GENETIC DIVERSITY, GENE DISCOVERY AND COMBINING ABILITY FOR YIELD AND FRUIT QUALITY TRAITS IN TOMATO (*Solanum lycopersicum* L.)

BY

YOSEF ALEMU BEKELE

(10804543)

THIS THESIS IS SUBMITTED TO THE UNIVERSITY OF GHANA, LEGON IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF DOCTOR OF PHILOSOPHY DEGREE IN PLANT BREEDING

WEST AFRICA CENTRE FOR CROP IMPROVEMENT

COLLEGE OF BASIC AND APPLIED SCIENCES

UNIVERSITY OF GHANA

LEGON

JULY 2024

DECLARATION

I hereby declare that this work is my original research and that all sources used for it have been duly acknowledged. I solemnly declare that this thesis is not submitted to any other institution anywhere for the award of academic degree.

YOSEF ALEMU BEKELE
(Student)

Prof. ERIC Y. DANQUAH
(Supervisor)

Prof. KWADWO OFORI
(Supervisor)

Dr. AGYEMANG DANQUAH
(Supervisor)

Dr. SELAMAWIT KETEMA
(Supervisor)

ABSTRACT

Tomato is a major vegetable crop in Ethiopia, contributing to nutrition, household income, and serving as a raw material for agro-processing and export markets. However, national tomato productivity remains low due to the prevalence of diseases, insect pests, and weeds. The shortage of improved varieties with high yield, superior fruit quality, and resistance to both biotic and abiotic stresses further limits production. Developing high-yielding OPVs and hybrids with superior fruit quality and tolerance to biotic and abiotic stresses is essential to enhance tomato productivity and utilization in Ethiopia. Therefore, the study aimed to develop high-yielding and fruit-quality varieties with large fruit size, and extended shelf life to contribute to improvement in productivity and utilization. The specific objectives were to: (i) assess the phenotypic diversity among tomato germplasm; (ii) determine the interrelationships among yield, yield-related, and fruit quality; (iii) determine molecular genetic diversity and population structure using SNP markers; (iv) identify QTLs or genomic regions associated with yield, yield-related, and fruit quality; (v) estimate combining ability and heterosis and identify parental lines for the production of hybrids that combine high yield and fruit quality. One hundred and forty-three (143) tomato genotypes were characterized for 15 quantitative and 18 qualitative traits for two seasons in 2021 and 2022. Most traits showed a highly significant difference ($p<0.01$) among tomato germplasm. The genotype-by-season interaction was also highly significant ($p<0.01$) for some traits. High (79–92%; 24–84%) to medium (41-46%; 12–13%) broad sense heritability and genetic advance as a percent of the mean, respectively, were detected for most traits. Fruit size traits (fruit weight, fruit length, fruit width, fruit shape index, and pericarp thickness), along with days to 50% flowering, fruit number per plant, fruits per cluster, and total soluble solids, showed strong genetic correlations with marketable and total yield. These correlations ranged from –0.552 to +0.55 and

were highly significant ($p < 0.01$). Principal component analysis showed that the first five components together accounted for 82.86% of the total variation. The first two components explained 51%, primarily contributed by fruits per plant, fruit weight, fruit width, fruit length, number of locules, and fruit skin thickness. Genetic diversity was also observed for 18 qualitative traits, including flower, growth pattern, and fruit traits. Genotypes were classified into three groups, with multiple correspondence analysis (MCA) biplots capturing 18.5% of the total variation. A molecular diversity analysis was conducted on 187 tomato germplasm, including 63 fresh-market, 112 processing, and 12 cherry types, using 4,729 high-quality SNP markers generated through DArTseq. The study revealed low to moderate gene diversity (0.04–0.50), polymorphic information content (0.04–0.37), and minor allele frequency (0.02–0.497). Two distinct population structures and admixtures comprising 6, 172, and 9 genotypes were detected. PCoA revealed two distinct genetic clusters and admixtures, with $PCoA_1$ (29.02%) and PCoA2 (10.78%) accounting for 39.8% of the total molecular genetic diversity. Analysis of molecular variance further elucidated high or moderate and highly significant genetic differentiation among the structured (FST = 0.496, $p \leq 0.001$) and pre-defined (FST = 0.134, $p \leq 0.001$) populations. The variation within a cluster (50%) was comparable to the variation accounted for among the population (50%) for the structured population; variation within the predefined population (87%) was much higher than among populations (13%). A genome-wide association study using 2,709 informative SNP markers identified 121 marker-trait associations ($-\log10p \geq 3.5$; FDR ≤ 0.05) related to ten traits. These associations corresponded to 104 QTLs, which were clustered based on linkage disequilibrium decay (0.157–28.826 Mbp at $r^2 = 0.2$) for each chromosome. Some of the identified QTLs were located in the vicinity of previously identified marker-trait associations, while others were new. A combining ability study was conducted using five male and five female

parental lines selected for phenotypic diversity in yield per plant, fruit size, fruit number per plant, fruit shape, and total soluble solids (TSS). The study revealed highly significant genetic variation ($p < 0.01$) among hybrids, parents, their specific combining ability (male × female), and interactions with the environment. The additive and dominance genetic variances were important, while the additive variance dominated for most traits. The highest positive midparent (81%) and better parent (75%) were recorded for fruit yield per plant for the hybrid G2xG7. Female parental line G2 was the best combiner for fruit yield and number per plant, and parental line G3 was the best for fruit number per plant. Male parental lines G7 and G10 were the best combiners for fruit weight, fruit wall thickness, and fruit width. Hybrids G2 x G7, G3 x G7, and G3 x G6 showed high positive specific combining ability effects, better performance *per se*, and heterosis over the mid and better parent for yield and related traits.

DEDICATION

I dedicate this work to my mother and my late father, whose nurturing guidance has been instrumental in helping me reach this stage of accomplishment. I also dedicate it to my wife, Hanna Fanta, and our children, Hemen Yosef and Yemariyam Yosef, whose unwavering support and encouragement have been invaluable throughout this journey.

ACKNOWLEDGMENTS

First and foremost, I am deeply grateful to Almighty God for His guidance and support throughout this journey.

I would like to express my heartfelt gratitude to my supervisors, Prof. Eric Danquah, Prof. Kwadwo Ofori, Dr. Agyemang Danquah, and Dr. Selamawit Ketema, for their unwavering support, encouragement, and constructive feedback. Despite their busy schedules, they generously dedicated significant time to critically reviewing and providing valuable comments on my manuscript, from the proposal stage to the final preparation of the thesis.

I am also sincerely thankful to the West African Centre for Crop Improvement (WACCI), the University of Ghana, and the German Academic Exchange Service (DAAD) for providing me with the opportunity to pursue this study and for their financial support. My gratitude also extends to the Ethiopian Institute of Agricultural Research (EIAR) and the Melkassa Agricultural Research Center (MARC) for granting study leave and access to research facilities.

I would like to acknowledge the staff of the Horticulture Department at MARC for their invaluable contributions to my research. In particular, I am grateful to Ms. Kokobe Taddese and Mrs. Tarikua Hailu for their assistance with field management and data collection.

I also wish to thank my friends, who contributed in various ways to the success of my research. Finally, I am deeply appreciative of my family for their constant encouragement, support, and hospitality throughout this journey.

TABLE OF CONTENTS

DECLARATION ---------- I

ABSTRACT ---------- II

DEDICATION ---------- V

ACKNOWLEDGMENTS ---------- VI

TABLE OF CONTENTS ---------- VII

LIST OF TABLES ---------- XIII

LIST OF FIGURES ---------- XVI

LIST OF ABBREVIATIONS ---------- XIX

CHAPTER ONE ---------- 1

1.0 GENERAL INTRODUCTION ---------- 1

CHAPTER TWO ---------- 6

2.0 LITERATURE REVIEW ---------- 6

2.1 Origin, domestication and genetic diversity of tomato ---------- 6

2.1.1 Origin and domestication ---------- 6

2.1.2 Genetic diversity of tomato ---------- 6

2.2 Genetics and breeding of tomato for yield and fruit quality ---------- 8

2.2.1 Genetics of yield and fruit quality ---------- 8

2.2.2 Floral biology and pollination mechanisms of tomato ---------- 11

2.2.3 Breeding objectives and approaches in tomato ---------- 13

2.3 Importance and methods of assessing genetic diversity ---------- 14

2.3.1 Morphological methods ---------- 15

2.3.2 Molecular methods ---------- 16

2.4 QTL mapping for yield and fruit quality ---------- 17

2.4.1 Genetic linkage analysis or family-based method ---------- 18

2.4.2 Association mapping using natural population ---------- 19

2.5 Factors affecting efficiency of QTL mapping ---------- 20

2.5.1 Population size and type of mapping population ---------- 20

2.5.2 Marker density, number and nature of markers ---------- 21

2.5.3 Methods of QTL mapping and statistical analysis ---------- 22

2.5.4 Level of precision in genotyping and phenotyping ---------- 23

2.5.5 Epistasis and QTL x E interaction ---------- 24

2.6 Hybrid breeding and heterosis for yield and fruit quality ---------- 24

2.6.1. Genetic basis of heterosis ---------- 25

2.6.2 Exploitation and manifestation of heterosis in tomato ---------- 27

2.6.3 Prediction of hybrid performance in tomato ---------- 28

2.6.3.1 Per se performance ---------- 28

2.6.3.2 Combining ability ---------- 29

2.6.3.3 Genetic distance ---------- 30

CHAPTER THREE ---------- 32

3.0 Phenotypic diversity among tomato (*Solanum lycopersicum* L.) germplasm ---------- 32

3.1 Introduction ---------- 32

3.2 Materials and methods ---------- 34

3.2.1 Study location and season ---------- 34

3.2.2 Experimental material and trial management ---------- 34

3.2.3 Data collected ---------- 35

3.2.4 Data analysis ---- 35
3.2.4.1 Genetic variances ---- 36
3.2.4.2 Broad sense heritability ---- 37
3.2.4.3 Genetic advance ---- 37
3.3.4.4 Genetic association among quantitative traits ---- 37
3.3.4.5 Cluster analysis ---- 38
3.2.4.6 Principal component analysis ---- 38
3.2.4.7 Qualitative traits ---- 38
3.3 Results ---- 39
3.3.1 Genetic variability for yield, yield related and fruit quality traits ---- 39
3.3.2 Variance components, heritability and genetic advance ---- 40
3.3.3 Genetic and phenotypic correlations among yield, yield related and fruit quality ---- 42
3.3.4 Hierarchical grouping of genotypes ---- 46
3.3.5. Principal components and relative contribution of traits ---- 48
3.3.6 Genetic variability based on qualitative traits ---- 50
3.3.6.1 Frequency distribution, chi-square test, and Shannon diversity index ---- 50
3.3.6.2 Cluster and multiple correspondence analyses (MCA) ---- 54
3.4 Discussion ---- 55
3.5 Conclusion ---- 61
CHAPTER FOUR ---- 63
4.0 Molecular genetic diversity and population structure of tomato germplasm ---- 63
4.1 Introduction ---- 63
4.2 Material and methods ---- 64

4.2.1 Plant materials, leaf sample preparation and high-throughput genotyping ---- 64
4.2.2 Data analysis ---- 66
4.2.2.1 Imputation and data filtering ---- 66
4.2.2.3 Genetic population structure analysis ---- 66
3.2.2.4 Genetic diversity parameters ---- 67
4.3 Result ---- 66
4.3.1 Level of SNP polymorphism ---- 72
4.3.2 Genetic diversity and population structure ---- 74
4.3.3 Genetic distance and weighted neighbor joining clustering ---- 75
4. 3.4 Analysis of molecular variance (AMOVA) ---- 77
4.4 Discussion ---- 80
4.5 Conclusion ---- 85
CHAPTER FIVE ---- 86
5.0 Genome wide association study for fruit yield and quality among tomato germplasm ---- 86
5.1 Introduction ---- 86
5.2 Material and methods ---- 88
5.2.1 Phenotyping ---- 88
5.2.2 Genotyping ---- 88
5.2.3 Imputation and SNP data filtering ---- 88
5.2.4 Linkage disequilibrium (LD) and decay ---- 89
5.2.5 Genome wide marker-trait association analysis ---- 89
5.3 Results ---- 90
5.3.1 Genetic variability ---- 90

5.3.2 Phenotypic correlations ---- 90

5.3.3 Population structure and kinship matrix ---- 91

5.3.5 Marker trait association ---- 93

5.3.5.1 Fruit length ---- 97

5.3.5.2 Fruit shape index ---- 98

5.3.5.3 Fruit width ---- 100

5.3.5.4 Fruit weight ---- 101

5.3.5.5 Marketable number per plant ---- 102

5.3.5.6 Number of locules ---- 104

5.3.5.7 Fruit skin thickness ---- 105

5.3.5.8 Plant height ---- 107

5.3.5.9 Total fruit number per plant ---- 107

5.3.5.10 Total soluble solid ---- 108

5.4 Discussion ---- 112

5.5 Conclusion ---- 120

CHAPTER SIX ---- 121

6.0 Combining ability and heterosis for yield and fruit quality in tomato ---- 121

6.1 Introduction ---- 121

6.2 Material and methods ---- 122

6.2.1 Study site ---- 122

6.2.2 Plant materials and progeny development ---- 123

6.2.3 Evaluation of progenies ---- 124

6.2.4 Data analysis ---- 124

6.3 Results ---------- 126
6.3.1 Genetic variability for yield, yield related and fruit quality ---------- 126
6.3.2 Components of genetic variance and heritability ---------- 129
6.3.3 Parental performance *pe se* and general combining ability effects ---------- 131
6.3.3.1 Yield and number of fruits per plant ---------- 131
6.3.3.2 Fruit weight and related traits ---------- 132
6.3.3.3 Plant height ---------- 133
6.3.3.4 Total soluble solid and juice volume ---------- 133
6.3.4 Hybrid performance *per se*, heterosis and specific combining ability effects ---------- 135
6.3.4.1 Yield and number of fruits per plant ---------- 135
6.3.4.2 Fruit weight and related traits ---------- 136
6.3.4.3 Total soluble solids and juice volume ---------- 137
6.3.4.4 Plant height and days to 50% flowering ---------- 138
6.4 Discussion ---------- 142
6.5 Conclusion ---------- 146
CHAPTER SEVEN ---------- 148
7 GENERAL CONCLUSIONS AND RECOMMENDATIONS ---------- 148
7.1 General conclusions ---------- 148
7.2 Recommendations ---------- 150
REFERENCES ---------- 152
APPENDICES ---------- 175

LIST OF TABLES

Table 3.1: Genetic variability of tomato germplasm based on 14 quantitative traits across seasons 41

Table 3.2: Genetic components of variance, heritability, and genetic advance of tomato genotypes across two seasons, 2021/2022 and 2022. 41

Table 3.3: Genotypic correlation (r_g) for yield, yield-related, and fruit quality 44

Table 3.4: Phenotypic correlation (r_p) for yield, yield-related, and fruit quality 45

Table 3.5: Characteristics of the three distinct clusters based on the mean value of the 15 quantitative traits 47

Table 3.6: The distribution of 143 tomato genotypes into three distinct clusters 47

Table 3.7: Eigenvectors and eigenvalues of the first five principal components for 15 quantitative traits of 143 tomato genotypes across seasons. 49

Table 3.8: Chi-square test and frequency distribution of flower traits 52

Table 3.9: Chi-square test and frequency distribution of growth pattern traits 52

Table 3.10: Chi-square test and frequency distribution of immature and mature stage fruit trait 53

Table 4.1: Types of SNP transition and transversion mutations 73

Table 4.2: Minor and major allele frequencies, gene diversity, observed heterozygosity, and polymorphic information content of 4729 SNPs across 187 tomato genotypes 74

Table 4.3: AMOVA for the structured and predefined population of tomato germplasm. 78

Table 4.4: Genetic diversity indices among the structured and pre-defined population of tomato germplasm 78

Table 4.5: The pairwise Nei genetic distance, genetic identity, gene flow, and FST among structured and pre-defined populations of tomato germplasm 80

Table 5.1: Range and broad-sense heritability of 14 quantitative traits of tomato genotypes..... 91

Table 5.2: Linkage disequilibrium (LD) decay for the 12 chromosomes (r^2 = 0.2) in tomato germplasm 95

Table 5.3: The number of MTAs and QTLs identified for 10 quantitative traits using three multi-locus models across 12 tomato chromosomes..... 96

Table 5.4: Common MTAs detected among traits..... 96

Table 5.5: Significant marker trait association for fruit length..... 97

Table 5.6: Significant marker trait associations for fruit shape index 99

Table 5.7: Significant marker trait associations for fruit width 100

Table 5.8: Significant marker trait associations for fruit weight 101

Table 5.9: Significant marker trait associations for marketable number per plant 103

Table 5.10: Significant marker trait associations for number of locules 105

FTable 5.11: Significant marker trait associations for fruit skin thickness 106

Table 5.12: Significant marker trait associations for plant height..... 108

Table 5.13: Significant marker trait association for total number of fruits per plant 110

Table 5.14: Significant marker trait associations for total soluble solids 111

Table 6.1: Description of the study area 123

Table 6.2: Description of parental lines used in the study..... 123

Table 6.3: Schematic of the North Carolina Design II (NCII) mating design, showing the 25 F_1 hybrids generated from a factorial cross between five male and five female parental lines……...124

Table 6.4: Means squares of 13 traits in tomato for two locations, MARC and DZARC, and combined 128

Table 6.5: Genetic components of variance and heritability of 13 traits in tomato for two locations, MARC and DzARC, and combined 130

Table 6.6*: Per se* performance of male and female parental lines with their respective general combining ability effects for 13 traits in tomato across two locations..... 134

Table 6.: Performance *per se* of parents and hybrids, the highest and lowest specific combining ability effects, mid parent and better parent heterosis for yield and quality 139

LIST OF FIGURES

Figure 3.1: Grouping of tomato genotypes into three distinct clusters. a) Gap statistics for determining the optimum number of clusters b) The three distinct clusters and genotype lists in each cluster. The red , green and blue colors indicate cluster I, cluster II and cluster II respectively; c) Biplots grouping genotypes into three: Blue is group 1, red is group 2, and orange is group 3. 48

Figure 3.2: a) Scree plot for the 10 principal components of 15 quantitative traits. b) cos2 shows the quality of representation of each trait in the principal component; a high value of cos2 indicates a good representation of the variable in the PCs, and these variables are the most important ones, while a low value of cos2 indicates the variable in the PCs is not perfectly represented. For a given variable, the sum of cos2 will be one if the variable in each PC is perfectly represented. 50

Figure 3.3: a) Average silhouette width for the optimum number of clusters; c) Dendrogram showing grouping of 143 tomato genotypes based on 18 qualitative traits. The blue, red and green colors showed cluster I, II and III respectively..... 54

Figure 3.4: Contribution of qualitative variables to the two PCs. 55

Figure 4.1: a) density and b) genome-wide distributions of SNPs across 12 chromosomes as determined by the DArTseq genotyping platform. 72

Figure 4.2: The proportion of nucleotide composition across the 12 chromosomes. 73

Figure 4.3: The genetic population structure of 187 tomato germplasm collections in Ethiopia. a) Structure harvester Delta K (Δk) = 2; b) a bar plot of the population structure of 187 tomato germplasm (red = cluster I, green = cluster II, and blue = admixtures); c) a weighted neighbor-joining tree; d) principal coordinates analysis (PCoA). 77

Figure 5.1: Heat map showing the association coefficients of 14 quantitative traits in tomato germplasm 92

Figure 5.2: Principal component analysis and kinship matrix for 127 tomato germplasm using 2709 SNP markers. a) PCA screen plot displaying the number of principal components and the percentage of variance explained. b) PCA of tomato germplasm based on two PCs; c) PCA based on three PCs, with each dot representing a single tomato genotype. d) a heat map of the kinship matrix between each pair of tomato germplasm. 93

Figure 5.3 :(a) linkage disequilibrium pattern deploying 2,709 SNP markers across 127 tomato germplasm collections in Ethiopia; b) plot of overall LD decay across 12 chromosomes at a fixed value of r^2 = 0.2, demonstrating the relationship between SNP base pair distance and LD value (r^2). The red line represents the genome-wide LD decay. The vertical (green) and horizontal (dark blue) lines represent the LD decay distance and half-LD value, respectively. LD decay equals 2.05 Mbp..... 95

Figure 5.4: a) The Manhattan plot and b) the Q-Q plot for fruit length using BLINK and Farm CPU models 98

Figure 5.5: a) Manhattan plot and b) Q-Q plot for fruit shape index using BLINK, FarmCPU and MLMM. 99

Figure 5.6: a) Manhattan plot and b) Q-Q plot for fruit width using BLINK and FarmCPU models 101

Figure 5.7: a) Manhattan plot and b) Q-Q plot for fruit weight using BLINK and FarmCPU models 102

Figure 5.8: a) Manhattan plot and b) Q-Q plot for marketable number of fruits per plant using BLINK, FarmCPU and MLMM models 104

Figure 5.9: a) Manhattan plot and b) Q-Q plot for number of locules using BLINK models 105

Figure 5.10: a) Manhattan plot and b) Q-Q plot for fruit skin thickness using BLINK, FarmCPU and MLMM models..... 106

Figure 5.11: a) Manhattan plot and b) Q-Q plot for plant height using BLINK, FarmCPU and MLMM models 109

Figure 5.12: a) Manhattan plot and b) Q-Q plot for total number of fruits per plant using BLINK, FarmCPU and MLMM models 111

Figure 5.13: a) Manhattan plot and b) Q-Q plot for TSS using BLINK, FarmCPU and MLMM models 112

LIST OF ABBREVIATIONS

AMOVA	Analysis of Molecular Variance
ANOVA	Analysis of Variance
Worldveg	World Vegetable Center
AFLP	Amplified Fragment Length Polymorphism
BP	Better Parent
BLUP	Best Linear Unbiased Predictor
BPH	Better Parent Heterosis
CSA	Central Statistics Agency
DART	Diversity Array Technology
DArTseq	Diversity Array Technology Sequencing
DzARC	Debrezeit Agricultural Research Center
FAOSTAT	Food and Agriculture Organization Statistics
F_1	First Filial Generation
FDR	False discovery rate
GAP	Good Agricultural Practices
GBS	Genotype-by-Sequencing
GCA	General Combining Ability
GCV	Genotypic Coefficient of Variance
GEI	Genotype by Environment Interaction
IPGRI	International Plant Genetic Resource Institute
ISSR	Inter Simple Sequence Repeat
Kg ha^{-1}	Kilogram per hectare

MAF	Minor Allele Frequency
MAR	Major Allele Frequency
MARC	Melkassa Agricultural Research Center
MATs	Marker Trait Associations
Mbp	Mega base pairs
MCA	Multiple Correspondence Analysis
MoANR	Ministry of Agriculture and Natural Resource
MP	Midparent
MPH	Midparent Heterosis
OPVs	Open Pollinated Varieties
PCA	Principal Component Analysis
PCoA	Principal Coordinate Analysis
PCV	Phenotypic Coefficient of Variance
PIC	Polymorphic Information Content
PVE	Percent variance explained
QTL	Quantitative Trait Loci
RAPD	Random Amplified Polymorphic DNA
SCA	Specific Combining Ability
SNP	Single Nucleotide Polymorphism
SSR	Simple Sequence Repeat
t ha^{-1}	ton per hectare
UPOV	International Union for the Protection of New Varieties of Plant

CHAPTER ONE

1.0 GENERAL INTRODUCTION

Tomato (*Solanum lycopersicum* L.) is an essential vegetable crop worldwide, mainly grown for nutrition and economic value and consumed as fresh and different processed products *(*Quinet *et al.*, 2019). It is the world's second most important vegetable after potato in terms of production (Sahu and Chattopadhyay, 2017; Quinet *et al*., 2019). It is an essential component of food and nutrition security in developing countries (Schreinemachers *et al*., 2018). Tomato is a rich source of vitamins A, C and E, minerals (potassium, magnesium, and calcium), lycopene and beta-carotene (Ali et al., 2021; Hedges & Lister, 2005; Le Van Trong et al., 2024). Globally, tomato production has been increasing, with over 5.17 million hectares of land being planted, with China, India, Turkey, the USA, and Italy being the top five producers. In Africa, 1.55 million hectares of tomato were planted, yielding 21.38 million metric tons (FAOSTAT, 2022).

Tomato is a diploid and highly self-pollinating species with a chromosome number of 2n = 2x = 24 and a genome size of 914 Mb (Barone *et al*., 2008). It belongs to the genus *Solanum* (Sahu & Chattopadhyay, 2017), the largest *Solanaceae* family, with over 2000 species distributed throughout tropical and subtropical regions (Kaunda & Zhang, 2019).

It is not well known when tomato was first introduced to Ethiopia, but cherry tomatoes have been grown in urban areas and small gardens (Desalegne, 2002). Over time, it has expanded to commercial production for domestic consumption, export markets, and agro-processing (Desalegne, 2002). Now, it is an important dietary ingredient, being a source of nutrition, income,

and foreign currency, and raw material for agro-industries (Ketema *et al*., 2017; Lule *et al*., 2024). It is produced by small-scale farmers and commercial farms, mainly under irrigated conditions. A total of 251,315 smallholder farmers were engaged in tomato production in 2022, covering 7,710 hectares from which 33,655 metric tons were produced (CSA, 2022).

Research on tomato in Ethiopia began in the 1970s, with the introduction of germplasm from different sources, mainly Worldveg, and screening and evaluation across locations for adaptation, yield, fruit quality, and resistance to diseases (Getachew *et al.*, 2021). Open-pollinated tomato varieties were released, and hybrid varieties were registered, contributing to productivity (MoANR, 2023). Open-pollinated varieties are declining in yield and have smaller fruit sizes than hybrid varieties. Although both OPVs and hybrid varieties have been produced, hybrid cultivars have become more popular and favored by producers due to their high yields and superior fruit qualities, such as large fruit size, firmness, shelf life, and market value. However, hybrid seed is expensive, and most farmers cannot afford to purchase it. Moreover, few of the hybrid varieties registered are accessible to farmers.

The national tomato productivity is significantly lower at 5.8 t ha^{-1} compared to the global average of 36.5 t ha^{-1}. The main production constraints include diseases (Gudero *et al*., 2018; Kesho & Tadesse, 2023; Wondirad *et al*., 2009), insect pests (Belete *et al*., 2024; Gashawbeza *et al*., 2009; Shiberu & Getu, 2017), parasitic weeds (Etagegnehu *et al*., 2009), limited availability of quality seed, a knowledge gap for Good Agricultural Practices (GAP), and limited improved varieties with high yield, fruit quality, and resistance to biotic and abiotic stresses (Mersha & Sime, 2022).

Improved tomato varieties (OPVs and hybrids) with a high yield and fruit quality, such as large fruit size, high firmness, long shelf life, and resistance to major diseases and insect pests, would contribute to improving production, productivity, and utilization in Ethiopia. Developing an efficient breeding strategy requires knowledge of the extent and pattern of genetic diversity and trait identification among tomato germplasm (Mohammadi and Prasanna, 2003; Bhanu, 2017; Xu *et al*., 2024). Morphological, cytological, biochemical, and molecular markers have been applied to analyze and understand genetic diversity in various crops (Nadeem *et al*., 2018). Currently, with the evolution of advanced genomic tools, molecular markers are widely applied in most crop plant species, including tomato (Bhanu, 2017; Chaudhary *et al*., 2024). Studies have been conducted to explore the genetic diversity of tomato through multivariate analysis of phenotypic traits (Mata-Nicolás *et al.*, 2020; Salim *et al.*, 2020) and molecular markers such as SSR (Eider *et al.*, 2020; Athinodorou *et al.*, 2021) and SNPs (Ayenan *et al*., 2021; Farinon *et al*., 2022; Mungai *et al*., 2021). In Ethiopia, few studies have been conducted using morphological traits involving a limited number of genotypes (Ene *et al*., 2022; Shushay *et al*., 2014). No scientific information is available on the molecular genetic diversity of tomato germplasm.

Molecular breeding methods such as marker-assisted selection and backcrossing have been applied in tomato breeding programs to supplement traditional breeding methods (Du *et al*., 2025; Joshi *et al*., 2011). Molecular markers that co-inherit with traits are first identified through linkage and genome-wide association studies (GWAS) and applied as an indirect selection technique to improve the efficiency and accuracy of quantitative trait selection. Such molecular markers identified in relation to agronomic traits, disease resistance, insect resistance, and abiotic resistance have been applied in marker assisted selection for many crops, including tomato (Du *et al*., 2025; Foolad, 2007).

The discovery of QTLs that account for a significant fraction of the phenotypic variation of tomato horticultural traits has assisted in the development of varieties that meet consumer demands and market niches (Adhikari *et al*., 2020). The identification of QTLs is significantly influenced by genetic background and interactions (GEI, epistasis), suggesting the significance of identifying specific or stable QTLs for effective marker-assisted selection (Jahani *et al*., 2019; YAO *et al*., 2016). Due to the polygenic nature of quantitative traits, more loci need to be scrutinized to fully understand their genetic bases (Adhikari *et al*., 2020). Furthermore, no comparable study on tomato germplasm from Ethiopia has been reported.

Since breeding for pure line or open-pollinated varieties has limited potential, one strategy for improving yield and fruit quality in tomato is to breed for heterosis (Tamta & Singh, 2018). Even though tomato is a self-pollinating crop, manifestations of heterosis for yield reach as high as 33 to 49% (Izzo *et al*., 2022). The development of hybrid varieties is resource-intensive, and high heterosis is not guaranteed in all parental combinations (Liu *et al*., 2021). Therefore, the use of appropriate predictive methods is crucial for the efficient identification of parental lines with high combining ability.

Main objective

- To contribute to the development of high-yielding, quality tomato varieties to enhance productivity and utilization in Ethiopia.

Specific objectives

1. To characterize tomato genotypes based on agro-morphological traits.
2. To analyze the interrelationships between agro-morphological and fruit quality traits.

3. To assess the genetic diversity and population structure of tomato genotypes using SNP markers.
4. To identify quantitative trait loci (QTLs) for yield, yield-related, and fruit quality traits.
5. To estimate combining ability and heterosis for yield and fruit quality.
6. To evaluate and identify superior parental lines for the development of high-yielding, high-quality hybrids.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Origin, domestication and genetic diversity of tomato

2.1.1 Origin and domestication

Modern tomato varieties found worldwide that vary in plant and fruit characteristics are the result of domestication, selection, and breeding (Bai & Lindhout, 2007; Lin *et al*., 2014). Domestication of wild species of crop plants, including tomato resulted in a significant change in the phenotype of various traits due to alterations in the genetic constitution, the domestication syndrome (John *et al*., 2006; Razifard *et al.*, 2020). The study of variation in both the wild and cultivated species offers an understanding of the historical process of domestication, adaptation, and breeding. Furthermore, it is important for effective and efficient germplasm management and exploitation in crop improvement programs.

Tomato is originated in the Andean, which nowadays includes parts of Chile, Bolivia, Ecuador, Peru, and Colombia (Mata-Nicolás *et al*., 2020b). Pre-domestication of tomato also started in the Andean region. The exact time and place of tomato domestication have been a matter of exciting debate for centuries (Bai and Lindhout, 2007; Razifard *et al*., 2020); for the Peruvian and Mexican centers of domestication. Historical records, anthropology, and genetic inferences indicate Mexico as the probable place of tomato domestication. *Solanum lycopersicum* var. *cerasiforme* (SLC) species and *Solanum pimpinellifolium* (SP) are the ancestors of cultivated tomato. SLC evolved from SP in Peru and Ecuador (Salim *et al*., 2020) and later spread northward, passing through secondary diversification to give rise to the cultivated tomato (Harlan, 1928) in Mexico. This was further supported through genome analysis (Razifard *et al*., 2020). Extensive domestication

continued in Europe in the 18th and 19th centuries (Sims, 1980); then, tomatoes became popular and diversified to the rest of the world. Selection for traits such as fruit color, size, and shape, including plant architecture (Ichihashi & Sinha, 2014), following the domestication process extensively reduced the genetic diversity of the cultivated tomato compared to its wild relatives (Du et al., 2025; Tieman *et al*., 2017). Since the early 20th century, a large number of tomato varieties have been developed that vary in fruit shape, size, and color (Bai and Lindhout, 2007; Du *et al*., 2025; Lin *et al.*, 2014).

2.1.2 Genetic diversity of tomato

Tomato (*Solanum lycopersicum* L.) is a member of the *Solanaceae* family and genus *Solanum* (Sahu & Chattopadhyay, 2017). The *Solanaceae* family comprises over 3000 species (Bai and Lindhout 2007). *Solanum* is the largest in the *Solanaceae* family, with over 2000 species distributed throughout tropical and subtropical regions (Kaunda & Zhang, 2019). The *lycopersicum* section in this genus consists of 13 species, 12 of which are wild relatives, and *Solanum lycopersicum* var. *lycopersicum*, the only cultivated species (Peralta and Spooner 2000; Ramírez-Ojeda *et al.*, 2021; Du *et al*., 2025). The self-incompatibility, outcrossing nature, and adaptation mechanisms, growing from sea level to over 3300 m, lead some wild species to harbor wide genetic diversity (Peralta and Spooner 2000; Sahu and Chattopadhyay 2017). Molecular diversity analysis indicated higher genetic diversity within wild tomato species than the cultivated tomato (Sahu and Chattopadhyay 2017; Tieman *et al*., 2017; Zhou *et al*. 2015). *Solanum pimpinellifolium* L. (SP), for example, retains huge genetic variation for flower and inflorescence size, fruit color, and extent of style exertion. The genetic variation in the cultivated tomato is less than 5% of the genetic variation available in its wild relatives.

Wild tomato species have played a significant role as sources of disease resistance, fruit color, and other desirable traits for tomato varieties worldwide (Mata-Nicolás *et al*., 2020a; Pessoa *et al*., 2023). The availability of huge genetic variation in *Solanum pimpinellifolium* L. (SP), for instance, and the possibility of cross-compatibility with the cultivated tomato make this wild species an essential source or reservoir of important traits (Rick *et al*., 1977; Lin *et al*., 2014; Wang *et al*., 2020). Improvements in fruit quality and resistant genes for biotic and abiotic stresses were transferred from *Solanum pimpinellifolium* L (SP). Fruit quality traits such as solid content, firmness, color, volatile compounds (Capel *et al*., 2015; Mata-Nicolás *et al*., 2020; Rambla *et al*., 2017), resistant genes for fungi, tomato leaf curl virus (Banerjee & Kalloo, 1987), *Alternaria solan*i, *Fusarium oxysporum*, and *Phytophtora infestans* (Hogenboom, 1972), were extensively introduced from wild species into the cultivated tomato. Over 82,393 tomato germplasm, mainly cultivated and wild species is maintained in various gene banks around the world (Ebert & Chou, 2015).

2.2 Genetics and breeding of tomato for yield and fruit quality

2.2.1 Genetics of yield and fruit quality

Since the early 20th century, different varieties of the cultivated tomato, varying in fruit and plant architecture, have been developed through breeding (Bai and Lindhout, 2007; Gao *et al*., 2019). Information on the genetic architecture, such as the extent and nature of gene action and interaction, number of genes, location, heritability, and genetic gain, is essential to devising an appropriate breeding procedure for their improvement (Said, 2014; Verma & Singh, 2018).

Fruit yield, which is quantitatively inherited and the primary objective of the tomato improvement program, is a complex trait controlled by the action and interaction of numerous genes and

environmental factors (Hernández-Bautista *et al.*, 2015; Jiang *et al*., 2013). Both additive and non-additive genetic components of variance are important in the inheritance of yield and its components (Goffar *et al.*, 2016).

The process of fruit formation in fruit-bearing plants like tomato, affects the variation in fruit yield. Tomato fruit yield is a function of different yield components (Rashidi *et al.*, 2009; Monamodi *et al.*, 2013; Ye *et al*., 2021), including fruit weight, length, diameter, and fruit wall thickness. These traits are correlated with fruit yield in tomato; further affected by inflorescence and fruit developmental processes that regulate cell layer numbers and volumes in the pericarps to determine fruit size. The number and volume of cell layers in the tomato fruit pericarp affect tomato fruit size (Sun *et al.,* 2023), which are functions of the frequency of cell division and expansion and duration of the cell cycle in the fertilized ovary (Ariizumi *et al.*, 2013). A longer period of cell division will result in larger fruits. Depending on the variety, the pericarp cell volume in certain fruits at the green mature stage was between 2000 and 22,000 times bigger than what was initially obtained during the ovary wall of the pre-anthesis (Cheniclet *et al.*, 2005). This information suggests that the size of the pericarp and pericarp thickness predominantly affect fruit size. The main determinants of tomato fruit yield are the capacity to bear fruit and the final cell number and fruit size (Ariizumi *et al.*, 2013). Furthermore, average fruit weight, fruit diameter, and fruit length have the strongest effects on tomato fruit yield. A small effect was also observed from % TSS and number of locules (Hernández-Bautista *et al.*, 2015). A molecular study of fruit yield indicated that it is largely determined by fruit size QTLs, with the remaining factors playing an additional role (Hernández-Bautista *et al.*, 2015; Ye *et al.,* 2021).

Fruit size and shape are important traits in many crops that affect yield, quality, marketability, and consumer preferences (Bertin, 2018). Despite a domestication-induced loss of genetic diversity for traits like stress tolerance and nutrition in tomato (Lin *et al*., 2014), human selection has driven a dramatic increase in variation for fruit size and shape. There is great variation in fruit size among wild and cultivated tomato species; wild species produce fruits as small as a pea, while cultivated varieties can reach up to 500 grams with diverse shapes (Grandillo *et al.*, 1999; Lippman & Tanksley, 2001). These traits are quantitatively inherited, and the complexity of this inheritance makes their understanding difficult at the molecular level.

QTL mapping analyses using populations derived from crosses between wild and cultivated tomatoes of varying sizes have indicated that over 30 QTLs collectively account for the majority of genetic variation in fruit size and shape, although their individual effects are not equal (Grandillo *et al*., 1999). Among these, approximately ten QTLs with large effects are responsible for major variations in fruit size. It is estimated that mutations in just six key loci were sufficient to drive the evolutionary transition from the small berries of wild progenitors to the large fruits of modern cultivated tomatoes. Notable among these are allelic variations at four specific loci fw1.1, fw2.2, fw3.1, and fw4.1 which were identified through QTL mapping of crosses between small-fruited wild and large-fruited cultivated species. These loci individually can have a major impact on fruit size, with some accounting for up to 30% of its variation. Furthermore, there is no clear distinction between QTLs governing fruit size and those governing shape in tomato, as their genetic control is often interconnected. The three major loci that play a predominant role in determining fruit shape are OVATE, SUN, and fs8.1, located on chromosomes 2 and 8 (Tanksley, 2004).

Fruit quality traits including pH, total soluble solids (TSS), color, viscosity, and lycopene content are critical determinants of tomato processing quality (Renquist & Reid, 1998). The expression of these traits, such as pH, %TSS, and lycopene content, is governed by both additive and non-additive gene effects (Dhaliwal & Chahal, 2005; Gao *et al*., 2019; Tieman *et al*., 2017; Ye *et al*., 2021).

Multiple studies have identified key quantitative trait loci (QTLs) governing tomato fruit morphology and quality. A genome-wide association study (GWAS) of 192 core tomato collections identified 41 QTLs for six fruit traits, explaining 7.11–37.64% of the phenotypic variance for fruit color (2 QTLs), shape (7), pericarp thickness (11), weight (13), width (10), and length (7) (Phan *et al*., 2019). Similarly, QTL mapping of an F2 population from a cross between *Solanum lycopersicum* and Solanum *pimpinellifolium* revealed approximately 25 significant QTLs controlling fruit length, diameter, weight, locule number, and Brix degrees (Hernández-Bautista *et al*., 2015). Complementing these mapping efforts, research on 368 cultivated and cherry tomato (*Solanum lycopersicum* var. *cerasiforme*) accessions has pinpointed four major genes underlying these QTLs: SUN and OVATE, which control elongated fruit shape, and FASCIATED (FAS) and LOCULE NUMBER (LN), which regulate flat fruit shape and locule number (Rodríguez *et al*., 2011).

2.2.2 Floral biology and pollination mechanisms of tomato

Understanding floral biology and pollination mechanisms is essential for selecting efficient breeding methods. The flower of the tomato occurs in a simple or complex inflorescence. The type of inflorescence could be cyme or raceme, but in most cases, it is cyme (Kaul, 1991). The number of flowers varies depending on the growing and environmental conditions. The reproductive part

of the plant comprises both the stamen and the carpel. Once the stamen produces the pollen grain, it will self-pollinate the female reproductive organ and get fertilized. Most commercial cultivars of tomato are self-pollinating; the stamens are found in the petal part of the flower, where each stamen fused surrounding the carpel, forming a yellow cylindrical shape favoring self-pollination of up to 94 to 99% (Kaul, 1991; Toni *et al*., 2020). Rarely, cross-pollination occurs with the help of insects like bumble bees, which carry pollen grains from one plant to another. Limited pollen is found on the surface of the anther but rather inside the anther in a hollow tube structure, resulting in the pollen being easily shaded. Pollen shade needs to be aided by outside forces and mechanisms like wind, insects, and an artificial vibrator, as was confirmed in greenhouse tomato production. The reproductive structures of tomato are highly responsive to high temperatures (Singh *et al*., 2016).

