UC Davis
UC Davis Electronic Theses and Dissertations

Title
Using Wild Tomato Species Introgressions to Breed for Water Stress Tolerance in Cultivated Tomato

Permalink
https://escholarship.org/uc/item/31n303s1

Author
Groh, Amy M

Publication Date
2022

Supplemental Material
https://escholarship.org/uc/item/31n303s1#supplemental

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Using Wild Tomato Species Introgressions to Breed for Water Stress Tolerance in Cultivated Tomato

By

AMY MICHELLE GROH

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Horticulture and Agronomy

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Dina St. Clair, Chair

Patrick Brown

Roger Chetelat

Committee in Charge

2022

ACKNOWLEDGEMENTS

Many people are involved in a dissertation besides the degree recipient and they deserve more than just thanks, but for now here is a simple thank you.

I would like to thank my major professor, Dina St. Clair, for her guidance and understanding throughout my degree. Thank you for the opportunity to learn and develop as a scientist and person. I would also like to thank my committee, Pat Brown and Roger Chetelat, for their input, advice, and time. Finally, I'd like to thank my lab mate Bryce for helping when I started in the lab and for his camaraderie over the years.

Next, I would like to thank some special family and friends. My parents, Jan and Randy, and my siblings, Allie and Jake, were supportive of my work even though it took me far away. They were always open to a call or visit and reminded me to have some fun. Brendan, Saarah, and Sam helped me get through the hard times, celebrate the good times, and persevere through finishing graduate school during a global pandemic.

There are many other people at UC Davis who helped me, taught me, had fun with me, and shared their passion for plant sciences with me. I am grateful for Antonia and Laura from SCOPE, Kalyn and Wencin from my dissertation work group, the field and greenhouse crews, and the professors who helped me figure things out along the way.

TABLE OF CONTENTS

Abstract……...iv

Introduction…….……1

Chapter 1……...….11

Chapter 2……43

Chapter 3……85

Conclusions and Future Directions……..…126

ABSTRACT

Given climate change and population growth, it is increasingly important to efficiently grow food crops such as tomato (*Solanum lycopersicum*) using less irrigation water, a limited resource. Cultivated tomato is susceptible to water stress. In contrast, wild tomato species such as *S. habrochaites* and *S. pennellii* are water stress tolerant. The overall goal of this dissertation research was to investigate specific chromosomal regions of *S. habrochaites* and *S. pennellii* that are associated with water stress tolerance in the field and their potential use in breeding cultivated tomato. Indirect measures of water stress tolerance were employed, including carbon isotope discrimination ($\Delta^{13}C$, an indirect measurement of water use efficiency, WUE), shoot dry weight, and yield to evaluate the various tomato genotypes derived from these two wild species that were grown under field conditions.

In chapter one, the goal was to fine map quantitative trait loci (QTL) for maturity, yield, shoot dry weight, and $\Delta^{13}C$ on chromosome 9 of *S. habrochaites* using a set of 21 sub-near-isogenic lines (sub-NILs) in a cultivated tomato background. These sub-NILs and control lines were evaluated in replicated field experiments grown under full and reduced irrigation treatments across two years. Significant QTL were detected for 20 trait-environment combinations. All QTL were partially or completely coincident. The presence of the *S. habrochaites* allele at QTL was associated with later maturity, lower yields, and greater WUE. Further mapping work is required before these QTL can be implemented in a marker assisted breeding program.

In chapter two, cultivated tomato lines containing wild species chromosome introgressions (introgression lines, ILs) on chromosomes 2, 5, and 9 from *S. habrochaites* and *S. pennellii* were studied in various combinations in hybrid progeny. The goal was to determine if hybrid combinations of ILs with introgressions from *S. habrochaites* and *S. pennellii* improved WUE of

genotypes compared to their parent ILs and/or inbred cultivated tomato varieties. A total of 35 genotypes (including two inbred tomato cultivars, homozygous ILs, heterozygous ILs, and hybrids between the two species ILs) were evaluated in the field for maturity, horticultural, yield, and WUE-related ($\Delta^{13}C$) traits across two years under full and reduced irrigation treatments. General and specific combining abilities were calculated for the hybrid offspring in a Design II mating design. Hybrids generally had best-parent heterosis for yield, but not $\Delta^{13}C$. Three genotypes had better WUE than both cultivated tomato controls for the genetic background and did not cause significant reductions in yield. Some introgressions that showed positive effects on WUE also exhibited negative effects on other traits, such as fruit weight or maturity. Specific introgressions from both wild species may prove useful for developing hybrid tomato varieties that have acceptable yields yet use less irrigation water in the field.

In chapter three, a *de novo* whole-genome sequence (WGS) of *S. habrochaites* accession LA1778 was generated using the PacBio RSII system and a Bionano optical map. Twelve pseudomolecules, corresponding to the 12 chromosomes of tomato, were constructed with a total length of approximately 959 Mb. The *S. habrochaites* WGS had high gene and sequence synteny with the *S. lycopersicum* reference annotation and sequence, although some inversions were detected. The *S. habrochaites* chromosome 9 sequence was used to determine the physical locations of previously mapped QTL for tolerance to rapid-onset water stress (*stm9*), WUE ($\Delta^{13}C$), and other horticultural traits. A list of potential genes within each QTL interval was generated using *S. lycopersicum* genes that mapped to each interval, predicted genes, as well as mapped transcript reads from *S. habrochaites* acc. LA1777. It is likely that many of these genes can be affected by water stress and some may play a role in the tolerant phenotypes. Further work is

required to determine which genes confer the slow-onset and rapid-onset water stress tolerant phenotypes.

Some of the wild genomic regions have both negative and positive effects on horticultural traits (i.e., linkage drag). Additional studies need to be conducted to fractionate the QTL that are in repulsion phase linkage. High-resolution QTL mapping can lead to the integration of positive WUE-related loci without negative horticultural effects via marker assisted selection. Collectively, these studies support the conclusion that *S. habrochaites* and *S. pennellii* have potential promising uses for breeding cultivated tomato with improved water stress tolerance.

INTRODUCTION

Cultivated tomato (*Solanum lycopersicum*) is an important vegetable crop globally and in the United States (Liedl et al., 2013). In 2019, tomatoes accounted for about 14% of the worldwide vegetable food supply (measured in kg per capita per year) with a gross production value of $102 billion USD globally (FAOSTAT, 2019). In the US, tomato is a major vegetable crop and is the second most consumed vegetable after potatoes (USDA ERS, 2019). In 2020, the US produced 12 million tonnes of tomatoes, with a value of about $15 billion USD (FAOSTAT, 2020).

Cultivated tomato is less genetically diverse than wild tomato species (Miller & Tanksley, 1990; Rick, 1983). Tomato underwent domestication in South America and Meso America, then was subjected to successive genetic bottlenecks during its introduction to Europe and to North America (Blanca et al., 2015; Labate et al., 2007; Miller & Tanksley, 1990). The limited amount of genetic diversity in cultivated tomato is likely due to several causes: genetic bottlenecks during domestication; self-pollination (unlike self-incompatible wild species); and breeding efforts focused on a narrow range of germplasm with the most desirable traits. Miller & Tanksley (1990) reported that for modern tomato cultivars, less than 10% of the genetic variation detected with DNA markers was within-accession variation. Modern tomato varieties are comparatively more genetically diverse than older cultivated tomato varieties primarily due to introgression of disease and pest resistance genes from wild species (Blanca et al., 2015).

The self-incompatible wild tomato species possess more genetic variation both within and between accessions than the self-compatible accessions (Miller & Tanksley, 1990). The obligate outcrossing species *S. habrochaites* and *S. pennellii* have some of the highest average genetic distances between and within accessions (Miller & Tanksley, 1990). Wild tomato species originate from a wide geographic range in western South America (e.g., Peru, Chile, and Ecuador) for which

they evolved adaptations to different environmental conditions, including low precipitation and cyclic droughts (Rick, 1973, 1983). Two accessions of *S. habrochaites*, LA1777 and LA1778, were collected at a high altitudes (3216 m and 3002 m, respectively) in a Peruvian mountain valley at sites approximately 5 km apart (https://tgrc.ucdavis.edu). This area has a low mean temperature (5.7 ºC), suggesting that these accessions have adaptations to chilling temperatures (https://tgrc.ucdavis.edu; Easlon et al., 2013). In contrast, *S. pennellii* acc. LA716 was collected from a dry, rocky, coastal area of Peru with an altitude of approximately 50 m and has been studied for water and salt stress tolerance due to the conditions in its natural environment (https://tgrc.ucdavis.edu).

Wild species relatives of crops are most often used in breeding as germplasm sources of biotic stress tolerances, but have also been employed to integrate quality traits and abiotic stress tolerances in cultivated crops (Dempewolf et al., 2017; Hajjar & Hodgkin, 2007). *S. habrochaites* and *S. pennellii* have many beneficial traits that can be used to improve cultivated tomato. *S. habrochaites* contains QTL associated with fruit color, soluble solids (brix), and firmness (Bernacchi et al., 1998; Monforte et al., 2001). *S. habrochaites* is also a source of biotic stress tolerances including resistance to *Alternaria solani* (early blight), *Phytophthora infestans* (late blight), tomato mosaic virus (*Tm-1* resistance gene), and a variety of insect pests (largely due to trichome-mediated defenses) (Brouwer & St. Clair, 2004; Foolad et al., 2002; Kennedy, 2003; Ohmori et al., 1996). Loci associated with tolerance to cold stress, chilling stress, and water stress have also been discovered in *S. habrochaites* (Arms et al., 2015; Chen et al., 2015; Lounsbery et al., 2016). A close relative of *S. habrochaites, S. pennellii,* is a source of fruit quality traits such as fruit soluble solids, color, and phenolics (Fridman et al., 2004; Liu et al., 2003; Rousseaux et al., 2005).

Wild species also carry negative effect alleles for various traits that are detrimental to the horticultural performance of elite crop varieties. It is not unusual for traits from wild species of tomato to exhibit negative linkage drag effects on horticultural and yield traits (Brouwer & St. Clair, 2004; Haggard et al., 2013). If linkage drag is due to repulsion phase linkages among QTL, recombination and subsequent segregation would be required to obtain individuals with coupling phase alleles to select for while breeding for improved cultivated tomato.

Molecular and genetic tools associated with wild tomato species facilitate the efficient and effective utilization of wild species in breeding (Dempewolf et al., 2017). Whole genome sequences (WGSs) can be used in various ways to aid in plant breeding. WGS can be used to detect genome-wide genetic variations (SNPs, indels, or other polymorphisms) that are necessary for determining marker-trait associations, identify candidate genes for traits, and for transcriptome analyses (Hu et al., 2018; Kim et al., 2020). Molecular markers (polymorphisms) associated with phenotypic variation in traits of interest and fine mapping can aid in the efficient integration of traits into existing varieties while potentially avoiding linkage drag via use of marker-assisted selection (MAS) (Collard & Mackill, 2008; Dempewolf et al., 2017). MAS can save resources, time, and sometimes be more precise than classical breeding because it may allow for earlier testing without phenotyping and can utilize background selection to avoid integration of undesirable sequences outside of a target QTL (Collard & Mackill, 2008).

Introgressions line (IL) libraries, which are sets of related inbred lines that each contain unique wild species introgressions, are another tool that can assist breeding. ILs enable association of traits to specific wild introgressions when compared to a cultivated tomato inbred control; they also facilitate integration of wild chromosome regions controlling valuable traits into cultivated tomato (Gur et al., 2004; Lippman et al., 2007). The University of California Davis Tomato

Genetics Resource Center (TGRC, https://tgrc.ucdavis.edu) maintains and distributes seed for tomato IL libraries containing introgressions from *S. habrochaites* acc. LA1777 and *S. pennellii* acc. LA716. Each library collectively represents the majority of the genome of each species (Eshed & Zamir, 1995; Monforte & Tanksley, 2000). The wild introgressions are in the genetic background of two inbred processing tomato varieties, E6203 and M82, for the *S. habrochaites* and *S. pennellii* ILs, respectively.

Fresh water is a limited resource in many parts of the world, such as the arid Western US that includes California. Climate change is associated with increasingly variable annual precipitation, which leads to unpredictable water availability for irrigated and non-irrigated agricultural production systems (Pathak et al., 2018). Tomato yields can be negatively affected by limited precipitation and reduced irrigation water availability (Nuruddin et al., 2003). Thus, growers have an increasing need for tomato varieties that are productive with low water inputs and are more resilient to highly variable weather conditions. Breeding tomato varieties that use less water yet maintain adequate yields will help growers contain production costs, enhance sustainability, and reduce water use by agriculture.

Breeding crops that use less water requires plant breeders to identify traits to target for selection. Water stress tolerance traits can be related to physiology, timing of growth, and/or indirect measures of water use efficiency (WUE). Early plant vigor, earlier planting varieties, and specific root architectures (depending on the environment and agricultural management practices) are some potential breeding selection targets (Schmidt & Gaudin, 2017; Tuberosa, 2012; Turner & Begg, 1981). WUE at the leaf level is defined as the ratio of carbon assimilation from photosynthesis to transpiration rate (Merchuk & Saranga, 2013). Agronomic WUE relates to the amount of biomass produced per unit of water transpired by a plant (Merchuk & Saranga, 2013).

A genotype producing the same yield of a crop using less water has greater (better) WUE. WUE can be challenging to measure directly in large field experiments, so indirect measures of WUE, such as carbon isotope discrimination ($\Delta^{13}C$), are often utilized in plant breeding evaluations (Merchuk & Saranga, 2013; Tuberosa, 2012). In this dissertation, I focus on identifying introgressions from wild tomato species that are associated with higher WUE in both water-stressed and non-stressed irrigation treatments in the field.

S. habrochaites and *S. pennellii* can be genetic sources of water stress-tolerance traits, and positive effect QTL alleles for these traits can be incorporated into cultivated tomato. Previous studies have identified specific regions of the genome in each species that may be associated with WUE. Truco et al. (2000) mapped QTL for shoot turgor maintenance under rapid-onset water stress caused by root chilling to chromosomes 5 and 9 of *S. habrochaites* LA1778 in an interspecific backcross population. The chromosome 9 QTL (designated *stm9*) was fine mapped by Arms et al. (2015) to a 0.32 cM region using a set of recombinant sub-near-isogenic lines (sub-NILs). The ability of *S. habrochaites* to maintain shoot turgor under chilling stress was related to the ability to close stomata and reduce stomatal conductance (Bloom et al., 2004).

Because *S. habrochaites* can control its stomatal aperture under rapid-onset water stress due to root chilling, it was hypothesized that it could exert stomatal control during slow-onset water stress due to deficit irrigation and exhibit greater WUE. Using the same set of sub-NILs as Arms et al. (2015), Lounsbery et al. (2016) mapped QTL for water stress-related traits, yield, and maturity-related traits on the short arm of chromosome 9 from *S. habrochaites* under full and deficit irrigation treatments. Lounsbery et al. (2016) mapped QTL for $\Delta^{13}C$ in this region of *S. habrochaites*. However, most of these QTL intervals were not fully resolved within the 1.28 cM region of chromosome 9 present in the sub-NILs. These QTL for WUE-related traits and

horticultural traits would not be feasible to use in breeding without additional fine mapping efforts to determine the amount of overlap and linkage among QTL associated with different traits.

In studies by other researchers, *S. pennellii* ILs with an introgression on chromosomes 2, 5, or 9 exhibited superior water stress tolerance compared to their cultivated tomato parent (Barrios-Masias et al., 2013; Gong et al., 2010; Rigano et al., 2016; Xu et al., 2008). In these studies, the water-stressed ILs showed higher survival rates, higher leaf dry matter content, and carbon isotope composition ($\delta^{13}C$, a value related to $\Delta^{13}C$) values that were more consistent with better WUE than the recurrent parent control.

The overall goal of this dissertation research was to create genomic tools and identify genotypes derived from wild tomato germplasm that can be useful for breeding improved WUE lines of cultivated tomato. The objective of chapter one was to fine map QTL for important horticultural and WUE-related traits on the short arm of chromosome 9 in *S. habrochaites*, a follow-up study to Lounsbery et al. (2016). Towards this goal, a new set of 21 sub-NILs with introgressions spanning 3.6 cM was generated and evaluated in the field under full and reduced water treatments for two years. Chapter two had three specific objectives. First, to determine if introgressions in ILs derived from two wild tomato species (*S. habrochaites* and *S. pennellii*) enhance WUE-related traits compared to their IL parents when combined together in hybrid progeny. Second, to determine if the ILs performed better for WUE in the field than the control cultivars. Third, to identify ILs (and their introgressions) that show promise for use in breeding hybrid tomatoes with improved WUE and without negative effects on horticultural traits. The objectives of chapter three were to create a high-quality WGS assembly for *S. habrochaites* acc. LA1778, and to identify potential candidate genes associated with chromosome 9 QTL intervals for *stm9*, $\Delta^{13}C$, and horticultural traits described by Arms et al. (2015) and Groh et al. (2022).

REFERENCES

Arms, E. M., Bloom, A. J., & St. Clair, D. A. (2015). High-resolution mapping of a major effect QTL from wild tomato *Solanum habrochaites* that influences water relations under root chilling. *Theoretical and Applied Genetics*, *128*(9), 1713–1724. https://doi.org/10.1007/s00122-015-2540-y

Barrios-Masias, F. H., Chetelat, R. T., Grulke, N. E., & Jackson, L. E. (2013). Use of introgression lines to determine the ecophysiological basis for changes in water use efficiency and yield in California processing tomatoes. *Functional Plant Biology*, *41*(2), 119–132.

Bernacchi, D., Beck-Bunn, T., Eshed, Y., Lopez, J., Petiard, V., Uhlig, J., Zamir, D., & Tanksley, S. (1998). Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from *Lycopersicon hirsutum*. *Theoretical and Applied Genetics*, *97*(3), 381–397.

Blanca, J., Montero-Pau, J., Sauvage, C., Bauchet, G., Illa, E., Díez, M. J., Francis, D., Causse, M., Van der Knaap, E., & Cañizares, J. (2015). Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *BMC Genomics*, *16*(1), 1–19.

Bloom, A. J., Zwieniecki, M. A., Passioura, J. B., Randall, L. B., Holbrook, N. M., & St. Clair, D. A. (2004). Water relations under root chilling in a sensitive and tolerant tomato species. In *Plant, Cell and Environment* (Vol. 27, pp. 971–979).

Brouwer, D., & St. Clair, D. (2004). Fine mapping of three quantitative trait loci for late blight resistance in tomato using near isogenic lines (NILs) and sub-NILs. *Theoretical and Applied Genetics*, *108*(4), 628–638.

Chen, H., Chen, X., Chen, D., Li, J., Zhang, Y., & Wang, A. (2015). A comparison of the low temperature transcriptomes of two tomato genotypes that differ in freezing tolerance: *Solanum lycopersicum* and *Solanum habrochaites*. *BMC Plant Biology*, *15*(1), 1–16.

Collard, B. C., & Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1491), 557–572.

Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., & Guarino, L. (2017). Past and future use of wild relatives in crop breeding. *Crop Science*, *57*(3), 1070–1082. https://doi.org/10.2135/cropsci2016.10.0885

Easlon, H. M., Asensio, J. S. R., St. Clair, D. A., & Bloom, A. J. (2013). Chilling-induced water stress: Variation in shoot turgor maintenance among wild tomato species from diverse habitats. *American Journal of Botany*, *100*(10), 1991–1999. https://doi.org/10.3732/ajb.1200508

Eshed, Y., & Zamir, D. (1995). An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics*, *141*(3), 1147–1162.

FAOSTAT. (2019). *FAOSTAT Data*. https://www.fao.org/faostat/en/#data

FAOSTAT. (2020). *FAOSTAT Data*. https://www.fao.org/faostat/en/#data

Foolad, M., Zhang, L., Khan, A. A., Niño-Liu, D., & Lin, G. (2002). Identification of QTLs for early blight (*Alternaria solani*) resistance in tomato using backcross populations of a *Lycopersicon esculentum* × *L. hirsutum* cross. *Theoretical and Applied Genetics*, *104*(6), 945–958. https://doi.org/10.1007/s00122-002-0870-z

Fridman, E., Carrari, F., Liu, Y.-S., Fernie, A. R., & Zamir, D. (2004). Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science*, *305*(5691), 1786–1789. https://doi.org/10.1126/science.1101666

Gong, P., Zhang, J., Li, H., Yang, C., Zhang, C., Zhang, X., Khurram, Z., Zhang, Y., Wang, T., Fei, Z., & Ye, Z. (2010). Transcriptional profiles of drought-responsive genes in modulating transcription signal transduction, and biochemical pathways in tomato. *Journal of Experimental Botany*, *61*(13), 3563–3575. https://doi.org/10.1093/jxb/erq167

Groh, A. M., Kubond, B. A., & St. Clair, D. A. (2022). Fine mapping of QTL for water use efficiency-related traits on chromosome 9 of *Solanum habrochaites* in the field. *Crop Science*. DOI: 10.1002/csc2.20828. (*in press*).

Gur, A., Zamir, D., & Dangl, J. (2004). Unused natural variation can lift yield barriers in plant breeding. *PLoS Biology*, *2*(10), e245.

Haggard, J. E., Johnson, E. B., & St. Clair, D. A. (2013). Linkage relationships among multiple QTL for horticultural traits and late blight (*P. infestans*) resistance on chromosome 5 introgressed from wild tomato *Solanum habrochaites*. *G3: Genes, Genomes, Genetics*, *3*(12), 2131–2146. https://doi.org/10.1534/g3.113.007195

Hajjar, R., & Hodgkin, T. (2007). The use of wild relatives in crop improvement: A survey of developments over the last 20 years. *Euphytica*, *156*(1), 1–13. https://doi.org/10.1007/s10681-007-9363-0

Hu, H., Scheben, A., & Edwards, D. (2018). Advances in integrating genomics and bioinformatics in the plant breeding pipeline. *Agriculture*, *8*(6). https://doi.org/10.3390/agriculture8060075

Kennedy, G. G. (2003). Tomato, pests, parasitoids, and predators: Tritrophic interactions involving the genus *Lycopersicon*. *Annual Review of Entomology*, *48*(1), 51–72.

Kim, K. D., Kang, Y., & Kim, C. (2020). Application of genomic big data in plant breeding: Past, present, and future. *Plants*, *9*(11). https://doi.org/10.3390/plants9111454

Labate, J. A., Grandillo, S., Fulton, T., Muños, S., Caicedo, A. L., Peralta, I., Ji, Y., Chetelat, R. T., Scott, J., Gonzalo, M. J., & others. (2007). Tomato. In *Vegetables* (pp. 1–125). Springer.

Liedl, B. E., Labate, J. A., Stommel, J. R., Slade, A., & Kole, C. (2013). *Genetics, genomics, and breeding of tomato*. CRC Press.

Lippman, Z. B., Semel, Y., & Zamir, D. (2007). An integrated view of quantitative trait variation using tomato interspecific introgression lines. *Current Opinion in Genetics & Development*, *17*(6), 545–552.

Liu, Y.-S., Gur, A., Ronen, G., Causse, M., Damidaux, R., Buret, M., Hirschberg, J., & Zamir, D. (2003). There is more to tomato fruit colour than candidate carotenoid genes. *Plant Biotechnology Journal*, *1*(3), 195–207. https://doi.org/10.1046/j.1467-7652.2003.00018.x

Lounsbery, J. K., Arms, E. M., Bloom, A. J., & St. Clair, D. A. (2016). Quantitative trait loci for water-stress tolerance traits localize on chromosome 9 of wild tomato. *Crop Science*, *56*(4), 1514–1525. https://doi.org/10.2135/cropsci2015.07.0432

Merchuk, L., & Saranga, Y. (2013). Breeding approaches to increasing water use efficiency. *Improving Water and Nutrient-Use Efficiency in Food Production Systems. John Wiley & Sons Inc, Hoboken*, 145–160.

Miller, J., & Tanksley, S. (1990). RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon. *Theoretical and Applied Genetics*, *80*(4), 437–448.

Monforte, A., Friedman, E., Zamir, D., & Tanksley, S. (2001). Comparison of a set of allelic QTL-NILs for chromosome 4 of tomato: Deductions about natural variation and implications for germplasm utilization. *Theoretical and Applied Genetics*, *102*(4), 572–590.

Monforte, A., & Tanksley, S. (2000). Fine mapping of a quantitative trait locus (QTL) from *Lycopersicon hirsutum* chromosome 1 affecting fruit characteristics and agronomic traits: Breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. *Theoretical and Applied Genetics*, *100*(3), 471–479.

Nuruddin, M. M., Madramootoo, C. A., & Dodds, G. T. (2003). Effects of water stress at different growth stages on greenhouse tomato yield and quality. *HortScience*, *38*(7), 1389–1393.

Ohmori, T., Murata, M., & Motoyoshi, F. (1996). Molecular characterization of RAPD and SCAR markers linked to the Tm-1 locus in tomato. *Theoretical and Applied Genetics*, *92*(2), 151–156. https://doi.org/10.1007/BF00223369

Pathak, T. B., Maskey, M. L., Dahlberg, J. A., Kearns, F., Bali, K. M., & Zaccaria, D. (2018). Climate change trends and impacts on California agriculture: A detailed review. *Agronomy*, *8*(3), 25.

Rick, C. M. (1973). Potential genetic resources in tomato species: Clues from observations in native habitats. In *Genes, enzymes, and populations* (pp. 255–269). Springer.

Rick, C. M. (1983). Genetic variability in tomato species. *Plant Molecular Biology Reporter*, *1*(2), 81–87.

Rigano, M. M., Arena, C., Matteo, A. D., Sellitto, S., Frusciante, L., & Barone, A. (2016). Eco-physiological response to water stress of drought-tolerant and drought-sensitive tomato genotypes. *Plant Biosystems*, *150*(4), 682–691. https://doi.org/10.1080/11263504.2014.989286

Rousseaux, M. C., Jones, C. M., Adams, D., Chetelat, R., Bennett, A., & Powell, A. (2005). QTL analysis of fruit antioxidants in tomato using *Lycopersicon pennellii* introgression lines. *Theoretical and Applied Genetics*, *111*(7), 1396–1408. https://doi.org/10.1007/s00122-005-0071-7

Schmidt, J. E., & Gaudin, A. C. M. (2017). Toward an integrated root ideotype for irrigated systems. *Trends in Plant Science*, *22*(5), 433–443. https://doi.org/10.1016/j.tplants.2017.02.001

Truco, M., Randall, L., Bloom, A., & St. Clair, D. (2000). Detection of QTLs associated with shoot wilting and root ammonium uptake under chilling temperatures in an interspecific backcross population from *Lycopersicon esculentum*× *L. hirsutum*. *Theoretical and Applied Genetics*, *101*(7), 1082–1092.

Tuberosa, R. (2012). Phenotyping for drought tolerance of crops in the genomics era. *Frontiers in Physiology*, *3 SEP*. https://doi.org/10.3389/fphys.2012.00347

Turner, N. C., & Begg, J. E. (1981). Plant-water relations and adaptation to stress. *Plant and Soil*, *58*(1), 97–131. https://doi.org/10.1007/BF02180051

USDA ERS. (2019). *Potatoes and tomatoes are the most commonly consumed vegetables*. https://www.ers.usda.gov/data-products/chart-gallery/gallery/chart-detail/?chartId=58340

Xu, X., Martin, B., Comstock, J. P., Vision, T. J., Tauer, C. G., Zhao, B., Pausch, R. C., & Knapp, S. (2008). Fine mapping a QTL for carbon isotope composition in tomato. *Theoretical and Applied Genetics*, *117*(2), 221–233. https://doi.org/10.1007/s00122-008-0767-6

CHAPTER 1

FINE MAPPING OF QTL FOR WATER USE EFFICIENCY-RELATED TRAITS ON CHROMOSOME 9 OF *SOLANUM HABROCHAITES* IN THE FIELD

Amy M. Groh, Bryce A. Kubond, and Dina A. St. Clair

Groh, A. M., Kubond, B. A., & St. Clair, D. A. (2022). Fine mapping of QTL for water use efficiency-related traits on chromosome 9 of *Solanum habrochaites* in the field. *Crop Science*. DOI: 10.1002/csc2.20828. (*in press*).

ABSTRACT

Fresh water is increasingly limited due to climate change. One way to reduce fresh water use by irrigated agriculture is to breed plants that can produce adequate yields with less water. A wild tomato, *Solanum habrochaites*, is more water stress tolerant than cultivated tomato, *S. lycopersicum*. Our study aimed to fine map quantitative trait loci (QTL) for maturity, yield, and water use efficiency-related (WUE) traits (shoot dry weight and carbon isotope discrimination, $\Delta^{13}C$) on chromosome 9 in *S. habrochaites* using a set of 21 sub-near-isogenic lines (sub-NILs). These sub-NILs and controls were evaluated in field experiments grown under full and reduced irrigation treatments across two years. Genotype by environment interactions were significant, so the data was separated into subsets of environments for further analyses. Significant QTL were detected for 20 trait-environment combinations. All QTL were partially or completely coincident. The presence of the *S. habrochaites* allele at QTL was associated with later maturity, lower yields, and greater WUE. However, yield and $\Delta^{13}C$ exhibited no or low correlations with each other, suggesting these traits could be selected for improvement independently during breeding. Two of

21 sub-NILs showed superior performance for WUE and yield compared to controls. These two genotypes may prove useful in breeding programs to improve the WUE of tomato cultivars without negative effects on yield.

INTRODUCTION

Fresh water is a limited resource in many parts of the world, such as the arid Western US that includes California. Climate change is associated with increasingly variable annual precipitation, which leads to unpredictable water availability for irrigated and non-irrigated agricultural production systems (Pathak et al., 2018). Thus, growers have an increasing need for crop varieties that are productive with low water inputs and are more resilient to highly variable weather conditions. Breeding crops that use less water yet maintain adequate yields will help growers contain production costs, enhance sustainability, and reduce water use by agriculture.

Cultivated tomato, *Solanum lycopersicum*, is an important vegetable crop worldwide. In 2020, over 186 million tonnes of tomatoes were produced globally, and China, India, Turkey, the United States, and Egypt were the top five producers (FAOSTAT, 2020). Within the US, tomato is economically important in California, with a value of about $1.17 billion USD in 2019 for fresh market and processing tomatoes combined (CDFA, 2020). In 2019, tomatoes accounted for over 247,000 harvested acres in California, most of which is produced in the Central Valley (CDFA, 2020). California has a Mediterranean climate with cool, wet winters and hot, dry summers. This climate requires irrigation for producing summer crops, such as tomato. Restricting irrigation water applications to less than the crop evapotranspiration rate can decrease yields in tomato varieties, especially when reduced at early growth stages (Albert et al., 2016; Nuruddin et al., 2003; Patanè & Cosentino, 2010). However, due to limited fresh water availability there is a need to breed tomato varieties that can produce acceptable yields with less applied water.

An approach to improving water stress tolerance is to increase crop water use efficiency (WUE). WUE is defined as the ratio of carbon assimilation from photosynthesis to transpiration rate and relates to the amount of biomass produced per unit of water transpired by a plant. Direct measurement of WUE in the field is expensive, technically difficult, and time consuming, so indirect measures are often used by field researchers (Jones, 2007; Tuberosa, 2012). Direct measures of photosynthetic rate and transpiration rate are also instantaneous snapshots of the plant's WUE. An instantaneous measure of WUE only provides data about a narrow moment in time at a particular growth stage, rather than an evaluation of how a genotype performs for WUE throughout a growing season. Martin and Thorstenson (1988) showed that instantaneous measures of WUE (net photosynthesis/transpiration) do not correlate well with season long measures of WUE and therefore should not be used as proxies.

Carbon isotope discrimination is an indirect measurement of WUE over an entire season and has been shown to be correlated with WUE in crops, including tomato (Farquhar & Richards, 1984; Martin et al., 1999; Martin & Thorstenson, 1988). During photosynthesis, there is discrimination against CO_2 molecules containing the heavier naturally occurring carbon isotope, ^{13}C, compared to the lighter, more common isotope, ^{12}C. When leaves reduce their stomatal aperture in response to water stress, they lose less water and reduce the concentration of CO_2 in the intracellular airspace (C_i) that is available for photosynthesis. This lower C_i leads to a decrease in discrimination against ^{13}C. As the stomatal aperture decreases, the rate of both photosynthesis and transpiration decrease, but photosynthesis decreases at a slower rate than transpiration, so the WUE (ratio of carbon assimilation to transpiration rate) increases (Farquhar & Sharkey, 1982). Thus, the ratio of Ber^{13}C to ^{12}C assimilated into plant biomass can be used to indirectly measure WUE over a growing season. Plants that have a higher isotopic composition ($\delta^{13}C$, the ratio of

^{13}C:^{12}C in a plant sample as compared to the Pee Dee Belemnite standard) are more water use efficient. The discrimination of ^{13}C ($\Delta^{13}C$) compares the $\delta^{13}C$ of the plant to that of the atmosphere and is negatively correlated with WUE. Genotypes with smaller $\Delta^{13}C$ values discriminate against ^{13}C less and are more water use efficient (Farquhar & Richards, 1984).

