

A chromosome-level genome assembly of *Coffea arabica* L. var. 'Kona Typica'

Haomin Lyu^{1,2}, Jinjin Song¹, Yanbin Yin³, Ming-Li Wang², Tracie Matsumoto¹, Vaishnavi Annabhemoju³, Lyndel W. Meinhardt⁴, Sunchung Park⁴, Seunghyun Lim⁴, Dapeng Zhang⁴, Qingyi Yu¹

¹Tropical Plant Genetic Resources and Disease Research Unit, Daniel K Inouye U.S. Pacific Basin Agricultural Research Center, Agricultural Research Service, U.S. Department of Agriculture, Hilo, Hawaii 96720, USA; ²Hawaii Agriculture Research Center, Kunia, Hawaii 96759, USA; ³Nebraska Food for Health Center, Department of Food Science and Technology, University of Nebraska-Lincoln, Lincoln, Nebraska 68588, USA; ⁴Sustainable Perennial Crops Laboratory, Agriculture Research Service, U. S. Department of Agriculture, Beltsville, Maryland 20705, USA

Introduction

Typica coffee was introduced to Hawaii from Guatemala in 1892, where it became the predominant variety in the Kona region. Today, over 90% of coffee produced in Kona is known as 'Kona Typica'. 'Kona Typica' is renowned for its exceptional quality and flavor, largely attributed to the unique geographic and climatic conditions of the region. However, despite its superior taste and quality, 'Kona Typica' is highly susceptible to pests and diseases, such as coffee leaf rust (CLR) and coffee berry disease (CBD). While Hawaii was among the last coffee-producing regions free of CLR, the disease was detected on Maui in February 2020 and subsequently spread to all other Hawaiian coffee-growing islands. Breeding programs for Typica coffee are prioritizing the improvement of disease and pest resistance to mitigate yield losses and economic impact. However, the lack of comprehensive molecular data on resistance genes remains a significant barrier to progress in coffee breeding programs. 'Typica' is one of the oldest and most culturally and genetically significant lineages of Arabica coffee. Deciphering the genome structure and organization of 'Kona Typica' coffee is critical for unraveling the genetic networks regulating important traits, thereby facilitating the development of superior varieties. In this study, we present a chromosome-level genome assembly of 'Kona Typica' coffee constructed using PacBio long-read sequencing and Omni-C data. The resulting genome comprises 22 chromosomes with a total size of 1.13 Gbp and a scaffold N50 of 50.50 Mbp. This high-quality reference genome provides valuable insights into the genomic features of 'Kona Typica' coffee and contributes to understanding the evolutionary history and domestication process of Arabica coffee genome.

Methods

Plant materials

The Arabica coffee variety 'Kona Typica' KO34 was planted in the field at Hawaii Agriculture Research Center, Kunia, Oahu. Young leaf tissue of KO34 was harvested for high molecular weight DNA extraction.

Library preparation for PacBio Sequencing and Omni-C sequencing

High molecular weight genomic DNA was extracted from isolated nuclei using lysis buffer with proteinase K, followed by phenol-chloroform extraction and isopropanol precipitation to purify DNA with minimal shearing and degradation. PacBio HiFi sequencing library was prepared using the SMRTbell Express Template Prep Kit 2.0. The DNA library was sequenced on a PacBio Sequel II System. For Hi-C sequencing, the library was prepared using the Omni-C™ Technology and sequenced on an Illumina HiSeqX system at Dovetail Genomics.

Genome assembly and quality assessment

The 'Kona Typica' genome was assembled using Hifiasm. Omni-C reads were used to phase haplotypes. Omni-C reads were aligned to the contig-level genome assembly using chromap. YaHs were used for scaffolding with default parameters. Hi-C contact matrices and editable Hi-C maps were generated using Juicer and Juicerbox facilitated visualization and manual error correction of Hi-C maps throughout the genome assembly. Minimap2 and GS-GapCloser were used to fill the N-gap. Purge_Haplotigs was used to eliminate haplotig duplications following the pipeline protocol. BUSCO analysis was conducted to assess the quality of the assembled genome.

Results

The final assembly consisted of 967 scaffolds with a total length of 1,172,725,641 bp and an N50 of 50.50 Mb. These scaffolds were anchored to 22 chromosomes with an accumulative length of 1,135,979,557 bp (Table 1, Fig. 1). BUSCO analysis revealed high genome completeness with 99.1% of BUSCO genes identified.

Table 1. Summary statistics of the 'Kona Typica' genome assembly.