Most wild tomato species cross pollinate due to a gametophytic self-incompatibility system, which prevents self-fertilization (Broz *et al*., 2021; Kaul, 1991). In some of these species, the pistil is physically exerted, extending beyond the male reproductive organs (stamens), a floral trait that further promotes cross-pollination. Furthermore, male sterility has been reported in certain wild relatives, which also enforces outcrossing. This reliance on cross-pollination was eliminated during the domestication of the cultivated tomato. As the crop migrated from its center of origin, its specific pollinators were absent. In this context, self-fertility provided a significant selective advantage by guaranteeing reliable fruit set, leading to the fixation of self-compatible alleles in modern cultivated varieties (Singh *et al*., 2016).

2.2.3 Breeding objectives and approaches in tomato

Fruit yield is the principal objective in tomato breeding (Yeon *et al*., 2024). However, breeding programs also prioritize other critical traits, including resistance to biotic and abiotic stresses, fruit quality, and appearance (external color, size, shape, uniformity). Additional key components are texture, firmness, flavor determined by a balance of sugars and acids—and nutritive value, such as vitamin, carotene, and lycopene content (Abewoy Fentik, 2017; Abhary & Rezk, 2021; Sharma et al., 2019).

The global tomato breeding programs have benefited greatly from both conventional and molecular breeding approaches (Du *et al*., 2025; Joshi *et al*., 2011). Conventional methods include germplasm introduction and adaptability studies, pure line and mass selection from landraces, and hybridization followed by selection using approaches such as pedigree, single-seed descent, and bulk methods, with hybridization followed by pedigree selection being the most commonly used. Interspecific hybridization (Jia *et al*., 2025; Vanlay *et al*., 2022), heterosis, and mutation breeding (Chaudhary *et al*., 2019) have also been widely applied in tomato improvement programs. Trait introgression from wild relatives has become increasingly important due to the loss of genetic diversity in cultivated tomato during domestication. Accordingly, backcross breeding and interspecific hybridization have been employed, particularly to transfer traits related to biotic (disease) and abiotic (salinity, drought) stress resistance (Abewoy Fentik, 2017; Abhary & Rezk, 2021b; Sharma *et al*., 2019;)

Marker-assisted selection and backcrossing are molecular breeding approaches applied in tomato breeding to complement conventional breeding methods (Hasan, 2021; Joshi *et al*., 2011). Molecular markers that co-inherit with traits are first identified and applied as an indirect selection

technique. Such molecular markers identified for disease resistance, insect resistance, and abiotic resistance have been applied in MAS for crops including tomato in developed countries (Foolad, 2007), but MAS is still in its infant stage in Africa, requiring a bigger outlook and investment. Similar techniques, such as marker-assisted back-crossing and recurrent selection, have been employed in tomato improvement programs in developed countries. These breeding techniques reduce costs and generation time, increase selection precision and efficiency, and accelerate genetic or selection gain (Osei *et al.*, 2019).

2.3 Importance and methods of assessing genetic diversity

Genetic diversity is the extent of genetic variation among individuals in a population or species (Brown, 1983; EBSCO Research Starters, 2023; Molecular Ecology & Evolution, 2024). Genetic diversity is essential for the survival of plant populations or species and crop improvement programs (Bahjat *et al.*, 2025; Govindaraj *et al.*, 2015; Henderson & Salt, 2017). Crop improvement relies on the extent and nature of available plant genetic resources to generate new varieties and improve the existing ones (Taylor & Duvick, 2008). Information on the amount and pattern of genetic variation among germplasms is essential, especially in the early stages of plant breeding, which helps to design breeding strategies (Da Silva *et al.*, 2017; Bhanu, 2017). From the beginning of agriculture, the genetic variability of crop species contributed to the improvement of basic food requirements (Govindaraj *et al.*, 2015). Genetic characterization is the detection of variation in populations that is due to differences in DNA sequence, specific genes, or modifying factors (Hanotte & Jianlin, 2005).

Various factors, including sexual recombination, mutation (gene mutation and epi-mutation), and gene flow, affect the genetic diversity of crop species (Ellegren & Galtier, 2016), with sexual

recombination being the primary cause (Bhanu, 2017). Selection (natural or artificial) and genetic drift are other factors affecting diversity through modifying allele frequency in populations. Selection, some forms of mating systems (inbreeding), and genetic drift reduce genetic diversity in populations, while mutation, outcrossing, and migration increase genetic diversity (Bhanu, 2017).

Morphological and molecular markers have been used to analyze the extent of genetic diversity in tomato (Ayenan *et al.*, 2021; Eider *et al.*, 2020; Ene *et al.*, 2022; Nadeem *et al.*, 2018; Salim *et al.*, 2020).

2.3.1 Morphological methods

Morphological markers, which include both qualitative and quantitative traits, are easily applicable as they require no advanced biochemical or molecular techniques (Kıymacı *et al*., 2024; Nadeem *et al*., 2018). Although they have been widely used to assess genetic diversity in numerous crops, their utility is constrained. Key limitations include their finite number and the fact that their expression can be significantly influenced by environmental conditions and plant developmental stage (Beyene *et al*., 2005).

Morphological characterization is essential for identifying and selecting promising genotypes with desirable traits for subsequent breeding and variety development programs (Salim *et al*., 2020; Grozeva *et al*., 2021). Numerous genetic diversity studies focusing on morphological, biotic, and abiotic stress traits have been conducted in tomato using phenotypic descriptors (Jáquez-Gutiérrez *et al*., 2019). Quantitative and qualitative morphological descriptors and markers including vegetative, developmental, and fruit traits—revealed considerable variation among 22 tomato inbred lines. This study also emphasized the importance of using a combination of morphological

traits to capture substantial variability rather than relying on a single character. Morphological traits such as hypocotyl color, hypocotyl pubescence, leaf type, and green shoulder tips proved to be valuable diagnostic characteristics (Salim *et al*., 2020). Similarly, phenotypic characterization of 52 tomato genotypes showed significant genetic variation in fruit weight, fruit number per plant, soluble solids, antioxidant activity, vitamin C, and total phenols (Grozeva *et al*., 2021). Moreover, significant genetic variation was detected for plant height, leaf width, leaf length, leaf area, fruit number per cluster, cluster number per plant, fruit per plant, fruit length, fruit width, taste, fruit weight, and fruit yield per plant (Hassan *et al*., 2021).

2.3.2 Molecular methods

Molecular markers are widely applied for genetic diversity studies (Atalawi *et al.,* 2025; Bhanu, 2017). These are nucleotide sequence polymorphisms present in different individuals that could be due to insertions, deletions, duplications, or translocations (Mondini *et al.*, 2009; Nadeem *et al.*, 2018). Molecular markers provide better power for the detection of the available diversity as they are not influenced by the environment. DNA markers such as RAPD, AFLP, ISSR, SSR, SNP, DART, and retrotransposons have been applied in genetic diversity assessment, each with its strengths and weaknesses (Hussain *et al*., 2025; Nadeem *et al.*, 2018). Among these, those that are codominant, evenly distributed throughout the genome, reproducible, and have the capacity to detect a large extent of polymorphism are preferred and applied in molecular diversity analysis and marker-assisted selection of different crops (Mondini *et al.*, 2009).

SNP markers are high-throughput and cost-effectively applied in map construction, genetic diversity analysis, genome-wide marker-trait association, marker-assisted selection, and genomic selection (Bahjat *et al*., 2025; Song *et al.*, 2023; Ganal *et al.*, 2009). SNPs have been preferred

because of their high level of polymorphism, wide genome coverage, locus specificity, high reproducibility, and fixed physical position on the chromosome (Yan *et al*., 2010). Information on phenotypic and molecular genetic diversity is beneficial for a better understanding of the organization of crop genetic diversity (Fisseha *et al.*, 2018).

Genetic diversity studies using molecular markers were conducted in tomato; including SSR (Gharsallah *et al.*, 2016; Hussain *et al*., 2025; Kaushal *et al.*, 2017; Korir *et al.*, 2014; Kovach & Egertsdotter, 2009; Kwon *et al.*, 2009), and SNPs (Celik *et al.*, 2017; Farinon *et al.*, 2022; Mungai *et al.*, 2021; Sim *et al.*, 2011; Mungai *et al.*, 2021) and contributed to tomato breeding programs.

2.4 QTL mapping for yield and fruit quality

Most traits of crops are controlled by many genes, whose effects are small but cumulative following quantitative inheritance. It is difficult to associate the observed phenotype with the genotypic values as they are affected by environment, QTLs by environment interaction, and epistasis (Hobby & Zhang, 2025; Semagn & Bjørnstad, 2010). Identification of genes or QTLs controlling quantitative traits has been difficult for genetics (Price, 2006).

Linkage analysis and association, or linkage disequilibrium mapping, are the two commonly applied QTL mapping methods to dissect quantitative traits in various cultivated crops. These methods have thus far identified molecular markers associated with QTLs that reveal the proportion of observed phenotypic variance in crops (Chen & Zhang, 2024; Xu *et al.*, 2017 Xiang *et al*., 2025). The two methods differ in mapping resolution and power, influenced by mapping populations. Linkage mapping is family-based mapping that depends on recombination during the development of the bi-parental or multi-parent mapping populations; association mapping uses natural populations with unknown relationships (Xu *et al.*, 2017).

2.4.1 Genetic linkage analysis or family-based method

QTL analysis aims to identify QTLs, their actions, interactions, and positions on chromosomes. In quantitative genetic analysis, it was possible to determine the combined effect of all QTLs responsible for the trait of interest; it has been difficult to determine the effects of each QTL. Currently, it is possible to identify molecular markers associated with a trait and individual QTL effects (Kearsey, 1998; Xiang *et al*., 2025). Linkage analysis have been conducted to dissect the genetic basis of quantitative traits of many crops (Sandhu *et al*., 2023; Yang *et al*., 2024) but only a limited number of QTLs identified have been tagged or cloned at the gene level (Price, 2006).

The initial step in linkage analysis is the development of a bi-parental or multi-parent population. The bi-parental populations are F_2, backcross, double haploids, recombinant inbreds, or near-isogeneic lines, each with strengths and weaknesses. A bi-parental population could be produced through the hybridization of diverse parents for the trait of interest, the phenotyping and genotyping of the mapping population, and the identification of QTLs using appropriate statistical procedures (Kearsey, 1998). The bi-parental population consists of very few genetic recombination events, resulting in low genetic variation and representing a small portion of the particular species used for mapping compared to the other population types. The low recombination frequency in the bi-parental population affects the QTL detection power, which is limited to 10 cM to 20 cM localization of QTL from a marker (Xu *et al.*, 2017; Zhu *et al.*, 2008a)

Multi-parent populations such as the NAM and MAGIC have higher genetic diversity, which leads to higher resolution and better QTL mapping. The MAGIC population, because it is developed through the intermating of many parental lines for multiple generations before the development of

inbred lines used for mapping, creates higher genetic variance and improves the precision of QTL mapping (Xu *et al*. 2017).

2.4.2 Association mapping using natural population

Association mapping, a new alternative approach that exploits historic recombination accumulated over several generations, gives the advantage of higher resolution and allele numbers than the traditional family-based mapping approaches (Zhu *et al.*, 2008). Association mapping uses natural populations where higher allelic richness, higher genetic diversity, and higher historical recombination are available, which provides a higher resolution for QTL mapping. The power of QTL mapping in this population depends on the extent of linkage disequilibrium, linkage decay, population structure, relatedness, population size, and allele frequency (Xu *et al*., 2017).

The main steps in natural population QTL mapping include the collection of a sample population that consists of landraces, wild relatives, released cultivars, and breeding lines; phenotyping of the natural population for the trait of interest; determination of heritability; genotyping using appropriate markers and genotyping platforms; estimating linkage disequilibrium, population structure and kinship; and finally, determination of the extent of association between the genotype and phenotype using appropriate statistical models.

Since their introduction to maize genetics (Thornsberry *et al.*, 2001), genome-wide association studies have been applied effectively to identifying genetic variants associated with a trait for many plant species, including rice, maize, wheat, barley, tomato, sorghum, soybean, watermelon, and other important crops (Flint-Garcia *et al.*, 2005; Getahun *et al*., 2025; Sandhu *et al*., 2023; Thornsberry *et al*., 2001; Yang *et al*., 2024). Similarly, GWAS was conducted in tomato to identify molecular markers associated with heat (Ruggieri *et al*., 2019); resistance to early blight (Chaerani

et al., 2007); fruit color (Mata-Nicolás *et al.*, 2020a); fruit dry weight, pH, TSS, vitamin C, sugar content, organic acid content (Albert *et al*., 2016); nutritional composition, ascorbic acid, fruit diameter and fruit weight (Zhang *et al.*, 2016); fruit flavor (Zhang *et al*., 2015); fruit metabolic traits including amino acid, ascorbic acid, and sugar content (Sauvage *et al.*, 2014); fruit shape, color, pericarp thickness, fruit weight, fruit height, and fruit width (Phan *et al.*, 2019).

2.5 Factors affecting efficiency of QTL mapping

Several factors could affect the efficiency and reliability of quantitative trait locus (QTL) mapping. These include population size, nature and type of mapping population, marker density and number, the position and effect size of QTLs, heritability of the trait and QTLs, precision in phenotyping and genotyping (error control), level of epistasis, QTL x environment interaction (Q x E), mapping approaches, model efficiency, and QTL (data) analysis.

2.5.1 Population size and type of mapping population

Simulation studies indicated that a large sample size improves the accuracy of QTL mapping. It also affects the number of QTLs, the level of additive genetic variance, and the total genetic variance. A large population size enables the detection of a higher number of QTLs and better detection capacity. On the other hand, small population sizes in QTL mapping result in reduced sensitivity of QTL detection, overestimation of QTL effects, and failure to detect QTL interactions (Vales *et al*., 2005). Moreover, it has limitations in detecting the precise position of QTLs on chromosomes (Raghavan, 2012).

The type of population applied in QTL mapping also affects QTL detection power, location on chromosomes, resolution, and precision of QTL mapping. The low recombination frequency in the bi-parental population, for example, affects the QTL detection power and enables QTL localization

further from the marker, such as from 10 cM to 20 cM. Multi-parent populations such as the NAM, MAGIC, and natural populations, on the other hand, have higher genetic diversity, which leads to higher resolution and better power and precision of QTL mapping (Xu *et al*., 2017).

2.5.2 Marker density, number and nature of markers

Marker density and the type of markers used in QTL mapping also affect its accuracy, indicating that optimum marker density needs to be used (Piepho, 2000). If a marker is far from the QTL, the power of detection decreases (Piepho, 2000). However, marker density beyond the optimum will not affect QTL mapping accuracy. A study on a double haploid population of maize generated from a bi-parental cross population using marker densities of 1 cM, 2 cM, and 5 cM indicated that high marker density neither improved nor reduced the QTL detection and genetic variance explained. However, the accuracy of the QTL location, the effect size of QTL, especially for large and small effect sizes, and the capacity to identify tightly linked QTL improved with increasing marker density from 5 cM to 1 cM (Stange *et al*., 2013).

The effect of marker density on QTL mapping depends on model types; for example, the marker density could not affect efficiency under the additive model while significantly affecting accuracy under the non-additive and epistatic models. Further, the marker density depends on the recombination frequency between the markers, genetic diversity, and linkage disequilibrium decay (LD). For example, cross-pollinated crops like maize possess high genetic diversity due to outcrossing and, hence, are in higher LD decay, requiring a higher marker density to detect significant associations. Whereas self-pollinating crops such as tomato have lower genetic diversity and low LD decay, which suggests lower marker density can detect significant marker associations. Generally, the use of optimum marker density depending on recombination

frequency, linkage disequilibrium, linkage disequilibrium decay, population type and size, crop species, QTL effect size of the trait of interest, method of QTL analysis, and model type to be used need to be considered during QTL mapping.

Regarding the type of markers, codominant markers are better for QTL mapping accuracy. Dominant markers detect low polymorphism; they don't have the power to detect heterozygous genotypes. The properties of markers such as abundance in the genome, polymorphism, dominance nature, etc. could also affect QTL detection. The SNP marker, for example, is abundant in the genome, polymorphic, codominant, and evenly distributed, which makes it the marker of choice for association studies.

2.5.3 Methods of QTL mapping and statistical analysis

Bi-parental linkage analysis and association or linkage disequilibrium mapping are the two most commonly used methods for QTL mapping (Da Silva *et al.,* 2017; Xu *et al*., 2017). The difference between the two methods is the mapping resolution and power, which are influenced by mapping populations (Xu *et al*., 2017). Linkage mapping results in a low resolution for QTL mapping. Association mapping can overcome this limitation with higher resolution because of the higher frequency of recombination in natural populations. Besides, it detects broader allele coverage due to the huge genetic variation in the natural population, but statistically, it has less power to detect rare alleles and epistatic interactions.

The statistical analysis techniques used to detect QTLs also determine their accuracy. The three common statistical techniques used in bi-parental linkage mapping are single marker analysis, interval mapping analysis, and composite interval mapping analysis, each of which has strengths

and limitations. The size of the QTL effect and the distance between markers and QTLs, for example, affect the efficiency of single marker analysis.

Regarding association mapping, various statistical analysis models are available. Among the models, the General Linear Model (GLM), the Mixed Linear Model (MLM), the Unified Mixed Model, the Efficient Mixed Model Association (EMMA), the Compressed Mixed Linear Model (CMLM), and the Fixed and Random Model Circulating Probability Unification (Farm CPU) are most commonly applied in association mapping. Each of these models has its strengths and weaknesses. For example, the GLM model has a lower power of detecting marker association when the mapping population panels are related, which leads to false positives. On the other hand, the MLM has better power to detect QTLs as it considers population structure and relatedness as covariates. However, the MLM becomes more restricted in some data sets and creates false negatives. The Farm CPU is a powerful model that better controls false positives and negatives (Kaler *et al.*, 2020). Generally, the use of an efficient statistical analysis technique or model is crucial. In GWAS analysis, the available models need to be tested, and the most powerful model that can reduce false positives and false negatives should be selected.

2.5.4 Level of precision in genotyping and phenotyping

Errors in genotyping will reduce the precision of the QTL mapping. The level of missing genotypes and marker alleles needs to be considered. The marker or genotype data should be cleaned for missing individuals and markers using appropriate statistical techniques for the analysis. Genotype imputation, for example, has been applied to improve the precision and power of QTL detection.

Phenotyping of the trait with high-level precision (error control) is required for the marker-trait association. Well-designed experiments under field or controlled conditions need to be conducted

to control error variance. Trials can be replicated across locations in representative environments. Data collection should also be done at an appropriate stage and using appropriate devices. During data analysis, the data should be tested for their normal distribution before the association studies. The heritability of the trait could be used as a measure of the precision of the experiment; high heritability indicates that environmental effects and errors are well controlled.

2.5.5 Epistasis and QTL x E interaction

Epistatic interactions between genes or markers influence trait expression. In addition to the significant markers that determine the traits, epistatic interaction between significant and non-significant markers also has a substantial contribution to trait expression. Therefore, considering epistatic interaction enable the detection of more markers. The epistasis effect may shift the normal distribution of traits, indicating the necessity of data transformation. The identification of QTLs is also highly influenced by genotype-environment interactions, suggesting the need to identify specific or stable QTLs (Jahani *et al.*, 2019; Wang *et al.*, 2014). Experiments should be conducted in more than one location, and appropriate models that can capture Q x E interaction need to be used so that the confounding effects of Q x E interaction in QTL mapping become reduced.

2.6 Hybrid breeding and heterosis for yield and fruit quality

Heterosis is an increase in the vigor of progenies derived from the mating of genetically distant parents from different species, an isolated population, or a selected strain of species (Kaeppler, 2012; Paril, 2024). Heterosis has been extensively exploited in many crops and has significantly contributed to attaining the world's food security through improved productivity and quality (Hochholdinger & Zhang, 2025; Timberlake, 2013; Paril, 2024).

Heterosis was recognized for the first time by Darwin (1876), but later elaborated as an F1 progeny with better performance than parents (East, 1908; Shull, 1908). The extent of heterosis depends on the genetic background of parents, traits, and environmental influences. Hybrid varieties are usually better in performance than their parents, as practically observed in many economically important cultivated crop species such as maize, sorghum, sunflower, rice, and various vegetable crops (Fujimoto *et al*., Hochholdinger & Zhang, 2025; 2018; Duvick, 1999); but the development of hybrid varieties takes more time and resources than breeding for pure or inbred lines.

2.6.1. Genetic basis of heterosis

Heterosis can be determined as the deviation of the F1 progeny from the average parental performance, usually termed mid-parent heterosis. The better parent heterosis is the difference between the F_1 and the better parent in the desired trait (Kaeppler, 2012). Conceptually, heterosis is an increase in heterozygosity, the opposite of inbreeding (an accumulation of homozygous alleles).

In a random mating population, deleterious alleles are preserved, but due to inbreeding as a result of population isolation, genetic drift, or mating among related individuals, homozygous deleterious alleles appear that reduce the vigor of the individuals or populations. Through the mating of divergent parents or individuals, it is possible to restore the lost vigor, as the parental combination masks the deleterious effects of recessive alleles (Kaeppler, 2012). Shull (1908) reported decreased yield performance of maize inbred lines, yet recovery in the hybrid combination of the two inbred lines used as parents in the hybrids.

Based on classical genetics, three hypotheses or concepts are available: dominance, over dominance, and epistasis (Hochholdinger & Zhang, 2025; Kaeppler, 2012; Paril, 2024; Yu *et al.*,

2021). The dominance concept underlines the complementary action of a favorable allele received from one of the parents in the heterozygote state in the F1 progeny, explaining that the favorable allele from one of the parents buffers or covers the deleterious action of the recessive or unfavorable allele from the other parent in the hybrid combination that leads to better performance (Yu *et al.*, 2021). This hypothesis predicts the possibility of identifying a pure line or inbred line with equal or better performance than the F1 hybrid if the deleterious effects of the recessive alleles are eliminated (Fujimoto *et al.*, 2018).

Over dominance is the presence of multiple alleles at a locus, and their interaction leads to better performance than their homozygous counterparts (Timberlake, 2013; Paril, 2024). In this situation, breeding populations and strategies that enhance heterosis will significantly maximize the performance of hybrid breeding. This leads to superior performance *per se* in the desired trait of heterozygous genotypes over homozygous parental genotypes at individual loci (Timberlake, 2013). The additive × additive epistasis interaction between a locus and entire genetic background is a major component and critical mechanism underlying heterosis especially the mid-parent heterosis, demonstrating that comprehensive genetic models must incorporate these interactions to accurately identify the genomic regions responsible for heterotic effects (Melchinger *et al.*, 2007). In maize, for example, epistasis is a dominant genetic force underlying the performance of superior hybrids, particularly in crosses between temperate and tropical lines. For critical agronomic traits like grain weight, epistatic interactions between genes were found to be more influential than individual dominance effects. The study identified hundreds of epistatic quantitative trait loci (QTLs) and revealed their mechanism: networks of interacting genes amplify heterosis by activating transcription processes essential for photosynthesis and protein synthesis (Sang *et al.*, 2022). This underscores that harnessing epistatic effects is key to accelerating hybrid breeding

programs. In general, three of the above models will contribute singly or in combination to the hybrid vigor.

2.6.2 Exploitation and manifestation of heterosis in tomato

Previously, the most frequently observed manifestations of heterosis in tomato were high total yield, better adaptation to unfavorable environmental conditions, uniformity, and greater plant vigor (Garg *et al*., 2013; Yordanov, 1983). Currently, heterosis breeding in tomato is for economic traits and complex disease resistance. This is true of hybrid tomato breeding in the Netherlands, France, and Bulgaria. Though tomato is a self-pollinating crop, a higher manifestation of heterosis for economic traits and disease resistance was observed. For the first time, heterosis was investigated for fruit yield, and the number of fruits per plant demonstrated that hybrid vigor for yield reached as high as 33 to 49% (Izzo *et al*., 2022). This situation is supported by the fact that cultivated tomato is derived from outcrossing wild relatives in the center of origin, but domestication and selection have resulted in a highly self-pollinating nature.

Heterosis breeding is employed in tomato to improve yield and quality since breeding for pure line or open-pollinating varieties has limited potential (Bhattarai *et al*., 2016; Ene *et al*., 2023; Emami *et al*., 2018; Garg *et al*., 2013; Izzo *et al*., 2022; Liu *et al*., 2021; Pavan *et al*., 2022; Solieman *et al*., 2013; Tamta & Singh, 2018). Heterosis (positive or negative) has been observed for various growth, yield, and quality traits. Negative heterosis was reported for plant height, number of branches per plant, days to first harvest, fruit length, fruit width, and number of seeds per plant, while positive heterosis was observed for fruit weight, number of fruits per plant, and fruit yield per plant. Over 28% of better parent heterosis for fruit yield per plant was observed (Tamta & Singh, 2018). Breeding for hybrid tomato is economical since substantial heterosis can be obtained

for yield, yield-contributing traits, and biotic and abiotic stresses and easier for hybrid seed production since adequate seed can be obtained even from a single fruit compared to other vegetable crops (Solieman *et al*., 2013).

2.6.3 Prediction of hybrid performance in tomato

The development of hybrid varieties requires a lot of resources and time. On the other hand, not every combination of parental lines in F1 progenies would provide heterosis when compared to their parents. Therefore, it is essential to identify parental combinations that would give better heterosis in the F1 progenies using appropriate prediction methods (Fujimoto *et al.*, 2018). Various prediction methods are available depending on the types of hybrids (single-cross or two-way crosses) and the inheritance of the traits. These methods predict the extent of the breeding value of the F1 progenies (Nyaga *et al.*, 2020). The *per se* performance of parental lines, combining ability studies, and heterotic grouping using morphological and molecular genetic distance are the common prediction methods applied in various crop improvement programs.

2.6.3.1 *Per se* performance

Selection of the best lines based on their per se performance for the desired trait is a poor predictor for complex traits. The *per se* performance prediction method is effective for simple traits governed mainly by additive gene action (Smith, 1986). Midparent value is another approach for heterosis prediction, but it is severely affected by the presence of dominance gene action. The correlation between the midparent value and hybrid performance is influenced by dominance effects.

2.6.3.2 Combining ability

Combining ability was first introduced in maize in 1942 (Sprague & Tatum, 1942). Since then, it has been widely applied in various crop improvement programs to identify lines or parents performing better based on the performance of their progenies (Fasahat, 2016). Combining ability refers to the capacity of parents to combine so that desirable genes or characters are transmitted to progenies. Also, a technique of determining the value of parents based on their progeny's performance. This could be achieved through the application of mating designs. If the performance of offspring is high for the desired trait, then it means the parents have good combining ability (Fasahat, 2016).

Parental lines that provide hybrid varieties with high heterosis are essential; thus, combining ability is employed to identify elite parental lines and hybrids. General combining ability and specific combining ability are used; the former is due to additive gene action, whereas the latter is due to non-additive gene action (Sprague & Tatum, 1942). The GCA of a parent or inbred line is determined as the average performance of all hybrids having that inbred line as a common parent. SCA is specific to a cross or hybrid and is calculated as the difference or deviation of the hybrid's performance from what is expected based on parental GCA (Sprague & Tatum, 1942).

Hybrid performance based on general combining ability is a promising phenotype-based prediction method. To have an accurate prediction using this approach, the magnitude of the GCA variance should be considerably larger than the SCA variance. This approach needs prior information on the GCA of the lines (Gowda *et al.*, 2010). The ratio of the variance of GCA to the variance of SCA can be used to predict hybrid performance; the predominance of σ^2 GCA to σ^2 SCA indicates

the effectiveness of identifying and selecting hybrids based on GCA effects. The relative importance of GCA and SCA in determining hybrid performance depends on traits and hybrids.

Identification of parental lines for crossing and estimation of the combining ability of parents requires the use of appropriate mating designs through which partitioning of additive and non-additive genetic effects can be done (Fasahat, 2016; Oakey *et al.*, 2007). To select the most appropriate mating designs, consideration of breeding objectives, time, space, cost, and other biological limitations is essential (Awata *et al.*, 2018; Muthoni & Shimelis, 2020).

2.6.3.3 Genetic distance

Genetic variation between parents is a prerequisite to obtaining high or better parent heterosis. Heterotic grouping is one of the methods to predict heterosis in hybrids. Studies on the relationship between the genetic distance of parents and heterosis have been conducted for different crops. Heterosis is correlated with genetic distance, within a specific range, while other studies showed no significant association between heterosis and genetic distance.

A linear relationship is reported between the genetic distances, considering all loci controlling the trait, and the extent of hybrid vigor (Falconer and Mackey 1996). Prediction of heterosis based on genetic distance (e.g., using molecular markers, for example) is suggested, but it has non-efficient prediction accuracy in some crops like sunflower (Reif *et al.*, 2013). This might be due to the absence of strong linkage or correlation between the QTLs responsible for the hybrid vigor and the molecular marker used to determine the genetic distances, the dissimilar linkage phases or stages among the QTL and marker alleles, and the effect of interallelic interactions (Charcosset *et al.*, 1991). Furthermore, the effectiveness of predicting hybrid performance using genetic distance depends on genetic backgrounds. Linkage disequilibrium between DNA markers and genes

determines the extent of the association between genetic distance and hybrid performance (Charcosset & Essioux, 1994).

Consideration of optimum genetic distance among parents in hybrid combinations, among others, is essential (Moll *et al.*, 1965). For high and mid-parent heterosis, consideration of gene action in terms of specific combining ability (presence of significant dominance and epistatic effect), genetic distance, and relative importance of the desired trait in each parent is essential. Moreover, consideration of the influence of the environment on the performance of the hybrids and their parents should be given attention. The environmental effect may mislead the decision on the relationship between heterosis and hybrid performance.

CHAPTER THREE

3.0 Phenotypic diversity among tomato (*Solanum lycopersicum* L.) germplasm

3.1 Introduction

The tomato breeding program aims to achieve a high fruit yield, resistance to biotic and abiotic stresses, and fruit quality such as fruit size, appearance, color, total soluble solids, texture, firmness, shelf life, flavor, and nutritional content (Abewoy Fentik, 2017; Abhary & Rezk, 2021a; Sharma *et al*., 2019). In Ethiopia, tomato breeding programs have released open-pollinated verities, and registered commercial hybrid varieties to improve productivity. However, open-pollinated varieties are less preferred by farmers due to declining productivity, smaller fruit sizes, and limited market value compared to hybrids. Although hybrids are preferred by growers for their high yield, large fruit size, better shelf life, and high market value, their seed is expensive, making it difficult for small-scale farmers to afford. Moreover, few of the registered hybrids are available to growers.

The productivity of tomato in Ethiopia remains low (6 t ha $^{-1}$) compared to the world average (36 t ha^{-1}). Diseases, insect pests, weeds, limited availability of quality seed, a knowledge gap for GAP, limited improved varieties, and resistance to biotic and abiotic stresses (Wondirad *et al*., 2009; Gashawbeza *et al*., 2009; Etagegnehu *et al*., 2009; Mersha & Sime, 2022) are among the major production constraints.

Tomato varieties (OPVs and hybrids) with high yield, fruit quality such as large fruit size, extended shelf life and thick pericarp, and resistance to diseases and insect pests would contribute to improving tomato productivity and utilization in Ethiopia. Establishing an effective breeding strategy necessitates understanding the degree and pattern of genetic diversity and trait identification among germplasm (Bhanu, 2017; Da Silva *et al*., 2017). Genetic diversity using

morphological, cytological, biochemical, and molecular markers has been studied in many cultivated crops (Franco *et al*., 2001; Nadeem *et al*., 2018). In tomato, highly heritable descriptors of seedling, plant, inflorescence, flower, fruit, and agronomic features have been applied (Figàs *et al*., 2015; IPGRI, 1992; UPOV, 2002) to study genetic diversity.

The primary challenges in crop improvement programs are genotype-environment interactions and trait interrelationships. Most economic quantitative traits of crops, such as tomato crop yield, have a complex inheritance influenced by multiple genes with little effects, environmental influences, and their interaction (Hernández-Bautista *et al*., 2015; Semagn & Asmund Bjørnstad, 2010). When phenotypic variation is partitioned into heritable and non-heritable components and the expected genetic advance is determined, promising parental materials and progenies can be identified. Multiple trait selection can also be effectively applied if information on the nature and strength of interactions between traits is available. In Ethiopia, phenotypic characterization of a small number of tomato germplasm was carried out (Ene *et al*., 2022; Shushay et al., 2014), but the information from a large number of germplasm would be more significant for breeding. As a result, the current study sought to evaluate the phenotypic diversity of 125 breeding lines and 18 improved cultivars maintained at the Melkassa Agricultural Research Center. The specific objectives were to:

- assess the extent and pattern of phenotypic diversity among tomato germplasm;
- determine genetic variability, heritability, genetic advance, and the relative importance of yield and fruit quality; and
- determine interrelationship between yield, yield related traits and fruit quality.

3.2 Materials and methods

3.2.1 Study location and season

The experiment was conducted at the Melkassa Agricultural Research Center (MARC) for two seasons. The first trial was carried out in the dry cool season (from the end of August 2021 to the end of January 2022), and the second experiment was carried out in the dry hot season (from February to the end of June 2022).

3.2.2 Experimental material and trial management

One hundred and forty-three (143) tomato genotypes made up of improved varieties (18) and breeding lines (125) were used for the study. For research purposes, these genotypes have been introduced since the 1970s from various sources, including WorldVeg and different countries.

Twenty-eight-day-old seedlings (3–4 leaf stage) were transplanted to the experimental field with a plot size of 6 m^2 (two rows of 3 m length and 2 m width) at a spacing of 100 cm by 50 cm between rows and plants, respectively. The experimental design was an 11 x 13 alpha lattice with two replications.

The water requirement of tomato plants was supplied through furrow irrigation, with the frequency of application being every other day for the first three weeks and every 5-7 days thereafter. All other necessary crop management practices for normal growth and development of tomato plants were applied uniformly to all experimental units. Inorganic fertilizers NPS (N = 19%, P205 = 46%, S = 7%) and urea (N = 46%) were applied. NPS was applied once, one day before transplanting, at a rate of 242 kg ha^{-1}, while urea was applied twice: three weeks after transplanting and six weeks after transplanting (flowering time), at a rate of 79 kg ha^{-1}. Diseases (early blight, powdery mildew,

and bacterial leaf spot) and insects (*Tuta absoluta*, spider mites, and white flies) were controlled through the application of recommended chemicals for the crop. Tomato plants were supported with wooden stakes for uniform and quality fruit production.

3.2.3 Data collected

Data were collected for 18 qualitative and 15 quantitative traits during vegetative, flowering, fruit set, and ripening stages, following International Plant Genetic Resource Institute (IPGRI, 1992) and the Union for the Protection of New Varieties of Plants (UPOV, 2002) descriptors and guidelines (Appendix Tables 3.1-3.4). The traits characterized included:

1. Plant morphology: Growth type, size, stem pubescence density, foliage density.
2. Leaf characteristics: Attitude, type, and anthocyanin coloration of veins.
3. Floral traits: Inflorescence type, corolla color, style position, and hairiness.
4. Fruit descriptors: External and internal color, firmness, green shoulder presence/intensity, shape (both general and blossom end).
5. Agronomic data: Number of days to 50% flowering, number of fruit sets per inflorescence, fruits per cluster, soluble solids, marketable and total yield (both weight and number per plant), and plant height.

3.2.4 Data analysis

Quantitative data were subjected to linear mixed model analysis using META-R software (Alvarado *et al*., 2020). The variance components were determined using linear mixed models with maximum residual likelihood (REML), implemented in the LME4-R package using the same software. Except for the overall mean, all effects were assumed to be random and normal, with particular effect variances. To construct BLUPs and broad sense heritability, the genotype effect was considered to be random. For the analysis, traits and environments with a broad sense of

heritability greater than 5% were considered. The linear mixed models for single location/season and multi-location/season are shown below.

$\boldsymbol{Y_{ijk}} = \mu + g_i + \pi_i + b_j(\pi_k) + \varepsilon_{ijk}$ single location or season

$\boldsymbol{Y_{ijkl}} = \mu + l_i + g_j + \pi_k(l_i) + b_l(\pi_k\, l_i) + g_j * l_i + \varepsilon_{ijkl}$ ..... across locations or seasons

Where $\mathbf{Y_{ijk}}$ = the phenotypic value of i^{th} genotype in block j and replication k or the trait of interest, g_j = the j^{th} genotype effect, μ = the overall mean of trait value of genotypes across replication and location or the overall mean effect, π_k = the effect of k^{th} replication with in i^{th} location, b_l = effect of the l^{th} incomplete blocks within the k^{th} replicate and i^{th} location, $g_j{*}l_i$ = the interaction effect between the j^{th} genotype and i^{th} location and ε_{ijkl} = random residual error or the effect of the error associated with j^{th} replication, k^{th} incomplete blocks and the i^{th} genotype which assumed to be iid (independently and identically normally distributed) with mean zero and variance σ^2.

3.2.4.1 Genetic variances

The phenotypic and genotypic variances and coefficients of variation were calculated as follows (Burton & DeVane, 1953):

$$Genotypic \text{ variance } (\sigma^2_G) = \frac{MS_G - MS_{G*E}}{E * r}$$

$$\text{Genotype by enviroment interaction } (\sigma^2_{G*E}) = \frac{MS_{G*E} - MS_E}{r}$$

$$Phenotypic\ variance\ (\sigma^2_P) = \sigma^2_G + \frac{\sigma^2_{G*E}}{E} + \frac{\sigma^2_E}{rE}$$

Where MS_G = Mean square of genotypes, MS_{G*E} = mean square of genotypes by season interaction, MS_E = mean square error, r=number of replications and e is the number of seasons.

$$\text{Phenotypic coefficient of variation (PCV)} = \frac{\sqrt{\sigma^2_P}}{x} \text{ x } 100$$

$$\text{Genotypic coefficient of variation (GCV)} = \frac{\sqrt{\sigma^2_G}}{x} \text{ x } 100$$

Where σ^2_P = phenotypic variance of traits, σ^2_G = genotypic variance among genotypes of traits, x = grand mean of traits across replication and seasons.

3.2.4.2 Broad sense heritability

Heritability and genetic advance was computed using genetic, phenotypic and environmental variances as described by (Johnson *et al*., 1955).

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{GE}^2}{E} + \frac{\sigma_E^2}{rE}}$$ .....accross seasons

$$\sigma_G^2 = \frac{MS_G - MS_{G*E}}{e * r}$$

$$\sigma_E^2 = MS_E$$

$$\sigma_P^2 = \sigma_G^2 + \frac{\sigma_{GE}^2}{eE} + \frac{\sigma_E^2}{rE}$$

Where MS_G = Mean square of genotypes, MS_{G*E} = mean square of genotypes by season interaction, MS_E = mean square error, r = number of replications and E is the number of seasons.

3.2.4.3 Genetic advance

The genetic advance was computed using the following formula, assuming 5% selection intensity:

$$GA = K * \sigma_P + H^2$$

$$GA\% = \frac{GA}{x}$$

Where H^2 = heritability in the broad sense, σ_P = phenotypic standard deviations, K = is a constant that vary depending on the selection intensity applied (k = 2.06 at 5% selection intensity), and x = population mean (the average phenotypic value of the trait).

3.2.4.4 Genetic association among quantitative traits

The same software META-R was used to compute the genetic and phenotypic correlation among traits (Alvarado *et al*., 2020).

$$\text{Phenotypic correlation } (r_P) = \frac{\text{Phenotypic coovariances of traits x and y}}{\sqrt{\sigma_{Px}^2 * \sigma_{Py}^2}}$$

$$\textbf{G}\text{enotypic correlation } (r_G) = \frac{\text{Genotypic coovariances of traits x and y}}{\sqrt{\sigma^2_{Gx} * \sigma^2_{Gy}}}$$

σ^2_{Px} = phenotypic variances of trait x, σ^2_{Py} = phenotypic variances of trait y, σ^2_{Gx} = genotypic variances of trait x, σ^2_{Gy}=

genotypic variances of trait y.

3.2.4.5 Cluster analysis

The mean values across seasons were employed for cluster analysis. The scale () function in the R software packages was employed to standardize the mean data. The same software's dist () function was used to determine Euclidean pairwise distance to carry out agglomerative hierarchical clustering. The grouping of genotypes was determined using the Ward.D2 technique and the hclust function of R software packages. The number of clusters was calculated using gap statistics and the fviz_nbclust function of R software.

3.2.4.6 Principal component analysis

Similarly, mean values across seasons were used to compute the principal component analysis. PCA based on the correlation matrix using the PCA () function of the FactoMineR package of the R software was determined. Similar to the hierarchical clustering method, the scaling was computed. PCA biplots were also generated using packages of the R software (Kassambara, 2017).

3.2.4.7 Qualitative traits

Eighteen qualitative traits that fall into three categories: flower traits (3), growth pattern traits (6), and fruit characteristics (9) were employed. The frequency was computed using SPSS version 2021. The Shannon diversity index, as described by Hutchenson (1970), and chi-square were

calculated using excel. Gower's coefficient of genetic distance and Ward. D2 clustering methods were used for the grouping of genotypes. Multiple Correspondence Analysis (MCA) was also generated for genotypes and traits. Various R software packages were used for the analysis (Kassambara, 2017).

$$H' = \sum_{i=1}^{n} pi(lnpi)$$

Where H'= Shannon weaver diversity index, pi= the proportion of the total number of individual accessions in the i^{th} class, ln = natural logarithm and n is the number of phenotypic classes for a given trait.

