA Mutschler

Affiliations

Anderson, TA, Zitter SM, De Jong DM and MA Mutschler*, Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853

DM Francis, Department of Horticulture and Crop Science, The Ohio State University, Wooster, OH 44691

Acknowledgments

TAA conceptualized laboratory and field experiments, completed formal analysis and investigation, acquired research funding, and wrote the manuscript. SMZ completed preliminary experiments and helped to develop pathology protocols. DMD assisted with genotyping and marker design. DMF provided sequence data for KASP maker development. DMF and MAM supervised and advised TAA and acquired research funding. All authors reviewed, edited, and approved the final manuscript.

Thanks to Dr. Michael Allen Gore, Professor at Cornell University and Dr. Deniz Akdemir, Statistical Consultant at the Cornell Statistical Consulting Unit for their advice on statistical methods. Dr. Tom Zitter, Cornell University Emeritus Professor of Plant Pathology, provided *Alternaria* isolates. Dr. Randolph Gardner, Emeritus Professor of Horticultural Science at North Carolina State University, supplied the seed of NC 1 CELBR, NC84-173, and C1943.

Abstract

Early blight is a widespread and problematic disease affecting tomatoes (*Solanum lycopersicum*). Caused by the fungal pathogen *A. linariae (syn. A. tomatophila*), symptoms include lesions on tomato stems, fruit, and foliage, often resulting in yield losses. Breeding tomatoes with genetic resistance would enhance production sustainability. Using cross-market breeding populations, we identified several quantitative trait loci (QTL) associated with early blight resistance. Early blight resistance putatively derived from 'Campbell 1943' and *S. habrochaites* PI 126445 was confirmed in modern fresh market tomato breeding lines. This resistance offered substantial protection against early blight stem lesions (collar rot) and moderate protection against defoliation. A distinctive and potentially novel form of early blight foliar resistance was discovered in a processing tomato breeding line and is probably derived from *S. pimpinellifolium* via 'Hawaii 7998'. Additional field trials validated the three most promising large-effect QTL, *EB-1.2*, *EB-5*, and *EB-9*. Resistance effects for *EB-5* and *EB-9* were consistent across breeding populations and environments, while *EB-1.2*'s effect was population-specific. Using genome-wide marker-assisted selection, we developed fresh market tomato lines that were near-isogenic for early blight QTL. Resistance in these lines was largely mediated by just two QTL, *EB-5* and *EB-9*, that together captured 49.0% and 68.7% of the defoliation and stem lesion variance, respectively. Our work showcases the value of mining cryptic introgressions in tomato lines, and across market classes, for use as additional sources of disease resistance.

CHAPTER 2 SUPPLEMENTAL MATERIALS

Supplement 2 Supplementary methods for growing early blight inoculum

Supplement 3 Tomato DNA sequences and CAPS markers associated with early blight disease

Supplement 4 Supplementary results tables S1-S9

Supplement 5 Supplementary figures S1-S7

Supplement 6 A compressed (.zip) file containing data tables S1-S7 and a table of contents

Introduction

Early blight is a severe disease of tomato (*Solanum lycopersicum* L.) and potato (*Solanum tuberosum* L.) grown in humid and semi-arid environments (Rotem 1994; Rotem and Reichert 1964). The necrotrophic fungal species *Alternaria linariae* (syn. *A. tomatophila*) causes early blight on tomato (Andersen et al. 2008; Brun et al. 2013; Gannibal et al. 2014; Rodrigues et al. 2010; Simmons 2000; Woudenberg et al. 2014). In contrast to biotrophic pathogens (which survive by sequestering nutrients from a living host), necrotrophic fungi secrete toxins that kill or weaken plant cells and then feed on the decaying tissue (Poland et al. 2009). Early blight lesions on tomato stems, leaves, and fruit begin as small, dark patches of dead tissue. Over time, the lesions expand into large, sunken, and irregular necrotic zones. Stem necrosis on young seedlings can cause plant collapse in a disease phase known as collar rot. Foliar necrosis leads to defoliation and sunscald of fruit and can lower marketable yields by more than 50% (Rotem 1994). Growers in areas where early blight is common, including the Northeast, Mid-Atlantic, and Midwest United States, must use multiple fungicide applications each season to control the disease (Patterson et al 2001, Percheur 2009, Zitter et al. 2005). Breeding tomatoes with genetic resistance offers a companion strategy to chemical control that can be more economically and environmentally sustainable.

Tomato breeders have employed genetic resistance to early blight disease in breeding programs since the 1940s (Andrus et al. 1942a, b). Some of this resistance was introgressed into advanced tomato breeding lines or cultivars, though the early lines, in particular, were often later maturing, had large and vigorous growth habits, or had reduced fruit yields (Barksdale 1969; Barksdale 1971; Barksdale and Stoner 1973; Barksdale and Stoner 1977; Barratt and Richards 1944; Chen and Mutschler 1988; Gardner 1988; Gardner and Shoemaker 1999; Locke 1949; Nash 1986; Nash and Gardner 1988b; Panthee and Gardner 2010). These

negative associations and a supposedly complex quantitative inheritance pose challenges to the incorporation of early blight resistance into tomato cultivars (Barksdale and Stoner 1973; Barksdale and Stoner 1977; Maiero et al. 1989, 1990a, b; Martin and Hepperly 1987; Nash and Gardner 1988a). However, this resistance is desirable due to the economic impact of early blight disease and because quantitatively inherited resistance is postulated to be more broadly effective than resistance derived from highly penetrant single genes (Poland et al. 2009).

Genetic dissection of early blight resistance in modern breeding lines would enable breeders to develop resistant tomato varieties efficiently using genotypic selection. Prior research identified QTL associated with resistance to early blight stem lesions and defoliation (Ashrafi and Foolad 2015; Chaerani et al. 2007; Foolad et al. 2002; Foolad et al. 2005; Zhang et al. 2003). However, these studies focused on QTL in interspecific crosses since it was difficult at that time to identify genetic polymorphism for QTL mapping within cultivated tomato. Recent improvements in marker resources and technologies including the SolCAP single nucleotide polymorphism (SNP) panel (Gill et al. 2019; Hamilton et al. 2012; Sim et al. 2012a, b) and the extraction of genotypic data from genotyping by sequencing (GBS) (Elshire et al. 2011) have made it easier to map QTL in commercially adapted tomatoes. This approach is advantageous to breeders since it exploits historical introgressions that, due to decades of breeding improvement, could have fewer genetic linkages between the resistance locus and loci negatively impacting horticultural performance (linkage drag) compared to their wild species counterparts.

We investigated the genetic determinants of early blight resistance in modern tomato breeding lines from two university breeding programs. Three Northeast-adapted fresh market tomato lines, CU151095-146, CU151011-146, and CU151011-170, were previously bred for resistance to early blight stem and foliar disease (Zitter et al. 2005; Zitter et al. 2011; Zitter and

Drennan 2007; Zitter and Drennan 2008**)**. Their resistance came from the elite breeding line NC 1 CELBR, which in turn resulted from breeding with Campbell 1943 (C1943) and *Solanum habrochaites* PI 126445 (Gardner 1988; Gardner and Panthee 2010; Gardner and Shoemaker 1999; Panthee and Gardner 2010). Two Midwest processing tomatoes, OH7536 and OH08-7663, were not bred for early blight resistance specifically, but are adapted to humid environments where early blight is common. In crossing tomatoes from these distinct market classes, we identified early blight resistance introgressions and used trait-linked genetic markers to study the combinatorial effects of resistance QTL in a contemporary fresh market background. Importantly, we found that a small number of QTL explained much of the early blight resistance in our tomatoes.

Materials and Methods

Germplasm for resistance characterization and mapping. Early blight resistance was measured in tomato breeding lines from three sources: fresh market lines from Cornell University (CU151011-146, CU151011-170 and CU151095-146), processing lines from The Ohio State University (OH08-7663 and OH7536) and fresh market lines from North Carolina State University (NC 1 CELBR, NC84-173). Mapping and validation of resistance QTL was accomplished in multiple populations and generations. Seven backcross (BC_1) families were created by crossing F_2 individuals of CU151011-170 x OH7536 to the processing line OH08-7663. Progeny were screened in a 2017 greenhouse trial at the Guterman Bioclimatic Lab in Ithaca, NY. An F_2 population (n = 276), CU151095-146 x OH08-7663, was transplanted at the Terwilliger Plant Pathology Research Farm (TPPRF) in Freeville, NY in 2017. F_2 plants from the 2017 field were self-pollinated, generating seed of $F_{2:3}$ families. $F_{2:3}$ families were genotypically selected using SNPs associated with stem and foliar resistance (see below) and were transplanted at TPPRF in 2018 and 2019.

Development of near-isogenic lines (NILs) for resistance QTL. Background genome selection (BGS) (Hospital et al. 1992) was initiated with 285 BC_1 plants from (OH08-7663 x CU151095-146) x CU151095-146 screened with 48 Kompetitive Allele Specific PCR (KASP) markers (Data S6, Supplement 6). A selected plant with a high proportion of the recurrent parent genome was self-pollinated to create a partially segregating BC_1F_2 population (n = 219). DNA of individual BC_1F_2 plants was extracted and genotyped with 15 PCR markers designed around specific QTL-linked SNPs. PCR protocols and polymorphic sequences are detailed in Supplement 3. A BC_1F_2 plant heterozygous for three QTL was self-pollinated, and 936 BC_1F_3 progeny were genotyped at 19 QTL-linked SNPs by a combination of amplicon sequencing and PCR markers (Data S7, Supplement 6). Sixty-six BC_1F_3 near-isogenic individuals were selected

for recombination at one or more QTL, and self-pollinated. The resulting $BC_1F_{3:4}$ families were planted in a replicated trial at TPPRF in 2019.

Genotyping of mapping populations. F_2 selections used to create the seven BC_1 families for an indoor stem screen were genotyped with 39 KASP markers distributed across the tomato genome (Data S1, Supplement 6). One-hundred and seventy-four F_2 plants at the extremes of the stem lesion and defoliation phenotype distributions were selectively genotyped (Darvasi and Soller 1992, Lebowitz et al. 1987). Reduced representation libraries were constructed with DNA from the F_2 plants using a PstI/MspI enzyme combination. Barcoded libraries were pooled and run on a single lane of an Illumina NextSeq 500 sequencer (Poland et al. 2012). Reads were aligned to the *S. lycopersicum* v3.0 genome (Tomato Genome Consortium 2012) with BWA (Li and Durbin 2009), variants were identified with the Tassel GBSv2 pipeline (Glaubitz et al. 2014) and named following the SL3.0CH[chromosome #]-[physical position on chromosome] convention. Genotypes were filtered using Tassel 5 (Bradbury et al. 2007) to remove < 8 read depths, multiple alleles, < 10% call rate across individuals, and < .05 minor allele frequencies (MAF). To fill gaps in genome coverage, 87 F_2 plants were genotyped again by amplicon sequencing of a 384 SNP panel (Gill et al. 2019), yielding an additional 235 SNPs after filtering as above. Genotypes were conservatively imputed with LDKNNI (Money et al. 2015) for use in linkage mapping or fully imputed with Beagle 5.1 (Browning et al. 2018) and default parameters for association mapping. Error correction of genotypes was done with ABHGenotypeR (Furuta et al. 2017) prior to use in linkage mapping.

Indoor disease screens. Indoor screens were performed as described by Barksdale 1968 with modifications. *Alternaria linariae* (formerly *A. tomatophila*) 'Gardner Mt. St.' (Frazer 2002) was grown as detailed in Supplement 2. Tomatoes were grown in 10 cm pots in a

greenhouse under 14h light at 27°C, 10h dark at 22°C, without fertilization for six weeks. Five milliliters of the *Alternaria* suspension at 6000 conidia/mL were applied to stems using a Preval® sprayer (Preval Sprayer Division, Yonkers, NY). Plants were moved every 12 hours between a 20°C unlit misting chamber and a 27°C greenhouse with ambient daylight for 7 days prior to evaluation of early blight stem lesions or defoliation.

Resistance was first characterized in the breeding lines CU151011-146, OH08-7663, and OH7536 using a completely randomized design with NC 1 CELBR and NC84-173 as resistant and susceptible controls, respectively. Early blight stem lesions were visually scored on a modified Franjul (1989) scale where 0 = no lesions, 1 = lesions < 1 mm across, 2 = lesions 1 to 5 mm across, 3 = elongated lesions > 5 mm, 4 = lesions girdling stem, 5 = plant collapse. Collar rot severity was scored on a reversed Reynard and Andrus (1945) scale, where 0 = no collar lesions and 4 = dead plant. Foliar disease was estimated by counting lesions > 2 mm in size (Barksdale 1969) and by scoring plants on a modified Horsfall-Barratt scale according to the percent of blighted foliage: 1 = 0%, 2 = 1 - 10%, 3 = 11 - 25%, 4 = 26 - 50%, and 5 = more than 50% (Lacy 1973).

A second indoor screen measured stem lesion resistance among seven BC_1 families and two F_1 hybrids between fresh market and processing tomatoes. Inoculation was as described above. Resistant parental controls included CU151011-170 and NC 1 CELBR, while the stem susceptible controls were OH08-7663 and OH7536. Each entry was replicated eight times, and the experiment was repeated four times. The percentage of stems affected by lesions was visually estimated on a 0-100% linear scale.

Field disease screens. Six-week old greenhouse-grown tomato seedlings were transplanted the first week of June. Fertilizer (650 kg/ha, 13-13-13 NPK) was applied two weeks before transplanting. Irrigation was through drip tape under black plastic mulch

covering beds. Plants were arbitrarily grouped into 6-8 plant plots with 120-150 cm between plants within a plot, 240 cm between plots within each row, and 275 cm between rows. $F_{2:3}$ families were established in a randomized complete block design (RCBD) with four and three replicates in 2018 and 2019, respectively. Over-replicated plots of CU151011-170, CU151095-146, OH7536, and OH08-7663, and the F_1 between CU151011-170 and OH7536 were randomly embedded in the experiment as controls. The 66 NIL families were placed in an RCBD design with three replicates in 2019. Because this was a test of QTL in the fresh-market background, the susceptible control was a fresh market NIL selected for homozygosity of susceptible alleles at resistance QTL (*EB-1.2*, *EB-5*, and *EB-9*, see below), while C1943 was also included as a resistant check.

Biological pests were limited by physical and chemical means. Weeds were controlled with black plastic mulch and a single application prior to transplant of the broadleaf herbicides Sandea® (Gowan), Metribuzin® (Loveland Products), and Dual Magnum® (Syngenta). Fungicides that target oomycete biological sites and are not thought to impact the lifecycle of *A. linariae* (Blum et al 2010; Mitani et al 2001a, b; Pasteris et al 2016; Toquin et al. 2007) were applied in rotation and included Orondis Ultra A® (Syngenta), Ranman® (Summitagro), and Presidio® (Valent). Fungicide application dates were based on a late blight disease forecast model (Small et al. 2015).

Basal stems and leaves of every plant were sprayed with 60 mL of *A. linariae* suspension using a backpack sprayer in the second week of July. Inoculum concentrations were 18,000 conidia/mL in 2017 and 11,000 conidia/mL in 2018 and 2019. Overhead irrigation was applied at 8 L/min, for 15 minutes, starting one hour after the inoculation and every other evening for one week, unless rain was forecast, to encourage infection.

Disease phenotyping began in early August when early blight symptoms were distinguishable on controls. Ratings were on a single-plant basis except for the $F_{2:3}$ and NIL trials, where observations were on a plot basis. Defoliation was estimated on a linear 0-100% scale on three occasions at ten and eight-day intervals for the 2017 F_2 and 2018/2019 $F_{2:3}$ populations, respectively. Defoliation was rated twice, sixteen days apart, for the BC_1F_2 trial and four times, 7 days apart, for the 2019 NIL experiment. Stem lesions were rated twice per season, 14 days apart, using the modified Franjul scale described above, except in the 2019 $F_{2:3}$ experiment, where one rating was done. Stem (stem-ADUPC) and foliar (foliar-AUDPC) area under the disease progress curves (Madden et al. 2007) were calculated from observations over a season. Internode lengths were estimated by averaging lengths (cm) of three adjacent basal internodes for three primary basal branches at the end of the season (mid-September). Yield potential was estimated by counting fruits larger than 5 cm on a plant in the third week of August.

Trait mapping: Univariate linkage mapping was done with R/qtl version 1.44-9 (Broman et al. 2003) using the multiple-QTL imputation framework (Arends et al. 2010). Genotypes were filtered against collinearity, unusual recombination fractions, and segregation distortion at $\alpha = .05$ Bonferroni-adjusted χ^2 thresholds. This yielded 290 high-quality markers for genetic map construction (Fig. S1, Supplement 5). QTL models for stem lesions and defoliation incorporated cofactors for plant location within a plot and covariates for yield potential and internode length (see below). Bayesian credible QTL confidence intervals ($\alpha = .05$) were expanded to the nearest identifiable SNPs in the tomato genome to define physical QTL boundaries. GEMMA (Zhou and Stephens 2014) was used to calculate likelihood ratio test statistics from a bivariate mixed-effects marker regression model for the joint stem and foliar

disease phenotype. False discovery rates (FDR; $\alpha = .05$) for both approaches were estimated by 5000 permutations.

Descriptive statistics: Statistical computations were done in R 3.5.3 (R Core Team 2017) for 64-bit Ubuntu Linux. Stem and foliar disease were modeled independently for each experiment, except in the bivariate association analysis. Experiments with family-wise replication were modeled with mixed linear regressions of the standard form $y = X\beta + Zu + e$ using lme4 (Bates et al 2015). Family or genotype class independent variables were treated as fixed effects and experimental blocks as random effects. QTL effects were similarly modeled, with fixed QTL genotype effects and random effects for block, family, and QTL x block interactions. QTL-by-block interactions were not significant at $p < .05$, so this term was dropped. Conditional R^2 of the model, multiplied by 100, gave the percentage of variance explained (PVE) by a QTL (Nakagawa et al. 2017). P-values for marker effects were determined by χ^2 tests on type III sums of squares. A QTL model for the BC_1F_2 experiment was fit by iteratively reweighted least-squares (robust) regression because of heteroskedastic residuals (Holland and Welsch 1977). Post-hoc tests were by the Tukey honest significant difference procedure ($\alpha = .05$) implemented in emmeans (Length 2017).

Results and Discussion

Preliminary resistance characterization in indoor screen

Mist-chamber experiments offered preliminary characterizations of early blight resistance in tomato lines from three breeding programs. The fresh market lines CU151011-146, CU151011-170, and NC 1 CELBR exhibited stem lesion and collar rot resistance that surpassed (p < .05) the susceptible OSU processing lines OH08-7663, OH7536, and the susceptible fresh market control NC84-173 (Tables S1 and S2, Supplement 4). Susceptible tomatoes developed elongated stem lesions that girdled stems and necrotized leaf axils, while resistant tomatoes had speck-sized lesions, typically under a millimeter across. OH08-7663 exhibited foliar resistance beyond that of the fresh market tomatoes in the indoor screens, with a smaller infected leaf area (p < .05) and fewer leaf lesions (p < .05). The distinct forms of resistance in the Cornell fresh market and the Ohio processing tomatoes suggested they had alternative and potentially complementary resistance-bearing introgressions.

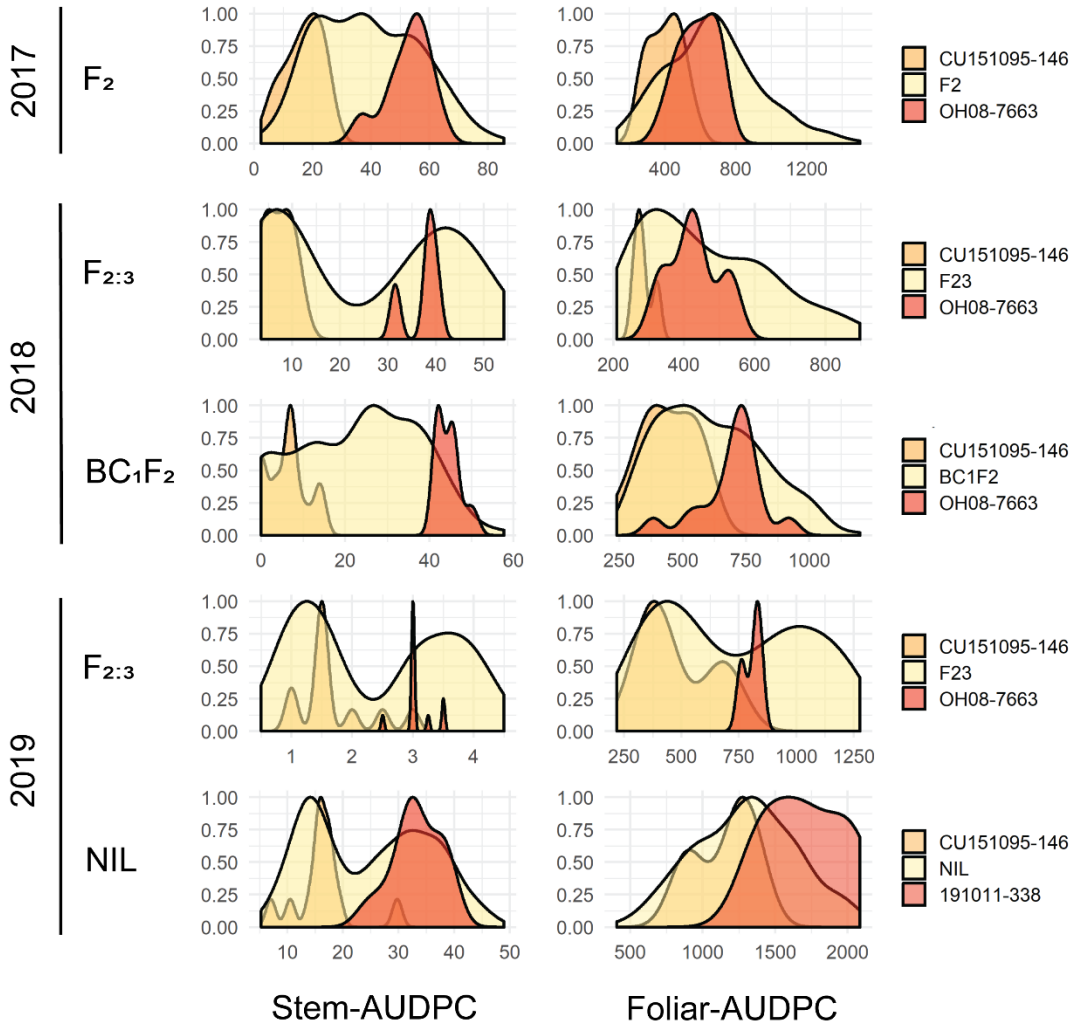
Preliminary mapping of stem resistance in an indoor screen

A mist chamber test of stem resistance in seven BC_1 families, for which we had low-density genotypic information, revealed a stem resistance QTL on chromosome 9. Just one marker, solcap_snp_sl_29188, exceeded the Bonferroni significance threshold (Table S3, Supplement 4). Stem resistance at this locus, which we designated *EB-9,* was contributed by the fresh market parent CU151011-170 (Fig. S2, Supplement 5). Because the number of markers and families was small, we could neither estimate *EB-9*'s effect nor rule out the possibility of other influential loci in the genome.