Wild species of plants may exhibit increased tolerance to abiotic stresses compared to their domesticated relatives (Bayuelo-Jiménez et al., 2002; Rick, 1983). These wild species can be a source of genes for beneficial traits that can be introgressed into cultivars via breeding programs. In tomato, the wild species *S. habrochaites* is more tolerant to various abiotic stresses, including chilling and deficit irrigation, compared to cultivated tomato (Lounsbery et al., 2016; Truco et al., 2000). Previous research in our lab has shown that chromosome 9 of *S. habrochaites* contains QTL associated with tolerance to rapid-onset water stress caused by root chilling and tolerance to slow-onset water stress (measured using $\Delta^{13}C$) from deficit irrigation in the field (Arms et al., 2015; Lounsbery et al., 2016). Wild species, including *S. habrochaites*, can also exhibit horticulturally undesirable traits such as late maturity and lower yields (Brouwer & St. Clair, 2004; Lounsbery et al., 2016). Thus, horticultural traits should be examined together with WUE-related traits to determine how the wild species introgressions may affect the horticultural performance of a genotype.

Lounsbery et al. (2016) mapped QTL for water stress-related traits, yield, and maturity-related traits on the short arm of chromosome 9 from *S. habrochaites*. Most of these QTL intervals were not resolved with the 18 sub-near isogenic lines (sub-NILs) representing a 1.28 cM region of chromosome 9 in an otherwise cultivated tomato background. Using these QTL for marker assisted selection is not feasible without additional fine mapping efforts to determine the amount of overlap and linkage among QTL associated with different traits.

The overall goal of the current study was to fine map QTL for important horticultural and WUE-related traits on the short arm of chromosome 9 in *S. habrochaites*. Towards this goal, a new set of 21 sub-NILs spanning 3.6 cM was generated for this study. The introgressions include the centromeric end of the region studied by Lounsbery et al. (2016) and extend further towards the centromere. The 21 sub-NILs were evaluated in the field under both deficit and normal irrigation over two years and the data was used for QTL fine mapping.

MATERIALS AND METHODS

Plant Materials

The *S. lycopersicum* fresh market tomato cultivar T5 and wild tomato *S. habrochaites* accession LA1778 were used as parents to create a set of near isogenic lines (NILs), as described in Goodstal et al. (2005). A near-isogenic line, NIL1322, was selected as one parent to create a new set of recombinant sub-NILs because it contained the *S. habrochaites* region studied in Lounsbery et al. (2016) and extended further towards the centromere of chromosome 9. T5 was crossed with NIL1322 and another backcross to T5 was performed, resulting in a BC_5 population. An individual that was heterozygous only in the introgressed region, 07GH0618, was then used to obtain the set of recombinant sub-NILs for this study. A BC_5S_2 population was generated by selfing 07GH0618 and was genotypically screened for recombinants within the introgressed region for use in QTL mapping. All seed source plants were grown in pots in a greenhouse on the UC Davis campus in Davis, California using standard horticultural practices.

Creation of SNP Markers

Markers for this chromosome 9 region were selected from the public datasets at Sol Genomics Network (SGN, solgenomics.net) and internally developed markers. Tomato genome

markers T0532, C2_At2g36480, C2_At5g06140, SSR73, CT283, TG10, T1617, T0641, T1148, T1908, T1673, Solyc09g009750 (i.e., SL9750), H14, and H307 were selected as candidates for genotyping to identify recombinant sub-NILs for this region. These markers were originally of various types including: CAPS, SSR, PCR-converted RFLP, COS, and SCAR. To convert these markers to single nucleotide polymorphic (SNP) markers for use in high-throughput genotyping, the relevant regions in T5 and NIL1322 were Sanger sequenced at the UC Davis College of Biological Science DNA Sequencing Facility. Subsequently, the sequences from the two species were compared and SNPs were identified using Bioedit (http://www.mbio.ncsu.edu/bioedit/bioedit.html). Using the ITAG 2.5 annotation, single copy genes were identified by BLASTing predicted gene sequences against the tomato reference genome version available at that time (v2.50; SGN, solgenomics.net). Primers were designed for the single copy genes to amplify 300-1000 bp anchored in two seperate exons of each gene, and SNPs were found in the intervening introns. A subset of 14 gene sequences amplified consistently and had polymorphisms suitable for development of SNP markers.

A multiplexed SNP genotyping assay was designed using Sequenom's MassARRAY Assay Design 3.1 Software. The presets for Single Base Extension (SBE) High-Multiplexing were used. A total of 22 markers were used for genotyping, including some that contained a single SNP (T1673, H307, T0532, CT283, TG10, T1148, T1617, T0641), two SNPs (H14, SSR73, SL9750, At5g06140), or three SNPs (At2g36480 and T1908). These were narrowed down to 14 markers based on which SNPs had the least missing data.

To identify recombinants for this target region of chromosome 9, the BC_5S_2 seedlings (2,588 individuals) were genotyped using 14 SNP markers using Sequenom's MassARRAY i-Plex Gold at the UC Davis Veterinary Genetic Lab (Supplemental Table S1.1). Approximately 164

recombinants were identified among the 2,588 individuals. Subsequently, marker-assisted selection was used to obtain a subset of 21 unique, homozygous sub-NIL recombinant genotypes that were evaluated phenotypically in the field. In the field experiments, controls included the recurrent parent, *S. lycopersicum* T5, and two chromosome 9 NILs, NIL1322 and NIL175 were included. NIL1322 has the larger introgression of the two NILs and covers the span of the region represented by the 21 sub-NILs. The introgression in NIL175 covers part of the introgression represented by our 21 sub-NILs and continues toward the telomere beyond marker T1673. All 24 genotypes are illustrated in Figure 1.1.

Field Experimental Design and Procedures

Field experiments were conducted at the UC Davis Plant Sciences Facilities in Davis, California during the summers of 2016 and 2017. A split plot design was implemented with water treatments as the main plots. Each main plot was separated into four blocks perpendicular to the length of the drip irrigation tape to account for variation in water pressure. Each split plot was replicated within a location and each field was replicated and re-randomized, for a total of four split plot experiments at two locations per year. Each subplot consisted of eight plants of the same genotype. The rows were 2.03 m apart and within row spacing between plants was 0.609 m. Two border rows of a fresh market tomato variety, Crimson, surrounded each experiment and separated the main plots.

All plant materials were seeded out in flats in a greenhouse in Davis five weeks prior to field transplant. Seedlings were hardened off in a lath house for one week prior to field transplanting by hand. The transplants were established in the field with sprinkler irrigation for two weeks then switched to drip irrigation for the remainder of the season. Subsurface drip

irrigation was used to precisely apply each water treatment. A total of 202 kg/ha of UN-32 (urea and ammonium nitrate) fertilizer was applied via fertigation.

The amount of water applied on a given day was calculated for each water treatment (full and reduced) based on the crop evapotranspiration rate (Et_c) as determined from the reference evapotranspiration rate (Et_o) from the California Irrigation Management Information System (CIMIS) Davis station and the canopy width of the tomato plants. Water was applied to the fields three times per week and accounted for the Et_c since the previous watering day. The full water treatment received 100% of Et_c and the reduced water treatment received 30% of Et_c.

Phenotyping

All trait data was obtained on a per subplot basis using methods similar to Lounsbery et al. (2016) (Table 1.1). Plant maturity was determined by days after transplanting to first green fruit (DAPG) and days after transplanting to first ripe fruit (DAPR). DAPG was defined as the number of days after transplanting when greater than 50% of the plants in a subplot had a green fruit at least one cm in diameter. DAPR was defined as the number of days after transplanting when greater than 50% of the plants in a subplot had at least one ripe fruit.

Harvest data was collected when the plants in a subplot had maximized the number of ripe fruit per plant but had relatively few rotting fruit in the lower canopy. Plants were destructively harvested by cutting two plants per subplot at the soil level and removing all fruit. Red and green fruit were separated and weighed to obtain the ripe fruit yield (RYLD) and the total yield (TYLD). Fresh shoots were placed in mesh onion bags and dried in a forced air dryer until they reached a consistent weight, then were weighed to obtain shoot dry weight (SDW).

Carbon isotope discrimination was determined in 2016 and 2017 at the end of the growing season by collecting eight terminal leaflets per subplot one to two weeks before destructive harvest.

Leaves were dried in an oven for two days at 48°C. The dried leaves were then ground into a fine powder, and 0.0012-0.0018g of ground leaves were placed into tin capsules. The trays of tin capsules were sent to the UC Davis Stable Isotope Facility (UCD SIF) for carbon isotope composition analysis with a mass spectrometer. The carbon isotope composition ($\delta^{13}C$) was determined by the UCD SIF using the following formula: $\delta^{13}C_{sample}$ (‰) = ($R_{sample}/R_{reference}$ - 1) x 1000, where R_{sample} = ratio of $^{13}C/^{12}C$ measured in the plant sample and $R_{reference}$ = ratio of $^{13}C/^{12}C$ measured in the reference standard (Vienna Pee Dee Belemnite standard). The carbon isotope discrimination ($\Delta^{13}C$) of a sample was calculated using the formula: $\Delta^{13}C_{sample}$ (‰) = ($\delta^{13}C_a$ - $\delta^{13}C_p$)/(1 + $\delta^{13}C_a$ /1000), where $\delta^{13}C_a = \delta^{13}C$ for the atmosphere and $\delta^{13}C_p = \delta^{13}C$ for the plant sample. The $\delta^{13}C_a$ value was assumed to be -8.00 ‰ (O'Leary, 1988).

Statistical Analysis

All statistical analysis was completed in R (version 3.6.2) (R Core Team, 2019). Shapiro Wilks Tests and Levene's Tests were performed on data for each trait to check for normality and homogeneity of variances, respectively. Analysis of variance (ANOVAs) were performed on data for each trait. Heterogeneity of variance was addressed by either transforming data or separating data by the main source of the heterogeneity of variance (i.e., water or location). If significant ($P \leq 0.05$) interactions were present, data was separated based on those factors and analyzed separately. The sub-analyses were completed using each trait-environment combination as appropriate.

All models were linear mixed models created using the lmer function in R. The original model for each trait included location, water treatment (water), genotype, and year as fixed effects and several combinations of block, location, replication, and year as random effects:

Trait ~ Location * Water * Genotype * Year + (Block.Location) + (Replication.Location) + (Block.Replication.Location) + (Block.Year) + (Location.Year) + (Replication.Year) + (Block.Location.Year) + (Replication.Location.Year) + (Block.Replication.Location.Year)

All factors within parentheses are random effects.

Significant year interactions ($P \leq 0.05$) were detected for all traits, so the data for each trait was separated by year for further analysis. The reduced model used for each year's dataset was:

Trait ~ Location * Water * Genotype + (Block.Location) + (Replication.Location) + (Block.Replication.Location)

As described above, the trait data often had to be further separated by location and/or water treatment due to heterogeneity of variance or significant interactions. The following models were used in those cases:

Model for sub-analysis by location: **Trait ~ Water * Genotype + (Block)**

Model for sub-analysis by water treatment: **Trait ~ Location * Genotype + (Block.Location) + (Replication.Location) + (Block.Replication.Location)**

Model for sub-analysis by location and water treatment: **Trait ~ Genotype + (Block)**

Mean separations (lsmeans) using Tukey pairwise comparisons were performed for each sub-analysis. Pearson's correlations between $\Delta^{13}C$ and the yield traits RYLD and TYLD and $\Delta^{13}C$ and SDW were conducted in R using the cor.test function.

Quantitative Trait Locus Mapping

A linkage map was created for this region of *S. habrochaites* chromosome 9 using the BC_5S_2 population. A total of 2588 individuals and 14 SNP markers were used to create the linkage

map as described in Arms et al. (2015) using JoinMap 4.0 (Van Ooijen, 2006). The linkage map was created using Kosambi's method and employed a 10.0 LOD threshold for the linkage tree.

QTL mapping for each trait-environment combination (i.e., data separated by year, location, and/or water treatment as appropriate) was performed using R/qtl (Broman et al., 2003). The trait data for each of the 21 sub-NILs was derived from least square means (lsmeans) of each trait-environment combination. The population was treated as a RIL population in R/qtl because the sub-NILs have an expected 1:1 segregation ratio at each marker (homozygous for each allele). Interval mapping was performed on each trait-environment combination using scanone (EM algorithm method and Kosambi map function). Trait-specific threshold LOD scores were calculated using 1000 permutations. QTL intervals were determined using a Bayesian Likelihood Interval (95% Credible Interval) expanded to markers. Estimates of percent phenotypic variance explained by each QTL (%var) were calculated using the formula $1 - 10^{\left(\frac{-2*LOD}{n}\right)}$, where n=21 (21 sub-NILs used for mapping) (Broman & Sen, 2009).

RESULTS

Statistical Analysis

ANOVAs were conducted on each trait and revealed significant ($P \leq 0.05$) interactions with year: genotype by year, location by year, and/or water by year. All further analyses were separated by year. Due to seed mix-ups of controls during planting that were discovered only after the plants set fruit in the field in 2016, the correct controls were not present in the field. All 2016 data reflects only the 21 sub-NILs.

For both years, all traits met the assumption of normality. For the 2016 dataset, all of the traits had significant sources of heterogeneity of variance in the Levene's Test, so the data were

either transformed (SDW was transformed using natural log) or a sub-analysis was performed. The sub-analysis involved separating the data by the source of variation in the Levene's test that was causing the heterogeneity of variance. For the 2017 dataset, only DAPR and $\Delta^{13}C$ met the assumptions for homogeneity of variance. DAPG was transformed using log10 and all other traits were separated into sub-analyses as described above. In some cases, ANOVAs revealed further need to separate data due to significant interactions (see Table 1.2 for the list of traits and environment combinations). Each trait-environment combination was given a coded name that included the trait abbreviation, year, and environmental separations (if necessary) separated by underscores (i.e., trait_year_location_water treatment).

Genotype main effects were significant for most trait-environment combinations except: $\Delta^{13}C$ _2016_red, SDW_2017_loc2_red, TYLD_2017_loc1_red, and TYLD_2017_loc2_red. Where loc# represents the location for the given year and red represents the reduced water treatment (see Table 1.2 for ANOVA Table).

A total of 28 trait-environment combinations were used for QTL mapping. In 2016, 16 trait-environment combinations were used and 12 were used in 2017 (see Table 1.2 for the full list of trait-environment combinations and ANOVA results).

Significant Pearson's correlation coefficients ($P \leq 0.05$) were found for comparisons between $\Delta^{13}C$ and TYLD for the full water treatments in both years (2016, r = -0.352; 2017, r = 0.307) and for $\Delta^{13}C$ and RYLD under the full water treatment in 2017 (r = 0.424) (Table 1.3). Correlations between $\Delta^{13}C$ and yield traits were not significant for any reduced water treatments. All Pearson's correlations between $\Delta^{13}C$ and SDW were significant in all trait-environment combinations (r values ranged from -0.242 to -0.447).

Quantitative Trait Locus Mapping

All 14 markers used for genotyping the sub-NILs were mapped to separate locations within a 3.6 cM region of chromosome 9 (Figure 1.2). The average distance between markers is 0.276 cM. The region spanned from marker T1673 to T0641. This region overlaps with the mapped region from Lounsbery et al. (2016) from marker T1673 to T0532.

Out of 28 trait-environment combinations, 20 had significant marker-trait associations ($P \leq 0.05$). The majority of these traits mapped within a 2.2 cM region between markers H14 and TG10 (Figure 1.2). Most of these QTL partially or completely overlap. No significant marker-trait associations were detected for RYLD_2016_red, RYLD_2017_full, TYLD_2016_full, TYLD_2016_red, TYLD_2017_loc1_red, TYLD_2017_loc2_red, Δ^{13}C_2016_full, or Δ^{13}C_2016_red.

The presence of the *S. habrochaites* allele at QTL increased the DAPG and DAPR (i.e., longer time to maturity) and reduced yields in all trait-environment combinations, which are horticulturally negative effects. The *S. habrochaites* allele at QTL increased the SDW and decreased the Δ^{13}C for each of the trait-environment combinations. Both are horticulturally positive *S. habrochaites* allele effects. The QTL for the maturity trait-environment combinations and the SDW-environment combinations explained relatively high proportions of trait phenotypic variance observed in these experiments (60-73% and 53-71%, respectively). Yield and Δ^{13}C did not have significant QTL in some environments and years. QTL for yield traits explained less than 50% of the phenotypic variance and the Δ^{13}C QTL for 2017 accounted for 48% of the variance (Table 1.4).

DISCUSSION

Coincidence of QTL

QTL for maturity, yield, and water stress tolerance-related traits were mapped to a 3.6 cM region of chromosome 9 of *S. habrochaites*. The majority of the QTL detected for each trait-environment combination are coincident and overlap at least partially, based on their 95% Bayesian Likelihood Intervals. Most QTL intervals begin at marker H14 or H307, which is consistent with results of Lounsbery et al. (2016). In this study, which includes sub-NILs with introgressions extending further towards the centromere of chromosome 9 in *S. habrochaites*, most of the QTL intervals end at marker TG10, with a few that end at marker SL9750. Our results suggest that there may be structural and/or genetic reasons why QTL are not detected beyond this marker. Some possible reasons include: an increase in heterochromatin content closer to the centromere, lower gene density, and/or other structural DNA differences such as highly repetitive DNA. After a visual inspection of the *S. lycopersicum* reference genome and annotation (SL4.0, ITAG 4.0; SGN, solgenomics.net), there is not an obvious difference in gene density before or after marker TG10. However, these observations may not be accurate for the *S. habrochaites* chromosome 9 sequence. *S. habrochaites*-specific gene annotations would help determine if there is a difference in gene density in the region containing the overlapping QTL compared to regions outside of it. There may be a higher amount of heterochromatin in this region in the *S. habrochaites* genome due to the proximity to the centromere or the presence of a heterochromatin island (Tomato Genome Consortium, 2012). In cultivated tomato, pericentric heterochomatin contains a much lower number of genes than the euchromatin (Wang et al., 2006). Additional research would be required to prove or disprove these hypotheses.

The majority of QTL for traits in different environments have highly overlapping intervals for each trait-environment combination. This suggests that some of the loci controlling these traits may be pleiotropic and/or the QTL associated with these traits are tightly linked (see the Implications for Breeding subsection). Interestingly, some LOD peaks were identified at different marker loci, suggesting that it may be possible to separate the QTL with additional recombinants for higher resolution mapping. The different location of LOD peaks was observed for different classes of traits as well. For example, $\Delta^{13}C$ has a peak at TG10 while the majority of the QTL for yield and maturity have peaks between H307 and At2g36480, which are 1.2 and 0.9 cM away from TG10, respectively.

The direction of *S. habrochaites* allelic effects at QTL for the traits in this study were consistent with those found in Lounsbery et al. (2016). Maturity and yield traits were negatively affected by the *S. habrochaites* allele and WUE-related traits were positively affected by the *S. habrochaites* allele in all trait-environment combinations. These results suggest that some loci may have pleiotropic effects and/or repulsion phase linkages among QTL that are causing negative linkage drag. If linkage drag is due to repulsion phase linkage among QTL, recombination and subsequent segregation would be required to obtain individuals with coupling phase linked QTL to select for practical use in breeding cultivated tomato. It is not unusual for traits relating to abiotic stress tolerance in wild species of tomato to exhibit negative linkage drag effects with horticultural and yield traits (Brouwer & St. Clair, 2004; Haggard et al., 2013). Linkage drag could present a challenge for using wild species in breeding but it is not insurmountable if the QTL fractionate with additional fine mapping. In addition, the percent phenotypic variance explained by the QTL for yield-environment combinations is relatively low, suggesting that the effects of linkage drag on yield may be minimal.

Complex Relationships between $\Delta^{13}C$, Yield, and the Environment

The presence of significant genotype by environment (GxE) interactions required the separation of trait data into subsets based on the environmental factors causing the interaction (year, water treatment, and/or location). GxE interactions are commonly observed for complex quantitative traits, such as WUE and yield, especially under differing water treatments (Albert et al., 2016; Bernardo, 2008; Des Marais et al., 2013). The lack of significant QTL for yield and $\Delta^{13}C$ across all environmental combinations suggests that there is also QTL by environment (QTLxE) interactions within this region. No significant QTL were detected for $\Delta^{13}C$ in 2016 and only three yield QTL were detected out of the nine yield-environment combinations tested. QTLxE and/or experimental error may also explain the differences in QTL intervals for a given trait in different environments. There is evidence of GxE and QTLxE for maturity, yield, and WUE-related traits in this region of chromosome 9 of *S. habrochaites*, yet the direction of allelic effects are consistent between the trait QTL detected in each environment. It would present a greater challenge to breeding if the allelic effects differed between the environments.

Arms et al. (2016) found a low positive correlation between $\Delta^{13}C$ and yield traits (RYLD and TYLD) in tomato chromosome 9 sub-NILs derived from *S. habrochaites*, indicating a negative correlation with WUE. In the present study, correlations between yield and $\Delta^{13}C$ were environmentally dependent. All statistically significant correlations were observed in the full water treatments only, and the direction of the correlation was dependent on year. In 2016, there was a low negative correlation between $\Delta^{13}C$ and TYLD ($r = -0.352$) indicating as WUE increased, total yields increased. In 2017 there were low to moderate positive correlations between $\Delta^{13}C$ and RYLD or TYLD ($r = 0.424$ and $r = 0.307$ respectively) indicating that as WUE increased, yields decreased. Our results suggest a complex relationship between yield and WUE that is dependent

on environmental factors. Under the reduced water treatment, the WUE and yield traits are consistently independent from one another, implying that they can be selected for separately in breeding programs targeting deficit irrigation systems. Dixon et al. (2019) also found a complex relationship between grain $\Delta^{13}C$ and grain yield in wheat. They observed low positive and neutral correlations between grain $\Delta^{13}C$ and grain yield, which was dependent on the field environment. The lack of significant correlations, or low correlations, between yield traits and $\Delta^{13}C$ in our study suggests that although the trait QTL are co-localizing, the underlying genes associated with those traits may be expressing independently of one another.

Arms et al. (2017) used a set of paired *S. habrochaites* NILs (including NIL175, a control for the sub-NILs in our experiment which contains part of the introgressed chromosome 9 region) to determine differential gene expression under fast onset water stress due to root chilling. Twenty genes were differentially expressed in the NILs prior to any chilling stress, and few of these genes were located in the introgressed region. Arms et al. observation suggested that the *S. habrochaites* introgressions had epistatic effects on the *S. lycopersicum* genome and primed the plants for an abiotic stress response. Interactions between the *S. habrochaites* introgressions and the rest of the *S. lycopersicum* genome are likely occurring in our sub-NILs as well, affecting the phenotypes we observed. Different introgressions in each of the sub-NILs may prime the *S. lycopersicum* genome distinctively and alter the phenotype whether under water stress or not. This hypothesis would need to be tested with future transcriptome studies of the sub-NILs subjected to water stress.

Implications for Breeding

The presence of wild *S. habrochaites* alleles at QTL detected in the sub-NILs of our study were associated with later maturity, lower yields, and improved WUE (see the Coincidence of QTL subsection). This observation was also a general trend in the trait-environment combination

lsmean data for each individual sub-NIL. Sub-NILs that ranked the highest for water stress tolerance-related traits tended to be ranked lower for maturity and yield traits, and visa versa (Supplemental Table S1.2). Introgressions from wild tomato species have been associated with lower yields in other studies, which could be a potential barrier in using introgression lines in breeding programs (Brouwer & St. Clair, 2004; Tanksley et al., 1996). However, we observed some exceptions to this trend that may be useful for future breeding.

Two sub-NILs, 13 and 19, had relatively low values for $\Delta^{13}C$ (indicating more WUE) and high values for yield (Supplemental Table S1.2). Sub-NIL 13 has later maturity than most other sub-NILs, but it also exhibits above average SDW values, a positive water stress tolerance-related trait. Sub-NIL 19 has earlier maturity, but has below average SDW. Both of these lines could potentially be utilized in breeding programs as sources of WUE without major reductions in yield. A marker-assisted backcrossing program could be used to transfer the introgressions from sub-NILs 13 or 19 into different cultivated tomato genetic backgrounds for evaluation of potential benefits. Most modern commercial tomato cultivars are F_1 hybrids, so it would also be interesting to determine if the introgressions have a different horticultural effect in a hybrid as homozygotes (i.e., both inbred parents contain the introgression) or as heterozygotes (i.e., only one inbred parent contains the introgression).

Although two of the sub-NILs may be useful for breeding for improved WUE in tomato, the QTL discovered in this study are unlikely to be directly deployable in a breeding program. As described previously, the QTL were not consistently detected across environments and are highly coincident. The data from our present study cannot determine if the coincidence of QTL is due to tight repulsion phase linkage of genes controlling these traits or pleiotropy. Additional research would be required. Fractionation of tightly linked QTL has been observed in other tomato studies

employing higher resolution mapping (Haggard et al., 2013; Johnson et al., 2012; Li et al., 2018). However, pleiotropy has also been observed in lines containing introgressions from wild species of tomato (Tanksley et al., 1996). If some of the genes controlling maturity, yield, and/or WUE-related traits are pleiotropic, then the QTL found in this study will not fractionate with higher resolution mapping. Further studies would be needed to test for the presence of pleiotropy in this region.

To map QTL in the region between markers H14 and SL9750 at a higher resolution, additional recombinants identified from a larger segregating population would be required to potentially separate QTL, assuming the absence of pleiotropy. However, repression of recombination in interspecific progeny can cause challenges in obtaining a suitable number of recombinants for high resolution mapping (Canady et al., 2006; Wijnker & de Jong, 2008). Repression of recombination has been reported in introgressions of wild species of tomatoes, especially smaller introgressions and in heterozygous segments (Chetelat et al., 2000; Lim et al., 2008; Rick, 1969, 1983). When our sub-NILs were being developed, the number of recombinants detected was also limited by polymorphic marker density as well as the availability and throughput of the lower cost genotyping technology available at that time. More recent advances in marker development and high genotyping throughput will improve opportunities for finding additional recombinants in this region.

In conclusion, wild species, such as *S. habrochaites*, are important sources of novel traits to improve cultivated crops, but they can also present challenges for use in breeding. This is especially true for quantitatively inherited traits that are often affected by the environment. In this study, we mapped QTL for 20 trait-environment combinations to a 3.6 cM introgressed region of *S. habrochaites* chromosome 9 using a set of 21 sub-NILs. The presence of the *S. habrochaites*

allele at QTL had negative horticultural effects on maturity and yield traits, and positive horticultural effects on $\Delta^{13}C$. These QTL were highly coincident and exhibited QTLxE, which limits their potential usefulness in breeding programs unless the QTL can be fractionated and confirmed with further fine mapping. Two sub-NILs exhibited improved WUE and horticulturally acceptable yields. These lines may be useful for breeding tomatoes that use less water without reducing yields.

ACKNOWLEDGMENTS

This research was supported by the California Tomato Research Institute and the US Department of Agriculture National Institute of Food and Agriculture competitive grants. We thank the St. Clair lab members and Stephanie Zullo for contributions to field and lab data collection. We also thank the UC Davis Plant Sciences field facility crew for assistance with field experiments.

REFERENCES

Albert, E., Gricourt, J., Bertin, N., Bonnefoi, J., Pateyron, S., Tamby, J. P., Bitton, F., & Causse, M. (2016). Genotype by watering regime interaction in cultivated tomato: Lessons from linkage mapping and gene expression. *Theoretical and Applied Genetics*, *129*(2), 395–418. https://doi.org/10.1007/s00122-015-2635-5

Arms, E. M., Bloom, A. J., & St. Clair, D. A. (2015). High-resolution mapping of a major effect QTL from wild tomato *Solanum habrochaites* that influences water relations under root chilling. *Theoretical and Applied Genetics*, *128*(9), 1713–1724. https://doi.org/10.1007/s00122-015-2540-y

Arms, E. M., Lounsbery, J. K., Bloom, A. J., & St. Clair, D. A. (2016). Complex relationships among water use efficiency-related traits, yield, and maturity in tomato lines subjected to deficit irrigation in the field. *Crop Science*, *56*(4), 1698–1710. https://doi.org/10.2135/cropsci2015.10.0661

Arms, E. M., Yan, Z., & St. Clair, D. A. (2017). Differential transcriptional regulation in roots of tomato near-isogenic lines in response to rapid-onset water stress. *Frontiers in Plant Science*, *8*, 166.

Bayuelo-Jiménez, J. S., Debouck, D. G., & Lynch, J. P. (2002). Salinity tolerance in Phaseolus species during early vegetative growth. *Crop Science*, *42*(6), 2184–2192.

Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Science*, *48*(5), 1649–1664.

Broman, K. W., & Sen, S. (2009). *A Guide to QTL Mapping with R/qtl* (Vol. 46). Springer.

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

Brouwer, D., & St. Clair, D. (2004). Fine mapping of three quantitative trait loci for late blight resistance in tomato using near isogenic lines (NILs) and sub-NILs. *Theoretical and Applied Genetics*, *108*(4), 628–638.

Canady, M. A., Ji, Y., & Chetelat, R. T. (2006). Homeologous recombination in *Solanum lycopersicoides* introgression lines of cultivated tomato. *Genetics*, *174*(4), 1775–1788.

CDFA. (2020). *California Agricultural Statistics Review, 2019-2020*. CDFA, California Agricultural Production Statistics. https://www.cdfa.ca.gov/Statistics/#:~:text=In%202020%20California's%20farms%20and,of%202.8%20percent%20from%202019.

Chetelat, R. T., Meglic, V., & Cisneros, P. (2000). A genetic map of tomato based on BC1 *Lycopersicon esculentum* x *Solanum lycopersicoides* reveals overall synteny but suppressed recombination between these homeologous genomes. *Genetics*, *154*(2), 857–867.

Des Marais, D. L., Hernandez, K. M., & Juenger, T. E. (2013). Genotype-by-environment interaction and plasticity: Exploring genomic responses of plants to the abiotic environment. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 5–29.

Dixon, L. S., Godoy, J. V., & Carter, A. H. (2019). Evaluating the utility of carbon isotope discrimination for wheat breeding in the Pacific Northwest. *Plant Phenomics*.

FAOSTAT. (2020). *Crops and livestock products*. FAOSTAT Data. https://www.fao.org/faostat/en/#data/QCL/visualize

Farquhar, G. D., & Sharkey, T. D. (1982). Stomatal conductance and photosynthesis. *Annual Review of Plant Physiology*, *33*(1), 317–345.

Farquhar, G., & Richards, R. (1984). Isotopic composition of plant carbon correlates with water-use efficiency of wheat genotypes. *Functional Plant Biology*, *11*(6), 539–552.

Goodstal, J. F., Kohler, G. R., Randall, L. B., Bloom, A. J., & St. Clair, D. A. (2005). A major QTL introgressed from wild *Lycopersicon hirsutum* confers chilling tolerance to cultivated tomato (*Lycopersicon esculentum*). *Theoretical and Applied Genetics*, *111*(5), 898–905.

Haggard, J. E., Johnson, E. B., & St. Clair, D. A. (2013). Linkage relationships among multiple QTL for horticultural traits and late blight (*P. infestans*) resistance on chromosome 5 introgressed from wild tomato *Solanum habrochaites*. *G3: Genes, Genomes, Genetics*, *3*(12), 2131–2146. https://doi.org/10.1534/g3.113.007195

Johnson, E. B., Haggard, J. E., & St. Clair, D. A. (2012). Fractionation, stability, and isolate-specificity of QTL for resistance to *Phytophthora infestans* in cultivated tomato (*Solanum lycopersicum*). *G3: Genes, Genomes, Genetics*, *2*(10), 1145–1159.

Jones, H. G. (2007). Monitoring plant and soil water status: Established and novel methods revisited and their relevance to studies of drought tolerance. *Journal of Experimental Botany*, *58*(2), 119–130. https://doi.org/10.1093/jxb/erl118

Li, J., Chitwood, J., Menda, N., Mueller, L., & Hutton, S. F. (2018). Linkage between the I-3 gene for resistance to Fusarium wilt race 3 and increased sensitivity to bacterial spot in tomato. *Theoretical and Applied Genetics*, *131*(1), 145–155. https://doi.org/10.1007/s00122-017-2991-4

Lim, G. T. T., Wang, G. P., Hemming, M. N., McGrath, D. J., & Jones, D. A. (2008). High resolution genetic and physical mapping of the I-3 region of tomato chromosome 7 reveals almost continuous microsynteny with grape chromosome 12 but interspersed microsynteny with duplications on *Arabidopsis* chromosomes 1, 2 and 3. *Theoretical and Applied Genetics*, *118*(1), 57–75. https://doi.org/10.1007/s00122-008-0876-2

Lounsbery, J. K., Arms, E. M., Bloom, A. J., & St. Clair, D. A. (2016). Quantitative trait loci for water-stress tolerance traits localize on chromosome 9 of wild tomato. *Crop Science*, *56*(4), 1514–1525. https://doi.org/10.2135/cropsci2015.07.0432

Martin, B., Tauer, C. G., & Lin, R. K. (1999). Carbon isotope discrimination as a tool to improve water-use efficiency in tomato. *Crop Science*, *39*(6), 1775–1783.

Martin, B., & Thorstenson, Y. R. (1988). Stable carbon isotope composition (δ13C), water use efficiency, and biomass productivity of *Lycopersicon esculentum*, *Lycopersicon pennellii*, and the F1 hybrid. In *Plant Physiol*, 88, 213–217.

Nuruddin, M. M., Madramootoo, C. A., & Dodds, G. T. (2003). Effects of water stress at different growth stages on greenhouse tomato yield and quality. *HortScience*, *38*(7), 1389–1393.