3.3 Results

3.3.1 Genetic variability for yield, yield related and fruit quality traits

Highly significant *($p<0.01$)* genetic differences were observed among genotypes for all traits (Table 3.1). Genotype-by-season interaction was also highly significant (*$p<0.01$*) for fruit weight, number of locules, number of marketable and total fruits per plant, marketable yield per plant, and fruit shape index.

The mean days to 50% flowering ranged from 32 to 36 days. The pericarp thickness ranged from 2.92 mm to 7.66 mm. The minimum (4.0) and maximum (4.96) values were recorded for fruit TSS. The marketable fruit number per plant ranged from 172 g to 18 g. The total number of fruits per plant ranged from 32 to 224. The maximum marketable and total yield per plant was 2967.35 g and 4851.9 g, respectively. Marketable and total yield per plant had a minimum value of 1969.03 g and 2669.46 g, respectively. The mean range of fruit weight was 23.53 g to 150.63 g. The maximum fruit width was 66.04 mm, while the minimum was 31.08 mm. Fruit length also ranged from 33.06 mm to 86.47 mm. Plant height ranged from 48.32 cm to 148.33 cm. The minimum fruit shape index was 0.69, while the maximum fruit shape index was 1.9.

3.3.2 Variance components, heritability and genetic advance

Variance components, along with estimates of heritability and genetic advance, are provided (Table 3.2). Variance components, heritability, and genetic advance are indicated in Table 3.2. A high phenotypic coefficient of variation (PCV) was observed compared to genotypic coefficient of variation (GCV), indicating a significant impact of environmental factors on trait expression. In 1973, Sivasubramanian and Menon categorized GCV and PCV as low (0–10%), medium (10–20%), and high (>20%), as cited in Chatara *et al.* (2023). The findings revealed a low PCV for days to 50% flowering (3.82%) and total soluble solids (8.69), a high PCV for pericarp thickness (20.71%), number of locules (30.15%), marketable number per plant (58.29%), marketable yield per plant (24.07%), total number of fruits per plant (54.08%), total yield per plant (22.53%), fruit weight (33.7%), and plant height (26.52%). A medium PCV was obtained for fruit length (19.75%), fruit width (14.74%), and number of fruits per cluster (13.12%).

GCV was high for number of locules (27.84%), marketable number per plant (51.88%), and total number of marketable fruits per plant (483.3%), fruit weight (31.08%), plant height (25.14%), and fruit shape index (21.02%). It was medium for pericarp thickness (19.14%), marketable yield per plant (13.03%), total yield per plant (14.49%), fruit length (18.79), fruit width (13.75%), and number of fruits per cluster (10.12%). However, low GCV was recorded for days to 50% flowering (2.58%) and total soluble solids (5.54%).

Table 3.1: Genetic variability of tomato germplasm based on 14 quantitative traits across seasons

Traits	Range		SE	Level of significance			CV (%)
	Minimum	Maximum		Genotype	Genotype *Season	Season	
DFF[1]	32	36	1.70	**	ns	*	5.0
PCTk[2]	2.92	7.66	0.86	**	ns	ns	15.9
NL[3]	2	7	0.63	**	**	ns	18.8
TSS[4]	4.07	4.96	0.59	**	ns	ns	13.4
MNP[5]	18	172	22.47	**	**	**	25.0
MYP[6]	1969.03	2967.35	792.09	**	**	ns	32.4
TNP[7]	32	224	30.80	**	**	ns	25.9
TYP[8]	2669.46	4851.9	1113.04	**	ns	ns	31.0
FWt[9]	23.53	150.63	19.37	**	*	ns	12.9
FL[10]	33.06	86.47	6.35	**	ns	ns	10.9
FWd[11]	31.08	66.04	4.85	**	ns	ns	9.4
PH[12]	48.32	148.33	10.37	**	ns	*	15.1
NFtC[13]	3	5	0.59	**	ns	ns	16.7
FSHI[14]	0.69	1.99	0.12	**	**	ns	10.5

** Highly significant ($p<0.01$), * significant ($p<0.05$), ns non-significant. SE = standard error, 1 = days to 50% flowering, 2 = pericarp skin thickness (mm), 3 = number of locules, 4 = total soluble solids, 5 = marketable fruit number per plant, 6 = marketable yield per plant (g), 7 = the total number of fruits per plant, 8 = total yield per plant (g), 9 = fruit weight (g), 10 = fruit length (mm), 11 = fruit width (mm), 12 = plant height (cm), 13 = number of fruits per cluster, and 14 = fruit shape index.

Table 3.2: Genetic components of variance, heritability, and genetic advance of tomato genotypes across two seasons, 2021/2022 and 2022.

Traits	σ^2_G	σ^2_{G*E}	σ^2_S	σ^2_e	σ^2_P	GCV (%)	PCV (%)	H^2	GA	GAM (%)
DFF[1]	0.77	0.4	64.74	2.88	1.69	2.58	3.82	0.46	0.83	2.44
PCTk[2]	1.08	0	0	0.74	1.27	19.14	20.71	0.85	1.82	33.51
NL[3]	0.87	0.1	0	0.4	1.02	27.84	30.15	0.85	1.63	48.75
TSS[4]	0.06	0	0	0.35	0.15	5.54	8.69	0.41	0.21	4.68
MNP[5]	513.91	17.24	227.11	505.04	648.79	51.88	58.29	0.79	36.89	84.42
MYP[6]	101098.5	174874.3	414478.5	627401.8	345386	13.02	24.07	0.29	189.95	7.78
TNP[7]	999.83	30.72	0	948.55	1252.33	48.33	54.08	0.80	51.98	79.44
TYP[8]	285954.1	109822.8	1754.38	1238868	650582.5	14.94	22.53	0.44	484.69	13.54
FWt[9]	755.46	77.54	30.31	375.31	888.05	31.08	33.7	0.85	48.18	54.48
FL[10]	119.37	4.77	0.96	40.32	131.83	18.79	19.75	0.91	20.37	35.04
FWd[11]	47.78	2.38	0.45	23.48	54.84	13.75	14.74	0.87	12.4	24.67
PH[12]	299.83	13.74	80.53	107.49	333.57	25.14	26.52	0.90	32.07	46.56
NFtC[13]	0.13	0	0	0.34	0.21	10.12	13.12	0.59	0.43	12.38
FSHI[14]	0.06	0	0	0.02	0.07	21.02	21.99	0.92	0.47	39.62

σ^2G = genetic variance, σ^2G*E = genotype-environment interaction variance, σ^2S = season variance, σ^2e = environmental variance, σ^2P = phenotypic variance, GCV (%) = genotypic coefficient of variance, PCV (%) = phenotypic coefficient of variance, H^2 = broad sense heritability, GA = genetic advance, and GAM (%) = genetic advance as percent of the mean., 1 = days to 50% flowering, 2 = pericarp skin thickness (mm), 3 = number of locules, 4 = total soluble solids, 5 = marketable fruit number per plant, 6 = marketable yield per plant (g), 7 = the total number of fruits per plant, 8 = total yield per plant (g), 9 = fruit weight (g), 10 = fruit length (mm), 11 = fruit width (mm), 12 = plant height (cm), 13 = number of fruits per cluster, and 14 = fruit shape index.

Robinson *et al*. (1949) highlighted the heritability of various traits in corn into three groups: low (0–30%), medium (30–60%), and high (>60%). High heritability was recorded for pericarp thickness (85%), number of locules (85%), marketable number per plant (79%), total number of fruits per plant (80%), fruit weight (85%), fruit length (91%), fruit width (87%), plant height (9%), and fruit shape index (92%). Medium heritability was recorded for days to 50% flowering (46%), total soluble solids (41%), total yield per plant (44%), and number of fruits per cluster (59%). Marketable yield per plant (29%) showed low heritability.

According to the classification by Johnson *et al.* (1955), genetic advance as a percent of the mean can be categorized as low (0–10%), medium (10–20%), and high (>20%). High genetic advance as percent of the mean was observed for pericarp thickness (33.51%), number of locules (48.75%), marketable number per plant (84.42%), total number of fruits per plant (79.44%), fruit weight (54.48%), fruit length (35.04%), fruit width (24.67%), plant height (46.56%), and fruit shape index (39.62%). Medium levels of genetic advance were found for total yield per plant (13.54%) and number of fruits per cluster (12.38%). However, low genetic advance as a percent of the mean was observed for days to 50% flowering (2.44%), total soluble solids (4.68%), and marketable yield per plant (7.78%).

3.3.3 Genetic and phenotypic correlations among yield, yield related and fruit quality

A highly significant ($p<0.01$) genetic correlation was observed among most of the traits (Table 3.3). Marketable yield was highly correlated with TNP (r_g = -0.4), FWt (r_g = +0.55), FL (r_g = +0.43), FWd (r_g = +0.31), FSHI (r_g = +0.34), DFF (r_g = +0.24), PCTk (r_g = +0.54), and MNP (r_g = -0.33). Similarly, a highly significant genetic correlation was observed among TYP and FWt (r_g = +0.39), FL (r_g = +0.23), FWd (r_g = +0.43), NFtC (r_g = -0.24), DFF (r_g = +0.34), PCTk (r_g = +0.38),

NL (r_g = +0.18), TSS (r_g = +0.18), MYP (r_g = +0.99), TNP (r_g = -0.42), and MNP (r_g = -0.42). Fruit size traits FWt (r_g =-0.825, r_g =-0.844), FL (r_g =-0.614, r_g = -0.670), FWd (r_g =-0.867, -r_g = 0.876), PCTk (r_g = -0.627, r_g = -0.669), and NL (r_g = -0.288, r_g =-0.328) had a highly significant negative correlation with MNP and TNP in that order. A significant negative correlation (r_g = -0.331) was observed between the marketable yield and number of marketable fruits per plant, and a negative correlation (r_g = -0.423) was found between the total yield and the total number of fruits per plant. However, a noteworthy positive genetic correlation was observed between fruit number per plant and NFtC (r_g = +0.76) and TSS (r_g = +0.41).

The majority of observed traits exhibited similar patterns of association at the phenotypic level (Table 3.4**)**. Marketable yield per plant demonstrated a highly significant (*p<0.01*) positive correlation with PCTk (rp = +0.4), FWt (r_p = +0.36), FL (r_p = +0.41), FWd (r_p = +0.25), and FSHI (r_p = +0.22). Conversely, there was a significant (*p<0.05*) negative correlation between NFIR (r_p = -0.17) and TSS (r_p = -0.17). DFF (r_p = 0.05), NL (r_p = 0.06), MNP (r_p = 0.07), PH (r_p = 0.08), and NFtC (r_p = -0.001) demonstrated no significant phenotypic associations. The trend in the association of total yield per plant with PCTk (r_p = +0.36), FWt (r_p = +0.46), FL (r_p = +0.36), FWd (r_p = +0.38), and PH (rp = +0.17) was similar to genotypic correlation. Notably, there was a significant negative correlation between TYP and NFIR (r_p = -0.21). Conversely, NFtC (r_p = -0.1), DFF (r_p = +0.12), NL (r_p = +0.11), TSS (r_p = +0.11), MNP (r_p = +0.0.07), and FSHI (r_p = +0.1) demonstrated no significant correlation. Similar to the genetic correlation, the association of MNP and TNP with fruit size traits such as FWt (r_p = -0.73), FL (r_p = +-0.51), FWd (r_p = -0.74), PCTk (r_p = -0.49), NL (r_p = -0.31), and DFF (r_p =-0.46) was found to be negative and highly significant (*p<0.01*). A positive and significant association (*p<0.05*) was observed between the number of fruits per plant and NFIR (r_p = +0.51), TSS (r_p = +0.19), and NFtC (r_p = +0.56).

Table 3.3: Genotypic correlation (r_g) for yield, yield-related, and fruit quality

Traits	DFF	PCTk	NL	TSS	MNP	MYP	TNP	TYP	FWt	FL	FWd	PH	NFtC	FSHI
DFF[1]	-													
PCTk[2]	0.52**	-												
NL[3]	-0.06[ns]	-0.32**	-											
TSS[4]	0.12*	-0.35**	0.05[ns]	-										
MNP[5]	-0.8**	-0.63**	-0.33**	0.41**	-									
MYP[6]	0.24**	0.54**	-0.09[ns]	0.2*	-0.33**	-								
TNP[7]	-0.5**	-0.67**	-0.29**	0.43**	0.99**	-0.4**	-							
TYP[8]	0.34**	0.38**	0.18*	0.18*	-0.42**	0.99**	-0.42**	-						
FWt[9]	0.52**	0.54**	0.49**	-0.22**	-0.83**	0.55**	-0.84**	0.58**	-					
FL[10]	0.36**	0.75**	-0.23**	-0.29**	-0.61**	0.42**	-0.67**	0.23**	0.6**	-				
FW[11]d	0.53**	0.41**	0.63**	-0.25**	-0.87**	0.31**	-0.88**	0.43**	0.91**	0.27**	-			
PH[12]	0.18*	-0.22**	0.05[ns]	0.68**	0.19*	-0.12[ns]	0.29**	0.08[ns]	-0.23**	-0.31**	-0.12[ns]	-		
NFtC[13]	-0.16[ns]	-0.22**	-0.66**	0.39**	0.76**	-0.06[ns]	0.7**	-0.24**	-0.76**	-0.38**	-0.76**	0.28**	-	
FSHI[14]	0.08[ns]	0.4**	-0.58**	-0.03[ns]	0.02[ns]	0.34[ns]	-0.02[ns]	0.08[ns]	-0.01[ns]	0.75**	-0.41**	-0.15[ns]	0.19[ns]	-

** Highly significant at *p<0.01*, * significant at p<0.05, **ns** non-significant, 1 = days to 50% flowering, 2 = pericarp skin thickness (mm), 3 = number of locules, 4 = total soluble solids, 5 = marketable fruit number per plant, 6 = marketable yield per plant (g), 7 = the total number of fruits per plant, 8 = total yield per plant (g), 9 = fruit weight (g), 10 = fruit length (mm), 11 = fruit width (mm), 12 = plant height (cm), 13 = number of fruits per cluster, and 14 = fruit shape index.

Table 3.4: Phenotypic correlation (r_p) for yield, yield-related, and fruit quality

Traits	DFF	NFIR	PCTk	NL	TSS	MNP	MYP	TNP	TYP	FWt	FL	FWd	PH	NFtC	FSHI
DFF[1]															
PCTk[2]	0.42**	-0.39**													
NL[3]	0.01ns	-0.15ns	-0.27**												
TSS[4]	0.07ns	0.18*	-0.26**	0.04ns											
MNP[5]	-0.46**	0.51**	-0.49**	-0.31**	0.19*										
MYP[6]	0.05ns	-0.17*	0.40**	-0.06ns	-0.17*	0.07ns									
TNP[7]	-0.45**	0.52**	-0.53**	-0.27**	0.24**	0.97**	0.02								
TYP[8]	0.11ns	-0.21*	0.36**	0.11ns	-0.11ns	-0.07ns	0.89	-0.04ns							
FWt[9]	0.37**	-0.57**	0.49**	0.48**	-0.21*	-0.73**	0.36	-0.74**	0.46**						
FL[10]	0.31**	-0.51**	0.68**	-0.15ns	-0.22**	-0.51**	0.41	-0.55**	0.36**	0.60**					
FWd[11]	0.36**	-0.45**	0.38**	0.60**	-0.19*	-0.74**	0.25	-0.73**	0.38**	0.89**	0.30**				
PH[12]	0.09ns	0.36**	-0.19*	0.02ns	0.40**	0.20*	0.08	0.29*	0.17*	-0.19*	-0.25**	-0.09ns			
NFtC[13]	-0.09ns	0.57**	-0.15ns	-0.50**	0.18*	0.56**	0.00	0.53**	-0.10ns	-0.57**	-0.26**	-0.57**	0.21*		
FSHI[14]	0.04ns	-0.16ns	0.34**	-0.51**	-0.04ns	0.03ns	0.22	0.00ns	0.09ns	-0.04ns	0.74**	-0.40**	-0.14ns	0.14ns	

** Highly significant, * significant, ns non-significant, ** Highly significant at $p<0.01$, * significant at $p<0.05$, ns non-significant, 1 = days to 50% flowering, 2 = pericarp skin thickness (mm), 3 = number of locules, 4 = total soluble solids, 5 = marketable fruit number per plant, 6 = marketable yield per plant (g), 7 = the total number of fruits per plant, 8 = total yield per plant (g), 9 = fruit weight (g), 10 = fruit length (mm), 11 = fruit width (mm), 12 = plant height (cm), 13 = number of fruits per cluster, and 14 = fruit shape index.

3.3.4 Hierarchical grouping of genotypes

Genotypes were grouped into three distinct clusters (Figure 3.1a & b). Cluster III was the largest with 74 genotypes, followed by Cluster II with 39 genotypes, and Cluster I with 29 genotypes (Table 3.6). The PCA biplots of individuals (Figure 3.1c) validated the grouping of genotypes into three clusters.

The characteristics of each cluster based on the average value of each trait (Table 3.5) revealed that cluster III had a relatively longer maturity period, fruits with thicker fruit walls, longer fruits, shorter plants, and a large fruit shape index, as well as a relatively high yield from a moderate number of fruits per plant and a large fruit size. Genotypes in cluster II were characterized by late flowering, small number of fruits per inflorescence, a small total and marketable number of fruits per plant, few fruits per cluster, large number of locules per fruit, and high marketable and total yield per plant. The high marketable and total yield in Cluster II was primarily due to larger fruit size, specifically greater fruit weight and width, compared to Clusters I and III. Despite having large number of marketable and total fruits per plant, cluster I was characterized by low marketable and total yield per plant, short flowering period, thin fruit wall thickness, and few locules per fruit. The small fruit size genotypes, which had the smallest fruit weight and length, were the cause of low yield per plant in this cluster.

Table 3.5: Characteristics of the three distinct clusters based on the mean value of the 15 quantitative traits

Character	Clusters		
	I	II	III
DFF[1]	33.26*	34.45**	34.28**
PCTK[2]	4.18*	5.28	6.04**
NL[3]	3*	4**	3
TSS[4]	4.57**	4.37*	4.38*
MNP[5]	75**	30*	37
MYP[6]	2217.53*	2501.99**	2503.18**
TNP[7]	109**	46*	56
TYP[8]	3261.49*	3833.94**	3574.25
FWt[9]	50.17*	114.97**	90.71
FL[10]	45.42*	56.66	64.39**
FWd[11]	42.02*	58.12**	49.66
PH[12]	83.5**	68.99	62.6*
NFtC[13]	3.82**	3.2*	3.53
FSHI[14]	1.11	0.98*	1.31**

**highest value, *lowest value, 1 = days to 50% flowering, 2 = pericarp skin thickness (mm), 3 = number of locules, 4 = total soluble solids, 5 = marketable fruit number per plant, 6 = marketable yield per plant (g), 7 = the total number of fruits per plant, 8 = total yield per plant (g), 9 = fruit weight (g), 10 = fruit length (mm), 11 = fruit width (mm), 12 = plant height (cm), 13 = number of fruits per cluster, and 14 = fruit shape index.

Table 3.6: The distribution of 143 tomato genotypes into three distinct clusters

Cluster	Number of genotypes	Genotypes
I	29	1,2,4,,8,9,10,11,14,17,18,19,21,23,25,28,31,32,33,34,46,53,54,55,56,57,58,61,63,70
II	39	3,6,7,13,15,16,20,26,36,43,44,45,47,50,59,60,65,76,78,81,84,89,92,95,98,99,100,101 ,103,105,113,132,133,134,135,136,137,142,143
III	74	5,12,22,24,27,29,30,35,38,39,40,41,42,48,49,51,52,62,64,66,67,68,69,71,72,73,74,75 ,77,79,80,82,83,85,86,87,88,90,91,93,94,96,97,102,104,106,107,108,109,110,111,11 2,114,115,116,117,118,119,120,121,122,123,124,125,126,127,128,129,130,131,138, 139,140,141

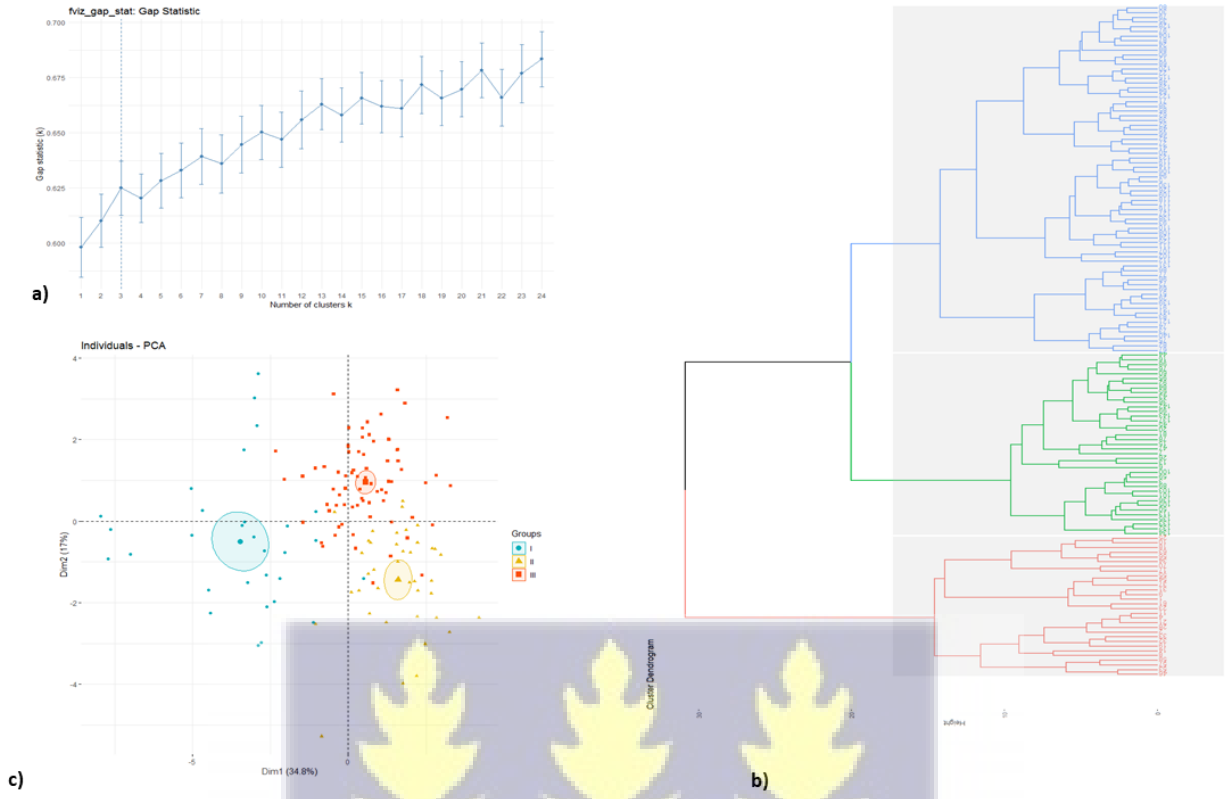

Figure 3.1: Grouping of tomato genotypes into three distinct clusters. a) Gap statistics for determining the optimum number of clusters b) The three distinct clusters and genotype lists in each cluster. The red, green and blue colors indicate cluster I, cluster II and cluster II respectively; c) Biplots grouping genotypes into three: Blue is group 1, red is group 2, and orange is group 3.

3.3.5. Principal components and relative contribution of traits

The first five principal components contributed 82.68% of the genetic variation (Table 3.7). The first principal component explained 34.82% of the total genetic variance. The number of marketable and total fruits per plant, fruit weight and width were the most contributing traits. The second principal component contributed 17% of the total variability and was influenced by number of locules, fruit wall thickness, fruit length, and fruit shape index. The third principal component, which accounted for 23.94% of the variance, was primarily influenced by plant height, the marketable and total yield per plant. The fourth principal component contributed 9.5% of the total variation, mainly by days to 50% flowering, TSS, and plant height. Finally, the fifth principal

component, which accounted for 7.42% of the total variation, was predominantly influenced by TSS.

Table 3.7: Eigenvectors and eigenvalues of the first five principal components for 15 quantitative traits of 143 tomato genotypes across seasons.

Variables	Principal components				
	PC1	PC2	PC3	PC4	PC5
Eigen value	5.22	2.55	2.1	1.43	1.11
Variance (%)	34.82	17	13.94	9.5	7.42
Cumulative variance (%)	34.82	51.82	65.76	75.26	82.68
	Eigen vectors				
DFF[1]	0.364	0.116	0.104	**0.646**	-0.384
PCTk[2]	0.600	**0.549**	0.063	0.172	-0.249
NL[3]	0.353	**-0.761**	0.180	-0.102	0.221
TSS[4]	-0.129	-0.057	-0.028	**0.591**	**0.683**
MNP[5]	**-0.848**	0.132	0.282	-0.221	0.094
MYP[6]	0.269	0.427	**0.781**	-0.237	0.069
TNP[7]	**-0.869**	0.102	0.273	-0.180	0.118
TYP[8]	0.358	0.256	**0.829**	-0.162	0.140
FWt[9]	**0.905**	-0.087	0.194	-0.027	0.061
FL[10]	0.661	**0.637**	-0.173	0.017	0.194
FWd[11]	**0.840**	-0.376	0.269	0.042	-0.106
PH[12]	-0.290	-0.136	**0.461**	**0.584**	0.203
NFtC[13]	-0.675	0.331	0.161	0.176	-0.278
FSHI[14]	0.035	**0.834**	-0.334	0.008	0.306
Mean	**7.837**	**4.819**	**4.453**	**3.449**	**3.382**

1 = days to 50% flowering, 2 = pericarp skin thickness (mm), 3 = number of locules, 4 = total soluble solids, 5 = marketable fruit number per plant, 6 = marketable yield per plant (g), 7 = the total number of fruits per plant, 8 = total yield per plant (g), 9 = fruit weight (g), 10 = fruit length (mm), 11 = fruit width (mm), 12 = plant height (cm), 13 = number of fruits per cluster, and 14 = fruit shape index.

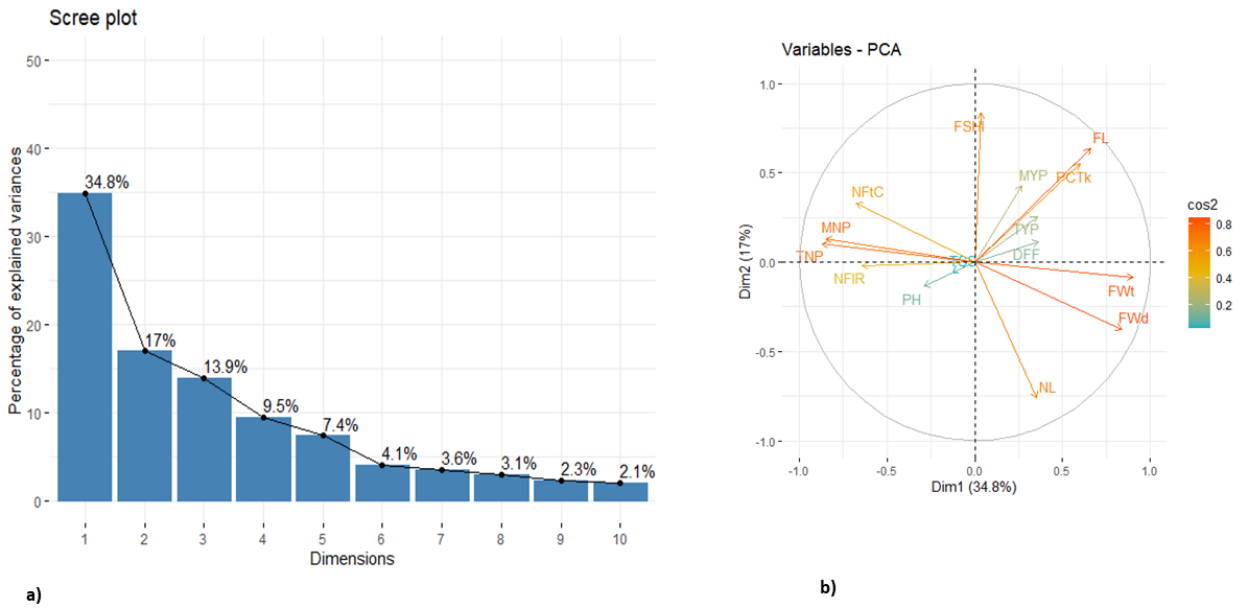

Figure 3.2: a) Scree plot for the 10 principal components of 15 quantitative traits. b) cos2 shows the quality of representation of each trait in the principal component; a high value of cos2 indicates a good representation of the variable in the PCs, and these variables are the most important ones, while a low value of cos2 indicates the variable in the PCs is not perfectly represented. For a given variable, the sum of cos2 will be one if the variable in each PC is perfectly represented.

3.3.6 Genetic variability based on qualitative traits

3.3.6.1 Frequency distribution, chi-square test, and Shannon diversity index

Analysis of floral traits revealed genetic diversity for flower type, style position, and hairiness (Table 3.8). About 95.1% of the genotypes had a uniparous inflorescence type, 68.5% had an inserted stigma position, and 80.4% had style hairiness. Some genotypes (4.2%, 27.3%, and 19.6%) had uniparous and multiparous inflorescence types, style positions at the same level as the stamen, and no style hairiness, respectively. The chi-square test indicated that all flower traits showed highly significant ($p<0.01$) deviation from expected distribution.

Genetic diversity was also detected for growth pattern traits such as growth type, plant size, foliage density, stem pubescence, and leaf type (Table 3.9). It was observed that 88.1%, 77.6%, 63.6%, 72%, 97.2%, and 55.9% had determinate growth habits, intermediate plant sizes, intermediate stem pubescence densities, intermediate foliage densities, semi-erect leaf attitudes, and standard leaf

types. The remaining 8.4%, 12.6%, 27.3%, 21%, 2.1%, and 26.6%, respectively, exhibited indeterminate growth habits, large plant sizes, sparse stem pubescence densities, and sparse foliage densities. The remaining genotypes showcased semi-indeterminate growth habits, small plant sizes, dense stem pubescence densities, dense foliage densities, horizontal leaf attitudes, and potato and Pimpinellifolium leaf types. The chi-square test indicated that all growth pattern traits showed highly significant ($p<0.01$) deviation from expected distribution.

Fruit shape among genotypes had 7 classes, where large proportions of the genotypes had cylindrical (oblong) fruit shapes (Table 3.10). The second and third largest proportions of genotypes had rounded and highly rounded fruit shapes. The remaining fruit shapes included slightly flattened (7.7%), plum-shaped (6.3%), flattened (4.9%), heart-shaped (2.1%), and pear-shaped (0.7%). A majority of the genotypes displayed desirable red pericarp (96.5%) and flesh color (95.8%), while the remaining exhibited pink (0.7%) and orange (2.8%) colors for both pericarp and flesh. Fruit firmness (58.7%) was firm, 40.6% had intermediate firmness, and 0.7% had low firmness. It was further observed that genotypes had a flat blossom end, with only 2.8% displaying an indented end and 18.9% exhibiting a pointed blossom end fruit shape.

The analysis of early-stage immature fruit traits (Table 3.10) revealed that a majority of the genotypes exhibited a light green fruit color. Specifically, 12.2% of the genotypes displayed a green fruit color, while 7% exhibited a whitish-green fruit color. Notably, 86.7% of the genotypes lacked a green shoulder, with the remaining 13.3% exhibiting this trait to varying degrees. The size of the green shoulder varied among the genotypes, with 3.5% displaying a small green shoulder, 7% displaying a medium green shoulder, and 2.1% displaying a large green shoulder. Furthermore, the intensity of the green shoulder varied as well, with 0.7% of genotypes exhibiting

a slight green shoulder, 4.2% exhibiting an intermediate green shoulder, and 8.4% exhibiting a strong green shoulder. The chi-square test indicated that all the immature and mature stage fruit traits showed highly significant ($p<0.01$) deviation from expected distribution.

Table 3.8: Chi-square test and frequency distribution of flower traits

Traits	Descriptors	Frequency distribution (%)	χ2	H'
Inflorescence type	Generally multiparous	0.7	245.79**	0.22
	Both	4.2		
	Generally uniparous	95.1		
Style position	Slightly exerted	4.2	91.14**	0.75
	Same level as stamen	27.3		
	Inserted	68.5		
Style hairiness	Absent	19.6	52.93**	0.49
	Present	80.4		

** Highly significant at $p<0.01$, **χ2** = chi-square, H' = Shannon diversity index

Table 3.9: Chi-square test and frequency distribution of growth pattern traits

Traits	Descriptors	Frequency distribution (%)	χ2	H'
Growth type	Semi-determinate	3.5	193.61**	0.44
	Indeterminate	8.4		
	Determinate	88.1		
Plant size	Small	9.8	126.39**	0.69
	Large	12.6		
	Intermediate	77.6		
Stem pubescence density	Dense	9.1	66.18**	0.86
	Sparse	27.3		
	Intermediate	63.6		
Foliage density	Dense	7	100.55**	0.75
	Sparse	21		
	Intermediate	72		
Leaf Attitude	Horizontal	0.7	262.55**	0.14
	Drooping	2.1		
	Semi erect	97.2		
Leaf type	Pimpinellifolium	8.4	85.17**	1.1
	Potato leaf	9.1		
	Peruvianum	26.6		
	Standard	55.9		

** Highly significant at $p<0.01$, **χ2** = chi-square, H' = Shannon diversity index

Table 3.10: Chi-square test and frequency distribution of immature and mature stage fruit trait

Traits	Code	Descriptors	Frequency distribution (%)	$\chi 2$	H'
Exterior color of immature fruits		Greenish-white	7	151.65**	0.6
		Green	11.2		
		Light green	81.8		
Presence of green shoulder		Present	13.3	77.1**	0.39
		Absent	86.7		
Intensity of green shoulder		Slight	0.7	292.16**	0.5
		Intermediate	4.2		
		Strong	8.4		
		Null	86.7		
Extent of green shoulder		Large	2.1	297.81**	0.5
		Small	3.5		
		Medium	7		
		Null	87.4		
Fruit firmness		Soft	0.7	75.62**	0.71
		Intermediate	40.6		
		Firm	58.7		
Fruit blossom end shape		Indented	2.8	135.79**	0.61
		Pointed	18.9		
		Flat	78.3		
Exterior color of mature fruit		Pink	0.7	256.88**	0.17
		Orange	2.8		
		Red	96.5		
Flesh color of pericarp		Pink	1.4	251.17**	0.2
		Orange	2.8		
		Red	95.8		
Predominant fruit shape		Pyriform (pear shape)	0.7	225.28**	1.5
		Heart shape	2.1		
		Flattened	4.9		
		Ellipsoid(plum shaped)	6.3		
		Slightly flattened	7.7		
		High rounded	11.2		
		Rounded	14.7		
		Cylindrical (long oblong)	52.4		

** Significant at $p<0.01$, $\chi 2$ = chi-square, H' = Shannon diversity index

The estimated Shannon diversity index for the 18 qualitative traits (Tables 3.8, 3.9, and 3.10) indicated that H' ranged from 0.14 for leaf attitude to 1.5 for fruit shapes. The H' was high for style position (0.75), plant size (0.69), stem pubescence density (0.86), foliage density (0.75), leaf

type (1.1), immature fruit color (0.6), fruit firmness (0.71), and fruit blossom end shape (0.61). Intermediate for style hairiness (0.49), growth type (0.44), the intensity of green shoulder (0.5), and extent of the green shoulder (0.5). It was lowest for inflorescence type (0.22), leaf attitude (0.14), presence of green shoulder (0.39), exterior color of mature fruits (0.17), and flesh color of pericarp (0.2).

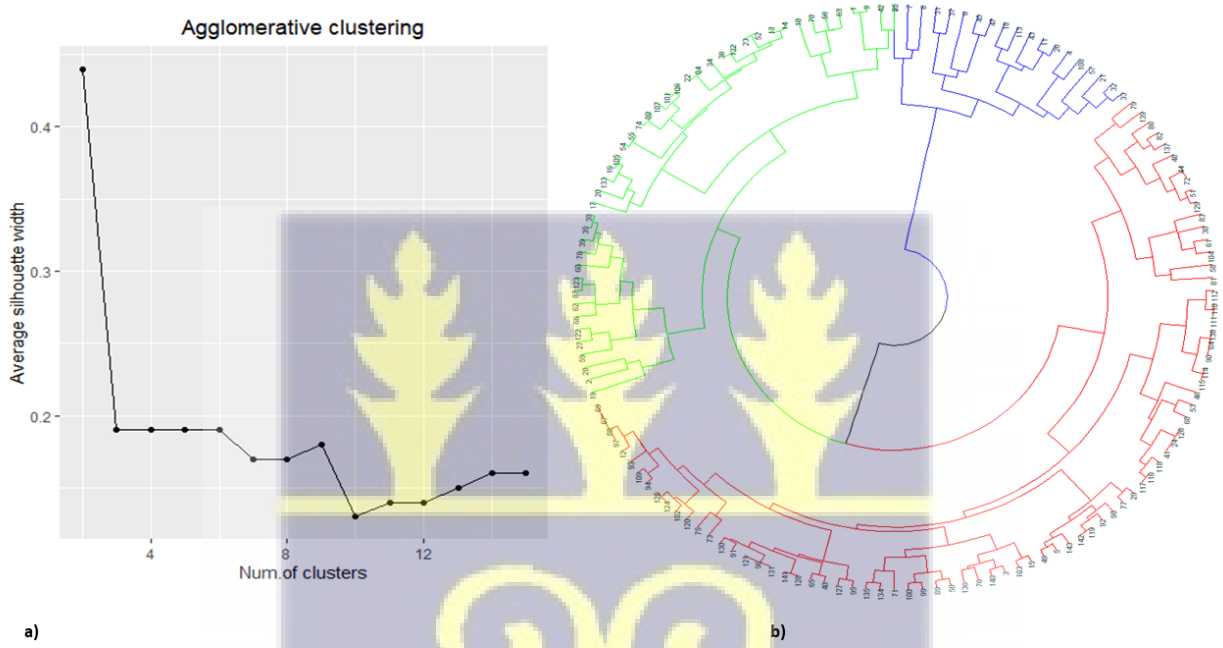

Figure 3.3. a) Average silhouette width indicating the optimal number of clusters. c) Dendrogram showing the grouping of 143 tomato genotypes based on 18 qualitative traits. Genotypes are color-coded by their final cluster assignment: Cluster I (blue), Cluster II (red), and Cluster III (green)

3.3.6.2 Cluster and multiple correspondence analyses (MCA)

Three clusters were identified based on qualitative traits (Figures 3.3 a, b, c). Group I included 19 genotypes, Cluster II had 80, and Cluster III had 44. The MCA biplots explained 18.5% of the variation mainly due orange skin and flesh color, drooping type petiole attitude, the presence of green shoulder being high, medium, and low extent and intensity of green shoulder, indeterminate growth habit, oblong and flattened fruit shape, foliage density, fruit firmness, and stem pubescence density being sparse and dense (Figure 3.4).

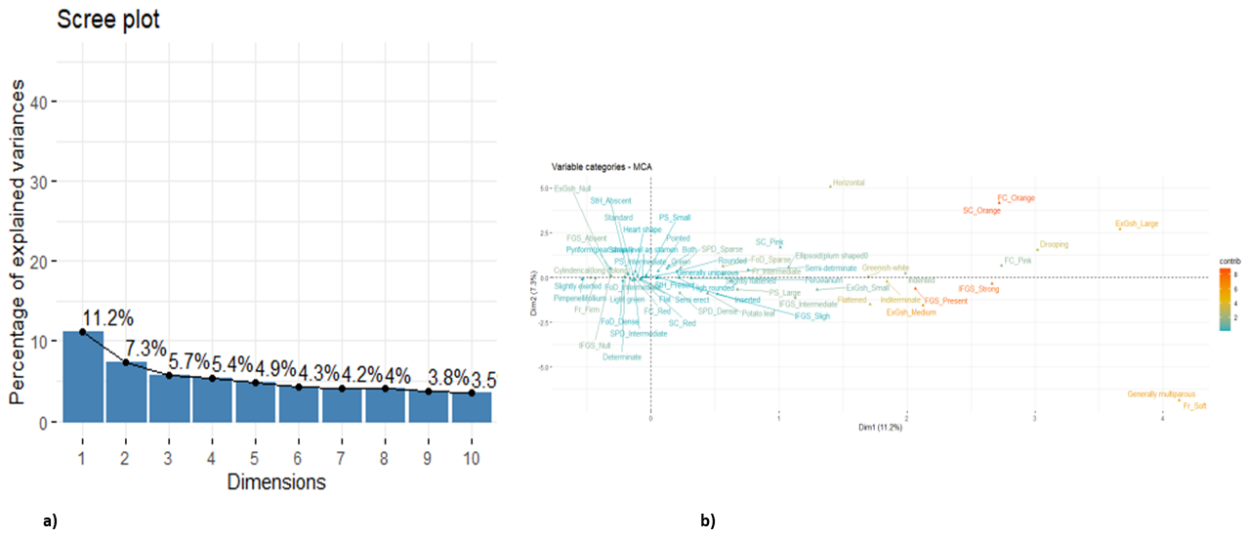

Figure 3.4: Contribution of qualitative variables to the two PCs

3.4 Discussion

Developing varieties with ideal phenotypes such as high yield, quality, and ability to withstand biotic and abiotic stress necessitates genetic variation and diversity in the desired traits (Franco et al., 2001; James, 2022). Knowing the breadth and structure of genetic diversity in accessible germplasm is the first and most significant operation in the early stage of plant breeding. Furthermore, crop improvement initiatives must consider the extent and nature of the gene action involved in the intended trait expression. This information is required for the sort of breeding procedure to be used, the type of cultivar to be developed, and where to begin and cease selection. The magnitude and direction of the relationship between traits are critical for their simultaneous selection and improvement. Several researchers found that a significant amount of genetic diversity exists in cultivated tomato and its wild relatives, while domestication and continuous selection for traits of interest have reduced genetic variation in cultivated tomato compared to the wild species (Bai & Lindhout, 2007b).