Early blight distributions and performance of parental controls

There were notable differences in stem and foliar resistance among the parental controls in three years of field experiments (Fig. 6). The fresh market tomato CU151095-146

Fig. 6 Scaled early blight stem lesion (stem-AUDPC) and defoliation (foliar-AUDPC) density distributions for five inoculated field trials in Freeville, NY. Parental controls are plotted separately from the experimental populations to enable visualization of transgressive segregation. Most of the transgressive segregation we observed was for early blight susceptibility, possibly because it was difficult to distinguish between increasingly resistant germplasm. CU151095-146 is a fresh market tomato with resistance to stem lesions and defoliation. OH08-7663 is a processing tomato with foliar early blight resistance. 191011-338 is a fresh market tomato genotypically selected for early blight susceptibility



had smaller areas under the stem lesion and defoliation progress curves than OH08-7663 (p < .05) for all but the foliar-AUDPC phenotype in 2019, which had the least replication (3 replicates) and was significant at p < .1 (Fig. S3 in Supplement 5). This finding was in contrast with the indoor test, where OH08-7663 boasted stronger foliar resistance, possibly because fresh market types produce more foliage on larger internodes, which is associated with foliar resistance in the field (Chen and Mutschler 1988). We observed transgressive segregation for resistance in some of the progeny populations, furthering the hypothesis of distinct underlying resistance loci. However, most of the transgressive segregation was on the susceptible end of the resistance spectrum.

Identification of early blight resistance QTL in an F_2 population

Two QTL analysis methodologies took advantage of the distinctiveness of the stem and foliar resistance phenotypes and the correlation between them (R^2 = 0.32, $F_{1,274}$ = 126.8, p < .001), in the 2017 F_2 experiment. The first approach used a linkage model to identify QTL associated with either stem or foliar disease. Covariates in this model accounted for factors known to impact early blight resistance, including edge-effects, yield potential, and internode lengths (Locke 1949; Chen and Mutschler 1988). This approach reduced the probability that QTL reflected morphological differences between fresh market and processing tomato that were associated with resistance. We also used a bivariate mixed-effects linear model to map stem and foliar resistance QTL jointly and as a proxy for *whole-plant* disease resistance.

Incompletely dominant QTL on chromosomes 1, 5, and 9 were statistically significant by both QTL mapping approaches and were thus the focus of subsequent experiments (Figs. 7, 8). Two partially linked chromosome 1 QTL, *EB-1.1* and *EB-1.2*, explained 6.7% and 4.9% of stem-AUDPC and foliar-AUDPC, respectively (Table 2). The SNP markers SL3.0CH01_86049848

Fig. 7 QTL associated with early blight disease in an intra-specific F_2 population. **a)** Multiple-QTL linkage analysis identified QTL associated with AUDPC for early blight stem lesions (solid grey lines) and defoliation (solid black lines) in the tomato genome. Horizontal lines indicate 5% FDR level, LOD_{FDR} = 2.02 for stem lesions (dotted grey line), LOD_{FDR} = 3.15 for defoliation (dotted black line). **b)** A bivariate mixed-effect linear model identified QTL associated with the joint stem and foliar AUDPC phenotypes, 5% FDR (solid black line) $-\log_{10}(p)$ = 3.73

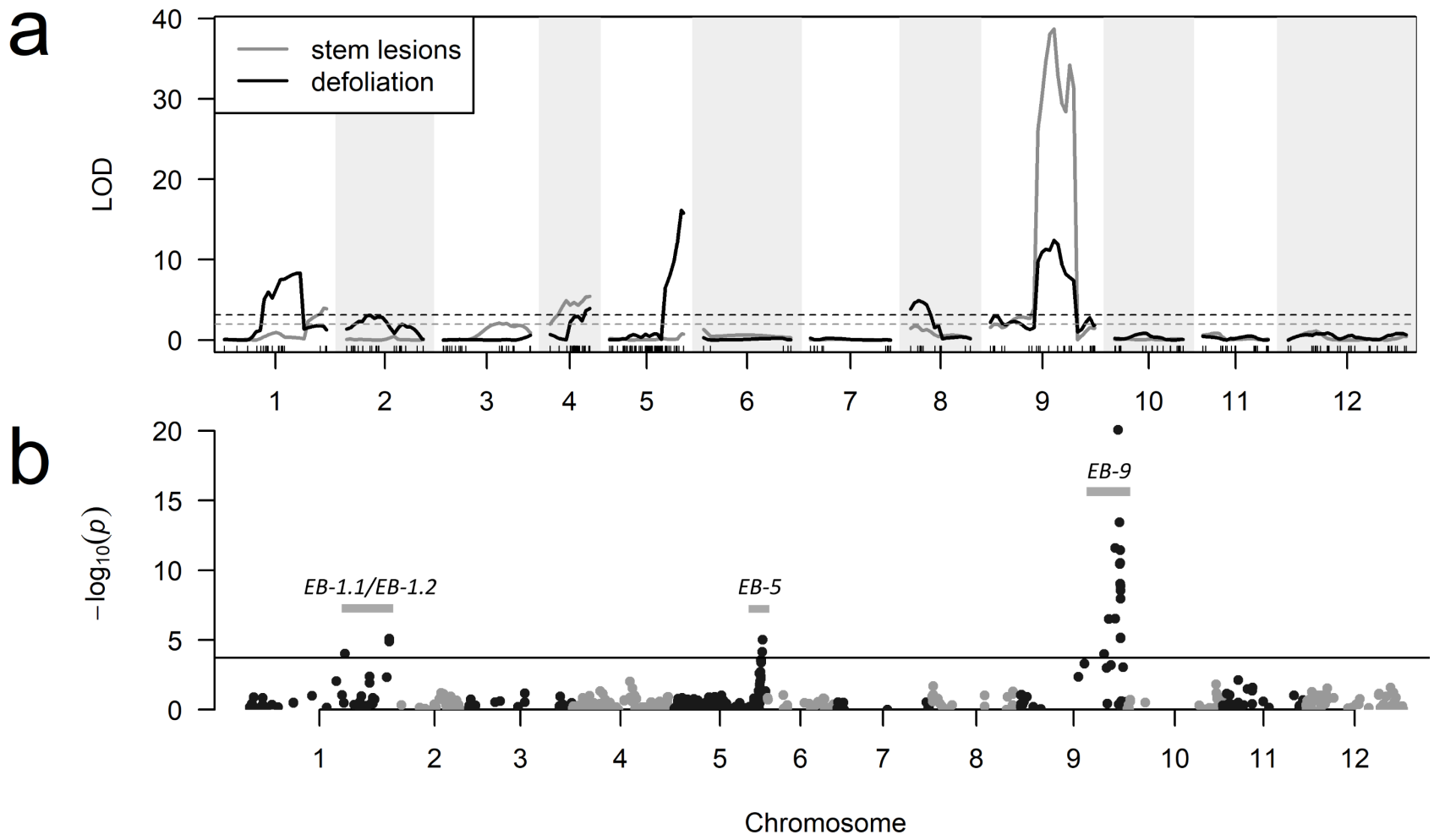


Fig. 8 Allele substitution effects for early blight QTL in three years of inoculated field trials. Genotypes were homozygous CU151095-146 (CU), OH08-7663 (OH), or heterozygous (HET) at QTL. QTL effects were significant at $p < .001$ except when noted by "NS". Additional details can be found in Table 2. Note that early blight stem lesions and defoliation were rated using scales of differing magnitudes and axes differ for clarity

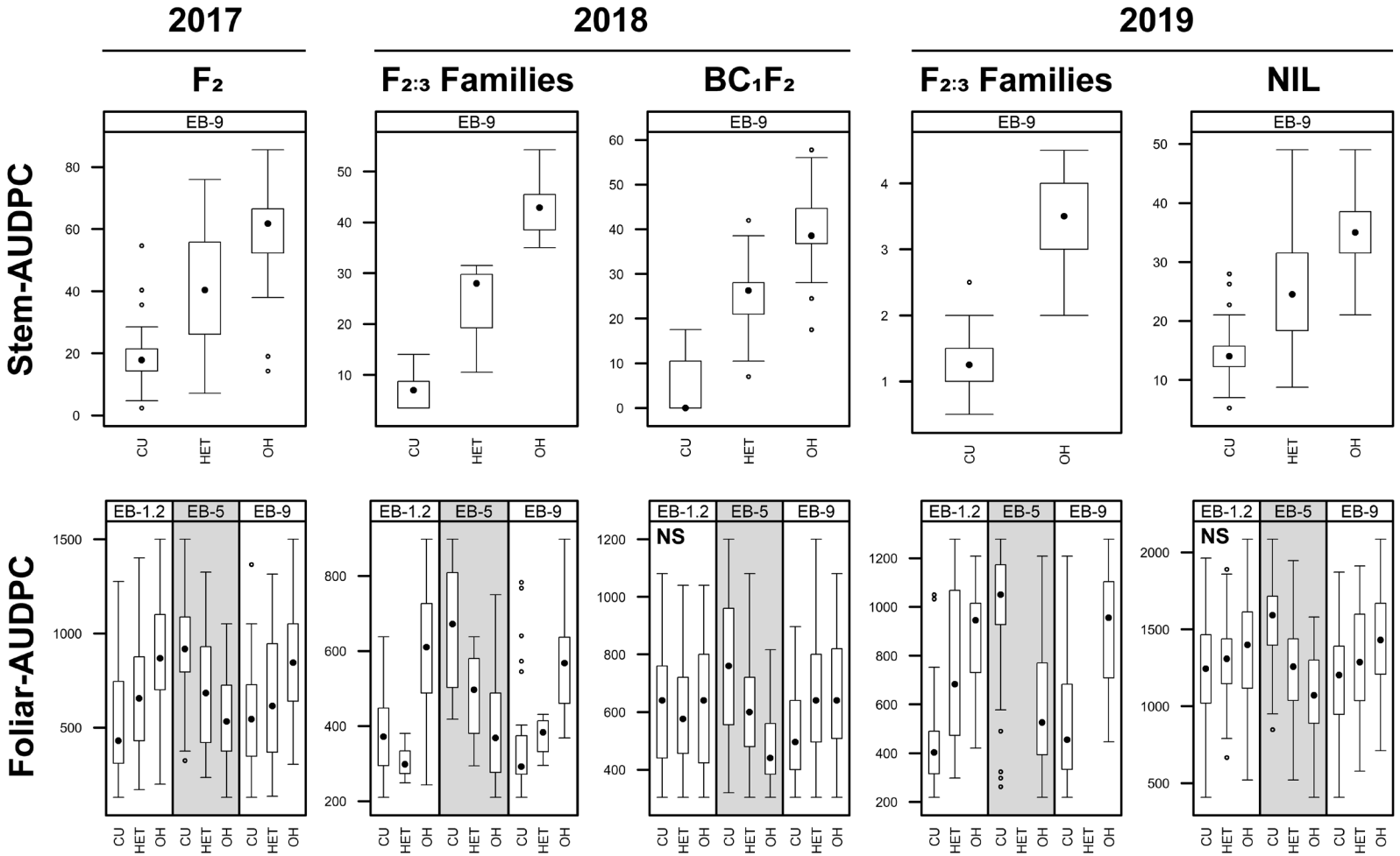


Table 2 QTL and molecular markers implicated in early blight disease resistance. An F_2 mapping experiment identified QTL associated with early blight stem lesions (stem-AUDPC) and defoliation (foliar-AUDPC). The three QTL with the greatest impacts (*EB-1.2*, *EB-5*, *EB-9*) on resistance were studied over the next two years in the context of a molecular breeding program. Polymorphic marker sequences and details can be found in Supplement 3

Year	Population	Trait	QTL	Marker[a]	Resistant Source	Chr	Pos	CI[b] (Mbp)	…
2017	F_2 (n=276)	stem lesions	*EB-1.1*	solcap_snp_sl_13399	CU151095-146	1	96.17	86.05 - 96.17	…
		defoliation	*EB-1.2*	solcap_snp_sl_34568	CU151095-146	1	85.02	81.03 – 94.46	…
		stem lesions	*EB-3*	SL3.0CH03_62848039	CU151095-146	3	62.85	5.23 – 71.57	…
		stem lesions	*EB-4*	SL3.0CH04_54163440	CU151095-146	4	54.16	14.90 – 54.42	…
		defoliation	*EB-4*	SL3.0CH04_54163440	CU151095-146	4	54.16	14.90 – 54.42	…
		defoliation	*EB-5*	SL3.0CH05_64366933	OH08-7663	5	64.37	64.16 – 64.37	…
		defoliation	*EB-8*	SL3.0CH08_2846199	OH08-7663	8	2.85	2.58 – 3.44	…
		stem lesions	*EB-9*	solcap_snp_sl_29188	CU151095-146	9	66.95	64.95 – 67.89	…
		defoliation	*EB-9*	solcap_snp_sl_29188	CU151095-146	9	66.95	61.99 – 67.89	…
2018	$F_{2:3}$ (n=60)	defoliation	*EB-1.2*	solcap_snp_sl_34568	CU151095-146	1	85.02	-	…
		defoliation	*EB-5*	SL3.0CH05_64366933	OH08-7663	5	64.37	-	…
		defoliation	*EB-9*	solcap_snp_sl_29188	CU151095-146	9	66.95	-	…
		stem lesions	*EB-9*	solcap_snp_sl_29188	CU151095-146	9	66.95	-	…
	BC_1F_2 (n=219)	defoliation	*EB-1.2*	SL3.0CH01_82710329	CU151095-146	1	82.71	-	…
		defoliation	*EB-5*	solcap_snp_sl_194	OH08-7663	5	64.69	-	…
		defoliation	*EB-9*	solcap_snp_sl_29188	CU151095-146	9	66.95	-	…
		stem lesions	*EB-9*	solcap_snp_sl_29188	CU151095-146	9	66.95	-	…
2019	$F_{2:3}$ (n=90)	defoliation	*EB-1.2*	solcap_snp_sl_34568	CU151095-146	1	85.02	-	…
		defoliation	*EB-5*	SL3.0CH05_64366933	OH08-7663	5	64.37	-	…
		defoliation	*EB-9*	solcap_snp_sl_29188	CU151095-146	9	66.95	-	…
		stem lesions	*EB-9*	solcap_snp_sl_29188	CU151095-146	9	66.95	-	…
	NILs (n=198)	defoliation	*EB-1.2*	solcap_snp_sl_34568	CU151095-146	1	85.02	-	…
		defoliation	*EB-5*	solcap_snp_sl_231	OH08-7663	5	64.96	-	…
		defoliation	*EB-9*	solcap_snp_sl_29188	CU151095-146	9	66.95	-	…
		stem lesions	*EB-9*	solcap_snp_sl_29188	CU151095-146	9	66.95	-	…

[a]Markers are those most strongly associated with resistance in any given experiment, with positions for the SL3.0 genome sequence (Tomato Genome Consortium 2012)

[b]95% confidence intervals were expanded to nearest markers mapped to the physical genome sequence

Table 2 *Continued*

Year	Population	Trait	QTL	…	Effect[c]	PVE	Statistic	df	P	LOD[d]
2017	F_2 (n=276)	stem lesions	*EB-1.1*	…	5.70	6.7	*F=11.27*	2	2.02×10^{-5}	4.98
		defoliation	*EB-1.2*	…	87.97	4.9	*F=12.79*	2	5.06×10^{-6}	5.67
		stem lesions	*EB-3*	…	3.31	2.1	*F=5.09*	2	6.78×10^{-3}	2.30
		stem lesions	*EB-4*	…	5.12	5.0	*F=12.14*	2	9.08×10^{-6}	5.35
		defoliation	*EB-4*	…	70.54	3.1	*F=7.64*	2	5.95×10^{-4}	3.45
		defoliation	*EB-5*	…	-126.35	11.0	*F=24.37*	2	2.03×10^{-10}	10.37
		defoliation	*EB-8*	…	-77.47	3.6	*F=9.11*	2	1.50×10^{-4}	4.09
		stem lesions	*EB-9*	…	12.00	26.4	*F=64.02*	2	~0	24.00
		defoliation	*EB-9*	…	99.69	7.3	*F=17.40*	2	8.12×10^{-8}	7.59
2018	$F_{2:3}$ (n=60)	defoliation	*EB-1.2*	…	81.24	18.0	χ^2=19.00	1	1.30×10^{-5}	-
		defoliation	*EB-5*	…	-144.64	23.4	χ^2=42.43	1	7.30×10^{-11}	-
		defoliation	*EB-9*	…	78-.79	18.3	χ^2=15.57	1	7.90×10^{-5}	-
		stem lesions	*EB-9*	…	18.06	93.1	χ^2=478.60	1	~0	-
	BC_1F_2 (n=219)	defoliation	*EB-1.2*	…	-0.46	0.5	*F*=0.09	2	0.92	-
		defoliation	*EB-5*	…	-88.14	20.4	*F*=40.30	2	1.73×10^{-15}	-
		defoliation	*EB-9*	…	48.64	7.8	*F*=15.38	2	6.00×10^{-7}	-
		stem lesions	*EB-9*	…	11.54	74.8	*F*=323.83	2	~0	-
2019	$F_{2:3}$ (n=90)	defoliation	*EB-1.2*	…	126.16	16.7	χ^2=21.90	2	1.76×10^{-5}	-
		defoliation	*EB-5*	…	-166.85	23.0	χ^2=57.90	1	2.76×10^{-14}	-
		defoliation	*EB-9*	…	140.88	15.7	χ^2=48.77	2	2.58×10^{-11}	-
		stem lesions	*EB-9*	…	1.10	82.9	χ^2=696.92	2	~0	-
	NILs (n=198)	defoliation	*EB-1.2*	…	64.90	1.1	χ^2=4.04	2	0.13	-
		defoliation	*EB-5*	…	-260.76	40.1	χ^2=148.27	2	~0	-
		defoliation	*EB-9*	…	150.32	15.0	χ^2=54.72	2	1.31×10^{-12}	-
		stem lesions	*EB-9*	…	10.23	68.7	χ^2=609.64	2	~0	-

[c]The average phenotypic effect of an allele substitution in an additive model. Positive estimates indicate CU151095-146 beneficial alleles

[d]Estimated 5% FDR (5000 permutations): LOD = 2.02 for stem lesions, LOD = 3.15 for defoliation

and SL3.0CH01_96165356 bounded the 95% positional confidence interval for *EB-1.1*, but because the latter SNP was the last polymorphic marker on the chromosome, a more appropriate interval might extend another ~2.3 Mbp to the end of chromosome 1. For this same reason, the chromosome 5 QTL, *EB-5*, could be confidently mapped to between SL3.0CH05_64158191 and the end of chromosome 5, about 2.3 Mbp away. The beneficial *EB-5* alleles contributed by the processing tomato OH08-7663 explained 11.0% of the defoliation variance, the largest of any single QTL. Once again, we identified a chromosome 9 QTL, centered on solcap_snp_sl_29188, that explained variance for both stem lesions (26.4%) and defoliation (7.3%).

Three QTL with weaker resistance effects were implicated in early blight resistance by linkage mapping alone (Table 2). These QTL included stem and foliar resistances on chromosomes 3 (*EB-3*) and 8 (*EB-8*), respectively. A chromosome 4 QTL (*EB-4*) was associated with stem and foliar resistance but had a positional confidence interval that spanned most of the chromosome. This imprecision is probably due to segregation distortion, as several chromosome 4 markers skewed towards the processing tomato (OH08-7663) alleles (Table S4, Supplement 4). Because these QTL had smaller effects, we did not genotypically select for them in our subsequent work, described below. However, a breeder hoping to incorporate greater early blight resistance into their breeding program might consider selecting for these QTL in addition to those on chromosomes 1, 5, and 9.

QTL validation in genotypically-selected $F_{2:3}$ families

To validate the most promising QTL, we genotypically selected F_2 individuals from the 2017 field and observed their $F_{2:3}$ progeny under early blight disease pressure in replicated field trials conducted in 2018 and 2019.