O'Leary, M. H. (1988). Carbon isotopes in photosynthesis. *Bioscience*, *38*(5), 328–336.

Patanè, C., & Cosentino, S. L. (2010). Effects of soil water deficit on yield and quality of processing tomato under a Mediterranean climate. *Agricultural Water Management*, *97*(1), 131–138. https://doi.org/10.1016/j.agwat.2009.08.021

Pathak, T. B., Maskey, M. L., Dahlberg, J. A., Kearns, F., Bali, K. M., & Zaccaria, D. (2018). Climate change trends and impacts on California agriculture: A detailed review. *Agronomy*, *8*(3), 25.

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rick, C. M. (1969). Controlled introgression of chromosomes of *Solanum pennellii* into *Lycopersicon esculentum*: Segregation and recombination. *Genetics*, *62*(4), 753.

Rick, C. M. (1983). Genetic variability in tomato species. *Plant Molecular Biology Reporter*, *1*(2), 81–87.

Tanksley, S., Grandillo, S., Fulton, T., Zamir, D., Eshed, Y., Petiard, V., Lopez, J., & Beck-Bunn, T. (1996). Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. *Theoretical and Applied Genetics*, *92*(2), 213–224.

Tomato Genome Consortium, xx. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, *485*(7400), 635.

Truco, M., Randall, L., Bloom, A., & St. Clair, D. (2000). Detection of QTLs associated with shoot wilting and root ammonium uptake under chilling temperatures in an interspecific backcross population from *Lycopersicon esculentum× L. hirsutum*. *Theoretical and Applied Genetics*, *101*(7), 1082–1092.

Tuberosa, R. (2012). Phenotyping for drought tolerance of crops in the genomics era. *Frontiers in Physiology*, *3 SEP*. https://doi.org/10.3389/fphys.2012.00347

Van Ooijen, J. (2006). JoinMap 4. *Software for the Calculation of Genetic Linkage Maps in Experimental Populations. Kyazma BV, Wageningen, Netherlands*, *33*.

Wang, Y., Tang, X., Cheng, Z., Mueller, L., Giovannoni, J., & Tanksley, S. D. (2006). Euchromatin and pericentromeric heterochromatin: Comparative composition in the tomato genome. *Genetics*, *172*(4), 2529–2540.

Wijnker, E., & de Jong, H. (2008). Managing meiotic recombination in plant breeding. *Trends in Plant Science*, *13*(12), 640–646.

TABLES AND FIGURES

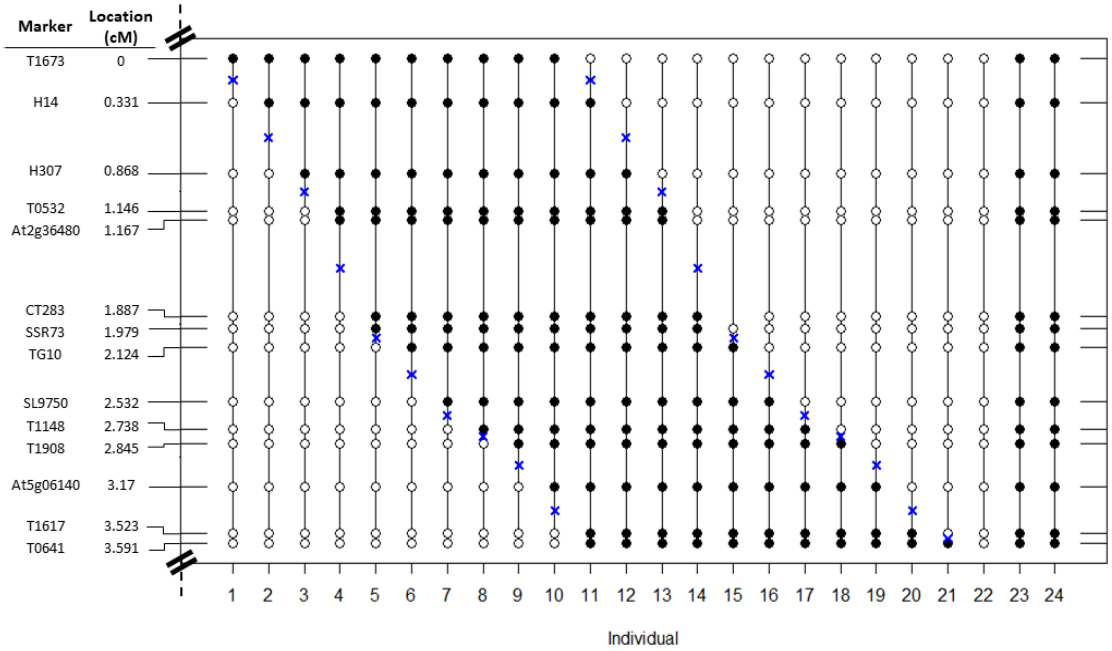


Figure 1.1. A graphical representation of the *S. habrochaites* chromosome 9 sub-NILs and control genotypes. White circles represent the *S. lycopersicum* cultivar T5 allele at a given marker locus. Black circles represent the *S. habrochaites* allele at a given marker locus. Individuals 1-22 have *S. lycopersicum* alleles at all loci outside of this region. A blue x indicates a recombination breakpoint in a given genotype. The linkage map is on the left with marker locations in cM. Each column represents graphical genotypes: sub-NILs are numbered 1-21; 22 is the recurrent parent T5; 23 is NIL1322 which contains a larger introgression of *S. habrochaites* (beyond the region shown here); and 24 is NIL175, which also contains a larger introgression of *S. habrochaites* that continues toward the telomere beyond marker T1673.

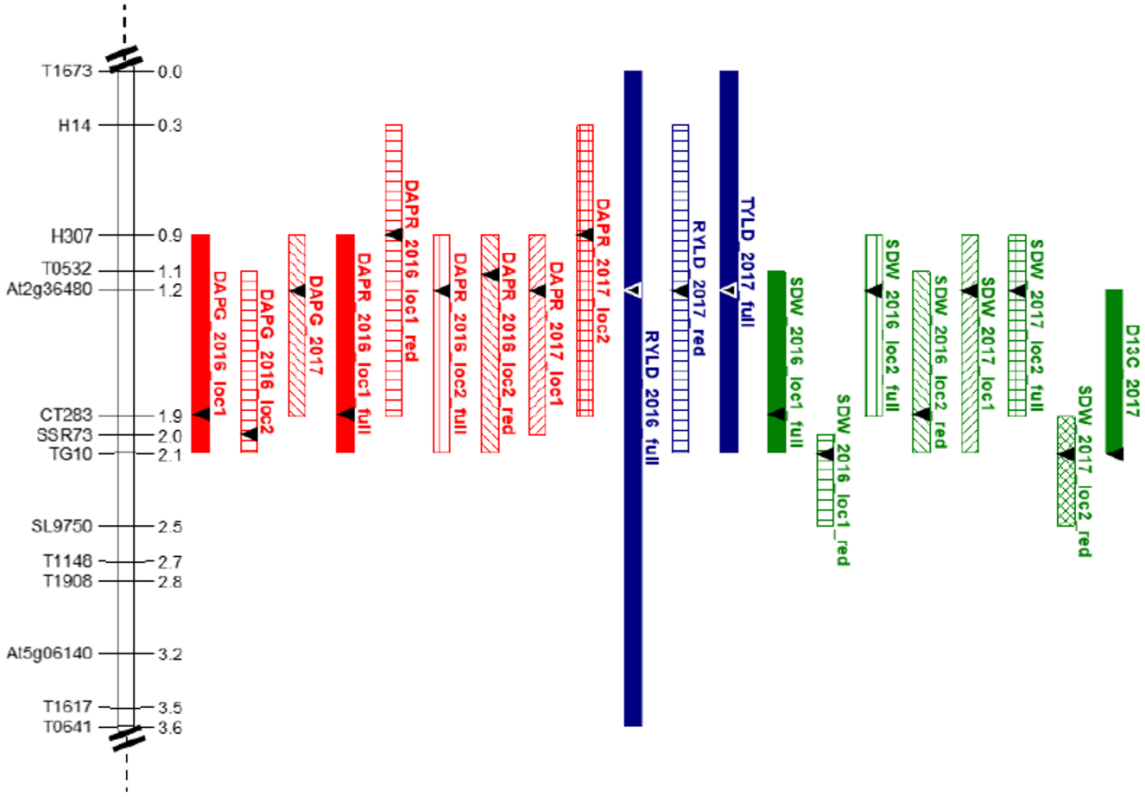


Figure 1.2. A linkage map and QTL map of an introgressed *S. habrochaites* region on the short arm of chromosome 9. The linkage map is shown on the left with marker positions in cM. Each trait-environment combination is written as the trait abbreviation, year, and environmental separations (if necessary) separated by underscores. The red trait-environment combinations represent the maturity traits (days after planting to first green fruit, DAPG; days after planting to first ripe fruit, DAPR), blue represents yield traits (ripe fruit yield, RYLD; total fruit yield, TYLD), and green represents WUE-related traits (shoot dry weight, SDW; carbon isotope discrimination, D13C). Bars represent the 95% credible intervals for each trait-environment combination. Triangles show the marker associated with the peak LOD score.

Table 1.1. Phenotypic traits, descriptions, and abbreviations for traits fine mapped to chromosome 9 of *S. habrochaites*.

Trait Category	**Abbreviation**	**Description**
Maturity	DAPG	Days After Planting to First Green Fruit
	DAPR	Days After Planting to First Ripe Fruit
Yield	RYLD	Ripe Fruit Yield (kg)
	TYLD	Total Fruit Yield (kg)
Water Use Efficiency-Related	SDW	Shoot Dry Weight (kg)
	$\Delta^{13}C$ or D13C†	Carbon Isotope Discrimination (‰)

† D13C is used in figures where use of symbols is limited

Table 1.2. Summary of ANOVA results for 21 *S. habrochaites* chromosome 9 sub-NILs and controls tested in the field in two years under full and reduced water treatments. The traits were separated into subsets of appropriate trait-environment combinations based on factors that caused heterogeneity of variance or significant interactions. Dashes indicate that that main effect or interaction was not applicable in the model for the trait-environment combination.

Trait	Year	Location	Water Treatment	F Value						
				Location	Water Treatment	Genotype	Location x Water	Location x Genotype	Water x Genotype	Location x Water x Genotype
DAPG[a]	2016	1	full + reduced	-	12.02***	22.63***	-	-	0.87 ns†	-
		2	full + reduced	-	4.22*	20.02***	-	-	1.04 ns	-
	2017	1+2	full + reduced	8.30 ns	9.56**	61.62***	0.49 ns	1.05 ns	1.01 ns	1.20 ns
DAPR[b]	2016	1	full	-	-	10.96***	-	-	-	-
		1	reduced	-	-	10.12***	-	-	-	-
		2	full	-	-	7.32***	-	-	-	-
		2	reduced	-	-	5.53***	-	-	-	-
	2017	1	full + reduced	-	0.45 ns	26.47***	-	-	1.07 ns	-
		2	full + reduced	-	13.57***	21.09***	-	-	1.03 ns	-
RYLD[c]	2016	1+2	full	6.56 ns	-	2.70***	-	0.63 ns	-	-
		1+2	reduced	0.55 ns	-	3.46***	-	1.35 ns	-	-
	2017	1+2	full	2.17 ns	-	4.51***	-	0.91 ns	-	-

		1+2	reduced	0.99 ns	-	1.82*	-	0.71 ns	-	-
TYLD[d]	2016	1+2	full	1.12 ns	-	3.48***	-	0.89 ns	-	-
		1+2	reduced	0.29 ns	-	2.89***	-	1.09 ns	-	-
	2017	1+2	full	11.88*	-	3.71***	-	0.83 ns	-	-
		1	reduced	-	-	1.60 ns	-	-	-	-
		2	reduced	-	-	1.00 ns	-	-	-	-
SDW[e]	2016	1	full	-	-	6.59***	-	-	-	-
		1	reduced	-	-	6.43***	-	-	-	-
		2	full	-	-	4.10***	-	-	-	-
		2	reduced	-	-	7.05***	-	-	-	-
	2017	1	full + reduced	-	9.43**	5.49***	-	-	1.40 ns	-
		2	full	-	-	2.60**	-	-	-	-
		2	reduced	-	-	1.53 ns	-	-	-	-
$\Delta^{13}C$[f]	2016	1+2	full	11.51 ns	-	2.87***	-	1.51 ns	-	-
		1+2	reduced	6.09 ns	-	1.42 ns	-	0.86 ns	-	-
	2017	1+2	full + reduced	1.45 ns	42.34***	18.79***	0.73 ns	1.24 ns	0.99 ns	1.16 ns

*** 0.001 significance level; ** 0.01 significance level; * 0.05 significance level; †ns: nonsignificant at the 0.5 significance level; - not applicable

[a] Days after planting to first green fruit, [b] days after planting to first ripe fruit, [c] ripe fruit yield, [d] total fruit yield, [e] shoot dry weight, [f] carbon isotope discrimination

Table 1.3. Pearson correlation results for various trait-environment combinations and carbon isotope discrimination ($\Delta^{13}C$) for *S. habrochaites* chromosome 9 sub-NILs and controls. Data for shoot dry weight (SDW), ripe fruit yield (RYLD), total fruit yield (TYLD), and $\Delta^{13}C$ was separated into subsets of appropriate trait-environment combinations based on factors that caused heterogeneity of variance or significant interactions.

Year	Traits	Location	Water Treatment	Pearson's Correlation Coefficient
2016	SDW x $\Delta^{13}C$	1	Full	-0.381**
	SDW x $\Delta^{13}C$	1	Reduced	-0.247*
	SDW x $\Delta^{13}C$	2	Full	-0.242*
	SDW x $\Delta^{13}C$	2	Reduced	-0.395**
	RYLD x $\Delta^{13}C$	1 + 2	Full	0.102 ns†
	RYLD x $\Delta^{13}C$	1 + 2	Reduced	-0.022 ns
	TYLD x $\Delta^{13}C$	1 + 2	Full	-0.352**
	TYLD x $\Delta^{13}C$	1 + 2	Reduced	0.023 ns
2017	SDW x $\Delta^{13}C$	1	Full + Reduced	-0.447**
	SDW x $\Delta^{13}C$	2	Full	-0.333*
	SDW x $\Delta^{13}C$	2	Reduced	-0.337*
	RYLD x $\Delta^{13}C$	1 + 2	Full	0.424**
	RYLD x $\Delta^{13}C$	1 + 2	Reduced	0.036 ns
	TYLD x $\Delta^{13}C$	1 + 2	Full	0.307**
	TYLD x $\Delta^{13}C$	1	Reduced	-0.130 ns
	TYLD x $\Delta^{13}C$	1	Reduced	-0.061 ns

** 0.01 significance level; * 0.05 significance level; †ns: nonsignificant at the 0.5 significance level

Table 1.4. QTL intervals and peak locations on the *S. habrochaites* chromosome 9 linkage map for each trait-environment combination. The *S. habrochaites* allele effect is denoted as positive (+) if it causes a horticulturally positive change to the trait phenotype and negative (-) if the *S. habrochaites* allele causes a horticulturally negative change to the trait phenotype. The threshold LOD is the minimum LOD value that is statistically significantly associated with a specific trait-environment combination ($P \leq 0.05$). The QTL peak is at the marker with the highest LOD score for a given trait-environment combination.

Trait	Year	Location	Water Treatment	*S. habrochaites* Allele Effect	QTL Peak	QTL Marker Interval (95% credible interval)	Peak LOD	Threshold LOD	Percent Phenotypic Variance Explained
DAPG[a]	2016	1	full + reduced	-	CT283	H307-TG10	5.93	1.91	0.73
		2	full + reduced	-	SSR73†	T0532-TG10	5.43†	1.77	0.70
	2017	1+2	full + reduced	-	At2g36480†	H307-CT283	5.68†	1.91	0.71
DAPR[b]	2016	1	full	-	CT283†	H307-TG10	4.52†	1.83	0.63
		1	reduced	-	H307†	H14-CT283	4.34†	1.88	0.61
		2	full	-	At2g36480†	H307-TG10	4.49†	1.74	0.63
		2	reduced	-	T0532	H307-TG10	4.21	1.90	0.60
	2017	1	full + reduced	-	At2g36480†	H307-SSR73	5.58†	1.91	0.71
		2	full + reduced	-	H307†	H14-CT283	5.31†	1.92	0.69
RYLD[c]	2016	1+2	full	-	At2g36480	T1673-T0641	2.51	1.85	0.42

	2017	1+2	reduced	-	At2g36480†	H14-TG10	2.73†	1.71	0.45
TYLD[d]	2017	1+2	full	-	At2g36480	T1673-TG10	1.70	1.62	0.31
SDW[e]	2016	1	full	+	CT283	T0532-TG10	4.96	1.77	0.66
		1	reduced	+	TG10	SSR73-SL9750	4.20	1.84	0.60
		2	full	+	At2g36480†	H307-CT283	5.63†	1.89	0.71
		2	reduced	+	CT283†	T0532-TG10	5.48†	1.85	0.70
	2017	1	full + reduced	+	At2g36480†	H307-TG10	3.61†	1.89	0.55
		2	full	+	At2g36480	H307-CT283	3.45	1.93	0.53
		2	reduced	+	TG10	CT283-SL9750	4.30	1.84	0.61
$\Delta^{13}C$[f]	2017	1+2	full + reduced	+	TG10†	At2g36480-TG10	2.99	2.01	0.48

[a] Days after planting to first green fruit, [b] days after planting to first ripe fruit, [c] ripe fruit yield, [d] total fruit yield, [e] shoot dry weight, [f] carbon isotope discrimination

†Peaks and peak LOD scores were forced to be at markers rather than locations and determined by the step value

SUPPLEMENTAL MATERIAL

Supplemental Table S1.1: Marker information for genotyping of chromosome 9 sub-NILs. This table includes the name of the marker, the forward primer sequences, reverse primer sequences, and extension primer sequences, and the Sequenom sequence including the SNP designated as [N/N], where each N represents the nucleotide present in the *S. habrochaites* and *S. lycopersicum* genotype, respectively.

Supplemental Table S1.2: Trait means for the 21 sub-NILs and three controls, separated into the appropriate trait-environment combinations (see Materials and Methods). Each trait-environment combination is written as the trait abbreviation, year, and environmental separations (if necessary) separated by underscores. The cell color of each trait mean indicates whether the value is more or less horticulturally favorable compared to the other values for that trait-environment combination. Values that are darker green are more horticulturally favorable (e.g., earlier maturity, higher yield, larger shoot dry weight, smaller $\Delta^{13}C$). Values that are darker red are less horticulturally favorable (e.g., later maturity, lower yield, smaller shoot dry weight, larger $\Delta^{13}C$).

CHAPTER 2

COMBINING ABILITY OF TOMATO INTROGRESSION LINES DERIVED FROM TWO WILD SPECIES FOR WATER USE EFFICIENCY AND YIELD

Amy M. Groh and Dina A. St. Clair

Groh, A. M., & St. Clair, D. A. (2022). Combining ability of tomato introgression lines derived from two wild species for water use efficiency and yield. *Crop Science*. (*submitted*).

ABSTRACT

Given climate change and population growth, it is increasingly important to efficiently grow crops such as tomato (*Solanum lycopersicum*) using less irrigation water. Tomato lines containing wild species chromosome introgressions (introgression lines, ILs) on chromosomes 2, 5, and 9 from *S. habrochaites* and *S. pennellii* were reported to be associated with greater water stress tolerance compared to cultivated tomato. ILs have been studied independently, but not in combination in a hybrid offspring. We aimed to determine if hybrid combinations of ILs with introgressions from *S. habrochaites* and *S. pennellii* improved water use efficiency (WUE) of genotypes compared to their parent ILs and/or cultivated tomato controls. A total of 35 genotypes including cultivated tomato controls, homozygous ILs, heterozygous ILs, and hybrids between the two species ILs were evaluated in the field for maturity, horticultural, yield, and WUE-related (carbon isotope discrimination, $\Delta^{13}C$) traits across two years under full and reduced water

treatments. Significant differences between genotypes for each trait-environment combination were detected. General (GCA) and specific (SCA) combining abilities were calculated for the hybrid offspring in a Design II mating design. Hybrids between two homozygous ILs generally did not have significantly better WUE than their best parent and $\Delta^{13}C$ had SCA effects of zero in most trait-environment combinations. However, two hybrids and one heterozygous IL had better WUE than both cultivated tomato controls for the genetic background without causing significant reductions in yield. Some introgressions that showed positive effects on WUE also exhibited negative effects on other traits such as fruit weight or maturity. Specific introgressions from both wild species may prove useful for developing hybrid tomato varieties that have acceptable yields yet use less irrigation water in the field.

INTRODUCTION

Fresh water is an increasingly limited resource for agricultural crop production. Precipitation, which can directly supply water to crops or replenish stored water that is used for irrigation, has become more variable annually due to climate change (Berg & Hall, 2017; Pathak et al., 2018). The variability in available water predicates the need for crops that can produce well under reduced water environments.

Cultivated tomato (*Solanum lycopersicum*) is an important vegetable crop globally and in the United States (Liedl et al., 2013). In 2019, tomatoes accounted for about 14% of the global vegetable food supply (measured in kg per capita per year) and had a gross production value of $102 billion USD globally (FAOSTAT, 2019). In the US, tomato is a major vegetable crop and is the second most consumed vegetable after potatoes (USDA ERS, 2019). In 2019, the US produced 12 million tonnes of tomatoes with a value of about $13 billion USD (FAOSTAT, 2019). Tomato yields can be negatively affected by limited precipitation and reduced irrigation water availability

(Nuruddin et al., 2003). When fresh water resources for crop irrigation are restricted in a given year, this leads to acreage reduction of irrigated tomatoes by growers. Acreage reduction presents a challenge for maintaining production volume and tonnage of both fresh market and processing tomatoes to meet the needs of consumers and food processors.

Crop cultivars can be bred to use less water by selection for water stress tolerance-related traits such as water use efficiency (WUE). WUE can be defined as a ratio of the rate of photosynthesis to the rate of transpiration in a plant. This trait is challenging and time-intensive to measure directly in a large field experiment, so indirect measures of WUE are often used (Jones, 2007; O'Leary, 1988; Tuberosa, 2012). One such measurement is carbon isotope discrimination ($\Delta^{13}C$), which relates the ratio of ^{13}C:^{12}C in plant tissues relative to that of the atmosphere (Farquhar & Richards, 1984; O'Leary, 1988). $\Delta^{13}C$ has been associated with season-long water use efficiency in many crops, including tomato (Farquhar & Richards, 1984; Martin & Thorstenson, 1988; Xu et al., 2008). Other indirect measures of WUE include yield and biomass (Tuberosa, 2012).

Wild species of tomato are more genetically diverse than cultivated tomato due to genetic bottlenecks that occurred during domestication (Blanca et al., 2015; Miller & Tanksley, 1990; Rick, 1983). Wild tomato species originate from a wide geographic range in South America for which they evolved adaptations to different environmental conditions, including low precipitation and cyclic droughts (Rick, 1973, 1983). All wild species of tomato, including *S. habrochaites* and *S. pennellii*, can be crossed successfully with cultivated tomato to varying degrees and produce viable progeny (Rick, 1979). Therefore, wild species can be utilized as sources of genetic variation for breeding specific traits in cultivated tomato (Tanksley & McCouch, 1997).

Tomato introgression lines (ILs) are inbred cultivated tomato genotypes with a chromosomal segment that has been replaced with a syntenic wild tomato genome segment (i.e., an introgression) via marker assisted backcrossing. ILs can be helpful for breeding because they permit association of traits to specific wild introgressions when compared to a cultivated tomato control; they also facilitate integration of wild chromosome regions controlling valuable traits into cultivated tomato (Gur et al., 2004; Lippman et al., 2007). A set of ILs for a given wild species accession, with each IL containing a unique wild chromosome introgression, is referred to as an IL library. The University of California Davis Tomato Genetics Resource Center (TGRC, tgrc.ucdavis.edu) maintains and distributes seed for tomato IL libraries containing introgressions from *S. habrochaites* accession LA1777 and *S. pennellii* accession LA716, each of which collectively represent the majority of the genome of each species (Eshed & Zamir, 1995; Monforte & Tanksley, 2000). The wild introgressions are in the genetic background of two inbred processing tomato varieties, E6203 and M82, for the *S. habrochaites* and *S. pennellii* ILs, respectively.

Previous research in our lab demonstrated that regions of chromosome 5 and 9 of *S. habrochaites* are associated with water stress tolerance-related traits. Truco et al. (2000) mapped decreased shoot wilting under rapid-onset water stress induced by root chilling in hydroponic tanks to chromosome 5 and 9 of *S. habrochaites* accession LA1778. Subsequently, a major effect quantitative trait locus (QTL) on chromosome 9 for shoot turgor maintenance (*stm9*) under root chilling was fined mapped with sub-near-isogenic lines (sub-NILs) by Arms et al. (2015). These same chromosome 9 sub-NILs were also evaluated in the field for tolerance to slow-onset water stress caused by deficit irrigation (Lounsbery et al., 2016). QTL associated with maturity, yield, and water stress tolerance-related traits (i.e., shoot dry weight, specific leaf area, and $\Delta^{13}C$) were detected and mapped to chromosome 9 (Groh et al., 2022; Lounsbery et al., 2016).

In studies by other researchers, four *S. pennellii* ILs exhibited superior water stress tolerance compared to their cultivated tomato parent: IL2-5, IL5-4, IL9-1, IL9-2-5 (Barrios-Masias et al., 2013; Gong et al., 2010; Rigano et al., 2016; Xu et al., 2008). In these studies, the water-stressed ILs showed higher survival rates, higher leaf dry matter content, and carbon isotope composition ($\delta^{13}C$, a value related to $\Delta^{13}C$) values that were more consistent with better WUE than the recurrent parent control, inbred cultivar M82. Most of these studies were performed in pots in greenhouses or covered outdoor experiments. These research findings suggest that these *S. pennellii* ILs would be interesting for further study of their potential for water stress tolerance in the field.

Interestingly, introgressions in some of the ILs from each species associated with water stress tolerance are syntenic, suggesting that these specific chromosomal regions contain genes and/or regulatory elements that affect how each species reacts to water stress. TA1544, a *S. habrochaites* chromosome 5 IL, includes the region with QTL for reduced shoot wilting under rapid-onset water stress discovered by Truco et al. (2000). This syntenic region is also present in *S. pennellii* IL5-4, which was reported to have better WUE than the cultivated tomato M82 (Xu et al., 2008). *S. habrochaites* ILs TA1324 and TA1325 have introgressions that include the region of *S. habrochaites* chromosome 9 studied in Lounsbery et al. (2016) and are also syntenic with IL9-1 in *S. pennellii*. Thus, *S. habrochaites* ILs TA1544, TA1324, and TA1325 are also interesting for further evaluation of their potential for WUE in the field.

ILs for chromosome 5 and 9 in both *S. habrochaites* and *S. pennellii* have been studied in association with water stress tolerance (as described above). However, chromosome 2 was reported to be associated with these traits only in *S. pennellii* ILs that were grown in pots (Gong et al., 2010). *S. habrochaites* IL TA1266 is partially syntenic to IL2-5 and was included in the

present study to evaluate if *S. habrochaites* alleles on chromosome 2 are potentially associated with WUE in the field. IL2-6 and IL9-1-2 partially overlap with some of the *S. pennellii* introgressions of interest on chromosome 2 and 9, respectively, and were also included in the present study.

ILs from both of these wild species have been studied independently for water stress tolerance and WUE characteristics. However, we are not aware of published studies of IL crosses that combine introgressions from different species into a single genotype (hybrid progeny) and subsequent evaluation for desirable trait phenotypes. Modern tomato varieties in commercial production are F_1 hybrids from the crossing of two inbred line parents. Hybrid varieties have several benefits to seed companies in comparison to inbred line varieties including reduced turn-around time to cultivar release and protection of intellectual property in the parent lines (Wehner, 1999). Hybrid varieties of crops may also display heterosis for one or more traits. Examples of heterosis have been observed in tomato when parental crosses include introgressions from wild species of tomato or mutant lines for specific genes (Eshed & Zamir, 1995; Krieger et al., 2010). Thus, crossing ILs from two wild species may confer heterosis for desirable traits controlled by genes in the introgressed regions or through epistatic interactions involving the introgressed regions.

One way to approach the evaluation of hybrid progeny and their parental genotype lines is mating designs. Mating designs are used to create structured populations with known genetic relationships among parents and progeny that share genetic covariances. Mating designs enable the estimation of per-trait general combining ability (GCA) and specific combining ability (SCA). GCA and SCA are used to determine the value of specific inbred lines as parents of hybrid progeny for superior horticultural or agronomic trait performance. Studies have shown that combining

ability can be an effective tool for making decisions about parent lines for breeding hybrids, including those involving wild introgressions (Thalapati et al., 2015; Zhang et al., 2016). For example, Zhang et al. (2016) found that GCA effects were greater than SCA effects for most traits in a diallel mating design involving ILs from wild cotton. They selected IL parents and specific crosses of ILs that were promising in breeding for cotton hybrids with improved fiber quality and yield. Similarly, combining ability could be a useful tool for determining if tomato ILs can serve as good parents in breeding hybrid tomato varieties.

Our study had three specific objectives. First, to determine if introgressions in ILs derived from two wild tomato species enhance WUE-related traits compared to their IL parents when combined together in hybrid progeny. Second, to determine if the ILs performed better for WUE in the field than the control cultivars. Third, to identify ILs (and their introgressions) that show promise for use in breeding hybrid tomatoes with improved WUE and without negative effects on horticultural traits.

MATERIALS AND METHODS

Plant materials

Seeds of ILs derived from wild *S. pennellii* accession LA716 and *S. habrochaites* accession LA1777 were obtained from the TGRC (https://tgrc.ucdavis.edu). Seed for the IL recurrent parent inbred line cultivars E6203 and M82 was sourced from maintained stocks in the St. Clair breeding program; these also served as controls. Four inbred *S. habrochaites* ILs and six *S. pennellii* ILs were selected for this study and were homozygous ILs (designated as HomILs, genotypes G1 ─ G11 in Table 2.1).

A factorial mating design (Design II) was followed to perform controlled crosses among potted plants in a greenhouse to obtain seed of hybrid progeny. A Design II mating design involves crossing a discrete set of designated male lines with a discrete set of female lines (Hallauer et al., 2010). Four *S. habrochaites* ILs (TA1266, TA1544, TA1324, and TA1325) served as the female parents and five *S. pennellii* ILs (IL2-5, IL2-6, IL5-4, IL9-1, and IL9-1-2) served as the male parents. Not all possible crosses in the Design II were performed due to the limited resources available for full evaluation in the field (see Table 2.1, genotypes G14 ─ G26 are part of the mating design). Crosses between the two species ILs are double introgression hybrids (denoted as DIHs). Each IL was also backcrossed to the appropriate parental cultivar for the genetic background of that IL (genotypes G27 ─ G36 in Table 2.1). The backcross progeny are heterozygous ILs (designated as HetILs). See Figure 2.1 for a schematic representation of the genotypes used in this study.

Source plants of each IL and cultivar were grown in 3 L pots in a greenhouse on the UC Davis campus in Davis, California under standard cultivation practices for tomato. Self-pollinated seed increases of ILs were obtained from these plants. Controlled crosses among the parental genotypes (ILs and cultivars) were performed in the late summer and fall of 2017 and 2018 in the greenhouse, and hybrid progeny seed was obtained.

Field experimental design and procedures

Field experiments were conducted at the UC Davis Plant Sciences Field Facilities in Davis, California during the summers of 2018 and 2019. Very little to no rainfall occurs in Davis during June through September. Each year, all 35 genotypes (Table 2.1, including cultivar controls, HomILs, DIHs, and HetILs) were grown in a replicated split plot design at two field locations per year. Locations were defined as fields with different cropping histories. Within each split plot

experiment, water treatment (full or reduced) was the main plot. Each main plot consisted of four blocks perpendicular to the length of the drip irrigation tape to account for variations in water pressure. Each genotype was transplanted in a subplot consisting of eight plants. Plants were on single row beds that were 2.03 m apart and plants were spaced 0.61 m apart within each row. There was a 0.91 m gap between the last plant of one plot and the first plant of the next to facilitate access for data collection. Two border rows of variety E6203 surrounded the perimeter of the experiment to minimize edge effects, and double border rows separated the main plots to isolate water treatment effects.

All plant materials were seeded into flats in the greenhouse about six weeks prior to transplant and grown using standard tomato cultivation practices. Flats were transferred to a lath house one week prior to transplant for hardening off. The plants were transplanted by hand into subplots and rooting established with sprinkler irrigation prior to switching to subsurface drip irrigation treatments. In 2018, irrigation water treatments were started about two weeks after the switch to drip. In 2019, there was a heavy May rain, which caused flooding in the fields and delayed the initiation of drip irrigation by three weeks. The soil moisture was monitored each year using eight soil moisture probes per field (two probes per main plot). In 2019, drip irrigation was initiated when the observed soil water tension readings doubled in magnitude within one week.

The amount of water applied per irrigation was calculated for each water treatment based on the crop evapotranspiration rate (Et_c). This was determined from the reference evapotranspiration rate (Et_o) from the California Irrigation Management Information System (CIMIS, cimis.water.ca.gov) Davis Station and the average canopy width of the tomato plants. Water was applied to the fields three times a week for the full growing season and accounted for the Et_c since the last watering day. The full water treatment received 100% of Et_c and the reduced

water treatment received 30% of Et_c. A total of 202 kg/ha of UN-32 (urea and ammonium nitrate) fertilizer was applied to each field via fertigation.