The current study explored the genetic diversity of 143 tomato genotypes for 15 quantitative and 18 qualitative traits. Highly significant ($p<0.01$) and significant ($p<0.05$) genetic variations among 15 quantitative traits were found. Genotype x season interaction had a significant effect on the expression of NL, MNP, MYP, TNP, FWt, and FSHI due to variations in temperature and relative humidity during the experimental seasons. Similarly, substantial genetic variation has been discovered across tomato genotypes based on 18 qualitative traits.

Previous studies have documented substantial genetic diversity within tomato germplasm for key agronomic traits, including fruit morphology (e.g., weight, dimensions, shape, and locule number), color, growth habit, and yield components (Daniela Ganeva, 2020; Figàs *et al*., 2015; Grozeva *et al*., 2021a; Mungai *et al*., 2021). The trait ranges reported in the literature such as a yield per plant of 99.79–5888 g, fruit weight of 1.23–576.6 g, and locule number of 1–18 closely align with the variation observed in the present study. Similarly, significant qualitative phenotypic variation has been identified for plant architecture (e.g., plant size, growth habit, stem pubescence, and leaf characteristics), floral traits (e.g., inflorescence type and style morphology), and fruit features (e.g., the presence and intensity of a green shoulder) (Figàs *et al*., 2015). Due to their high heritability and stable expression across environments, these qualitative traits are highly suitable for use as phenotypic markers in targeted breeding programs. Furthermore, their consistent expression makes them particularly valuable for Distinctness, Uniformity, and Stability (DUS) testing, a system routinely used for cultivar identification (Sushma *et al*., 2021).

Phenotypic selection alone is often unreliable due to confounding environmental influences; therefore, partitioning observed variance into its heritable and non-heritable components is fundamental for precise breeding. The magnitude of environmental influence can be quantified

using phenotypic (PCV) and genotypic (GCV) coefficients of variation. In this study, the PCV for quantitative traits ranged from 3.82% (DFF) to 58.29% (MNP), while the GCV for the same traits ranged from 2.58% to 51.88%. These findings are consistent with previous reports; for instance, Hassan *et al*. (2021) reported high PCV and GCV for plant height and fruit number per plant, but moderate values for fruit length, fruit width, and yield. Similarly, Rasheed *et al*. (2023) reported high coefficients for the number of fruits per cluster, single fruit weight, and yield, with moderate values for plant height, days to 50% flowering and number of locules.

The high GCV and PCV observed for most traits in our study suggest that phenotypic selection would be effective for their improvement. Conversely, the low coefficients recorded for DFF and total soluble solids (TSS) indicate limited genetic variation and a greater influence of environment on their expression, rendering direct selection less effective for these traits (Alam *et al*., 2022; Chatara *et al*., 2023). Furthermore, the narrow gap between GCV and PCV for traits such as DFF, PCTk, NL, TSS, FWt, FL, and FWd implies a minimal environmental impact on their phenotypic expression. In contrast, the wider gap for TNP, MYP and TYP points to a significant environmental effect modulating these traits.

Heritability quantifies the proportion of total phenotypic variance attributable to genetic variance, though it is important to note that not all genetic components are heritable or fixable in subsequent generations. Consequently, selection strategies that integrate heritability estimates with corresponding genetic advance are more effective (Alam *et al*., 2022). In the present study, broad-sense heritability (H^2) ranged from 29% for MYP to 92% FSHI. The high (H^2 = 79–92%) to moderate (H^2 = 41–46%) heritability observed for most quantitative traits, with the exception of MYP, indicates that the phenotypic variation among genotypes is primarily due to their inherent

genetic differences. Similarly, the genetic advance as a percent of the mean (GAM) varied substantially, from 2.44% for DFF to 84.42% for MNP. For most traits, the combination of high-to-moderate heritability and high GAM (24–84%) suggests that their expression is controlled predominantly by additive gene effects, indicating a high potential for genetic gain through direct phenotypic selection (Chatara *et al*., 2023). This finding aligns with previous research; for instance, Hassan *et al*. (2021) reported high heritability and GAM for fruit length, fruits per plant, and plant height, and moderate values for total soluble solids (TSS). Likewise, Rasheed *et al*. (2023) documented high to moderate heritability and genetic gain for key yield components. Conversely, the low heritability and genetic advance for MNP in our study likely reflect a significant influence of environmental variation on this trait. The low expected genetic advance for DFF, meanwhile, can be attributed to its limited genetic variation, as evidenced by its low genotypic and phenotypic coefficients of variation.

The genetically complex interrelationships among many quantitative traits can complicate breeding and reduce the genetic gain from selection. However, a thorough understanding of the nature and degree of these associations can enhance selection efficiency (Neyhart *et al*., 2019). In this study, a highly significant genetic correlation was observed among most of the 14 horticultural traits. The strong, positive genetic correlations of MYP) and TYP with key quantitative traits such as FWt, FL, FWd, FSH, DFF, PCTk indicate that selection for any of these traits can concurrently improve yield in tomatoes. This aligns with reports of a significant positive correlation between yield and both fruit length and width (Mungai *et al*., 2021).

Conversely, the negative correlation observed between fruit number per plant (MNP, TNP) and yield (MYP, TYP) contrasts with some previous findings (Ghosh *et al*., 2010; Souza *et al*., 2012).

This discrepancy may be attributed to the prevalence of genotypes with high fruit numbers but small fruit sizes in the current study. The findings of the current study are supported by Henareh *et al*. (2015), who also reported a significant negative correlation, which they attributed to the inclusion of many cherry tomato genotypes with small fruits. The strong association between fruit size traits and high marketable and total yield implies that genotypes with larger fruits can achieve higher yields than those with more numerous, smaller fruits, a conclusion supported by similar research (Hernández-Bautista *et al*., 2015).

Furthermore, a negative genotypic relationship was found between yield and TSS, consistent with earlier reports (Henareh *et al*., 2015). In contrast to the findings of Henareh *et al*. (2015), days to first flowering (DFF) showed a significant positive association with yield, suggesting that late-maturing genotypes, potentially those developing larger fruits, may achieve higher yields. For traits of high economic value that have low heritability or are difficult to measure directly, such as fruit yield, indirect selection offers a viable strategy. Genetic gain can be achieved by selecting for easily measurable, highly heritable traits that are genetically correlated with yield (Souza *et al*., 2012).

Principal component analysis (PCA) is a widely used method for dimensionality reduction, enabling the investigation of maximal genetic diversity within large datasets (Das *et al*., 2017). In the current study, the first five principal components collectively explained 82.68% of the total genetic variation for 15 quantitative traits across 143 tomato genotypes. PCA biplots facilitate the visualization of variables and genotypes in a reduced dimension, helping to eliminate incongruent fluctuations and highlight the principal characteristics within the dataset (Alam *et al*., 2021). The key contributors to the observed genetic variation in the current study include MNP, TNP, FL,

FSHI, PCTk, FWt, and Fwd. Consequently, these traits should be prioritized during parental selection. These findings are consistent with previous research. For instance, Cebolla-Cornejo *et al*. (2013) reported that the first six principal components accounted for 78.73% of the overall variation, primarily in traits like cluster number per plant, locule number, fruit weight, and fruit number per plant. Other studies have shown that the first two PCs explained 51.2% of the variation, largely attributed to fruit size and shape, while three PCs captured 71.6% of the cumulative variance, with PC1 alone contributing 50% from traits including days to flowering, maturity, and fruit morphology. Hussain *et al*. (2018) found that PC1 and PC2 explained 37.12% of the variation, driven by phenological and fruit characteristics. Similarly, Ganeva (2020) demonstrated that the first five PCs accounted for 87.5% of the cumulative variance for plant and fruit traits, with the first two PCs contributing 32% and 18.1%, respectively.

Multivariate analysis is a valuable tool for selection in the early stages of breeding, where numerous traits must be evaluated simultaneously (Barth *et al*., 2022). Furthermore, crosses between parents from divergent gene pools can generate heterosis and transgressive segregants, enhancing genetic gain (Geng *et al*., 2021). In this study, cluster analysis based on 15 quantitative traits classified the 143 tomato genotypes into three distinct groups. Cluster III was the largest (n=74), characterized by a long fruit maturity period, thicker pericarps, longer fruits, shorter plants, a high fruit shape index, and high yield derived from a moderate number of large-sized fruits. Cluster II (n=39) was defined by late flowering, fewer fruits per inflorescence and cluster, lower total and marketable fruit numbers, a high number of locules per fruit, and the highest marketable and total yield per plant. In contrast, Cluster I (n=30) comprised genotypes with a short flowering time, thin fruit walls, few locules, and low marketable and total yield. This cluster also exhibited

the highest values for total soluble solids (TSS), plant height, and number of fruits per cluster, but the smallest fruit size traits (weight, length and width).

Similarly, multiple correspondence analysis (MCA) of 18 qualitative traits also grouped the genotypes into three distinct clusters. The primary traits driving this variation were orange skin and flesh color, a drooping petiole attitude, the presence and varying intensity/extent of a green shoulder, indeterminate growth habit, and oblong or flattened fruit shapes. This clustering of tomato germplasm based on quantitative and qualitative traits is consistent with findings from other scholars (Daniela Ganeva, 2020; Henareh *et al*., 2015; Sharma *et al*., 2006). The valuable traits identified across the clusters such as fruit size, pericarp thickness, yield, fruit shape, firmness, and skin color highlight the potential of utilizing these genetically diverse genotypes as parental lines in targeted crossing programs or other pre-breeding activities for tomato improvement.

3.5 Conclusion

The current findings confirm the existence of highly significant genetic variation among the tomato genotypes for the 15 quantitative and 18 qualitative traits studied. For most quantitative traits, high to moderate estimates of broad-sense heritability and genetic advance indicate that their expression is governed primarily by additive gene action, confirming that phenotypic selection would be effective for their improvement. Furthermore, the combination of substantial heritability and genetic advance, coupled with strong positive genetic correlations of fruit weight, length, width, shape index, days to flowering, and pericarp thickness with marketable and total yield, strongly suggests that simultaneous improvement in these traits will directly enhance fruit yield in tomatoes. Conversely, these beneficial fruit traits were negatively correlated with the number of fruits per plant. Multivariate analyses substantiated the genetic diversity, with cluster analysis

grouping the 143 genotypes into three distinct clusters, and principal component analysis (PCA) revealing that the first five principal components accounted for 82.68% of the total variation. The PCA biplots identified FWt, FL, PH, NFtC, NFIF, PCTk, MNP, TNP, NL and FSHI as the key traits driving this genetic variation. This study provides a comprehensive genetic characterization of a diverse tomato germplasm collection, specifically identifying the key heritable yield components and their interrelationships. It pinpoints a core set of measurable traits fruit size and architecture as the primary drivers of yield, offering a practical framework for breeders to enhance selection efficiency in tomato improvement programs.

CHAPTER FOUR

4.0 Molecular genetic diversity and population structure of tomato germplasm

4.1 Introduction

Tomato is a significant vegetable crop in Ethiopia, serving as a source of income, nutrition, raw materials for agro-industries and foreign currency. However, its productivity is limited by diseases, pests, and a lack of varieties with high-yield, fruit quality, and resistance to biotic and abiotic stresses. Research has led to the development of open-pollinated and hybrid varieties, but open-pollinated varieties are declining in yield and hybrid seeds are costly and limited in availability.

Developing varieties that combine high yield, fruit quality and resistance to various biotic and abiotic stresses contributes to improving the productivity of tomato in Ethiopia. In this regard, determining the quantity and pattern of genetic variation found in germplasm would help in the identification of superior genotypes or breeding lines (Govindaraj *et al*., 2023). Morphological and molecular markers have been used to analyze the extent and pattern of genetic diversity (Nadeem *et al*., 2018). Morphological markers are limited in number and influenced by various factors such as, growth stages and environmental factors (Bekele & Bekele, 2014).

Currently, with the evolution of advanced genomic tools, molecular markers are widely applied for diversity assessment in most crop plant species (Bhanu, 2017). As compared to morphological markers, molecular markers provide better power to explore genetic diversity and could be reproducible in any plant part, growth stage, or environment. DNA markers including RAPD, AFLP, ISSR, SSR, SNPs, DART, and retrotransposons have been applied in genetic diversity assessment, each with its own strengths and weaknesses (Nadeem *et al.*, 2018). Amongst these, SNPs are high-throughput and cost-effective technologies that have been applied in map

construction, genetic diversity analysis, genome-wide marker trait association, marker assisted selection, and genomic selection (Xia *et al*., 2019). SNPs have been preferred because of their high level of polymorphism, wide genome coverage, locus specificity, high reproducibility, and fixed physical position on the chromosome (Yan *et al.*, 2010).

Studies have been conducted to explore the genetic diversity of tomato through multivariate analysis of phenotypic traits (Ghosh *et al*., 2010; Grozeva *et al*., 2021b; Hassan *et al*., 2021; Marefatzadeh-Khameneh *et al*., 2021; Mata-Nicolás *et al*., 2020b; Rasheed *et al*., 2023; Salim *et al*., 2020; Wang *et al*., 2020) and molecular markers such as SSR (Athinodorou, Foukas, Tsaniklidis, Kotsiras, & Chrysargyris, 2021; Benor *et al*., 2008; Eider *et al*., 2020; Gharsallah *et al*., 2016; Kaushal *et al*., 2017; Korir e*t al*., 2014; Luisa *et al*., 2012; Zhou *et al*., 2015) and SNPs (Celik *et al*., 2017; Farinon *et al*., 2022; Mungai *et al*., 2021; Sim *et al*., 2011). In Ethiopia, few phenotypic diversity studies were conducted involving a limited number of genotypes (Ene *et al*., 2022; Shushay *et al*., 2014), but no scientific information is available on the molecular genetic diversity. This study was conducted to determine the level of genetic diversity, genetic relationships, and population structure among 187 tomato germplasm.

4.2 Material and methods

4.2.1 Plant materials, leaf sample preparation and high-throughput genotyping

One hundred and eighty-seven tomato genotypes consisting of fresh market (63), processing (112) and cherry types (12) were genotyped using the DArTseq genotyping platform. Leaf samples were collected from seedlings grown in a screen house using 128 cavity seedling trays filled with peat moss media. Five young leaf discs (60–80 mg leaf discs) were taken from four 28-day-old seedlings per genotype using a leaf cutting punch and mat. The leaf cutting mat and punch were

sterilized using 70% alcohol every time after the completion of sample collection from each genotype. Leaf discs were placed in 96-well plates and oven-dried at 35 °C for 12 hours. The dried leaf samples were sent to SEQART Africa, located at the International Livestock Research Institute in Nairobi, for DArTseq SNP genotyping.

DNA extraction was done using the Nucleomag plant DNA extraction kit (Macherey-Nagel, 2014). The genomic DNA quality and quantity were checked on 0.8% agarose gel electrophoresis and the Nanodrop spectrophotometer. The genomic DNA extracted was in the range of 50–100 ng/ul. Genomic DNA libraries were constructed according to Kilian *et al.* (2012). DArTseq complexity reduction method through digestion of genomic DNA using a combination of PstI and MseI enzymes and ligation of barcoded adapters and common adapters, followed by PCR amplification of adapter-ligated fragments. Next-generation sequencing was carried out using Hiseq2500.

Genotyping was performed using Diversity Arrays Technology's DArTseq platform, a high-throughput Genotyping-by-Sequencing (GBS) method (https://www.diversityarrays.com, accessed on 24 January 2024). The DArTsoft14 proprietary pipeline was used to score two distinct types of markers from the sequencing data: (1) SilicoDArT markers, which represent Presence/Absence Variations (PAVs) of specific DNA fragments and are scored in a binary format (1/0 for presence/absence), and (2) SNP markers, which represent single nucleotide polymorphisms and are scored as co-dominant genotypes. Both sets of markers were aligned to the tomato reference genome SL4.0 (Hosmani *et al.*, 2019) to determine their physical chromosomal positions. The SilicoDArT markers, being dominant, are highly suitable for genetic

fingerprinting and broad-scale population structure analysis, while the co-dominant SNP markers provide higher resolution for detailed genetic studies (Sansaloni *et al*., 2010).

4.2.2 Data analysis

4.2.2.1 Imputation and data filtering

An initial set of 18,229 single SNP markers could be extracted from the DArTseq genotyping platform. Monomorphic SNPs were removed, while polymorphic SNPs, including those with unknown positions on chromosomes, were retained for the study. After imputation of the SNPs with Beagle (Browning *et al*., 2021), 4729 highly informative SNPs (51.9%) were employed for the analysis. SNPs with >20% missing sites, <2% minor allele frequency, and >10% heterozygosity were removed using TASSEL software (Buckler *et al*. 2011). Subsequently, 187 genotypes and 4729 SNPs were employed for the molecular diversity and population structure analyses.

4.2.2.2 Level of SNP polymorphism

The number of SNPs on each chromosome was determined. The average nucleotide composition and rate of transition and transversion mutation was computed using the MEGA software (Kumar *et al*., 2018). The SNP density within 1 Mb window size was also determined using MC Package in R software. The polymorphism information content (PIC), minor (MAF) and major (MAR) allele frequencies, observed heterozygosity (H_o) and expected heterozygosity (H_e) were computed using Power marker v 3.25 (Liu & Muse, 2005).

4.2.2.3 Genetic population structure analysis

Genetic population structure, assuming the admixture model and correlated allele frequencies, was computed using the STRUCTURE software version 2.3.4 (Evanno *et al*., 2005). The structure software applies the Bayesian algorithm for individual grouping, which utilizes the Markov Chain Monte Carlo (MCMC) estimation method. Individuals are pre-allocated to a pre-specified number of clusters, followed by re-grouping of individuals based on the extent of variant frequencies estimated in each group (Porras-Hurtado *et al.*, 2013). The analysis was repeated five times independently for each value of K (k = 1 to 10). These replicates are essential to verify that the results are consistent and robust, not dependent on the random starting point of a single analysis. The structure was run with a 50,000 burn-in period followed by 100,000 MCMC iterations after the burn-in period. The Δk was applied to determine the most probable number of subpopulations (K groupings) using web-based structure harvester software (Earl, 2012). Population grouping was inferred based on the highest delta (K) value from the five independent analyses for each k value. Membership of each individual genotype to the inferred sub-population was determined with a membership coefficient of $Q \geq 0.65$, otherwise considered as an admixture. A bar plot was generated to visualize the population structure of the germplasm with the same software. Principal coordinate analysis (PCoA) and Weighed Neighbor-joining tree using simple matching (Mohammadi & Prasanna, 2003) dissimilarity index of allelic data using 3000 bootstraps were computed using Darwin 5.0 software (Perrier & Jacquemoud-Collet, 2020).

3.2.2.4 Genetic diversity parameters

The genetic diversity within and among populations was determined using 4729 highly informative SNPs, for the structured populations and predefined groups (fresh market, processing and cherry types). AMOVA, which is applied for hierarchical partitioning of the genetic diversity among and within population was computed following the distance Condom-Allelic option using

the GenAlEx 6.5 software plugin for Excel (Peakall & Smouse, 2012). F-statistics, including the FST which provides the total standardized genetic differentiation among population, were computed to generate information regarding genetic diversity within and among populations. The following genetic indices: number of different alleles (Na), proportion of polymorphic loci (Pp), effective number of alleles (Ne), Shannon's Information Index (I), observed heterozygosity (Ho), expected heterozygosity (He), and unbiased expected heterozygosity (uHe) were analyzed. Pairwise gene flow, Nei's genetic distance and F_{ST} among populations were also computed using the same software.

$$Ne = \frac{1}{\sum p_i^2}$$

$$I = -1 * \sum (p_i * Ln\,(p_i))$$

$$Ho = \frac{\text{Number of Hets in loci}}{N}$$

$$uHe = \frac{(2N)}{(2N-1) * H_e}$$

$$F = \frac{H_e - H_o}{H_e} = 1 - \left(\frac{H_o}{H_e}\right)$$

$$F_{is} = \frac{(\text{Mean } H_e - \text{Mean } H_o)}{\text{Mean } H_e}$$

$$F_{it} = \frac{H_t - \text{Mean } H_o}{H_t}$$

$$F_{ST} = \frac{(H_t - \text{Mean } H_e)}{H_t}$$

$$Nm = \frac{\left[\left(\frac{1}{FST}\right) - 1\right]}{4}$$

$$Ht = \sum tp_i^2$$

Where *pi* is the frequency of the i^{th} allele for the population, $\sum pi^2$ is the sum of the squared population *allele* frequencies, mean H_e = average expected heterozygosity across the populations, mean H_o = average observed heterozygosity across the populations, H_t = the total expected heterozygosity, *tpi* is the frequency of the i^{th} allele and tpi^2 is the squared allele frequencies for the total

4.3 Result

4.3.1 Level of SNP polymorphism

After filtering and removing monomorphic markers on chromosomes, 187 genotypes and 4729 SNPs were retained for molecular genetic diversity analysis. SNP markers were distributed along all the 12 chromosomes. About 227 SNPs were found to be with no known positions on chromosomes. The highest number of SNPs, 795 (16.8%), was distributed on chromosome 6, while the smallest number of SNPs, 101 (2.13%), was distributed on chromosome 10 (Figure 4.1b). The average SNPs per chromosome was 375. Similarly high density of SNP was observed on chromosomes 6, 2, 1, 9 and 4 (Figure 4.1a).

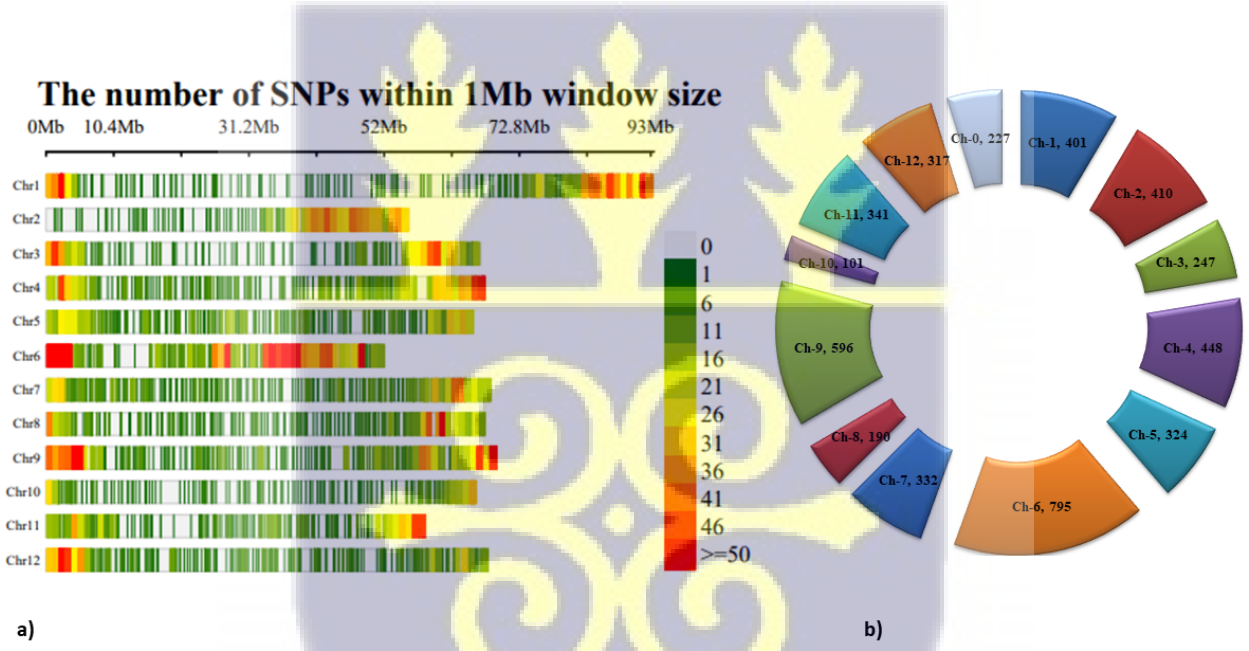

Figure 4.1: a) density and b) genome-wide distributions of SNPs across 12 chromosomes as determined by the DArTseq genotyping platform.

Figure 4.2: The proportion of nucleotide composition across the 12 chromosomes. A = Adenine, T = Thymine, C = Cytosine, G = Guanine

Table 4.1: Types of SNP transition and transversion mutations

	Transitions (si)			Transversion (sv)			si/sv
Substitutions	A/G	T/C	A/T	A/C	T/G	C/G	1.41
Number of alleles	1392	1374	596	489	502	376	
Frequency (%)	29.4	29.0	12.6	10.3	10.6	8.0	
Total	2766 (58.5%)			1963 (41.5%)			

The proportion of nucleotide compositions of A and T was higher compared to those of G and C (Figure 4.2). Additionally, the study showed that transition mutations (TC and AG) were more prevalent (2766, 58.45%) than transversion mutations (1963, 41.5%) (Table 4.1). The AG transition type was slightly more common than the TC transition. The transversion mutation with the highest level of AT was followed by TG, AC, and CG.

The average polymorphism information content value was 0.143, with a range of 0.04 to 0.37. Over 56.5% of SNPs had a PIC value of 0.2 (Table 4.2). The overall range of gene diversity (H_e)

was 0.04 to 0.50, with an average of 0.163., 73.4% of the SNPs had gene diversity value of 0.2. The heterozygosity that was observed had an average of 0.031 and varied from 0.00 to 0. 099.The MAF varied from 0.02 to 0.497 with the average of 0.11. In most cases, it was discovered that the observed heterozygosity was lower than the expected heterozygosity. The SNP markers exhibited substantial gene diversity, PIC and MAF among the tomato germplasm resources in Ethiopia.

Table 4.2: Minor and major allele frequencies, gene diversity, observed heterozygosity, and polymorphic information content of 4729 SNPs across 187 tomato genotypes

Parameter	Minimum	Maximum	Mean
Major allele frequency	0.503	0.98	0.89
Minor allele frequency	0.02	0.497	0.11
Availability/reproducibility	0.83	1	0.99
Gene diversity	0.04	0.5	0.163
Observed heterozygosity	0	0.099	0.031
Polymorphic information content	0.04	0.37	0.143

4.3.2 Genetic diversity and population structure

The genetic population structure of 187 tomato germplasm was analyzed using 4729 SNP markers aimed at identifying the genetic differentiation and relationship. The study revealed two genetically distinct populations (Figure 4.3a) and admixtures, with cluster I and cluster II containing 6 (3.2%) and 172 (91.98%) genotypes respectively (Figures 4b). The remaining 9

(4.81%) genotypes were classified as admixtures. The cherry tomatoes were classified in cluster I, introduced from RDA, South Korea. On the other hand, most of the genotypes in Cluster II consisted of fresh market and processing types. The admixture group had a mix of cherry and fresh market types. Similarly, the PCoA analysis partitioned the germplasm into two well differentiated genetic groups, with $PCoA_1$ (29.02%) and $PCoA_2$ (10.78%) accounting for a significant portion of the total genetic diversity (39.8%) (Figure 4.3d).

4.3.3 Genetic distance and weighted neighbor joining clustering

The Neighbor-joining tree was determined using a simple matching dissimilarity index of allelic data using 3000 bootstraps. The highest genetic distance (0.54323) was found between genotype 166 (cherry tomato) and genotype 179, while the smallest distance (0.00744) was found between genotypes 153 and 156 (Figure 4.3c).

In the pre-defined populations, such as in the processing sub-group, the maximum distance (0.31301) was between genotypes 174 and 145, while the lowest (0.015483) was between genotypes 117 and 116. In the fresh market sub-group, genotypes 183 and 171 had the greatest genetic distance (0.41204), while genotypes 156 and 153 had the smallest genetic distance (0.00741935). In the cherry subgroup, the largest distance (0.452258) was identified between genotypes 02 and 166, while the smallest (0.0744086) was detected between genotypes 163 and 166.

In structured populations, such as population 1 (cherry tomato), the greatest genetic distance (0.19172) was between 162 and 171 genotypes, while the smallest (0.074408) was between genotypes 163 and 166. In subpopulation 2, the greatest distance (0.2965591) was between genotypes 37 and 145 and the smallest (0.0074193) between genotypes 153 and 156. In the

admixture groups, the largest genetic distance (0.3973118) was between genotypes 25 and 174, and the shortest (0.05473118) genetic distance was between genotypes 174 and 176.

The weighted Neighbor-Joining method revealed three distinct genetic clusters, with cluster 1 comprising 56 genotypes, cluster 2 with 3 genotypes and cluster 3 with 128 genotypes.

Out of the 56 genotypes in cluster 1, 47 were lines developed at MARC, 4 were released varieties, and 5 were lines introduced from other sources. All of these lines are of processing types and have oblong fruit shapes, except for the released variety 48 and line 21, which exhibit a round fruit shape. Cluster-1 has been further splited into two subpopulations, with 19 genotypes in subpopulation 1 and the remaining 37 genotypes in subpopulation 2. The three genotypes in cluster 2 exhibited similar plant and fruit morphology, which were introduced from India.

The third cluster contained 128 genotypes, which were further divided into two major subgroups. Subgroup-1 consisted of nine fresh market genotypes, three released varieties (47, 45, and 43) and six lines developed at MARC. Genotypes in this subgroup are distinguished mostly by their huge, spherical fruit form. The second subgroup had 119 genotypes, which were divided into two sub-clusters, with sub-cluster-1 consisting of 26 genotypes. The 26 genotypes in this subgroup comprised the 6 cherry and 7 admixture cherry types grouped by the model-based structure analysis, as well as 13 genotypes introduced from Worldveg (3 genotypes), lines generated at MARC (7 lines), and an unknown source (3 genotypes).

The remaining sub-cluster of cluster-3 consisted of 93 genotypes of fresh market (37 genotypes) and processing (56 genotypes) types brought from various sources, including Worldveg, South Korea's RDA, and MARC lines. Ten varieties released for cultivation in Ethiopia (46, 42, 57, 6,

53, 56, 52, 54, 55, and 49) were grouped within this sub-cluster This sub-cluster included ten released varieties (46, 42, 57, 6, 53, 56, 52, 54, 55, and 49) in Ethiopia. The Weighted neighbor joining classification did not clearly distinguish tomato germplasm with respect to breeding status and fruit morphology, such as size and shape, seed source, or market types.

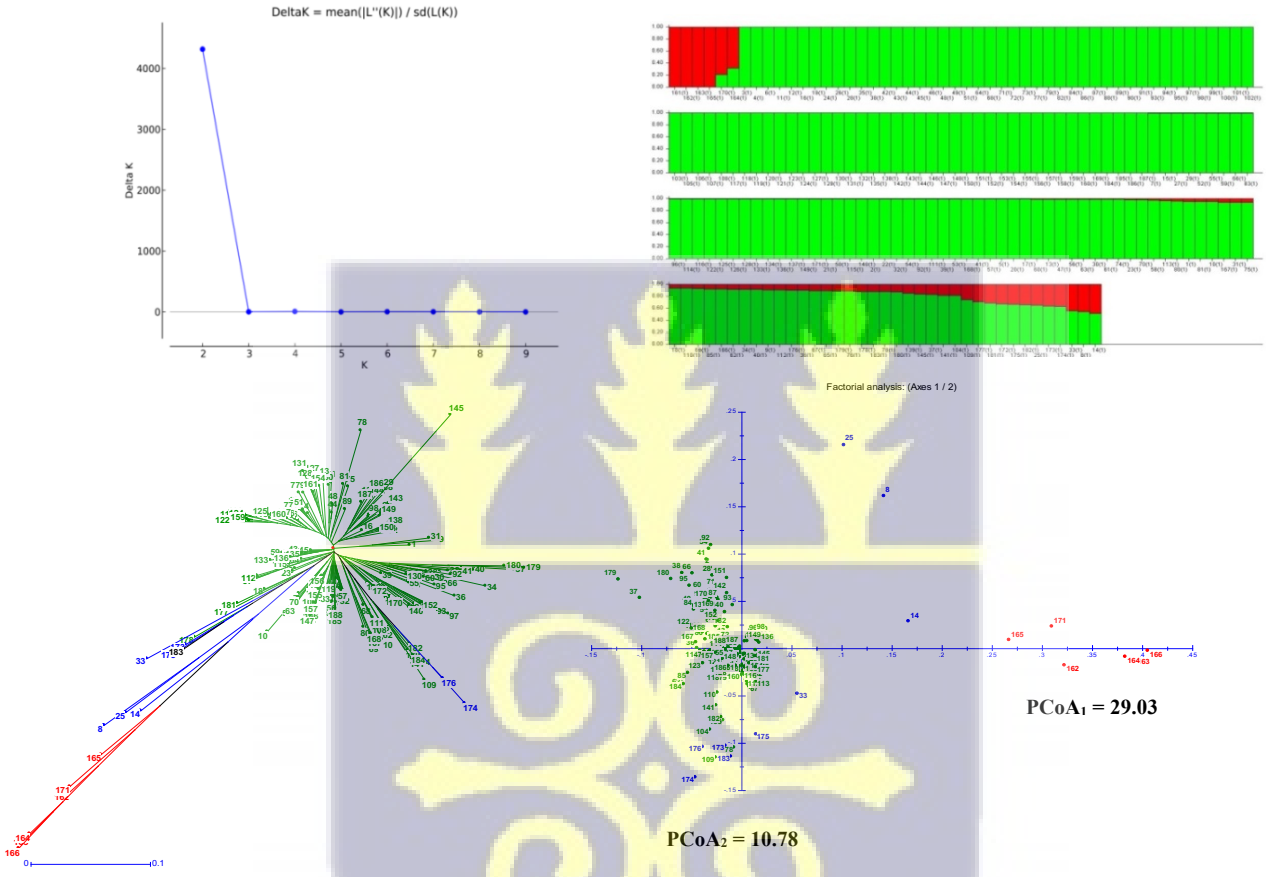

Fig. 4 Population genetic structure of 187 Ethiopian tomato germplasm: (a) ΔK-determined optimal clusters (K=2); (b) STRUCTURE plot (Cluster I: red, II: green, admixed: blue); (c) weighted neighbor-joining tree; (d) Principal Coordinates Analysis (PCoA)

4. 3.4 Analysis of molecular variance (AMOVA)

The AMOVA was computed based on the two structured distinct populations and the admixture as inferred from the population structure analysis, and the pre-defined market types, such as fresh market, processing, and cherry, to investigate the level of genetic distinctiveness for future

breeding works. Variation within the cluster (50%) was comparable to the variation accounted among population (50%) for the structured population, whereas the variation within in predefined population (87%) was much higher than among populations (13%) (Table 4.3).

Table 4.3: AMOVA for the structured and predefined population of tomato germplasm.

Source	DF	SS	MS	Est.	%	F-statistics
Pre-defined population						
Among population	2	10831.7	5415.87	49.52	13	F_{ST} = 0.134 (0.001*)
Among Individual	184	104283.6	566.76	246.88	67	F_{IS} = 0.772 (0.001*)
Within Individual	187	13649.5	72.99	72.99	20	F_{IT} = 0.802 (0.001*)
Total	373	128764.8		369.39	100	Nm = 1.615
Structured population						
Among population	2	17746.2	8873.08	296.19	50	F_{ST} =0.496 (0.001*)
Among individual	184	97369.13	529.2	228.1	38	F_{IS} = 0.758 (0.001*)
Within individual	187	13649.5	72.2	72.2	12	F_{IT} = 0.878 (0.001*)
Total	373	128764.8		597.3	100	Nm = 0.254

** P(rand >= data)*

Analysis of genetic diversity indices among the structured populations revealed that Population 3 possessed the highest gene diversity (0.25), followed by Population 1 (0.14), a group comprised of cherry tomato accessions (Table 4.4).

Table 4.4: Genetic diversity indices among the structured and pre-defined population of tomato germplasm

Population	N	Na	Ne	I	H_o	H_e	uH_e	% of PL
Pre-defined population								
Processing	112	1.858	1.180	0.221	0.028	0.127	0.128	85.8
Fresh market	63	1.970	1.162	0.226	0.032	0.123	0.124	97.0
Cherry	12	1.779	1.491	0.411	0.064	0.277	0.290	77.9

Total	62	1.869	1.277	0.286	0.041	0.176	0.181	86.9
Structured population								
Population -1	6	1.5	1.22	0.22	0.1	0.14	0.15	47.4
Population -2	172	1.9	1.17	0.22	0.03	0.12	0.12	89.4
Population -3	9	1.8	1.41	0.38	0.04	0.25	0.26	75.4
Total	62	1.7	1.26	0.27	0.056	0.17	0.18	70.7

N = number of samples, Na = number of different alleles, PL= polymorphic loci, Ne = effective number of alleles, I = Shannon's information index, H_o = observed heterozygosity, H_e = expected heterozygosity, uH_e = unbiased expected heterozygosity.

Genotypes in population 2 had the lowest expected heterozygosity (0.12). In terms of observed heterozygosity, population 1 had the highest value (0.1), followed by population 3 (0.04) and population 2 (0.03). Population 2 had the highest percentage of polymorphic loci (89.4%), followed by population 3 (75.4%), while population 1 had the lowest percentage (47.4%). The Shannon diversity index was highest for population 3 (0.38), followed by population 2 (0.22) and population 1 (0.22). Additionally, the number of effective alleles was highest for population 3 (1.41), followed by population 1 (1.22) and population 2 (1.17). Similarly, genetic diversity indices among the predefined population indicated that the highest expected heterozygosity was detected in cherry tomatoes (0.277), followed by the processing types (0.127), and fresh market tomatoes (0.123)

In the predefined populations, the fresh market and processing types had the lowest Nei's genetic distance (0.006) and F_{ST} (0.013) due to the high gene flow (N_m = 18.968) (Table 4.5). The highest Nei's genetic distance (0.112) and FST (0.126) were identified between fresh market and cherry types, followed by between processing and cherry types, Nei's genetic distance (0.104) and F_{ST} (0.115), and relatively low gene flow ranging from 1.732 to 1.929.

The highest Nei genetic distance was observed between population 1 and population 2 (0.347), followed by between population 1 and population 3 (0.277) for the structured population. The

lowest genetic distance was observed between population 2 and population 3. A similar pattern was observed for the genetic differentiation value between structured populations, while the pairwise Nei genetic identity and gene flow showed an opposite trend between each pair of populations.

Table 4.5: The pairwise Nei genetic distance, genetic identity, gene flow, and FST among structured and pre-defined populations of tomato germplasm

		Nei D	Nei I	Nm	F_{ST}
Pre-defined population					
Processing	Fresh market	0.006	0.994	18.968	0.013
Fresh market	Cherry	0.112	0.901	1.732	0.126
Processing	Cherry	0.104	0.894	1.929	0.115
Structured population					
Population-1	Pop-2	0.347	0.707	0.125	0.667
Population-1	Pop-3	0.277	0.758	0.339	0.424
Population- 2	Pop-3	0.065	0.937	0.684	0.268

Nei D = Nei genetic distance, Nei I = Nei genetic identity, Nm = Number of migrants, F_{ST} = fixation index

4.4 Discussion

Molecular genetic diversity analysis is a cornerstone of modern crop improvement, providing the means to identify distinct alleles for utilization in breeding programs. A thorough characterization of this variation is essential for uncovering novel allelic sources and for the strategic selection of complementary parental lines. In tomato, as in other crops, genetic diversity and population structure have traditionally been assessed using various molecular markers across a range of genetic resources, including breeding lines, landraces, and wild relatives (Benor *et al*., 2008; Korir *et al*., 2014; Gharsallah *et al*., 2016). Contemporary research, however, increasingly relies on

Single Nucleotide Polymorphism (SNP) markers due to their high density and reproducibility. Sourced from advanced genotyping platforms such as whole-genome sequencing, array-based technologies, and genotyping-by-sequencing (GBS), SNPs provide a robust framework for detailed genetic architecture analysis. This enables a more precise resolution of phylogenetic relationships and the identification of unique, valuable alleles.