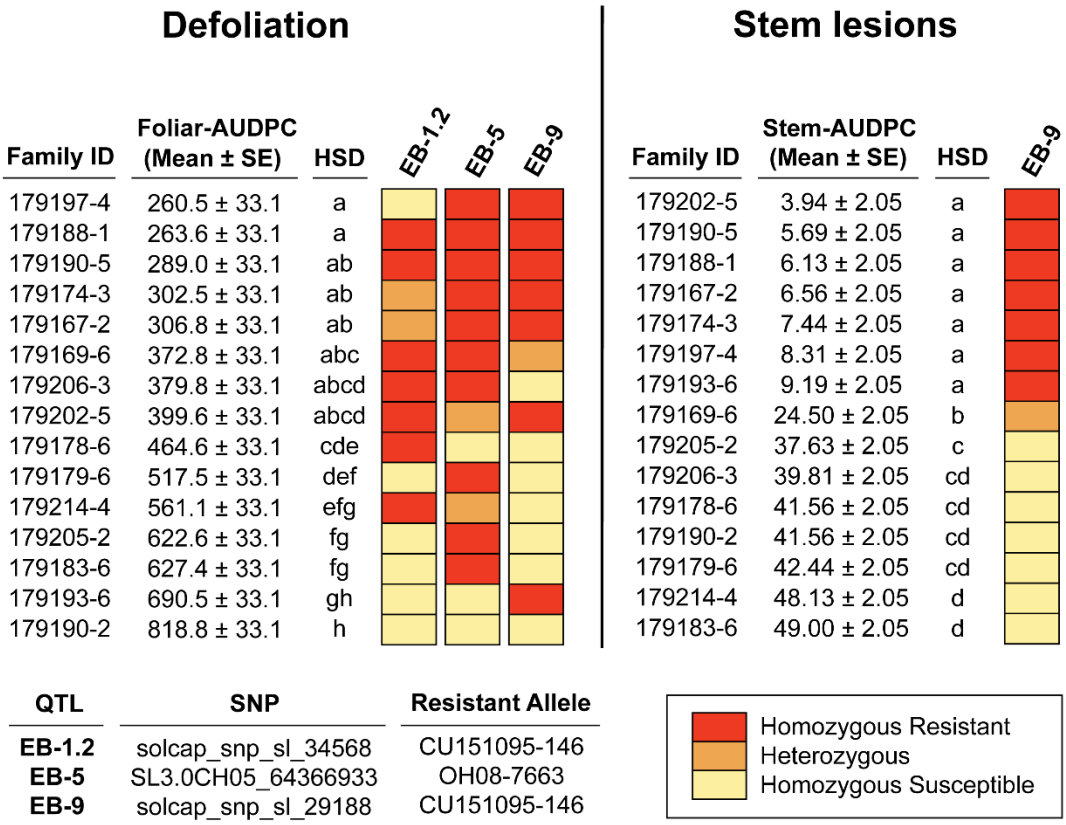
2018 $F_{2:3}$ trial: Tests of 15 $F_{2:3}$ families, differentially selected for combinations of alleles at *EB-1.2*, *EB-5,* and *EB-9*, showed that pyramiding resistance alleles enhanced early blight resistance. QTL effect estimates were larger than for the F_2 experiment, due to family replication and plot-based (as opposed to single plant) disease ratings (Table 2). The *EB-9* genotype explained 93.1% (p ~ 0) of the phenotypic variance for early blight stem lesions. Defoliation was again associated with *EB-1.2*, *EB-5*, and *EB-9*. Together, the main effects of these three QTL explained 70.0% of the defoliation variance (p ~ 0). Families selected for few or no resistance alleles had among the highest rates of defoliation, while those with multiple resistance alleles fared significantly better (Fig. 9). The combination of *EB-5* and *EB-9* offered levels of foliar resistance comparable to CU151095-146 and to $F_{2:3}$ families with all three QTL.

2019 $F_{2:3}$ trial: Other work in 2018 called into question the location and effect of the chromosome 1 QTL *EB-1.2* (see next section), prompting us to repeat the $F_{2:3}$ selection experiment in 2019. To confirm *EB-1.2*'s position, we screened 30 families that were recombinant along chromosome 1 and were homozygous at markers near *EB-5* and *EB-9*. A genome-wide association scan confirmed that the SNP most predictive of *EB-1.2* foliar resistance was solcap_snp_sl_34568 (Table 1, Table S5 in Supplement 4). Markers at *EB-5* and *EB-9* also remained predictive of foliar resistance. Together, the three markers for *EB-1.*2, *EB-5*, and *EB-9* explained 67.1% of the phenotypic variance for defoliation (p ~ 0).

QTL characterization in fresh market breeding lines.

To evaluate QTL effects in a commercially relevant background, we backcrossed *EB-1.*2, *EB-5*, and *EB-9* into the fresh market breeding line CU151095-146. Genome-wide marker-assisted selection facilitated the rapid recovery of market class and efficient introgression of resistance QTL, yielding a BC_1F_2 population with a mean recurrent parent genome content of 86.3 ± 6.0% (Data S6, Supplement 6), according to genetic marker data. After screening the

Fig. 9 Genotypic selection of F_2 parents at *EB-1.2*, *EB*-5, and *EB-9* yielded $F_{2:3}$ families differing for early blight stem lesions, $\chi^2 = 1509$, $p < .001$, and defoliation, $\chi^2 = 739$, $p < .001$ in a 2018 field trial. $F_{2:3}$ families are ordered from resistant to susceptible for early blight disease to allow visualization of QTL effects on early blight disease resistance. Families with the same letter under the Tukey honest significant difference (HSD) column are not different at the 95% confidence level. QTL-linked marker details are given in the table at bottom



QTL	SNP	Resistant Allele
EB-1.2	solcap_snp_sl_34568	CU151095-146
EB-5	SL3.0CH05_64366933	OH08-7663
EB-9	solcap_snp_sl_29188	CU151095-146

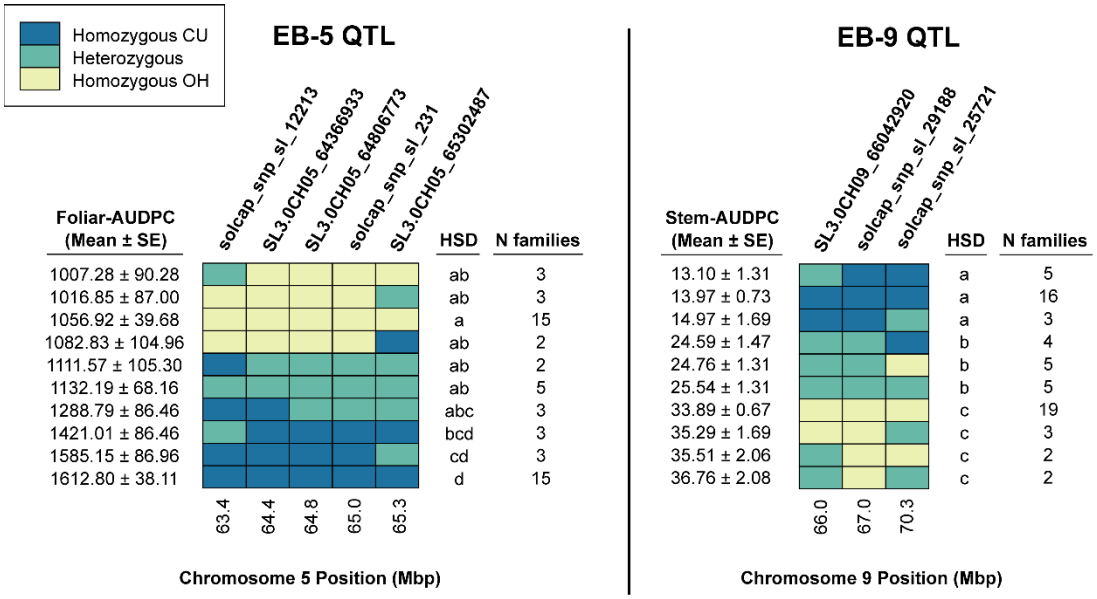
BC_1F_2 population under early blight disease pressure in 2018, we proceeded with additional inbreeding and genotypic selection (see *methods*) to create fresh market NILs for testing in a replicated trial the following year.

2018 BC_1F_2 population: QTL effects in the fresh-market background were consistent with prior findings for *EB-5* and *EB-9*, but not for *EB-1.2*. The *EB-9* marker solcap_snp_sl_29188 was highly predictive of stem lesions, explaining 74.8% of the phenotypic variance (Table 2). Curiously, there was little evidence of an association between *EB-1.*2 and defoliation, while there continued to be strong evidence for *EB-5* and *EB-9.* Foliar resistance at *EB-9* was recessive in this screen, contrasting with the incompletely dominant effects observed in other experiments (Fig. 8).

2019 NIL trial: Screening of fresh market breeding lines near-isogenic for QTL confirmed that *EB-5* and *EB-9* explained much of the variation for early blight resistance. Together, *EB-5* and *EB-9* explained 49.0% of the defoliation variance ($p \sim 0$), while *EB-9* alone explained 68.7% of the stem lesion variance (Table 2). We found weak evidence for foliar resistance at *EB-1.2*. Since three prior experiments mapped *EB-1.2* to a location within the recombinant chromosome 1 region (Table 1; Data S7, Supplement 6), we conclude that *EB-1.2*'s effect is population-dependent. It is not apparent what factors underlie *EB-1.2*'s inconsistent effect, or why this QTL was undetectable in our fresh market tomato background.

New markers developed for this experiment (Table S1, S2, Supplement 3) allowed us to refine the locations of *EB-5* and *EB-9* in the physical tomato genome. Analysis on genotype classes (2 or more NIL families with identical genotypes within previously-defined QTL intervals) gave strong evidence that the causative variant(s) underlying *EB-5* lie very near solcap_snp_sl_231, and between 64.4 and 65.3 Mbp on chromosome 5 of the SL3.0 tomato genome sequence (Fig. 10). Similarly, there was substantial evidence that *EB-9* resides

Fig. 10 NILs that fell into contrasting genotype classes facilitated a finer mapping of early blight resistance QTL. Markers at solcap_snp_sl_29188 (*EB-9*) and solcap_snp_sl_231 (*EB*-5) were most predictive of stem and foliar resistance, respectively. Genotype classes are ordered from most resistant to susceptible for early blight disease, and differences are indicated at the Tukey-adjusted 95% confidence level (HSD). *EB-5* (**left panel**) is most likely located between SL3.0CH05_64366933 and SL3.0CH05_65302487 on chromosome 5. *EB-9* (**right panel**) is between SL3.0CH09_66042920 and solcap_snp_sl_25721 on chromosome 9. Marker positions are for the *S. lycopersicum* 3.0 genome sequence (Tomato Genome Consortium 2012). "N families" is the number of families in each genotype class



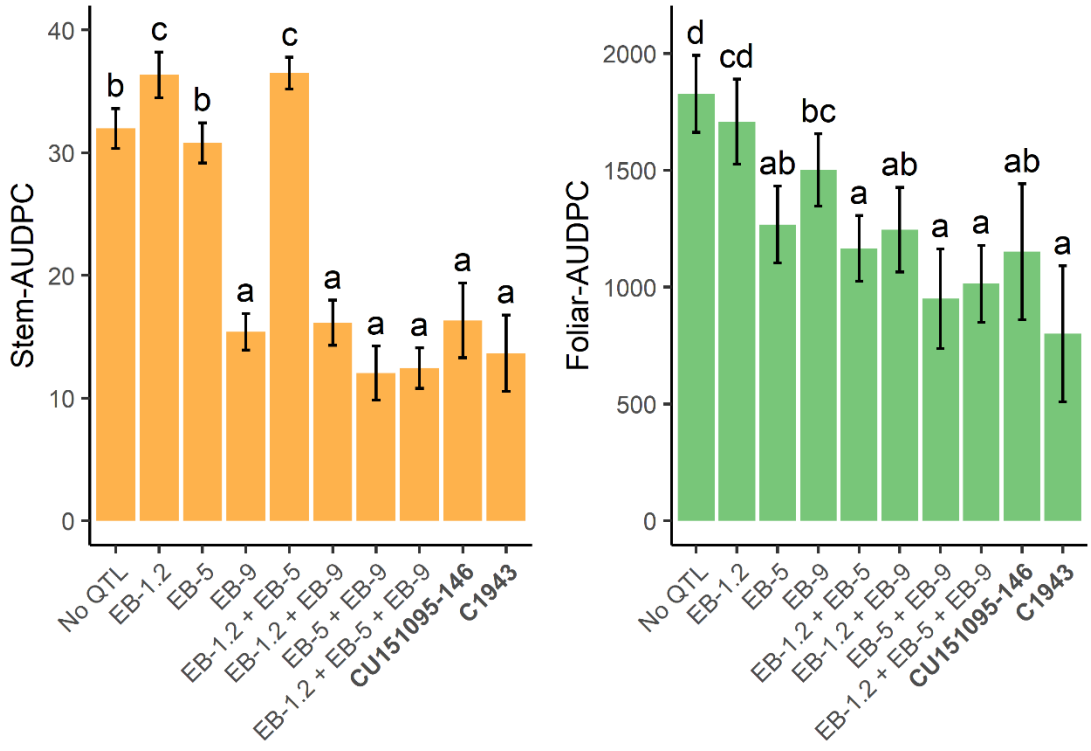
between 66.0 and 70.3 Mbp on chromosome 9, near solcap_snp_sl_29188. The *EB-9* interval could be further narrowed to 66.0 - 67.9 Mbp on chromosome 9 when considering the partially overlapping QTL intervals from the 2017 F_2 and the 2019 NIL populations. The polymorphic DNA sequences underlying these QTL-linked markers, and their updated positions in the SL4.0 genome, can be found in Supplement 3.

Pyramiding QTL in fresh market tomatoes enhanced foliar early blight resistance but offered little additional benefit to stem resistance beyond that of *EB-9* alone (Fig. 11). NILs homozygous for *EB-9* had stem-AUDPC scores equivalent to the ancestral resistance donor C1943. Interestingly, C1943 boasted the highest foliar resistance in the experiment, though the result was not statistically different from NILs homozygous for *EB-5*. This result suggests incomplete introgression of all C1943 resistance loci in modern lines or a reduction in resistance efficacy in modern lines with more compact growth habits. NILs homozygous for both *EB-5* and *EB-9* boasted a 36.7 ± 6.3% (t_{25} = 4.79, p = *.001*) reduction in foliar-AUDPC relative to NILs homozygous for just *EB-9*. The *EB-5* + *EB-9* NIL combination was not different at $p < .05$ from the *EB-5* NILs for foliar disease, despite a 25.1 ± 6.5% (t_{25} = 2.67, p = 0.18) reduction in foliar-AUDPC, unless the nonsignificant effect of *EB-1.2* was dropped from the QTL model (Table S8, Supplement 4).

Evidence for QTL x QTL interactions

There was little consistent evidence for important epistatic interactions between QTL in this work. Two marginally significant QTL x QTL interactions in the 2017 F_2 experiment explained 2.1% and 3.9% of the stem and foliar disease variance, respectively (Table S6, Supplement 4; Figs. S5, S6, Supplement 5). The 2018 $F_{2:3}$ trial gave evidence of interactions between all pairwise combinations of *EB-1.2*, *EB-5*, and *EB-9*. Here, the interaction between *EB-5* and *EB-9*, χ^2 = 4.88, p = .03 fit a synergistic resistance model (Fig. S7, Supplement 5). In

Fig. 11 Lines near-isogenic for early blight QTL were created in the background of the fresh market breeding line CU151095-146. Early blight stem lesions (**left**) and defoliation (**right**) are compared for NILs homozygous for zero (susceptible benchmark), one, two, or three early blight QTL, as determined from alleles at markers solcap_snp_sl_34568, solcap_snp_sl_231, and solcap_snp_sl_29188. CU151095-146 (recurrent parent) and C1943 (ancestral resistance donor) are shown at right (bolded). Error bars show 95% confidence intervals, and letters indicate significant differences at the Tukey-adjusted 95% confidence level. Note that early blight stem lesions and defoliation were rated using scales of differing magnitudes



contrast, the interactions between *EB-1.2* and *EB-9*, $\chi^2 = 5.31$, $p = .02$, and between *EB-1.2* and *EB-5*, $\chi^2 = 4.28$, $p = .04$, showed diminishing improvements to early blight resistance as resistant alleles were combined. Together, these interactions captured an additional 8.7% of the defoliation variance over the additive model. However, these findings were not corroborated by the 2018 BC_1F_2, 2019 $F_{2:3}$, or 2019 NIL trials, where no suggestive ($p < 0.2$) interactions were observed. Thus, early blight resistance QTL can show epistatic deviations from a quantitative inheritance model, but these interactions appear to be environmentally dependent and explain a relatively small proportion of the disease variance.

Comparison to classical and molecular studies

Our finding that stem lesion and collar rot resistance is simply inherited is consistent with prior classical and molecular studies. Reynard and Andrus (1944) thought a single recessive gene conferred stem resistance in 'Devon Surprise'. Similarly, Maiero et al. (1990b) postulated that one gene was responsible for collar rot resistance in C1943. Chaerani et al. (2007) found a QTL of comparable effect size to *EB-9* on Chromosome 9 of *S. arcanum* LA2157 that explained 34.5% of the total phenotypic variation for stem lesions. While we cannot rule out a distant *S. arcanum* source, their reported genetic distances suggest the QTL is on a different chromosome 9 arm than *EB-9*, and thus these QTL are probably distinct.

The nature of the *EB-9* locus is consistent with prior studies of C1943-derived resistance. We found *EB-9* to underlie stem, and to a lesser extent, foliar resistance – a characteristic feature of C1943 resistance (Barksdale and Stoner 1977). This feature is what enabled Gardner (1990) to predict foliar resistance in the field via an indoor stem seedling screen. Gardner concluded that stem and foliar resistances in C1943 must be due to the same or closely linked gene(s). In contrast, stem and foliar resistance are uncorrelated in other early blight resistance sources, including 67B833 and 68B134 (Andrus and Reynard 1942; Barksdale

and Stoner 1973). We postulate that the same gene or genes underlying *EB-9* are responsible for early blight resistance in C1943, although the resistance mechanism(s) remain unknown.

It is unclear whether *EB-9*'s simultaneous contributions to both stem and foliar resistance are due to a shared resistance mechanism or a correlation between stem health and overall plant fitness. Stem lesions at a petiole can cause the rapid collapse of whole leaves and kill entire branches. Damage to stems would presumably reduce the flow of nutrients throughout the plant vasculature, weakening the plant. Thus, infected stems might result in conditions favorable to the necrotrophic fungus, hastening defoliation. The strong and largely tissue-specific resistance offered by *EB-9* presents an intriguing system in which to study resistance to necrotrophy in tomato.

The genomic location and contributing parent of *EB-5* foliar resistance gives insight into this QTL's origin. Early blight resistance QTL near *EB-5* are found in every QTL study on the topic to date, with sources ranging from PI 126445 and NCEBR-1 (a *S. lycopersicum* breeding line with PI 126445 in its pedigree) to the wild accession LA2157 (Foolad et al. 2002; Zhang et al. 2003; Chaerani et al. 2007; Ashrafi et al. 2015). Therefore, *EB-5* falls within a hot spot for early blight resistance in *Solanum*. In our work, *EB-5* was derived from OH08-7663, where it is in coupling phase linkage with the bacterial spot (*Xanthomonas sp.*) race T1 resistance locus *Rx-3*, located approximately 2 Mbp away on the same chromosome. *Rx-3* was introgressed into OH08-7663 from Hawaii 7998 (H7998), a tomato with unknown *S. pimpinellifolium* introgressions (Liabeuf 2016). Therefore, we hypothesize H7998, and possibly *S. pimpinellifolium*, to be the source of *EB-5*. This linkage may also explain why Yang and Francis (2005) found a correlation between bacterial spot and early blight resistance in materials derived from H7998.

Our serendipitous discovery of a large-effect foliar resistance QTL in OH08-7663 is intriguing since a breeder or pathologist would be unlikely to characterize OH08-7663 as highly resistant to early blight, as this line can sustain heavy foliar losses in the field. However, our preliminary greenhouse work observed foliar resistance in OH08-7663, which we then repeatedly mapped to the *EB-5* locus in the field. This observation suggests that it would be beneficial to look for early blight resistance introgressions in other adapted materials displaying potentially valuable, if not exceptional, resistance. Such introgressions would also have a lower potential for linkage drag compared with introgressions from wild species. There likely are additional cases of undetected but desirable genes being unknowingly introgressed into tomato from wild species during interspecific breeding for unrelated traits. Because resistance to necrotrophic pathogens such as early blight is often incomplete, it is conceivable that such resistances could have persisted, undetected, during breeding for environmental adaptivity and stability.

The sources of the remaining QTL identified in this paper are not as clear as for *EB-5* and *EB-9*. *EB-1.2*'s location is similar to that of previously identified foliar resistances in PI 126445 and LA2157 (Foolad et al. 2002; Chaerani et al. 2007). Because PI 126445 is in the distant pedigree of the CU fresh-market tomatoes, we suspect that *EB-1.2* may come from this *S. habrochaites* accession. A pericentromeric chromosome 4 QTL for foliar resistance in PI 126445 (Zhang et al. 2003) may underlie our *EB-4* QTL. It is also possible that *EB-4* is an artifact of segregation distortion resultant from *gamete eliminator* (Rick 1965), which is present in OH08-7663 (D. Francis, Per Comm). *EB-8*, with resistance derived from OH08-7663, is close to a previously identified foliar resistance QTL in PI 126445 (Foolad et al. 2002; Zhang et al. 2003). However, this accession is not in the known pedigree of Ohio processing tomatoes.

A common discrepancy among early blight genetic studies concerns the measurement of resistance gene action and inheritance. Maiero et al. (1990b) found C1943 collar rot resistance to be incompletely recessive, while we observed mid-parent stem resistance at *EB-9*. Barksdale and Stoner (1977) thought recessive genes conferred C1943 foliar resistance, Maiero (1989) considered inheritance additive with epistasis, and we found foliar resistance to be primarily mediated by a small number of additive QTL. These discrepancies are not surprising given the variety of experimental methods used in the early blight literature and the potentially complex nature of resistance to necrotrophic pathogens. The challenge that environmental variability poses to phenotypic selection of early blight resistance underlies the importance of developing robust genotypic selection strategies.

It is notable that prior QTL mapping works found early blight resistance to be inherited in a highly quantitative manner, while we found it to be oligogenic and possibly even simply inherited. The greatest distinction between our work and the prior QTL mapping efforts of others is that we mapped resistance within a cultivated background, while others employed interspecific crosses between cultivated tomatoes and wild species (Ashrafi and Foolad 2015; Chaerani et al. 2007; Foolad et al. 2002; Foolad et al. 2005; Zhang et al. 2003). Such wide crosses would have been segregating for domestication and improvement loci that can influence the phenotypic measurements of early blight disease. This segregation alone could underly the discrepancy between our findings. Alternatively, it is possible that our parental lines share many early blight resistance loci. However, the significant degree of transgressive segregation for susceptibility that we observed in progeny populations suggests otherwise.