Trait phenotyping

All traits were assessed on a per-subplot basis similarly to Groh et al. (2022). In 2019 the heavy May rain and subsequent field flooding also caused stunting of plants at the north west end of location 1. Consequently, location 1, split plot 1, block 4 was dropped from the experiment to exclude the stunted plants from analyses of all traits. Location 1 in 2019 was also not included in the days to first green fruit measurement because flooding prevented access to the field plots.

Plant maturity was measured as days after planting to first green fruit (DAPG) and days after planting to first ripe fruit (DAPR). DAPG was defined as when greater than 50% of the plants in a subplot had a green fruit of at least one cm in diameter. DAPR was when greater than 50% of the plants in a subplot had at least one ripe fruit. After all plants in a subplot had ripe fruit, two measurements of the subplot width or height were averaged together to acquire the average canopy width (ACW) and average canopy height (ACH), respectively. The weight of 15 fruit (FW15) was measured from 15 random fruit collected per subplot. Plants were destructively harvested at the end of the growing season by cutting two plants per subplot at the soil level and removing the fruit. The red fruit and green fruit were weighed separately to obtain the ripe fruit yield (RYLD) and added together to get the total yield (TYLD). The two fresh shoots per subplot (minus the fruits) were placed in mesh nylon onion bags, weighed to obtain shoot fresh weight (SFW), and then dried in a large field forced-air dryer for several weeks until they reached a consistent weight. Dry shoot biomass was then obtained by weighing the dried shoots (SDW). Harvest index (HI) was calculated using the formula: HI = RYLD/(SFW + TYLD).

$\Delta^{13}C$ was determined by collecting eight terminal leaflets per subplot one to two weeks before destructive harvest at the end of the season. Leaves were dried in an oven for one to two days at 48°C, then ground into a fine powder with mortar and pestle. A sample of 0.0012 ─ 0.0018g of ground leaves per subplot was placed into tin capsules and sent to the UC Davis Stable Isotope Facility (SIF, https://stableisotopefacility.ucdavis.edu) for carbon isotope composition analysis. SIF calculated $\delta^{13}C$ using the formula: $\delta^{13}C = (R_{sample}/R_{reference} - 1) \times 1000$, where R_{sample} is the ratio of $^{13}C/^{12}C$ measured in the plant sample and $R_{reference}$ is the ratio of $^{13}C/^{12}C$ measured in the reference (Vienna Pee Dee Belemnite standard). $\Delta^{13}C$ was calculated using the following formula: $\Delta^{13}C = (\delta^{13}C_a - \delta^{13}C_p)/(1 + \delta^{13}C_a/1000)$, where $\delta^{13}C_a$ is the $\delta^{13}C$ for the atmosphere (assumed to be -8.00 ‰) and $\delta^{13}C_p$ is the $\delta^{13}C$ for the plant sample (Farquhar & Richards, 1984; O'Leary, 1988).

Statistical analysis

All data analyses were completed in RStudio (v3.6.2) (R Core Team, 2019). To check that the trait data met the assumptions of ANOVAs, Shapiro Wilks Tests and Levene's Tests were performed on each trait dataset. Trait data that exhibited heterogeneity of variance was transformed using a square root (ACW, SFW, SDW, RYLD, and TYLD) or logit (HI) transformation and/or separated into sub-analyses by the factor causing the heterogeneity of variance.

ANOVAs were performed on each trait dataset to determine the presence of significant ($P \leq 0.05$) main effects and interactions. If a significant interaction was present and it's sums of squares were close in magnitude to the factor with the highest sums of squares (genotype often had the highest sums of squares), then the data was separated by the environmental factor causing the interaction and analyzed separately. If the sums of squares of an interaction was small in magnitude compared to the magnitude of the main effects, or the interaction had a *P* value less

than, but close to 0.05, then the data was not separated. The full model used for the analysis of each trait dataset was:

Trait ~ location * water treatment * genotype * year + (block.location) + (replication.location) + (block.replication.location) + (block.year) + (location.year) + (replication.year) + (block.location.year) + (replication.location.year) + (block.replication.location.year)

All of the interactions in parentheses were treated as random effects. The * between the fixed effects indicate that their main effects and interactions were included in the model. If a trait dataset needed to be separated into sub-analyses, then the factors that caused the heterogeneity of variance or that were involved in a significant interaction were removed from the model. Mean separations using Tukey pairwise comparisons were performed using the emmeans (estimated marginal means) function and back-transformed for presentation in tables. Pearson's and Spearman Rank Correlations were obtained using the cor and cor.test functions in R.

Genotype emmeans were ranked based on horticulturally favorable values for each trait. Rank 1 is the genotype with the most favorable emmean value for a given trait-environment combination and rank 35 is a genotype with the least favorable emmean value for a given trait-environment combination. For the maturity traits DAPG and DAPR, earlier maturity is more desirable. A suitable canopy size for processing tomato is a medium height and width to facilitate tractor-based field operations and mechanical fruit harvest. Some ILs are semi-determinant and tend to have a larger plant size than is horticulturally acceptable, so smaller ACW and ACH was considered more favorable horticulturally. For the traits FW15, SFW, SDW, RYLD, TYLD, and HI, larger values are more desirable. A smaller $\Delta^{13}C$ value is favorable (indicates better WUE).

Heterosis

The potence ratio (H_{PR}), an estimation for heterosis, was obtained per trait for all parent-offspring combinations, using the formula: H_{PR} = (F1 - ((BP + WP)/2))/((BP - WP)/2), where F1 is the mean value of the hybrid offspring of a cross (HetIL or DIH), BP is the mean value of the best parent, and WP is the mean value of the worst parent (de Vienne & Fiévet, 2020). The best parent and worst parent were determined based on which value was more horticulturally favorable for a given trait, e.g., earlier maturity, larger fruit, smaller canopies, larger SFW and SDW, higher yields, and lower $\Delta^{13}C$. This calculation of heterosis was used because it doesn't depend on the scale or range of values of the parents and values can be compared between traits, crosses, and environments (de Vienne & Fiévet, 2020). The potence ratio was interpreted using the following scale modified from de Vienne & Fievet (2020): -0.2 to 0.2, no heterosis (additive gene action); -0.2 to -1.2 or 0.2 to 1.2, negative or positive mid-parent heterosis, respectively; <-1.2 or >1.2, worst-parent or best-parent heterosis, respectively.

Combining ability

Combining ability analysis per trait was conducted using data for the DIH individuals in a Design II mating design ANOVA. Shapiro Wilks Tests and Levene's Tests were performed on every trait dataset for this subset of genotypes to check for normality and homogeneity of variances, respectively. If the assumptions for ANOVA were not met, then the data was transformed using a square root (FW15, ACW, SFW, SDW, RYLD, and TYLD) or logit (HI) transformation. The REML method was implemented using the sommer package and mmer function using the following model (Covarrubias-Pazaran, 2016, 2022).

Trait ~ year + location + year.location + block.year.location + (female.male) + (female) + (male)

Factors in parentheses were considered random. The function randef was used to extract the effects of female:male (SCA), female (GCA), and male (GCA). If significant ($P \leq 0.05$) interactions were detected, then the data was separated into trait-environment combinations based on the factor causing the interaction and analyzed separately.

Heritability

Heritability was calculated based on the combining ability model and the trait-environment subsets, as appropriate. The additive variance (Va) from the male or female was calculated as: $Va_{parent} = (4/(1 + F)) * V_{parent}$, where $(4/(1 + F)) = 2$ when $F = 1$ for a completely inbred parent (Covarrubias-Pazaran, 2022; Hallauer et al., 2010). The dominance variance (Vd) was calculated as: $Vd = (4/(1 + F)^2) * V_{female:male}$, where $(4/(1 + F)^2) = 1$, when $F = 1$ for completely inbred parents. The total genetic variance (Vg) was calculated as: $Vg = Va_{female} + Va_{male} + Vd$ (Covarrubias-Pazaran, 2022). This calculation assumes that there is either no epistatic variance or that Vd includes dominance and epistatic variance (the non-additive components of genetic variance). Broad sense heritability (BSH) and narrow sense heritability (NSH) were calculated as: $BSH = Vg/(Vg + Ve)$ and $NSH = (Va_{male} + Va_{female})/(Vg + Ve)$ (Fehr, 1987). The ratio of NSH/BSH was calculated to determine the proportion of Vg due to Va. A value close to one indicated that most of Vg is due to Va. A value less than one indicates that there is non-additive variance contributing to Vg.

RESULTS

All trait datasets were normally distributed. Most datasets needed to be separated into trait environment subsets due to heterogeneity of variance or interactions, as described above. ANOVAs were performed on 26 total trait-environment combinations: DAPG_2018_loc1, DAPG_2018_loc2, DAPG_2019_loc2_full, DAPG_2019_loc2_red, DAPR_2018, DAPR_2019, FW15_full, FW15_red, ACW_full, ACW_red, ACH_2018_full, ACH_2018_red, ACH_2019_full, ACH_2019_red, SWW_full, SWW_red, SDW_full, SDW_red, RYLD_full, RYLD_red, TYLD_full, TYLD_red, HI_2018, HI_2019, Δ^{13}C_full, Δ^{13}C_red. The trait acronym is followed by the environmental factors used to subset the trait datasets (separated by underscores). These environmental factors include year (2018 or 2019), full or reduced water treatments (full or red, respectively), and/or location (loc1 or loc2). For example, Δ^{13}C_red represents the carbon isotope discrimination dataset for both years and locations for the reduced water treatment only.

All trait-environment combinations had significant main effects of genotype (Table 2.2). For most trait-environment combinations, M82 and E6203 were not significantly different from one another. There were two notable exceptions: FW15-environment combinations and Δ^{13}C-environment combinations. Mean separations for each trait are shown in Supplemental Table S2.1. The phenotypic rankings for each genotype and the emmeans of each genotype grouped by parent-offspring combinations are in Supplemental Table S2.2 and Supplemental Table S2.3, respectively.

Correlations

The general trends for Pearson's and Spearman Rank correlations were the same for both, but Spearman Rank had fewer significant correlations ($P \leq 0.05$) (Figure 2.2). RYLD and TYLD under the reduced water treatment were moderately to highly positively correlated with HI in both years (Pearson's r value: 0.68 – 0.71), but under the full water treatment those traits had either no or low correlations with HI (Pearson's r value: 0 or 0.39). RYLD and TYLD were moderately negatively correlated with maturity traits in all environmental subsets (plants that matured earlier tended to have higher yields) (Pearson's r value: -0.34 – -0.65).

$\Delta^{13}C$ did not show many significant correlations with other trait-environment combinations. $\Delta^{13}C$ under the reduced water treatment had a moderate to high negative correlation with FW15, ACW, SFW, and SDW (Pearson's r value: -0.39 – -0.71). $\Delta^{13}C$ under the reduced water treatment also had a moderate positive Pearson's correlation with HI in both years (Pearson's r value: 0.44 – 0.50), but this was not significant in the Spearman Rank correlation. Maturity traits also had moderate to high positive correlations with ACW, SFW, and SDW (Pearson's r value: 0.48 – 0.80), indicating that later maturing plants tended to have larger canopy sizes and shoot weights.

Heterosis

The maturity trait-environment combinations exhibited positive mid-parent or best-parent heterosis for most HetIL or DIH genotypes as determined by H_{PR} (Table 2.3, Supplemental Table S2.4). FW15 generally exhibited H_{PR} values consistent with negative or positive mid-parent heterosis ($-1.20 < H_{PR} < -0.20$ or $0.2 < H_{PR} < 1.20$). A few genotypes had FW15 H_{PR} values close to zero (suggesting no heterosis) and genotypes under the reduced water treatment tended to have

positive H_{PR} values. The canopy size traits (ACW and ACH) had highly variable H_{PR} values, but worst-parent heterosis was more common than the other types of heterosis. The DIHs with TA1325 or TA1544 as a parent generally had positive mid-parent or best-parent heterosis ($H_{PR} > 0.20$) for SFW and SDW. Most crosses had best-parent heterosis ($H_{PR} > 1.20$) for yield traits, with the exception of G30 (E6203xTA1324), which had H_{PR} values less than -1.0 for TYLD and RYLD under both water treatments. $\Delta^{13}C$ tended to exhibit more moderate values of H_{PR}.

Combining ability analysis of design II mating design

The male parents exhibited non-zero GCA effects in all trait-environment combinations, but the female parents had GCA effects of zero for some traits in 2018, but not in 2019 (Supplemental Table S2.5). G1 (IL2-5), G3 (IL5-4), G9 (TA1544), and G10 (TA1325) had the strongest horticulturally positive GCA effect on $\Delta^{13}C$ (a negative GCA effect for $\Delta^{13}C$ indicates a positive effect on WUE). The 2019 full water treatment was the only $\Delta^{13}C$-environment combination to have SCA effects. In general, the SCA effects for $\Delta^{13}C$ were lower in magnitude than the GCA effects. G16 (TA1544xIL5-4) and G20 (TA1325xIL9-1) had the strongest horticulturally positive SCA effects for $\Delta^{13}C$_2019_full.

All SCA values for yield traits (RYLD, TYLD, and HI) were zero in 2019. In 2018, we observed patterns of the relative magnitude of the yield GCA and SCA values associated with water treatments. For the full water treatment, the SCA for RYLD and TYLD were similar in magnitude to the GCA values. For the reduced water treatment, SCA values for RYLD and TYLD tended to be larger than the GCA values. G10 (TA1325) had the strongest positive GCA effect on yield traits in most years and water treatment. HomIL parent lines with introgressions on chromosome 5, G9 (TA1544) and G3 (IL5-4), tended to have negative GCA effects on yield traits.

G17 (TA1325xIL2-5) and G18 (TA1325xIL2-6) tended to have the highest positive SCA effects for RYLD and TYLD.

For FW15 and maturity trait-environment combinations, SCA tended to be lower in magnitude than GCA. G3 (IL5-4) had the most positive GCA effects for FW15 (an increase in fruit weight) and the *S. pennellii* chromosome 2 ILs (G1 (IL2-5) and G2 (IL2-6)) had the most negative GCA effects. Most parents had a positive horticultural effect on maturity (a negative GCA means less days to maturity) except for G3 (IL5-4) and G9 (TA1544) which had strong negative effects on maturity.

DAPG, DAPR, FW15, ACW, and SDW had moderate to high BSH and NSH (0.57 – 0.93) under all environmental combinations and a NSH/BSH ratio very close to one (Supplemental table S2.6). ACH had very low heritability (0.000 ─ 0.220), which may be partially due to the difficulty of obtaining consistent measurements for this trait. SFW and HI had moderate heritability (0.551 ─ 0.692 and 0.360 ─ 0.678, respectively) and NSH/BSH values close to one. RYLD had low heritability in 2018 (0.044 ─ 0.168) but moderate heritability in 2019 (0.455 ─ 0.542). TYLD had low heritability (0.048 ─ 0.361). The NSH/BSH values for RYLD and TYLD were less than one in 2018 for both water treatments and were particularly low in the reduced water treatment (0.262 and 0.320, respectively). $\Delta^{13}C$ had low to moderate heritability in each trait-environment combination (0.291 ─ 0.518) and NSH/BSH was high in all environment combinations.

DISCUSSION

Combining introgressions from two wild species did not cause a synergistic effect on WUE-related traits

In general, combining introgressions from two wild species into hybrid progeny did not result in a synergistic positive effect on WUE-related traits**.** Only two out of 13 DIHs had best-parent heterosis for $\Delta^{13}C$ in the reduced water treatment. The remaining DIHs tended to have positive mid-parent heterosis or H_{PR} values close to zero, which suggests additivity or dominance (incomplete or complete). Some studies have also observed evidence of predominantly additive gene action for $\delta^{13}C$ or $\Delta^{13}C$, while others have reported dominance effects of $\delta^{13}C$-associated QTL (Martin et al., 1989; Rebetzke et al., 2006; Xu et al., 2008). G24 (TA1324xIL5-4) was the only DIH to have statistically significant lower $\Delta^{13}C$ (i.e., higher WUE) than its best-parent, but it was not significantly different from the cultivated tomato control, E6203. Furthermore, most $\Delta^{13}C$-environment combinations had SCA effects of zero for every DIH. The SCA values observed for $\Delta^{13}C$ in 2019 in the full water treatment were generally lower in magnitude than the GCA values.

$\Delta^{13}C$ had NSH/BSH ratios close to 1 (>0.94) indicating that most of the genetic variance observed for this trait and environments was due to additive genetic variance. These results suggest that the hybrid offspring more closely resemble their parents for WUE-related traits. Although most DIHs did not perform better than their best HomIL parent for WUE-related traits, some did perform better than the cultivated tomato controls. These genotypes may prove useful for improving WUE in cultivated tomato. This will be discussed further below.

The observed heterosis for yield is most likely due to the wild introgressions

Most HetILs and DIHs exhibited best-parent heterosis for yield, and many had best-parent heterosis for maturity traits (DAPG and DAPR). The best-parent heterosis for yield and other traits observed in this study is unlikely to be due to interactions between the two cultivar genetic backgrounds, E6203 and M82, because the HetILs (backcrosses to the inbred recurrent parent) also showed best-parent heterosis. If heterosis was solely due to the hybridization of E6203 and M82, then HetILs wouldn't exhibit best-parent heterosis and all DIHs would exhibit similar levels of heterosis. Collectively, our results suggest that the observed best-parent heterosis for yield is likely due to inter-genic interactions among the introgressions from each species and/or interactions among the introgressions and the cultivar genetic backgrounds. In addition, the two cultivars are more genetically similar to each other than they are to either wild species genome (Miller & Tanksley, 1990; Park et al., 2004; Rick, 1983). Previous studies focused on hybrids between *S. pennellii* ILs and various cultivated tomatoes did not report interactions between introgressions and inbred cultivated tomato on traits (Eshed & Zamir, 1995; Gur et al., 2004). These prior studies suggest that the phenotypic effects of the introgressions may not depend on the specific *S. lycopersicum* genetic background.

Some introgressions may be useful for breeding hybrid tomato cultivars with improved WUE

Two introgressions from *S. habrochaites* were associated with increased WUE without significant reduction of yield, compared to cultivar controls E6203 and M82. HomIL G9 (TA1544), which contains an *S. habrochaites* chromosome 5 introgression, had the lowest $\Delta^{13}C$ value and, by inference, had the best WUE of all other genotypes in this study. HetIL G28 (E6203xTA1544) had significantly lower $\Delta^{13}C$ than both E6203 and M82 in the reduced water

treatment. This suggests that the introgression in TA1544 conferred a significant increase in WUE when it was either homozygous or heterozygous. G28 (E6203xTA1544) did not show significantly different yields than the controls in either water treatment. However, G9 (TA1544) did have significantly lower yields than the controls in the reduced water treatment and had negative GCA effects on yield traits. The TA1544 introgression may be useful for breeding improved WUE lines that do not exhibit a decrease in yield if it is present as only one copy in the hybrid offspring.

The horticulturally unacceptable values for some traits observed in G9 (TA1544), such as later maturity, larger canopy size, and lower yield, may be associated with the *SP5G* gene in this introgressed region of chromosome 5. A previous study found that the *S. pennellii* allele at the *SP5G* locus was associated with larger plant height and lower ripe fruit yields at harvest (Jones et al., 2007). Their finding is congruent with our high canopy size measurements and field observations for G3 (IL5-4) and G9 (TA1544), which each contain syntenic regions of chromosome 5. The *S. pennellii* allele of *SP5G* is also associated with day-length sensitivity, later maturity, and later yield (Jones et al., 2007; Soyk et al., 2017). It's possible that the presence of the *S. habrochaites* allele at *SP5G* has a similar negative phenotypic effect on these traits.

A preliminary analysis of an alignment of the *SP5G* sequence from *S. lycopersicum* (SL4.0, ITAG4.1; Sol Genomics Network, SGN, solgenomics.net), *S. pennellii* (Bolger et al., 2014; accessed from SGN), and *S. habrochaites* (LA1778; unpublished data) was conducted. Each of the four coding sequence (CDS) for *SP5G* were aligned using the Muscle algorithm with default settings in MEGA v10.2.6 and compared (data not shown) (Kumar et al., 2018). Translated amino acid sequences were also aligned and compared for nonsense and missense mutations, stop codons, etc. The first CDS had three missense mutations. Interestingly, the fourth CDS had a stop codon early on in the amino acid sequence that was present in *S. pennellii* and *S. habrochaites*, but not in

S. lycopersicum. If this stop codon is causing the phenotypic differences between *S. pennellii* and *S. lycopersicum*, then it may also cause later maturity and lower yields in *S. habrochaites*. Soyk et al. (2017) also compared the amino acid sequences of *SP5G* to various wild species and found that *S. habrochaites* (LA1777) has many amino acid substitutions in common with *S. pennellii* (LA716). Further studies would be needed to determine if the *S. habrochaites SP5G* allele is also associated with later maturity and lower yield.

G10 (TA1325), which contains a *S. habrochaites* chromosome 9 introgression, had significantly lower $\Delta^{13}C$ than both controls in the reduced water treatment and was significantly lower than M82 in the full water treatment. The HetIL containing the same introgression, G29 (E6203xTA1325), had significantly lower $\Delta^{13}C$ than M82, but not E6203, in both water treatments. Neither G10 (TA1325) nor G29 (E6203xTA1325) had significantly different yield values than the cultivated controls, indicating the introgression did not reduce yields. G10 (TA1325) had the strongest horticulturally beneficial GCA effects for yield and $\Delta^{13}C$ in most environment combinations (Supplemental Table S2.5). These results suggest that G10 (TA1325) has beneficial effects on $\Delta^{13}C$ and yield when used as a parent and could be beneficial to employ in a hybrid breeding program for improving WUE.

Specific hybrids had improved WUE and/or yield over controls

Some HetILs and DIHs had better WUE than the cultivar controls and provide insights into introgression combinations that may be useful in a breeding program for improved WUE in cultivated tomato. G16 (TA1544xIL5-4), G19 (TA1325xIL5-4), and G28 (E6203xTA1544) had significantly lower $\Delta^{13}C$ values (more WUE) than E6203 and M82 in the reduced water treatment. These three genotypes did not have significantly different yields than either control. Therefore, these hybrids are significantly more WUE than the controls without a significant reduction in yield.

The introgressions associated with these genotypes could be deployed for improving WUE in cultivated tomato in combination or individually, as discussed below in the Conclusion.

The DIH G17 (TA1325xIL2-5) had very high performance *per se* for most traits, but it also exhibited smaller values for FW15, which is undesirable in tomato breeding and suggests linkage drag. G17 (TA1325xIL2-5) had the best yield *per se* in both water treatments, was ranked very high for $\Delta^{13}C$ (Supplemental Table S2.2), and had positive SCA effects for most traits. However, it was only significantly different from M82, not E6203, for $\Delta^{13}C$ and yield. G17 (TA1325xIL2-5) also had significantly smaller fruit weight than both controls and G1 (IL2-5) has a strong negative GCA effect on FW15 in all trait-environment combinations. The low fruit weight observed in genotypes containing at least one copy of the IL2-5 introgression is likely related to three QTL associated with fruit weight discovered in this region (Eshed & Zamir, 1995). The combination of the TA1325 and IL2-5 introgressions have potential use to improve WUE and yield in a hybrid tomato breeding program, however the alleles for low fruit weight from *S. pennellii* would need to be selected against.

The specific environmental conditions play a major role in WUE-related studies

Crop phenotypes for quantitative traits are affected by the environment, including plant growing conditions. Therefore, plant materials should be evaluated in target breeding environments to determine if there are interactions between the conditions and the phenotypes of interest. Conducting experiments on plants in pots in a greenhouse can be convenient, but results from potted plant experiments cannot be extrapolated to open field conditions. Plants respond differently to water deficit, drought and other treatments in potted experiments due to various factors including root restriction, soil media type, pot temperature, and rooting zone depth (Comstock et al., 2005; Passioura, 2006; Poorter et al., 2012; Wu et al., 2011). Most prior

experiments on water stress tolerance using *S. habrochaites* and *S. pennellii* ILs were conducted in potted greenhouse or potted covered field experiments rather than in open field environments (Gong et al., 2010; Rigano et al., 2016; Truco et al., 2000; Xu et al., 2008). Thus, the results from potted experiments are not directly comparable to our field-based study. Here we compared water stress tolerant HomILs to their cultivated tomato controls in field experiments with plants grown in soil with unrestricted root zones under slow-onset water stress (season-long water deficit).

Our results differ from the potted greenhouse experiment by Xu et al. (2008) who found that IL5-4 was the only IL out of 50 that had a significantly better WUE (measured by $\delta^{13}C$) than M82. In contrast, *S. pennellii* HomILs (G1 ─ G6: IL2-5, IL2-6, IL9-2, IL9-1-2, IL9-2-5) and HetILs (G31 ─ G36) generally had poor WUE in the field. *S. pennellii* HomILs and HetILs were among the poorest performers for $\Delta^{13}C$ and none had significantly different $\Delta^{13}C$ values than M82, with the exception of G1 (IL2-5) (Supplemental Table S2.1). Most of these HomILs and HetILs did not have $\Delta^{13}C$ values that were significantly different from each other. These differing results illustrate the importance of evaluating quantitative traits in the field for effective plant breeding of field-grown crops.

The HomIL G3 (IL5-4) introgression may be useful for breeding improved WUE cultivars when crossed with specific introgressions. HomIL G3 (IL5-4) did not have greater WUE than M82, but it did have a high GCA effects for $\Delta^{13}C$, especially under reduced water. The two most WUE DIHs (described in subsection: Specific hybrids had improved WUE and/or yield over controls) had G3 (IL5-4) as a parent. However, the HetIL G33 (M82xIL5-4) did not have significantly different $\Delta^{13}C$ than M82 suggesting that, under field conditions, the IL5-4 introgression increases WUE via interactions with certain *S. habrochaites* introgressions.

CONCLUSION

Combining introgressions from two water stress-tolerant wild tomato species did not enhance WUE of hybrid offspring compared to their parents, yet some hybrids performed better than the cultivated tomato varieties that served as the genetic backgrounds for the IL libraries. Interestingly, DIHs and HetILs tended to have best-parent heterosis for yield. This suggests that although wild species of tomatoes tend to have lower yields than cultivated tomato, they also contain alleles that have positive effects on yield and could be utilized in breeding hybrid cultivars. TA1544, TA1325, IL5-4, and IL2-5 may be useful for developing hybrid tomato varieties with improved WUE without negative effects on yield. Since some of these introgressions are on different chromosomes, they could be utilized in a hybrid breeding program in two ways. Introgressions could be transferred into the same parent line and crossed with a second parent line that has other valuable horticultural traits. Alternatively, different introgressions could be transferred into maternal and paternal lines separately via backcrossing, then the parent lines could be crossed with each other to create a hybrid. Although the introgressions mentioned above have beneficial effects on WUE, some do exhibit negative linkage drag for lower fruit weight and later maturity. For IL2-5 and IL5-4, potential causes of the linkage drag observed in our study have been QTL mapped by others and marker assisted selection could be used to select for cultivated tomato alleles at these loci.

ACKNOWLEDGMENTS

This research was supported by the California Tomato Research Institute and the Jastro-Shields Research Scholarship Award. We thank the St. Clair lab members for contributions to field and lab data collection. We also thank the UC Davis Plant Sciences field facility crew for assistance with field experiments. We thank Mitchell Feldmann for his analysis advice.

REFERENCES

Arms, E. M., Bloom, A. J., & St. Clair, D. A. (2015). High-resolution mapping of a major effect QTL from wild tomato *Solanum habrochaites* that influences water relations under root chilling. *Theoretical and Applied Genetics*, *128*(9), 1713–1724. https://doi.org/10.1007/s00122-015-2540-y

Barrios-Masias, F. H., Chetelat, R. T., Grulke, N. E., & Jackson, L. E. (2013). Use of introgression lines to determine the ecophysiological basis for changes in water use efficiency and yield in California processing tomatoes. *Functional Plant Biology*, *41*(2), 119–132.

Berg, N., & Hall, A. (2017). Anthropogenic warming impacts on California snowpack during drought. *Geophysical Research Letters*, *44*(5), 2511–2518. https://doi.org/10.1002/2016GL072104

Blanca, J., Montero-Pau, J., Sauvage, C., Bauchet, G., Illa, E., Díez, M. J., Francis, D., Causse, M., Van der Knaap, E., & Cañizares, J. (2015). Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *BMC Genomics*, *16*(1), 1–19.

Bolger, A., Scossa, F., Bolger, M. E., Lanz, C., Maumus, F., Tohge, T., Quesneville, H., Alseekh, S., Sørensen, I., Lichtenstein, G., Fich, E. A., Conte, M., Keller, H., Schneeberger, K., Schwacke, R., Ofner, I., Vrebalov, J., Xu, Y., Osorio, S., … Fernie, A. R. (2014). The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nature Genetics*, *46*(9), 1034–1038. https://doi.org/10.1038/ng.3046

Comstock, J. P., McCouch, S. R., Martin, B. C., Tauer, C. G., Vision, T. J., Xu, Y., & Pausch, R. C. (2005). The effects of resource availability and environmental conditions on genetic rankings for carbon isotope discrimination during growth in tomato and rice. *Functional Plant Biology*, *32*(12), 1089–1105. https://doi.org/10.1071/FP05117

Covarrubias-Pazaran, G. (2016). Genome-assisted prediction of quantitative traits using the R package sommer. *PloS One*, *11*(6), e0156744.

Covarrubias-Pazaran, G. (2022). Quantitative genetics using the sommer package. *R Found. Stat. Comput., Vienna. Https://Cran. r-Project. Org/Web/Packages/Sommer/Vignettes/Sommer. Pdf (Accessed 6 June. 2022)*.

de Vienne, D., & Fiévet, J. B. (2020). The pitfalls of heterosis coefficients. *Plants*, *9*(7), 875.

Eshed, Y., & Zamir, D. (1995). An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics*, *141*(3), 1147–1162.

FAOSTAT. (2019). *FAOSTAT Data*. https://www.fao.org/faostat/en/#data

Farquhar, G., & Richards, R. (1984). Isotopic composition of plant carbon correlates with water-use efficiency of wheat genotypes. *Functional Plant Biology*, *11*(6), 539–552.

Fehr, W. R. (1987). *Principles of cultivar development. Volume 1. Theory and technique.* Macmillan publishing company.

Gong, P., Zhang, J., Li, H., Yang, C., Zhang, C., Zhang, X., Khurram, Z., Zhang, Y., Wang, T., Fei, Z., & Ye, Z. (2010). Transcriptional profiles of drought-responsive genes in modulating transcription signal transduction, and biochemical pathways in tomato. *Journal of Experimental Botany*, *61*(13), 3563–3575. https://doi.org/10.1093/jxb/erq167

Groh, A. M., Kubond, B. A., & St. Clair, D. A. (2022). Fine mapping of QTL for water use efficiency-related traits on chromosome 9 of *Solanum habrochaites* in the field. *Crop Science*. DOI: 10.1002/csc2.20828. (*in press*).

Gur, A., Zamir, D., & Dangl, J. (2004). Unused natural variation can lift yield barriers in plant breeding. *PLoS Biology*, *2*(10), e245.

Hallauer, A. R., Carena, M. J., & Miranda Filho, J. de. (2010). *Quantitative genetics in maize breeding* (Vol. 6). Springer Science & Business Media.

Jones, C. M., Rick, C. M., Adams, D., Jernstedt, J., & Chetelat, R. T. (2007). Genealogy and fine mapping of obscuravenosa, a gene affecting the distribution of chloroplasts in leaf veins, and evidence of selection during breeding of tomatoes (*Lycopersicon esculentum*; Solanaceae). *American Journal of Botany*, *94*(6), 935–947.

Jones, H. G. (2007). Monitoring plant and soil water status: Established and novel methods revisited and their relevance to studies of drought tolerance. *Journal of Experimental Botany*, *58*(2), 119–130. https://doi.org/10.1093/jxb/erl118

Krieger, U., Lippman, Z. B., & Zamir, D. (2010). The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nature Genetics*, *42*(5), 459–463.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547.

Liedl, B. E., Labate, J. A., Stommel, J. R., Slade, A., & Kole, C. (2013). *Genetics, genomics, and breeding of tomato*. CRC Press.

Lippman, Z. B., Semel, Y., & Zamir, D. (2007). An integrated view of quantitative trait variation using tomato interspecific introgression lines. *Current Opinion in Genetics & Development*, *17*(6), 545–552.

Lounsbery, J. K., Arms, E. M., Bloom, A. J., & St. Clair, D. A. (2016). Quantitative trait loci for water-stress tolerance traits localize on chromosome 9 of wild tomato. *Crop Science*, *56*(4), 1514–1525. https://doi.org/10.2135/cropsci2015.07.0432

Martin, B., Nienhuis, J., King, G., & Schaefer, A. (1989). Restriction fragment length polymorphisms associated with water use efficiency in tomato. *Science*, *243*(4899), 1725–1728.

Martin, B., & Thorstenson, Y. R. (1988). Stable carbon isotope composition (δ13C), water use efficiency, and biomass productivity of *Lycopersicon esculentum*, *Lycopersicon pennellii*, and the F1 hybrid. *Plant Physiology*, *88*(1), 213-217.