The current study investigated the molecular genetic diversity and population structure of 187 tomato accessions using 4,729 informative SNP markers derived from DArTseq. Previous research, often utilizing smaller SNP sets, established that wild species and geographically isolated landraces not modern cultivars harbor the greatest genetic diversity in tomato (Kawicha *et a*l., 2023; Labate, 2021; Pailles *et al*., 2017). These divergent populations, which exhibit high genetic variation, are therefore considered vital reservoirs of novel alleles for breeding more resilient and productive cultivated tomatoes. This work builds upon past efforts that employed substantial numbers of SNPs to characterize tomato germplasm (Ayenan *et al*., 2021; Farinon *et al*., 2022; Kandel *et al*., 2019; Mungai *et al*., 2021; Sim *et al*., 2012). In the present analysis, the average of 375 SNPs per chromosome indicates a comprehensive distribution of markers across the genome. The density of these informative markers reflects the potential genetic variation available for discovery. Furthermore, the SNP density was observed to be higher in the telomeric regions than in the centromeric regions of most chromosomes, a pattern consistent with previous reports (Ayenan *et al*., 2021).

As part of an assessment of the total genetic diversity in the tomato germplasm panel, the polymorphism information content (PIC), minor allele frequency (MAF), expected heterozygosity (He), and observed heterozygosity (Ho) were evaluated. The overall ranges of gene diversity (He:

0.04–0.50), Ho (0.00–0.031), MAF (0.02–0.497), and PIC (0.04–0.37) indicated that the SNP markers displayed considerable genetic diversity across the tomato germplasm. These values align, in most cases, with the findings of Ayenan *et al*. (2021). The observed heterozygosity was lower than the expected heterozygosity, which is consistent with the autogamous nature of tomato.

The transition-to-transversion mutation ratio is commonly used as a general criterion for evaluating SNP quality (Guo *et al*., 2012, 2013). This study identified a higher proportion of transition mutations (TC and AG: 58.45%) compared to transversion mutations (41.5%), yielding a ratio of 1.41:1. A similar pattern of transition-to-transversion mutations has been previously reported in tomato (Ayenan *et al*., 2021), pepper (Solomon *et al*., 2019; Tamisier *et al*., 2020; Taranto *et al*., 2016), and potato (Simko *et al*., 2006).

Model-based structure analysis categorized the 187 tomato germplasm accessions into two distinct genetic groups. The majority of genotypes (172, 91.98%) were assigned to Cluster II, which consisted predominantly of fresh market and processing types. In contrast, only six genotypes (3.2%) were assigned to Cluster I, all of which were cherry tomatoes. The remaining nine genotypes (4.81%) were identified as admixtures, showing genetic contributions from cherry, fresh market, and processing types. Principal Coordinate Analysis (PCoA) corroborated this genetic structure, revealing two distinct clusters that aligned with the taxonomic separation between Solanum lycopersicum var. lycopersicum and Solanum lycopersicum var. cerasiforme. The first two principal coordinates, PCoA1 (29.02%) and PCoA2 (10.78%), together accounted for 39.8% of the total molecular genetic diversity. This pattern is consistent with previous research. For instance, a PCoA analysis of 952 tomato accessions from *Solanum pimpinellifolium* (SP*), Solanum lycopersicum* var. *cerasiforme* (SLC), and *Solanum lycopersicum* var. *lycopersicum* (SLL) using

2313 markers also classified the germplasm into three distinct groups corresponding to each species. In that study, PCoA1 (34.02%) and PCoA2 (5.6%) collectively explained 39.62% of the genetic diversity (Blanca *et al*., 2015).

The findings of this study indicate that, with the exception of the separation of cherry types, neither the model-based (structure analysis) nor the non-parametric approaches (PCoA and weighted neighbor-joining) classified the germplasm into distinct fresh market and processing types. An Analysis of Molecular Variance (AMOVA) further confirmed the lack of clear differentiation, revealing negligible genetic divergence (F_{ST} = 0.013) between these two groups, which can be attributed to a high degree of gene flow (N_m = 19). This genetic homogeneity is likely a consequence of breeding history, where genetic improvement has historically involved the exchange of adaptive traits between fresh market and processing types, including introgression from related and wild species (Bai and Lindhout, 2007).

Consistent with our results, a similar study of 190 tomato germplasm accessions from 31 countries identified three genetic clusters using model-based methods but also found no clear genetic distinction among pre-defined categories such as fresh market, processing, ornamental, breeding lines, landraces, and home garden types (Labate, 2021). This further supports the conclusion that genetic variation in tomato germplasm is poorly associated with a priori groupings based on market type.

In contrast to the findings of the current study, previous research has reported genetic differentiation between fresh market and processing tomato types. For instance, Wang *et al*. (2016) identified significant genetic structure among 348 vintage and contemporary tomato accessions, with a subgrouping of vintage varieties by fruit color that may reflect specific market niches.

Similarly, Sim *et al*. (2011) found genetically distinct groups among 70 tomato germplasms representing contemporary (processing and fresh market), vintage, and landrace varieties, which aligned with predefined market niches and historical age, although a high degree of admixture was noted between the fresh market and processing groups. Blanca *et al*. (2015) also reported a distinction between fresh market and processing types, with further subgrouping within each category. This observed genetic differentiation in cultivated tomato is likely a result of breeding efforts targeting specialized market niches. The discrepancy between our results and these earlier studies may be attributed to differences in the germplasm sampled, as the specific subpopulation selected for analysis can significantly influence the detection of genetic structure.

AMOVA was performed using two sets of groupings: the pre-defined market types and the two genetic sub-populations, including the admixture group, identified by the structure analysis. In both scenarios, the analysis revealed moderate to high and statistically significant genetic differentiation among the subpopulations. Within the pre-defined groups, a moderate ($F_{ST} = 0.143$) and highly significant ($p \leq 0.001$) genetic differentiation was observed between the cherry types and either the fresh market or processing types. In contrast, the differentiation between fresh market and processing types was negligible ($F_{ST} = 0.013$), confirming the lack of a clear genetic distinction based on market type in this germplasm set. This result aligns with the findings from the model-based structure and weighted neighbor-joining analyses. The low level of differentiation was further supported by a very low Nei's genetic distance (0.006) between the fresh market and processing groups, which can be explained by the high gene flow ($Nm = 18.968$) observed between them.

Gene diversity (He) was highest in the cherry types (He = 0.277), followed by the processing (He = 0.21) and fresh market (He = 0.123) subpopulations. This pattern is consistent with the findings of Blanca *et al.* (2015), who reported the highest gene diversity in *Solanum pimpinellifolium* (SP), followed by *Solanum lycopersicum* var. *cerasiforme* (SLC) and *Solanum lycopersicum* var. *lycopersicum* (SLL). Furthermore, the same study noted a slightly higher gene diversity within the processing (He = 0.096) compared to the fresh market (He = 0.091) subgroups of the cultivated tomato.

4.5 Conclusion

A comprehensive genomic analysis of 187 Ethiopian tomato germplasm, using 4,729 SNP markers, revealed a hierarchical population structure defined by two primary genetic clusters. The predominant cluster (92.0% of germplasm) encompasses both fresh market and processing types, exhibiting remarkably low genetic differentiation (F_{ST} = 0.013) and high gene flow (Nm = 19). This indicates extensive germplasm exchange during breeding. In contrast, a distinct minor cluster (3.2% of germplasm), dominated by cherry tomatoes, shows significant divergence from the main group (F_{ST} > 0.115). The overall collection displays low-to-moderate genetic diversity (He = 0.163; PIC = 0.143), consistent with global cultivated tomatoes and reflecting domestication bottlenecks and self-pollination. Notably, cherry tomatoes exhibit substantially higher diversity (He = 0.277) than processing (He = 0.127) or fresh market types (He = 0.123). As the first detailed genomic characterization of Ethiopia's tomato germplasm, this study establishes a critical foundation to: (1) identify molecular markers for key agronomic traits (yield, quality, stress resistance), and (2) strategically exploit the existing diversity within and between clusters for breeding. This enables the development of locally adapted varieties resilient to Ethiopia's unique

growing challenges and market demands, directly addressing the country's significant productivity gap.

CHAPTER FIVE

5.0 Genome wide association study for fruit yield and quality among tomato germplasm

5.1 Introduction

The tomato improvement program focuses on increasing yield and quality. Although traditional breeding methods have led to the development of high-yielding and better-quality tomato varieties, they have been constrained by the time required and the complexity of trait inheritance. Fruit yield is influenced by fruit size, number of fruits, TSS, etc. (Monamodi *et al*., 2013; Rasheed, 2017). Molecular studies indicate that fruit yield is largely determined by fruit size QTLs (Hernández-Bautista *et al*., 2015). Understanding the genetic basis of these traits is crucial for improvement through molecular plant breeding. Recent advancements in molecular breeding, specifically genome-wide association studies (GWAS), provide a new approach to examining the genetic factors that underlie complex traits.

The tomato fruit size and shape genetic architecture have been studied (Bertin *et al*. 2009; FOGLE and CURRENCE 1950; Grandillo, Ku, and Tanksley 1999; Ibarbia and Lambeth 1969). Fruit-size QTLs, such as fw1.1, fw2.1, fw2.2, fw3.1, fw3.2, and fw11.3, were identified and mapped. FW2.2 (Frary *et al*., 2000), FAS (FW11.3) (Z. Huang & van der Knaap, 2011), LC (Muños *et al*., 2011), FW3.2 (Chakrabarti *et al*., 2013), and Fw11.3 (Mu *et al*., 2017) were fine-mapped and cloned. Four to 17 loci were known to determine fruit shape and number of locules (Rodríguez *et al*.,

2011). Among these genes, SUN and OVATE are associated with elongated fruit shapes, while FAS and LC determine the number of locules and flat fruit shapes (Rodríguez *et al.*, 2011). Similarly, 25 QTLs for fruit length, fruit diameter, fruit weight, fruit yield, locule number, and brix degree that accounted for 4.19% to 12.67% of phenotypic variation were identified (Hernández-Bautista *et al.*, 2015) and plant height (Zhou *et al.*, 2016).

GWAS has also been widely applied to identify genomic regions closely associated with yield, yield-related, and fruit quality traits in tomato (Kim *et al.*, 2021; Liu, Jiang and Li, 2023; Phan *et al.*, 2019; Shah *et al.*, 2024;Wang *et al.*, 2020). For instance, 51 QTLs were identified for fruit color, fruit shape, pericarp thickness, fruit weight, fruit width, and fruit length, explaining 7.11 to 37.64% of the phenotypic variance (Phan *et al*., 2019). Similarly, 33 significant QTLs were identified for fruit weight, fruit width, fruit length, fruit shape index, fruit pericarp thickness, locule number, firmness, and total soluble solids (Kim *et al*., 2021). Furthermore, 44 marker-trait associations were detected for 19 metabolic traits(Sauvage *et al*., 2014).

Quantitative traits such as yield and fruit quality are polygenic in nature, and the identification of their genetic bases using molecular tools is greatly influenced by the genetic background and interactions, including gene–environment interactions (GEI) and epistasis. Therefore, identifying additional and stable QTLs is essential to deepen our understanding of these genetic bases and to support their effective utilization in breeding programs (Wang et al., 2016). Moreover, no similar study has been conducted on Ethiopian tomato germplasm. Hence, this study aims to identify QTLs that can make a significant contribution to the improvement of tomato breeding programs in Ethiopia. Specifically, the objective was to identify QTLs associated with yield, yield-related traits, and fruit quality traits.

5.2 Material and methods

5.2.1 Phenotyping

One hundred and forty-three tomato germplasms, including 125 breeding lines and 18 improved varieties, as described in Chapter 3, were employed for the study. Field phenotyping and method of data analysis were identical to those outlined in Chapter 3. Fourteen traits with high broad-sense heritability and a highly significant positive genetic correlation with fruit yield were employed. Days to 50% flowering (DFF), fruit pericarp thickness (PCTk), number of locules (NL), total soluble solids (TSS), fruit weight (FWt), fruit length (FL), fruit width (FWd), plant height (PH), number of marketable fruit per plant (MNP), total number of fruit per plant (TNP), fruit shape index (FSHI), marketable yield per plant (MYP), and total yield per plant (TYP) were considered.

5.2.2 Genotyping

One hundred and forty-three tomato germplasms that were part of 188 tomato germplasm genotyped at SEQART Africa, as described in Chapter 4, were utilized for genome-wide association studies (GWAS). The procedures for DNA extraction and genotyping are thoroughly discussed in Chapter 4.

5.2.3 Imputation and SNP data filtering

An initial set of 18,229 single SNP markers was extracted from the DArTseq genotyping platform. Monomorphic and SNPs with unknown positions on chromosomes were removed. SNP markers with a missingness of >10% and individuals or taxa with >10% genotype missingness were

discarded. Moreover, SNPs with minor allele frequency of ≤3% and sites and individuals with ≥5% heterozygosity were identified using TASSEL software (Buckler et al., 2011) and were discarded from further analysis. Finally, 127 genotypes and 2,709 informative SNPs were employed for GWAS analysis.

5.2.4 Linkage disequilibrium (LD) and decay

Linkage disequilibrium between all pairs of markers on each chromosome was determined using TASSEL software v5.0 (Buckler *et al*., 2011). A total of 2,709 SNPs with MAF greater than 3% were used. The same software was used to calculate pairwise r^2 and the physical distance between all pairs of markers on each chromosome. R software was used to plot r^2 values with physical distance, and LD decay was computed using a scatter plot with r^2 set at 0.2.

5.2.5 Genome wide marker-trait association analysis

Genome-wide marker-trait association analysis was conducted using the genetic analysis software, Genomic Association and Prediction Integrated Tool (GAPIT) package version 3 (J. Wang and Zhang 2021), in the R software. Multi-locus models (Wang and Zhang 2021), such as Multiple Loci Mixed Linear Mixed Model (MLMM), Fixed and Random Model Circulating Probability Unification (FarmCPU) (Xiaolei Liu *et al*., 2016), and Bayesian information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) (Huang *et al*. 2019), were used to identify significant marker-trait associations. Principal component analysis (PCA) and pairwise genetic relatedness (kinship matrix) were employed as covariates in the MLMM and Farm CPU multilocus models to control false positives due to unknown population structure and family relatedness. The PCA and kinship matrix were computed using TASSEL version 5.0 software (Buckler *et al.*, 2011).

Results from GWAS were examined and graphically visualized through the Quantile-Quantile (Q-Q) and Manhattan plots; the Q-Q plot graphically visualizes the fitness of the models and the presence of significant marker-trait associations, while the Manhattan plot depicts significant associated genomic regions on each chromosome. Manhattan and Q-Q plots were computed using the same package, GAPIT, in the R software environment. The default p-value cutoff ($\alpha = 0.01$) of the Bonferroni correction implemented in the GAPIT software was used to identify significant associations between SNP markers and traits in the Manhattan plots

5.3 Results

5.3.1 Genetic variability

A highly significant ($p<0.01$) genetic variation was detected among genotypes for the 14 quantitative traits (Chapter 3). The genotype-by-season interaction was also highly significant ($p<0.01$) for NL, MNP, MYP, TNP, TYP, FWt, and FSHI. The range between minimum and maximum values indicated the presence of genetic variability among the 14 quantitative traits (**Table 5.1**). Broad-sense heritability ranged from 29% to 92%: high (79–92%) for PCTk, NL, MNP, TNP, FWt, FL, FWd, PH, and FSHI; medium (41–59%) for DFF, TSS, TYP, and NFtC; and low for MYP (29%).

5.3.2 Phenotypic correlations

Highly significant ($p<0.01$) phenotypic correlation was detected for most of the traits (Figure 5.1). Marketable yield exhibited a strong correlation with TYP (r_p = +0.893), FWt (r_p = +0.359), FL (r_p = +0.408), FWd (r_p = +0.248), FSH (r_p = +0.218), PCTk (r_p = +0.402), and TSS (r_p = -0.171). Furthermore, a highly significant ($p<0.01$) phenotypic correlation was observed between total yield per plant and FWt (r_p = +0.461), FL (r_p = +0.358), FWd (r_p = +0.384), NFtC (r_p = -0.104), DFF (r_p = +0.108), PCTk (r_p =+0.361), NL (r_p = +0.11), TSS (r_p = -0.107), and PH (r_p = 0.167).

The study also identified highly significant positive correlation among PCTk, FWt, FL, FWd, and DFF. On the other hand, fruit size traits such as FWt (r_p = -0.729, r_p = -0.740), FL (r_p =-0.506, r_p =-0.548), FWd (r_p =-0.739, r_p =-0.734), PCtK (r_p =-0.492, r_p =-0.527), and NL (r_p =-0.310, r_p =-0.272) exhibited a highly significant negative correlation with MNP and TNP, respectively. The phenotypic correlation between marketable yield and marketable number per plant (r_p = 0.071) and total yield and total number of fruits per plant (r_p = -0.037) was non-significant.

Table 5.1: Range and broad-sense heritability of 14 quantitative traits of tomato genotypes

Traits	Range	CV (%)	H_b^*
DFF[1]	32-36	4.98	46
PCTk[2]	3–8	15.87	85
NL[3]	2-7	18.84	85
TSS[4]	4.07-4.96	13.35	41
MNP[5]	18-172	25.0	79
MYP[6]	1969.03-2967.35	32.4	29
TNP[7]	32-224	25.9	80
TYP[8]	2669.46-4852	31.0	44
FWt[9]	23.53-150.63	12.91	85
FL[10]	33.06-86.47	10.92	91
FWd[11]	31.08-66.04	9.44	87
PH1[2]	48.32-148.33	15.06	90
NFCt[13]	3-5	16.7	59
FSHI[14]	0.69 -1.99	10.48	92

Hb = heritability in the broad sense, CV = coefficient of variation; 1 = days to 50% flowering, 2 = pericarp skin thickness (mm), 3 = number of locules, 4 = total soluble solids, 5 = marketable fruit number per plant, 6 = marketable yield per plant (g), 7 = the total number of fruits per plant, 8 = total yield per plant (g), 9 = fruit weight (g), 10 = fruit length (mm), 11 = fruit width (mm), 12 = plant height (cm), 13 = number of fruits per cluster, and 14 = fruit shape index.

5.3.3 Population structure and kinship matrix

Principal component analysis revealed the existence of three subpopulations (Figures 5.2a, b, and c). The screen plot was used to determine the significance of the three principal components, which explained the largest cumulative percentage variance of 30%. The kinship matrix heat map exhibited three distinct groups (Figure 5.2d).

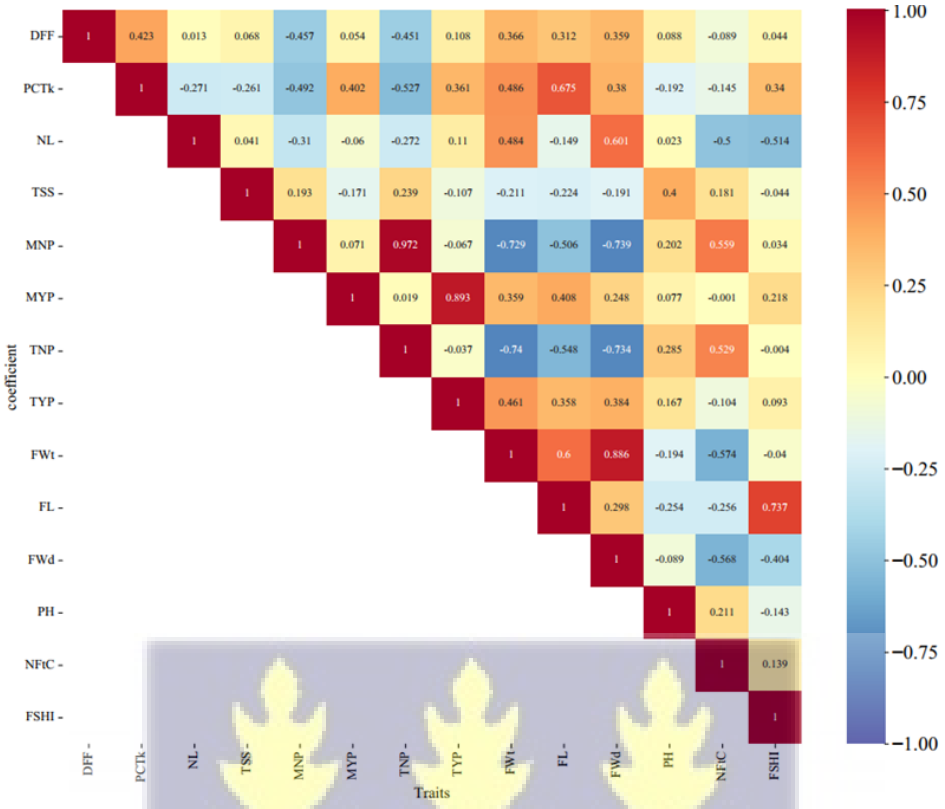

Figure 5.1: Heat map showing the association coefficients of 14 quantitative traits in tomato germplasm

5.3.4 Linkage disequilibrium (LD) and decay

In the genome-wide linkage disequilibrium plot (Figure 5.3a), SNPs exhibiting substantial LD are highlighted in red, signifying that these alleles are co-inherited. In the current study, the r^2 ranged from 0 to 1, with an average of 0.22 (Figure 5.3a). The entire LD decay value at $r^2 = 0.2$ was 2.052 Mbp (Figure 5.3b), ranging between 0.157 Mbp for chromosome 07 and 28.826 Mbp for chromosome 09 (Table 5.2).

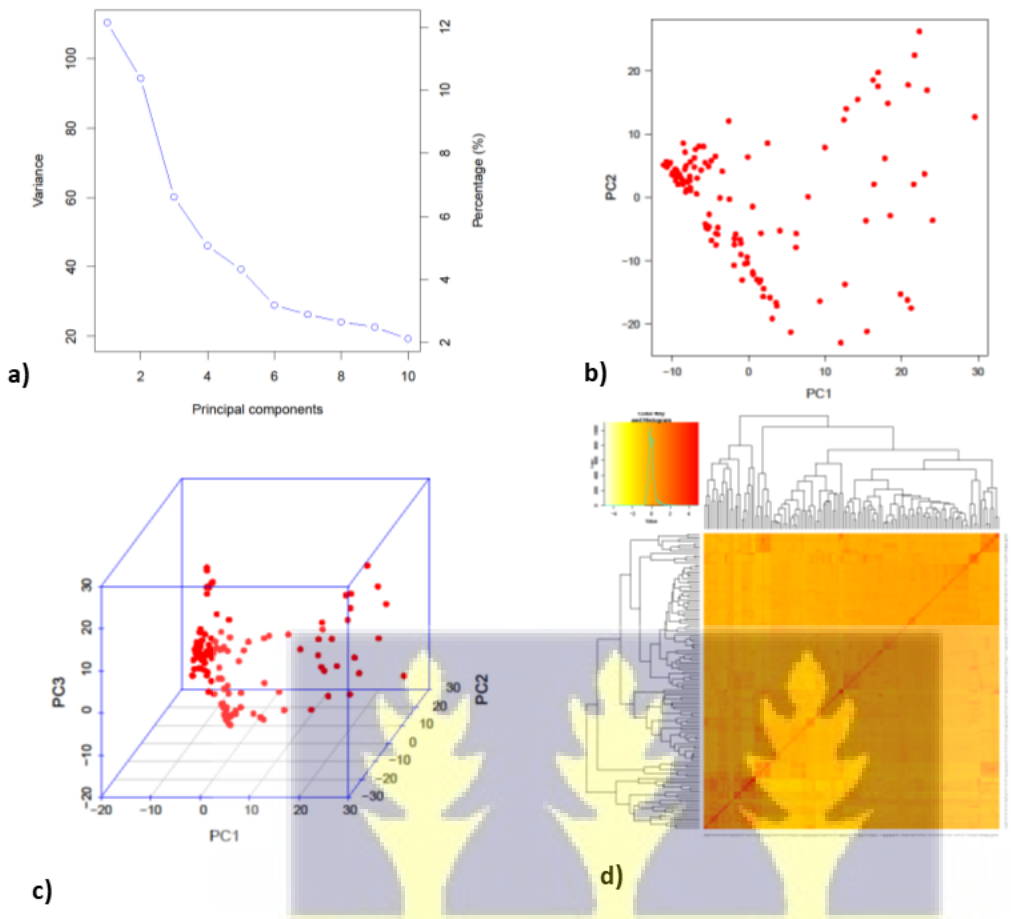

Figure 5.2: Principal component analysis and kinship matrix for 127 tomato germplasm using 2709 SNP markers. a) PCA screen plot displaying the number of principal components and the percentage of variance explained. b) PCA of tomato germplasm based on two PCs; c) PCA based on three PCs, with each dot representing a single tomato genotype. d) a heat map of the kinship matrix between each pair of tomato germplasm.

5.3.5 Marker trait association

In a genome-wide association study using three multi-locus mixed models, 121 significant MTAs with false discovery rate (FDR) adjusted p-values less than 0.05 (Table 5.3) were identified for ten quantitative traits over two seasons. The 121 MTAs were associated with 104 significant QTLs based on the LD decay for each chromosome. However, no significant MTAs were found for MYP, TYP, NFTC, or DFF. The highest number of MTAs (23) was detected on chromosome 01, and the lowest was on chromosome 05.

Furthermore, 29 common MTAs were discovered using two or three multilocus models. Sixteen common MTAs, or pleiotropic loci, were also identified (Table 5.4). For example, TNP and MNP had nine marker-trait associations, PH and MNP had one, and FL and FSHI had two. Furthermore, one marker-trait association was found between PH, TNP, and MNP, one between FWd and MNP, one between FWd and TNP, and one between FSHI and TNP.

The Manhattan scatter plot (Figure 5.4a) displayed the significant SNPs along their corresponding positions on chromosomes. The expected -log (p-value) across all markers compared to the observed -log (p-value) is indicated in the Q-Q plot (Figure 4b). The null hypothesis that there is no significant association or causal polymorphism was rejected for most traits except for MYP, TYP, NFtC, and DFF. The Q-Q plot shows a straight line with a sharp upward deviation, indicating that the multi-locus models adequately controlled both false positives and negatives, signifying real significant MTAs. This is because the population structure (principal component analysis) and the relatedness (kinship matrix) were captured by the models. The non-significant marker p-values have a uniform distribution due to the absence of linkage disequilibrium (LD) between them; the p-values of the significant markers deviate from the straight line, as indicated in the figures.

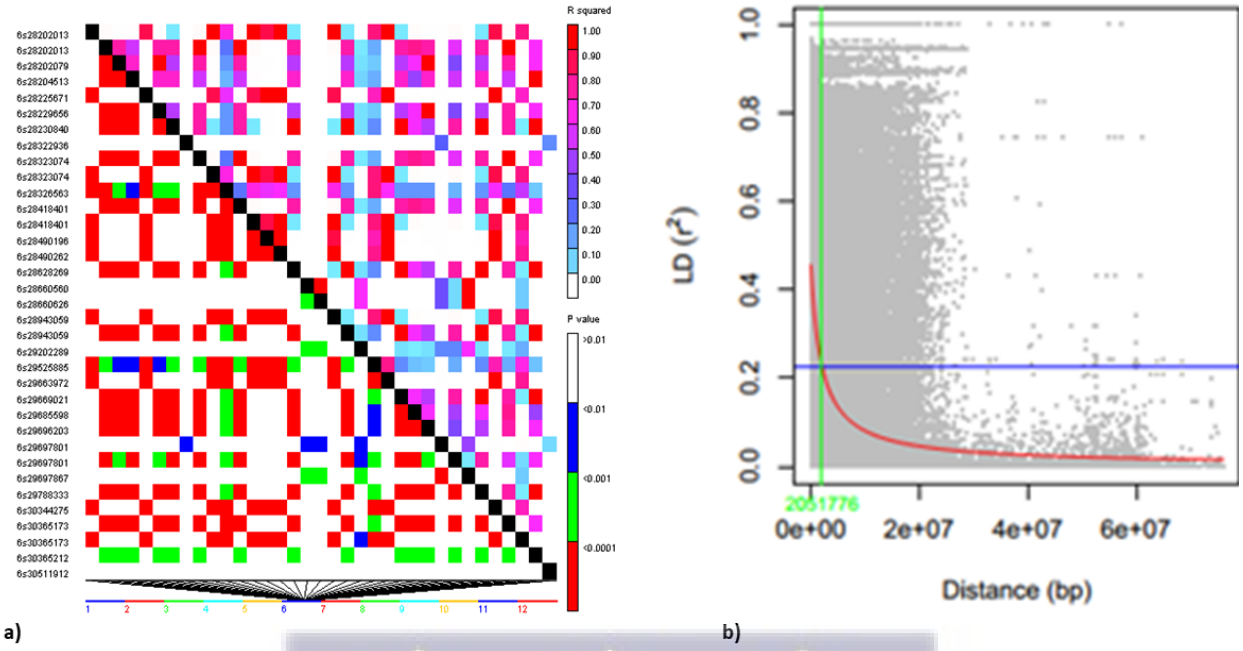

Figure 5.3 :(a) linkage disequilibrium pattern deploying 2,709 SNP markers across 127 tomato germplasm collections in Ethiopia; b) plot of overall LD decay across 12 chromosomes at a fixed value of r^2 = 0.2, demonstrating the relationship between SNP base pair distance and LD value (r^2). The red line represents the genome-wide LD decay. The vertical (green) and horizontal (dark blue) lines represent the LD decay distance and half-LD value, respectively. LD decay equals 2.05 Mbp.

Table 5.2: Linkage disequilibrium (LD) decay for the 12 chromosomes (r^2 = 0.2) in tomato germplasm

Chromosome	LD decay (Mbp)
01	1.513
02	0.370
03	0.412
04	1.288
05	14.429
06	1.591
07	0.157
08	0.281
09	28.83
10	0.391
11	1.118
12	5.530
Overall	**2.052**

Table 5.3: The number of MTAs and QTLs identified for 10 quantitative traits using three multi-locus models across 12 tomato chromosomes.

Trait	Number of MTAs	Number of QTLs
FL	7	6
FSHI	8	8
FWd	7	7
FWt	3	3
MNP	25	19
NL	4	4
PCTk	7	7
PH	20	17
TNP	32	26
TSS	8	7
Total	**121**	**104**

FL = fruit length (mm), FSHI = fruit shape index, FWd = fruit width, FWt = fruit weight (g), MNP = marketable fruit number, NL = number of locules, PCTk = pericarp skin thickness (mm), PH = plant height (cm), TNP = total number of fruits per plant and TSS = total soluble solids.

Table 5.4: Common MTAs detected among traits

Trait	Chromosomes	Position (base pair)	Number of MTAs
TNP, MNP	7,10,6,11,4,1	26,773,776; 3,767, 938; 2, 518, 858; 42,339, 982; 52, 699, 841; 56, 075, 153; 62, 261, 938; 84, 019, 411; 92, 092, 536	9
FL, FSHI	2	18,268,340; 49,548,307	2
MNP, PH	3	51,605,018	1
FW, TNP	2	55,409,913	1
FSHI,TNP	4	62,541,098	1
MNP, FWd	8	63,558,538	1
TNP, MNP, PH	1	92,127,603	1
Total			**16**

FL = fruit length (mm), FSHI = fruit shape index, FWd = fruit weight (g), FWd = fruit width, MNP = marketable fruit number, PCTk = pericarp skin thickness (mm), PH = plant height (cm) and TNP = total number of fruits per plant.

5.3.5.1 Fruit length

Seven marker-trait associations ($p \leq 4.4x10^{-5}$, $FDR < 0.05$), which could be clustered into 6 QTLs, were identified on chromosomes 2, 4, 6, and 9 (Table 5.5). Two MTAs on chromosome 09 at 68.706 Mbp and 69.066 Mbp were clustered into one QTL. The BLINK model detected three loci, and the FarmCPU model detected the remaining. The phenotypic variance explained ranged between 1.5% and 7.0%. The most significant marker-trait association (4698369), with a p-value of less than 1.1 x 10^{-6}, was identified on chromosome 2 at 18,268,340 bp. It accounted for the highest phenotypic variance (7.0%). Locus 7839232 on the same chromosome was associated with the second-highest phenotypic variance explained, 5.2%, with a p-value smaller than 3.2 x 10^{-6}. Locus 100096164 on chromosome 6 was identified to account for 4.8% of the phenotypic variance, with a p-value of less than 6.9 x 10^{-6}. The remaining loci (7841051, 7834983, 7755338, and 4695365) corresponded with phenotypic variations of 3.4%, 2.8%, 2.3%, and 1.5%, respectively. Three of the loci showed a negative effect on the expression of fruit length, while the remaining four showed a positive effect.

Table 5.5: Significant marker trait association for fruit length

SNP	Allele	Ch.	Position	P-value	MAF	Effect	Model	PVE	FDR
4698369	T/C	2	18268340	1.1E-06	0.17	-5.4	FarmCPU	7.0	0.00
7839232	T/C	2	54813995	3.2E-06	0.04	5.3	FarmCPU	5.2	0.01
100096164	A/G	6	26787671	6.9E-06	0.06	-5	FarmCPU	4.8	0.01
7841051	C/T	4	3225147	4.4E-05	0.04	5.8	BLINK	3.4	0.04
7834983	C/T	2	49548307	8.6E-06	0.15	-3.9	BLINK	2.8	0.00
7755338	G/T	9	69065763	2.0E-06	0.15	4.2	BLINK	2.3	0.00
4695365	C/T	9	68706460	1.1E-06	0.10	4.3	FarmCPU	1.5	0.00

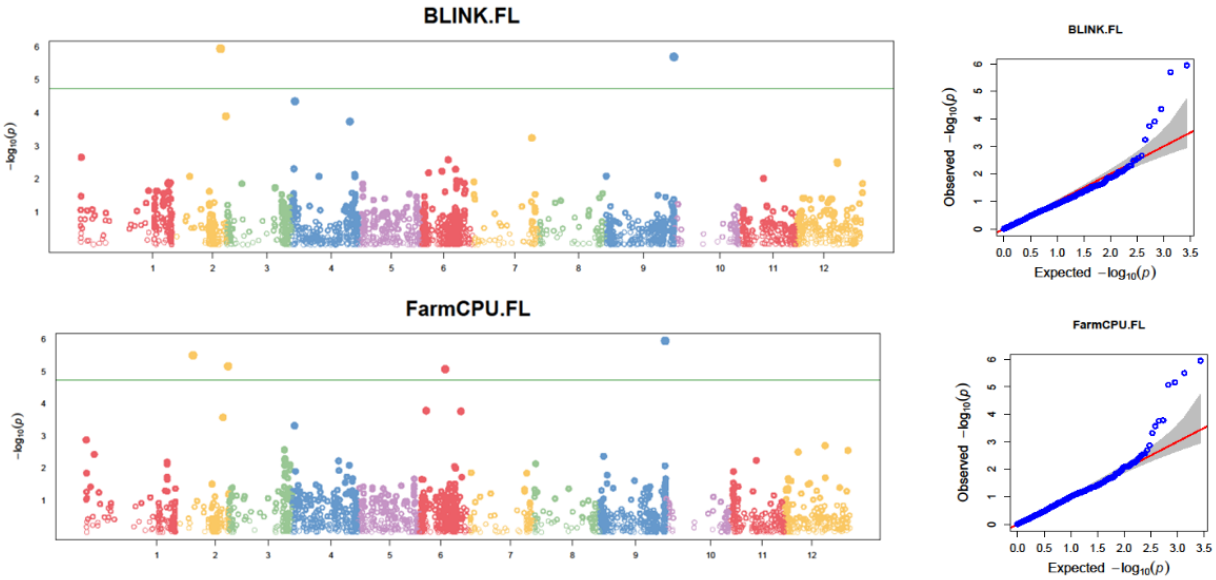

Figure 5.4: a) The Manhattan plot and b) the Q-Q plot for fruit length using BLINK and Farm CPU models

5.3.5.2 Fruit shape index

Eight significant ($p<8.9 \times 10^{-5}$, FDR ≤ 0.05) marker-trait associations, or QTLs, were detected on chromosomes 1, 2, 4, 6, 7, 9, and 11 (Table 5.6). The proportion of phenotypic variance explained ranged from 0.5% to 6.7%. The most significant marker-trait association (100058987, $p<1.1 \times 10^{-12}$) was identified on chromosome 6, at position 43,184,064 bp. It explained 4.1% of the phenotypic variance. Three of the multi-locus models were able to detect this locus. Locus 7850579 was identified on chromosome 4 at position 62,541,098 bp with a p-value of 1.0×10^{-7}. It explained the highest phenotypic variance (6.7%). The next highest phenotypic variance (3%) explained was due to locus 7834983, detected on chromosome 11. The remaining loci (7834983, 78834983, 7986772, 100087688, 4697721, and 4698369) explained phenotypic variances of 1.5%, 1%, 0.8%, 0.6%, and 0.5%, respectively. Four of the detected loci had a negative influence, while the remaining four loci had a favorable effect on the expression of the fruit shape index in tomato.

Table 5.6: Significant marker trait associations for fruit shape index

SNP	Allele	Ch.	Position	P-value	MAF	Effect	Model	PVE	FDR
7850579	C/A	4	62541098	1.0E-07	0.04	0.19	BLINK	6.7	0.00
100058987	T/G	6	43184064	1.1E-12	0.09	-0.19	BLINK +Farm CPU +MLMM	4.1	0.00
7835756	C/T	11	53574006	2.6E-05	0.07	0.19	MLMM	3	0.04
7834983	C/T	2	49548307	1.6E-10	0.17	-0.14	BLINK +Farm CPU	1.5	0.00
7986772	C/T	1	8778193	1.3E-06	0.04	0.25	BLINK	1	0.00
100087688	A/T	7	1724842	3.4E-05	0.11	0.08	BLINK	0.8	0.02
4697721	C/A	9	68951923	8.9E-05	0.14	-0.07	BLINK	0.6	0.03
4698369	T/C	2	18268340	1.2E-09	0.04	-0.29	BLINK+ MLMM	0.5	0.00

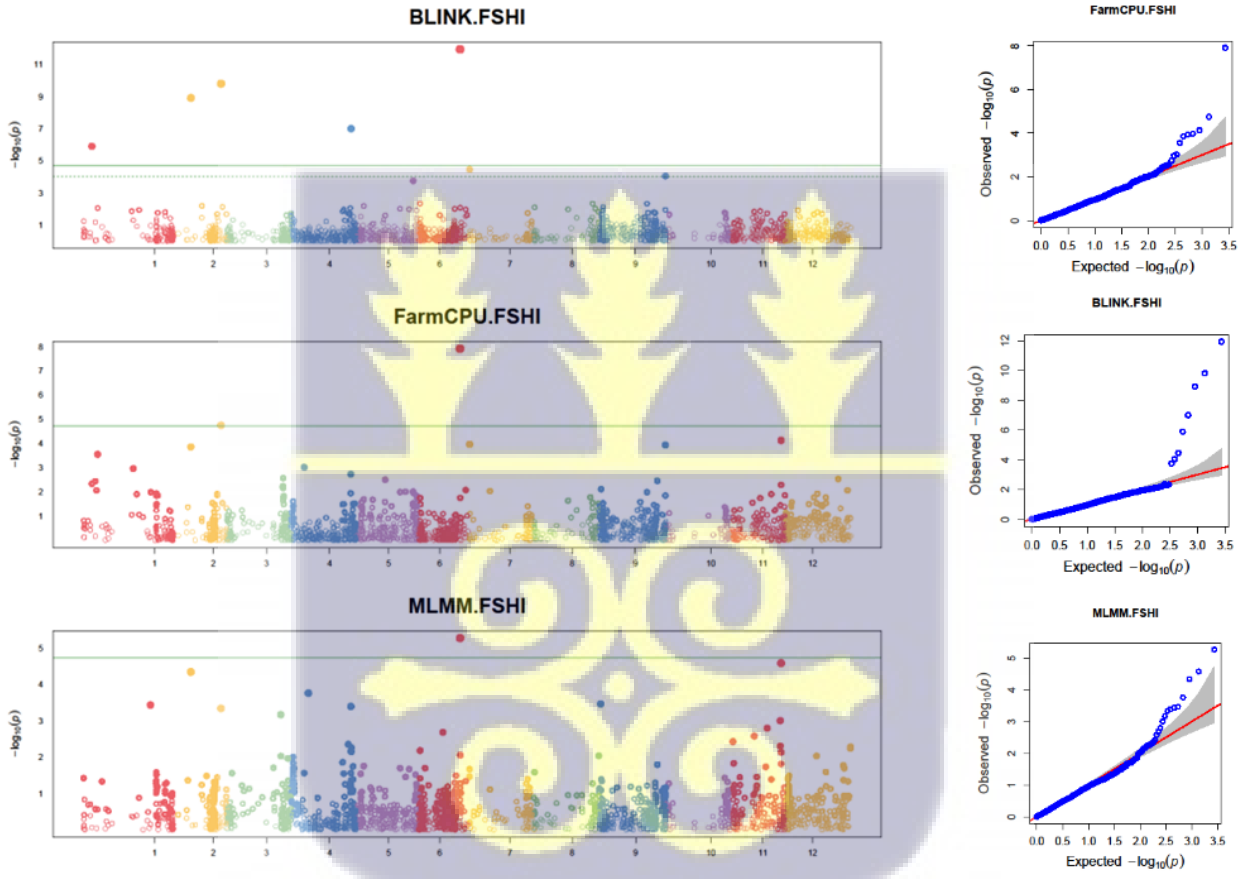

Figure 5.5: a) Manhattan plot and b) Q-Q plot for fruit shape index using BLINK, FarmCPU and MLMM.