Breeders aiming to incorporate early blight resistance at minimal cost could focus on deploying *EB-5* and *EB-9*. These two QTL offered consistent and complimentary resistances to early blight stem and foliar disease phases. Selection for these QTL would also present

opportunities to incorporate linked resistances to other pathogens. *EB-5* is linked in coupling phase to *Rx-3* in the inbred line OH08-7663 (recombination distance unknown), and *EB-9* is loosely linked to *Ph-3* (*Phytophthora infestans*) in CU151095-146 and NC 1 CELBR (40 cM in our map). *EB-1.1* and *EB-1.2* might also enhance early blight resistance, but our data did not find resistance effects in all genetic backgrounds. In general, resistance QTL were incompletely dominant, so breeders developing hybrids involving a single resistance donor would benefit from deploying multiple resistance QTL.

Conclusion

In this work, we identified and characterized useful early blight resistance introgressions in contemporary tomato breeding lines and presented molecular markers that can be used to select for these introgressions. A logical next step might be to define these introgressions. A comparison of the genome sequences of modern resistant tomatoes and the putative ancestral donors should be capable of defining the approximate introgression boundaries. Such an analysis would facilitate the identification of polymorphic markers that are unique or distinctive to ancestral sources, and thus would be portable across breeding programs and germplasm

REFERENCES

Andersen B, Dongo A, Pryor BM (2008) Secondary metabolite profiling of *Alternaria dauci*, *A. porri*, *A. solani*, and *A. tomatophila*. Mycological Research 112:241–250. https://doi.org/10.1016/j.mycres.2007.09.004

Andrus CF, Reynard GB, Jorgensen H, Eades J (1942a) Collar rot resistance in tomatoes. Journal of Agr. Research 65(7):339-346

Andrus CF, Reynard GB, Wade BL (1942b) Relative resistance of tomato varieties, selections, and crosses to defoliation by *Alternaria* and *Stemphylium*. USDA Circular No. 652:1-23

Arends D, Prins P, Jansen RC, Broman KW (2010) R/qtl: High-throughput multiple QTL mapping. Bioinformatics 26:2990-2992. https://doi.org/10.1093/bioinformatics/btq565

Ashrafi H, Foolad MR (2015) Characterization of early blight resistance in a recombinant inbred line population of tomato: II. Identification of QTLs and their co-localization with candidate resistance genes. Adv. Stud. in Bio. 7(4):149-168. http://dx.doi.org/10.12988/asb.2015.41163

Barksdale TH (1969) Resistance of tomato seedlings to early blight. Phytopathology 59:443-446.

Barksdale TH (1971) Field evaluation for tomato early blight resistance. Plant Dis. Rep. 55:807-809.

Barksdale TH, Stoner AK (1973) Segregation for horizontal resistance to tomato early blight. Plant Dis. Rep. Vol 57(11):964-965

Barksdale TH, Stoner AK (1977) A study of the inheritance of tomato early blight resistance. Plant Dis. Rep. Vol 61(2):63-65

Barratt RW, Richards MC (1944) Physiological maturity in relation to *Alternaria* blight in the tomato. Phytopathology 34:997

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

Blum M, Boehler M, Randall E, Young V, Csukai M, Kraus S, Moulin F, Scalliet G, Avrova AO, Whisson SC, Fonne-Pfister R (2010) Mandipropamid targets the cellulose synthase-like PiCesA3 to inhibit cell wall biosynthesis in the Oomycete plant pathogen, *Phytophthora infestans*. Molecular Plant Pathology 11(2):227-243. https://doi.org/10.1111/j.1364-3703.2009.00604.x

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635. https://doi.org/10.1093/bioinformatics/btm308

Broman KW, Wu H, Sen Ś, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889-890. https://doi.org/10.1093/bioinformatics/btg112

Browning BL, Zhou Y, Browning SR (2018) A one-penny imputed genome from next generation reference panels. Am J Hum Genet 103(3):338-348. http://doi.org/10.1016/j.ajhg.2018.07.015

Brun S, Madrid H, Gerrits van den Ende, Andersen B, Marinach-Patrice C, Mazier D, Sybren De Hoog G (2013) Multilocus phylogeny and MALDI-TOF analysis of the plant pathogenic species *Alternaria dauci* and relatives. Fungal Biology 117: 32–40. https://doi.org/10.1016/j.funbio.2012.11.003

Chaerani R, Smulders MJM, van der Linden CG, Vosman B, Stam P, Voorrips RE (2007) QTL identification for early blight resistance (*Alternaria solani*) in a *Solanum lycopersicum* x *S. arcanum* cross. Theoretical and Applied Genetics 114:439-450. https://doi.org/10.1007/s00122-006-0442-8

Chen ZX, Mutschler MA (1988) Tolerance in tomato to early blight (*Alternaria solani* (Ell. And Martin) Sor.). Tomato Gen. Coop. Rep. 38:13-15

Darvasi A, Soller M (1992) Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. Theoretical and Applied Genetics 85(2-3):353-359

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 19(1):11-15

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

Foolad MR, Sharma A, Ashrafi H, Lin G (2005) Genetics of early blight resistance in tomato. Acta Horticulturae 695:397-406. https://doi.org/10.17660/ActaHortic.2005.695.49

Foolad MR, Zhang LP, Khan AA, Niño-Liu D, Lin GY (2002) Identification of QTLs for early blight (*Alternaria solani*) resistance in tomato using backcross populations of a *Lycopersicon esculentum* x *L. hirsutum* cross. Theoretical and Applied Genetics 104:945-958. https://doi.org/10.1007/s00122-002-0870-z

Franjul CA (1989) Response of tomato, *Lycopersicon esculentum* to foliage and stem infection by early blight, *Alternaria solani* (Ellis and Martin) Jones and Grout. Dissertation, Cornell University

Frazer JT (2002) Two species of *Alternaria* cause early blight of potato (*Solanum tuberosum*) and tomato (*Lycopersicon esculentum*). Dissertation, Cornell University

Furuta T, Ashikari M, Jena KK, Doi K, Reuscher S (2017) Adapting genotyping-by-sequencing for rice F_2 populations. G3: Genes, Genomes, Genetics 7(3):881-893. https://doi.org/10.1534/g3.116.038190

Gannibal PB, Orina AS, Mironenko NV, Levitin MM (2014) Differentiation of the closely related species, *Alternaria solani* and *A. tomatophila*, by molecular and morphological features and aggressiveness. European Journal of Plant Pathology 139:609–623. https://doi.org/10.1007/s10658-014-0417-6

Gardner RG (1988) NC EBR-1 and NC EBR-2 early blight resistant tomato breeding lines. HortScience 23(4):779-781

Gardner RG (1990) Greenhouse disease screen facilitates breeding resistance to tomato early blight. HortScience 25(2):222-223. https://doi.org/10.21273/HORTSCI.25.2.222

Gardner RG, Panthee DR (2010) NC 1 CELBR and NC 2 CELBR: early blight and late blight-resistant fresh market tomato breeding lines. HortScience 45(6):975-976. https://doi.org/10.21273/HORTSCI.45.6.975

Gardner RG, Shoemaker PB (1999) 'Mountain Supreme' early blight-resistant hybrid tomato and its parents, NC EBR-3 and NC EBR-4. HortScience 34(4):745-746. https://doi.org/10.21273/HORTSCI.34.4.745

Gill U, Scott JW, Shekasteband R, Ogundiwin E, Schuit C, Francis DM, Sim S-C, Smith H, Hutton SF (2019) Ty-6, A major *Begomovirus* resistance gene on chromosome 10, is effective against tomato yellow leaf curl virus and tomato mottle virus. Theoretical and Applied Genetics. 132:1543–1554. https://doi.org/10.1007/s00122-019-03298-0

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014) TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLOS ONE 9(2):e90346. https://doi.org/10.1371/journal.pone.0090346

Hamilton JP, Sim SC, Stoffel K, Van Deynze A, Buell CR, Francis DM (2012) Single nucleotide polymorphism discovery in cultivated tomato via sequencing by synthesis. The Plant Genome 5:17-29. https://doi.org/10.3835/plantgenome2011.12.0033

Holland PW, Welsch RE (1977) Robust regression using iteratively reweighted least-squares. Communications in Statistics - Theory and Methods 6(9)813-827. https://doi.org/10.1080/03610927708827533

Hospital F, Chevalet C, Mulsant P (1992) Using markers in gene introgression breeding programs. Genetics 132(4):1199-1210

Lacy ML (1973) Control of septoria leafspot of celery with systemic and nonsystemic fungicides. The Plant Dis. Rep. 57:425-428

Length R (2017) emmeans: estimated marginal means, aka least-squares means. R package version 1.0. https://CRAN.R-project.org/package=emmeans. Accessed 23 June 2019

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-60. https://doi.org/10.1093/bioinformatics/btp324

Liabeuf D (2016) Development of processing tomato lines resistant to *Xanthomonas gardneri*: from screening to breeding. Dissertation, The Ohio State University

Lebowitz, RJ, Soller M, Beckmann, JS (1987) Trait-based analyses for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. Theoretical and Applied Genetics 73(4):556-562. https://doi.org/10.1007/bf00289194

Locke SB (1949) Resistance to early blight and *Septoria* leaf spot in the genus *Lycopersicon*. Phytopathology 39:829-836.

Madden LV, Hughes G, van den Bosch F (2007) The study of plant disease epidemics. The American Phytopathological Society Press, St. Paul

Money D, Gardner K, Migicovsky Z, Schwaninger H, Zhong GY, Myles S (2015) LinkImpute: fast and accurate genotype imputation for nonmodel organisms. G3: Genes, Genomes, Genetics 5(11):2383-2390. https://doi.org/10.1534/g3.115.021667

Maiero M, Ng TJ, Barksdale TH (1989) Combining ability estimates for early blight resistance in tomato. Journal of the American Society of Horticultural Science 114(1):118-121

Maiero M, Ng TJ, Barksdale TH (1990a) Genetic resistance to early blight in tomato breeding lines. HortScience 25(3):344-346

Maiero M, Ng TJ, Barksdale TH (1990b) Inheritance of collar rot resistance in the tomato breeding lines C1943 and NC EBR-2. Phytopathology 80(12):1365-1368

Martin FW, Hepperly P (1987) Sources of resistance to early blight, *Alternaria solani*, and transfer to tomato, *Lycopersicon esculentum*. Journal Agric. Univ. P.R. 71(1):85-95

Mitani S, Araki S, Takii Y, Ohshima T, Matsuo N, Miyoshi H (2001a) The Biochemical Mode of Action of the Novel Selective Fungicide Cyazofamid: Specific Inhibition of Mitochondrial Complex III in *Phythium spinosum*. Pesticide Biochemistry and Physiology 71(2):107-115. https://doi.org/10.1006/pest.2001.2569

Mitani S, Araki S, Yamaguchi T, Takii Y, Ohshima T, Matsuo N (2001b) Antifungal activity of the novel fungicide Cyazofamid against *Phytophthora infestans* and other plant pathogenic fungi *in vitro*. Pesticide Biochemistry and Physiology 70(2):92-99. https://doi.org/10.1006/pest.2001.2541

Nakagawa S, Johnson PCD, Schielzeth H (2017) The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. Journal Royal Society Interface 14:20170213. https://doi.org/10.1098/rsif.2017.0213

Nash AF (1986) Tomato early blight resistance derived from *Lycopersicon hirsutum* Humb. and Bonpl. P.I. 126445. Dissertation, North Carolina State University

Nash AF, Gardner RG (1988a) Heritability of tomato early blight resistance derived from *Lycopersicon hirsutum* PI 126445. Journal Amer. Soc. Hort. Sci. 113(2):264-268

Nash AF, Gardner RG (1988b) Tomato early blight resistance in a breeding line derived from *Lycopersicon hirsutum* PI 126445. The Amer. Phyt. Soc. 72:206-209

Patterson JM, Nokes SE, Bennett MA, Riedel RE (2001) Evaluation of residual chlorothalonil levels on processing tomato foliage using the TOM-CAST Spray Program. Applied Engineering in Agriculture 17(4):445-448. https://doi.org/10.13031/2013.6459

Panthee DR, Gardner RG (2010) 'Mountain Merit': A late blight-resistant large-fruited tomato hybrid. HortScience 45(10):1547-1548. https://doi.org/10.21273/HORTSCI.45.10.1547

Pasteris RJ, Hanagan MA, Bisaha JJ, Finkelstein BL, Hoffman LE, Gregory V, Andreassi JL, Sweigard JA, Klyashchitsky BA, Henry YT, Berger RA (2016) Discovery of Oxathiapiprolin, a new Oomycete fungicide that targets an oxysterol binding protein. Bioorganic & Medicinal Chemistry 24(3):354-361. https://doi.org/10.1016/j.bmc.2015.07.064

Percheur RJ (2009) Ohio Vegetable Production Guide, The Ohio State University Extension, Bulletin 672.

Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC, Nelson RJ (2009) Shades of gray: the world of quantitative disease resistance. Trends in Plant Sci. 14(1):1-29. https://doi.org/10.1016/j.tplants.2008.10.006

Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLOS ONE 7(2):e32253. https://doi.org/10.1371/journal.pone.0032253

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rick C (1965) Abortion of male and female gametes in the tomato determined by allelic interaction. Genetics 53: 85-96

Rodrigues TTMS, Berbee ML, Simmons EG, Cardoso CR, Reis A, Maffia LA, Mizubuti ESG (2010) First report of *Alternaria tomatophila* and *A. grandis* causing early blight on tomato and potato in Brazil. New Disease Reports 22:28. http://dx.doi.org/10.5197/j.2044-0588.2010.022.028

Rotem J (1994) Crop and Yield Losses. In: The genus *Alternaria*: biology, epidemiology, and pathogenicity. The Amer. Phyto. Society, St. Paul, pp 185-200

Rotem J, Reichert I (1964) Dew - a principal moisture factor enabling early blight epidemics in a semi-arid region of Israel. Plant Dis. Rep. 48:211-215

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

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

Simmons EG (2000) *Alternaria* themes and variations: species on *Solanaceae*. *Mycotaxon* 75:1-115

Small IM, Joseph L, Fry WE (2015) Development and implementation of the BlightPro decision support system for potato and tomato late blight management. Comp. Elec. In Agric. 115:57-65. https://doi.org/10.1016/j.compag.2015.05.010

Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature. 485(7400):635–641. https://doi.org/10.1038/nature11119

Toquin V, Barja F, Sirven C, Bffa R (2007) Fluopicolide, a new anti-oomycetes fungicide with a new mode of action inducing perturbation of a spectrin-like protein. In: Modern crop protection compounds. Wiley, Weinheim, pp 675-682. https://doi.org/10.1002/9783527619580.ch19

Woudenberg JHC, Truter M, Groenewald JZ, Crous PW (2014) Large-spored *Alternaria* pathogens in section *Porri* disentangled. Stud. Mycol. 79:1–47. https://doi.org/10.1016/j.simyco.2014.07.003

Yang W, Francis DM (2005) Marker-assisted selection for combining resistance to bacterial spot and bacterial speck in tomato. J. Am. Soc. Hortic. Sci. 130:716-721. https://doi.org/10.21273/JASHS.130.5.716

Zhang LP, Lin GY, Niño-Liu D, Foolad, MR (2003) Mapping QTLs conferring early blight (*Alternaria solani*) resistance in a *Lycopersicon esculentum* x *L. hirsutum cross* by selective genotyping. Mol. Breeding 12:3-19. https://doi.org/10.1023/A:1025434319940

Zhou X, Stephens M (2014) Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nature Methods 11(4):407–409. https://doi.org/10.1038/nmeth.2848

Zitter TA, Drennan JL (2007) Using host resistance and light fungicides to control early blight on tomato. Plant Dis. Manag. Rep. 2:60

Zitter TA, Drennan JL (2008) Using host resistance and lite fungicides to control early blight and *Septoria* leaf spot on tomato. Plant Dis. Manag. Rep. 2:61

Zitter TA, Drennan JL, Mutschler MA, Kim MJ (2005) Control of early blight of tomato with genetic resistance and conventional and biological sprays. Acta Horticulturae 695

Zitter TA, Zitter SM, Mutschler MA, McKay SP (2011) Using host resistance and reduced-risk fungicides to control early blight and *Septoria* leaf spot on tomato. Plant Dis. Manag. Rep. 5:076

CHAPTER 3

WHOLE-GENOME INTROGRESSION DETECTION AND HAPLOTYPE ANALYSIS REVEALS EARLY BLIGHT RESISTANCE IN MODERN TOMATO BREEDING LINES TRACES TO 'DEVON SURPRISE' AND 'HAWAII 7998'

Anderson TA*, Sudermann MA*, Francis DM, Smart C, DeJong DM, Mutschler, MA

* Co-first author

Affiliations:

Anderson TA, DeJong DM, Mutschler, MA, Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, 240 Emerson Hall, Ithaca, NY 14853

Sudermann MA, Smart C, Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, 135 Plant Science Building, Ithaca, NY

Francis DM, Department of Horticulture and Crop Science, The Ohio State University, 1680 Madison Avenue, Wooster, OH 44691

Acknowledgements:

TAA and MAS contributed equally to all aspects of the design, execution, analysis, and writing of this chapter. Revisions were offered by DM Francis, C Smart, DM DeJong, and MA Mutschler. Thanks to Dr. Randy Gardner, NCSU, and Dr. Esther van der Knaap, UGA, for providing seed of relevant tomato breeding lines and accessions. Thank you to Dr. Gregory Martin, BTI, for providing sequence for CU151011-146.

Abstract

We sequenced the genomes of landmark accessions with early blight disease resistance and used efficient inference of local ancestry to define cryptic introgressions underlying disease resistance in modern tomato (*Solanum lycopersicum* L.) breeding lines. The genome sequences for eleven accessions from the historical early blight resistance breeding pedigrees are reported, including NC 1 CELBR, *Campbell 1943*, and *Solanum habrochaites* PI 126445. We trace the source of the early blight resistance quantitative trait locus (QTL) *EB-9* backwards from *Campbell 1943* to the *Ailsa Craig*-derived vintage tomato *Devon Surprise*. The foliar resistance QTL *EB-5* was traced back to *Hawaii 7998*. We failed to detect strong evidence for PI 126445 introgressions in modern breeding lines, despite it being a commonly cited resistance source. Identification of the shared ancestral haplotypes along the historical breeding pedigree enabled fine mapping of *EB-5* and *EB-9* and the identification of candidate variants that could underlie early blight disease resistance. We surveyed the re-sequenced genomes of an additional 769 sequenced accessions, predicting *EB-9* resistance in several vintage tomatoes, several accessions of *S. lycopersicum* var. *cerasiforme,* and *S. pimpinellifolium* PI 37009. In contrast, there was little evidence of foliar *EB-5* resistance among sequenced accessions. We also defined introgressions underlying bacterial spot resistance at *Rx-3* and *QTL-11* in modern breeding lines. Our work makes it easier to breed for effective resistance to early blight disease in cultivated tomato and demonstrates an efficient method to leverage growing sequence resources for value prediction in publicly available germplasm.

CHAPTER 3 SUPPLEMENTAL MATERIALS

Supplement 7 Supplementary tables (.zip)

Table S1 Detailed Information on all tomato lines included in this study, including taxonomic information, PCA coordinates, and cluster membership (.csv)

Table S2 Full list of gene annotations that fall within the refined *EB-5* interval SL4.0ch05:62566094 - SL4.0ch05:63401898 (.csv)

Table S3 Full list of gene annotations that fall within the refined *EB-9* interval SL4.0ch09:62599611 - SL4.0ch09:62943349 (.csv)

Table S4 Table of high-quality SNP markers within the *EB-5* ancestral haplotype from *Hawaii 7998* (.csv)

Table S5 Table of high-quality SNP makers within the *EB-9* ancestral haplotype from *Devon Surprise* (.csv)

Supplement 8 Python 3 code for haplotype clustering and R code for visualizations (.zip)

README.txt - instructions and information on the contents of supplement 8

cluster_haplotypes.py - Script for generating windowed haplotype clusters

parallelize_haplotype_clustering.sh – A bash/shell script for parallelizing by chromosome

haplotype_visualization.Rmd - Haplotype cluster analysis and visualization script and example outputs as html (haplotype_visualization.html)

example_file – several example files provided for tutorial purposes (see README.txt)

Supplement 9 Supplementary figures S1 to S11

Introduction

Early blight is caused by the necrotrophic fungus *A. linariae* (syn. *Alternaria tomatophila*). It is a widespread and damaging disease of tomato (*Solanum lycopersicum* L.) (Rotem 1994; Woudenberg et al. 2014). The disease can girdle the stems of young transplants causing plant collapse in a disease phase known as collar rot (Pritchard and Porte 1921). On mature plants, lesions can form on the fruit, stems, or leaves, causing branch collapse, defoliation, and reductions in marketable fruit yields under warm, humid conditions (Barksdale 1971). Prior studies found early blight disease resistance in tomato to be quantitatively inherited, suggesting it is unlikely to conform to the classic gene-for-gene hypothesis (Barksdale and Stoner 1973, 1977; Martin and Hepperly 1987; Nash and Gardner 1988; Maiero et al. 1989, 1990). However, several tomato breeding lines have demonstrated a substantial degree of resistance to early blight (Adhikari et al. 2017). The resistance in these lines can be paired with modest chemical control to minimize early blight yield losses in the field (Zitter et al. 2005; Zitter and Drennan 2007, 2008). Nevertheless, difficulties in obtaining reliable phenotype data and the complex resistance inheritance make the early blight pathosystem particularly well-suited to efficient marker-based selection (Adhikari et al. 2017).