Miller, J., & Tanksley, S. (1990). RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon. *Theoretical and Applied Genetics*, *80*(4), 437–448.

Monforte, A. J., & Tanksley, S. D. (2000). Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: a tool for gene mapping and gene discovery. *Genome*, *43*(5), 803-813.

Nuruddin, M. M., Madramootoo, C. A., & Dodds, G. T. (2003). Effects of water stress at different growth stages on greenhouse tomato yield and quality. *HortScience*, *38*(7), 1389–1393.

O'Leary, M. H. (1988). Carbon isotopes in photosynthesis. *Bioscience*, *38*(5), 328–336.

Park, Y. H., West, M. A., & St. Clair, D. A. (2004). Evaluation of AFLPs for germplasm fingerprinting and assessment of genetic diversity in cultivars of tomato (*Lycopersicon esculentum L.*). *Genome*, *47*(3), 510–518.

Passioura, J. B. (2006). The perils of pot experiments. *Functional Plant Biology*, *33*(12), 1075–1079.

Pathak, T. B., Maskey, M. L., Dahlberg, J. A., Kearns, F., Bali, K. M., & Zaccaria, D. (2018). Climate change trends and impacts on California agriculture: A detailed review. *Agronomy*, *8*(3), 25.

Poorter, H., Bühler, J., van Dusschoten, D., Climent, J., & Postma, J. A. (2012). Pot size matters: A meta-analysis of the effects of rooting volume on plant growth. *Functional Plant Biology*, *39*(11), 839–850.

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rebetzke, G., Richards, R., Condon, A. G., & Farquhar, G. (2006). Inheritance of carbon isotope discrimination in bread wheat (*Triticum aestivum L.*). *Euphytica*, *150*(1), 97–106.

Rick, C. (1979). *Biosystematic studies in Lycopersicon and closely related species for Solanum*.

Rick, C. M. (1973). Potential genetic resources in tomato species: Clues from observations in native habitats. In *Genes, enzymes, and populations* (pp. 255–269). Springer.

Rick, C. M. (1983). Genetic variability in tomato species. *Plant Molecular Biology Reporter*, *1*(2), 81–87.

Rigano, M. M., Arena, C., Matteo, A. D., Sellitto, S., Frusciante, L., & Barone, A. (2016). Eco-physiological response to water stress of drought-tolerant and drought-sensitive tomato genotypes. *Plant Biosystems*, *150*(4), 682–691. https://doi.org/10.1080/11263504.2014.989286

Soyk, S., Müller, N. A., Park, S. J., Schmalenbach, I., Jiang, K., Hayama, R., Zhang, L., Van Eck, J., Jimenez-Gomez, J. M., & Lippman, Z. B. (2017). Variation in the flowering gene SELF PRUNING 5G promotes day-neutrality and early yield in tomato. *Nature Genetics*, *49*(1), 162–168.

Tanksley, S. D., & McCouch, S. R. (1997). Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science*, *277*(5329), 1063–1066.

Thalapati, S., Guttikonda, H., Nannapaneni, N. D., Adari, P. B., Surendhar Reddy, C., Swamy, B., Batchu, A. K., Basava, R. K., Viraktamath, B., & Neelamraju, S. (2015). Heterosis and combining ability in rice as influenced by introgressions from wild species *Oryza rufipogon* including qyld2. 1 sub-QTL into the restorer line KMR3. *Euphytica*, *202*(1), 81–95.

Truco, M., Randall, L., Bloom, A., & St. Clair, D. (2000). Detection of QTLs associated with shoot wilting and root ammonium uptake under chilling temperatures in an interspecific backcross population from *Lycopersicon esculentum*× *L. hirsutum*. *Theoretical and Applied Genetics*, *101*(7), 1082–1092.

Tuberosa, R. (2012). Phenotyping for drought tolerance of crops in the genomics era. *Frontiers in Physiology*, *3 SEP*. https://doi.org/10.3389/fphys.2012.00347

USDA ERS. (2019). *Potatoes and tomatoes are the most commonly consumed vegetables*. https://www.ers.usda.gov/data-products/chart-gallery/gallery/chart-detail/?chartId=58340

Wehner, T. (1999). Heterosis in vegetable crops. *Genetics and Exploitation of Heterosis in Crops*, 387–397.

Wu, Y., Huang, M., & Warrington, D. N. (2011). Growth and transpiration of maize and winter wheat in response to water deficits in pots and plots. *Environmental and Experimental Botany*, *71*(1), 65–71.

Xu, X., Martin, B., Comstock, J. P., Vision, T. J., Tauer, C. G., Zhao, B., Pausch, R. C., & Knapp, S. (2008). Fine mapping a QTL for carbon isotope composition in tomato. *Theoretical and Applied Genetics*, *117*(2), 221–233. https://doi.org/10.1007/s00122-008-0767-6

Zhang, J., Wu, M., Yu, J., Li, X., & Pei, W. (2016). Breeding potential of introgression lines developed from interspecific crossing between upland cotton (*Gossypium hirsutum*) and *Gossypium barbadense*: Heterosis, combining ability and genetic effects. *PLoS One*, *11*(1), e0143646.

TABLES AND FIGURES

Table 2.1. List of 35 genotypes evaluated in the field for horticultural and water use efficiency-related traits and their genotype codes. The cultivars E6203 and M82 (G12 and G13, respectively) are controls for the genetic background. A genotype is designated as a homozygous IL (HomIL) if it is an inbred IL that is homozygous for the designated introgression. A double introgression hybrid (DIH) is the hybrid of a cross between two different HomILs. A heterozygous IL (HetIL) is the cross between a control for the genetic background and a HomIL. The accession numbers are derived from the Tomato Genetics Resource Center (https://tgrc.ucdavis.edu) where the HomILs were sourced from.

Genotype Code †	Designation	Source of introgression(s)	Introgression Chromosome	IL/Cross/Cultivar	Accession
G1	HomIL	*S. pennellii*	2	IL2-5	LA4040
G2	HomIL	*S. pennellii*	2	IL2-6	LA4041
G3	HomIL	*S. pennellii*	5	IL5-4	LA4057
G4	HomIL	*S. pennellii*	9	IL9-1	LA4078
G5	HomIL	*S. pennellii*	9	IL9-1-2	LA4079
G6	HomIL	*S. pennellii*	9	IL9-2-5	LA4082
G7	HomIL	*S. habrochaites*	2	TA1266	LA3922
G9	HomIL	*S. habrochaites*	5	TA1544	LA3943
G10	HomIL	*S. habrochaites*	9	TA1325	LA3957
G11	HomIL	*S. habrochaites*	9	TA1324	LA3956
G12	Control	-	control	E6203	-
G13	Control	-	control	M82	-
G14	DIH	*S. habrochaites*, *S. pennellii*	2 x 2	TA1266 x IL2-5	-
G15	DIH	*S. habrochaites*, *S. pennellii*	2 x 2	TA1266 x IL2-6	-
G16	DIH	*S. habrochaites*, *S. pennellii*	5 x 5	TA1544 x IL5-4	-
G17	DIH	*S. habrochaites*, *S. pennellii*	9 x 2	TA1325 x IL2-5	-
G18	DIH	*S. habrochaites*, *S. pennellii*	9 x 2	TA1325 x IL2-6	-

G19	DIH	*S. habrochaites*, *S. pennellii*	9 x 5	TA1325 x IL5-4	-
G20	DIH	*S. habrochaites*, *S. pennellii*	9 x 9	TA1325 x IL9-1	-
G21	DIH	*S. habrochaites*, *S. pennellii*	9 x 9	TA1325 x IL9-1-2	-
G22	DIH	*S. habrochaites*, *S. pennellii*	9 x 2	TA1324 x IL2-5	-
G23	DIH	*S. habrochaites*, *S. pennellii*	9 x 2	TA1324 x IL2-6	-
G24	DIH	*S. habrochaites*, *S. pennellii*	9 x 5	TA1324 x IL5-4	-
G25	DIH	*S. habrochaites*, *S. pennellii*	9 x 9	TA1324 x IL9-1	-
G26	DIH	*S. habrochaites*, *S. pennellii*	9 x 9	TA1324 x IL9-1-2	-
G27	HetIL	*S. lycopersicum, S. habrochaites*	2	E6203 x TA1266	-
G28	HetIL	*S. lycopersicum, S. habrochaites*	5	E6203 x TA1544	-
G29	HetIL	*S. lycopersicum, S. habrochaites*	9	E6203 x TA1325	-
G30	HetIL	*S. lycopersicum, S. habrochaites*	9	E6203 x TA1324	-
G31	HetIL	*S. lycopersicum*, *S. pennellii*	2	M82 x IL2-5	-
G32	HetIL	*S. lycopersicum*, *S. pennellii*	2	M82 x IL2-6	-
G33	HetIL	*S. lycopersicum*, *S. pennellii*	5	M82 x IL5-4	-
G34	HetIL	*S. lycopersicum*, *S. pennellii*	9	M82 x IL9-1	-
G35	HetIL	*S. lycopersicum*, *S. pennellii*	9	M82 x IL9-1-2	-
G36	HetIL	*S. lycopersicum*, *S. pennellii*	9	M82 x IL9-2-5	-

† G8 was eliminated due to limited field space and was not included in field evaluation

Table 2.2. Summary of ANOVA results for 35 genotypes including cultivar controls, homozygous introgression lines (ILs), heterozygous ILs, and double introgression hybrids. Each trait was measured in field-grown plants across two years under full and reduced water treatments. Trait data was separated into sub-analyses based on factors that caused heterogeneity of variance or significant interactions

Trait	Year	Location	Water treatment	F value Year	F value Location	F value Water treatment	F value Genotype	F value Location x Water	F value Location x Genotype	F value Location x Year	F value Water x Genotype	F value Genotype x Year	F value Location x Water x Genotype	F value Location x Genotype x Year
DAPG[a]	2018	1	full + reduced	-	-	0.60 NS	37.94 ***	-	-	-	0.79 *	-	-	-
		2	full + reduced	-	-	11.03 ***	30.96 ***	-	-	-	1.27 NS	-	-	-
	2019	2	full	-	-	-	8.17 ***	-	-	-	-	-	-	-
		2	reduced	-	-	-	13.73 ***	-	-	-	-	-	-	-
DAPR[b]	2018	1 + 2	full + reduced	-	40.73 ***	8.39 **	111.87 ***	5.56*	2.19 ***	-	1.45 *	-	1.12 NS	-
	2019	1 + 2	full + reduced	-	3.26 NS	22.31 ***	56.94 ***	9.68 **	1.76 **	-	1.04 NS	-	1.22 NS	-
FW15[c]	2018 + 2019	1 + 2	full	0.54 NS	3.95 NS	-	168.27 ***	-	0.82 NS	2.65 NS	-	3.13 ***	-	1.21 NS
		1 + 2	reduced	0.00 NS	6.87 NS	-	121.92 ***	-	1.27 NS	0.27 NS	-	1.60 *	-	0.69 NS
ACW[d]	2018 + 2019	1 + 2	full	0.00 NS	0.14 NS	-	66.91 ***	-	1.01 NS	0.21 NS	-	2.17 ***	-	0.85 NS
		1 + 2	reduced	2.37 NS	3.55 NS	-	53.07 ***	-	9.87 NS	9.61 **	-	2.68 ***	-	1.82 **
ACH[e]	2018	1 + 2	full	-	6.83 NS	-	3.23 ***	-	0.96 NS	-	-	-	-	-
		1 + 2	reduced	-	7.41 NS	-	2.74 ***	-	1.98 **	-	-	-	-	-
	2019	1 + 2	full	-	34.61 NS	-	7.35 ***	-	1.17 NS	-	-	-	-	-
		1 + 2	reduced	-	21.48 *	-	6.43 ***	-	1.60 *	-	-	-	-	-
SFW[f]	2018 + 2019	1 + 2	full	0.00 NS	0.14 NS	-	17.22 ***	-	1.53 *	0.12 NS	-	1.33 NS	-	1.60 *
		1 + 2	reduced	0.06 NS	0.21 NS	-	9.69 ***	-	1.11 NS	0.34 NS	-	1.58 *	-	0.85 NS
SDW[g]	2018 + 2019	1 + 2	full	0.00 NS	0.27 NS	-	22.19 ***	-	1.57 *	0.20 NS	-	2.16 ***	-	1.55 *
		1 + 2	reduced	0.01 NS	0.15 NS	-	18.86 ***	-	1.29 NS	0.21 NS	-	1.27 NS	-	1.46 NS
RYLD[h]	2018 + 2019	1 + 2	full	0.01 NS	0.02 NS	-	5.00 ***	-	1.01 NS	0.02 NS	-	1.48 *	-	1.15 NS

		1 + 2	reduced	0.05 NS	0.07 NS	-	5.66 ***	-	1.37 NS	0.04 NS	-	1.68 *	-	1.50 *
TYLD[i]	2018 + 2019	1 + 2	full	0.06 NS	0.38 NS	-	4.46 ***	-	1.09 NS	0.17 NS	-	1.44 NS	-	1.03 NS
		1 + 2	reduced	0.06 NS	0.08 NS	-	5.79 ***	-	1.29 NS	0.08 NS	-	1.79 **	-	1.41 NS
HI[j]	2018	1 + 2	full + reduced	-	0.12 NS	3.96 *	7.36 ***	8.68 **	1.10 NS	-	1.37 NS	-	1.10 NS	-
	2019	1 + 2	full + reduced	-	1.38 NS	9.31 **	11.43 ***	0.00 NS	1.11 NS	-	0.94 NS	-	0.85 NS	-
$\Delta^{13}C$[k]	2018 + 2019	1 + 2	full	2.11 NS	0.31 NS	-	22.72 ***	-	1.30 NS	0.34 NS	-	1.85 **	-	0.79 NS
		1 + 2	reduced	0.00 NS	0.00 NS	-	31.95 ***	-	0.81 NS	0.00 NS	-	1.63 *	-	1.25 NS

*** 0.001 significance level; ** 0.01 significance level; * 0.05 significance level; †ns: nonsignificant at the 0.5 significance level; - not applicable

[a] Days after planting to first green fruit, [b] days after planting to first ripe fruit, [c] weight of 15 fruit, [d] average canopy width, [e] average canopy height, [f] shoot fresh weight, [g] shoot dry weight, [h] ripe fruit yield, [i] total fruit yield, [j] harvest index, [k] carbon isotope discrimination

Table 2.3. Potence ratio (H_{PR}) values for each double introgression hybrid (DIH) or heterozygous introgression line (HetIL) genotype for a subset of trait-environment combinations. The color of a H_{PR} value represents the range of heterosis that the value falls into: dark blue, best-parent heterosis (H_{PR} >1.2); light blue, positive mid-parent heterosis (0.2 ─ 1.2); yellow, no heterosis (additive gene action) (-0.2 ─ 0.2); negative mid-parent heterosis (-1.2 ─ -0.2); worst-parent heterosis (<-1.2). The trait acronyms are: days after planting to first ripe fruit (DAPR), weight of 15 fruit (FW15), shoot dry weight (SDW), ripe fruit yield (RYLD), harvest index (HI), and carbon isotope discrimination ($\Delta^{13}C$). The trait acronym is followed by the environmental factors used to subset the trait datasets (separated by underscores). These environmental factors include year (2018 or 2019) and/or full or reduced water treatments (full or red, respectively).

IL/Cross	Genotype Code	DAPR_2018	DAPR_2019	FW15_full	FW15_red	SDW_full	SDW_red	RYLD_full	RYLD_red	HI_2018	HI_2019	Δ13C_full	Δ13C_red
TA1266 x IL2-5	G14	1.09	1.66	-0.53	-0.65	-0.16	-0.98	3.85	0.23	0.84	0.79	-1.29	-2.36
TA1266 x IL2-6	G15	13.32	2.17	-0.44	-0.45	-0.02	1.31	1.10	0.76	0.46	0.54	0.20	0.08
TA1544 x IL5-4	G16	1.63	0.51	0.85	1.99	8.70	6.74	0.22	1.54	-0.20	-1.22	0.19	0.43
TA1325 x IL2-5	G17	1.92	2.55	-0.67	-0.73	0.10	-0.04	3.69	15.06	4.64	1.55	-0.17	-0.72
TA1325 x IL2-6	G18	3.99	2.07	-0.60	-0.48	3.90	11.58	10.77	25.17	2.06	1.46	0.22	0.50
TA1325 x IL5-4	G19	0.62	0.59	-0.54	0.10	0.51	0.78	5.35	1.06	-1.88	0.22	0.91	1.90
TA1325 x IL9-1	G20	3.97	7.21	-0.21	-0.04	-1.23	0.20	1.74	12.71	2.08	1.68	0.54	0.41
TA1325 x IL9-1-2	G21	4.93	6.63	0.04	-0.12	-0.11	-1.05	7.54	21.75	3.49	3.04	0.89	0.27
TA1324 x IL2-5	G22	1.34	1.03	-0.58	-0.67	-0.34	-0.63	2.85	1.18	-0.05	0.56	-0.24	-0.54
TA1324 x IL2-6	G23	207.00	8.46	-0.38	-0.48	0.33	0.77	2.05	4.83	-1.48	1.03	0.59	0.13
TA1324 x IL5-4	G24	0.31	0.34	0.01	0.55	0.11	0.70	18.57	3.24	0.27	0.43	2.70	9.12
TA1324 x IL9-1	G25	1.64	0.95	-0.18	-0.05	-0.08	-0.15	1.44	14.68	0.24	0.99	0.30	0.88
TA1324 x IL9-1-2	G26	1.80	1.32	0.00	0.25	-0.55	-1.40	16.88	2.51	0.16	4.53	2.43	0.87
E6203 x TA1266	G27	0.42	1.55	0.06	0.49	-0.95	-0.36	3.72	-0.20	4.98	-0.38	-2.43	0.48
E6203 x TA1544	G28	0.13	-0.47	1.16	0.96	-0.15	0.66	3.19	1.86	0.93	0.26	-0.64	0.30
E6203 x TA1325	G29	3.05	1.58	-1.12	5.44	0.58	0.92	13.29	7.26	1.74	-1.12	-2.82	0.10
E6203 x TA1324	G30	11.00	-0.83	-17.80	15.62	-2.34	0.00	-4.58	-1.12	0.86	-0.04	0.42	18.00
M82 x IL2-5	G31	0.72	1.31	-0.39	-0.49	0.26	-0.23	8.62	3.37	0.69	1.89	-0.50	-0.46
M82 x IL2-6	G32	16.33	4.35	-0.30	-0.30	-0.30	0.84	1.37	3.29	-20.09	3.89	-0.55	-1.42
M82 x IL5-4	G33	-0.19	-0.11	0.72	1.24	0.01	0.72	7.42	1.97	0.39	1.19	4.24	0.59
M82 x IL9-1	G34	0.96	0.97	-0.21	0.20	-0.31	-0.43	0.96	42.65	1.03	6.13	0.93	7.20
M82 x IL9-1-2	G35	1.46	1.83	1.17	0.21	-1.46	-1.67	0.48	-0.61	-0.09	1.23	0.00	6.77
M82 x IL9-2-5	G36	-0.09	1.60	1.48	1.96	0.37	-0.32	1.00	22.37	0.20	188.24	-4.72	-2.08

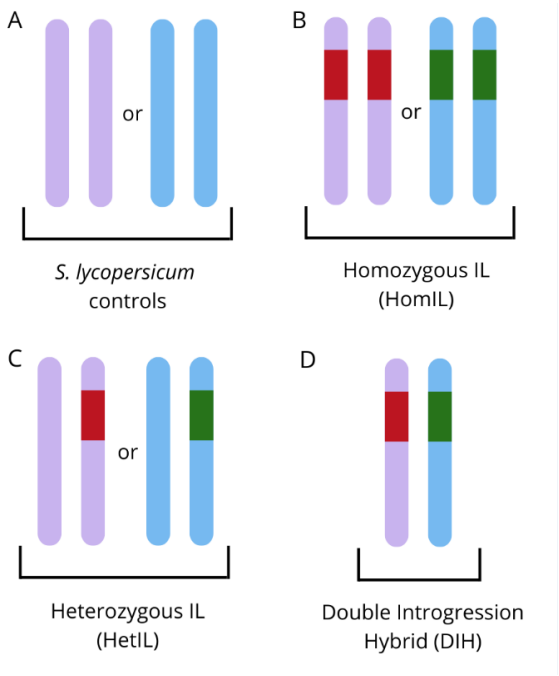
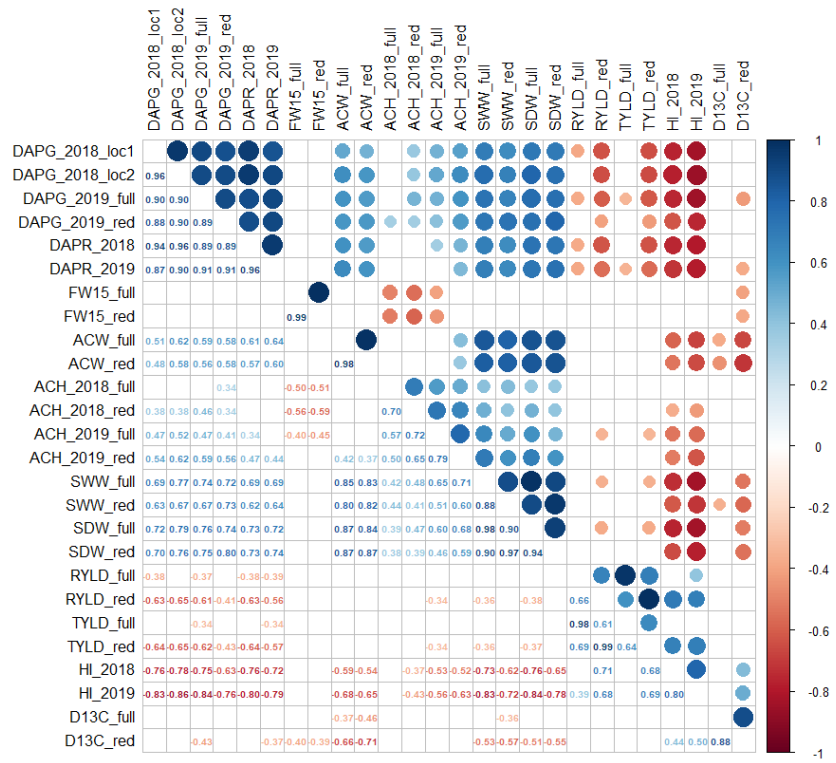
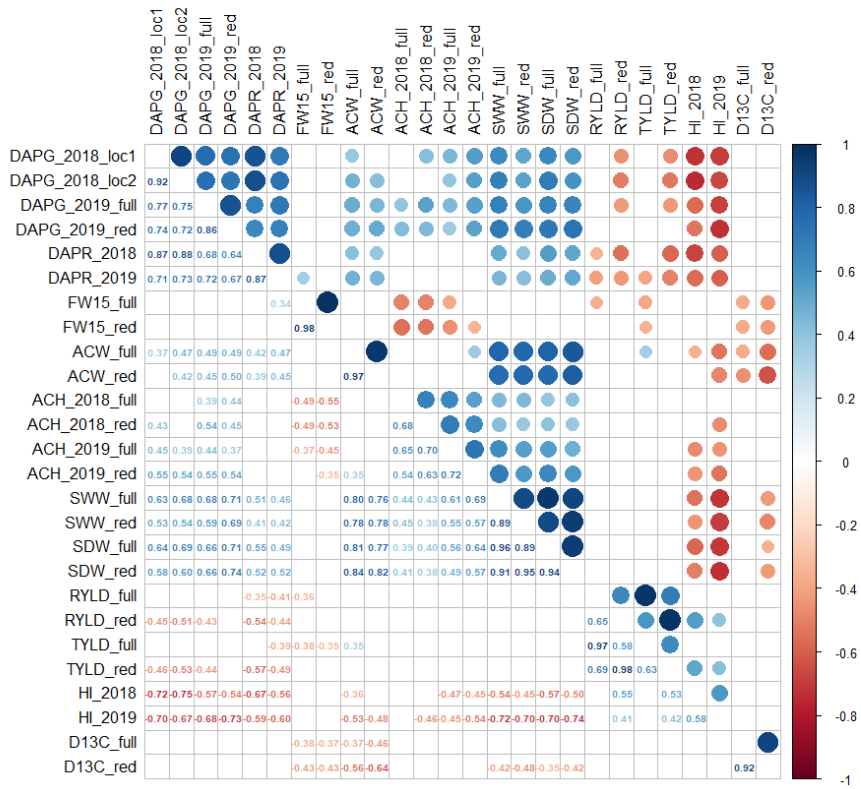


Figure 2.1. Schematic representation of genotypes used in this study. Each pair of ovals represents a diploid chromosome pair. Rectangles within the chromosomes represent a wild species introgression. (A) The inbred line cultivar controls for the genetic background. Purple represents the E6203 genetic background and blue represents the M82 genetic background. (B) Homozygous ILs (HomILs) are inbred lines containing two copies of an introgression from *S. habrochaites* (red) or *S. pennellii* (green). (C) Heterozygous ILs (HetILs) are the offspring of backcrosses to the control for the genetic background and contain only one copy of an introgression. (D) Double Introgression Hybrids (DIHs) are the result of crossing two HomILs (*S. habrochaites* HomIL x *S. pennellii* HomIL). These genotypes contain one copy of an introgression from each species.


DAPG_2018_loc1
DAPG_2018_loc2
DAPG_2019_full
DAPG_2019_red
DAPR_2018
DAPR_2019
FW15_full
FW15_red
ACW_full
ACW_red
ACH_2018_full
ACH_2018_red
ACH_2019_full
ACH_2019_red
SWW_full
SWW_red
SDW_full
SDW_red
RYLD_full
RYLD_red
TYLD_full
TYLD_red
HI_2018
HI_2019
D13C_full
D13C_red
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

A


DAPG_2018_loc1
DAPG_2018_loc2
DAPG_2019_full
DAPG_2019_red
DAPR_2018
DAPR_2019
FW15_full
FW15_red
ACW_full
ACW_red
ACH_2018_full
ACH_2018_red
ACH_2019_full
ACH_2019_red
SWW_full
SWW_red
SDW_full
SDW_red
RYLD_full
RYLD_red
TYLD_full
TYLD_red
HI_2018
HI_2019
D13C_full
D13C_red
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

B

Figure 2.2. Pearson (A) and Spearman Rank (B) correlations for each trait-environment combination. The upper triangle is a heat map with the color of the circles representing positive (blue) or negative (red) correlations, and the size of the circles represents the magnitude of the corresponding r value (larger is closer to 1 or -1). The lower diagonal shows the numerical values for all pairwise r values. The color of the lower diagonal corresponds with the color of the upper diagonal. Only significant ($P \leq 0.05$) r values are represented in the figures. The trait acronyms are: days after planting to first green fruit (DAPG), days after planting to first ripe fruit (DAPR), weight of 15 fruit (FW15), average canopy width (ACW), average canopy height (ACH), shoot fresh weight (SFW), shoot dry weight (SDW), ripe fruit yield (RYLD), total fruit yield (TYLD), harvest index (HI), carbon isotope discrimination ($\Delta^{13}C$). The trait acronym is followed by the environmental factors used to subset the trait datasets (separated by underscores). These environmental factors include year (2018 or 2019), full or reduced water treatments (full or red, respectively), and/or location (loc1 or loc2).

SUPPLEMENTAL MATERIAL

Supplemental Table S2.1: ANOVA mean separations for each trait-environment combination. The trait acronyms are: days after planting to first green fruit (DAPG), days after planting to first ripe fruit (DAPR), weight of 15 fruit (FW15, kg), average canopy width (ACW, cm), average canopy height (ACH, cm), shoot fresh weight (SFW, kg), shoot dry weight (SDW, kg), ripe fruit yield (RYLD, kg), total fruit yield (TYLD, kg), harvest index (HI), carbon isotope discrimination ($\Delta^{13}C$, ‰). The trait acronym is followed by the environmental factors used to subset the trait datasets (separated by underscores). These environmental factors include year (2018 or 2019), full

or reduced water treatments (full or red, respectively), and/or location (loc1 or loc2). The back-transformed emmeans are displayed where appropriate. Each trait has the genotypes with the most horticulturally favorable value for that trait as the first row of the table and the genotype with the least horticulturally favorable value for the trait as the last row of the table. The letters in the Groups column show which genotypes are significantly different from one another. Genotypes that share at least one letter in the Groups column are not significantly different from each another. Horticultural favorability was defined as follows: earlier maturity, larger fruit, smaller canopies, larger SFW and SDW, higher yields, and lower $\Delta^{13}C$. Red genotypes are the cultivated tomato controls (E6203 and M82). Green genotypes are significantly different from one cultivated tomato control. Blue genotypes are significantly different from both cultivated tomato controls.

Supplemental Table S2.2: Phenotypic ranking of each genotype (including inbreds, hybrids, and cultivated tomato controls) for each trait-environment combination under two water treatments. The trait acronyms are: days after planting to first green fruit (DAPG), days after planting to first ripe fruit (DAPR), weight of 15 fruit (FW15), average canopy width (ACW), average canopy height (ACH), shoot fresh weight (SFW), shoot dry weight (SDW), ripe fruit yield (RYLD), total fruit yield (TYLD), harvest index (HI), carbon isotope discrimination ($\Delta^{13}C$). Environmental factors include year (2018 or 2019), full or reduced water treatments (full or red, respectively), and/or location (loc1 or loc2). Colors represent the value of the ranking for a given trait. A ranking of 1 is the most horticulturally favorable genotype for a given trait (the most saturated green) and a ranking of 35 is the least horticulturally favorable genotype for a given trait (the most saturated red). Earlier maturity (DAPG and DAPR), larger fruit (FW15), larger SFW and SDW, higher yield

(RYLD and TYLD), higher HI, and lower $\Delta^{13}C$ values were considered more horticulturally favorable.

Supplemental Table S2.3: Emmeans of 35 genotypes for each trait-environment combination grouped by parents and offspring. The back-transformed emmeans are displayed where appropriate. The trait acronyms are: days after planting to first green fruit (DAPG), days after planting to first ripe fruit (DAPR), weight of 15 fruit (FW15, kg), average canopy width (ACW, cm), average canopy height (ACH, cm), shoot fresh weight (SFW, kg), shoot dry weight (SDW, kg), ripe fruit yield (RYLD, kg), total fruit yield (TYLD, kg), harvest index (HI), carbon isotope discrimination ($\Delta^{13}C$, ‰). The trait acronym is followed by the environmental factors used to subset the trait datasets (separated by underscores). These environmental factors include year (2018 or 2019), full or reduced water treatments (full or red, respectively), and/or location (loc1 or loc2). The colors represent whether the value is more horticulturally favorable for a given trait. The most saturated green value for a given trait-environment combination has the most favorable value and the most saturated red is the least favorable. Earlier maturity (DAPG and DAPR), larger fruit (FW15), larger biomass (SFW and SDW), higher yield (RYLD and TYLD), higher HI, and lower $\Delta^{13}C$ values were considered more horticulturally favorable.

Supplemental Table S2.4: Potence ratio (H_{PR}) values for each DIH or HetIL genotype for all trait-environment combinations. The color of a H_{PR} value represents the range of heterosis that the value falls into: dark blue, best-parent heterosis (H_PR >1.2); light blue, positive mid-parent heterosis (0.2 ─ 1.2); yellow, no heterosis (additive gene action) (-0.2 ─ 0.2); negative mid-parent heterosis (-1.2 ─ -0.2); worst-parent heterosis (<-1.2). The trait acronyms are: days after planting to first

green fruit (DAPG), days after planting to first ripe fruit (DAPR), weight of 15 fruit (FW15), average canopy width (ACW), average canopy height (ACH), shoot fresh weight (SFW), shoot dry weight (SDW), ripe fruit yield (RYLD), total fruit yield (TYLD), harvest index (HI), carbon isotope discrimination ($\Delta^{13}C$). Environmental factors include year (2018 or 2019), full or reduced water treatments (full or red, respectively), and/or location (loc1 or loc2).

Supplemental Table S2.5: General and specific combining ability (GCA and SCA) effects of the Design II mating design for a subset of trait-environment combinations. The trait acronyms are: days after planting to first green fruit (DAPG), days after planting to first ripe fruit (DAPR), weight of 15 fruit (FW15), average canopy width (ACW), average canopy height (ACH), shoot fresh weight (SFW), shoot dry weight (SDW), ripe fruit yield (RYLD), total fruit yield (TYLD), harvest index (HI), carbon isotope discrimination ($\Delta^{13}C$). Environmental factors include year (2018 or 2019), full or reduced water treatments (full or red, respectively), and/or location (loc1 or loc2). Color represents whether the effect is horticulturally favorable (green) or unfavorable (red). Earlier maturity (negative effect on DAPR), larger fruit (positive effect on FW15), larger SDW, higher yield (positive effect on RYLD), higher HI (positive effect on HI), and lower $\Delta^{13}C$ (negative effect on $\Delta^{13}C$) were considered more horticulturally favorable. Uncolored cells have a GCA or SCA value of zero. The magnitude of the GCA and SCA effects was affected by transformations required to meet the assumption of homogeneity of variance (square root: FW15, ACW, SFW, SDW, RYLD, and TYLD; logit: HI).