5.3.5.3 Fruit width

Seven significant marker-trait associations or QTLs ($p < 1.4 \times 10^{-5}$, FDR ≤ 0.01) were identified on chromosomes 3, 4, 5, 8, 9, and 12, with the phenotypic variance explained ranging from 0.4% to 11.7% (Table 5.7). The locus 7844021 on Ch. 8 at position 63,558,538 bp identified by the BLINK and FarmCPU models was the most significant ($p < 5.0 \times 10^{-15}$), and explained the highest phenotypic variance (11.7%). The next most significant MTA identified by the same two models was locus 100060851, located on ch.08, which explained 9.1% of the phenotypic variance. Two loci (10006111667 and 15980448) explained the next-highest phenotypic variance of 7%. One of these loci (10006111667) was the second most significant MTA identified on chromosome 9 at position 1,839,962 bp. The remaining three loci (7834643, 100102183, and 7941886) explained 4.1%, 5.9%, and 0.4% of the phenotypic variance. Four of the identified loci showed a negative effect on fruit width, while the remaining three had a positive effect.

Table 5.7: Significant marker trait associations for fruit width

SNP	Allele	Ch.	Position	P-value	MAF	Effect	Model	PVE	FDR
7834643	C/A	3	56569746	9.5E-06	0.04	3.8	BLINK	4.1	0.00
100102183	C/T	4	62046988	2.1E-08	0.04	4.8	BLINK +Farm CPU	5.9	0.00
7941886	G/A	5	2192871	6.6E-06	0.39	1.5	BLINK	0.4	0.00
7844021	C/T	8	63558538	5.0E-15	0.04	-7.4	BLINK +FarmCPU	11.7	0.00
100060851	T/C	8	35491023	4.5E-10	0.05	-4.8	BLINK +Farm CPU	9.1	0.00
10006111667	C/T	9	1839962	3.3E-13	0.06	-5.5	BLINK	7.0	0.00
15980448	A/C	12	38160310	1.4E-05	0.03	-3.5	BLINK	7.0	0.01

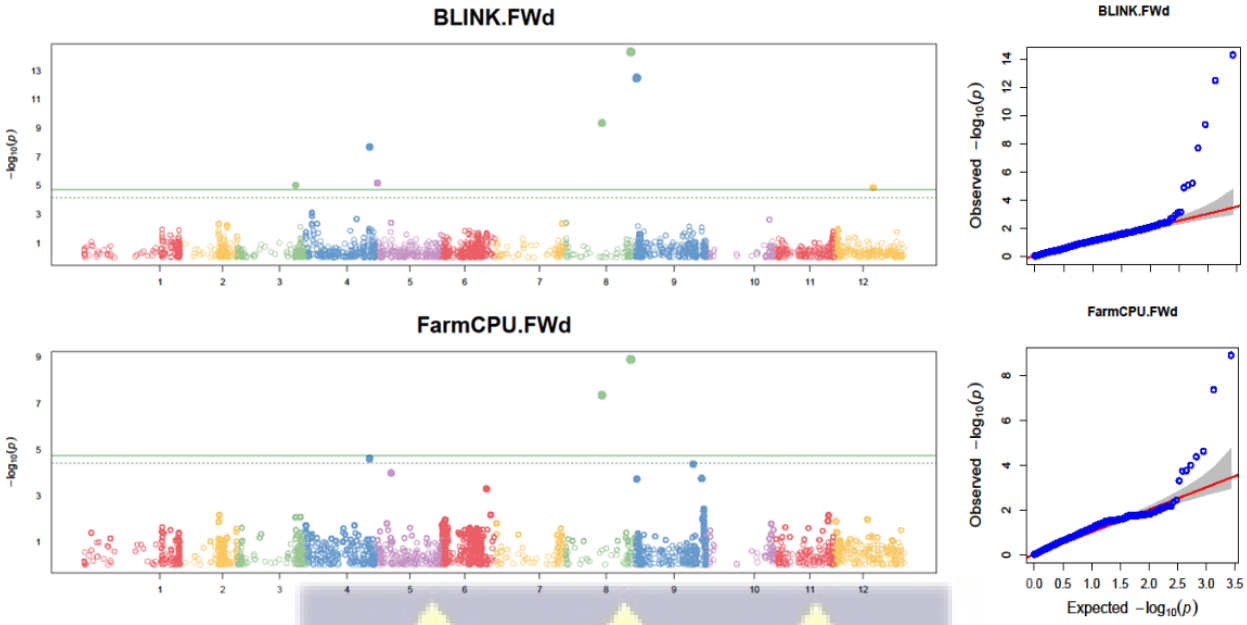

Figure 5.6: a) Manhattan plot and b) Q-Q plot for fruit width using BLINK and FarmCPU models

5.3.5.4 Fruit weight

Three marker-trait associations, or QTLs, with high statistical significance ($p<1.4 \times 10^{-5}$, FDR ≤ 0.05), were identified on chromosomes 2, 4, and 6 (Table 5.8). The percent phenotypic variance explained ranged from 0% to 2%. One of the associations, located on ch. 4 at position 56,782,433 bp and named locus 7761773, had the highest impact on fruit weight, explaining 2% of the phenotypic variance. However, it showed an effect of -6.8 on fruit weight expression. The remaining had an effect of 15.6 and 10.1 on fruit weight expression.

Table 5.8: Significant marker trait associations for fruit weight

SNP	Allele	Ch.	Position	P-value	MAF	Effect	Model	PVE	FDR
8019591	T/A	2	55409913	3.2E-07	0.07	15.6	FarmCPU	0	0.00
7761773	G/A	4	56782433	6.0E-06	0.23	-6.8	FarmCPU	2.0	0.02
7760727	A/T	6	47390620	1.4E-05	0.13	10.1	BLINK	0.5	0.04

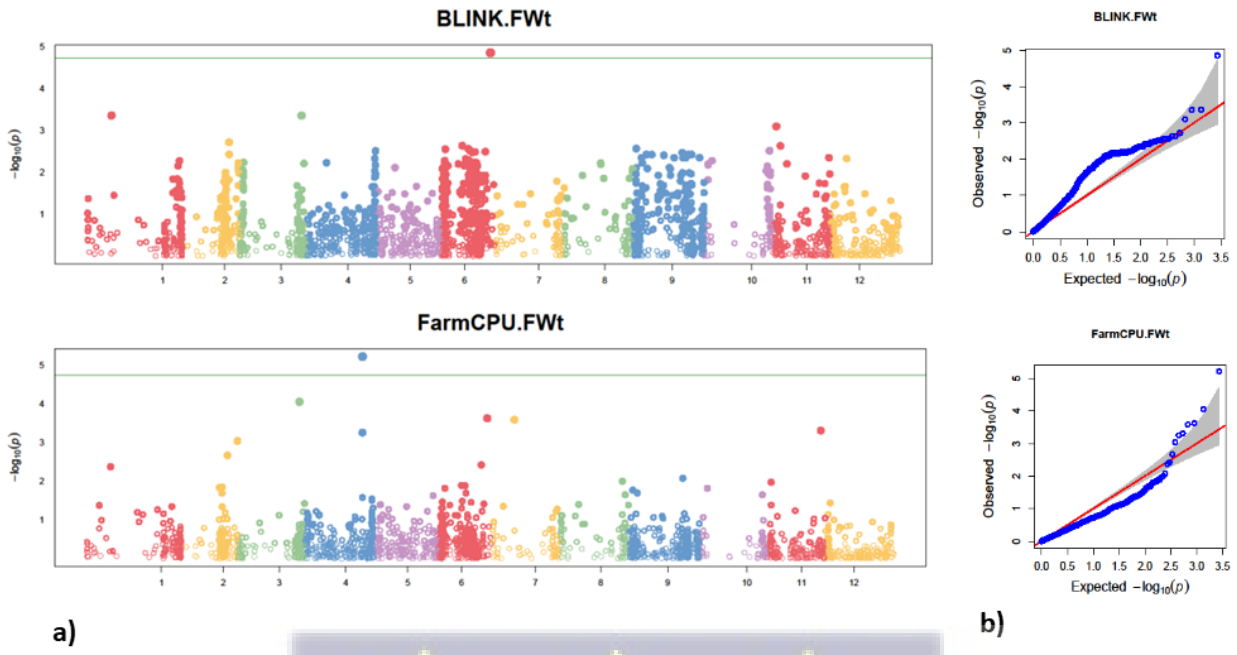

Figure 5.7: a) Manhattan plot and b) Q-Q plot for fruit weight using BLINK and FarmCPU models

5.3.5.5 Marketable number per plant

A total of twenty-five significant marker-trait associations (MTAs) were detected across chromosomes 01, 02, 03, 04, 06, 07, 08, 09, 10, and 11 (Table 5.9). The MTAs were grouped into 19 QTLs based on chromosome-wise LD decay. Seven MTAs on chromosome 01 that spanned 91.733 Mbp to 92.128 Mbp were grouped into 1 QTL. These MTAs had a p-value less than 9.0×10^{-5} and an FDR of ≤ 0.05. The phenotypic variance explained by these associations varied from 0% to 19.3%. Locus 783374 on ch. 01 at position 84,019,411 bp had the highest phenotypic variance explained (19.3%). The second-highest phenotypic variance explained (18.5%) was by locus 7764934, identified on ch. 01 at position 91732918. This locus was also the most significant MTA, with a p-value less than 1.0×10^{-15}. Three multi-locus models identified two loci (4698475 and 7840400) that explained 12.6% and 7.6% of the phenotypic variance. Twelve loci showed a positive effect, while the remaining had a negative effect.

Table 5.9: Significant marker trait associations for marketable number per plant

SNP	Allele	Ch.	Position	P-value	MAF	Effect	Model	PVE	FDR
7833748	A/G	1	84019411	7.0E-12	0.04	30.45	FarmCPU	19.3	0.00
7764934	G/A	1	91732918	1.0E-15	0.04	-59.3	MLMM +BLINK	18.5	0.00
7861514	C/G	1	92051538	1.0E-09	0.05	-40.8	MLMM + BLINK	15.0	0.00
7834395	T/A	7	2518858	3.0E-27	0.04	-28.7	BLINK +MLMM	14.5	0.00
4700003	A/G	4	62261938	4.0E-08	0.05	13.47	FarmCPU + BLINK	12.7	0.00
4698475	T/C	7	26773776	8.0E-14	0.02	-30.7	BLINK+FarmCPU+MLMM	12.6	0.00
7756687	T/G	6	42339982	7.0E-12	0.03	-16.2	BLINK	8.3	0.00
7840400	C/T	3	51605018	2.0E-12	0.06	15.74	BLINK+FarmCPU+MLMM	7.6	0.00
7834873	C/A	10	3767938	2.0E-07	0.04	13.93	FarmCPU	5.4	0.00
7852998	T/C	4	56684075	5.0E-05	0.09	-7.45	FarmCPU	5.3	0.01
7841432	A/G	3	63867280	6.0E-06	0.08	-8.59	FarmCPU	5.2	0.00
4702429	C/T	12	62859968	4.0E-05	0.04	-7.56	BLINK +MLMM	5.0	0.01
7844021	C/T	8	63558538	7.0E-10	0.04	11.62	BLINK	4.4	0.00
7834970	G/T	2	55619445	1.0E-07	0.05	10.91	FarmCPU	4.0	0.00
7834355	G/A	11	4286608	3.0E-08	0.04	-13.8	MLMM	3.5	0.00
100060948	G/A	11	56075153	2.0E-15	0.13	-10.3	BLINK	2.7	0.00
7838831	T/C	1	91888138	6.0E-05	0.05	-10	FarmCPU	2.1	0.02
7850743	G/A	1	92127603	9.0E-05	0.06	13.71	BLINK	2.0	0.02
4704929	T/A	9	52625779	9.0E-06	0.17	7.53	FarmCPU	1.7	0.00
7855962	G/T	7	58172885	5.0E-06	0.02	16.84	FarmCPU	0.3	0.00
7840048	T/C	11	52699841	5.0E-06	0.08	7.53	BLINK	0.2	0.00
7839010	T/A	3	60900986	2.0E-04	0.13	-4.21	BLINK	0.2	0.03
4701825	C/T	1	92086196	9.0E-05	0.05	-23.1	BLINK	0.0	0.02
7851641	T/C	1	92092536	9.0E-05	0.05	23.14	BLINK	0.0	0.02
4703064	T/G	1	92121592	9.0E-05	0.05	23.14	BLINK	0.0	0.02

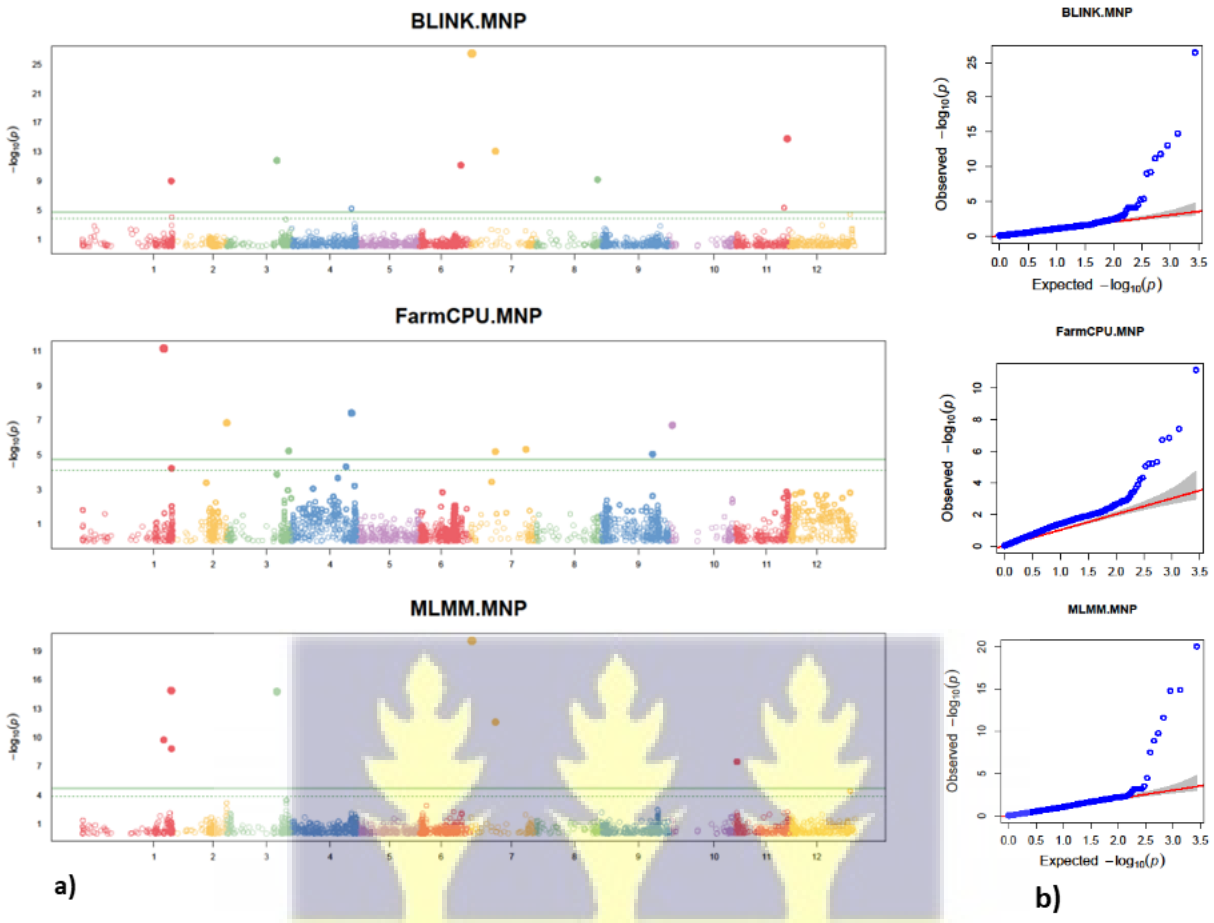

Figure 5.8: a) Manhattan plot and b) Q-Q plot for marketable number of fruits per plant using BLINK, FarmCPU and MLMM models

5.3.5.6 Number of locules

Four statistically significant MTAs or QTLs ($p< 5.9 \times 10^{-6}$, FDR ≤ 0.01) were identified on chromosomes 01, 10, and 1 (Table 5.10). The percentage of phenotypic variance explained by these loci ranged from 0.8% to 25.9%. The most significant MTA ($p<5.6x10^{-8}$) that detected on ch. 01 at position 79,849,107 bp accounted for the highest percentage of phenotypic variance (25.9%). The second most significant MTA was 8005855, explaining 4.2% of the phenotypic variance. The remaining two MTAs, 15980273 and 4700934, were associated with a phenotypic variance of 1.1% and 0.8%, respectively. All loci exhibited a negative effect on the number of locules per plant.

Table 5.10: Significant marker trait associations for number of locules

SNP	Allele	Chromosome	Position	P-value	MAF	Effect	Model	PVE	FDR
4694895	T/C	1	79849107	5.6E-08	0.03	-0.85	BLINK	25.9	0.00
8005885	C/A	10	60993603	5.5E-10	0.04	-0.96	BLINK	4.2	0.00
15980273	G/A	11	55919437	2.1E-06	0.10	-0.45	BLINK	1.1	0.00
4700934	A/G	11	39121508	5.9E-06	0.13	-0.41	BLINK	0.8	0.00

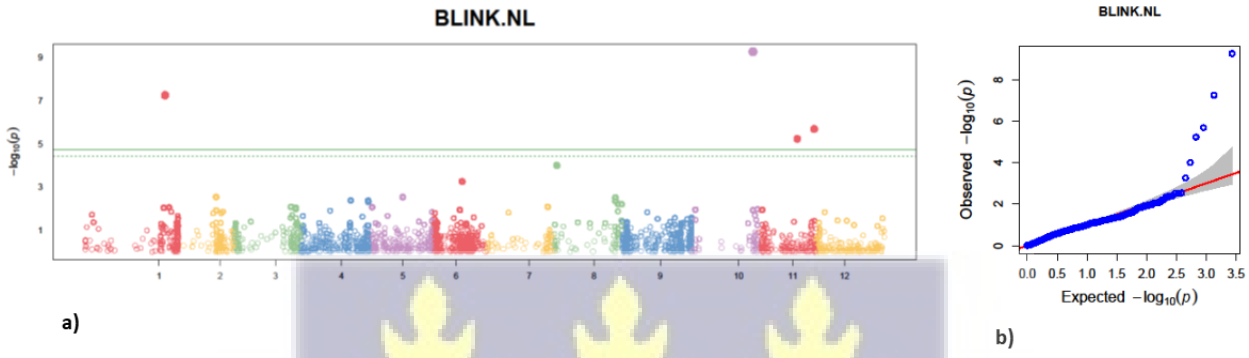

Figure 5.9: a) Manhattan plot and b) Q-Q plot for number of locules using BLINK models

5.3.5.7 Fruit skin thickness

A total of seven marker-trait associations, or QTLs, were identified for fruit skin thickness, with a significance level of $p<1.2x\ 10^{-4}$ and a false discovery rate (FDR) of ≤ 0.05 (Table 5.11). These associations were located on chromosomes 2, 3, 4, 6, 10, and 11, and their corresponding phenotypic variances ranged from 1.18% to 11.02%. Among these associations, locus 100062613 on ch. 2, at position 54,890,103 bp, explained the highest phenotypic variance of 11.02%. FarmCPU and MLMM multi-locus models identified the loci. The second-highest phenotypic variance (5.33%) was explained by loci 7841559 on ch. 10, at position 63,023,727 bp. This locus was the most significant one ($p<9.9x10^{-10}$). Loci 7763327 on ch. 6 and 7841841 on ch. 4 comparably explained equivalent phenotypic variances of 4.42% and 4.05%, respectively. The remaining loci (100086481, 8006152, and 15981408) explained 3.3%, 1.65%, and 1.18% of the

phenotypic variance. Notably, except for loci 8006152 and 7763327, which had a favorable effect on fruit skin thickness, the remaining loci showed a negative effect.

Table 5.11: Significant marker trait associations for fruit skin thickness

SNP	Allele	Ch.	Position	P-value	MAF	Effect	Model	PVE	FDR
100062613	T/C	2	54890103	9.5E-07	0.04	-0.78	BLINK +FarmCPU	11.02	0.00
7841559	G/T	10	63023727	9.9E-10	0.12	-0.47	FarmCPU + LMM	5.33	0.00
7763327	C/T	6	3173036	6.4E-05	0.12	0.52	FarmCPU	4.42	0.04
7841841	G/A	4	24376360	8.4E-05	0.11	-0.3	FarmCPU	4.05	0.04
100086481	A/G	3	59152008	1.7E-07	0.23	-0.24	FarmCPU + BLINK	3.3	0.00
8006152	G/A	11	4040148	1.2E-04	0.06	0.31	FarmCPU	1.65	0.05
15981408	C/T	4	64497030	7.9E-05	0.19	-0.2	FarmCPU	1.18	0.04

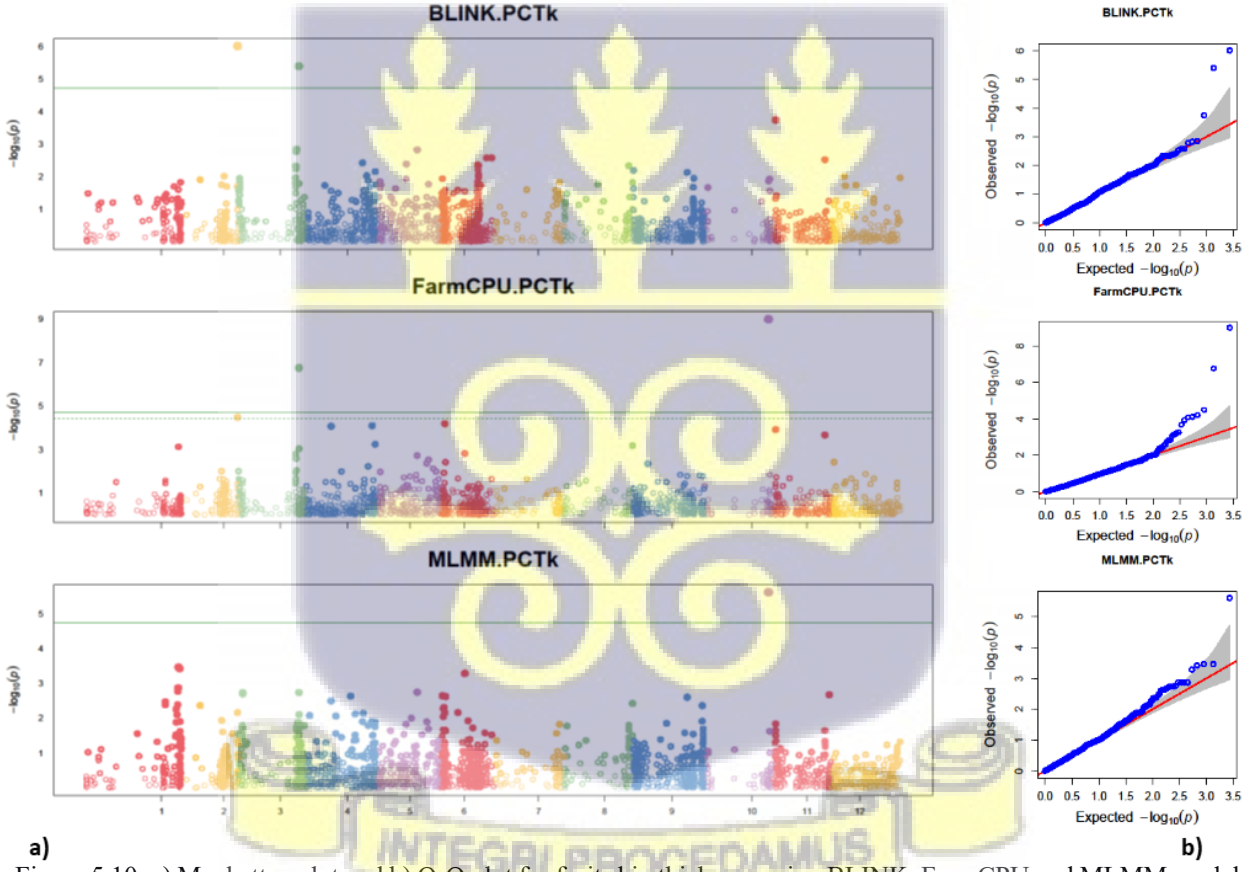

Figure 5.10: a) Manhattan plot and b) Q-Q plot for fruit skin thickness using BLINK, FarmCPU and MLMM models

5.3.5.8 Plant height

Twenty significant ($p < 2.1 \times 10^{-4}$ FDR ≤ 0.01) MTAs or 17 QTLs were identified on chromosomes 01, 02, 03, 04, 06, 08, 10, 11, and 12 by three multi-locus models (Table 5.12). A total of six MTAs on chromosomes 01, 04, and 06 were clustered into three QTLs based on LD decay on each chromosome. The phenotypic variance explained ranged from 0% to 32.5%. Loci 7762862 was the most significant MTA ($p < 3.3 \times 10^{-25}$) and explained the highest phenotypic variance (32.5%). It was located on Ch. 1 at position 1,355,364 bp. The next two most significant associations were 100088666 ($p < 3.1 \times 10^{-20}$) and 7840400 ($p < 3.7 \times 10^{-23}$), explaining a phenotypic variance of 24.1% and 22.1%, respectively. Two identified loci (4694949 and 4694933) had no notable phenotypic variance. The remaining 15 loci explained a phenotypic variance that ranged from 0.1% to 14.7%. Among the identified loci, 12 showed a favorable effect, while the remaining 8 had a negative effect.

5.3.5.9 Total fruit number per plant

A total of 32 marker-trait associations (MTAs) or 26 significant QTLs ($p < 3.0 \times 10^{-4}$, FDR ≤ 0.05) were identified on chromosomes 01, 02, 03, 04, 06, 07, 08, 09, 10, and 11, with the variance in the phenotype explained ranging from 0% to 25% (Table 5.13). Twelve MTAs on chromosomes 01, 03, 04, and 11 were clustered into six QTLs based on the LD decay determined for each chromosome. Locus 7756687 accounted for the highest phenotypic variance (25%) detected on ch. 6 at position 42,339,982 bp, with a p-value of less than 1×10^{-19}. This locus was detected by both the BLINK and FarmCPU models. The most significant MTA was found in the genomic region 78337448 ($p < 2x10^{-24}$) on ch. 01 and position 84,019,411 bp, accounting for 17% of the second-highest phenotypic variance. The three subsequent most significant marker-trait associations were 10008796 ($p < 2x10^{-23}$), 8019591 ($p < 9x10^{-22}$), and 7850579 ($p < 1x10^{-20}$),

explaining 14.5%, 10.7%, and 10% of the phenotypic variance, respectively. Nineteen of the MTAs had a negative effect, while the remaining had a positive effect. A combination of the two models detected only five genomic regions.

Table 5.12: Significant marker trait associations for plant height

SNP	Allele	Ch.	Position	P-value	MAF	Effect	Model	PVE	FDR
7762862	G/A	1	13553641	3.3E-25	0.03	30.4	BLINK +Farm CPU	32.5	0.00
100088666	A/T	6	48980638	3.1E-20	0.04	30.8	MLMM	24.1	0.00
7840400	C/T	3	51605018	3.7E-23	0.06	35.7	MLMM	22.1	0.00
7759348	G/A	1	50069111	3.4E-18	0.04	-19.7	BLINK + Farm CPU	14.7	0.00
7837211	A/T	6	44611654	1.0E-16	0.02	20.1	BLINK + Farm CPU	10.6	0.00
7850743	G/A	1	92127603	2.8E-11	0.06	-17.0	BLINK	6.1	0.00
7851120	A/G	11	52765945	1.8E-12	0.02	22.1	FarmCPU + BLINK +MLMM	5.7	0.00
100087517	T/A	6	33562947	2.1E-08	0.03	19.3	MLMM	5.6	0.00
7851381	G/A	1	347519	4.3E-06	0.04	-11.0	MLMM	5.5	0.00
4697003	T/G	12	2808715	2.1E-04	0.04	4.1	BLINK	4.1	0.04
100060025	G/A	2	40542182	1.1E-10	0.05	-16.8	MLMM	3.8	0.00
7845956	G/C	4	340451	2.7E-07	0.07	-7.5	BLINK	3.6	0.00
4696053	T/C	6	43328388	1.0E-12	0.07	9.7	BLINK +Farm CPU	3.3	0.00
8019365	A/T	12	19469964	1.2E-06	0.32	-2.6	BLINK	0.3	0.00
4703584	T/C	1	93020900	3.7E-06	0.10	-4.0	BLINK	0.3	0.00
100096808	G/A	8	58855538	1.9E-06	0.34	2.7	BLINK	0.2	0.00
100065233	C/T	10	6908226	1.4E-04	0.21	2.3	BLINK	0.2	0.03
4696893	C/T	6	27876078	5.7E-05	0.19	5.1	FarmCPU	0.1	0.03
4694949	G/A	4	327591	4.0E-12	0.05	13.7	FarmCPU	0.0	0.00
4694933	A/C	12	3471158	1.1E-04	0.06	-8.2	BLINK	0.0	0.03

5.3.5.10 Total soluble solid

Eight significant ($p<1.1\ x10^{-4}$, FDR ≤ 0.05) MTAs or 7 QTLs were identified on chromosomes 01, 06, 09, and 12, with phenotypic variance ranging from 1.3% to 11.7% (Table 5.14). Based on the LD decay, 2 MTAs clustered between 2.193 Mbp and 3.012 Mbp on chromosome 09 were considered 1 QTL. The genomic region 4704667, detected on ch.01 at position 837,897 bp, explained the highest phenotypic variance (11.7%), as identified by three models. The subsequent two loci, 100058756 (7.2%) and 8006159 (4.3%), detected on chromosomes 12 and 6,

respectively, explained the next two highest phenotypic variances. The remaining MTAs (7758160, 7856928, 7855688, 4703044, and 7839485) explained 3%, 2.7%, 2.6%, 2.3%, and 1.3% of the phenotypic variance, respectively. Five loci had a positive effect on TSS, while the remaining had a negative effect.

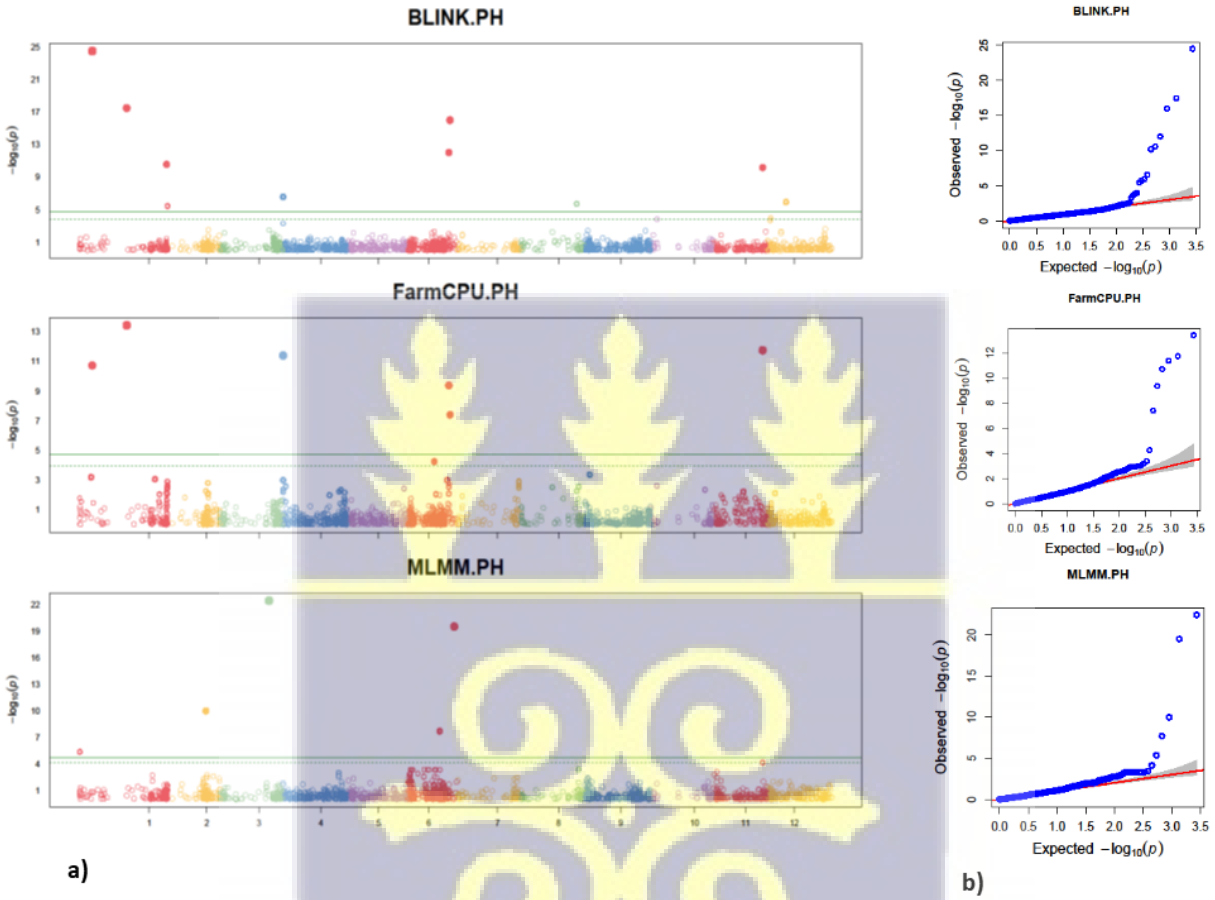

Figure 5.11: a) Manhattan plot and b) Q-Q plot for plant height using BLINK, FarmCPU and MLMM models

Table 5.13: Significant marker trait association for total number of fruits per plant

SNP	Allele	Ch.	Position	P-value	MAF	Effect	Model	PVE	FDR
7756687	T/G	6	42339982	1E-19	0.03	-39.8	BLINK +Farm CPU	25	0.00
7833748	A/G	1	84019411	2E-24	0.04	43.2	MLMM	17	0.00
100087096	T/C	8	60826673	2E-23	0.04	36.6	BLINK +Farm CPU	14.5	0.00
7834395	T/A	7	2518858	2E-15	0.04	-22.4	BLINK	13	0.00
7759794	T/G	1	91892058	4E-13	0.05	-48	MLMM	11.5	0.00
8019591	T/A	2	55409913	9E-22	0.07	-29.5	MLMM +BLINK	10.7	0.00
7850579	C/A	4	62541098	1E-20	0.04	26	BLINK +MLMM	10	0.00
4700003	A/G	4	62261938	2E-15	0.05	22	FarmCPU	9.7	0.00
7834448	T/C	9	60539897	1E-14	0.04	-21	BLINK	9	0.00
100097203	G/A	6	23820719	3E-07	0.02	-20.3	MLMM	8.4	0.00
4696531	A/T	1	92760509	4E-10	0.02	-36	MLMM	8.2	0.00
7836779	G/A	3	56605649	2E-11	0.07	-20.2	FarmCPU	8	0.00
4698475	T/C	7	26773776	6E-09	0.02	-37.5	MLMM	6.6	0.00
7834873	C/A	10	3767938	3E-06	0.04	15.8	FarmCPU	6.3	0.00
7858444	T/C	3	63004234	4E-11	0.03	21.6	FarmCPU	5.2	0.00
7850743	G/A	1	92127603	2E-09	0.06	20.8	BLINK +FarmCPU	3.7	0.00
7987501	A/T	3	57773468	2E-04	0.02	15.3	MLMM	3.4	0.04
7851641	T/C	1	92092536	3E-06	0.05	30.8	MLMM	3.3	0.00
7850671	G/A	9	2806753	5E-09	0.08	-12.8	FarmCPU	3.2	0.00
4700875	T/C	3	65860991	3E-10	0.09	-16.2	MLMM	2.8	0.00
4697624	C/T	1	87772818	5E-06	0.03	-14.8	BLINK	1.7	0.00
100062554	T/G	11	56291441	3E-04	0.04	15	MLMM	1.3	0.05
4702591	G/T	3	57456080	4E-08	0.13	-13.1	FarmCPU	1.2	0.00
7850872	T/C	9	8766479	1E-08	0.14	-9.2	BLINK	1.1	0.00
100060948	G/A	11	56075153	8E-06	0.13	-6	FarmCPU	0.7	0.00
7834269	G/A	7	65956443	2E-04	0.04	9.5	FarmCPU	0.2	0.05
7838175	T/C	3	58479400	8E-05	0.03	15.3	MLMM	0.1	0.02
7840048	T/C	11	52699841	2E-04	0.08	-8.1	MLMM	0.1	0.03
7841523	G/C	3	57967710	5E-05	0.03	17.2	MLMM	0	0.01
4704697	A/T	3	58941531	5E-05	0.03	-17.2	MLMM	0	0.01
7758311	C/T	3	59045750	5E-05	0.03	-17.2	MLMM	0	0.01
100059661	A/G	3	56721358	2E-04	0.04	-14.3	MLMM	0	0.03

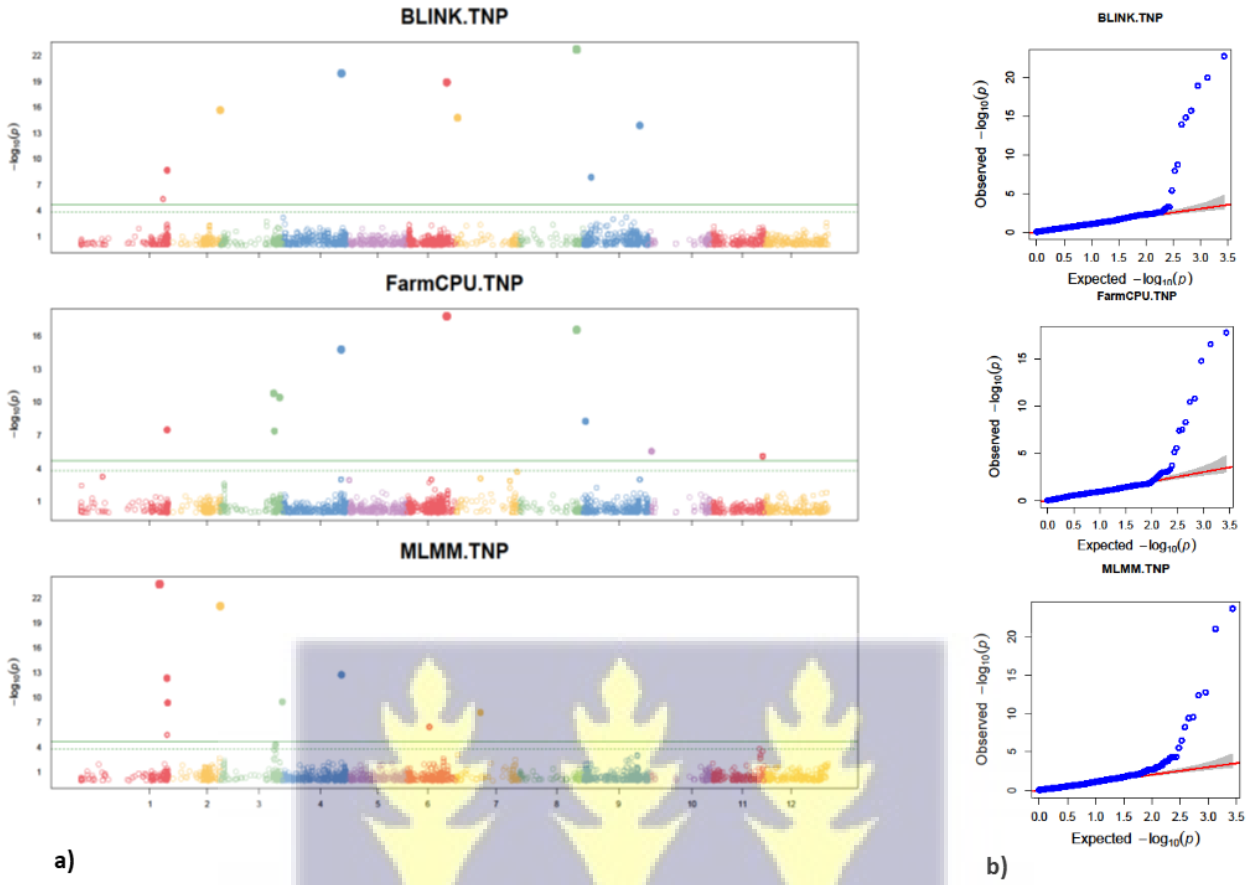

Figure 5.12: a) Manhattan plot and b) Q-Q plot for total number of fruits per plant using BLINK, FarmCPU and MLMM models

Table 5.14: Significant marker trait associations for total soluble solids

SNP	Allele	CHR	Position	P-value	MAF	Effect	Model	PVE	FDR
4704667	C/T	1	837897	4.0E-06	0.02	0.18	BLINK +Farm CPU +MLMM	11.7	0.01
100058756	G/T	12	25009549	3.4E-05	0.03	-0.12	BLINK, Farm CPU, MLMM	7.2	0.02
8006159	T/G	6	18101028	3.2E-05	0.02	0.15	FarmCPU	4.3	0.02
7758160	G/A	9	58743355	6.2E-06	0.17	0.1	BLINK + Farm CPU	3	0.01
7856928	A/G	9	2193014	1.6E-05	0.04	0.12	FarmCPU	2.7	0.02
7855688	A/G	6	22708228	7.4E-05	0.11	-0.09	FarmCPU	2.6	0.03
4703044	G/A	1	91074884	2.7E-05	0.08	-0.08	BLINK	2.3	0.02
7839485	G/A	9	3012303	1.1E-04	0.06	0.08	FarmCPU	1.3	0.04

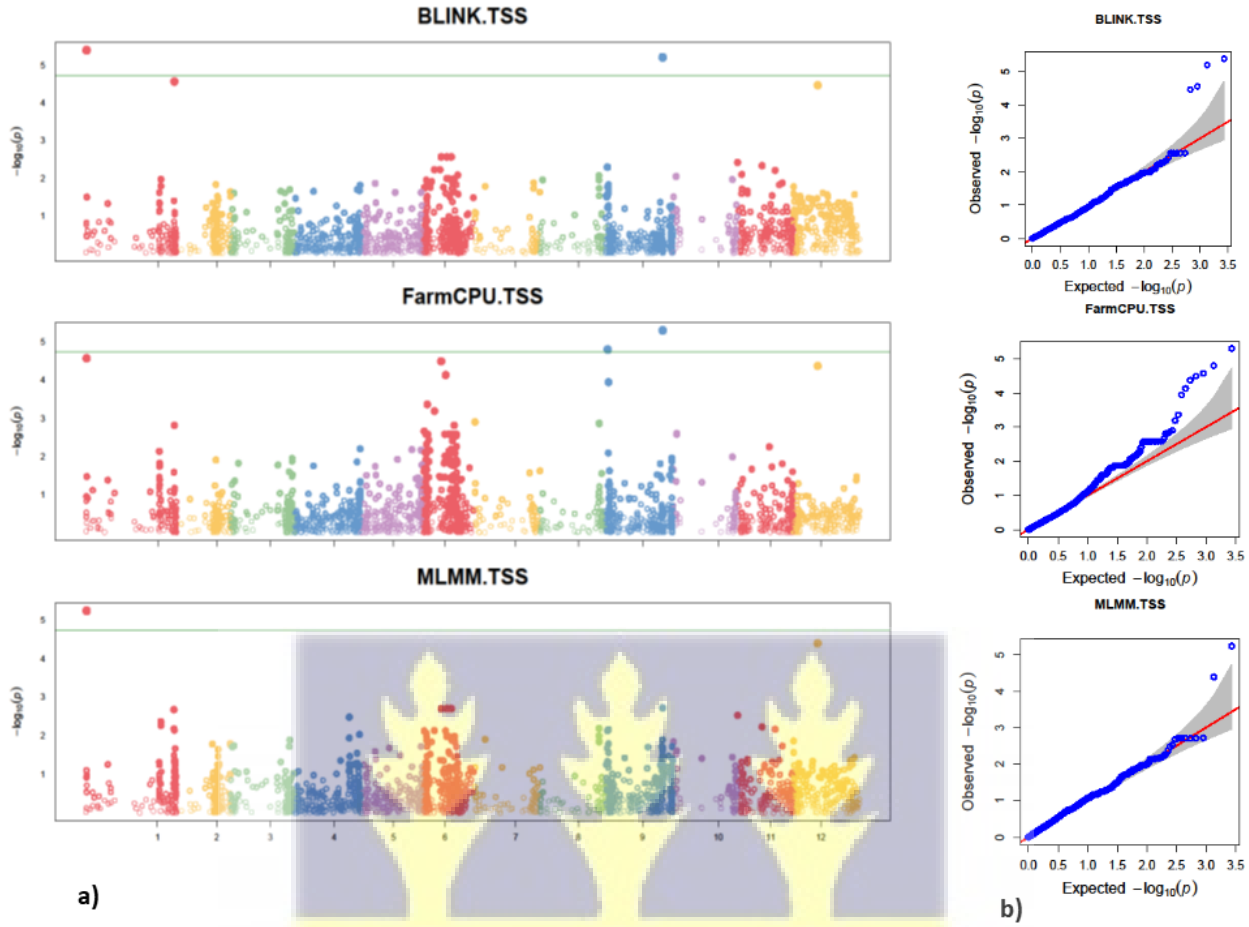

Figure 5.13: a) Manhattan plot and b) Q-Q plot for TSS using BLINK, FarmCPU and MLMM models

5.4 Discussion

Effective crop improvement programs require the integration of phenotype-based breeding with molecular approaches to efficiently exploit available genetic diversity. Most traits are polygenic, controlled by numerous loci with small and cumulative effects, and are influenced by environmental and genotype interactions, which make phenotype-based selection challenging (Ikram & Chardon, 2010). The application of molecular markers in plant improvement programs has been facilitated by the development of a wide range of molecular tools and resources over the years (Soriano, 2020). The discovery and use of molecular markers in marker-assisted breeding

have greatly enhanced the efficiency of plant breeding by improving time, cost-effectiveness, precision, and genetic gain (Foolad & Panthee, 2012; Ibitoye & Akin-Idowu, 2011).