Much of the early blight resistance in cultivated tomato is thought to originate from the mid-century canning tomato breeding line *Campbell 1943* and the wild *Solanum habrochaites* accession PI 126445. *Campbell 1943* exhibits strong resistance to early blight stem lesions and collar rot and moderate resistance to defoliation in the field (Gardner 1990; Chapter 2). While the source of resistance in Campbell 1943 is unknown, it is possible that Dr. George B. Reynard of Campbell's tomato seed leveraged his knowledge of a simply-inherited collar rot resistance in the tomato *Devon Surprise*, from his prior role as a USDA researcher in the early 1940s, to create *Campbell 1943* (Reynard and Andrus 1944, Anon. 2016). An appendix

to the Interim Report of the Committee on Varietal Pedigrees of the Tomato Genetics Cooperative suggests that *Devon Surprise* was itself a 'mutant' of the vintage variety *Ailsa Craig*. Subsequent breeding efforts by Dr. Randolph Gardner at North Carolina State University (NCSU) transferred the *Campbell 1943* resistance into the fresh market tomato background of NC EBR 2 in the 1980s (Barksdale 1971) (Gardner 1988). Separately, Gardner introgressed foliar early blight resistance from *Solanum habrochaites* PI 126445, into fresh market tomato, creating NC EBR 1 (Gardner 1988). Resistance from NC EBR 1 and NC EBR 2 was then combined in NC 1 CELBR, which has been used widely and is the source of stem and foliar early blight resistance for several Cornell tomato breeding lines (Gardner and Panthee 2010).

We recently characterized quantitative trait loci (QTL) conferring partial early blight resistance in Cornell fresh market and The Ohio State (OSU) processing tomato breeding lines (Chapter 2). Our work suggested a somewhat oligogenic resistance inheritance in the cultivated background, where a large proportion of the phenotypic variance for resistance can be explained by a relatively small number of QTL. These QTL represent a readily deployable resistance resource for breeding programs, but further work is needed to resolve their genomic positions and to develop reliable predictive markers. Two of these QTL, *EB-5* and *EB-9*, conferred appreciable levels of early blight resistance in the cultivated background and were effective across all growing environments. The *EB-5* locus was associated with resistance to defoliation, while *EB-9* was associated with stem lesion, collar rot, and moderate foliar resistance. We hypothesized that *EB-9* traces to *Campbell 1943* and possibly *Devon Surprise*, while *EB-5* traces to the unimproved fresh market tomato breeding line *Hawaii 7998.* A third QTL on chromosome 1, *EB-1.2*, was also associated with early blight defoliation in some populations, but subsequent evaluations suggested the QTL was associated with an increase in plant size that might explain the connection with early blight resistance in the processing

tomato background (Anderson T, unpublished observations). We speculated that *EB-1.2* could be derived from PI 126445.

Our goal was to use comparative genomics to clarify the ancestry of early blight resistance in modern tomatoes and to define the ancestral introgressions underlying *EB-5* and *EB-9.* Decades of breeding for early blight resistance should have whittled down the resistance introgressions from the ancestral sources. Thus, we hypothesized that identification of the shared ancestral introgression among resistant tomatoes would enable fine-mapping of *EB-5* and *EB-9*. To do this, we sequenced the genomes of nine landmark early blight resistant tomato breeding lines and the wild PI 126445 resistance donor and analyzed these data together with 769 existing genome sequences for a variety of tomato and wild species accessions. We developed a flexible and computationally efficient protocol for identifying introgressions through the detection of homologous haplotypes in sequenced accessions. Using this method, we delineated the boundaries of the ancestral resistance introgressions and predicted early blight resistance in publicly available accessions. The refined QTL intervals helped to identify ancestry-specific molecular markers and possible causative resistance loci for *EB-5* and *EB-9* that will empower future molecular breeding efforts for early blight resistance.

Materials and Methods

Plant material

Nine tomatoes representing over 80 years of breeding for early blight resistance were sequenced for our work (Table 3). The Cornell tomato breeding program provided seed of the elite tomato breeding lines CU151095-146, CU151011-146, CU191357, and CU201041. Tomatoes CU151095-146 and CU151011-146 are full-sib fresh market inbreds with a complex pedigree and phenotypically selected early blight resistance tracing to NC 1 CELBR. CU191357 has additional early blight resistance (*EB-5*) from OH08-7663 in the CU151095-146 background. CU201041 is a processing tomato breeding line in the background of OH08-7663 with introgressed resistance (*EB-1.1*|*EB-1.2* + *EB-9*) from CU151095-146. Seed of *Campbell 1943* and the fresh market lines NC EBR 1, NC EBR 2, and NC 1 CELBR were obtained from Dr. Gardner at NCSU and were described previously. Seed of *S. habrochaites* PI 126445 and *Devon Surprise* were requested from the Northeast Regional PI Station USDA, ARS Plant Genetic Resources Unit in Geneva, NY. An unrelated Cornell line with known introgressions from *Solanum pennellii* LA0716, CU17NBL, was grown and sequenced for use in the validation of introgression detection methods. Plants were seeded into 10 cm pots under 14h light at 27°C, 10h dark at 22°C at the Guterman Bioclimatic Lab in Ithaca, NY.

Acquisition of whole genome sequence de novo and from existing studies

DNA was extracted from 85g of leaf tissue from four-week-old plants using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA). Multiplexed genomic DNA libraries were prepared for sequencing using the Nextera Flex platform (Illumina Inc., San Diego, CA, USA) and sequenced on a NextSeq500 (paired-end 2 x 150bp) by the Cornell Institute of Biotechnology Genomics Facility. Two plants of NC 1 CELBR were sequenced because we previously observed residual heterozygosity in the seed lot. These sequences are to be deposited in the National

Center for Biotechnology Information (NCBI) Sequence Read Archives (SRA) [*see upcoming peer-reviewed publication*]. Raw sequences for *Hawaii 7998*, OH08-7663, OH7536, and OH88119 were provided by Dr. Francis, OSU. Sequence for CU151011-146 was provided by Dr. Martin of the Boyce Thompson Institute for Plant Research. An additional 764 sequences for tomato lines, F_1 hybrids, and wild species accessions were obtained from the NCBI SRA deposits SRP094624, SRP045767, ERP004618, and SRP150040 (Table S1, Supplement 7) (100 Tomato Genome Sequencing Consortium et al. 2014; Lin et al. 2014; Tieman et al. 2017; Gao et al. 2019).

Sequence handling

Raw fastq files were mapped to the SL4.0 *Heinz 1706* reference tomato genome using BWA-MEM 0.7.17 (Li 2013; Fernandez-Pozo et al. 2015; Hosmani et al. 2019). Variants were called jointly for all 780 sequences with the GATK4 HaplotypeCaller 4.1.4.1 and filtered using VCFtools 0.1.13 and BCFtools (Li et al. 2009; Danecek et al. 2011; Poplin et al. 2018). Single nucleotide polymorphisms (SNPS) were retained at sites for which the proportion of missing data was less than 20% and where the minor allele frequency exceeded 0.05. A site was removed if it was multi-allelic or if greater than 50% of samples were heterozygous. We removed sequences that did not possess genotypes for at least 20% of variants. Data were phased and imputed with Beagle 5.0 (Browning and Browning 2007; Browning et al. 2018).

Whole-genome genetic variance decomposition and sample clustering

The whole-genome genotype matrix of 780 sequences was pruned for redundant SNPs within 100 bp physical windows using BCFtools (to reduce impacts of wild-species divergence), then scaled and centered using Scikit-learn version 0.23.2 for Python 3 on 64-bit Linux CentOS 7 (Li et al. 2009; Pedregosa et al. 2011). Scikit was used to calculate principal components and to hierarchically cluster sequences using Ward distance criterium and a range of k = {3-20}

clusters with Scikit. A value of *k*=10 clusters was empirically chosen for its balance of accuracy and sensitivity when cluster membership was cross-referenced with passport data and by scree visualization. Clusters were named according to "majority rules" taxonomic info from passport data and were visualized with ggplot2 in R 3.6.0 (R Core Team 2020)

Determination of introgression boundaries by sliding window haplotype clustering analysis

Haplotype homology was detected on a sliding window using a custom Python 3 script built around Scikit-learn 0.23.2 (Supplement 8). The script windowed the genomes of all sequences, using physical distance, and used hierarchical agglomerative clustering to identify haplotype clusters. The threshold distance, *d*, under which clusters are merged, was determined by maximizing the silhouette score. If two or more sequences were clustered together (indicating sequence similarity) their haplotypes were considered homologous for the genomic window. We assume homology (as opposed to sequence similarity) within the narrow genetic base of cultivated tomato. We visualized haplotypes using ggplot2 in R. The analysis was repeated with the window sizes {100 Kb, 250 Kb, 500 Kb, 1 Mb}, corresponding step sizes of {25 Kb, 100 Kb, 100 Kb, 250 Kb}, and *d* ranges of {2-80, 2-100, 20-120, 20-200}.

Evaluating evidence for introgressions with coalescent tree inference

Further evidence for introgression was evaluated by inferring haplotype trees under the coalescent model. Variants falling within the physical boundaries of the ancestral *EB-5* or *EB-9* introgressions were included in the analyses unless we specified a genome-wide analysis, in which case filtered SNP variants for all chromosomes were supplied. We estimated relationship trees using the SVDquartets program implemented in PAUP*, with 100 bootstrap replicates (Swofford 2003; Chifman and Kubatko 2014, 2015).

Introgression boundary refinement using pairwise ancestry painting

Introgression boundaries for *EB-5* and *EB-9* were identified by localized pairwise

ancestry painting with Ruby scripts from *Dsuite* (Malinsky et al. 2020). We identified polymorphic sites in the sequences for two contrasting accessions that were fixed in opposite directions. Genotypes for a panel of samples were then "painted" based on whether they shared an allele with one of two contrasting accessions. Three contrasts were examined: *Devon Surprise* vs. Heinz 1706, *Devon Surprise* vs. NC 84173, and *Devon Surprise* vs. *Yellow Pear*. For *EB-9*, the contrasts included *Hawaii 7998* vs. Heinz 1706, *Hawaii 7998* vs. CU151095-146, and *Hawaii 7998* vs. *Yellow Pear*. We visualized shared alleles with *ggplot2* for R.

Prediction of resistance in sequenced accessions

Hierarchical agglomerative clustering of all samples for the refined *EB-5* and *EB-9* genomic windows was repeated for 7 partially overlapping haplotype windows at each resistance locus (shown in Table 4). If a sequence was clustered with the putative resistance donor for 6-7 of these iterations, the sample to which that sequence belonged was predicted to have resistance with "high confidence". If a sequence was clustered with the donor in 3-5 or 1-2 iterations, resistance was predicted with medium or low confidence, respectively. Fine-scale evidence for resistance donor homology was obtained by windowing the *EB-5* or *EB-9* interval using a 250 Kb window size and 25 Kb step size.

Identification of gene annotations with SNP variants

Refined QTL intervals were surveyed for putative causative loci underlying *EB-5* and *EB-9*. The Integrative Genome Viewer (IGV) was used initially to visualize SNP density and SNP patterns at sites within the introgression boundaries (Thorvaldsdóttir et al. 2013; Robinson et al. 2017). SNPeff was used to annotate variants that fell within and outside of predicted genes and to screen for functional changes in reference to the SL4.0 reference genome and the ITAG4.1 gene annotations (Cingolani et al. 2012; Hosmani et al. 2019). Variants were considered to possibly underlie resistance if the alleles were mutually exclusive in known

resistant (*EB-5* or *EB-9*) and susceptible (lacking *EB-5* and/or *EB-9*) sets of accessions.

Identification of polymorphic SNP markers in ancestral sequences

SNP variants were selected from the *EB-5* interval SL4.0ch05:62,350,391 – 63,450,391 and the *EB-9* interval SL4.0ch09:62,452,852 – 63,002,852 for which an allele was shared among early blight resistant accessions from the hypothesized introgression pathway for either *EB-9* (Fig. 13A) or *EB-5* (Fig. 14A), and was absent from relevant susceptible tomatoes, as described in the results. Flanking SNP sequences are from the SL4.0 genome sequence.

Results

Re-sequencing depth across sources

Read depths for 780 genome sequences aligned to the SL4.0 reference varied primarily according to the data source. Mean depths across SNP variants ranged from 5.5x to 24.3x for the newly sequenced early blight resistant accessions (Table 3). Among these, *S. habrochaites* accession PI 126445 had the lowest mean depth because of its large genome size and distant relationship to the tomato reference. Depths for sequences from OSU ranged from 8.3x to 11.9x for OH88119 and OH08-7663, respectively. Sequences from ERP004618 had relatively high average depths from 12.6x to 39.0x, but average depths were lower, 0.3x to 13.5x, for sequences from SRP045767. From SRP094624, the depth ranged between 2.1x to 22.2x, and from SRP150040 the range was 0.9x to 39.9x (Table S1, Supplement 7).

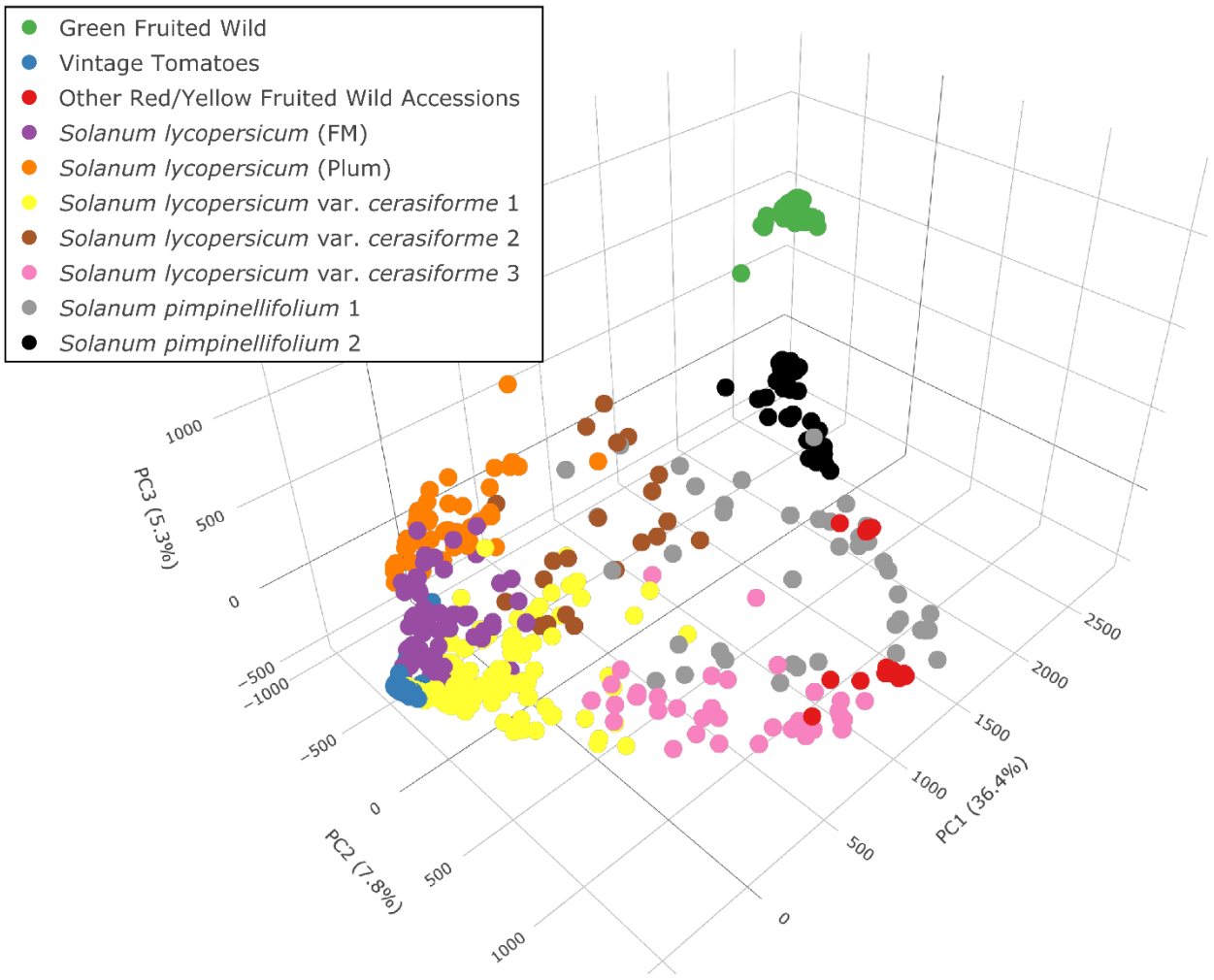
Classification and demographic history of sequenced accessions

We summarized the contents of the sequence dataset by performing hierarchical agglomerative clustering of all 780 whole-genome sequences. The resulting clusters broadly agreed with the known demography for tomato and its wild relatives. When the sequences were projected onto a 3-dimensional principal component (PCA) space (Fig. 12), the data were roughly circumscribed by *S. pimpinellifolium*, *S. lycopersicum* var. *cerasiforme*, cultivated vintage tomato accessions, and relatives of tomato from the green-fruited clade. The primary axis of variation reflected genetic differentiation between cultivated tomato and its wild progenitor *S. pimpinellifolium*. The second principal component reflected additional variation within the red-fruited clade, including that among the canning/processing and fresh market classes. The third principal component reflected the divergence between the green and red-fruited *Solanum* clades. The green-fruited wild species cluster incorporated all accessions of *S. chmielewskii, S. arcanum, S. neorickii, S. hualyasense, S. peruvianum, S. corneliomuelleri, S.*

Table 3 Brief descriptions of core sequenced accessions discussed throughout this work, including the market class and species of the accession, the number of independent sequences with the same name, and the average sequenced depth across filtered SNP variants mapped to the SL4.0 genome

Entry Name	Tomato Type	Species	Sequence from	Number of Sequenced Accessions	Mean SNP Variant Depth
Devon Surprise	Vintage	*S. lycopersicum*	This study	1	21.0
Campbell 1943	Early processing	*S. lycopersicum*	This study	1	20.0
NC EBR 1	Fresh market	*S. lycopersicum*	This study	1	15.6
NC EBR 2	Fresh market	*S. lycopersicum*	This study	1	11.2
NC 1 CELBR	Fresh market	*S. lycopersicum*	This study	2	17.0, 24.3
CU191357	Fresh market	*S. lycopersicum*	This study	1	20.5
CU201041	Processing	*S. lycopersicum*	This study	1	18.8
PI 126445	Wild species	*S. habrochaites*	This study	1	5.5
CU151095-146	Fresh market	*S. lycopersicum*	This study	1	16.7
CU151011-146	Fresh market	*S. lycopersicum*	BTI	1	27.9
Hawaii 7998	Fresh market	*S. lycopersicum*	OSU	2	10.8, 4.5
OH7536	Processing	*S. lycopersicum*	OSU	1	11.2
OH08-7663	Processing	*S. lycopersicum*	OSU	1	11.9
OH88119	Processing	*S. lycopersicum*	OSU	1	8.3
M82	Processing	*S. lycopersicum*	SRP045767	1	5.0
NC 84173	Fresh market	*S. lycopersicum*	SRP045767	1	5.8
Heinz 1706	Processing	*S. lycopersicum*	SRP045767	1	4.4
Brandywine	Vintage	*S. lycopersicum*	ERP004618	2	33.4, 6.5
Yellow Pear	Vintage	*S. lycopersicum*	SRP094624	1	5.1

Fig. 12 Whole-genome classification of 780 tomato and related wild species accessions projected onto a 3-dimensional PCA space. Samples were classified by Ward agglomerative hierarchical clustering (*k*=10) and clusters were named according to majority rules taxonomic passport information



chilense, S. habrochaites, and *S. pennellii*. The red and yellow-fruited wild cluster had accessions of *S. galapagense, S. cheesmaniae, S. pimpinellifolium*, and *S. lycopersicum* var. *cerasiforme*. Most of the genetic variance in the dataset was within *S. pimpinellifolium* and *S. lycopersicum* var. *cerasiforme*, for which there was also considerable overlap and ambiguity of cluster membership, reflecting the complex history of tomato domestication. The vintage tomato cluster had the least genetic variation, despite comprising 35.0% of the accessions in the dataset. Greater genetic diversity within the fresh market and processing tomato clusters reflected efforts by breeders to improve disease resistance and horticultural performance through interspecific hybridization and introgression.

Evaluation of a windowed haplotype classification method

We developed a custom introgression detection method based on the identification of homologous windows (haplotypes) that is both simple and computationally efficient. Briefly, sequences are windowed and stepped according to user specifications. Next, a hierarchical agglomerative clustering model is built using Ward distance. Sequences in a window are then clustered together into homologous haplotypes if their pairwise distance falls under a distance threshold. The program determines the distance threshold, *d*, by iterating through a user-specified range and identifying the value of *d* that maximizes the mean silhouette coefficient for all sequences. This approach allows the clustering algorithm to account for unequal information content across genomic windows. The code is built on Scikit-learn's clustering module and is available for download as a Python 3 command-line script in Supplement 8.

The optimal range for *d* depends on the size of the genomic window under consideration. Because the genetic information content and mean pairwise distance among samples increases along with window size, the optimal value of *d* that returns the maximal number of meaningful clusters will also increase (Fig. S1, Supplement 9). A value of *d* that is

too small for a genomic window may cause homologous haplotypes to be split into separate clusters as a result of minor genotyping errors or residual heterozygosity in seed lots. A *d* value that is too large can result in the clustering of divergent haplotypes. Visualization of *d* values returned by the analysis can give insight into the chosen *d* range, with an ideal *d* distribution, in our experience, being uniform to somewhat left-skewed (Fig. S2, Supplement 9).

We evaluated our introgression analysis methodology by detecting introgressions in the well-characterized breeding line CU17NBL that is known to have six *S. pennellii* LA0716 introgressions across five chromosomes (MA Mutschler, unpublished). Using the cluster data from our windowed analysis, we defined LA0716 introgressions as those where the haplotypes from breeding line CU17NBL clustered with LA0716 but not with two cultivated breeding lines without LA0716 introgressions: M82 and NC 1 CELBR. With a 100 Kb window size, 25 Kb step size, and *d* range of {2-80}, the clustering algorithm identified 12 putative LA0716 introgressions, which is greater than the true number of introgressions (Fig. S3, Supplement 9). All distance thresholds returned for the erroneous windows fell in the top 10% of the *d* distribution, indicating little distinguishing genetic information in these windows. Seven putative introgressions were detected, including all six known introgressions, when the input parameters were changed to a 250 Kb window size, 100 Kb step size, and *d* range of {2-100}. With a window size of 500 Kb or greater, a step size of 100 Kb, and *d* range {20-200}, we detected all but the smallest known introgression on chromosome 7. Thus, larger window sizes showed decreased sensitivity to the identification of smaller introgressions, but greater accuracy in the detection of known introgressions.