Assembly characteristics	Value
Number of contigs	2,138
Contig size (bp)	1,214,654,331
Contig N50 (bp)	49,070,996
Number of scaffolds	967
Scaffold size (bp)	1,172,725,641
Scaffold N50 (bp)	50,495,782
Scaffold N90 (bp)	40,978,299
Number of chromosomes	22
Total length of anchored sequences on chromosomes (bp)	1,135,979,557
Total length of unanchored sequences (bp)	36,746,084
Proportion of unanchored sequences	3.13%
Complete BUSCOs (Number / %)	1,599 / 99.1
Complete and single-copy BUSCOs (Number / %)	80 / 5.0
Complete and duplicated BUSCOs (Number / %)	1,519 / 94.1
Fragmented BUSCOs (Number / %)	6 / 0.4
Missing BUSCOs (Number / %)	9 / 0.5

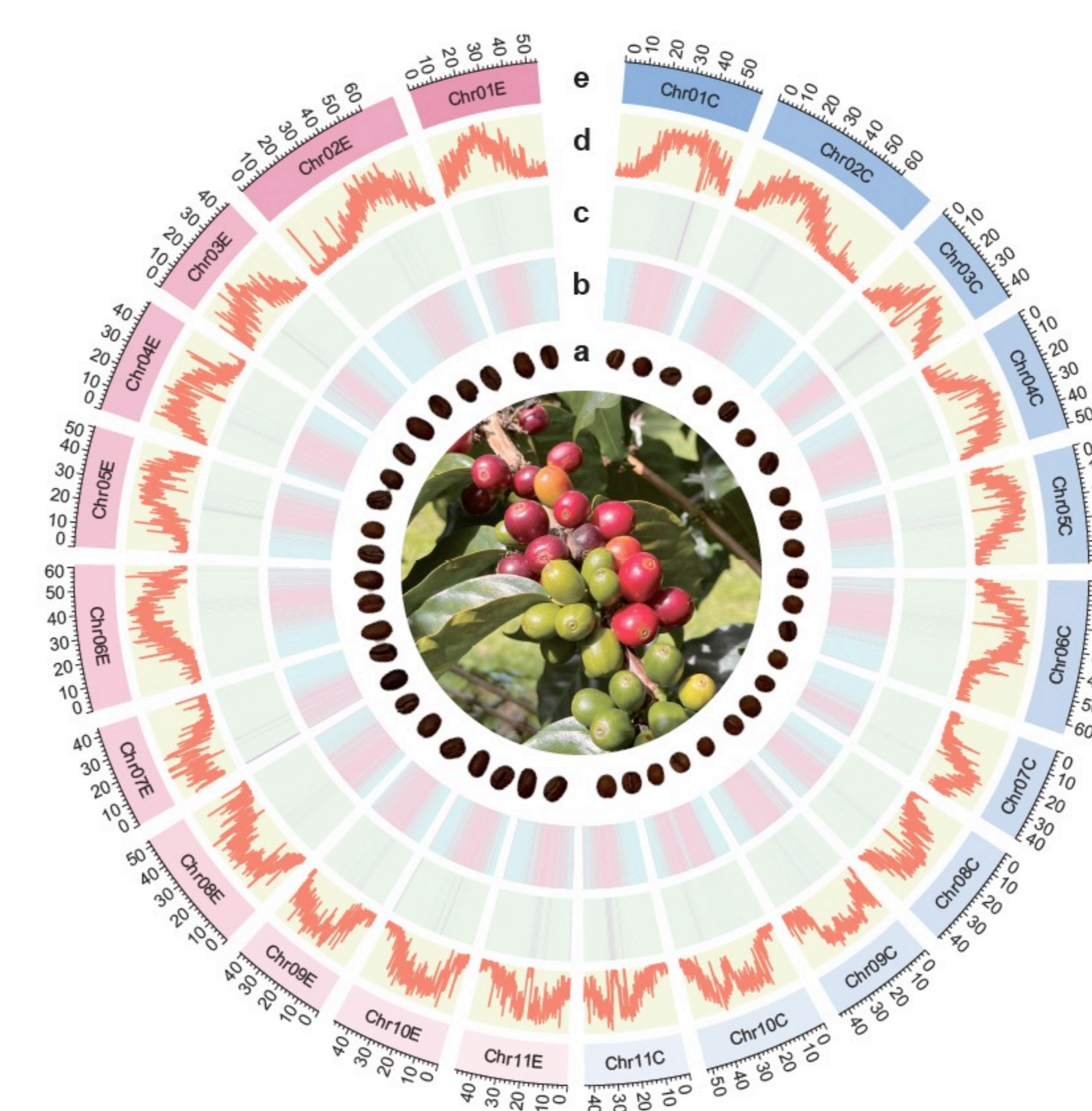


Fig. 1 Circos plot illustrating features of the chromosome-level assembly of 'Kona Typica'. **a.** Comparison of roasted Arabica (left) and Robusta (right) coffee beans. **b.** gene density (low to high, indicated by blue to red gradient); **c.** GC content; **d.** TE content; **e.** chromosomes, with megabase pair (Mbp) scale shown above. All density plots were generated using a 100 kbp sliding window size.