Supplemental Table S2.6: Broad (BSH) and narrow (NSH) sense heritability values of all trait-environment combinations for genotypes in the Design II mating design combinations. The trait

acronyms are: days after planting to first green fruit (DAPG), days after planting to first ripe fruit (DAPR), weight of 15 fruit (FW15), average canopy width (ACW), average canopy height (ACH), shoot fresh weight (SFW), shoot dry weight (SDW), ripe fruit yield (RYLD), total fruit yield (TYLD), harvest index (HI), carbon isotope discrimination ($\Delta^{13}C$). The trait acronym is followed by the environmental factors used to subset the trait datasets (separated by underscores). Environmental factors include year (2018 or 2019) and/or full or reduced water treatments (full or red, respectively). Color of the BSH and NSH values represents the magnitude of those values (green is high, yellow is intermediate, and red is low). If the NSH/BSH ratio is equal to one, then the cell is uncolored, if the ratio is less than one, then the value is blue.

CHAPTER 3

WHOLE-GENOME SEQUENCE OF WILD TOMATO (*SOLANUM HABROCHAITES)* AND IDENTIFICATION OF CANDIDATE GENES FOR QTL ASSOCIATED WITH WATER STRESS TOLERANCE ON CHROMOSOME 9

Amy M. Groh, Tingting Zhu, Le Wang, Ming-Cheng Luo, Dina A. St. Clair

ABSTRACT

Solanum habrochaites, a wild tomato species, is a source of abiotic and biotic stress tolerances that can be used in breeding for improved cultivated tomato (*S. lycopersicum*). A *de novo* whole-genome sequence (WGS) of *S. habrochaites* accession LA1778 was generated using the PacBio RSII system and a Bionano optical map. Twelve pseudomolecules, corresponding to the 12 chromosomes of tomato, were constructed, with a total length of approximately 959 Mb. The *S. habrochaites* WGS had high gene synteny with the *S. lycopersicum* reference genome, although some inversions were detected. QTL for tolerance to rapid-onset water stress (*stm9*) and slow-onset water stress (carbon isotope discrimination, $\Delta^{13}C$) were mapped in prior studies to the short arm of chromosome 9 of *S. habrochaites* acc. LA1778. Here, the *S. habrochaites* chromosome 9 sequence was used to determine the physical locations of QTL for *stm9*, $\Delta^{13}C$, and other horticultural traits. A list of potential genes within each QTL interval was generated using *S. lycopersicum* genes that mapped to each interval, predicted genes, as well as mapped transcript reads from *S. habrochaites* acc. LA1777. It is likely that many of these genes can be affected by

water stress and some may play a role in conferring the tolerant phenotypes. A subset of genes of particular interest are discussed here. Several genes in each QTL interval remain unclassified or were present in only the *S. habrochaites* sequence. Further work is necessary to determine which genes confer the slow and rapid-onset water stress tolerant phenotypes of *S. habrochaites*.

INTRODUCTION

Cultivated tomato, *Solanum lycopersicum*, is a major vegetable crop globally. The global gross production value for tomatoes in 2020 was $103 billion USD (FAOSTAT, 2020). Tomato production will be essential to maintain for the growing global population due to its cultural and nutritional significance (Peralta & Spooner, 2007). Tomatoes are widely used in both processed and fresh forms around the world and are a good source of vitamins A, C, and carotenoids including lycopene, which are beneficial for human health (Labate et al., 2007; Peralta & Spooner, 2007).

Crop production, including tomato, faces challenges with climate change and increased pathogen pressures (Bebber et al., 2013; Pathak et al., 2018). One way to address production issues arising from abiotic and biotic stresses is through breeding for genetic tolerance to stresses. Genetic diversity in crop wild relatives provides essential source materials for introducing novel phenotypes to cultivated crops through breeding efforts (Bai & Lindhout, 2007; Dempewolf et al., 2017).

Wild tomato species have ample genetic diversity in comparison to cultivated tomato, which has relatively narrow genetic variation due to domestication bottlenecks (Miller & Tanksley, 1990; Rick, 1983). Self-incompatible wild species, which includes *S. habrochaites* and *S. pennellii*, are among the most diverse (Miller & Tanksley, 1990). *S. habrochaites* and *S. pennellii* are green-fruited wild species that are more distantly related to *S. lycopersicum* and other red-

fruited species. Fortunately, all wild tomato species can be crossed successfully with *S. lycopersicum* to produce viable progeny. Thus, wild species can be used as essential sources of horticulturally important traits and phenotypes in cultivated tomato breeding.

Wild tomato species contain genes controlling fruit quality, horticultural traits, biotic and abiotic stress tolerances, and other agriculturally valuable traits that can be introgressed into cultivated tomato (Labate et al., 2007). *S. habrochaites* contains QTL associated with fruit color, soluble solids (brix), and firmness (Bernacchi et al., 1998; Monforte et al., 2001). *S. habrochaites* is also a source of biotic stress tolerances including resistance to *Alternaria solani* (early blight), *Phytophthora infestans* (late blight), tomato mosaic virus (*Tm-1* resistance gene), and a variety of insect pests (largely due to trichome-mediated defenses) (Brouwer & St. Clair, 2004; Foolad et al., 2002; Kennedy, 2003; Ohmori et al., 1996). Loci associated with tolerance to cold stress, chilling stress, and water stress have also been discovered in *S. habrochaites* (Arms et al., 2015; Chen et al., 2015; Lounsbery et al., 2016). A close relative of *S. habrochaites, S. pennellii,* is a source of fruit quality traits such as fruit soluble solids, color, and phenolics (Fridman et al., 2004; Liu et al., 2003; Rousseaux et al., 2005).

Whole-genome sequence (WGS) of plant species can aid in a variety of genomic analyses that can be utilized in crop improvement. These analyses include identifying genome-wide genetic variation (SNPs, indels, or other polymorphisms) that can be used for determining marker-trait associations, identifying candidate genes for traits, and transcriptome analyses (Hu et al., 2018; Kim et al., 2020). There are publicly available genome assemblies for various tomato species including: *S. lycopersicum* cv. Heinz 1706, *S. pennellii* accessions LA716 and LA5240, *S. habrochaites* acc. LA1353, *S. lycopersicoides* acc. LA2951, and *S. pimpinellifolium* acc. LA1670

and LA2093 (Bolger et al., 2014; Powell et al., 2022; Schmidt et al., 2017; Seong et al., 2022; Takei et al., 2021; Tomato Genome Consortium, 2012; Wang et al., 2020).

A new draft genome for *S. habrochaites* acc. LA1353 was published by Seong et al. (2022), yet there is no publicly available WGS for *S. habrochaites* acc. LA1778. LA1353 was collected approximately 270 km from LA1778 in a different province in Peru (Tomato Genetics Resource Center, TGRC, https://tgrc.ucdavis.edu). The distance and mountainous terrain between the sites from which the accessions were collected from most likely severely limited or prevented direct gene flow via pollen transported by insects. Landis et al. (2021) showed that accessions geographically close to LA1353 (acc. LA1352 and LA1354) were not part of the same population cluster as LA1777 and LA1778 based on genome-wide RAD-seq markers. This suggests that these accessions are from genetically distinct populations. In contrast, *S. habrochaites* acc. LA1777 was collected at a site located approximately 5 km from LA1778 within in the same Peruvian mountain valley (TGRC, https://tgrc.ucdavis.edu). Given the proximity and pollinator activity, the populations of LA1778 and LA1777 likely exchanged genes with each other frequently via either native bees or via seed transport, and thus are closely related accessions. The available introgression line (IL) library (a collection of cultivated tomato inbred lines containing an introgression from a wild species) derived from LA1777 represents a large portion of the wild species genome and has been used widely for QTL mapping studies (Groh & St. Clair, 2022; Liu et al., 2012; Monforte & Tanksley, 2000; Thapa et al., 2015). LA1778 WGS will be useful for researchers studying LA1777 who would benefit from full genome assemblies.

Studies have been conducted to explore the genetic basis of abiotic stress tolerances in *S. habrochaites* acc. LA1778. LA1778 was collected at a high altitude (3002 m) site in Peru with low mean temperature (5.7 ºC), suggesting that LA1778 has adaptations to chilling temperatures

(https://tgrc.ucdavis.edu; Easlon et al., 2013). Truco et al. (2000) mapped QTL for shoot turgor maintenance under rapid-onset water stress caused by root chilling to chromosomes 5 and 9 of *S. habrochaites* LA1778 in an interspecific back cross population. The chromosome 9 QTL (designated *stm9*) was fine mapped by Arms et al. (2015) to a 0.32cM region using a set of recombinant sub-near-isogenic lines (sub-NILs). The ability of *S. habrochaites* to maintain shoot turgor under root chilling stress was associated with reduced stomatal conductance, suggesting that it may also affect slow-onset water stress due to deficit irrigation (Bloom et al., 2004). In subsequent studies, chromosome 9 sub-NILs derived from LA1778 were tolerant to slow-onset water stress in the field caused by a 70% reduction in normal irrigation (Groh et al., 2022; Lounsbery et al., 2016). QTL for carbon isotope discrimination ($\Delta^{13}C$), which is a water use efficiency (WUE) related trait, and other horticultural traits were mapped to a 3.6 cM region of *S. habrochaites* chromosome 9 with sub-NILs (Groh et al., 2022). The QTL associated with $\Delta^{13}C$ did not coincide with the *stm9* locus but was located nearby.

QTL for *stm9* and $\Delta^{13}C$ on chromosome 9 of LA1778 are tightly linked to or are coincident with QTL for undesirable horticultural traits (Groh et al., 2022; Lounsbery et al., 2016). Negative linkage drag has been observed when using wild species in breeding to improve cultivated tomato (Brouwer & St. Clair, 2004; Haggard et al., 2013). Linkage drag may be overcome if the QTL for the traits of interest are not pleiotropic and the repulsion phase linkage among the loci controlling the traits can be broken through recombination. We can use the QTL marker intervals for each trait, information from a *S. habrochaites* LA1778 WGS assembly, and available transcript data to predict candidate genes for these traits. Candidate genes that are associated with a change in trait phenotype can serve as specific targets for introgression in a marker-assisted breeding program.

This approach can allow breeders to introgress specific alleles of each gene into elite germplasm while potentially avoiding linkage drag from alleles with negative horticultural effects.

This study had two main objectives. First, to create a high-quality WGS assembly for *S. habrochaites* LA1778. Second, to identify potential candidate genes associated with chromosome 9 QTL intervals for *stm9*, $\Delta^{13}C$, and horticultural traits described by Arms et al. (2015) and Groh et al. (2022).

MATERIALS AND METHODS

Plant material for whole-genome sequencing

Seed from *S. habrochaites* acc. LA1778 was originally obtained from the TGRC at UC Davis (https://tgrc.ucdavis.edu). A single plant of LA1778 was selected (designated HS34) and clonally propagated via vegetative tip cuttings in greenhouses at the UC Davis Plant Sciences Greenhouse Facility in Davis, CA (Arms et al., 2015; Truco et al., 2000). HS34 plants were maintained in 3 L pots using standard horticultural greenhouse procedures for tomato.

Whole-genome sequencing and assembly using Single Molecule Real-Time (SMRT) technology

Young leaf tissue from *S. habrochaites* acc. LA1778-HS34 plants was collected and high quality, high molecular weight DNA was extracted as described in Dvorak et al. (1988). The sequencing libraries were prepared using a standard protocol from Pacific Biosciences (PacBio, Menlo Park, CA, USA). Libraries were sequenced at the UC Davis DNA Technologies Core on the PacBio long read sequencer (model RSII) with P6 polymerase binding and C4 chemistry kits. The reads were assembled into contigs using Canu v1.4 (Koren et al., 2017). The tool Quiver was

used to generate the consensus sequences of the genome based on assembled contigs (Chin et al., 2013).

Bionano optical map construction

Bionano optical map construction to assist scaffolding was conducted as described in Zhu et al. (2018). A brief overview, with the differences in protocol noted, is described here. Young leaves of *S. habrochaites* LA1778-HS34 were collected and ultra high-molecular-weight (uHMW) DNA was isolated by Amplicon Express (Pullman, WA, USA). The uHMW DNA was labeled at specific sequence motifs using the nicking endonuclease Nt.BspQI (New England BioLabs, Ipswich, MA) chosen according to the reference *S. lycopersicum* genome SL3.0 (Tomato Genome Consortium, 2012; the Sol Genomics Network, SGN). The nicked DNA molecules were stained as described in detail in Luo et al. (2017). The Irys instrument and Irys Chip (Bionano Genomics, San Diego, CA) were used to image the stained DNA sample. Molecules that were greater than 180 kb and passed the quality control parameters were aligned, clustered, and assembled into a consensus map using the Bionano Genomics assembly pipeline as described by (Lam et al., 2012).

Scaffolding using Bionano optical maps

Scaffolding was conducted similarly to Zhu et al. (2018). Sequence contigs were digested in silico according to the Nt.BspQI restriction sites by using the Knickers software (Bionano Genomics). To check conflicts between sequence assembly and the optical map, the digested sequences were aligned to the optical map using the software RefAligner (Bionano Genomics). Visualization of the alignments were performed in IrysView (Bionano Genomics). The contigs that were not consistent with the optical map were resolved accordingly. Scaffolding of the refined contigs by the guidance of the optical map was performed using the Stitch algorithm (Shelton et al., 2015). The filtering parameters of Stitch were trained to be suitable for the *S. habrochaites*

genome and Stitch was performed in iterations until no additional super-scaffolds were produced. After each iteration of Stitch, the potentially chimeric scaffolds flagged by the program were manually checked and resolved if necessary. Software packages used in these scaffolding steps can be obtained from Bionano Genomics (https://bionanogenomics.com/support/software-downloads/).

Pseudomolecule construction

The scaffolds were checked for chimeras using various linkage maps in SGN (https://solgenomics.net) and resolved accordingly (Fernandez-Pozo et al., 2015). Gaps between neighboring scaffolds were filled with 1,000 N's, and each pseudomolecule started at the telomeric end of the short chromosome arm.

Gene synteny analysis and whole-genome alignment

Coding sequences (CDS) of 35,768 genes from ITAG3.2 of *S. lycopersicum* were mapped to the 12 pseudomolecules of the wild tomato genome using BLAT with default parameters (Kent, 2002). Using an in-house script, the best hit of each CDS sequence on the 12 pseudomolecules was kept for further analysis based on the identity and the coverage of CDS sequence. According to the location of each gene on the wild tomato genome, the synteny analysis between the genomes of *S. habrochaites* and cultivated tomato was performed using MCScanX with default parameters (Wang et al., 2012).

The WGS of *S. habrochaites* and cultivated tomato (SL3.0) were aligned using minimap2 (Li, 2018). The minimum length of alignment was set to 10 kb. The dot plot was plotted using an R script.

Identifying annotated and predicted genes within the QTL intervals and confirming with transcript data

The physical location of QTL *stm9* was determined with the gene used to anchor marker H358 from Arms et al. (2015), Solyc09g008310, and the marker T1673 (which overlaps with the marker set used in Groh et al., 2022) (Figure 3.1). Markers from Groh et al. (2022) and the sequence for Solyc09g008310 were BLASTed against the *S. lycopersicum*, *S. habrochaites*, and *S. pennellii* chromosome 9 sequences to obtain physical positions of the SNPs in each genome assembly (Zhang et al., 2000). *S. pennellii* gene annotation for these regions were from Bolger et al. (2014) (accessed on SGN, solgenomics.net). Most trait QTL for yield, maturity, and WUE were located within the marker interval H307-TG10, a genetic distance of 1.25 cM (subsequently referred to as the highly coincident QTL interval). The $\Delta^{13}C$ QTL interval ranged from At2g36480 to TG10, a genetic distance of 0.95 cM.

A list of genes located within the *stm9* region, and within the highly coincident (markers H307 and TG10) QTL intervals was identified. This list was extracted from the genes that mapped to the *S. habrochaites* genome from the ITAG3.2 *S. lycopersicum* annotation (as described in the subsection Gene synteny analysis and whole-genome alignment). Gene function descriptions were sourced from the SGN FTP site. Unknown or predicted genes were annotated using the National Center for Biotechnology Information (NCBI) webtool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to identify their potential functions, if available (Zhang et al., 2000).

To attempt to determine whether *S. habrochaites* contains novel genes within the QTL intervals on chromosome 9 that are not present in the *S. lycopersicum* genome, genes were predicted in these regions and confirmed with publicly available transcript data, as described below. The *S. habrochaites* chromosome 9 sequence data was uploaded to the Galaxy web

platform (https://usegalaxy.org) (Afgan et al., 2018). Augustus (v3.4.0) was used to predict genes within the QTL intervals using the default settings and model organism *S. lycopersicum* (Stanke et al., 2008).

RNAseq data for *S. habrochaites* LA1777 seedlings from Koenig et al. (2013) was used to confirm predicted genes and to detect other novel genes not described by the initial ITAG 3.2 mapping or the gene prediction. LA1777 is an accession that is closely related to LA1778 due to the close geographical proximity of their sites of origin and high rate of cross-pollination in the wild (TGRC). Paired end reads for transcripts (Study Accession: PRJNA192978; Run Accessions: SRR786504, SRR786505, and SRR786506) from European Nucleotide Archive (ENA, https://www.ebi.ac.uk/ena/browser/home) were loaded into the Galaxy web platform and mapped to *S. habrochaites* chromosome 9 using TopHat (v2.1.1) (Kim et al., 2013). The parameters used in TopHat were the same as described in Koenig et al. (2013): -g 1 --segment-mismatches 1 -a 8 -m 1 -i 14 -I 10000. The resulting BAM files were merged using MergeSamFiles and downloaded locally (Broad Institute, n.d.). The mapped reads and *S. habrochaites* chromosome 9 sequence were loaded into Integrative Genomics Viewer (IGV v2.13.2) for manual comparison of the mapped transcript reads to the mapped ITAG3.2 reads and the predicted genes (Thorvaldsdóttir et al., 2013).

Predicted genes that were not supported by mapped transcripts were not considered further. Predicted genes and mapped transcripts that did not collocate with *S. lycopersicum* ITAG3.2 mapped genes (i.e., genes unique to *S. habrochaites*) were blasted against the NCBI database to determine potential gene function and the *S. lycopersicum* genome (SL4.0) to determine if they had strong matches in the cultivated tomato genome. A list of proposed genes for the region was composed by combining the list of ITAG3.2 mapped genes that did not overlap, predicted genes

that were confirmed by transcript data, and mapped transcripts that did not overlap with any other proposed genes.

Identifying candidate genes for *stm9*, Δ13C, and other trait QTL on chromosome 9

Specific genes and transcription factors identified in these regions that were found to be associated with water stress tolerance in the literature were compared between the *S. habrochaites* genome and the *S. lycopersicum* genome. The CDS of one candidate gene was aligned in MEGA X (v10.2.6) using the Muscle algorithm with default settings and compared visually (Kumar et al., 2018). The translated amino acid sequences were also compared for differences in amino acids and stop codons.

RESULTS AND DISCUSSION

S. habrochaites de novo whole-genome assembly and optical map assisted scaffolding

The *S. habrochaites* LA1778-HS34 genome was sequenced using 121 PacBio SMRT Cells, generating 120.4 Gb (~80x genome coverage) of raw sequence data (Figure 3.2). The average insert read length of the subreads was 11,107 bp. A total of 1,099,694,196 bp sequences in 4,182 contigs were generated (Table 3.1, TM_CTG_V0). The contig N50 size of the assembly was 1.10 Mb, with the longest contig of 14.46 Mb. The genomic consensus tool, Quiver, was used to generate the consensus sequences of the genome, which resulted in the final assembly of 1,094,915,996 bp in 4,097 contigs. The contig N50 size slightly increased to 1.12 Mb and the largest contig size increased slightly to 14.52 Mb (Table 3.1, TM_CTG_V0.1).

A total of 161 Gb of data containing 515,098 raw molecules were generated and were used to *de novo* assemble an optical map (Table 3.2). These raw molecules were assembled into 995 contigs, with an N50 size of 1.53 Mb and a total length of 1.05 Gb (Table 3.2). The *S. habrochaites*

sequence assembly was validated by aligning and comparing it to the optical map. Of the 4,097 contigs in the sequence assembly, 796 (~844Mb) had sufficient restriction sites to permit validation by the optical map. Of these, 39 contigs showed conflicts and were disjoined accordingly, which increased the total number of contigs to 4,136 (Table 3.1, TM_CTG_V1).

Scaffolds were generated by using the Stitch algorithm with the optical map as a guide. This reduced the number of scaffolds to 3,695 (Table 3.1, TM_SCF). The total length of these scaffolds was 1,120,945,628 bp, with N50 of 2,762,933 bp and 2.33% N's. The *S. habrochaites* genome is approximately 1.5x the size of the *S. lycopersicum* genome (~950 Mb) as determined by flow cytometry (Arms et al., 2015). The difference between the approximated genome size (~1.43 Gb) and our sequence assembly (1.12 Gb) is likely due to the difficulty of sequencing and assembling long repeat regions near the centromeres.

Various tomato linkage maps in SGN were used to validate the final scaffolds (TM_SCF) and anchor scaffolds onto chromosomes. A total of 1,035 scaffolds containing two or more annotated genes were anchored onto 12 chromosomes (Table 3.3). These scaffolds were linked with 1000 N's as a spacer to generate pseudomolecules. The total length of the pseudomolecules was 959,003,884 bp, which accounts for 85.63% of the genome assembly.

Gene synteny analysis

We attempted to map 35,768 genes from the *S. lycopersicum* annotation (ITAG3.2) to the 12 pseudomolecules of the *S. habrochaites* WGS. Of these, 34,839 genes were located on the 12 pseudomolecules. The synteny analysis was performed using the 34,839 genes mapped to the *S. habrochaites* genome and the 35,768 genes from *S. lycopersicum*. A high consistency was detected in synteny dot plots as well as for the whole-genome alignment between the *S. habrochaites* genome and the cultivated tomato genome (Figures 3.3 and 3.4). However, some inconsistencies

were detected in certain regions, especially around centromeres. The sum of 349 syntenic blocks (defined as ≥ five genes) were detected between the *S. habrochaites* genome and *S. lycopersicum* genome (Figure 3.3). Of these 349 syntenic blocks, the orientations of 85 were reversed on the homologous chromosome pairs between the wild and cultivated tomato. These inversions were also detected in the whole-genome alignment of these two genomes (Figure 3.4). The inversions could have occurred during the phylogenetic history of tomato speciation or during domestication and crop improvement (Huang & Rieseberg, 2020; Wang et al., 2020). A recent *S. pimpinellifolium* genome analysis identified 28 inversions containing numerous genes related to biotic and abiotic stress tolerances (Wang et al., 2020). Alternatively, the inversions observed in this and other studies of wild tomato WGS assemblies could be due to assembly error in either the cultivated tomato or *S. habrochaites* genome. Further analysis would be needed to determine if these inversions are due to sequencing errors or are real, and if they are biologically relevant.

Chromosome 9 assembly and physical marker location

The length of *S. habrochaites* chromosome 9 is 87,670,881 bp, including 2,130,549 bp (2.43%) of N's (Table 3.3). A total of 2,441 genes from the *S. lycopersicum* annotation ITAG3.2 were mapped onto the chromosome. The *stm9* QTL interval, linkage map for the full 3.59 cM sub-NIL interval (T1673 ─ T0641), and the $\Delta^{13}C$ QTL interval from Groh et al. (2022) were compared to the marker's physical positions on three *Solanum* species genomes: *S. lycopersicum* (SL4.0), *S. habrochaites* LA1778-HS34 (newly assembled), and *S. pennellii* LA716 (whole-genome shotgun sequence from SGN BLAST; Bolger et al., 2014). Almost all the markers were identified in all three genomes, except markers H14 and T1673 (Table 3.4). Marker H14 could not be located by BLASTing the marker sequence to the *S. habrochaites* assembly. Marker T1673 was not found in the expected syntenic location in the *S. pennellii* genome, although a similar match was found

much closer to the telomere of chromosome 9. All other markers were in similar physical genome locations in each species. The physical locations of markers CT283 and SSR73 were not in the same relative order as the linkage map from Groh et al. (2022). This difference in order could be due to genotyping errors in the population used to make the linkage map, experimental error due to population size, or sequence assembly errors.

The QTL interval for *stm9* described by Arms et al. (2015) ranges from the annotated gene Solyc09g008310 to marker T1673, which is approximately 322 kb (1,776,418 bp ─ 2,098,817 bp) in the *S. lycopersicum* (SL4.0) genome and about 380 kb (1,853,030 bp – 2,233,206 bp) in the *S. habrochaites* genome. The entire chromosome 9 region examined in Groh et al. (2022) represented 3.59 cM, which corresponds to approximately 1.76 Mb in the *S. lycopersicum* genome (SL4.0) and 1.86 Mb in the *S. habrochaites* assembly (Table 3.5). Based on Groh et al. (2022), the marker interval H307 to TG10 contains the largest number of coincident QTL. Almost all of the QTL mapped in Groh et al. (2022) fully or partially overlap with this interval. This highly coincident QTL interval spans 1.26 cM, which corresponds to approximately 579 kb in the *S. lycopersicum* genome (SL4.0), 661 kb in the *S. habrochaites* assembly, and about 672 kb in the *S. pennellii* genome (SGN; Bolger et al., 2014) (Table 3.5). The physical locations of all three QTL intervals (*stm9*, $\Delta^{13}C$, and the highly coincident QTL interval) are represented in Figure 3.1.

It is not surprising that the QTL intervals are slightly larger in the wild species genomes. The *S. habrochaites* and *S. pennellii* genomes have been estimated to be 1.43 Gb (about 1.5x the cultivated tomato genome; Arms et al., 2015) and 1.2 Gb (about 1.3x the cultivated tomato genome; Bolger et al., 2014), respectively, while the *S. lycopersicum* genome is about 950 Mb. The larger genome size of the wild species is likely due to long regions of repetitive DNA that may be shorter or missing in the cultivated tomato genome. Interestingly, the differences in the

physical length of the QTL interval is not proportional to the size differences between the genomes (Table 3.5). This could be due to repeats localizing closer to the centromere rather than being spread throughout the genome.

Chromosome 9 *stm9* QTL interval and candidate genes

The *stm9* QTL interval of *S. habrochaites* contains 38 proposed genes. Of these, 25 mapped to the *S. habrochaites* genome from the ITAG3.2 annotation, nine were predicted genes, and four were mapped LA1777 transcripts (Supplemental Table S3.1) (Koenig et al., 2013). Any of these genes could be involved in the *stm9* phenotype. A plant's molecular response to abiotic stress can be complex, can involve many genes, and can depend on other environmental conditions (Des Marais & Juenger, 2010). A few of the most interesting genes that may be associated with variation in the *stm9* phenotype (i.e., candidate genes) are discussed here.

Several lines of evidence point to a xyloglucan endotransglucosylase/hydrolase gene (XTH, Solyc09g008320) as a potential candidate for *stm9*. XTHs are involved in cell wall processes and play a role in tolerances to freezing and salt stress in *Arabidopsis thaliana* (Ishida & Yokoyama, 2022). Previous studies showed that homologous XTH genes in pepper, another Solanaceous crop, have an effect on tolerance to drought, salt, and cold (Cho et al., 2006; Choi et al., 2011). Transgenic tomato plants containing a XTH gene from pepper (*CaXTH3*) exhibited improved tolerance to water and salt stress due to a reduction in stomatal aperture in the transgenic plants (Choi et al., 2011). Previous research indicated that stomatal control is involved in the *stm9* phenotype in *S. habrochaites*, suggesting that XTH could be involved in the response to rapid-onset water stress caused by root chilling (Bloom et al., 2004).

Transcriptome data from previous studies also supports the hypothesis that XTH is responsible for phenotypic differences observed under rapid-onset water stress due to root chilling.

Arms et al. (2017) conducted a transcriptome study using two NILs derived from *S. habrochaites* (LA1778): one contained the *S. habrochaites* allele (root chilling tolerant) at the *stm9* locus and the other contained the *S. lycopersicum* allele (susceptible). XTH was differentially expressed between susceptible and tolerant NILs at all time points before and after root chilling. The tolerant NIL also exhibited differential expression for XTH between the prechilling timepoints and two timepoints after chilling, and had higher transcript levels of XTH than the susceptible NIL. Interestingly, other XTHs were also shown to be differentially expressed between *S. pennellii* ILs (including IL9-1, which contains a chromosome 9 introgression) and a cultivated tomato control under water stress (Gong et al., 2010).

XTH CDS sequences were aligned and compared among all three tomato species, and nonsynonymous and nonsense mutations were present in the *S. habrochaites* and *S. pennellii* sequences (Figure 3.5). However, *S. habrochaites* and *S. pennellii* did not have the same sequence variations. Easlon et al. (2013) observed that *S. pennellii* was more susceptible to root chilling than *S. habrochaites* in a phenotypic evaluation of tomato accessions including wild species. This suggests that the gene or genes causing the resistant *stm9* phenotype in *S. habrochaites* likely has a different allele in *S. pennellii,* causing it to be susceptible. XTH has different CDS sequence variations between *S. habrochaites* and *S. pennellii*, which further supports that it may be associated with the *stm9* phenotype (Figure 3.5).

Glycosyltransferase genes can also be involved in abiotic stress tolerance. Arms et al. (2017) reported that glycosyltransferase (Solyc09g008510) had significantly greater expression in the chilling tolerant tomato NIL than the susceptible NIL at all timepoints both before and after root chilling. This observation is consistent with previous studies indicating that upregulation of glycosyltransferases in rice increases tolerance to water and chilling stress (Liu et al., 2021; Shi et

al., 2020). In rice, the upregulation of a glycosyltransferase gene increased water stress tolerance, reduced stomatal aperture, and affected the expression of many other stress response genes (Liu et al., 2021). It is possible that the glycosyltransferase gene in the *stm9* interval of *S. habrochaites* may be involved in activating stress response genes in other parts of the genome when induced by root chilling.

R2R3MYB transcription factor 22 (Solyc09g008390) could also be a cause of variation of the *stm9* phenotype. MYB transcription factors are known to be involved in tolerance to salt, cold, and/or water stress in other plants, including *Arabidopsis* and rice (Liang et al., 2005; Yang et al., 2012). In *Arabidopsis*, expression of the gene *AtMYB61,* part of the R2R3MYB transcription factor family, has been associated with changes in stomatal aperture, which is relevant to the *stm9* phenotype, as mentioned previously (Bloom et al., 2004; Liang et al., 2005). However, this gene was not differentially expressed under the conditions of the root chilling experiment in Arms et al. (2017). Additionally, none of the LA1777 transcript reads mapped to the location of this gene, suggesting that it was not being transcribed in the seedlings (Koenig et al., 2013). However, it is possible that this gene would be transcribed under different conditions such as root chilling or water stress, at other plant developmental stages, and/or in other plant tissues.

Chromosome 9 Δ13C QTL interval and associated genes

The marker interval containing the Δ^{13}C QTL (At2g36480 ─ TG10) ranges from about 2497345 ─ 2926980 bp in *S. lycopersicum* and 2719281 ─ 3182960 bp in the *S. habrochaites* genome (Table 3.4) (Groh et al., 2022). We found 50 proposed genes in this interval of *S. habrochaites*. These genes include 33 genes mapped from ITAG3.2, 15 predicted genes that were confirmed by transcript reads, and two regions with mapped transcripts that were not explained by any other proposed genes (Supplemental Table S3.1). Of these 50 genes, many are potentially

associated with water stress tolerance or other changes in phenotype caused by water stress. A few particularly interesting candidates for WUE and $\Delta^{13}C$ are discussed here.

Fructose-1,6-bisphosphate aldolase (FBA, Solyc09g009260 and Solyc09g009280) has been associated with water and temperature stress in previous studies of plants, including tomato and *Arabidopsis* (Cai et al., 2016; Lu et al., 2012). In tomato, FBA has been shown to respond to high and low temperature stress in a study by Cai et al., (2016). Furthermore, an increase in expression of Solyc09g009260 (*SlFBA7*) increased activity of Calvin cycle enzymes and the net photosynthetic rate. Genes that effect $\Delta^{13}C$ measurements are most likely related to stomatal opening and closing or photosynthesis, so *SlFBA7*'s relationship to the Calvin cycle suggests it is a potential candidate. Gong et al. (2010) also found that FBA genes are downregulated in IL2-5 (an *S. pennellii* chromosome 2 IL) under water stressed conditions.

Growth-regulating factor 3 (GRF3, Solyc09g009200) may be a candidate gene for $\Delta^{13}C$ or another trait QTL that is coincident with this interval. GRFs are a family of transcription factors that regulate growth in plants, especially in new tissues such as developing leaves, flowers, and roots (Omidbakhshfard et al., 2015). In *Arabidopsis*, a GRF (*AtGRF7*) suppresses stress-responsive genes that also slow plant growth, such as the transcription factor dehydration-responsive element binding protein 2A (DREB2A) (Kim et al., 2012; Omidbakhshfard et al., 2015). GRF3 in tomato may also suppress stress-responsive genes. If this is true, we hypothesize that the *S. habrochaites* GRF3 allele may be regulated differently than the *S. lycopersicum* allele. If the transcription of GRF3 in *S. habrochaites* is reduced more quickly under water stress, it could lead to higher transcription of downstream genes that help the plant cope with water stress and lead to the greater WUE seen in Groh et al. (2022). GRF3 may have downstream effects causing

both the positive WUE QTL and the negative horticultural QTL (this will be discussed further in the following subsection). However, future work would be needed to test this hypothesis.