Early blight resistance underlying *EB-9* in modern tomatoes traces to *Devon Surprise*

We found compelling evidence for introgression from *Devon Surprise* in modern early blight resistant breeding lines. To demonstrate the value of our windowed introgression

detection method over SNP-based methods, we began by identifying SNPs in the genome where an allele was both 1) shared among landmark accessions along the putative *EB-9* introgression pathway (Fig. 13A) and 2) was absent from three tomatoes susceptible to early blight stem lesions: NC EBR 1, NC 84173, and *Brandywine*. In total, 616 SNPs across twelve chromosomes fit this specified pattern (Fig. 13B). In contrast, our windowed haplotype clustering method found just three haplotypes fitting the pattern (Fig. 13C). One of these haplotypes fell within the previously defined *EB-9* interval on chromosome 9 (Fig. 13D). Chromosome-level visualizations showed the decay of the chromosome 9 introgression from *Devon Surprise* as it was transferred through *Campbell 1943* into a modern cultivated background. The other two haplotypes fell on chromosomes 8 and 12 and did not coincide with resistance QTL (Fig. S4, Supplement 9). However, these haplotypes were also homologous with OH08-7663, which we had used as a QTL mapping parent.

Evidence for a *Devon Surprise* introgression underlying *EB-9* was confirmed by estimating haplotype relationships with coalescent-based species tree inference. Using genotype data for the refined *EB-9* interval (Table 4), the tree gave strong evidence for a shared local ancestry at *EB-9* among the landmark early blight stem resistant breeding lines and *Devon Surprise* (Fig. 13E). The inferred *EB-9* tree closely approximated the putative breeding pedigree (Fig. 13A), reflecting modifications to the introgression over decades of breeding. Furthermore, breeding lines with the resistant *EB-9* sequence were apparently more similar to *S. habrochaites* PI 126445 at the *EB-9* locus than to other *S. lycopersicum* accessions. This branching behavior suggests a wild species origin for *EB-9*, rather than a cultivated one. Interestingly, our haplotype analysis indicated that *Devon Surprise* and *Ailsa Craig* overlapped

Fig. 13 Fine-mapping *EB-9* stem and foliar early blight resistance by comparative sequencing. **A.** *EB-9* Resistance is hypothesized to be derived from the vintage *Devon Surprise*. Landmark tomato lines along the *EB-9* introgression pathway are shown in orange (many generations are omitted). **B.** Visualization of SNP variants (orange bars) in the tomato genome that fit the expected pattern of introgression (i.e. an allele is shared among all resistant tomatoes from the hypothesized pedigree in panel A, but is absent from the early blight stem-susceptible controls NC EBR 1, NC 84173 and *Brandywine*). **C.** Haplotypes that fit the pattern of introgression based on a custom 250 Kb sliding window analysis are colored orange, showing homology and putative introgression from *Devon Surprise*. **D.** Zoomed-in visualizations of the three putative introgressions from *Devon Surprise*. Red lines indicate the low-confidence introgression limits for the window (window center ± ½*window step size), while blue lines indicate the outer window edges. The prior QTL mapping boundaries for *EB-9* are shown. **E.** Evidence of shared haplotypes for 20 tomatoes relevant to this study, including early blight stem resistant lines, lines with foliar early blight resistance only, and a few famous tomato lines. The hierarchical tree contains bootstrap support values for the average *EB-9* genomic interval estimates (see Table 4)

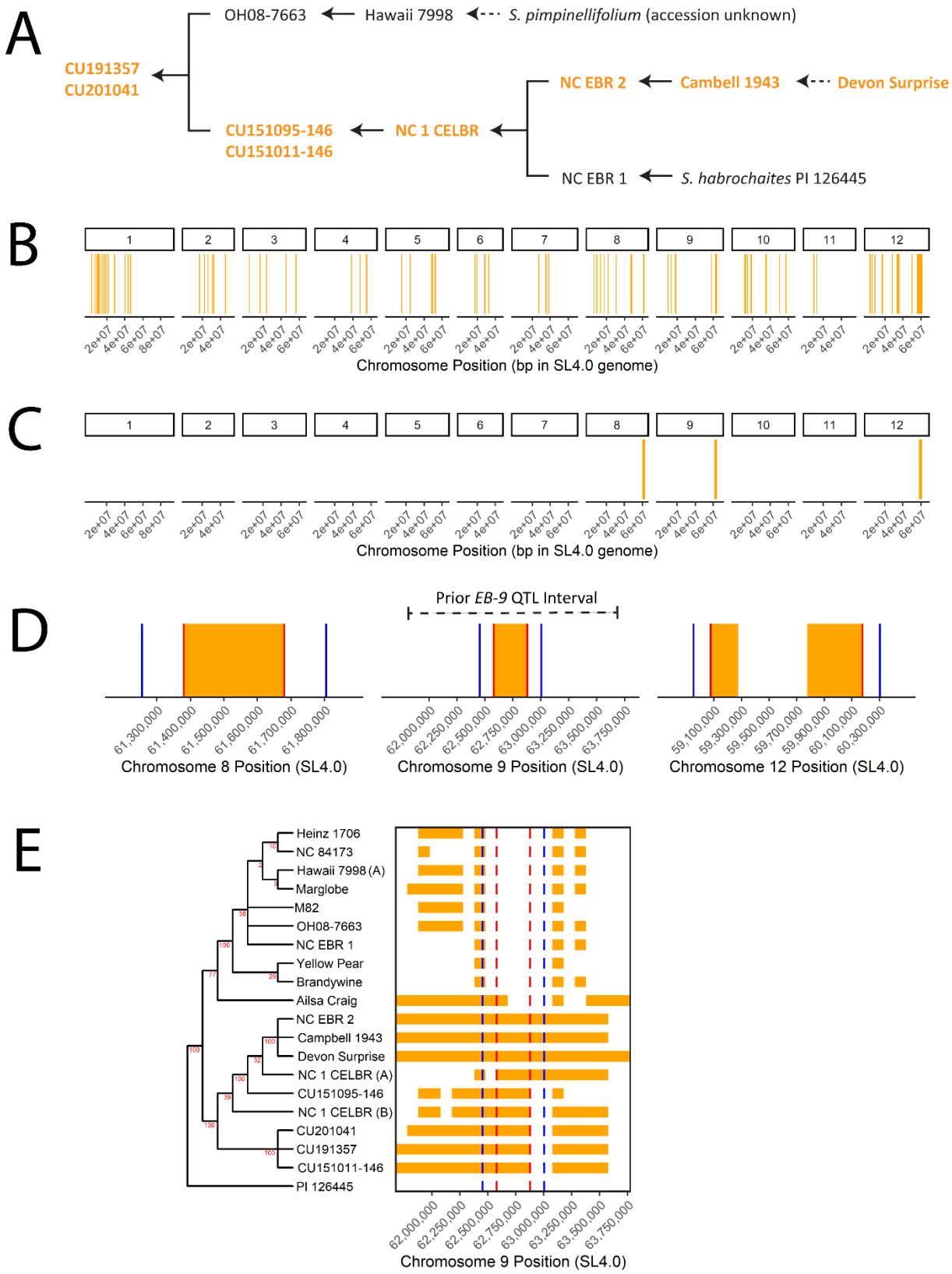

A
OH08-7663
Hawaii 7998
S. pimpinellifolium (accession unknown)
CU191357
CU201041
NC EBR 2
Cambell 1943
Devon Surprise
CU151095-146
CU151011-146
NC 1 CELBR
NC EBR 1
S. habrochaites PI 126445
B
1
2
3
4
5
6
7
8
9
10
11
12
Chromosome Position (bp in SL4.0 genome)
C
Chromosome Position (bp in SL4.0 genome)
D
Prior EB-9 QTL Interval
Chromosome 8 Position (SL4.0)
Chromosome 9 Position (SL4.0)
Chromosome 12 Position (SL4.0)
E
Heinz 1706
NC 84173
Hawaii 7998 (A)
Marglobe
M82
OH08-7663
NC EBR 1
Yellow Pear
Brandywine
Ailsa Craig
NC EBR 2
Campbell 1943
Devon Surprise
NC 1 CELBR (A)
CU151095-146
NC 1 CELBR (B)
CU201041
CU191357
CU151011-146
PI 126445
Chromosome 9 Position (SL4.0)

Table 4 Comparison of *EB-5* and *EB-9* QTL boundaries as determined from three different mapping approaches. **Traditional QTL mapping** in populations derived from the cross of CU151095-146 and OH08-7663 gave initial, course boundaries for the early blight resistance QTL. **Pairwise chromosome painting** was used to delimit the ancestral introgression boundaries from *Devon Surprise* (*EB-9*) and *Hawaii 7998* (*EB-5*), narrowing the QTL intervals. This was done by "painting" alleles for a panel of accessions as matching one of two possible donors (e.g. *Devon Surprise* vs. NC 84173) across the QTL intervals and then visually estimating the shared boundaries among resistant accessions. A custom **windowed haplotype clustering** methodology gave similar introgression boundaries. Our approach enabled multi-way contrasts by using agglomerative hierarchical clustering to identify genomic windows that were shared among resistant accessions and divergent from known susceptible accessions. All positions are base pairs in the SL4.0 tomato reference genome sequence

	EB-5 (Chr 5)			***EB-9*** (Chr 9)		
	Upper Boundary	**Lower Boundary**	**QTL Interval**	**Upper Boundary**	**Lower Boundary**	**QTL Interval**
QTL Mapping[a]						
Minimum boundaries	62,700,265	63,842,577	1,142,312	61,819,509	63,679,761	1,860,252
Pairwise Ancestry Painting[b]						
Devon Surprise vs. NC 84173	-	-	-	62,599,611	62,943,349	343738
Devon Surprise vs. *Yellow Pear*	-	-	-	62,599,611	62,943,349	343738
Devon Surprise vs. *Heinz 1706*	-	-	-	62,599,611	62,943,349	343738
Hawaii 7998 vs. CU151095-146	62,728,321	63,401,903	673,582	-	-	-
Hawaii 7998 vs. *Yellow Pear*	62,699,938	63,401,903	519,858	-	-	-
Hawaii 7998 vs. *Heinz 1706*	62,417,592	63,401,903	701,638[d]	-	-	-
Windowed Clustering[c]						
100 Kb sliding window	62,412,891	63,412,891	712,626[d]	62,540,352	62,940,352	400,000
250 Kb sliding window	62,350,391	63,200,391	500,126[d]	62,452,852	63,002,852	550,000
500 Kb sliding window	62,350,391	63,450,391	750,126[d]	62,252,852	63,052,852	800,000
1 Mb sliding window	62,350,391	63,650,391	950,126[d]	61,877,852	63,627,852	1,750,000

a. Boundaries determined from Chapter 2
b. Pairwise ancestry painting looks for evidence of introgression from one entry relative to a contrasting entry for each SNP. We present three contrasts for *EB-5* and *EB-9* for comparative purposes
c. A conservative range is given based on the physical position of the outer window edges
d. In this case, the upper boundary of the ancestral introgression extends beyond the previously determined upper QTL boundary. The QTL interval was calculated by subtracting the upper QTL boundary from the lower boundary determined from sequence mapping

for just a portion of the *EB-9* interval (Fig. 13E). Estimation of the whole-genome species tree[4] for a relevant subset of accessions confirmed a close genetic relationship between *Devon Surprise* and *Ailsa Craig* (Fig. S5, Supplement 9), supporting the assertion that *Devon Surprise* is a spontaneous mutant or offtype *Ailsa Craig*.

Delineation of the ancestral *Devon Surprise* introgression boundaries enabled fine-mapping of the *EB-9* resistance locus. Introgression boundaries were determined by two methods, with broad agreement. First, we used SNP-based pairwise chromosome painting to estimate the *EB-9* boundary. Three contrasts were examined between *Devon Surprise* and one of three accessions lacking stem lesion resistance (Table 4). All three contrasts delimited the ancestral *EB-9* haplotype between SL4.0ch09:62,599,611 and SL4.0ch09:62,945,798 (Table 4; Fig. S6, Supplement 9). These boundaries represent an 81.4% reduction over the *EB-9* interval from QTL mapping. The introgression boundaries were also estimated using our haplotype clustering method. Using a 250 kb window size, we delimited *EB-9* to between SL4.0ch09:62,452,852 and SL4.0ch09:63,002,852, which represents a 70.4% reduction in the QTL interval (Fig. 13D). Visualization of the pairwise cluster-based homology with *Devon Surprise* reveals that NC 1 CELBR was segregating for the upper portion of the *Devon Surprise* introgression (NC 2 CELBR was sequenced twice), delimiting its upper bound (Fig. 13E). The lower boundary was delimited by CU151095-146, which had lost the lower portion of the *Devon Surprise* introgression.

To investigate the effect of varying the windowing size on our clustering approach to introgression boundary estimation, we repeated the analysis for *EB-9* using window sizes of 100 Kb, 250 Kb, 500 Kb, and 1 Mb. There was consistent evidence for *Devon Surprise* ancestry

[4] Using coalescent-based tree inference with all filtered SNP variants (across all chromosomes)

at *EB-9* regardless of the window size. However, varying the parameter affected the trade-off between clustering accuracy and the estimated introgression size. A window size of 100 Kb resulted in noisy haplotype data, with many discontinuous haplotypes sharing homology with *Devon Surprise.* This noise posed a challenge to accurate boundary estimation. We combatted this issue by setting a minimum of 15 SNPs per window for the clustering analysis, ensuring that haplotype clusters were not estimated for data-poor genomic windows. Larger window sizes gave increasingly accurate estimations of haplotype homology, but also resulted in wider and increasingly conservative introgression intervals (Table 4). We found the 250 Kb window size to be a good compromise for our data, as reflected by the small *EB-9* interval in Table 4, the clear homology visualizations in Fig. 13E, and the desirable distribution of *d* scores (Fig. S2, Supplement 9).

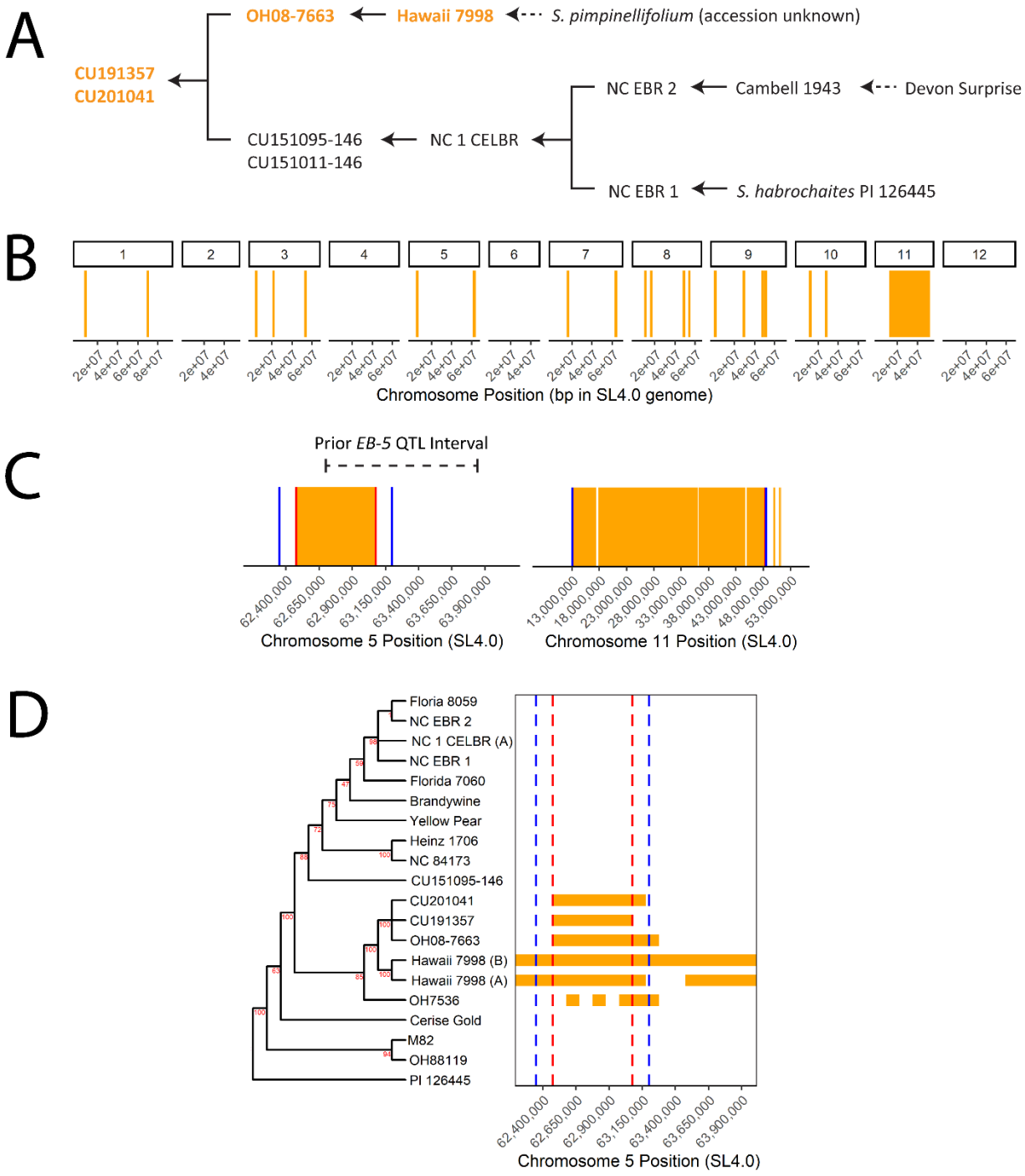
A survey of all 780 sequences identified 48 sequences sharing homology with *Devon Surprise* for at least one candidate *EB-9* interval (Fig. S7, Supplement 9). Of the 48 sequences, nine belonged to the landmark breeding lines already known to have stem lesion resistance. Four additional accessions shared homology with *Devon Surprise* for all seven of the candidate intervals investigated (see Table 4) and are predicted to have *EB-9* early blight resistance. These accessions were the vintage tomatoes *Gardner's Delight* (two sequences), *Monplaisir*, *Yellow Perfection*, and *Katinka Cherry*. Twenty and 16 accessions clustered with *Devon Surprise* in 3-5 and 1-2 of the candidate intervals, respectively, and may have *EB-9* stem lesion resistance. All accessions that had similar sequences to *Devon Surprise* at *EB-9* were either *S. lycopersicum* or *S. lycopercium* var. *cerasiforme*, except for one sequence belonging to *S. pimpinellifolium* PI 370093. Fine-scale windowing (250 Kb) of the full *EB-9* interval for these sequences revealed at least 7 sub-haplotypes with distinctive patterns of homology (Fig. S7, Supplement 9).

Early blight and bacterial spot resistance traces to introgressions from *Hawaii 7998*

Our cluster-based introgression detection method identified several genomic windows shared among *EB-5* breeding lines and absent from the early blight susceptible tomatoes OH88119, NC 84173 and *Brandywine* (Fig. 14A-B). One of these windows fell within the QTL interval for *EB-5* (Fig. 14C). It was possible to trace the decay in the *Hawaii 7998* introgression through OH08-7663 into the fresh market breeding line CU191357. The localized *EB-5* species tree inference (Fig. 14D) confirmed evidence for *Hawaii 7998* as the source of *EB-5* foliar resistance in OH08-7663 and the derived breeding lines CU191357 and CU201041.

The *EB-5* QTL interval was refined by delimiting the shared ancestral *Hawaii 7998* introgression boundaries in resistant breeding lines using pairwise SNP chromosome painting and haplotype clustering. Pairwise SNP contrasts gave estimated upper boundaries for *EB-5* between SL4.0ch05:62,417,592 and SL4.0ch05:62,728,321, depending on the contrast, while the lower boundary was consistently estimated at SL4.0ch05:63,401,903 (Table 4, Fig. S8, Supplement 9). These intervals represent a 38.6 - 54.5% reduction in the *EB-5* interval relative to those obtained from QTL mapping. Our haplotype clustering analysis with a 250 Kb window size gave an *EB-5* interval between SL4.0ch05:62,350,391 and SL4.0ch05:63,200,391 (Fig. 14C). The upper boundary of the ancestral introgression was shared by the resistant breeding lines OH08-7663, CU191357, and CU201041, while the lower boundary was delimited by CU191357 (Fig. 14D). The upper haplotype boundary exceeded that obtained from QTL mapping. With the QTL boundary marking the upper limit of the *EB-5* interval, the refined *EB-5* QTL interval represents a 56.2% reduction relative to QTL mapping. Varying the windowing parameters, as done for *EB-9*, gave similar results, with the 250 Kb and 500 Kb window sizes offering the best compromise between clustering accuracy and overly conservative introgression intervals.