We identified structural variations between the Arabica coffee genomes of 'Kona Typica' and 'ET-39' (Fig. 2). A total of 88 inversions and 385 translocations were revealed, while 524 duplications and approximately 17.27 Mb genomic regions were unique to the 'Kona Typica' genome.

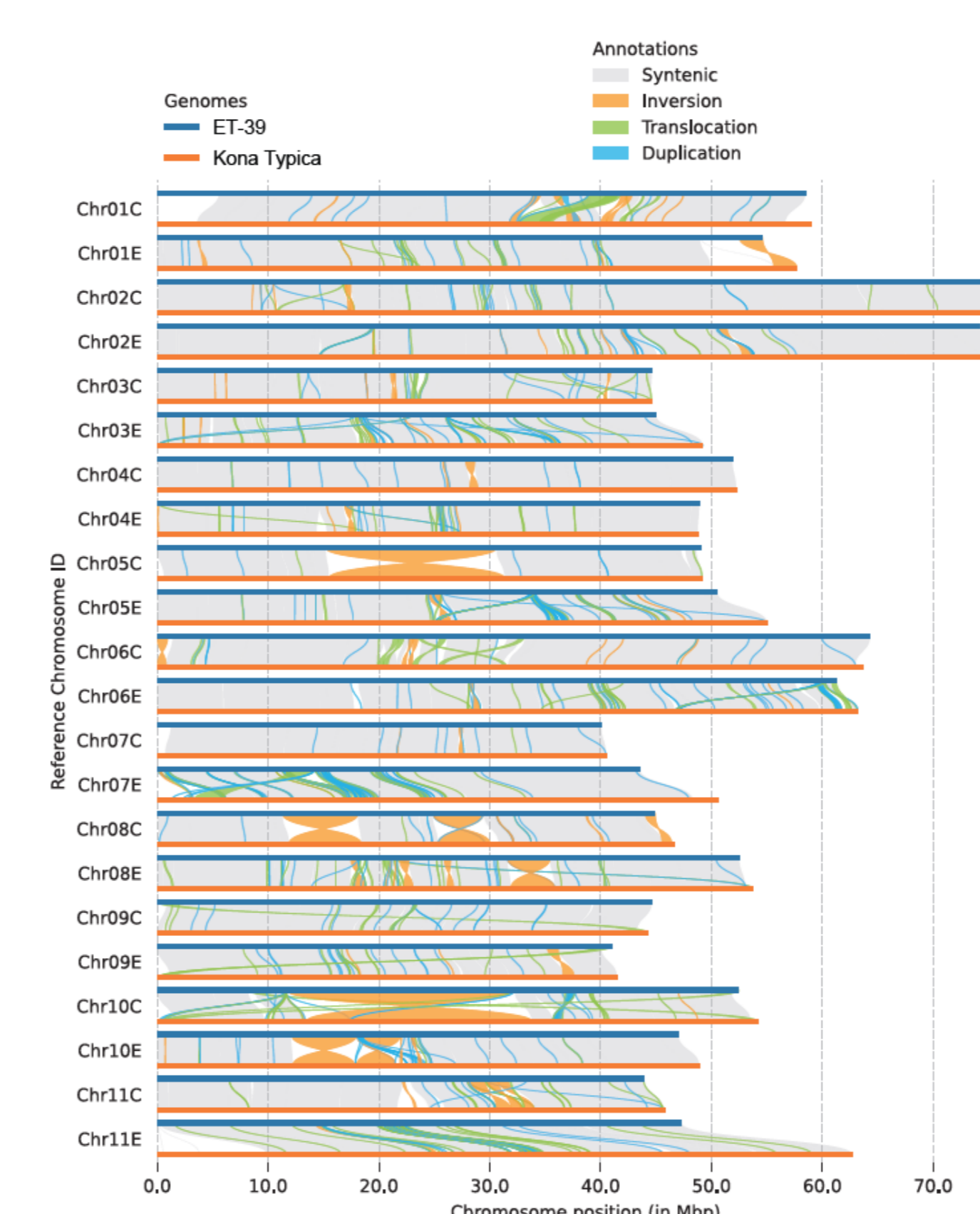


Fig. 2 Collinearity and structural variation analysis of the genome assembly of 'Kona Typica'. The ref indicates the 'ET-39' genome, while 'Typica' genome is the query.

Acknowledgements

This work was supported by National Institute of Food and Agriculture (NIFA) – Specialty Crop Research Initiative (SCRI) Grant 2022-51181-38241 to MW, TM, DZ, and QY.