Highly coincident QTL interval

Next to the $\Delta^{13}C$ QTL, but within the highly coincident QTL interval (marker interval H307 ─ TG10, see Figure 3.1 and Supplemental Table S3.1), is a heat stress transcription factor A3 (HSFA3, Solyc09g009100). HSFs regulate abiotic stress-responsive genes in plants, including heat shock proteins (HSP) (Guo et al., 2016). HSFA3 is one of 24 HSF genes in tomato and has been shown to be responsive to water stress in *Arabidopsis* (Guo et al., 2016; Sakuma et al., 2006). Transcription of HSFs in *S. pennellii* ILs are also affected by water stress (Gong et al., 2010). HSFs have been studied mostly under heat stress conditions, yet crops in the field often experience more than one type of abiotic stress. The chromosome 9 sub-NILs used for QTL mapping by Groh et al. (2022) were grown in the field in Davis, CA which has a Mediterranean climate with hot, dry summers. HSFA3 could be involved in responding to more than one environmental stress experienced by the sub-NILs, causing phenotypic effects on traits such as maturity and/or yield.

HSFA3 may be a candidate gene for maturity QTL that are associated with this region. Li et al. (2013) performed a variety of experiments using transformed *Arabidopsis* plants that overexpressed the HSFA3 gene from *S. lycopersicum* (*SlHSFA3*). The authors reported that the transformed lines had delayed flowering under heat stress. Therefore, *SlHSFA3* can affect time to flowering, which subsequently affects time to development of green fruit and ripe fruit. It is possible that the *S. habrochaites* allele for HSFA3 has a stronger effect on time to flowering and fruit development under abiotic stress than the *S. lycopersicum* allele, resulting in the negative association with maturity-related traits in this QTL interval (Groh at al., 2022).

Interestingly, studies in *Arabidopsis* showed HSFA3 expression is dependent on DREB2A under water and heat stress (Sakuma et al., 2006; Schramm et al., 2008). It is possible that GRF3 (described above in the subsection discussing the $\Delta^{13}C$ QTL) is affecting the expression of HSFA3 via regulation of DREB2A. The regulation of DREB2A and the response of HSFA3 to DREB2A may depend on whether a genotype contains the *S. habrochaites* or *S. lycopersicum* allele for the HSFA3 and GRF3 genes. It is possible there could be a complex relationship among the alleles at genes in this region, depending on which alleles are present and which environmental stressors are acting on the genotypes.

Gibberellin 20-oxidase (GA20ox, Solyc09g009110) is a gene involved in the biosynthesis of gibberellin (GA), a hormone that affects plant growth. It may be a candidate gene for the QTL for horticultural traits in this region. GA is important for promoting expansion of plant organs as well as promoting vegetative and reproductive development (Yamaguchi, 2008). Bioactive GA levels can be reduced by water stress in tomato, which can lead to increased water stress tolerance and reduced shoot growth (Nir et al., 2014).

Reduction in bioactive GA under water stress is partially due to reduced expression of GA20ox in tomato plants (Shohat et al., 2021). Shohat et al. (2021) showed that plants with loss-of-function mutants of two GA20ox that were exposed to water stress maintained their relative water content longer than wild type plants, but they also had smaller stems and leaves. The greater water stress tolerance seen by plants with lower GA levels is due to reduced water loss via smaller plant size or reduced stomatal aperture (Shohat et al., 2021). Shohat et al. (2021) found that reduced expression of GA20ox did not affect the transpiration rate of plants, suggesting that the reduced water loss was caused by smaller leaf areas rather than effects on stomatal aperture. Other GA deactivating genes are likely responsible for the decreased levels of GA in guard cells and

subsequent closing of stomata (Shohat et al., 2021). The GA20ox found in the highly coincident QTL interval may also be down regulated by water stress, leading to lower levels of GA. However, Groh et al. (2022) observed that plants with the *S. habrochaites* alleles in this QTL interval tended to have larger shoot dry weights than plants with the *S. lycopersicum* allele. If the *S. habrochaites* allele at GA20ox is affecting trait phenotypes that were mapped to this region, it's effects may be different than those observed in the tomato studies described above. For example, if the transcription of GA20ox was reduced in reproductive tissues, but not leaf tissues, it could reduce the development of fruit (causing reduced yields and later maturity) rather than reduced plant growth.

Proposed genes unique to *S. habrochaites*

Some of the genes identified in the *stm9*, $\Delta^{13}C$, and highly coincident QTL marker intervals on chromosome 9 were only found in the *S. habrochaites* genome sequence. These genes did not overlap with genes mapped to the *S. habrochaites* genome from the ITAG3.2 annotation of *S. lycopersicum*. These proposed distinctive *S. habrochaites* genes were predicted by the Augustus software, or were transcripts from LA1777 (Koenig et al., 2013) that mapped to unique positions within the three QTL intervals of interest on chromosome 9. A tomato pan-genome consisting of sequence data from many accessions of the red-fruited species (*S. lycopersicum*, *S. pimpinellifolium, S. cheesmaniae,* and *S. galapagense*) showed that the reference annotation for *S. lycopersicum* was missing 4,873 genes that were present in at least one of the pan-genome accessions (Gao et al., 2019). Thus, it's likely that *S. habrochaites* contains genes that aren't present in the reference *S. lycopersicum* annotation.

The 36 *S. habrochaites*-specific sequences were BLASTed in NCBI, and only a few had predicted functions, with the remainder uncharacterized (Supplementary Table S3.1).

Interestingly, many of these predicted functions were the same as annotated genes at other loci nearby, although they may have only had partial matches to these predicted functions. For example, in the *stm9* QTL interval there is a mapped transcript (2,170,267 ─ 2,172,043 bp) that is predicted to code for pectate lyase 12 based on a BLAST result. The annotated gene pectate lyase (Solyc09g008380) is also located within this QTL interval (1,942,580 ─ 1,947,340 bp). It is unknown whether some of these predicted genes and transcripts are duplicated genes, genome assembly errors, or a result of bias from mapping the RNAseq reads to chromosome 9 only. However, some of these potential genes in the *S. habrochaites* genome may be candidates for *stm9*, $\Delta^{13}C$, and other horticultural traits mapped to chromosome 9. Transcription data for *S. habrochaites* LA1778 under water stress and functional analysis of these proposed genes would help determine if they are involved in trait phenotypes of interest.

Potential future work

Knowing precisely which gene or genes are causing a trait phenotype can be useful for breeding purposes but it is not essential in an introgression-based breeding strategy. Having markers with tight linkage to a region controlling a trait of interest can be sufficient for employing marker assisted breeding if the markers are not also linked to loci with negative effects on horticultural traits. Trait QTL from wild species can be closely linked to loci conferring negative horticultural traits (i.e., linkage drag), which was observed in Groh et al. (2022) and other QTL studies utilizing wild tomato species (Brouwer & St. Clair, 2004; Haggard et al., 2013; Prohens et al., 2017). To integrate the QTL for beneficial traits into cultivated varieties, the linkages between the positive and negative effect loci must be broken to recover desirable recombinants. Alternatively, a transgenic approach such as cis-genesis can be used to integrate desirable genes into cultivated varieties (Prohens et al., 2017). Thus, methods that help reduce the number of

potential candidate genes for $\Delta^{13}C$ and other horticultural traits can facilitate the transfer of genes controlling desirable traits from wild species into cultivated varieties.

Future work involving these QTL under full and reduced water treatments could provide insights into which of the proposed genes are causal for *stm9*, $\Delta^{13}C$, and the other horticultural traits QTL mapped by Arms et al. (2015) and Groh et al. (2022). Although these QTL have been fine-mapped on the short arm of chromosome 9 of *S. habrochaites* using sub-NILs, there are still tens of genes located within these QTL intervals that are potential candidate genes. The QTL intervals could be further refined and/or coincident QTL could be fractionated using high-resolution mapping. QTL high-resolution mapping involves obtaining additional recombinants, creating a high-density genetic map, phenotyping the recombinants, using the data to narrow QTL intervals and facilitate identification of putative causal genes. High-resolution mapping has assisted identification of candidate genes for QTL controlling important traits in crops including tomato and wheat (Gonda et al., 2019; Kuzay et al., 2019).

Transcriptome data can provide evidence to support potential candidate genes for $\Delta^{13}C$ and other horticultural traits (Nguyen et al., 2019; Salvi & Tuberosa, 2005). RNA-seq data could be collected from water stress-susceptible (*S. lycopersicum* alleles) and tolerant (*S. habrochaites* alleles) tomato NILs under full and reduced water treatments in the field (similar to the genotypes used by Arms et al., 2017). Genes within the introgressions that are differentially expressed between the two NILs under different water treatments could be candidate genes, although more work would need to be done to verify candidates found with this method. Candidate genes for trait QTL may be validated using a variety of methods including using genetic transformation to knock-out, knock-down, or overexpress a candidate gene (Jiang et al., 2012; Pflieger et al., 2001; Salvi & Tuberosa, 2005). However, these methods might not definitively determine if a gene of interest

is the causal gene in the case of complex traits (Flint & Mott, 2001; Pflieger et al., 2001). Gene sequence or differential expression of transcripts are not the only potential differences between the species under water stress. Alternative splicing can be induced under abiotic stress and could be another path for further investigation in these genes and their transcripts (Chen et al., 2015).

CONCLUSION

The high-quality *S. habrochaites* WGS will be a useful tool for scientists and breeders interested in utilizing *S. habrochaites* acc. LA1778 and the LA1777 IL library. The genome can be used for comparative genomics with other *Solanaceous* species and variant calling for trait mapping studies using *S. habrochaites* introgressions. Using the WGS in combination with gene predicting software and transcript data identified potential candidate genes in the QTL intervals for *stm9*, $\Delta^{13}C$, and other horticultural traits. More work will need to be done to determine which genes may be causing changes in phenotypes under water stress. Precise locations of QTL are important for using wild species alleles for breeding to avoid linkage drag for important horticultural traits.

ACKNOWLEDGMENTS

This research was supported by the California Tomato Research Institute. We thank the UC Davis Plant Sciences greenhouse facility crew for assistance with maintaining plant materials in the greenhouse.

REFERENCES

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V.,

Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.*, *46*(W1), W537–W544. https://doi.org/10.1093/nar/gky379

Arms, E. M., Bloom, A. J., & St. Clair, D. A. (2015). High-resolution mapping of a major effect QTL from wild tomato *Solanum habrochaites* that influences water relations under root chilling. *Theoretical and Applied Genetics*, *128*(9), 1713–1724. https://doi.org/10.1007/s00122-015-2540-y

Arms, E. M., Yan, Z., & St. Clair, D. A. (2017). Differential transcriptional regulation in roots of tomato near-isogenic lines in response to rapid-onset water stress. *Frontiers in Plant Science*, *8*, 166.

Bai, Y., & Lindhout, P. (2007). Domestication and breeding of tomatoes: What have we gained and what can we gain in the future? *Annals of Botany*, *100*(5), 1085–1094. https://doi.org/10.1093/aob/mcm150

Bebber, D. P., Ramotowski, M. A. T., & Gurr, S. J. (2013). Crop pests and pathogens move polewards in a warming world. *Nature Climate Change*, *3*(11), 985–988. https://doi.org/10.1038/nclimate1990

Bernacchi, D., Beck-Bunn, T., Eshed, Y., Lopez, J., Petiard, V., Uhlig, J., Zamir, D., & Tanksley, S. (1998). Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from *Lycopersicon hirsutum*. *Theoretical and Applied Genetics*, *97*(3), 381–397.

Bloom, A. J., Zwieniecki, M. A., Passioura, J. B., Randall, L. B., Holbrook, N. M., & St. Clair, D. A. (2004). Water relations under root chilling in a sensitive and tolerant tomato species. In *Plant, Cell and Environment* (Vol. 27, pp. 971–979).

Bolger, A., Scossa, F., Bolger, M. E., Lanz, C., Maumus, F., Tohge, T., Quesneville, H., Alseekh, S., Sørensen, I., Lichtenstein, G., Fich, E. A., Conte, M., Keller, H., Schneeberger, K., Schwacke, R., Ofner, I., Vrebalov, J., Xu, Y., Osorio, S., … Fernie, A. R. (2014). The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nature Genetics*, *46*(9), 1034–1038. https://doi.org/10.1038/ng.3046

Broad Institute (n.d.). Picard. In *Broad Institute, GitHub repository*. GitHub. http://broadinstitute.github.io/picard/

Brouwer, D., & St. Clair, D. (2004). Fine mapping of three quantitative trait loci for late blight resistance in tomato using near isogenic lines (NILs) and sub-NILs. *Theoretical and Applied Genetics*, *108*(4), 628–638.

Cai, B., Li, Q., Xu, Y., Yang, L., Bi, H., & Ai, X. (2016). Genome-wide analysis of the fructose 1, 6-bisphosphate aldolase (FBA) gene family and functional characterization of FBA7 in tomato. *Plant Physiology and Biochemistry*, *108*, 251–265.

Chen, H., Chen, X., Chen, D., Li, J., Zhang, Y., & Wang, A. (2015). A comparison of the low temperature transcriptomes of two tomato genotypes that differ in freezing tolerance: *Solanum lycopersicum* and *Solanum habrochaites*. *BMC Plant Biology*, *15*(1), 1–16.

Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E. E., Turner, S. W., & Korlach, J. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods*, *10*(6), 563–569. https://doi.org/10.1038/nmeth.2474

Cho, S. K., Kim, J. E., Park, J.-A., Eom, T. J., & Kim, W. T. (2006). Constitutive expression of abiotic stress-inducible hot pepper CaXTH3, which encodes a xyloglucan endotransglucosylase/hydrolase homolog, improves drought and salt tolerance in transgenic *Arabidopsis* plants. *FEBS Letters*, *580*(13), 3136–3144.

Choi, J. Y., Seo, Y. S., Kim, S. J., Kim, W. T., & Shin, J. S. (2011). Constitutive expression of CaXTH3, a hot pepper xyloglucan endotransglucosylase/hydrolase, enhanced tolerance to salt and drought stresses without phenotypic defects in tomato plants (*Solanum lycopersicum* cv. Dotaerang). *Plant Cell Reports*, *30*(5), 867–877.

Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., & Guarino, L. (2017). Past and Future Use of Wild Relatives in Crop Breeding. *Crop Science*, *57*(3), 1070–1082. https://doi.org/10.2135/cropsci2016.10.0885

Des Marais, D. L., & Juenger, T. E. (2010). Pleiotropy, plasticity, and the evolution of plant abiotic stress tolerance. *Annals of the New York Academy of Sciences*, *1206*(1), 56–79.

Dvorak, J., McGuire, P. E., & Cassidy, B. (1988). Apparent sources of the A genomes of wheats inferred from polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. *Genome*, *30*(5), 680–689.

Easlon, H. M., Asensio, J. S. R., St. Clair, D. A., & Bloom, A. J. (2013). Chilling-induced water stress: Variation in shoot turgor maintenance among wild tomato species from diverse habitats. *American Journal of Botany*, *100*(10), 1991–1999. https://doi.org/10.3732/ajb.1200508

FAOSTAT. (2020). *FAOSTAT Data*. https://www.fao.org/faostat/en/#data

Fernandez-Pozo, N., Menda, N., Edwards, J. D., Saha, S., Tecle, I. Y., Strickler, S. R., Bombarely, A., Fisher-York, T., Pujar, A., & Foerster, H. (2015). The Sol Genomics Network (SGN)—From genotype to phenotype to breeding. *Nucleic Acids Research*, *43*(D1), D1036–D1041.

Flint, J., & Mott, R. (2001). Finding the molecular basis of quatitative traits: Successes and pitfalls. *Nature Reviews Genetics*, *2*(6), 437–445. https://doi.org/10.1038/35076585

Foolad, M., Zhang, L., Khan, A. A., Niño-Liu, D., & Lin, G. (2002). Identification of QTLs for early blight (*Alternaria solani*) resistance in tomato using backcross populations of a *Lycopersicon esculentum* × *L. hirsutum* cross. *Theoretical and Applied Genetics*, *104*(6), 945–958. https://doi.org/10.1007/s00122-002-0870-z

Fridman, E., Carrari, F., Liu, Y.-S., Fernie, A. R., & Zamir, D. (2004). Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science*, *305*(5691), 1786–1789. https://doi.org/10.1126/science.1101666

Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D. M., Burzynski-Chang, E. A., Fish, T. L., Stromberg, K. A., Sacks, G. L., Thannhauser, T. W., Foolad, M. R., Diez, M. J.,

Blanca, J., Canizares, J., Xu, Y., van der Knaap, E., Huang, S., Klee, H. J., … Fei, Z. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nature Genetics*, *51*(6), 1044–1051. https://doi.org/10.1038/s41588-019-0410-2

Gonda, I., Ashrafi, H., Lyon, D. A., Strickler, S. R., Hulse-Kemp, A. M., Ma, Q., Sun, H., Stoffel, K., Powell, A. F., Futrell, S., Thannhauser, T. W., Fei, Z., Van Deynze, A. E., Mueller, L. A., Giovannoni, J. J., & Foolad, M. R. (2019). Sequencing-based bin map construction of a tomato mapping population, facilitating high-resolution quantitative trait loci detection. *The Plant Genome*, *12*(1), 180010. https://doi.org/10.3835/plantgenome2018.02.0010

Gong, P., Zhang, J., Li, H., Yang, C., Zhang, C., Zhang, X., Khurram, Z., Zhang, Y., Wang, T., Fei, Z., & Ye, Z. (2010). Transcriptional profiles of drought-responsive genes in modulating transcription signal transduction, and biochemical pathways in tomato. *Journal of Experimental Botany*, *61*(13), 3563–3575. https://doi.org/10.1093/jxb/erq167

Groh, A. M., Kubond, B. A., & St. Clair, D. A. (2022). Fine mapping of QTL for water use efficiency-related traits on chromosome 9 of *Solanum habrochaites* in the field. *Crop Science*. DOI: 10.1002/csc2.20828. (*in press*).

Groh, A. M., & St. Clair, D. A. (2022). Combining ability of tomato introgression lines derived from two wild species for water use efficiency and yield. *Crop Science*. (*submitted*).

Guo, M., Liu, J.-H., Ma, X., Luo, D.-X., Gong, Z.-H., & Lu, M.-H. (2016). The plant heat stress transcription factors (HSFs): Structure, regulation, and function in response to abiotic stresses. *Frontiers in Plant Science*, *7*, 114.

Haggard, J. E., Johnson, E. B., & St. Clair, D. A. (2013). Linkage relationships among multiple QTL for horticultural traits and late blight (*P. infestans*) resistance on chromosome 5 introgressed from wild tomato *Solanum habrochaites*. *G3: Genes, Genomes, Genetics*, *3*(12), 2131–2146. https://doi.org/10.1534/g3.113.007195

Hu, H., Scheben, A., & Edwards, D. (2018). Advances in integrating genomics and bioinformatics in the plant breeding pipeline. *Agriculture*, *8*(6). https://doi.org/10.3390/agriculture8060075

Huang, K., & Rieseberg, L. H. (2020). Frequency, origins, and evolutionary role of chromosomal inversions in plants. *Frontiers in Plant Science*, *11*, 296.

Ishida, K., & Yokoyama, R. (2022). Reconsidering the function of the xyloglucan endotransglucosylase/hydrolase family. *Journal of Plant Research*, 1–12.

Jiang, Y., Cai, Z., Xie, W., Long, T., Yu, H., & Zhang, Q. (2012). Rice functional genomics research: Progress and implications for crop genetic improvement. *Biotechnology Advances*, *30*(5), 1059–1070. https://doi.org/10.1016/j.biotechadv.2011.08.013

Kennedy, G. G. (2003). Tomato, pests, parasitoids, and predators: Tritrophic interactions involving the genus *Lycopersicon*. *Annual Review of Entomology*, *48*(1), 51–72.

Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Research*, *12*(4), 656–664.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, *14*(4), 1–13.

Kim, J.-S., Mizoi, J., Kidokoro, S., Maruyama, K., Nakajima, J., Nakashima, K., Mitsuda, N., Takiguchi, Y., Ohme-Takagi, M., Kondou, Y., Yoshizumi, T., Matsui, M., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2012). Arabidopsis GROWTH-REGULATING FACTOR7 functions as a transcriptional repressor of abscisic acid– and osmotic stress–responsive genes, including DREB2A. *The Plant Cell*, *24*(8), 3393–3405. https://doi.org/10.1105/tpc.112.100933

Kim, K. D., Kang, Y., & Kim, C. (2020). Application of genomic big data in plant breeding: Past, present, and future. *Plants*, *9*(11). https://doi.org/10.3390/plants9111454

Koenig, D., Jiménez-Gómez, J. M., Kimura, S., Fulop, D., Chitwood, D. H., Headland, L. R., Kumar, R., Covington, M. F., Devisetty, U. K., Tat, A. V., & others. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proceedings of the National Academy of Sciences*, *110*(28), E2655–E2662.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, *27*(5), 722–736.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547.

Kuzay, S., Xu, Y., Zhang, J., Katz, A., Pearce, S., Su, Z., Fraser, M., Anderson, J. A., Brown-Guedira, G., DeWitt, N., Peters Haugrud, A., Faris, J. D., Akhunov, E., Bai, G., & Dubcovsky, J. (2019). Identification of a candidate gene for a QTL for spikelet number per spike on wheat chromosome arm 7AL by high-resolution genetic mapping. *Theoretical and Applied Genetics*, *132*(9), 2689–2705. https://doi.org/10.1007/s00122-019-03382-5

Labate, J. A., Grandillo, S., Fulton, T., Muños, S., Caicedo, A. L., Peralta, I., Ji, Y., Chetelat, R. T., Scott, J., Gonzalo, M. J., & others. (2007). Tomato. In *Vegetables* (pp. 1–125). Springer.

Lam, E. T., Hastie, A., Lin, C., Ehrlich, D., Das, S. K., Austin, M. D., Deshpande, P., Cao, H., Nagarajan, N., Xiao, M., & Kwok, P.-Y. (2012). Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nature Biotechnology*, *30*(8), 771–776. https://doi.org/10.1038/nbt.2303

Landis, J. B., Miller, C. M., Broz, A. K., Bennett, A. A., Carrasquilla-Garcia, N., Cook, D. R., Last, R. L., Bedinger, P. A., & Moghe, G. D. (2021). Migration through a major Andean ecogeographic disruption as a driver of genetic and phenotypic diversity in a wild tomato species. *Molecular Biology and Evolution*, *38*(8), 3202–3219. https://doi.org/10.1093/molbev/msab092

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100.

Li, Z., Zhang, L., Wang, A., Xu, X., & Li, J. (2013). Ectopic overexpression of *SlHsfA3*, a heat stress transcription factor from tomato, confers increased thermotolerance and salt hypersensitivity in germination in transgenic *Arabidopsis*. *PLOS ONE*, *8*(1), 1–16. https://doi.org/10.1371/journal.pone.0054880

Liang, Y.-K., Dubos, C., Dodd, I. C., Holroyd, G. H., Hetherington, A. M., & Campbell, M. M. (2005). AtMYB61, an R2R3-MYB transcription factor controlling stomatal aperture in *Arabidopsis thaliana*. *Current Biology*, *15*(13), 1201–1206.

Liu, H., Ouyang, B., Zhang, J., Wang, T., Li, H., Zhang, Y., Yu, C., & Ye, Z. (2012). Differential modulation of photosynthesis, signaling, and transcriptional regulation between tolerant and sensitive tomato genotypes under cold stress. *PLOS ONE*, *7*(11), 1–16. https://doi.org/10.1371/journal.pone.0050785

Liu, Q., Dong, G.-R., Ma, Y.-Q., Zhao, S. M., Liu, X., Li, X.-K., Li, Y., & Hou, B. (2021). Rice glycosyltransferase gene UGT85E1 is involved in drought stress tolerance through enhancing abscisic acid response. *Frontiers in Plant Science*, *12*, 790195–790195.

Liu, Y.-S., Gur, A., Ronen, G., Causse, M., Damidaux, R., Buret, M., Hirschberg, J., & Zamir, D. (2003). There is more to tomato fruit colour than candidate carotenoid genes. *Plant Biotechnology Journal*, *1*(3), 195–207. https://doi.org/10.1046/j.1467-7652.2003.00018.x

Lounsbery, J. K., Arms, E. M., Bloom, A. J., & St. Clair, D. A. (2016). Quantitative trait loci for water-stress tolerance traits localize on chromosome 9 of wild tomato. *Crop Science*, *56*(4), 1514–1525. https://doi.org/10.2135/cropsci2015.07.0432

Lu, W., Tang, X., Huo, Y., Xu, R., Qi, S., Huang, J., Zheng, C., & Wu, C. (2012). Identification and characterization of fructose 1,6-bisphosphate aldolase genes in *Arabidopsis* reveal a gene family with diverse responses to abiotic stresses. *Gene*, *503*(1), 65–74. https://doi.org/10.1016/j.gene.2012.04.042

Luo, M.-C., Gu, Y. Q., Puiu, D., Wang, H., Twardziok, S. O., Deal, K. R., Huo, N., Zhu, T., Wang, L., Wang, Y., McGuire, P. E., Liu, S., Long, H., Ramasamy, R. K., Rodriguez, J. C., Van, S. L., Yuan, L., Wang, Z., Xia, Z., … Dvořák, J. (2017). Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature*, *551*(7681), 498–502. https://doi.org/10.1038/nature24486

Miller, J., & Tanksley, S. (1990). RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. *Theoretical and Applied Genetics*, *80*(4), 437–448.

Monforte, A., Friedman, E., Zamir, D., & Tanksley, S. (2001). Comparison of a set of allelic QTL-NILs for chromosome 4 of tomato: Deductions about natural variation and implications for germplasm utilization. *Theoretical and Applied Genetics*, *102*(4), 572–590.

Monforte, A. J., & Tanksley, S. D. (2000). Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: a tool for gene mapping and gene discovery. *Genome*, *43*(5), 803-813.

Nguyen, K. L., Grondin, A., Courtois, B., & Gantet, P. (2019). Next-generation sequencing accelerates crop gene discovery. *Trends in Plant Science*, *24*(3), 263–274. https://doi.org/10.1016/j.tplants.2018.11.008

Nir, I., Moshelion, M., & Weiss, D. (2014). The A rabidopsis GIBBERELLIN METHYL TRANSFERASE 1 suppresses gibberellin activity, reduces whole-plant transpiration and promotes drought tolerance in transgenic tomato. *Plant, Cell & Environment*, *37*(1), 113–123.

Ohmori, T., Murata, M., & Motoyoshi, F. (1996). Molecular characterization of RAPD and SCAR markers linked to the Tm-1 locus in tomato. *Theoretical and Applied Genetics*, *92*(2), 151–156. https://doi.org/10.1007/BF00223369

Omidbakhshfard, M. A., Proost, S., Fujikura, U., & Mueller-Roeber, B. (2015). Growth-regulating factors (GRFs): A small transcription factor family with important functions in plant biology. *Molecular Plant*, *8*(7), 998–1010. https://doi.org/10.1016/j.molp.2015.01.013

Pathak, T. B., Maskey, M. L., Dahlberg, J. A., Kearns, F., Bali, K. M., & Zaccaria, D. (2018). Climate change trends and impacts on California agriculture: A detailed review. *Agronomy*, *8*(3), 25.

Peralta, I. E., & Spooner, D. M. (2007). History, origin and early cultivation of tomato (*Solanaceae*). *Genetic Improvement of Solanaceous Crops*, *2*, 1–27.

Pflieger, S., Lefebvre, V., & Causse, M. (2001). The candidate gene approach in plant genetics: A review. *Molecular Breeding*, *7*(4), 275–291. https://doi.org/10.1023/A:1011605013259

Powell, A. F., Feder, A., Li, J., Schmidt, M. H.-W., Courtney, L., Alseekh, S., Jobson, E. M., Vogel, A., Xu, Y., Lyon, D., Dumschott, K., McHale, M., Sulpice, R., Bao, K., Lal, R., Duhan, A., Hallab, A., Denton, A. K., Bolger, M. E., … Usadel, B. (2022). A *Solanum lycopersicoides* reference genome facilitates insights into tomato specialized metabolism and immunity. *The Plant Journal*, *110*(6), 1791–1810. https://doi.org/10.1111/tpj.15770

Prohens, J., Gramazio, P., Plazas, M., Dempewolf, H., Kilian, B., Diez, M. J., Fita, A., Herraiz, F. J., Rodriguez-Burruezo, A., Soler, S., & others. (2017). Introgressiomics: A new approach for using crop wild relatives in breeding for adaptation to climate change. *Euphytica*, *213*(7), 1–19.

Rick, C. M. (1983). Genetic variability in tomato species. *Plant Molecular Biology Reporter*, *1*(2), 81–87.

Rousseaux, M. C., Jones, C. M., Adams, D., Chetelat, R., Bennett, A., & Powell, A. (2005). QTL analysis of fruit antioxidants in tomato using *Lycopersicon pennellii* introgression lines. *Theoretical and Applied Genetics*, *111*(7), 1396–1408. https://doi.org/10.1007/s00122-005-0071-7

Sakuma, Y., Maruyama, K., Qin, F., Osakabe, Y., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2006). Dual function of an Arabidopsis transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. *Proceedings of the National Academy of Sciences*, *103*(49), 18822–18827.

Salvi, S., & Tuberosa, R. (2005). To clone or not to clone plant QTLs: Present and future challenges. *Trends in Plant Science*, *10*(6), 297–304.

Schmidt, M. H. W., Vogel, A., Denton, A. K., Istace, B., Wormit, A., Geest, H. van de, Bolger, M. E., Alseekh, S., Maß, J., Pfaff, C., Schurr, U., Chetelat, R., Maumus, F., Aury, J. M., Koren, S., Fernie, A. R., Zamir, D., Bolger, A. M., & Usadel, B. (2017). De novo assembly of a new *Solanum pennellii* accession using nanopore sequencing. *Plant Cell*, *29*(10), 2336–2348. https://doi.org/10.1105/tpc.17.00521

Schramm, F., Larkindale, J., Kiehlmann, E., Ganguli, A., Englich, G., Vierling, E., & Von Koskull-Döring, P. (2008). A cascade of transcription factor DREB2A and heat stress transcription factor HsfA3 regulates the heat stress response of Arabidopsis. *The Plant Journal*, *53*(2), 264–274. https://doi.org/10.1111/j.1365-313X.2007.03334.x

Seong, K., Shaw, C. L., Seo, E., Li, M., Krasileva, K. V., & Staskawicz, B. (2022). A draft genome assembly for the heterozygous wild tomato *Solanum habrochaites* highlights haplotypic structural variations of intracellular immune receptors. *BioRxiv*, 2022.01.21.477156. https://doi.org/10.1101/2022.01.21.477156

Shelton, J. M., Coleman, M. C., Herndon, N., Lu, N., Lam, E. T., Anantharaman, T., Sheth, P., & Brown, S. J. (2015). Tools and pipelines for BioNano data: Molecule assembly pipeline and FASTA super scaffolding tool. *BMC Genomics*, *16*(1), 734. https://doi.org/10.1186/s12864-015-1911-8

Shi, Y., Phan, H., Liu, Y., Cao, S., Zhang, Z., Chu, C., & Schläppi, M. R. (2020). Glycosyltransferase OsUGT90A1 helps protect the plasma membrane during chilling stress in rice. *Journal of Experimental Botany*, *71*(9), 2723–2739.

Shohat, H., Cheriker, H., Kilambi, H. V., Illouz Eliaz, N., Blum, S., Amsellem, Z., Tarkowská, D., Aharoni, A., Eshed, Y., & Weiss, D. (2021). Inhibition of gibberellin accumulation by water deficiency promotes fast and long-term 'drought avoidance' responses in tomato. *New Phytologist*, *232*(5), 1985–1998. https://doi.org/10.1111/nph.17709

Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, *24*(5), 637–644.

Takei, H., Shirasawa, K., Kuwabara, K., Toyoda, A., Matsuzawa, Y., Iioka, S., & Ariizumi, T. (2021). De novo genome assembly of two tomato ancestors, *Solanum pimpinellifolium* and *Solanum lycopersicum* var. Cerasiforme, by long-read sequencing. *DNA Research*, *28*(1). https://doi.org/10.1093/dnares/dsaa029

Thapa, S. P., Miyao, E. M., Michael Davis, R., & Coaker, G. (2015). Identification of QTLs controlling resistance to *Pseudomonas syringae* pv. Tomato race 1 strains from the wild tomato, *Solanum habrochaites* LA1777. *Theoretical and Applied Genetics*, *128*(4), 681–692. https://doi.org/10.1007/s00122-015-2463-7

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*(2), 178–192.

Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, *485*(7400), 635.

Truco, M., Randall, L., Bloom, A., & St. Clair, D. (2000). Detection of QTLs associated with shoot wilting and root ammonium uptake under chilling temperatures in an interspecific backcross population from *Lycopersicon esculentum× L. hirsutum*. *Theoretical and Applied Genetics*, *101*(7), 1082–1092.

Wang, X., Gao, L., Jiao, C., Stravoravdis, S., Hosmani, P. S., Saha, S., Zhang, J., Mainiero, S., Strickler, S. R., Catala, C., Martin, G. B., Mueller, L. A., Vrebalov, J., Giovannoni, J. J., Wu, S., & Fei, Z. (2020). Genome of *Solanum pimpinellifolium* provides insights into structural variants during tomato breeding. *Nature Communications*, *11*(1), 5817. https://doi.org/10.1038/s41467-020-19682-0

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., & Guo, H. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, *40*(7), e49–e49.