In this study, 2,709 SNP markers and 143 tomato germplasms were used to identify genomic regions associated with yield, yield-related traits, and fruit quality. Phenotypic analysis revealed highly significant genetic variation across 14 quantitative traits. Genotype-by-season interactions had a significant effect on the expression of traits such as LN, MYP, FWt, and FSHI. Broad-sense heritability values ranged from 29% to 92%, indicating that the observed phenotypic variation was largely attributable to genetic factors. Moreover, highly significant genetic correlations were observed among most of the 14 traits. Molecular characterization further demonstrated considerable genetic diversity among the germplasm (Chapter 4), suggesting strong potential for identifying molecular markers linked to these traits

One approach to determining the optimal marker density for genetic analysis is to evaluate linkage disequilibrium (LD) decay based on physical genetic distance. The extent of LD decay varies across species, chromosomes, populations, and estimation methodologies (Sim *et al*., 2012). Using the entire genome to calculate average LD decay, Robbins *et al*. (2011) reported 6–8 cM across all varieties, 6–14 cM for processing tomatoes, and 3–16 cM for fresh-market tomatoes. In the present study, with the available marker density across the 12 chromosomes, an overall LD decay of 2.052 Mbp at $r^2 = 0.2$ was observed, with variations among chromosomes. Chromosome 07 had an LD decay of 0.157 Mbp, whereas chromosomes 05 and 09 exhibited 14.43 Mbp and 28.83 Mbp, respectively. Similar studies reported low LD decay values for chromosomes 05 (47.2 Mbp) and 09 (42.5 Mbp) (Liu *et al*., 2017). Consistent with the current findings, Ruggieri *et al*. (2014) observed an LD of 665 kbp for the entire genome, ranging from 287 kbp on chromosome 02 to

1965 kbp on chromosome 11. Another study reported LD decay from 0.8 cM on chromosome 11 to 35.2 cM on chromosome 6 for processing tomatoes, and from 3.8 cM on chromosome 04 to 47.6 cM on chromosome 11 for fresh-market tomatoes (Sim *et al*., 2012). Esposito *et al*. (2020) found that LD decay in 288 germplasm accessions from different countries, using 32,799 SNP markers, was less than 5 kb.

Genome-wide association studies (GWAS) are prone to false positives due to population structure and family relatedness. To address this issue, several statistical approaches have been developed. Among them, the Mixed Linear Model (MLM) is widely regarded as the most effective, as it accounts for both population structure and family relatedness. However, MLM has limitations in detecting genomic regions, which can lead to false negatives due to the confounding effects of population structure, relatedness, and QTLs (Xiaolei *et al*., 2016). The Multi-Locus Mixed Model (MLMM) partially controls for the confounding effects of kinship and marker testing. In contrast, FarmCPU completely removes confounding effects while controlling for model overfitting and false positives. An improved version of FarmCPU, BLINK, eliminates the requirement for evenly distributed markers and instead utilizes information from linkage disequilibrium (Huang *et al*., 2019).

In population genetics, principal components (PCs) are commonly used to represent population structure. These PCs can be derived from genetic markers across the genome (McVean, 2009) or from the Q matrix, which indicates the proportion of individuals in subpopulations. To identify marker–trait associations (MTAs), the present study employed three multilocus models MLMM, FarmCPU, and BLINK using three kinship matrices and the first three PCs, which accounted for the largest proportion of cumulative variance (30%). Using these three multilocus models, a total

of 121 significant DArTseq SNP markers (FDR ≤ 0.05), corresponding to 104 QTLs associated with ten traits across two seasons, were identified.

Fruit length has been reported to be highly correlated with fruit yield in tomatoes and exerts a strong influence on overall yield (Hernández-Bautista *et al*., 2015). Consistent with these reports, our findings showed a positive and highly significant genetic correlation between fruit length and fruit yield, suggesting that QTLs identified for fruit length could be utilized to improve yield in tomato breeding programs. In this study, a total of seven marker–trait associations (FDR ≤ 0.05) were detected, explaining 1.5% to 7% of the phenotypic variance on chromosomes 02, 04, 06, and 09. Previous research has identified three significant genomic loci linked to fruit length on chromosomes 04 and 08, accounting for 0 to 42.6% of the phenotypic variance (Xiaolei *et al*., 2016). Notably, one of the MTAs (7841051) identified on chromosome 04 at position 3.23 Mb in our study was located 0.25 Mb and 2.02 Mb away from previously reported markers SLA7841051 and SLA798303, respectively (Kim *et al*., 2021).

Fruit size is a major determinant of yield in tomato, and a significant co-location of QTLs between yield and fruit size traits has been reported, suggesting that these QTLs may operate within the same expression pathway controlling yield (Hernández-Bautista *et al*., 2015). Findings of the current study revealed highly significant genetic variation and a positive, highly significant genetic association between fruit weight and yield. In most cases, high-yielding genotypes exhibited larger fruit sizes. The genome-wide association study detected three significant marker–trait associations (FDR ≤ 0.05) for fruit weight on chromosomes 2, 4, and 6, explaining 0% to 2% of the phenotypic variance. Previously, seven SNP loci significantly associated with fruit weight were identified on chromosomes 1, 2, 4, 8, and 10, accounting for 2.91% to 26.2% of the phenotypic variance (Kim

et al., 2021). Mata-Nicolás *et al*. (2020) identified four SNP loci on chromosomes 2, 7, 9, and 12 using 163 diverse tomato accessions, including cultivated tomato, *Solanum lycopersicum* var. *cerasiforme*, and *Solanum pimpinellifolium*. One SNP locus (Solyc02g091330, position 52.7 Mb) on chromosome 2 in our study was located approximately 2.7 Mb from SNP 7853862 at position 55.4 Mb. Similarly, a study involving 174 genotypes (123 cherry and 51 large-fruited tomatoes) comprising landraces, cultivars, and breeding lines identified 17 MTAs for fruit weight on chromosomes 1, 2, 3, 5, 6, 8, 9, and 11, explaining 2.94% to 16.05% of the phenotypic variance. One MTA (SSR306, position 56 Mb) on chromosome 4 in that study was located 0.78 Mb from SNP Sl7761773 in our findings (Zhang *et al*., 2016). Another study using 212 cultivated tomato germplasms and 24,428,210 SNPs identified a locus on chromosome 2, which was 7.7 Mb from locus 8019591 in our study (Liu *et al*., 2023).

Fruit width, like fruit length, influences fruit size in tomatoes and is a key determinant of yield variation. In this study, fruit diameter and yield exhibited highly significant genetic variation and a strong positive association. Fruit diameter was also found to contribute to fruit size and weight, showing a highly significant positive correlation with fruit weight. A total of seven significant marker–trait associations (FDR ≤ 0.01) for fruit width were detected on chromosomes 03, 04, 05, 08, 09, and 12, explaining 0.4% to 11.7% of the phenotypic variance. Previous studies on interspecific populations derived from cultivated tomato and *Solanum pimpinellifolium* using 31 polymorphic SSR markers identified nine fruit width QTLs on chromosomes 1, 2, 3, 4, and 7, explaining 4.19% to 12.09% of the phenotypic variance (Hernández-Bautista *et al*., 2015). Differences between studies may be due to variations in genetic background, marker type and density, and the phenotyping environment. For instance, one QTL on chromosome 03 (SSR111) is located 3.57 Mbp from locus 7834643 identified in the current study, and a QTL on chromosome

04 (SSR306) is approximately 3.5 Mbp from the genomic region Sl100102183 detected here. Similarly, Kim *et al*. (2021) reported seven genomic regions associated with fruit diameter on chromosomes 1, 2, 3, 9, and 12, accounting for 0.24% to 22.94% of phenotypic variance. Two regions in the current study 7834643 (56.6 Mb) and 10006111667 (1.84 Mb) on chromosomes 03 and 09 are 1.56 Mb and 1.68 Mb, respectively, from SLA805140 (55.04 Mb) on chromosome 03 and 807083 (0.16 Mb) on chromosome 09. An additional study using 174 tomato genotypes (123 cherry and 51 large-fruited), including landraces, cultivars, and breeding lines, identified 22 MTAs on chromosomes 1, 2, 3, 5, 6, 7, 8, 9, and 11, explaining 0.96% to 10.6% of phenotypic variance. Four of these regions on chromosomes 04, 08, and 09 correspond closely to loci identified in the current study (Zhang *et al*., 2016). Specifically, TGS292 on chromosome 04 is 3.38 Mb from locus 100102183, TES718 on chromosome 08 is 2.4 Mb from locus 100060851, and TES36 and TES166 on chromosome 09 are 1.11 Mb and 1.26 Mb, respectively, from locus 10006111667 detected in the present study.

Eight significant marker–trait associations (FDR ≤ 0.05) for fruit shape index were detected on chromosomes 01, 02, 04, 06, 07, 09, and 11, explaining 0.5% to 6.7% of the phenotypic variance. In a previous study, four genomic regions significantly associated with fruit shape index were identified on chromosomes 02, 03, 04, and 12, accounting for 3.19% to 31.46% of the phenotypic variance (Kim *et al*., 2021). Notably, one genomic region identified on chromosome 04 in the current study is located only 2.36 Mb away from the corresponding region reported by Kim *et al*. (2021).

Total soluble solids (TSS) is a quantitative trait that exhibits genetic variability in tomato germplasm. In this study, eight highly significant MTAs ($p < 1.1 \times 10^{-4}$, FDR ≤ 0.05) were

identified on chromosomes 01, 06, 09, and 12, explaining 1.3% to 11.7% of the phenotypic variance. Similarly, Kim *et al*. (2021) identified a genomic region (SLA8126289, position 62.64 Mb) on chromosome 09 associated with TSS, accounting for 28.73% of the phenotypic variance, located 3.64 Mb from the region identified in the current study. Seven additional MTAs were detected on chromosomes 02, 05, 06, 07, 08, and 09. A genomic region on chromosome 06 at position 19.1 Mb was closely located to locus 7855688 (position 22.7 Mb) in the current study. Likewise, the locus TGS1032 (position 30.1 Mb) on chromosome 09 identified by Zhang *et al*. (2016) was closely linked to locus 7839485 (position 30.12 Mb) in findings of the current study. An additional study using 212 cultivated tomato accessions identified four SNP loci associated with TSS on chromosomes 01, 09, and 11, explaining 19% to 24% of the phenotypic variance (Liu *et al*., 2023). One SNP locus on chromosome 09 at 4.17 Mb was located approximately 2 Mb from SNP locus 7856928 (position 2.2 Mb) in the current study and 1.19 Mb from locus 7839485. Another locus on chromosome 01 at position 94.2 Mb was situated 3 Mb from locus 4703044 identified in the current findings.

Genetic variation exists in locule number among cultivated tomatoes, ranging from 2 to 10. This trait is important for seed content in fruits and is quantitatively inherited, influencing both tomato fruit shape and size. A large portion of the phenotypic variation in locule number is controlled by two genomic regions, faciated (fas) and locule number (lc), which exhibit epistatic interactions. The faciated QTLs have been cloned, and genes within these regions contribute to increased fruit size in modern tomato cultivars (Muños *et al*., 2011).

In the present study, four statistically significant MTAs ($p < 5.9 \times 10^{-6}$, FDR ≤ 0.01) were identified on chromosomes 01, 10, and 11, explaining 0.8% to 25.9% of the phenotypic variance.

Kim *et al*. (2021) reported genomic regions associated with locule number on chromosomes 02, 03, 06, and 10, explaining 0% to 15.38% of the phenotypic variance. One of the genomic regions in the current study (8005885) on chromosome 10 is located only 1.7 Mb from the region SLA795305, suggesting colocalization. Mata-Nicolás *et al*. (2020) reported four genomic regions associated with locule number on chromosomes 01, 02, and 11. Two of their regions, solyc11g071600 and solyc11g071310 on chromosome 11, are located 0.7 Mb and 1.06 Mb, respectively, from locus 55919437 in the current study. Similarly, Hernández-Bautista *et al*. (2015) identified three significantly associated genomic regions on chromosomes 01 and 03, explaining 3.26% to 4.69% of the phenotypic variance. In another study using 123 tomato germplasm panels and 7,720 SNPs, two loci on chromosomes 10 and 11 colocalized with regions in findings of the current study. For example, locus 8005885 on chromosome 10 is 2.3 Mbp from Solcap-1981, while locus 15980273 on chromosome 11 is 0.84 Mbp from Solcap504 (Sacco *et al*., 2015).

Seven significant marker–trait associations (FDR ≤ 0.05) for PCTk were identified on chromosomes 02, 03, 04, 06, 10, and 11, explaining 1.1% to 11.02% of the phenotypic variance. Previously, four genomic regions on chromosomes 02, 09, and 12 were reported, accounting for 0.38% to 2.92% of the phenotypic variance (Kim *et al*., 2021). Additionally, twelve genomic regions significantly associated with PCTk were identified on chromosomes 01, 05, 06, 07, 08, 09, and 11, explaining 10.53% to 14.62% of the phenotypic variance. Notably, one genomic region on chromosome 03 (SJS1419, position 55.83 Mb) colocalized with a region in the current study (SNP locus 7836779, position 53,934,779 bp), located 3.32 Mb apart (Phan *et al*., 2019b). Furthermore, one genomic region (Solcap2392) identified on chromosome 11 by Sacco *et al*. (2015) is situated 0.43 Mb from locus SL8006152 detected in the present study.

Twenty highly significant marker–trait associations (FDR ≤ 0.01) for plant height were identified on chromosomes 01, 02, 03, 04, 06, 08, 10, 11, and 12, explaining 0% to 32.5% of the phenotypic variance. Notably, one MTA on chromosome 06 accounted for 24% of the phenotypic variance in a study using 212 cultivated tomatoes and 24,428,210 SNPs (Liu *et al*., 2023). Two loci identified in the current study (100088666 and 7837211) are located 2.6 Mb and 1.8 Mb, respectively, from the corresponding loci reported by Liu *et al*. (2023).

5.5 Conclusion

A total of 121 marker–trait associations (MTAs; $-\log_{10}p \geq 3.5$; FDR ≤ 0.05) were identified for ten traits, which were clustered into 104 quantitative trait loci (QTLs) based on linkage disequilibrium (LD) decay (0.157–28.826 Mbp) at $r^2 = 0.2$ for each chromosome. Several QTLs for fruit length (FL), pericarp thickness (PCTk), fruit shape index (FSHI), fruit weight (FWt), fruit width (FWd), locule number (NL), and total soluble solids (TSS) were located near previously reported MTAs, whereas the remaining loci represent novel genomic regions. Across all traits, 16 common MTAs were identified, with maximum MTAs observed for MNP and TNP (9) and the minimum for PH (1). No significant MTAs were detected for marketable yield per plant (MYP), total yield per plant (TYP), days to first flowering (DFF), or number of fruits per cluster (NFtC). Importantly, QTLs associated with fruit size traits FWt, FWd, FL, PCTk, and NL exhibited positive and highly significant genetic correlations with fruit yield per plant, indicating their potential utility in yield improvement. These findings expand the current understanding of the genetic architecture of tomato yield and quality traits, providing new molecular targets for marker-assisted selection and offering valuable insights for breeding programs aimed at developing high-yielding and improved-quality tomato cultivars.

CHAPTER SIX

6.0 Combining ability and heterosis for yield and fruit quality in tomato

6.1 Introduction

Tomato is a commonly grown vegetable crop in Ethiopia, cultivated by small-scale and commercial farmers. It is a source of nutrition, income, foreign currency, and raw materials for agro-industries (Desalegne, 2002; Ketema *et al*., 2017). Open-pollinated and commercial hybrid varieties, entirely imported, are produced. Open-pollinated varieties are declining in productivity and have smaller fruit sizes than hybrid varieties. Because of the high yields, large fruit size, better shelf life, and high market value, growers prefer hybrid varieties. However, the seeds of the hybrids are expensive for small-scale farmers. Furthermore, few of the hybrid varieties are available.

The national productivity of tomato is low (6 t ha^{-1}). This is largely due to diseases, pests, parasitic weeds, limited access to high-quality seeds, knowledge gaps for GAP, and a lack of improved varieties resistant to biotic and abiotic stresses. Studies including Wondirad *et al.* (2009), Gashawbeza *et al*. (2009), Etagegnehu *et al.* (2009), and Mersha and Sime (2022) have highlighted these challenges.

Hybrid varieties with high yield, large fruit size, high pericarp thickness, better shelf life, and resistance to biotic and abiotic stresses would contribute to improving tomato productivity in the country. Identifying superior parental lines is a crucial aspect of developing hybrid varieties. Breeding experience has shown that elite parents with high performance *per se* may not necessarily produce superior hybrids (Riedelsheimer *et al*., 2012).

The general combining ability and the specific combining ability are used to identify parents for superior hybrids with high heterosis, albeit varying depending on the trait and cross-combinations.

Parents with a high GCA effect can produce either high or low SCA effect hybrids, and vice versa. Determining the association between the GCA of parental lines and the SCA of the hybrid and the relative importance of GCA and SCA enhances heterosis breeding efficiency.

Combining ability has been widely applied in various crop improvement programs, including tomato (Ayenan *et al*., 2022; Basté *et al*., 2010; Ene *et al*., 2023; Fortuny *et al*., 2021; Izzo *et al*., 2022; Z. Liu *et al*., 2021; Pavan *et al*., 2022; Shalaby, 2013; Tamta & Singh, 2018). Breeding for heterosis has been employed in tomato to improve yield, quality, and biotic and abiotic stress resistance (Ene *et a*l., 2023; Z. Liu *et al*., 2021; Solieman *et al*., 2013; Tamta & Singh, 2018). It is economical due to the high heterosis for most traits, including fruit yield, fruit number, TSS, fruit weight, pericarp thickness, fruit firmness (Solieman *et al*., 2013), and the considerably high seed yield for hybrid seed production (Tamta & Singh, 2018). In Ethiopia, few studies were conducted (Ene *et al*., 2023) to determine the level of combining ability and heterosis to identify parental lines and promising hybrid varieties. The present study was proposed with the following objectives:

1. to determine the combining ability for fruit yield, yield related and fruit quality
2. Identify potential parents and hybrids with combined high fruit yield and quality

6.2 Material and methods

6.2.1 Study site

The experiment was conducted in two stages. The first stage involved crossing selected male and female parental lines to produce F_1 progenies. The second stage involved field evaluation of the progenies, parents and standard check varieties. Crossing was conducted at the Melkassa Agricultural Research Center (MARC) under field conditions using irrigation from January to May

2023. The field evaluation was carried out at two locations (Table 6.1), MARC and Debrezeit Agricultural Research Center (DzARC), using irrigation from September to January 2024.

Table 6.1: Description of the study area

Location	Altitude (m)	Annual Rain fall (mm)	Soil type	Temperature (°C)	
				Min	Max
Melkassa	1550	818	Sandy loam	14	29
Debrezeit	1900	851	Alfisols/Mollisols	8.9	28.3

6.2.2 Plant materials and progeny development

Ten tomato parental lines, consisting of five males and five females (Table 6.2), were identified based on two seasons of field characterization (as described in Chapter 3). The parental lines were selected based on diversity of fruit yield, fruit number per plant, and fruit size, which were the traits that contributed the most to phenotypic variations. During selection, phenotypic genetic distance, a significant and positive correlation of traits with fruit yield, high to medium broad sense heritability, and genetic advance were considered. A 5 × 5 North Carolina Design II (NCII) mating scheme was employed, crossing five female and five male parental lines to generate 25 F_1 hybrids for evaluation (Table 6.3). The protocol of Worldveg (Opeña *et al.*, 2001) was followed for emasculation, pollen collection, hybridization, and seed extraction.

Table 6.2: Description of parental lines used in the study

Male	Cluster	Fruit shape	Growth habit	Fruit number per plant	Fruit size	Type
G6	1	Oblong	Semi-indeterminate	High	Medium	Processing
G7	2	Oblong	Determinate	Medium	Large	Processing
G8	1	Pear	Determinate	High	Small	Processing
G9	1	Globe	Semi-indeterminate	High	Medium	Fresh market
G10	3	Oblong	Determinate	Low	Large	Processing
Female						
G1	3	Globe	Determinate	Low	Large	Fresh market
G2	2	Oblong	Determinate	Low	Large	Processing
G3	2	Oblong	Determinate	Medium	Large	Processing
G4	2	Oblong	Determinate	Medium	Large	Processing
G5	2	Oblong	Determinate	Low	Large	Processing

Table 6.3: Schematic of the North Carolina Design II (NCII) mating design, showing the 25 F_1 hybrids generated from a factorial cross between five male and five female parental lines

Male/Female	G1	G2	G3	G4	G5
G6	G1 x G6	G2 x G6	G3 x G6	G4 x G6	G5 x G6
G7	G1xG7	G2 x G7	G3 x G7	G4 x G7	G5 x G7
G8	G1xG8	G2 x G8	G3 x G8	G4 x G8	G5 x G8
G9	G1xG9	G2 x G9	G3 x G9	G4 x G9	G5 x G9
G610	G1xG10	G2 x G10	G3 x G10	G4 x G10	G5 x G10

6.2.3 Evaluation of progenies

Twenty-five F_1 progenies, ten parental lines, two standard checks varieties, one of which was an OPV, and one popular commercial hybrid variety were evaluated for yield and fruit quality. The field experiment was arranged in a 6 x 6 alpha lattice design. Each plot consisted of two rows, measuring 3 m in length and 2 m in width, giving a total area of 6 m^2 and containing 12 plants. Spacing of 100 cm between rows and 50 cm between plants was used. Fertilizer was applied as per the national recommendation for tomato production (242 kg ha^{-1} NPS and 79 kg ha^{-1} urea). All other necessary cultural practices and crop protection measures were applied uniformly to all plots.

Data were collected for plant height (PH), days to 50% flowering (DFF), marketable (MNP), and total number (TNP) of fruits per plant, marketable (MYP), and total yield (TYP) per plant, fruit weight (FWt), fruit width (FWd), fruit length (FL), pericarp thickness (PCTk), number of locules (NL), TSS, and juice volume (JV). (Appendix Table 3.1).

6.2.4 Data analysis

Analysis of variance was carried out to determine the differences among the F_1 progenies, male and female parents, male x female, and their interaction with the environment. The AGDR software was used for the analysis. The Noth Carolina Design II model was used to generate the GCA for males and females, and the SCA for hybrids. Mixed-effects model was implemented, where genotypes (males, females, and hybrids) and locations were treated as fixed effects, while

all the interactions (male x location, female x location, male x female x location) were treated as random effects. The same software was used to estimate GCA and SCA effects, variance components, and broad and narrow sense heritability. The components of variance, along with the general combining ability (GCA) effects of the male and female lines and the specific combining ability (SCA) effects of the crosses and test for significance, were estimated according to the methods described by Dabholkar (1999).

Male GCA effect (g_i) = $\frac{xi..}{rf} - \frac{x...}{mfr}$

Female GCA effect (g_j) = $\frac{xj..}{rm} - \frac{x...}{mfr}$

Cross (M x F) SCA effect (s_{ij}) = $\frac{xij.}{r} - \frac{xi..}{fr} - \frac{xj..}{mr} + \frac{x...}{mfr}$

Where g_i = the GCA effect of the i^{th} male line and g_j = the GCA effect of the j^{th} female line, $X_{i..}$ = sum of the i^{th} male lines; x_j = sum of the j^{th} female lines; $x^{...}$ = grand sum of crosses, s_{ij} = SCA effect of the ij^{th} cross, $x_{ij.}$ = sum of the i x j cross, m = number of male lines, f = number of female lines, and r = number of replications.

The significance of the GCA and SCA effects for each male and female line, and the crosses was determined by dividing the corresponding GCA and SCA effects by their respective standard errors. The ratio was then compared with the tabular t-value, using the error degree of freedom. Where the GCA and SCA effects were significant, the difference between each line or cross effect was tested using the standard error of the difference (SEd).

SE (GCA of male) = $\sqrt{\frac{MSe}{rf}}$

SE (GCA of female) = $\sqrt{\frac{MSe}{rm}}$

SE (SCA of M x F) = $\sqrt{\frac{MSe}{r}}$

SEd (g_i-g_j) male = $\sqrt{\frac{2MSe}{fr}}$

SEd (g_i-g_j) female = $\sqrt{\frac{2MSe}{mr}}$

SEd (s_{ij}-s_{kl}) M x F = $\sqrt{\frac{2MSe}{r}}$

Where SE = standard error, SEd = standard error of difference, MSe = mean square error, g_i = GCA effect of male or female line i, g_j = GCA effect of male or female line j, s_{ij} = SCA effect of cross male line i and female line j, and s_{kl} = SCA effect of cross male line k and female line l.

Mid parent and better parent were computed using the following formula. The standard heterosis for the open pollinated and hybrid standard check varieties was not computed, due to the fact that their performance was very low.

$$MPH = \frac{F_1 - MP}{MP} \times 100\%$$

$$BPH = \frac{F_1 - BP}{BP} \times 100\%$$

Where MPH = midparent heterosis, BPH = better parent heterosis, F_1 = the mean of the hybrid, MP = the average of the mean value of the male and female parents of the respective F_1 hybrid.

6.3 Results

6.3.1 Genetic variability for yield, yield related and fruit quality

The variance due to hybrids pooled across locations revealed highly significant ($p < 0.01$) and significant differences ($p < 0.05$) for all traits (Table 6.4). The combined ANOVA for the combining ability indicated that the variance attributed to females was highly significant for all traits except for days to 50% flowering (DFF) and total soluble solid (TSS). Similarly, it was highly significant ($p < 0.01$) for most traits and significant ($p < 0.05$) for DFF among male parental lines, except TSS. The male-female interaction genetic variance was also highly significant for most traits except fruit length, plant height, number of locules, and fruit width.

The hybrid-by-location interaction variance was highly significant ($p<0.01$) for total and marketable fruit number and yield per plant, fruit skin thickness, juice volume, and TSS. The partitioning of the F_1 with environment interaction indicated that the variance due to male x location interaction was highly significant ($p<0.01$) for the total and marketable number of fruits, yield per plant, and juice volume. Similarly, the female x location interaction variance was also highly significant ($p<0.01$) for total and marketable fruit number and yield per plant, juice volume, and average fruit weight. The male-female-location interaction was significant ($p<0.05$) for total and marketable fruit number and yield per plant, plant height, and TSS.

Table 6.4: Means squares of 13 traits in tomato for two locations, MARC and DZARC, and combined

Sources of variation	Df	MYP	TYP	MNP	TNP	DFF	PH	FWt	FWd	FL	PCTk	NL	TSS	JV
Site	1	5.2**	6.86**	3776.8**	5294.5**	112.4**	685.4**	2557**	18.3	586 **	17 **	0	2.1**	57240.6**
R(site)	2	0.1	0.09	5.2	5.4	3.4	315	267.9	6.2	15	0.1	0.1	0	5780.6
Genotypes	24	1.0**	1.03**	1252.2**	1306.8**	7.1 *	1006.6**	3087**	58 **	458 **	1.8 **	2.4**	0.45**	93181.5**
Male	4	0.5**	0.44**	4299.4**	4121.8**	10.3 *	3168.4**	13250**	170**	2149 **	2.8 **	8.0**	0.29	351455.0**
Female	4	2.2**	2.28**	2117.2**	2500.7**	2.9	699.3**	1947**	85 **	391 **	4.1 **	4.7**	0.46*	87259.1**
Male: Female	16	0.8**	0.86**	274.2**	304.6**	7.4 *	542.9	831**	23.5	52	1.0 **	0.5	0.49**	30093.8**
Site: Genotypes	24	1.5**	1.53**	360.1**	368.3**	5.3	196.2	195	15.1	27	0.4 **	0.3	0.38*	13176.8**
Site: Male	4	1.2**	1.17**	507.4**	519.2**	6.5	186.3	28	15	27	0.2	0.1	0.13	20584.0**
Site: Female	4	1.9**	1.80**	429.1**	321.9**	7.9	133.8	604**	23.5	41	0.3	0.5	0.4	33923.1**
Site: Male: Female	16	1.5**	1.55**	305.9**	342.1**	4.4	214.3*	134	13.1	23	0.5	0.2	0.43**	6138.4
Residuals	48	0.14	0.12	33.7	39.1	3.4	113	152.8	22.3	26	0.33	0.3	0.16	5635

** Significant at p< 0.01, * significant at p<0.05; MYP = marketable fruit yield (kg) per plant, TYP = total yield (kg) per plant; MNP = marketable number of fruits per plant; TNP = total number of fruits per plant. Fruit FWt = single fruit weight (g), FWd = fruit width (mm), FL = fruit length (cm), PCTk = pericarp thickness (mm), NL = number of locules, DFF = days to 50% flowering, PH = plant height (cm), %TSS =total soluble solids, and JV = juice volume (ml).

6.3.2 Components of genetic variance and heritability

The genetic parameters, including variance components, their contributions, and broad- and narrow-sense heritability, are provided (Table 6.5). The additive and dominance genetic variances played a significant role in controlling most of the traits, as evidenced by the highly significant GCA and SCA mean squares across locations. For the traits TNP (53%, 32%), MNP (57%, 28%), FWt (72%, 11%), FWd (49%, 24%), FL (78%, 14%), PCTk (25%, 37%), PH (52%, 12%), NL (55%, 32%), and JV (63%, 16%), the GCA variance contributions of the male and female parental lines was high, with male parental lines having a greater effect on most of the traits, except for PCTk. On the other hand, SCA variance predominated for DFF (69%), MYP (54%), TYP (56%), and TSS (73%). The broad-sense heritability was generally high to moderate, ranging from 48% to 92%. High to moderate narrow sense heritability (32 - 82%) was also detected for most traits, except for DFF (1%), TYP (6%), MYP (7%), and TSS (0%). Traits with high narrow sense heritability, such as FWt, FL, and NL, were less susceptible to environmental influences, making them suitable for early generational selection to improve their performance. Conversely, traits with low narrow-sense heritability should be subjected to subsequent breeding and selection programs to improve their performance.

Table 6.5: Genetic components of variance and heritability of 13 traits in tomato for two locations, MARC and DzARC, and combined

Trait/source of variance	σ^2_{Hybrid}	σ^2_M	σ^2_F	σ^2_{M*F}	σ^2_A	σ^2_D	σ^2_e	h^2_b (%)	h^2_n (%)	Contribution			$\frac{(2\sigma^2 GCA}{2\sigma^2 GCA + S}$
										Male	Female	Male*Female	
DFF	0	0.15	0	0.99	0	3.97	4.37	0.48	0.0	0.24	0.07	0.69	0.23
TYP	0.025	0	0.07	0.18	0.14	0.74	0.83	0.50	0.06	0.07	0.37	0.56	0.44
MYP	0.03	0	0.07	0.17	0.11	0.68	0.82	0.49	0.07	0.09	0.37	0.54	0.45
TNP	150.3	191	109.8	66.4	601.3	265.5	203.7	0.81	0.56	0.53	0.32	0.16	0.90
MNP	146.7	201.3	92.2	60.1	586.8	240.5	196.9	0.81	0.57	0.57	0.28	0.15	0.91
FWt	338.4	620.96	55.83	169.48	1353.59	677.93	173.83	0.92	0.61	0.72	0.11	0.18	0.89
FWd	5.18	7.31	3.06	0.31	20.74	1.24	22.27	0.5	0.47	0.49	0.24	0.27	0.99
FL	60.9	104.9	16.9	6.6	243.6	26.5	26.1	0.91	0.82	0.78	0.14	0.08	0.97
PCTk	0.12	0.088	0.152	0.175	0.48	0.7	0.382	0.75	0.307	0.25	0.37	0.38	0.73
PH	69.55	131.27	7.82	107.47	278.2	429.9	154.6	0.82	0.32	0.52	0.12	0.36	0.72
TSS	0.0	0.0	0.0024	0.08	0.0	0.33	0.27	0.55	0.0	0.11	0.17	0.73	0.06
NL	0.2913	0.279	0.211	0.045	1.165	0.179	0.32	0.81	0.70	0.55	0.32	0.14	0.96
JV	9463.2	16068.1	2858.3	6114.7	37852.7	24458.8	9405.9	0.87	0.53	0.63	0.16	0.22	0.86

σ^2Hybrid = variance due to crosses, σ^2_M= variance due to male line, σ^2_F=variance due to female line, σ^2_{M*F}= variance due to male-female interaction, σ^2_A= additive genetic variance, σ^2_D=dominance genetic variance, σ^2_e = environmental variance, h^2_b = broad sense heritability, h^2_n= narrow sense heritability, M= male, F=female, M*F= male-female interaction.

6.3.3 Parental performance *pe se* and general combining ability effects

6.3.3.1 Yield and number of fruits per plant

Performance *per se*, for total and marketable yield per plant, varied between 2.67 kg and 3.42 kg for TYP, 2.57 kg and 3.27 kg for MYP among females; 2.63 kg and 3.71 kg for TYP, and 2.53 kg and 3.6 kg for MYP among male parental lines (Table 6.6). The female parental lines showed highly significant ($p<0.01$) and significant ($p<0.05$) positive or negative GCA effects for TYP and MYP, while non-significant GCA effects were observed among the male parental lines. The GCA effects ranged from -0.37 to 0.51 for TYP in female parental lines and -0.16 to 0.17 in male parental lines. Female parental lines G1 and G5 exhibited significant negative GCA effects for both traits. In contrast, female line G2 demonstrated the highest and most significant positive GCA effect, identifying it as the best general combiner for yield per plant. Among the male lines, G6 and G7 showed the highest positive GCA effects for both traits.

Mean number of marketable fruits per plant (MNP) and total number of fruits per plant (TNP) ranged from 28 to 43 and 32 to 52 for female parental lines respectively. Similarly, the male parental lines showed a wider range of variation for the traits: 30 to 123 for MNP and 32 to 128 for TNP. Highly significant positive or negative GCA effects were observed among the female and male parental lines. Critically, the best combiner parents were identified as female lines G2 and G4 and male lines G6, G8, and G9, which showed highly significant positive GCA effects for both traits. Conversely, lines G1 and G3 (females) and G7 and G10 (males) showed highly significant negative GCA effects, identifying them as poor combiners.

6.3.3.2 Fruit weight and related traits

For fruit size traits such as fruit weight (FWt), fruit width (FWd), and fruit length (FL), a broad range of genetic variation was observed in the *per se* performance of male and female parental lines. Among female parental lines, FWt, FWd, and FL varied from 82.65 g to 163.3 g, 44.43 mm to 67.28 mm, and 51.38 mm to 71.25 mm, respectively. Among the male parental lines, it ranged from 49.95 g to 140.7 g, 35.9 mm to 51.8 mm, and 44.95 mm to 84.52 mm. The range of GCA effects among the female parental lines was -12.97–13.95, -2.74–2.57, and -7.8–2.74 for FWt, FWd, and FL, respectively. Similarly, among the male paternal lines, the GCA effects varied from -19.44 to 35.41, -3.62 to 4.27, and 9.05 to 13.54 for FWt, FWd, and FL, respectively. Highly significant positive GCA effects were observed for female line G2 and male lines G7 and G10 for FWt, male line G10 for FWd, and male lines G7 and G10 for FL. Highly significant negative GCA effects were observed for female line G4 and male lines G6, G8, and G9 for FWt; male line G6 for FWd; and female line G1 and male lines G6, G8, and G9 for FL. The lines with the largest favorable GCA effects were G7 (FL), G10 (FWd), and G7 and G10 (FWt) among male lines, and lines G2 (FWt), G1 (FWd), and G3 (FL) among female parental lines. The best general combiner parents for each trait were identified based on the magnitude of favorable GCA effects. For fruit length (FL), the superior combiners were male lines G7 and G10. For fruit width (FWd), male line G10 was the best combiner. Regarding fruit weight (FWt), the superior lines were male lines G7 and G10, along with female line G2.

The mean values for fruit wall thickness ranged from 3.79 mm to 6.72 mm and 4.16 mm to 5.47 mm for female and male parental lines, respectively. None of the female parental lines showed significant GCA effects, but line G7 among male parents showed highly significant GCA effect for PCTk. The mean number of locules ranged from 2.33 to 6.6 in female parental lines and from

2.15 to 4.75 in male parental lines. Female parental line G1 showed a positive and highly significant GCA effect for the number of locules. Female parental lines G4 and G5 and male lines G6 showed a negative and significant GCA effects for number of locules.