Fig. 14 Fine-mapping *EB-5* foliar early blight resistance by comparative sequencing. **A.** *EB-5* resistance was hypothesized to be derived from *S. pimpinellifolium*, but we found little evidence for such a putative donor. Landmark tomato lines along the *EB-5* resistance breeding pathway are shown in orange. **B.** Haplotypes that fit the pattern of introgression (i.e. an allele is shared among all resistant tomatoes from the hypothesized pedigree in panel A, but is absent from the early blight susceptible controls OH88119, NC 84173 and *Brandywine*) based on a 250 Kb sliding window analysis are colored orange, showing homology and putative introgression from *Hawaii 7998***. C.** Zoomed-in visualizations of the two putative introgressions from *Hawaii 7998*. Red lines indicate the low-confidence introgression limits for the window (window center ± ½*window step size), while blue lines are located at the window edges. The prior QTL mapping boundaries for *EB-5* are shown. **D.** Evidence of shared haplotypes for 20 tomatoes relevant to this study, including early blight resistant lines and famous tomatoes. The hierarchical tree contains bootstrap support values for the average *EB-5* genomic interval estimates (see Table 4)


A
OH08-7663
Hawaii 7998
S. pimpinellifolium (accession unknown)
CU191357
CU201041
NC EBR 2
Cambell 1943
Devon Surprise
CU151095-146
CU151011-146
NC 1 CELBR
NC EBR 1
S. habrochaites PI 126445
B
1
2
3
4
5
6
7
8
9
10
11
12
Chromosome Position (bp in SL4.0 genome)
C
Prior EB-5 QTL Interval
Chromosome 5 Position (SL4.0)
Chromosome 11 Position (SL4.0)
D
Floria 8059
NC EBR 2
NC 1 CELBR (A)
NC EBR 1
Florida 7060
Brandywine
Yellow Pear
Heinz 1706
NC 84173
CU151095-146
CU201041
CU191357
OH08-7663
Hawaii 7998 (B)
Hawaii 7998 (A)
OH7536
Cerise Gold
M82
OH88119
PI 126445
Chromosome 5 Position (SL4.0)

We found little evidence for *EB-5* resistance in accessions from outside of the Cornell or OSU breeding programs. Using our haplotype clustering method, we identified twelve sequences from outside our breeding programs with homology to *Hawaii 7998* for the refined *EB-5* intervals. However, these accessions clustered with *Hawaii 7998* for just one or two of the seven *EB-5* intervals investigated (Fig. S9, Supplement 9). Furthermore, fine-scale (250 Kb) windowing of the full *EB-5* interval revealed little continuous homology with *Hawaii 7998,* suggesting that these haplotypes were relatively divergent.

The chromosome 5 introgression from *Hawaii 7998* in modern breeding lines also overlapped with the putative location of the *Rx-3* bacterial spot (*Xanthomonas sp.*) disease resistance locus. Fine-map data for *Rx-3* has not been published, but the QTL is known to be located between approximately 61.9 Mb and 63.2Mb on chromosome 5 of SL4.0 (DF Francis, pers. comm.). Both OH7536 and OH08-7663 have *Rx-3* resistance tracing to *Hawaii 7998* (Yang and Francis 2005; Sim et al. 2015). Given the partially overlapping *Hawaii 7998* introgressions in these two lines, we could use our clustering analysis with a 250 kb window size to delineate the *Rx-3* interval to between SL4.0ch05:62,850,391 and SL4.0ch05:63,400,391. We believe that CU191357 and CU201041 also have *Rx-3* resistance, which could further narrow the interval, but neither of these lines has yet been tested against the bacterial spot pathogen.

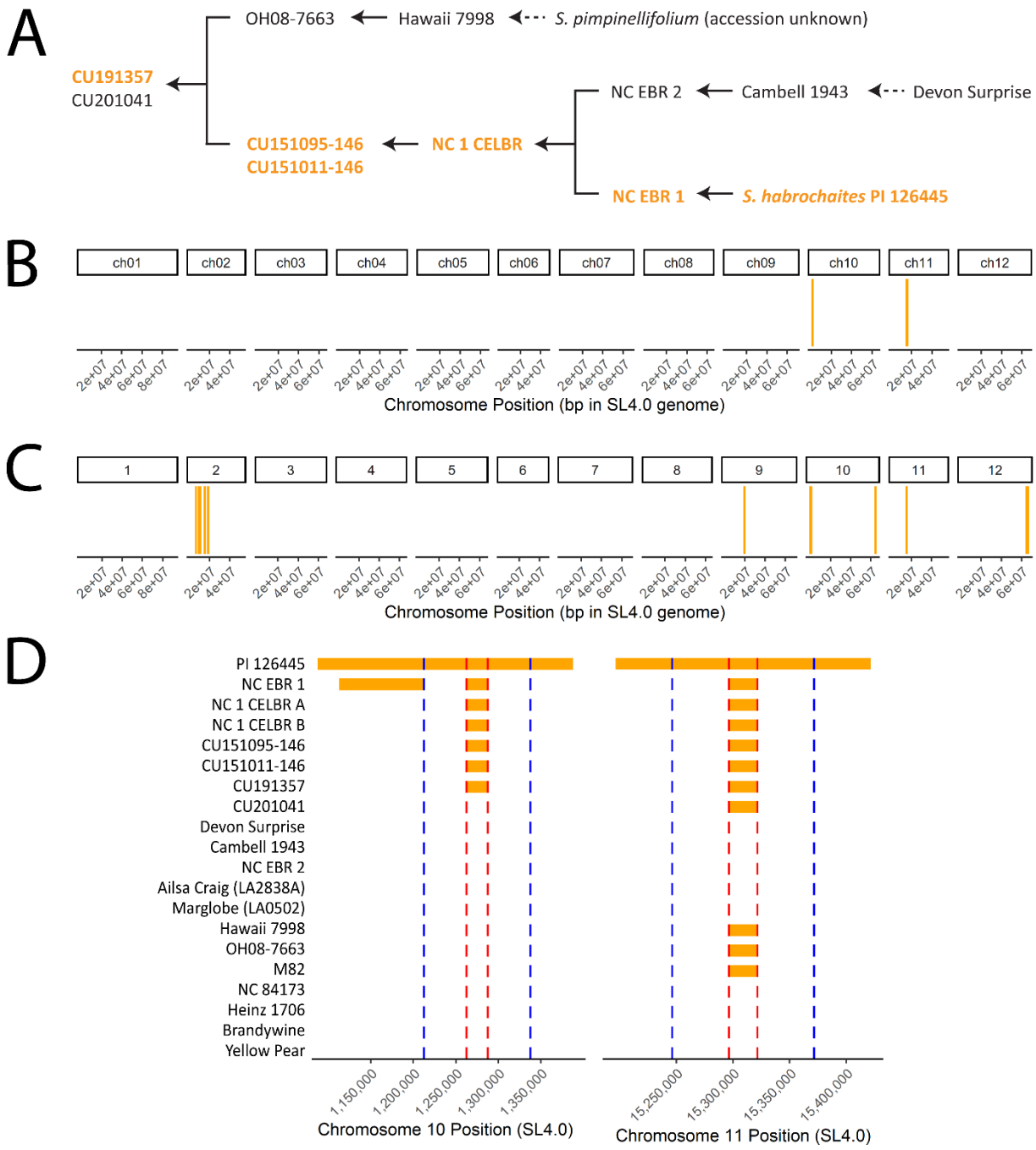
Several early blight resistant breeding lines shared homology with *Hawaii 7998* for a large centromeric chromosome 11 interval that contains the bacterial spot quantitative resistance locus *QTL-11.* Using our windowed analysis with a 250 Kb window size, we estimated the shared haplotype among breeding lines known to have *QTL-11* to extent from SL4.0ch11:13,059,026 to SL4.0ch11:48,509,026 (Fig. S10, Supplement 9). Because this QTL is hypothesized to be derived from *S. pimpinellifolium*, we looked among all *S. pimpinellifolium* accessions in our dataset for homology with *Hawaii 7998* for the centromeric haplotype. One

S. pimpinellifolium accession, LA0722, was homologous with *Hawaii 7998* for the entire haplotype window (Fig. S11, Supplement 9). We also found evidence of homology for the *QTL-11* centromeric interval in a total of 58 sequences, 38 of which were clustered with *Hawaii 7998* for all seven of the *QTL-11* intervals we investigated. Most of these accessions were modern cultivated tomatoes, including NC 1 CELBR, M82, *Micro Tom*, *Florida 7060*, *Florida 8059*, *Peto 9543*, *Jelly Bean Hybrid*, and *Mountain Spring VFF Hybrid*.

Limited evidence for introgressions from *S. habrochaites* PI 126445 in modern breeding lines

We observed limited evidence for shared PI 126445 introgressions in the genomes of early blight resistant breeding lines (Fig. 15). Our cluster-based introgression detection method failed to identify any haplotypes common among the landmark breeding lines from the putative introgression pathway (Fig. 15A) and PI 126445 with our default window size of 250 Kb. To increase the detection sensitivity, we reduced the window size to 100 Kb. This analysis revealed homologous haplotypes on chromosomes 10 and 11 that were shared among breeding lines with putative PI 126445 ancestry but were absent from NC 84173, *Brandywine*, and NC EBR 2 (Fig. 15B). Neither interval overlapped with our previously mapped early blight resistance QTL. We further investigated the likelihood of an *S. habrochaites* introgression underlying these haplotypes using principal component analysis. Breeding lines with putative PI 126445 ancestry received PCA coordinates for the chromosome 10 haplotype that were close to other *S. habrochaites* accessions in the dataset and relatively distant from most cultivated tomatoes, suggesting an *S. habrochaites* ancestry. In contrast, PCA analysis for the chromosome 11 haplotype placed the resistant breeding lines alongside many cultivated accessions. In addition, the chromosome 11 haplotype fell within the centromeric *QTL-11* interval with homology to *Hawaii 7998*, described above. Because we saw little evidence of recombination near the chromosome 11 centromere, this result is probably spurious.

Fig. 15 Investigating evidence of *S. habrochaites* PI 126445 introgressions in early-blight resistant breeding lines. **A.** PI 126445 is thought to be the donor early blight foliar resistance in many modern tomato breeding lines. Landmark tomato lines along the early blight resistance breeding pathway are shown in orange. **B.** Haplotypes that fit the pattern of introgression (i.e. an allele is shared among all resistant tomatoes from the hypothesized pedigree in panel A, but are absent from tomatoes without PI 126445 in their pedigree, NC 84173, *Brandywine* and NC EBR 2) based on a 100 Kb sliding window analysis are colored orange, showing homology and putative introgression from PI 126445**. C.** Evidence for introgressions in NC EBR 1 that are not in NC 84173, Brandywine, and NC EBR 2. **D.** Evidence of shared haplotypes for 20 tomatoes relevant to this study


A
OH08-7663
Hawaii 7998
S. pimpinellifolium (accession unknown)
CU191357
CU201041
NC EBR 2
Cambell 1943
Devon Surprise
CU151095-146
CU151011-146
NC 1 CELBR
NC EBR 1
S. habrochaites PI 126445
B
ch01
ch02
ch03
ch04
ch05
ch06
ch07
ch08
ch09
ch10
ch11
ch12
Chromosome Position (bp in SL4.0 genome)
C
1
2
3
4
5
6
7
8
9
10
11
12
Chromosome Position (bp in SL4.0 genome)
D
PI 126445
NC EBR 1
NC 1 CELBR A
NC 1 CELBR B
CU151095-146
CU151011-146
CU191357
CU201041
Devon Surprise
Cambell 1943
NC EBR 2
Ailsa Craig (LA2838A)
Marglobe (LA0502)
Hawaii 7998
OH08-7663
M82
NC 84173
Heinz 1706
Brandywine
Yellow Pear
Chromosome 10 Position (SL4.0)
Chromosome 11 Position (SL4.0)

To determine whether a PI 126445 region was introgressed into NC EBR 1 but subsequently lost during breeding, we repeated the 100 Kb windowed haplotype search using a relaxed search pattern. We sought haplotypes from PI 126445 that were shared with just NC EBR 1 but not with NC 84173, *Brandywine*, and NC EBR 2. This pattern returned additional putative introgressions from PI 126445 on chromosomes 2, 9, 10, and 12 (Fig. 15C). Principal component investigation revealed the chromosome 12 introgression was the most likely of these haplotypes to be derived from *S. habrochaites*.

Identification of putative causative loci

Boundaries for *EB-5* and *EB-9* were confirmed visually with IGV, showing greater SNP density and a conserved variant pattern within the introgression boundaries for tomatoes containing the proposed *EB-5* or *EB-9* introgressions. We identified 82 SNP variants within *EB-5,* 27 of which fell within 12 predicted genes potentially underlying resistance. Within *EB-9*, 90 SNP variants were identified with 16 variants located in eight putative genes. The remaining SNPS were located in intergenic regions (Tables S2 and S3, Supplement 7).

We considered any (possibly multiple) genes in the *EB-5* or *EB-9* intervals to potentially underlie early blight disease resistance if it contained SNP variants that were mutually exclusive in the known resistant and susceptible sets of accessions presented above. The predicted functions of such genes were varied, as were the functional effects of the variants. Within *EB-5, Solyc05g053980.1* putatively encoded a plant resistance protein, while others encoded enzymes including gibberellin 2-oxidase 1, a galactosyltransferase family protein, a protein phosphatase-2C, a DNA (Cytosine-5)-methyltransferase, and a serine hydroxymethyltransferase (Table 5). Two genes encoded plant self-incompatibility proteins. Putative protein-coding genes containing variants on chromosome 9 encoded two potassium transporters, a metal tolerance protein C1, cation efflux protein, a peptide chain release factor

1, eukaryotic translation initiation factor 4G, an F-box protein, and a 2-oxoglutarate and Fe(II)-dependent oxygenase protein 1 (Table 6). Many of the variants were located in introns and classified as modifiers. Some variants were missense variants with predicted moderate effects, and a smaller number of variants were either synonymous or located in the 5' untranslated region, with low putative impacts on gene function (Tables 5 and 6).

Several intergenic variants were clustered between genes potentially underlying resistance. For example, within *EB-5*, 16 variants resided between Solyc05g053900.4.1 (putative function: eukaryotic aspartyl protease family protein) and Solyc05g053910.1 (putative function: phospholipase A1) (Tables S2 and S3, Supplement 7). Within *EB-9*, 17 intergenic variants existed between Solyc09g160100.1 (putative function: F-box protein) and Solyc09g074580.1 (putative function of protein: glutaredoxin). Another cluster of 15 variants was located between Solyc09g161650.1 (putative function: Eukaryotic translation initiation factor 4G) and Solyc09g074780.3 (putative function: Protein indeterminate-domain 2).

Polymorphic SNPs within the *EB-5* and *EB-9* ancestral haplotypes

Breeding for early blight resistance at *EB-5* or *EB-9* may be more reliable if trait-linked genetic markers are located within the ancestral haplotypes from *Hawaii 7998* or *Devon Surprise* in modern breeding lines. These haplotypes had shown no signs of recombination over many breeding generations and are thus tightly linked to the causative resistance loci. We identified high-quality SNPs that are specific to the ancestral haplotypes based on whether they matched the expected allele patterns given the hypothesized introgression pathways discussed above; their positions and flanking sequences are reported in Tables S4 and S5 of Supplement 7. Several SNPs fell in gene annotations and have the potential to be causal. The alleles for all 780 sequences in the dataset are also provided so that markers can be screened for their usefulness in related genetic backgrounds.

Table 5 Highlighted variants and gene annotations from snpEFF that fall within the refined *EB-5* interval (SL4.0ch05:62566094 - SL4.0ch05:63401898) for the SL4.0 genome and ITAG4.1 gene annotation

Genes	Gene Location (bp in Chr. 5)	SNP Location (bp in Chr. 5)	Variant Description	Predicted Variant Impact	Predicted Function
Solyc05g053080.2	62602656-62605054	62602681	5' UTR	Low	Unknown
		62602843	Missense	Moderate	
		62602847	Missense	Moderate	
		62603260	Intron	Modifier	
		62603437	Intron	Modifier	
		62603449	Intron	Modifier	
		62603689	Intron	Modifier	
		62603792	Intron	Modifier	
		62603828	Intron	Modifier	
		62603916	Intron	Modifier	
Solyc05g053260.3	62784717-62801391	62792542	Intron	Modifier	DNA (Cytosine-5)-methyltransferase DRM2
Solyc05g053280.3	62803504-62809045	62805073	Intron	Modifier	Galactosyltransferase family protein
Solyc05g053290.3	62811779-62814389	62811946	Missense	Moderate	Protein phosphatase-2C
Solyc05g053340.5	62856750-62863385	62858495	Intron	Modifier	Gibberellin 2-oxidase 1
		62858789	Intron	Modifier	
Solyc05g053450.3	62911678-62913123	62912425	Intron	Modifier	Late embryogenesis abundant protein 1-like
		62912458	Intron	Modifier	
		62912463	Intron	Modifier	
		62912551	Intron	Modifier	
		62912599	Intron	Modifier	
Solyc05g160450.1	62938728-62939307	62938903	Missense	Moderate	Plant self-incompatibility S1
Solyc05g160460.1	62940755-62941326	62941209	Missense	Moderate	Plant self-incompatibility S1
		62941241	5'UTR	Low	
Solyc05g053600.3	63054527-63061342	63059261	Intron	Modifier	Pleiotropic drug resistance protein
Solyc05g053770.4	63182158-63190019	63189871	5' UTR	Low	Myb-like protein X
Solyc05g053810.3	63231432-63235896	63233342	Intron	Modifier	Serine hydroxymethyltransferase
Solyc05g053980.1	63362638-63363006	63362833	Synonymous	Low	Plant resistance protein

Table 6 Highlighted variants and gene annotations from snpEFF that fall within the refined *EB-9* interval (SL4.0ch09:62599611 - SL4.0ch09:62943349) for the SL4.0 genome and ITAG4.1 gene annotation

Genes	Gene Location (bp in Chr. 9)	SNP Location (bp in Chr. 9)	Variant Description	Predicted Variant Impact	Predicted Function
Solyc09g160100.1	62625587-62625973	62625766	Missense	Moderate	F-box protein
Solyc09g074650.3	62684784-62691573	62689139	Intron	Modifier	Peptide chain release factor 1
Solyc09g074740.2	62770894-62773833	62771619 62771967	Intron Intron	Modifier Modifier	Cation efflux family protein
Solyc09g074750.3	62776208-62780230	62780127	Missense	Moderate	Metal tolerance protein C1
Solyc09g161650.1	62797984-62800125	62798499 62798512 62798613 62798884 62799362 62799480 62799508 62800102	Intron Intron Intron Intron Intron Intron Intron Missense	Modifier Modifier Modifier Modifier Modifier Modifier Modifier Moderate	Eukaryotic translation initiation factor 4G
Solyc09g074790.3	62817971-62822507	62819042	Missense	Moderate	Potassium transporter
Solyc09g074800.2	62835326-62840489	62839284	Missense	Moderate	Potassium transporter
Solyc09g074920.3	62933032-62937363	62937049	Missense	Moderate	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein

Discussion

Introgression has left characteristic traces of DNA scattered throughout the genomes of modern tomatoes (Causse et al. 2013; Menda et al. 2014; Blanca et al. 2015). Defining introgressions in modern breeding lines presents breeders with several advantages, including the ability to predict transgressive segregation from the introgressions and ancestry of the parents and the opportunity to fine-map loci using the accumulation of ancestral recombinations. Furthermore, valuable introgressions in modern breeding lines are more readily deployable than those from wild species, as linkage drag on horticultural traits has often been reduced or eliminated through breeding. Defining introgressions can also give insight into the reliability of trait-linked genetic markers, since genomic regions that have remained unbroken over decades of breeding are tightly linked to any causative genes. Therefore, markers within or flanking the ancestral introgressions of modern lines should be reliable in the context of molecular breeding.

Efficient haplotype-based introgression detection

Several methods have been proposed for the detection of introgressions in genome sequences. The simplest approach involves pairwise chromosomal painting (Fu et al. 2015; Der Sarkissian et al. 2015). However, this approach describes alleles relative to just two contrasting donors on which the results are highly dependent. Chromosome painting must be repeated for several pairwise contrasts in order to estimate introgression boundaries for complex, multi-parent breeding germplasm. Introgressions can alternatively be found by identifying stretches of high sequence polymorphism relative to a reference genome and then cross-referencing these polymorphic regions with those of an introgression-free tomato such as *Yellow Pear* (Menda et al. 2014; Strickler et al. 2015). However, this approach only detects introgressions from wild species. The popular ABBA-BABA test statistic can offer evidence for

introgression from a specific ancestor (Green et al. 2010). However, we found that this statistical approach (as implemented in Malinsky et al. (2020)) lacks the sensitivity to detect cryptic introgressions from chromosome-scale data and was highly dependent on the designated outgroup (not shown). Furthermore, the ABBA-BABA test can be unreliable at the local ancestry level (Martin et al. 2015). Several other parametric methods that infer local ancestry or tree topology could be adapted to detect introgression (Huelsenbeck and Ronquist 2001; Sankararaman et al. 2008; Maples et al. 2013; Martin and Belleghem 2017; Salter-Townshend and Myers 2019). But these sophisticated methods can be slow for large numbers of taxa and may require the user to supply unknown information like the number of mixing groups, taxa clusters, genetic maps, or reference panels. Furthermore, these methods typically require specialized input file formats that are atypical of genomic breeding data.

To address these challenges, we used a simple non-parametric hierarchical clustering method to group haplotypes along a sliding window. By imposing logical requirements relating to the putative ancestry (and resistance susceptibility) on the resulting cluster data, we identified putative introgressions tracing to specific ancestors. Our method takes as input a standard chromosome-level VCF and can handle genomic variants for hundreds of sequences thanks to its reliance on the efficient clustering algorithm in Python 'Scikit-learn'. A key feature of our approach is that it leverages the available genetic information in each window to determine the optimal threshold for grouping haplotypes. This feature offers several advantages. First, the method does not require the user to pre-specify the number of clusters to output. Instead, the user can broadly calibrate the analysis to the desired level of clustering sensitivity by inputting a range of distance thresholds. Second, the optimal number of haplotype clusters can vary across genomic windows as a function of the underlying genetic variance. This enables the method to adapt to varying genetic diversity and recombination

frequency across windows (without specification of genetic distances). Lastly, the clustering accuracy benefits from a greater number of input sequences.

Characterization and prediction of early blight disease resistance in sequenced accessions

Our work clarified the ancestry of early blight disease resistance in several modern tomato lines, offering insights that will enable breeders to better utilize available tomato genetic resources. Whole genome sequencing of a small number of resistant accessions allowed us to trace stem and foliar resistance underlying *EB-9* back more than 80 years to the early 20th century cultivar *Devon Surprise*. Similarly, we traced early blight resistance from *EB-5* to the mid-century breeding line *Hawaii 7998*, and to the same haplotype that contains *Rx-3* bacterial spot resistance (Yang and Francis 2005). We used the ancestral introgression boundaries to fine map the resistance loci and to report polymorphic sequences, specific to these haplotypes, that can be used to reliably select for early blight resistance in tomato.