Yamaguchi, S. (2008). Gibberellin metabolism and its regulation. *Annual Review of Plant Biology*, *59*(1), 225–251. https://doi.org/10.1146/annurev.arplant.59.032607.092804

Yang, A., Dai, X., & Zhang, W.-H. (2012). A R2R3-type MYB gene, *OsMYB2*, is involved in salt, cold, and dehydration tolerance in rice. *Journal of Experimental Botany*, *63*(7), 2541–2556.

Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, *7*(1–2), 203–214.

Zhu, T., Hu, Z., Rodriguez, J. C., Deal, K. R., Dvorak, J., Vogel, J. P., Liu, Z., & Luo, M.-C. (2018). Analysis of *Brachypodium* genomes with genome-wide optical maps. *Genome*, *61*(8), 559–565.

TABLES AND FIGURES

Table 3.1. Properties of the *S. habrochaites* acc. LA1778 genome sequence in various iterations of assembling and scaffolding. TM_CTG_V0 represents the initial assembly by Canu. TM_CTG_V0.1 represents the assembly after polishing using the Quiver consensus tool. TM_CTG_V1 represents the assembly validated and corrected by the optical map. TM_SCF is the scaffolding result (scaffolds) after applying the Stitch algorithm.

	TM_CTG_V0	TM_CTG_V0.1	TM_CTG_V1	TM_SCF
NUMBER OF SEQUENCES	4,182	4,097	4,136	3,695
TOTAL SIZE (BP)	1,099,694,196	1,094,915,996	1,094,915,996	1,120,945,628
SEQUENCE N50 (BP)	1,095,134	1,115,370	1,058,661	2,762,933
MAX LENGTH (BP)	14,462,597	14,523,410	14,523,410	17,557,643
N%	0	0	0	2.33

Table 3.2. Properties of the Bionano optical map of the *S. habrochaites* acc. LA1778 genome.

Enzyme	Nt.*Bsp*QI
# Molecules	515,098
Molecule N50	319 Kb
Molecule total length	161 Gb
# Contigs	995
Map total length	1.05 Gb
Contig N50	1.53 Mb

Table 3.3. Comparison of chromosomes between the assemblies of *S. lycopersicum* (SL3.1, ITAG3.2) and the newly assembled *S. habrochaites* acc. LA1778. The *S. habrochaites* gene counts are based on the number of genes from ITAG3.2 that were successfully mapped to the assembly.

	***S. habrochaites* acc. LA1778 Assembly**					***S. lycopersicum ITAG3.2***				
Chr #	**Len (bp)**	**N_Len (bp)**	**Eff_Len (bp)**	**N %**	**Gene #**	**Len (bp)**	**N_Len (bp)**	**Eff_Len (bp)**	**N %**	**Gene #**
Chr1	125,421,613	7,741,658	117,679,955	6.17%	4,400	98,455,869	10,450,621	88,005,248	10.61%	4,442
Chr2	69,897,999	1,303,876	68,594,123	1.87%	3,526	55,977,580	7,271,751	48,705,829	12.99%	3,576
Chr3	75,907,451	1,641,908	74,265,543	2.16%	3,164	72,290,146	10,761,438	61,528,708	14.89%	3,483
Chr4	76,732,650	865,700	75,866,950	1.13%	2,852	66,557,038	5,940,647	60,616,391	8.93%	2,844
Chr5	81,709,027	2,125,466	79,583,561	2.60%	2,826	66,723,567	5,118,953	61,604,614	7.67%	2,531
Chr6	57,605,876	1,915,082	55,690,794	3.32%	2,798	49,794,276	5,833,871	43,960,405	11.72%	2,939
Chr7	82,629,036	1,723,994	80,905,042	2.09%	2,427	68,175,699	5,722,367	62,453,332	8.39%	2,551
Chr8	70,865,583	1,807,224	69,058,359	2.55%	2,326	65,987,440	5,715,954	60,271,486	8.66%	2,529
Chr9	87,670,881	2,130,549	85,540,332	2.43%	2,441	72,906,345	7,493,324	65,413,021	10.28%	2,552
Chr10	82,686,178	1,290,037	81,396,141	1.56%	2,644	65,633,393	2,518,649	63,114,744	3.84%	2,576
Chr11	67,481,045	1,882,234	65,598,811	2.79%	2,505	56,597,135	5,856,292	50,740,843	10.35%	2,441
Chr12	80,396,545	2,513,565	77,882,980	3.13%	2,745	68,126,176	5,999,612	62,126,564	8.81%	2,486
Chr0	-	-	-	-	-	20,852,292	3,036,007	17,816,285	14.56%	818
Total	959,003,884	26,941,293	932,062,591	2.81%	34,654	807,224,664	78,683,479	728,541,185	9.75%	35,768

Table 3.4. Comparison of the *stm9* QTL interval and the *S. habrochaites* acc. 1778 chromosome 9 linkage map from Groh et al. (2022) to the physical positions of the markers in the assembled genomes of *S. lycopersicum* (SL4.0), *S. habrochaites* acc. 1778 (newly assembled), and *S. pennellii* (whole-genome shotgun sequence from SGN BLAST; Bolger et al., 2014). The *stm9* QTL ranges from the gene Solyc09g008310 to the marker T1673 (highlighted in blue). The region of chromosome 9 containing the most highly coincident QTL from Groh et al. (2022) ranges from marker H307 to TG10 (highlighted in yellow and green). The $\Delta^{13}C$ QTL region ranges from At2g36480 to TG10 (highlighted in green).

Marker	**Linkage Map Position (cM)**	***S. lycopersicum* (SL4.0) bp position of the SNP [a]**	***S. habrochaites* bp position the SNP [a]**	***S. pennellii* bp position of the SNP [a]**
Solyc09g008310	-	1776418	1853030	1799276
T1673	0.00	2098817	2233206	766379 [b]
H14	0.33	2256902	- [c]	2264543
H307	0.87	2347670	2521643	2358943
T0532	1.15	2414054	2636454	2427251
At2g36480	1.17	2497345	2719281	2518486
CT283	1.89	2874797 [d]	3120786 [d]	2972654 [d]
SSR73	1.98	2755822 [d]	2997656 [d]	2851411 [d]
TG10	2.12	2926980	3182960	3031460
SL9750	2.53	3269552	3474729	3341200
T1148	2.74	3416789	3617690	3497969
T1908	2.85	3475385	3673869	3554402
At5g06140	3.17	3526253	3725846	3609043
T1617	3.52	3792816	4006514	3885604
T0641	3.59	3865813	4101447	3972894

a: The positions listed for Solyc09g008310 are the fist bp of the gene sequence from SGN

b: The pb position of T1673 in *S. pennellii* is much different than the other two genomes because the BLAST results only revealed a partial match for the marker at this position

c: The bp position for H14 could not be found using BLAST in the *S. habrochaites* assembly

d: The order of markers is based on the linkage map created by Groh et al. (2022). The physical positions of markers CT283 and SSR73 in all three genomes may reflect a potential error in the ordering of the linkage map.

Table 3.5. Physical ranges and lengths of the full chromosome 9 sub-NIL region, the interval with the most coincident QTL, and the $\Delta^{13}C$ QTL from Groh et al. (2022) for three tomato species. The marker interval for each region is noted in parentheses.

		Full sub-NIL Region (markers T1673-T0641)			**Highly Coincident QTL Interval (markers H307-TG10)**			**$\Delta^{13}C$ QTL Interval (markers At2g36480-TG10)**		
	Total Length of Chr9	**Physical range (bp)**	**Length of region (bp)**	**% of Chr 9 length**	**Physical range (bp)**	**Length of region (bp)**	**% of Chr 9 length**	**Physical range (bp)**	**Length of region (bp)**	**% of Chr 9 length**
S. lycopersicum	68,513,564	2,098,817 – 3,865,813	1,766,996	2.58	2,347,670 – 2,926,980	579,310	0.85	2,497,345 – 2,926,980	429,635	0.63
S. habrochaites	87,670,881	2,233,206 – 4,101,447	1,868,241	2.13	2,521,643 – 3,182,960	661,317	0.75	2,719,281 – 3,182,960	463,679	0.53
S. pennellii	84,057,508	766,379 – 3,972,894 [a]	3,206,515 [a]	3.81	2,358,943 – 3,031,460	672,517	0.8	2,518,486 – 3,031,460	512,974	0.61

a: The physical length of the full sub-NIL region is skewed for *S. pennellii* because the first marker only had a partial match at the recorded position. This may not reflect the true length of the full sub-NIL region.

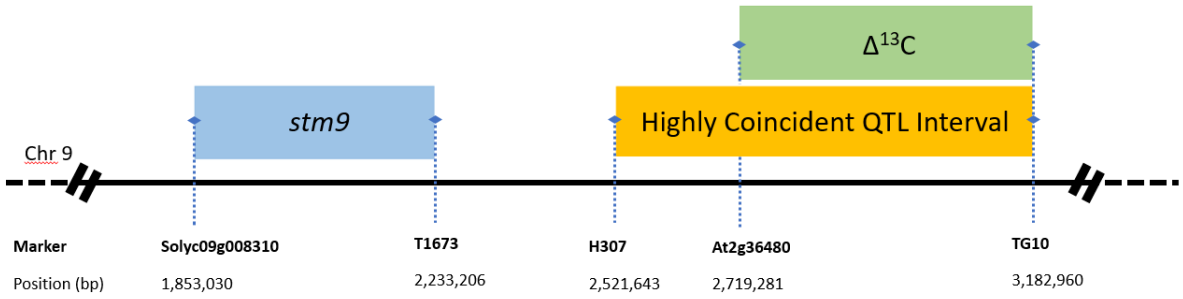


Figure 3.1 Schematic representation of the *S. habrochaites* acc. LA1778 chromosome 9 region containing QTL for the traits shoot turgor maintenance under root chilling (*stm9*), carbon isotope discrimination ($\Delta^{13}C$), and the highly coincident QTL interval. The black horizontal line represents a segment of the short arm of *S. habrochaites* chromosome 9. Each colored bar represents a QTL interval. The markers and physical positions of the start and end of each QTL interval are noted below the corresponding blue dotted lines.

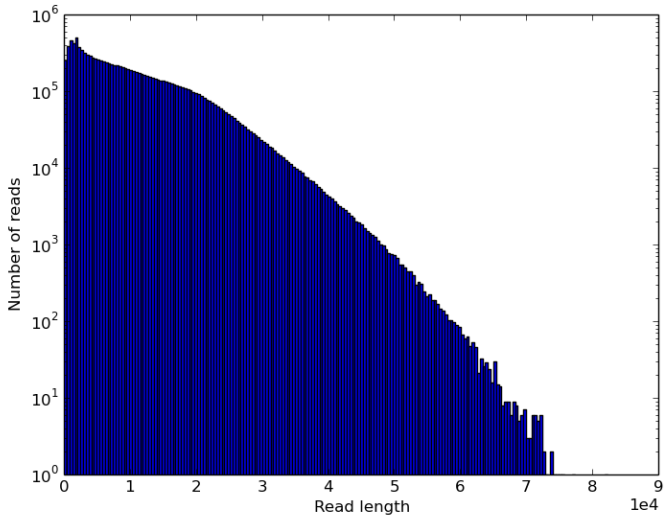


Figure 3.2. The profile of raw sequencing data for *S. habrochaites* acc. LA1778 from the PacBio RSII system. The read length (horizontal axis) is in 1e4 bp.

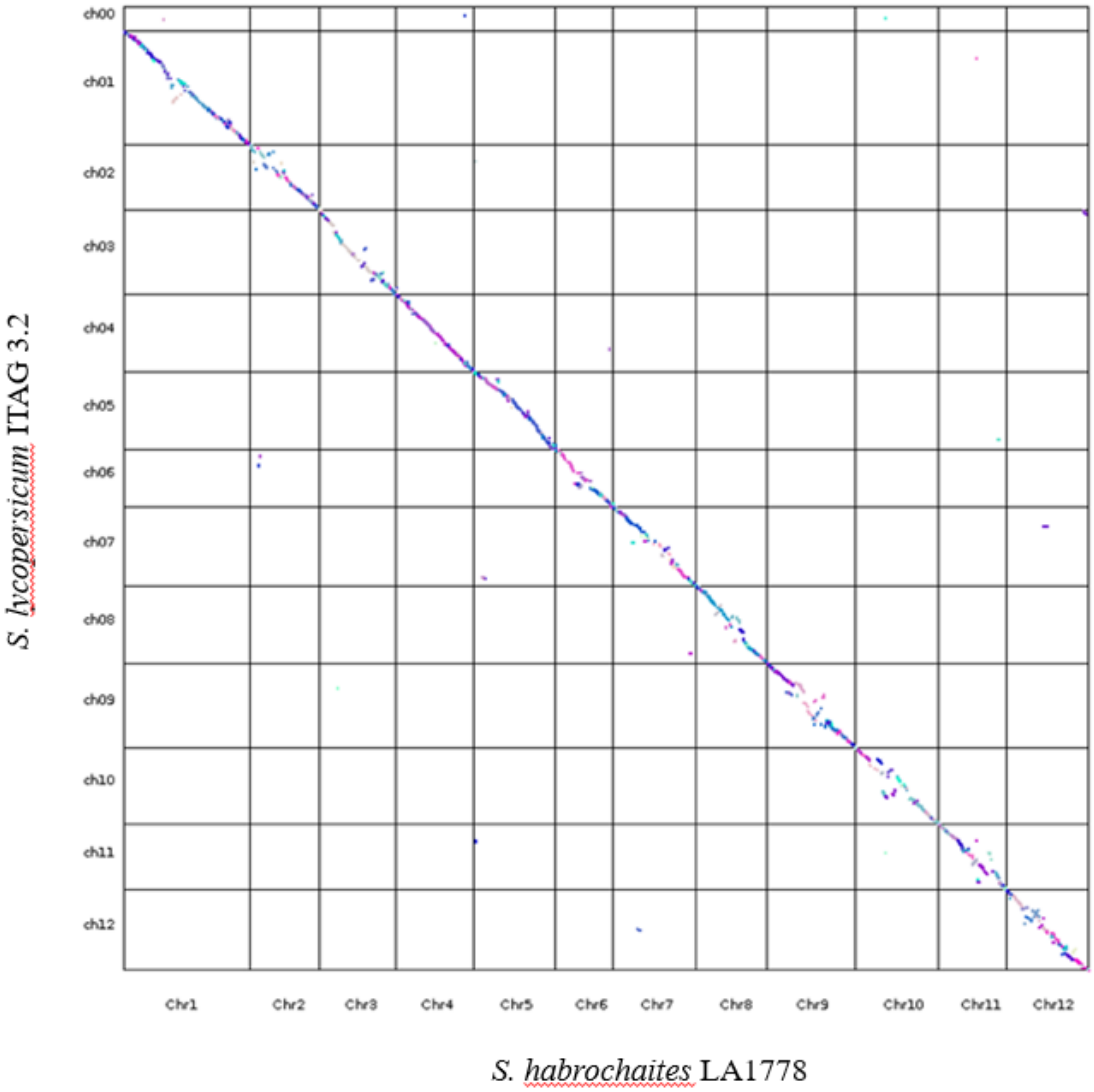


Figure 3.3. Gene synteny dot plots between *S. habrochaites* acc. LA1778 pseudomolecules and pseudomolecules of the reference genome of cultivated tomato, *S. lycopersicum*. The beginnings of the pseudomolecules are to the upper left.

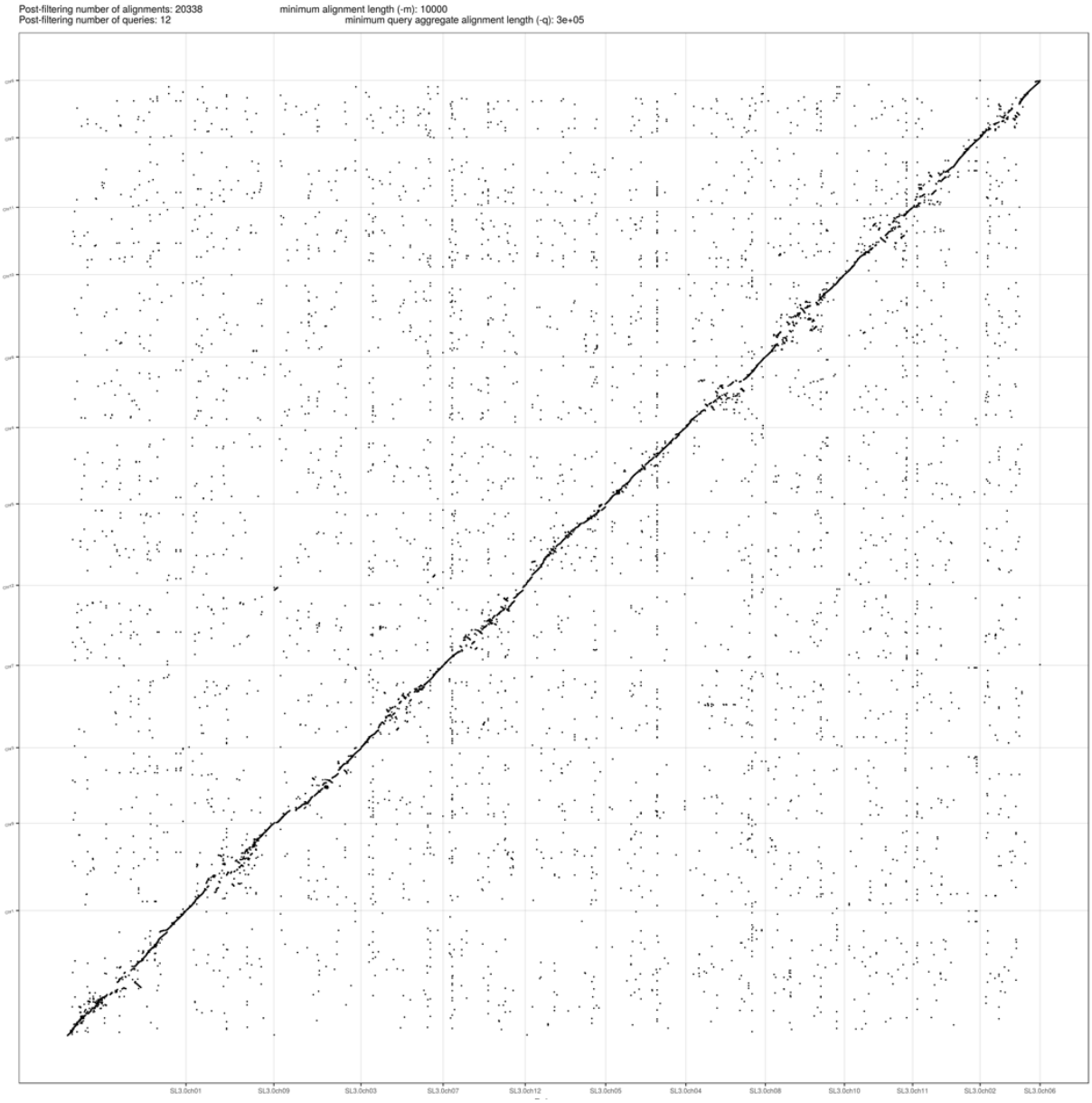


Figure 3.4. Dot-plot of the whole-genome alignment of the *S. habrochaites* acc. 1778 pseudomolecules (vertical axes) against the reference genome cultivated tomato pseudomolecules (SL3.0) (horizontal axes). The beginnings of the pseudomolecules are to the bottom or left.

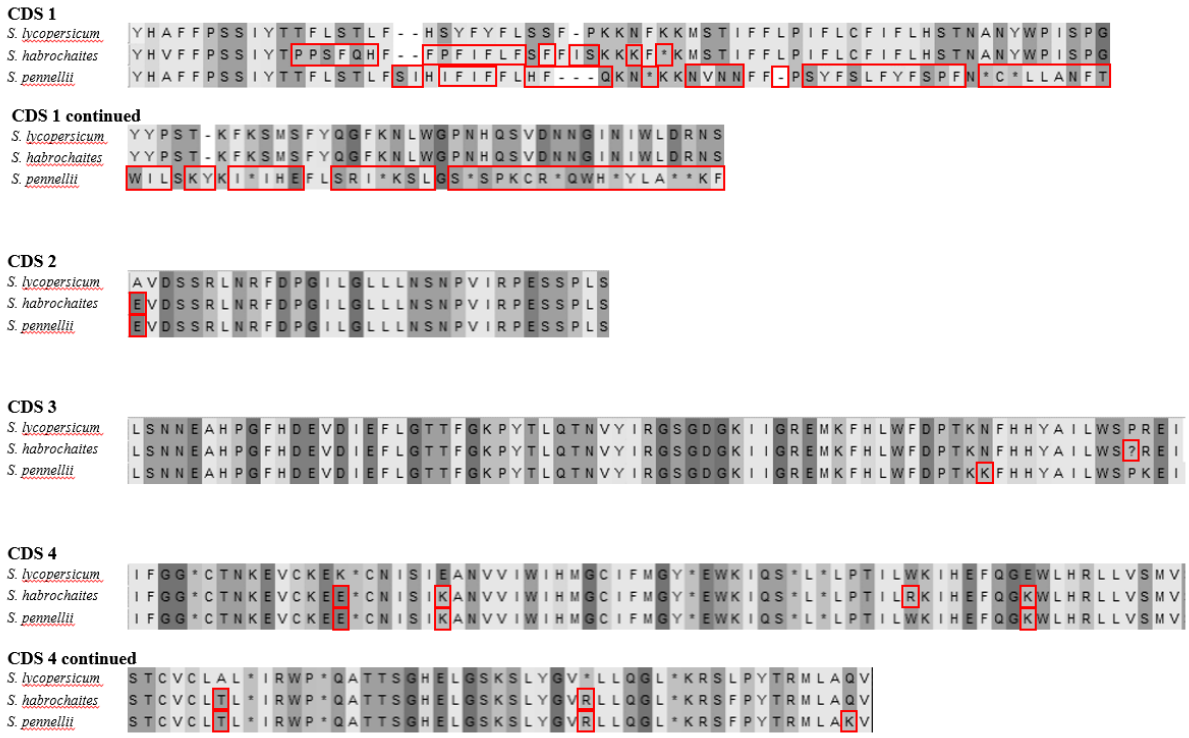


Figure 3.5 Alignment of the amino acid translations of the four coding sequences (CDS) for the xyloglucan endotransglucosylase/hydrolase gene (XTH). The *S. lycopersicum* SL4.0 amino acid sequence is in the top row of each CDS, the second sequence is from *S. habrochaites* acc. LA1778, and the bottom sequence is from *S. pennellii* (SGN, Bolger et al., 2014). Stop codons are represented by the * symbol and gaps are represented by dashes. Red boxed highlight amino acids that differ from *S. lycopersicum*.

SUPPLEMENTAL MATERIAL

Supplemental Table S3.1: List of genes and gene descriptions within the chromosome 9 QTL intervals for *stm9*, $\Delta^{13}C$, and the highly coincident QTL interval. The source of the gene (mapped genes from *S. lycopersicum,* Augustus predicted genes, or mapped transcripts), the start and stop bp positions, gene IDs (if available), and descriptions (if available) are noted. Blue highlighted gene descriptions were interesting candidates for each trait QTL.

CONCLUSIONS AND FUTURE DIRECTIONS

The overall goal of this dissertation research was to investigate specific chromosomal regions of *S. habrochaites* and *S. pennellii* that are associated with water stress tolerance-related traits for their potential use in improvement of water stress tolerance in cultivated tomato. Molecular and genomic tools were produced for *S. habrochaites* acc. LA1778, including a *de novo* whole-genome sequence (WGS) for this species and a quantitative trait locus (QTL) linkage map for chromosome 9, to facilitate the integration of positive effect QTL alleles into cultivated tomato. It was also shown that certain *S. habrochaites* and *S. pennellii* introgressions in introgression lines (ILs) could be utilized in a hybrid cultivar breeding program to improve the water use efficiency (WUE) of cultivated tomato without having negative effects on yield.

There can be challenges in using wild species in cultivar breeding pipelines, including linkage drag and genotype by environment interactions (GxE). Negative effect wild alleles for horticultural traits, such as low yields, low fruit weight, and late maturity, can be tightly linked to or pleiotropic to the positive effect alleles for WUE in *S. habrochaites* and *S. pennellii*. If these traits are not pleiotropic, then repulsion phase linkages will need to be broken for these introgressions to be used in cultivated tomato breeding. In Groh et al. (2022), *S. habrochaites* chromosome 9 QTL with negative horticultural effects on yield and maturity are highly coincident with the positive effect QTL for WUE. High-resolution mapping may help fractionate these QTL if the traits are not pleiotropic. Additionally, QTL and putative causal genes for low fruit weight on chromosome 2 and later maturity and yield on chromosome 5 of *S. pennellii* have been studied previously (Eshed & Zamir, 1995; Frary et al., 2000; Jones et al., 2007; Soyk et al., 2017). DNA markers from these earlier studies can be utilized to identify recombinants that have the beneficial

alleles associated with increased WUE without the negative effect alleles associated with maturity or fruit weight.

GxE was a complicating factor when fine mapping the trait QTL in Groh et al. (2022). The datasets had to be separated into subsets based on environmental factors, and not all of the subsets had significant QTL for a given trait. This finding suggests that the genes underlying WUE and yield in these regions are affected by additional environmental factors besides water availability. Interestingly, Groh & St. Clair (2022) did not observe the same degree of GxE for WUE-related traits and yield as in Groh et al. (2022). This is likely due to variation in environment between the years in which each study was conducted. Future experiments on water use efficiency using these lines should be replicated over years and environments to better estimate the effect of the environment on the observed phenotypes.

Introgressions from two wild tomato species were shown to have positive effects on horticultural traits such as yield. In Groh & St. Clair (2022), genotypes containing at least one introgression from a wild species exhibited heterosis for yield. Heterozygous IL genotypes exhibited similar levels of heterosis as the double introgression hybrids, suggesting that the observed heterosis is most likely due to interactions between the introgressions with each other or with the genetic background. Some introgressions and combinations of introgressions also had positive general and specific combining abilities for yield. This suggests that although wild species of tomato have low yield, certain genomic regions contain favorable alleles that can have a positive effect on yield phenotypes.

The studies described in this dissertation provide further evidence of the potential value of using wild species of tomato for improving cultivated tomato. Specific introgressions from *S. habrochaites* and *S. pennellii* (chromosomes 2, 5, and 9) that contain genes associated with WUE

were identified. These introgressions are good targets for further study and potential uses in breeding pipelines. Future work is required to determine if linkage drag can be overcome in some of these regions. Exploring these various avenues of investigation may lead to the improvement of WUE in cultivated tomato.

Future Directions

The genetic resources (QTL linkage maps and WGS), genotypes (sub-near-isogenic lines, sub-NILs), and conclusions from this dissertation supports future research on employing *S. habrochaites* and *S. pennellii* to improve WUE in cultivated tomato. In Groh et al. (2022), sub-NILs 13 and 19 had better WUE and yield than the cultivated tomato control, cultivar T5, and could be utilized directly in a breeding program. The *S. habrochaites* introgressions from these sub-NILs could be introgressed into modern germplasm using marker assisted backcrossing. Elite inbred lines containing these introgressions could be tested in the field under full and reduced water conditions to determine if they have improved WUE without reduced yield. The sub-NILs could also be crossed with other elite parent lines to determine if they would combine well in a hybrid breeding program. In Groh & St. Clair (2022), TA1325 (a *S. habrochaites* chromosome 9 introgression that is syntenic to the regions in sub-NILs 13 and 19) had better WUE as a homozygous IL than as a heterozygous IL. This suggests that these sub-NIL introgressions may perform better as homozygotes, although hybrid varieties may still benefit from having one copy of the introgressions. Depending on the outcomes of field testing these hybrids, breeders can decide if one or both parent lines of a hybrid would need to contain the introgression to improve WUE.

To utilize the QTL reported in Groh et al. (2022) in a marker-assisted breeding program, the highly coincident QTL for WUE, yield, and maturity would need to be mapped at a higher

resolution. The WGS for *S. habrochaites* could be used to identify sequence variants (single nucleotide polymorphisms, SNPs) between *S. habrochaites* and *S. lycopersicum* to identify additional recombinants and create a higher density linkage map of the QTL containing interval. However, small introgressions of wild species may experience repression of recombination, a potential complication (discussed in further detail in Groh et al., 2022). Higher resolution mapping could fractionate non-pleiotropic QTL, which would facilitate their use in marker-assisted back crossing, and also narrow down the regions containing potential candidate genes for each trait.

Transcriptome analysis of water stress susceptible and tolerant lines under normal and reduced water conditions in the field could provide data on which genes in these QTL intervals are candidate genes for $\Delta^{13}C$, yield, and maturity traits. The *S. habrochaites* sub-NIL 10 (the sub-NIL from Groh et al. (2022) containing the largest *S. habrochaites* introgression, a water stress-tolerant genotype) and cultivar T5 (water stress-susceptible) could be informative genotypes for comparison in a transcriptome study. However, designing a transcriptome study for understanding WUE in the field would be challenging. One must decide which tissues to extract RNA from and which timepoints are appropriate for evaluating the traits of interest. Genes associated with $\Delta^{13}C$ are likely to be expressed in the leaves because photosynthesis and transpiration rate (related to stomatal aperture) are related to the $\Delta^{13}C$ values. However, genes related to maturity and yield may be differentially expressed in the reproductive tissues rather than leaves. Genes that are differentially expressed between the two species will also vary based on life stage. $\Delta^{13}C$ and yield are complex traits that are a culmination of season-long genetic and environmental effects. Collecting data at various time points and from various plant tissues is essential to determine how the transcriptome may affect these phenotypes. Transcriptome data from a variety of tissues and

life stages would also contribute towards creating a *S. habrochaites* specific genome-wide annotation that can be utilized by the broader scientific community.

Introgressions from chromosomes 2, 5, and 9 of *S. habrochaites* and *S. pennellii* (TA1544, TA1325, IL5-4, and IL2-5) were shown to improve WUE in cultivated tomato in various combinations as homozygotes, heterozygotes, or combined into a double introgression hybrid (Groh & St. Clair, 2022). These introgressions are relatively large and some contain negative effect alleles for horticultural traits (discussed above). A logical next step for these genotypes would be to fine map the QTL associated with their introgressions to narrow down the regions associated with WUE and other horticultural traits. Fractionating QTL in repulsion phase linkages can generate a set of recombinant lines that may be more readily be used in a breeding pipeline. These smaller mapped regions should also be tested in combining ability studies, especially for introgressions that improved WUE in combination with other introgressions (i.e., IL5-4 and IL2-5). NILs containing portions of the introgressions from IL5-4 and IL2-5 could be crossed to an IL containing a larger introgression and phenotyped in the field. For example, a set of IL2-5 NILs (NIL2-5) could be crossed with TA1325 to create a set of TA1325xNIL2-5 hybrid genotypes. The NILs could also be crossed with NILs from the other introgressions, but this strategy would require more crosses. These hybrids could be tested in the field similarly to the methods used in Groh & St. Clair (2022) to determine if any hybrids show improved WUE and yield without the significant reduction in fruit weight or other horticultural traits. Furthermore, TA1325 had good general combining ability for $\Delta^{13}C$ and the QTL could be fine mapped and integrated into modern hybrid breeding pipelines. Lounsbery et al. (2016) and Groh et al. (2022) have already fine mapped two QTL for $\Delta^{13}C$ in this chromosome 9 introgressions, but there could be additional beneficial QTL within this region.

REFERENCES

Eshed, Y., & Zamir, D. (1995). An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics*, *141*(3), 1147–1162.

Frary, A., Nesbitt, T. C., Frary, A., Grandillo, S., Knaap, E. van der, Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K. B., & Tanksley, S. D. (2000). *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. *Science*, *289*(5476), 85–88. https://doi.org/10.1126/science.289.5476.85

Groh, A. M., Kubond, B. A., & St. Clair, D. A. (2022). Fine mapping of QTL for water use efficiency-related traits on chromosome 9 of *Solanum habrochaites* in the field. *Crop Science*. (*in press*).

Groh, A. M., & St. Clair, D. A. (2022). Combining ability of tomato introgression lines derived from two wild species for water use efficiency and yield. *Crop Science*. (*submitted*).

Jones, C. M., Rick, C. M., Adams, D., Jernstedt, J., & Chetelat, R. T. (2007). Genealogy and fine mapping of obscuravenosa, a gene affecting the distribution of chloroplasts in leaf veins, and evidence of selection during breeding of tomatoes (*Lycopersicon esculentum*; Solanaceae). *American Journal of Botany*, *94*(6), 935–947.

Lounsbery, J. K., Arms, E. M., Bloom, A. J., & St. Clair, D. A. (2016). Quantitative trait loci for water-stress tolerance traits localize on chromosome 9 of wild tomato. *Crop Science*, *56*(4), 1514–1525. https://doi.org/10.2135/cropsci2015.07.0432

Soyk, S., Müller, N. A., Park, S. J., Schmalenbach, I., Jiang, K., Hayama, R., Zhang, L., Van Eck, J., Jimenez-Gomez, J. M., & Lippman, Z. B. (2017). Variation in the flowering gene SELF PRUNING 5G promotes day-neutrality and early yield in tomato. *Nature Genetics*, *49*(1), 162–168.