Our analysis predicted *EB-9* resistance in 48 of 769 sequenced accessions from outside our pedigree, suggesting that early blight stem lesion resistance is common in tomato. Fine-scale homology detection revealed a distinct set of sub-haplotypes for *EB-9* that will enable further fine-mapping of the resistance locus in our upcoming work. We hypothesize that only a subset of these haplotypes will contain the causative resistance locus, as several were nonhomologous with *Devon Surprise* for the genomic window containing the marker solcap_snp_sl_29188, which we previously determined to be tightly linked to *EB-9*. Most of the predicted accessions were *S. lycopericum* var. *cerasiforme,* but one was classified as *S. pimpinellifolium*. This is unsurprising, as several accessions of *S. pimpinellifolium* are early blight resistant (Martin and Hepperly 1987; Thirthamallappa and Lohithaswa 2000; Ashrafi and Foolad 2015). Unfortunately, the last major screen of tomato germplasm for stem lesion/collar rot resistance (that we are aware of) was done by CF Andrus and colleagues in 1942. Of the

115 accessions they screened, 36 were tolerant or resistant to stem lesions. Most were from England and Europe, and included both *Ailsa Craig* and *Devon Surprise*, which we confirmed to be highly related. In our study, a majority of the predicted *EB-9* accessions were from the Universidad Politécnica de Valencia, Spain, adding to the idea that early blight stem resistance is prevalent among European tomatoes. Intriguing, the genomes of *Devon Surprise* and *Ailsa Craig* are homologous for just a small portion of the refined *EB-9* interval, presenting an opportunity to narrow the *EB-9* interval further by confirming stem resistance in these tomatoes.

We found the *EB-5* resistance haplotype to be rare among sequenced accessions. This was unexpected, as the breeder of *Hawaii 7998*, Dr. J.C. Gilbert, was known to have been intermating tomato and *S. pimpinellifolium* for the purposes of disease resistance. Therefore, we expected to find evidence of a *S. pimpinellifolium* source for *EB-5*. Re-sequencing additional *S. pimpinellifolium* accessions might help to uncover the *EB-5* resistance source. Further refinement of the *EB-5* interval would also help to trace the introgression history of *EB-5* and to predict *EB-5* in additional sequenced accessions.

Curiously, we saw little evidence of introgression from *S. habrochaites* PI 126445 in modern tomato. We detected possible PI 126445 introgressions on chromosomes 10 and 11 of early blight resistant breeding lines, but only after lowering the sensitivity of the detection method. The most promising of these loci was the putative introgression on chromosome 10. However, our prior QTL mapping population would have been segregating for this putative introgression (as it was absent from OH08-7663), but we did not find a QTL associated with early blight resistance on this chromosome. Breeding done by Dr. Gardner to eliminate linkage drag associated with PI 126445 introgressions might have reduced the size of the introgression below the limit of detection for our method. Alternatively, the PI 126445 accession we

sequenced could be different from that used in breeding. Finally, genotyping challenges stemming from the comparison of divergent genomes may have influenced our results.

One shortcoming of QTL mapping in biparental populations is the limited genetic base for QTL detection. In our prior work, we mapped early blight resistance in a population founded by the processing tomato breeding line OH08-7663 and the fresh market line CU151095-146. As both of these tomatoes are regionally adapted inbreds with some degree of early blight resistance, it is possible that our QTL mapping missed important resistance loci that were shared by both parents and were therefore undetectable in our work. Here, we identified two such introgressions from *Devon Surprise* that could be involved in early blight resistance on chromosomes 8 and 12 (Fig. S4, Supplement 9). Any association between these haplotypes and early blight resistance can only be determined through disease trials. Nevertheless, our haplotype analysis method demonstrates that detecting introgressions in parental lines can help to broaden the impact and relevance of narrowly focused QTL mapping papers.

Prevalence of the *QTL-11* haplotype among sequenced accessions

Hawaii 7998 has a centromeric chromosome 11 QTL (*QTL-11*) that confers broad-spectrum quantitative resistance to bacterial spot. To our surprise, we detected homology between the centromeric *QTL-11* haplotype in *Hawaii 7998* and as many as 58 accessions, including LA0722 and breeding lines from Cornell, NCSU and the University of Florida breeding programs (Fig. S10, Supplement 9). Importantly, this centromeric haplotype is homologous between *Hawaii 7998,* CU151095-146, and OH08-7663. The OSU breeding program recently introgressed *QTL-11* into OH08-7663 to enhance bacterial spot resistance in that line (Sim et al. 2015). Subsequently, we used flanking SNPs to transfer *QTL-11* from OH08-7663 into the fresh market background of CU151095-146 (Anderson et al. 2021). Upon completing the lines,

we failed to detect enhanced resistance to bacterial spot in the fresh market background. Our analysis can explain this peculiar result because it suggests that *QTL-11* was already in CU151095-146, despite the fact that we never intentionally bred for bacterial spot resistance, nor knowingly crossed with a *QTL-11* donor. Thus, there is a clear benefit to characterizing the cryptic introgressions in parental breeding materials.

Identification of possible causative loci underlying *EB-5* and *EB-9*

Refinement of the *EB-5* interval empowered the identification of possible causative resistance loci. It should be noted that multiple genes or intergenic variants could possibly underly resistance, as our understanding of the early blight pathosystem remains poor. Of particular interest within *EB-5* was *Solyc05g053980*, which encodes a plant resistance protein (Table 5). While *Solyc05g053980* could underlie early blight *EB-5* disease resistance, it may instead underlie *Rx-3* bacterial spot resistance, as we found that these two loci have overlapping positional intervals and are likely from the same *Hawaii 7998* introgression. Furthermore, *Rx-3* is thought to be a classic R gene conferring hypersensitive resistance, which contrasts with the quantitative nature of *EB-5* resistance (Yang and Francis 2005). Additional experiments are necessary to determine whether any relationship exists between *Solyc05g053980* and either bacterial spot or early blight disease resistance.

Five predicted genes within *EB-5* encoded enzymes (Table 5). Several of these enzymes may play a role in signaling pathways and in triggering plant defense mechanisms. One putative gene encodes a DNA Cytosine-5 DNA Methyltransferase. This enzyme plays an important role in DNA methylation, genome protection and regulation, gene expression, secondary metabolism, and plant development (Wang et al. 2016). Previous studies have shown that UDP-glycosyltransferases (UGTs) glycosylate metabolites and phytohormones when plants encounter biotic and abiotic stresses. Some UGTs may play a role in jasmonic acid

and abscisic acid signaling pathways. Several UGTs have also been studied in the context of the functional role they play in the hypersensitive response (Le Roy et al. 2016; Rehman et al. 2018). The gene encoding a phosphatase-2C enzyme may also play a role in ABA signaling, responding to abiotic and biotic stresses, and immune activation and suppression (Sugimoto et al. 2014; Singh et al. 2016). The putative gene encoding a serine hydroxymethyltransferase has been shown to play a role in the photorespiratory pathway, which could also have an impact on both abiotic and biotic stress. In experiments with *Arabidopsis*, serine hydroxymethyltransferase SHMT1 was shown to help restrict cell death induced by pathogens (Moreno et al. 2005). Any of these genes (or multiple genes) could be involved in quantitative resistance to early blight disease but would require additional experiments to elucidate a link.

Less is known about the remaining genes potentially underlying *EB-5* resistance. For example, a gene encoding gibberellin 2-oxidase 1 was identified. The enzyme is involved in the catabolic pathway for gibberellins, which influence plant growth and development, including fruit set and seed development (Chen et al. 2016). Few studies directly address the possible role of this enzyme in disease resistance. There was also a gene of unknown function, *Solyc05g053080*, that contained 10 SNP variants that were only present in the known resistant set of accessions (Table 5).

Fewer putative genes within *EB-9* were identified, but several appear to encode proteins connected to signaling pathways and defense responses (Table 6). A putative gene for an F-box protein was identified. In previous work on tomato and tobacco, a conserved F-box protein was shown to help regulate plant hormone signaling and to be involved in plant defense responses (van den Burg et al. 2008). Another putative gene encodes the 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase enzyme, which may also help to control defense responses to nectrotrophs and herbivorous insects. Not only does the enzyme play a

role in oxygenation and hydroxylation reaction in plants, but some proteins in the family can hydroxylate and deactivate gibberellic acid, auxin, and defense hormones such as salicylic acid and jasmonic acid (Caarls et al. 2017).

Other putative genes within *EB-9* encode potassium transporters, a cation efflux protein, a metal tolerance protein and a eukaryotic translation initiation factor. Potassium helps mediate osmotic pressure and the opening and closing of stomata. Membrane potential and enzyme activity are also regulated, in part, by potassium transporters. There may be any number of ways that potassium transport, potassium availability, stomatal closure, or physiological responses could influence resistance to *A. linariae* (Melotto et al. 2017; Ragel et al. 2019). Metal tolerance proteins are cation efflux transporters that play roles in homeostasis, but little is known about their role in defense against fungal pathogens (Ricachenevsky et al. 2013). Eukaryotic translation initiation factors play important roles as susceptibility factors for virus infection, but less is known about the role they might play in early blight disease resistance (Piron et al. 2010).

Conclusion

We advanced the study of early blight resistance in tomato by narrowly defining two valuable QTL underlying early blight disease in modern cultivated breeding lines. This was accomplished not by traditional fine mapping, but by pairing historical pedigree information with publicly available genetic resources, including germplasm maintained by gene banks and breeders and the growing library of re-sequenced *Solanum* genomes. Our efficient method to identify shared introgressions in breeding materials and to predict local homology in public germplasm could be extended to any number of crops, provided the existence of a high-quality pedigree and sufficient whole genome sequence resources. We identified several genes potentially related to early blight resistance within the boundaries of the ancestral *EB-5* and *EB-9* introgressions. Because so little is known about the early blight pathosystem (and quantitative resistance to fungal pathogens more generally), further experiments are necessary to identify causal links between the genic polymorphisms we have highlighted and *in planta* early blight disease resistance. Future investigations into the expression levels of promising genes within *EB-5* and *EB-9* during disease progression would offer insight into the early blight pathosystem and improve our understanding of quantitative disease resistance.

REFERENCES

100 Tomato Genome Sequencing Consortium, Aflitos S, Schijlen E, et al (2014) Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. Plant Journal 80(1):136–148. https://doi.org/10.1111/tpj.12616

Adhikari P, Oh Y, Panthee DR (2017) Current status of early blight resistance in tomato: an update. International Journal of Molecular Sciences 18(10). https://doi.org/10.3390/ijms18102019

Anon (2016) Obituaries. In: Delaware Valley Ornithological Club. https://dvoc.org/wp/wp-content/uploads/2016/12/C72_73Obituaries.pdf. Accessed 2 Oct 2020

Ashrafi H, Foolad MR (2015) Characterization of early blight resistance in a recombinant inbred line population of tomato: II. Identification of QTLs and their co-localization with candidate resistance genes. Advanced Studies in Biology 7(4):149–168. https://doi.org/10.12988/asb.2015.41163

Barksdale TH (1971) Field evaluation for tomato early blight resistance. Plant Dis Report 55:807-809

Barksdale TH, Stoner AK (1973) Segregation for horizontal resistance to tomato early blight. Plant Dis Report 57(11):964-965

Barksdale TH, Stoner AK (1977) A study of the inheritance of tomato early blight (*Alternaria Solani*) resistance (fungal diseases). Plant Dis Report 61(2):63-65

Blanca J, Montero-Pau J, Sauvage C, et al (2015) Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. BMC Genomics 16:257. https://doi.org/10.1186/s12864-015-1444-1

Browning BL, Zhou Y, Browning SR (2018) A one-penny imputed genome from next-generation reference panels. Am J Hum Gen 103(3):338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Gen 81(5):1084–1097. https://doi.org/10.1086/521987

Caarls L, Elberse J, Awwanah M, et al (2017) Arabidopsis JASMONATE-INDUCED OXYGENASES down-regulate plant immunity by hydroxylation and inactivation of the hormone jasmonic acid. PNAS 114(24):6388–6393. https://doi.org/10.1073/pnas.1701101114

Causse M, Desplat N, Pascual L, et al (2013) Whole genome resequencing in tomato reveals variation associated with introgression and breeding events. BMC Genomics 14:791. https://doi.org/10.1186/1471-2164-14-791

Chen S, Wang X, Zhang L, et al (2016) Identification and characterization of tomato gibberellin 2-oxidases (GA2oxs) and effects of fruit-specific SlGA2ox1 overexpression on fruit and seed growth and development. Hort Res 3:1–9. https://doi.org/10.1038/hortres.2016.59

Chifman J, Kubatko L (2014) Quartet Inference from SNP Data Under the coalescent Model. Bioinformatics 30:3317–3324. https://doi.org/10.1093/bioinformatics/btu530

Chifman J, Kubatko L (2015) Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. J Theor Biol 374:35–47. https://doi.org/10.1016/j.jtbi.2015.03.006

Cingolani P, Platts A, Wang LL, et al (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w ^{1118} ; iso-2; iso-3. Fly 6:80–92. https://doi.org/10.4161/fly.19695

Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Der Sarkissian C, Ermini L, Schubert M, et al (2015) Evolutionary genomics and conservation of the endangered Przewalski's horse. Curr Biol 25:2577–2583. https://doi.org/10.1016/j.cub.2015.08.032

Fernandez-Pozo N, Menda N, Edwards JD, et al (2015) The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. Nucleic Acids Res 43:D1036–D1041. https://doi.org/10.1093/nar/gku1195

Fu Q, Hajdinjak M, Moldovan OT, et al (2015) An early modern human from Romania with a recent Neanderthal ancestor. Nature 524:216–219. https://doi.org/10.1038/nature14558

Gao L, Gonda I, Sun H, et al (2019) The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. Nat Genet 51:1044–1051. https://doi.org/10.1038/s41588-019-0410-2

Gardner R (1988) NC EBR-1 and NC EBR-2 early blight resistant tomato breeding lines. HortScience 23:779–781

Gardner RG (1990) Greenhouse disease screen facilitates breeding resistance to tomato early blight. HortScience 25:222–223. https://doi.org/10.21273/HORTSCI.25.2.222

Gardner RG, Panthee DR (2010) NC 1 CELBR and NC 2 CELBR: early blight and late blight-resistant fresh market tomato breeding lines. HortScience 45:975–976. https://doi.org/10.21273/HORTSCI.45.6.975

Green RE, Krause J, Briggs AW, et al (2010) A draft sequence of the Neandertal genome. Science 328:710–722. https://doi.org/10.1126/science.1188021

Hosmani PS, Flores-Gonzalez M, Geest H van de, et al (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

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755. https://doi.org/10.1093/bioinformatics/17.8.754

Le Roy J, Huss B, Creach A, et al (2016) Glycosylation Is a major regulator of phenylpropanoid availability and biological activity in plants. Front Plant Sci 7:735. https://doi.org/10.3389/fpls.2016.00735

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997 [q-bio]

Li H, Handsaker B, Wysoker A, et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lin T, Zhu G, Zhang J, et al (2014) Genomic analyses provide insights into the history of tomato breeding. Nature Genetics 46:1220–1226. https://doi.org/10.1038/ng.3117

Maiero M, Ng TJ, Barksdale T (1990) Genetic resistance to early blight in tomato breeding lines. HortScience 25:344–346

Maiero M, Ng TJ, Barksdale TH (1989) Combining ability estimates for early blight resistance in tomato. J Am Soc Hortic Sci 114(1):118-121

Malinsky M, Matschiner M, Svardal H (2020) Dsuite - fast D-statistics and related admixture evidence from VCF files. Molecular Ecology Resources 2020:1–12. https://doi.org/10.1111/1755-0998.13265

Maples BK, Gravel S, Kenny EE, Bustamante CD (2013) RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am J Hum Genet 93:278–288. https://doi.org/10.1016/j.ajhg.2013.06.020

Martin FW, Hepperly P (1987) Sources of resistance to early blight, *Alternaria solani*, and transfer to tomato, *Lycopersicon esculentum*. 1 71:85–95. https://doi.org/10.46429/jaupr.v71i1.6927

Martin GB, Brommonschenkel SH, Chunwongse J, et al (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262:1432–1436. https://doi.org/10.1126/science.7902614

Martin SH, Belleghem SMV (2017) Exploring Evolutionary relationships across the genome using topology weighting. Genetics 206:429–438. https://doi.org/10.1534/genetics.116.194720

Martin SH, Davey JW, Jiggins CD (2015) Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. Mol Biol Evol 32:244–257. https://doi.org/10.1093/molbev/msu269

Menda N, Strickler SR, Edwards JD, et al (2014) Analysis of wild-species introgressions in tomato inbreds uncovers ancestral origins. BMC Plant Biology 14:287 https://doi.org/10.1186/s12870-014-0287-2

Moreno JI, Martín R, Castresana C (2005) Arabidopsis SHMT1, a serine hydroxymethyltransferase that functions in the photorespiratory pathway influences resistance to biotic and abiotic stress. Plant J 41:451–463. https://doi.org/10.1111/j.1365-313X.2004.02311.x

Nash AF, Gardner RG (1988) Heritability of tomato early blight resistance derived from *Lycopersicon hirsutum* PI 126445. J Am Soc Hortic Sci 113(2):264-268

Pedregosa F, Varoquaux G, Gramfort A, et al (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830

Piron F, Nicolaï M, Minoïa S, et al (2010) An induced mutation in tomato eIF4E leads to immunity to two Potyviruses. PLOS ONE 5:e11313. https://doi.org/10.1371/journal.pone.0011313

Poplin R, Ruano-Rubio V, DePristo MA, et al (2018) Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv 201178. https://doi.org/10.1101/201178

Pritchard F, Porte W (1921) Collar-rot of tomato. J Agric Res 21:179–184

R Core Team (2020) A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Ragel P, Raddatz N, Leidi EO, et al (2019) Regulation of K+ Nutrition in Plants. Front Plant Sci 10:281. https://doi.org/10.3389/fpls.2019.00281

Rehman HM, Nawaz MA, Shah ZH, et al (2018) Comparative genomic and transcriptomic analyses of Family-1 UDP glycosyltransferase in three *Brassica* species and *Arabidopsis* indicates stress-responsive regulation. Scientific Reports 8:1875. https://doi.org/10.1038/s41598-018-19535-3

Reynard G, Andrus C (1944) Inheritance of resistance to the collar rot phase of *Alternaria solani* of tomato. Phytopathology 35:25-36

Ricachenevsky FK, Menguer PK, Sperotto RA, et al (2013) Roles of plant metal tolerance proteins (MTP) in metal storage and potential use in biofortification strategies. Front Plant Sci 4:144. https://doi.org/10.3389/fpls.2013.00144

Robinson JT, Thorvaldsdóttir H, Wenger AM, et al (2017) Variant review with the Integrative Genomics Viewer. Cancer Res 77:e31–e34. https://doi.org/10.1158/0008-5472.CAN-17-0337

Rotem J (1994) The Genus *Alternaria*: Biology, Epidemiology, and Pathogenicity. APS Press

Salter-Townshend M, Myers S (2019) Fine-scale inference of ancestry segments without prior knowledge of admixing groups. Genetics 212:869–889. https://doi.org/10.1534/genetics.119.302139

Sankararaman S, Sridhar S, Kimmel G, Halperin E (2008) Estimating local ancestry in admixed Populations. Am J Hum Genet 82:290–303. https://doi.org/10.1016/j.ajhg.2007.09.022

Sim S-C, Robbins MD, Wijeratne S, et al (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

Singh A, Pandey A, Srivastava AK, et al (2016) Plant protein phosphatases 2C: from genomic diversity to functional multiplicity and importance in stress management. Crit Rev in Biotechnol 36:1023–1035. https://doi.org/10.3109/07388551.2015.1083941

Strickler SR, Bombarely A, Munkvold JD, et al (2015) Comparative genomics and phylogenetic discordance of cultivated tomato and close wild relatives. PeerJ 3:e793. https://doi.org/10.7717/peerj.793

Sugimoto H, Kondo S, Tanaka T, et al (2014) Overexpression of a novel Arabidopsis PP2C isoform, AtPP2CF1, enhances plant biomass production by increasing inflorescence stem growth. J Exp Bot 65:5385–5400. https://doi.org/10.1093/jxb/eru297

Swofford D (2003) PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts

Thirthamallappa, Lohithaswa HC (2000) Genetics of resistance to early blight (*Alternaria solani* Sorauer) in tomato (*Lycopersicon esculentum* L.). Euphytica 113:187–193. https://doi.org/10.1023/A:1003929303632

Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178–192. https://doi.org/10.1093/bib/bbs017

Tieman D, Zhu G, Resende MFR, et al (2017) A chemical genetic roadmap to improved tomato flavor. Science 355:391–394. https://doi.org/10.1126/science.aal1556

van den Burg HA, Tsitsigiannis DI, Rowland O, et al (2008) The F-Box Protein ACRE189/ACIF1 regulates cell death and defense responses activated during pathogen recognition in tobacco and tomato. Plant Cell 20:697–719. https://doi.org/10.1105/tpc.107.056978

Wang P, Gao C, Bian X, et al (2016) Genome-wide identification and comparative analysis of cytosine-5 DNA methyltransferase and demethylase families in wild and cultivated peanut. Front Plant Sci 7:7. https://doi.org/10.3389/fpls.2016.00007

Woudenberg JHC, Truter M, Groenewald JZ, Crous PW (2014) Large-spored *Alternaria* pathogens in section *Porri* disentangled. Stud Mycol 79:1–47. https://doi.org/10.1016/j.simyco.2014.07.003

Yang WC, Francis DM (2005) Marker-assisted selection for combining resistance to bacterial spot and bacterial spark in tomato. J Am Soc Hortic Sci 130(5):716-721

Zitter T, Drennan J (2007) Using host resistance and light fungicides to control early blight on tomato. Plant Dis. Manag. Rep. 2:60. Plant Dis Manag Rep 2:60

Zitter T, Drennan J (2008) Using host resistance and lite fungicides to control early blight and Septoria leaf spot on tomato. Plant Dis Manag Rep 2:61

Zitter TA, Drennan JL, Mutschler MA, Kim MJ (2005) Control of early blight of tomato with genetic resistance and conventional and biological sprays. Acta Horticulturae 695:181–190. https://doi.org/10.17660/ActaHortic.2005.695.20