6.3.3.3 Plant height

Mean plant height ranged from 62.95 cm to 82.45 cm for female parental lines and from 65.85 cm to 101.5 cm for male parental lines. GCA effects ranged from -7.42 to 7.32 for females and -9.88 to 18.9 for males. Female line G3 and male lines G6 and G9 displayed significant positive GCA effects for plant height, identifying them as the best general combiners for this trait. Female lines G1 and male lines G7, G8, and G10 showed negative and significant GCA effects.

6.3.3.4 Total soluble solid and juice volume

The mean performance ranged from 4.75 to 5.48 for females and 4.48 to 6 for males. GCA effects ranged from -0.16 to 0.22 for females and -0.13 to 0.16 for males. No significant GCA effects were observed among male and female parental lines for total soluble solid (TSS), but female line 5 and male line 7 possessed the highest positive GCA effects. Regarding juice volume (JV), the mean value ranged from 408.75 ml to 877.5 ml in females and 218.75 ml to 685 ml in males. The general combining ability (GCA) effects for juice volume exhibited a wide range, from -96.63 to 63.00 in female lines and -112.13 to 172.38 in male lines. Among these, female line G2 and male lines G7 and G10 were identified as the best general combiners for this trait, based on their highly significant positive GCA effects. A highly significant negative GCA effects was recorded among female parental lines G4 and G5 and male parental lines G6, G8, and G9.

Table 6.6*: Per se* performance of male and female parental lines with their respective general combining ability effects for 13 traits in tomato across two locations

Parent	MYP		TYP		MNP		TNP		FWt		FWd	
	Mean	GCA	Mean	GCA	Mean	GCA	Mean	GCA	Mean	GCA	Mean	GCA
Female												
G1	3.27	-0.23*	3.42	-0.24*	29.79	-14.08**	31.41	-14.67**	155.8	1.26	67.28	2.57
G2	2.57	0.53**	2.67	0.51**	36.83	8.76**	38.39	10.65**	143.59	13.95**	55.02	1.10
G3	2.68	0.08	3.01	0.11	27.7	-6.49**	31.67	-7.56**	163.3	2.24	60.85	0.31
G4	2.75	-0.05	2.94	-0.02	56.72	10.17**	62.03	10.65**	82.65	-12.97**	44.43	-2.74
G5	3.1	-0.33**	3.35	-0.37**	42.51	1.64	52.39	0.94	104.55	-4.48	48.1	-1.24
Male												
G6	2.79	0.11	2.82	0.09	68.61	7.25**	70.17	7.45**	64.78	-18.07**	40.61	-3.62*
G7	2.74	0.16	2.87	0.17	49.32	-7.98**	54.02	-6.27**	99.55	19.62**	45.88	0.86
G8	3.6	0.06	3.71	0.05	122.51	15.17**	128.02	13.95**	49.95	-17.52**	35.9	-0.13
G9	2.69	-0.13	2.78	-0.16	72.14	7.08**	76.03	6.90**	57.75	-19.44**	44.27	-1.38
G10	2.53	-0.21	2.63	-0.15	29.75	-21.52**	31.52	-22.07**	140.7	35.41**	51.28	4.27**
SE		0.12		0.11		1.84		1.98		3.91		1.49
SEd($g_i - g_j$)		0.167		0.155		2.596		2.796		5.528		2.112

Table 6.6 (continued)

Parent	FL		PCTk		NL		DFF		PH		TSS		JV	
	Mean	GCA	Mean	GCA	Mean	GCA	Mean	GCA	Mean	GCA	Mean	GCA	Mean	GCA
Female														
G1	51.38	-7.80**	3.79	-0.776**	6.6	0.695**	37.4	-0.04	62.95	-7.42*	4.98	-0.16	877.5	38.13
G2	66.02	1.71	5.84	0.012	3.16	0.06	40.59	-0.29	73.1	-3.71	4.78	0.03	649.38	63.00**
G3	71.25	2.74	5.66	0.304	4.01	0.15	38.03	0.66	73.05	7.32*	5.18	-0.12	842.5	34.38
G4	69.02	2.48	5.65	0.147	2.35	-0.54**	40.88	-0.24	82.45	4.15	4.75	0.04	408.75	-96.63**
G5	62.54	0.87	6.72	0.312	2.33	-0.42*	42.47	-0.09	65.76	-0.35	5.48	0.22	539.38	-38.98
Male														
G6	63.44	-3.91*	4.96	0.20	2.15	-0.55**	42.5	0.56	99.44	18.9**	5.73	-0.01	305	-105.63**
G7	84.52	13.54**	5.47	0.53**	3.15	-0.03	41.45	-0.34	73.4	-9.2**	5.48	0.16	570	111.63**
G8	56.32	-9.05**	4.16	-0.15	2.65	-0.23	38.72	-0.04	65.85	-6.77*	4.58	-0.13	218.75	-112.13**
G9	44.95	-8.93**	4.35	-0.13	2.85	-0.31	42.47	-0.99	101.5	6.96*	6	0.07	295	-66.25**
G10	73.57	8.35**	4.43	-0.45*	4.75	1.07	40.42	0.81	71.6	-9.88**	5.13	-0.09	685	172.38**
SE		1.61		0.18		0.17		0.58		3.36		0.13		23.74
SEd ($g_i - g_j$)		2.28		0.257		0.245		0.825		4.754		0.179		33.571

6.3.4 Hybrid performance *per se*, heterosis and specific combining ability effects

6.3.4.1 Yield and number of fruits per plant

Data for *per se* performance, specific combining ability (SCA) effects, mid-parent heterosis, and better-parent heterosis are provided (Tables 6.7 & 6.8). Hybrids exhibited better *per se* performance in most of the traits across locations. TYP and MYP ranged from 2.96 kg to 5.02 kg and 2.83 kg to 4.88 kg, respectively. Most of the hybrids (over 80%) for TYP and over 88% for MYP showed positive heterosis. The MPH and BPH for TYP ranged from -1% to 81% and -12% to 75%, respectively, while it ranged from 1% to 84% and -8% to 79% for MYP.

The SCA for TYP and MYP ranged from -0.82 to 0.56 and -0.83 to 0.53, respectively. Hybrid G2 x G7 showed a significant positive SCA effect, whereas hybrids G3 x G9, G2 x G8, and G4 x G7 showed highly significant negative SCA effects for TYP and MYP. Hybrids G3 x G6, G3 x G7, G1 x G10, and G4 x G9 showed high non-significant SCA effect for both traits. Hybrids G2 x G7, G3 x G7, and G3 x G6 showed 81%, 41%, and 49% MPH. Similarly, these hybrids exhibited 75%, 37%, and 46% heterosis over the better parent. Hybrid G2 x G7 showed a significant positive SCA effect. Hybrid G2 x G7 was identified as the best specific combiner for yield, as it exhibited the highest values for mid-parent heterosis (81%) and better-parent heterosis (75%), coupled with a significant positive SCA effect.

The average MNP and TNP varied from 36 to 95 and from 34 to 91 respectively. MPH and BPH also varied from -25% to 74% and from -52% to 49% for TNP, respectively, and from -21% to 64% and from -51% to 42% for MNP. The SCA ranged from -19.3 to 11.2 and from -19.4 to 9.9 for TNP and MNP, respectively. Significant positive SCA effects were recorded for hybrids G1 x G10, G5 x G9, G2 x G7, and G3 x G8, while hybrids G2 x G8, G2 x G10, and G3 x G9 showed

negative SCA effects for TNP. Hybrids G2 x G9, G1 x G10, G5 x G9, and G3 x G8 showed a positive and significant SCA effect for MNP, while hybrid G4 x G9 showed a negative and significant effect on this trait. Though no significant positive SCA effect was detected for hybrids G3 x G7 and G3 x G6, they showed higher positive SCA effects for TNP and MNP. Hybrids G2 x G7, G3 x G7, and G3 x G6 exhibited 74%, 38%, and 42% MPH, respectively. The BPH was also 49%, 9%, and 3% for these hybrids. Hybrids G2 x G7, G1 x G10, and G5 x G9 were identified as the most promising for increasing the number of fruits per plant.

6.3.4.2 Fruit weight and related traits

Fruit weight and width ranged from 68.8 g to 167.6 g and 42.8 mm to 57.11 mm, respectively. Midparent heterosis varied from -38% to 20% for fruit weight, while it ranged from -14% to 26% for fruit width. Similarly, the better parent heterosis varied from -58% to 17% for fruit weight and -31% to 14% for fruit width. The SCA effect varied from -27.37 to 20.2 and from -3.51 mm to 4.66 mm for fruit weight and width, respectively. Hybrids G2 x G10 and G4 x G6 showed a significant positive SCA effect, while hybrids G4 x G7 showed a highly significant negative SCA effect for fruit weight. None of the hybrids showed a significant SCA effect on fruit width. Although none significant SCA effects were detected for fruit weight among G2 x G7 and G3 x G7 hybrids, it was high and positive. The highest midparent (20%) and better parent (17%) heterosis were recorded for hybrids G2 x G7 and G2 x G10, respectively.

The fruit pericarp thickness (PCTk), fruit length (FL), and number of locules (NL) exhibited considerable variability among hybrids. FL ranged from 47.83 to 82.85 mm, PCTk from 4.05 to 6.31 mm, and NL was consistently low, ranging from 2 to 5. Mid-parent heterosis (MPH) and better-parent heterosis (BPH) ranged from -18% to 10% and -25% to 4% for FL, from -21% to

25% and -30% to 31% for PCTk, and from -28% to 20% and -52% to 11% for NL, respectively. Specific combining ability (SCA) effects also varied, spanning from -6.65 to 4.63 for FL and -0.77 to 0.94 for PCTk. For PCTk, hybrid G1 x G8 was the top performer, exhibiting the highest heterosis (25% MPH, 20% BPH), while G4 x G10 and G2 x G6 were confirmed as superior specific combiners due to their significant positive SCA effects. For FL, no significant SCA effects were detected; however, G2 x G7 and G4 x G10 showed the highest MPH (10%) and BPH (3%), respectively, suggesting limited potential for specific genetic improvement. Regarding NL, no significant SCA effects were observed, and the predominantly negative SCA effects in hybrids such as G2 x G6, G5 x G8, G4 x G10, and G3 x G9 were detected.

6.3.4.3 Total soluble solids and juice volume

Mean values for juice volume (JV) and total soluble solids (TSS) ranged from 372.62 to 889.96 ml and 4.38° to 6.03°, respectively. Highly significant specific combining ability (SCA) effects were observed for both traits. Based exclusively on significant SCA effects, hybrid G5 x G9 was identified as the superior combiner for TSS, while hybrids G1 x G9 and G2 x G10 were the best for JV. The heterosis for JV ranged from -37% to 33% for mid-parent (MPH) and -58% to 29% for better-parent (BPH). For TSS, the heterosis was narrower, with MPH from -19% to -10% and BPH from -24% to 5%. Notably, hybrid G2 x G10, one of the best combiners for JV, also exhibited the highest heterosis for this trait (33% MPH, 29% BPH). Hybrid G2 x G7 showed high positive SCA effects that were not statistically significant, yet it demonstrated appreciable heterosis for JV (20% MPH, 13% BPH) and TSS.

6.3.4.4 Plant height and days to 50% flowering

Mean values for plant height (PH) and days to 50% flowering (DFF) ranged from 63.39 to 121.09 cm and 38 to 42 days, respectively. Mid-parent heterosis (MPH) and better-parent heterosis (BPH) for PH ranged from -9% to 57% and -22% to 33%, while for DFF, they varied from -6% to 11% and -4% to 17%, respectively. Hybrid G3 x G8 was the best combiner for plant height, based on its significant positive SCA effect. In contrast, hybrid G3 x G9 exhibited a significant negative SCA effect for this trait. For DFF, no hybrids showed significant SCA effects, though G4 x G8 and G1 x G10 were notable for having the lowest (least negative) SCA effects.

Table 6.7: Performance *per se* of parents and hybrids, the highest and lowest specific combining ability effects, mid parent and better parent heterosis for yield and quality

Trait	Hybrid	F1	P1	P2	SCA	MPH	BPH	Trait	Hybrid	F1	P1	P2	SCA	MPH	BPH
TYP	G2xG7	5.02	2.67	2.87	0.56*	0.81	0.75	DFF	G1xG8	42.15	37.40	38.73	2.49	0.11	0.13
	G4xG9	4.42	3.01	2.82	0.43	0.52	0.47		G4xG10	41.95	40.88	35.81	1.59	0.09	0.17
	G3xG6	4.04	2.94	2.78	0.43	0.41	0.37		G5xG9	39.90	42.47	34.56	1.24	0.04	0.15
	G5xG8	3.88	3.35	3.71	0.40	0.10	0.05		G1xG6	41.37	37.40	42.50	1.14	0.04	0.11
	G1xG10	3.74	3.42	2.63	0.34	0.24	0.09		G2xG10	41.23	40.59	38.93	0.89	0.04	0.06
	G3xG7	4.39	3.01	2.87	0.32	0.49	0.46		G4xG6	41.06	40.88	37.06	0.84	0.05	0.11
	G1xG6	3.18	3.42	2.82	-0.47	0.02	-0.07		G3xG9	40.29	38.03	37.68	0.74	0.06	0.07
	G3xG9	3.09	3.01	2.78	-0.65*	0.07	0.03		G3xG7	40.70	38.03	38.31	0.59	0.07	0.07
	G2xG8	3.53	2.67	3.71	-0.82**	0.11	-0.05		G2xG7	39.76	40.59	39.87	0.54	-0.01	0.00
	G4xG7	3.12	2.94	2.87	-0.82**	0.07	0.06		G3xG6	39.58	38.03	38.62	-1.56	0.03	0.04
MYP	G2xG7	4.88	2.57	2.74	0.53*	0.84	0.78		G4xG8	40.88	36.44	38.66	-1.81	-0.02	0.04
	G3xG6	4.33	2.68	2.79	0.48	0.58	0.55		G1xG10	37.40	40.43	38.92	-2.11	-0.01	0.03
	G1xG10	3.89	2.75	2.69	0.41	0.43	0.45	PH	G3xG8	109.30	73.05	65.85	22.72**	0.57	0.33
	G4xG9	3.60	3.27	2.53	0.38	0.24	0.10		G4xG9	112.13	82.45	101.50	14.99	0.22	0.10
	G3xG7	4.25	2.68	2.74	0.35	0.57	0.58		G2xG7	84.75	73.10	71.60	12.32	0.17	0.14
	G3xG9	2.94	2.68	2.69	-0.66*	0.10	0.10		G2xG10	84.50	73.10	73.40	11.39	0.15	0.15
	G2xG8	3.54	2.57	3.60	-0.71**	0.15	-0.02		G1xG9	96.46	62.95	101.50	10.90	0.17	-0.05
	G4xG7	2.95	2.75	2.74	-0.83**	0.07	0.07		G3xG6	123.04	73.05	99.44	10.79	0.43	0.24
TNP	G1xG10	41.34	31.41	31.52	11.17*	0.31	0.31		G3xG7	76.15	73.05	73.40	-8.00	0.04	0.04
	G5xG9	83.98	52.39	76.03	9.23*	0.31	0.10		G2xG9	79.15	73.10	101.50	-10.12	-0.09	-0.22
	G2xG7	80.37	38.39	54.02	9.08*	0.74	0.49		G2xG8	63.05	73.10	65.85	-12.49	-0.09	-0.16
	G3xG8	82.29	31.67	128.02	8.99*	0.03	-0.36		G3xG9	83.90	73.05	101.50	-16.41*	-0.04	-0.17
	G2xG9	92.08	38.39	76.03	7.62	0.61	0.21	PCTk	G4xG10	5.90	5.65	4.43	0.94*	0.17	0.08
	G3xG7	58.98	31.67	54.02	5.90	0.38	0.09		G2xG6	6.30	5.84	4.96	0.82*	0.17	0.31
	G3xG6	72.22	31.67	70.17	5.37	0.42	0.03		G1xG8	4.98	3.79	4.16	0.63	0.25	0.20
	G2xG8	83.06	38.39	128.02	-8.45	0.00	-0.35		G2xG7	6.17	5.84	5.47	0.36	0.09	0.12
	G2xG10	45.88	38.39	31.52	-9.62*	0.31	0.20		G3xG7	6.35	5.66	5.47	0.25	0.14	-0.06
	G3xG9	47.01	31.67	76.03	-19.3**	-0.13	-0.38		G5xG7	6.35	6.72	5.47	0.24	0.04	-0.03
MNP	G2xG9	89.11	36.83	72.14	9.88*	0.64	0.24		G4xG9	5.49	5.65	4.35	0.20	0.10	0.04
	G1xG10	37.25	29.79	29.75	9.49*	0.25	0.25		G3xG6	5.91	5.66	4.96	0.14	0.11	0.09
	G5xG9	80.58	42.51	72.14	8.48*	0.41	0.12		G2xG8	4.66	5.84	4.16	-0.48	-0.07	-0.06
	G3xG8	80.52	27.70	122.51	8.46*	0.07	-0.34		G4xG7	5.30	5.65	5.47	-0.65	-0.05	-0.30
	G3xG7	55.29	27.70	49.32	6.39	0.44	0.12		G2xG10	4.06	5.84	4.43	-0.77	-0.21	-0.30

Table 6.7 (continued)

	Cross	F1	P1	P2	SCA	MPH	BPH	Trait	Hybrid	F1	P1	P2	SCA	MPH	BPH
MNP	G3xG6	70.48	27.70	68.61	6.34	0.46	0.03	NL	G3xG10	4.90	4.01	4.75	0.42	0.12	0.03
	G2xG7	70.40	36.83	49.32	6.25	0.63	0.43		G2xG8	3.50	3.16	2.65	0.41	0.20	0.11
	G4xG9	27.70	72.14	44.60	-19.4**	-0.11	-0.38		G5xG9	2.90	2.33	2.85	0.37	0.12	0.02
FL	G4xG6	66.20	69.02	63.44	4.63	0.00	-0.04		G1xG10	5.35	6.60	4.75	0.32	-0.06	-0.19
	G2xG7	82.85	66.02	84.52	4.59	0.10	-0.02		G4xG9	2.70	2.35	2.85	0.29	0.04	-0.05
	G1xG8	50.36	51.38	56.32	4.21	-0.06	-0.11		G3xG6	3.15	4.01	2.15	0.29	0.02	-0.21
	G5xG8	58.40	62.54	56.32	3.57	-0.02	-0.07		G2xG7	3.45	3.16	3.15	0.16	0.09	0.09
	G5xG7	80.24	62.54	84.52	2.82	0.09	-0.05		G3xG7	3.40	4.01	3.15	0.02	-0.05	-0.15
	4GxG10	76.09	69.02	73.57	2.26	0.07	0.03		G2xG6	2.40	3.16	2.15	-0.37	-0.10	-0.24
	2GxG10	74.95	66.02	73.57	1.88	0.07	0.02		G3xG9	2.70	4.01	2.85	-0.40	-0.21	-0.33
	G1xG9	47.94	51.38	44.95	1.66	0.00	-0.07		G5xG8	2.20	2.33	2.65	-0.41	-0.12	-0.17
	G3xG6	63.19	71.25	63.44	1.36	-0.06	-0.11		G4xG10	3.20	2.35	4.75	-0.59	-0.1	-0.38
	G3xG7	80.53	71.25	84.52	1.24	0.03	-0.05	JV	G1xG9	667.50	877.50	295.00	171.50**	0.14	-0.24
	G2xG8	50.16	66.02	56.32	-5.52	-0.18	-0.24		G2xG10	885.00	649.38	685.00	125.50*	0.33	0.29
	G4xG7	72.46	69.02	84.52	-6.56	-0.06	-0.14		G4xG6	417.50	408.75	305.00	95.62	0.17	0.02
FWt	G2xG10	167.60	143.59	140.70	20.20*	0.18	0.17		G3xG7	745.00	842.50	570.00	74.88	0.05	-0.12
	G4xG6	85.35	82.65	64.78	18.35*	0.16	0.03		G4xG8	382.50	408.75	218.75	67.12	0.22	-0.06
	G1xG9	95.90	155.80	57.75	16.04	-0.10	-0.38		G5xG7	636.25	539.38	570.00	39.37	0.15	0.12
	G2xG7	145.50	143.59	99.55	13.89	0.20	0.01		G2xG7	731.25	649.38	570.00	32.50	0.20	0.13
	G4xG9	76.60	82.65	57.75	10.97	0.09	-0.07		G1xG6	375.00	877.50	305.00	-81.63	-0.37	-0.57
	G3xG7	130.50	163.30	99.55	10.60	-0.01	-0.20		G3xG9	397.50	842.50	295.00	-94.75	-0.30	-0.53
	G3xG6	76.85	163.30	64.78	-5.36	-0.33	-0.53		G4xG7	402.50	408.75	570.00	-136.63**	-0.18	-0.29
	G4xG7	77.35	82.65	99.55	-27.3**	-0.15	-0.22	TSS	G5xG9	6.03	5.48	6.00	0.62*	0.05	0.10
FWd	G4xG8	50.59	44.43	35.90	4.66	0.26	0.14		G2xG7	5.65	4.78	5.48	0.35	0.10	0.03
	G1xG9	53.03	67.28	44.27	3.03	-0.05	-0.21		G3xG10	5.20	5.18	5.13	0.30	0.01	0.00
	G2xG10	57.11	55.02	51.28	2.95	0.07	0.04		G5xG6	5.60	5.48	5.73	0.28	0.00	-0.02
	G5xG6	46.04	48.10	40.61	2.10	0.04	-0.04		G3xG6	5.10	5.18	5.73	0.12	-0.06	-0.11
	G1xG7	53.80	67.28	45.88	1.57	-0.05	-0.20		G3xG7	5.00	5.18	5.48	-0.15	-0.06	-0.09
	G2xG7	52.17	55.02	45.88	1.43	0.03	-0.05		G3xG9	4.55	5.18	6.00	-0.51	-0.19	-0.24
	G3xG10	54.28	60.85	51.28	0.91	-0.03	-0.11		G1xG6	4.38	4.98	5.73	-0.56*	-0.18	-0.24
	G3xG6	46.38	60.85	40.61	0.89	-0.09	-0.24								
	G3xG7	49.80	60.85	45.88	-0.16	-0.07	-0.18								
	G4xG10	47.37	44.43	51.28	-2.95	-0.01	-0.08								
	G1xG8	47.73	67.28	35.90	-3.51	-0.07	-0.29								

F1 = hybrid performance; p1 = female parent performance, p2 = male parent performance. MYP = marketable fruit yield (kg) per plant, TYP = total yield (kg) per plant, MNP = marketable number of fruits per plant; TNP = total number of fruits per plant, fruit FWt = single fruit weight (g), FWd = fruit width (mm), FL = fruit length (cm), PCTk = pericarp thickness (mm), NL = number of locules, DFF = days to 50% flowering, PH = plant height (cm), %TSS =total soluble solids, JV = juice volume (ml).

Table 6.8: Range of performance *per se*, heterosis and SCA effects among crosses

Trait	Performance *pe se* of crosses	MPH (%)	BPH (%)	SCA effect
TYP	2.96-5.02	-1-81	-12-75	-0.82-0.56
MYP	2.83-4.88	1-84	-8-78	-0.83-0.53
TNP	36.63-94.5	-25-74	-52-49	-19.3-11.2
MNP	33.55-91.32	-21-64	-51-42	-19.4-9.9
DFF	37.59-42.15	-6-11	-4-17	-2.11-2.49
PH	63.39-121.09	-9-57	-22-33	-16.41-22.72
PCTk	4.05-6.31	-21-25	-30-31	-0.77-0.94
FL	47.83-82.85	-18-10	-25-3	-6.56-4.63
FWt	68.8-167.6	-38-20	-58-17	-27.34-20.2
FWd	42.8-57.11	-14-26	-31-14	-3.51-4.66
NL	2.20-5.35	-28-20	-52-11	-0.59-0.42
JV	372.5-885.96	-37-33	-57-29	-136.63-171.5
TSS	4.38-6.03	-19-10	-24-5	-0.56-0.62

MYP = marketable fruit yield (kg) per plant, TYP = total yield (kg) per plant; MNP = marketable number of fruits per plant; TNP = total number of fruits per plant. fruit FWt = single fruit weight (g), FWd = fruit width (mm), FL = fruit length (cm), PCTk = pericarp thickness (mm), NL = number of locules, DFF = days to 50% flowering, PH = plant height (cm), %TSS =total soluble solids, and JV = juice volume (ml).

6.4 Discussion

Understanding the gene action governing a trait is fundamental for developing effective breeding strategies and methodologies (Fasahat, 2016; Yu *et al*., 2020). The analysis of combining ability serves as a valuable tool in this regard, providing critical insights into the type and magnitude of gene action influencing a trait. In hybrid breeding, where the identification of superior parental lines and hybrids is essential, combining ability studies are particularly important (Zhang *et al*., 2024). This analysis is based on two key concepts: general combining ability (GCA), which reflects additive genetic variance, and specific combining ability (SCA), which encompasses non-additive genetic variance (Fasahat, 2016). The relative importance of GCA and SCA can vary significantly, differing not only for various traits within the same parental line or hybrid but also among different parents or hybrids for the same trait (Liu *et al*., 2021).

The findings of the current combining ability analysis revealed highly significant ($p < 0.01$) or significant ($p < 0.05$) general combining ability (GCA) variances for the majority of the traits studied. Similarly, the variance due to specific combining ability (SCA) was highly significant ($p < 0.01$) for most traits, with the exceptions of fruit length, plant height, number of locules, and fruit width. These findings are partially consistent with previous research; for instance, Izzo *et al*. (2022) and Liu *et al*. (2021) also reported highly significant GCA and SCA variances for a similar set of traits. However, their observation of significant SCA for plant height, fruit weight, fruit length, and number of locules contrasts with our results. Emami *et al*. (2018) likewise documented highly significant GCA and SCA variances for yield per plant, fruit weight, fruit number per plant, and days to 50% flowering.

Conversely, a study by Pavan *et al*. (2022) found that pericarp thickness, number of locules, and plant height did not exhibit significant GCA variance, while they observed highly significant SCA variance for fruit weight, length, width, and yield per plant. These discrepancies are likely attributable to differences in the genetic background of the parental lines used in the various studies. The presence of highly significant GCA and SCA variances in our study indicates that both additive and non-additive gene actions are important in controlling these traits (Fasahat, 2016). However, for most traits, additive genetic variance was predominant, suggesting that early selection in segregating generations would be an effective strategy for their genetic improvement (Bhattarai *et al*., 2016).

The current study determined that GCA variance, representing additive genetic effects, played a significant role in controlling various tomato traits, including TNP, MNP, FWt, FWd, FL, PCTk, PH, NL, and JV. This aligns with other studies that have also identified additive gene action as the primary influence for traits such as fruit weight, fruit length, fruit width, and the number of fruits per plant (Izzo et al., 2022). In contrast to our findings, Pavan et al. (2022) reported that non-additive genetic variance is more prevalent in determining fruit number per plant, pericarp thickness, number of locules, fruit length, and fruit width. Furthermore, the findings of the current study indicated that SCA variance contributed significantly to traits such as DFF, MYP, TYP, and TSS. This is consistent with other studies that have highlighted the crucial role of non-additive genetic variance in governing yield per plant (Bhattarai *et al*., 2016; Emami *et al*., 2018; *Izzo et al*., 2022; Pavan *et al*., 2022) and days to 50% flowering (Izzo *et al*., 2022). However, some reports emphasize the importance of additive gene action for days to 50% flowering (Emami *et al*., 2018). In conclusion, breeding for heterosis would be an effective strategy for significantly improving traits with high SCA variances.

Understanding the extent of GCA and SCA effects in breeding materials is crucial for identifying elite parental lines for subsequent breeding programs (Wang *et al*., 2023; Yu *et al*., 2020). The GCA effects of a parental line reflect its additive and additive x additive genetic value (Fasahat, 2016; Gaballah *et al*., 2022). Significant, favorable GCA effects for a given trait indicate a parent's capacity to transmit favorable alleles to its progeny, adapt to a wide range of environments, and exhibit high potential for genetic improvement of that trait (Fasahat, 2016; Gaballah *et al*., 2022). However, not all crosses between parents with high GCA effects will result in high-performing progeny for the desired trait, as gene interactions may deviate from expectations (Kumar & Bharathi, 2009).

In the present study, the positive and significant GCA effect of female parent G2 and male parental lines G7 and G6 contributed to increased TYP and MYP in the hybrids. Similarly, the high and positive significant GCA effect among female parental lines G2 and G4, and male lines G6, G8, and G9 contributed to increased fruit number per plant in the hybrids. Consequently, hybrids G4 x G9, G4 x G8, and G2 x G9 had a higher number of fruits per plant. Regarding fruit size traits, a combination of lines with the largest favorable GCA effects such as G7 (FL), G10 (FWd), and G7 and G10 (FWt) among male lines, and G2 (FWt), G1 (FWd), and G3 (FL) among female parental lines would produce hybrids with increased fruit weight, length and width. Hybrids G4 x G6 and G5 x G6 in our study produced a small number of locules per fruit. This indicated that crosses of parental lines with low significant negative GCA effect would produce a small number of locules per fruit. Increased plant height could be obtained by a combination of parental lines with high and significant GCA effects, among female line G3 and male lines G6 and G9.

The SCA effect is a decisive factor in selecting the best hybrids among crosses to exploit heterosis or proceed selections from segregating generations (Kumar & Bharathi, 2009). High SCA effects could be obtained from crosses involving parents with (high x high), high x low, low x high, or low x low GCA effects (Fasahat, 2016; Kumar & Bharathi, 2009). In the present finding, highly significant and significant SCA effects were obtained for most of the traits except fruit length, fruit width, days to 50% flowering, and number of locules. Hybrid G2 x G7 showed a significant positive SCA effect; whereas hybrids G3 x G9, G2 x G8, and G4 x G7 showed highly significant negative SCA effects for TYP and MYP. The hybrid G2 x G7 was a (high x low) parental combination, implying the existence of additive x dominance genetic interaction. Hybrids G3 x G6, G3 x G7, G1 x G10, and G4 x G9 were a low x low GCA effect parental combination indicating dominance x dominance genetic interaction. Most crosses with significantly low SCA effects had one of their parents with low GCA effects. A similar trend was observed for the majority of crosses for the traits considered.

Higher heterosis for fruit yield and quality is needed for hybrid tomato commercialization. The results of the study suggest that crosses exhibited superior performance *per se* in comparison to their midparent and best parent across a range of traits. We found crosses with high MPH and BPH for most of the traits which implied the possibility to exploit heterosis. The highest positive heterosis was recorded for TYP (81%, 75%) and MYP (84%, 78%) for MPH and BPH for hybrid G2 x G7. The lowest positive MPH and BPH heterosis was recorded for TSS (10%, 5%).

The presence of heterosis was reported for fruit weight, yield per plant, fruit number per plant, and fruit quality (Ayenan *et al.*, 2022; Basté *et al*., 2010; Ene *et al*., 2023; Fortuny *et al*., 2021; Krieger *et al*., 2010; Shalaby, 2013; Solieman *et a*l., 2013). Emami *et al.*,(2018) found heterosis of 37.75%,

-22.95%, 55,04%, and 3.05% for fruit yield per plant, fruit weight, fruit number per plant, and days to 50% flowering. Liu *et al.*,(2021) also reported MPH of 27.72%, 49.26%, 80.53%, and 42.08% mid-parent heterosis for plant height, fruit weight, fruit number per plant, and total soluble solids.

A high and positive correlation was observed between hybrids performance *per se* and the SCA effect for the majority of the traits. Similarly, the SCA effect showed a strong correlation with MPH and BPH. The cross G2 x G7 had the highest and highly significant SCA effect (0.56), the highest fruit yield per plant (5.02 kg), and the highest MPH (81%) and BPH (75%), demonstrating their strong association with each other.

6.5 Conclusion

This study clarifies the relative importance of additive and non-additive gene actions for key agronomic traits in tomato, providing a foundational knowledge for targeted breeding strategies. The analysis revealed highly significant variances for both general (GCA) and specific (SCA) combining ability, underscoring the roles of both additive and dominance gene effects. Additive genetic effects were predominant for traits such as TNP, MNP, FWt, FWd, FL, PCTk, PH, NL, and JV, whereas non-additive effects were more influential for DFF, MYP, TYP, and TSS. A key contribution of this work is the identification of elite parental combiners. Among the female lines, G2 was the best general combiner for yield and yield-related traits (TYP, MYP, TNP, FWt, and JV). Lines G4 and G3 were the best combiners for fruit number per plant. Among the male lines, G6 was superior for fruit number per plant, number of locules, and plant height; G7 for FWt, FL, PCTk, and JV; and G10 for fruit width, fruit weight, fruit length, pericarp thickness, and JV.

Furthermore, the hybrids G2 x G7, G3 x G7, and G3 x G6 were identified as particularly promising, demonstrating high positive SCA effects, superior performance *per se*, and significant heterosis over both the mid-parent and better parent for critical traits like yield, fruit number, single fruit weight, pericarp thickness, and TSS. This precise elucidation of gene action and the identification of superior parents and specific hybrid combinations provide a direct and valuable roadmap for enhancing tomato productivity and quality in future breeding programs.

CHAPTER SEVEN

7 General conclusions and recommendations

7.1 General conclusions

The phenotypic characterization of 143 tomato germplasm for two seasons revealed a highly significant genetic variation among tomato genotypes based on 15 most quantitative and 18 most qualitative traits. The genotype-environment interaction was also highly significant for the number of locules, marketable fruit number per plant, marketable yield per plant, total fruit number per plant, total yield per plant, fruit weight, and fruit shape index. High to medium broad-sense heritability and genetic advance were detected for most traits. Highly significant positive genetic correlations were found between fruit yield per plant and fruit weight, fruit length, fruit width, fruit shape index, days to 50% flowering, and pericarp thickness.

Multivariate analysis grouped 143 tomato genotypes into three clusters; the first five PCS accounted for 82.86% of the total variation due to fruit weight, fruit length, plant height, number of fruits per cluster, pericarp thickness, marketable and total number of fruits per plant, number of locules, and fruit shape index. Fruit weight, fruit length, fruit width, pericarp thickness, number of locules, and fruit shape index could be targeted to improve fruit yield in tomatoes due to their high heritability, genetic advance, and strong and positive association with yield per plant. Improving tomato productivity could be achievable through the hybridization of tomato genotypes from different clusters while taking the specific traits of each cluster into account. The numerous desirable quantitative and qualitative characteristics observed among the genotypes could be used in tomato breeding programs to use favorable breeding materials from different clusters as parental lines or for pre-breeding activities.

A molecular genetic diversity study using 4729 DArTseq-derived SNPs revealed two distinct genetic groups and admixtures among 187 tomato germplasm. The structure analysis partitioned only cherry tomatoes among the different market types, such as cherry, fresh market, and processing. The grouping was not in line with the pre-defined market groups. A high level of gene flow was detected between the fresh market and processing market types. The considerable genetic variation available within subpopulations, including cherry, fresh market, and processing tomatoes, can be utilized in tomato breeding programs in Ethiopia.

A genome-wide association study using 127 tomato genotypes and 2709 SNP markers identified a total of 121 MTAs ($-\log10p \geq 3.5$; FDR ≤ 0.05) that could be clustered into 104 QTLs based on LD decay (0.157–28.826 Mbp) for each chromosome at $r^2 = 0.2$ for ten traits across two seasons. Some of the QTLs identified for fruit length, pericarp thickness, fruit shape index, fruit weight, fruit width, number of locules, and total soluble solids were located near previously discovered MTAs, while the remaining were new. Across various traits, 16 common MTAs were identified, with marketable fruit number per plant and total number of fruits per plant; and plant height having the highest (9) and lowest (1) MTAs, respectively. No significant MTA was identified for marketable and total yield per plant, days to 50% flowering, and number of fruits per cluster. However, QTLs identified for fruit weight, fruit width, fruit length, pericarp thickness, and number of locules could be employed for fruit yield improvement since they have a positive and highly significant genetic correlation with fruit yield per plant, subject to validation in various genetic backgrounds and environments.

A combining ability study using progenies derived from crosses among five female and male parental lines, selected based on diversity for fruit yield, fruit number per plant, and fruit size,

showed highly significant GCA and SCA variances for most traits, indicating the importance of additive and dominance gene effects in controlling the traits. Additive genetic variances were dominant for marketable and total number of fruits per plant, fruit weight, fruit width, fruit length, pericarp thickness, plant height, number of locules, and juice volume. The non-additive genetic variance was higher for days to 50% flowering, marketable and total yield per plant, and total soluble solids. Female parental line G2 was the best combiner for total yield and marketable yield per plant, number of fruits per plant, fruit weight, and juice volume, and lines G4 and G3 were the best combiners for fruit number per plant. Male parental line G6 was the best combiner for fruit number per plant, number of locules, and plant height. Line G7 was the best combiner for fruit weight and length, pericarp thickness, and juice volume. Line G10 was the best combiner for fruit width, weight and length, pericarp thickness, and juice volume. Hybrids G2 x G7, G3 x G7, and G3 x G6 showed high positive SCA effects, better performance per se, and heterosis over the mid and better parent for yield and number of fruits per plant, single fruit weight, pericarp thickness, and TSS.

7.2 Recommendations

1. Information on the phenotypic and molecular genetic diversity from this study would be employed for varietal development in tomato breeding in Ethiopia.
2. The identified QTLs need to be validated for different genetic backgrounds and in more locations to be employed in marker-assisted breeding in the tomato breeding program.
3. The hybrids need to be tested across locations to identify the best-adapted and stable hybrids for release and production in Ethiopia.
4. Generate comprehensive information on combining ability, heterotic grouping, and heterosis using a large number of parental lines across locations.

5. Strengthening hybrid variety development in Ethiopia would be beneficial since high heterosis is detected for yield, related traits, and fruit quality in the germplasm.

\

REFERENCES

Abewoy Fentik, D. (2017). Review on Genetics and Breeding of Tomato (Lycopersicon esculentum Mill). Advances in Crop Science and Technology, 05(05). https://doi.org/10.4172/2329-8863.1000306

Abhary, M., & Rezk, A. A. (2021b). Advances in Plant Breeding Strategies: Vegetable Crops. In Advances in Plant Breeding Strategies: Vegetable Crops (Issue March 2022). Springer Nature Switzerland AG 2021. https://doi.org/10.1007/978-3-030-66961-4

Adhikari, P., McNellie, J., & Panthee, D. R. (2020). Detection of quantitative trait loci (Qtl) associated with the fruit morphology of tomato. Genes, 11(10), 1–15. https://doi.org/10.3390/genes11101117

Alam, M. A., Rahman, M., Ahmed, S., Jahan, N., Khan, M. A. A., Islam, M. R., Alsuhaibani, A. M., Gaber, A., & Hossain, A. (2022). Genetic Variation and Genotype by Environment Interaction for Agronomic Traits in Maize (Zea mays L.) Hybrids. Plants, 11(11), 1–16. https://doi.org/10.3390/plants11111522

Alam, M. S., Tester, M., Fiene, G., & Mousa, M. A. A. (2021). Early growth stage characterization and the biochemical responses for salinity stress in tomato. Plants, 10(4), 1–20. https://doi.org/10.3390/plants10040712

Ali, M. Y., Sina, A. A. I., Khandker, S. S., Neesa, L., Tanvir, E. M., Kabir, A., Khalil, M. I., & Gan, S. H. (2021). Review nutritional composition and bioactive compounds in tomatoes and their impact on human health and disease: A review. Journal of Food Biochemistry, 2021(1). https://doi.org/10.1155/2021/14119

Albert, E., Segura, V., Gricourt, J., Bonnefoi, J., Derivot, L., & Causse, M. (2016). Association mapping reveals the genetic architecture of tomato response to water deficit: focus on major fruit quality traits. Journal of Experimental Botany, 67(22), 6413–6430. https://doi.org/10.1093/jxb/erw411

Alvarado, G., Rodríguez, F. M., Pacheco, A., Burgueño, J., Crossa, J., Vargas, M., Pérez-Rodríguez, P., & Lopez-Cruz, M. A. (2020). META-R: A software to analyze data from multi-environment plant breeding trials. Crop Journal, 8(5), 745–756. https://doi.org/10.1016/j.cj.2020.03.010

Ariizumi, T., Shinozaki, Y., & Ezura, H. (2013). Genes that influence yield in tomato. Breeding Science, 63(1), 3–13. https://doi.org/10.1270/jsbbs.63.3

Alatawi, I., Xiong, H., Alkabkabi, H., Chiwina, K., Luo, Q., Ling, K.-S., Qu, Y., Du, R., & Shi, A. (2025). Genetic diversity and population structure of tomato (Solanum lycopersicum) from the USDA-GRIN germplasm collection. Agronomy, *15*(1), 22. https://doi.org/10.3390/agronomy15010022

Athinodorou, F., Foukas, P., Tsaniklidis, G., Kotsiras, A., Chrysargyris, A., Delis, C., Kyratzis, A. C., Tzortzakis, N., & Nikoloudakis, N. (2021). Morphological diversity, genetic characterization, and phytochemical assessment of the cypriot tomato germplasm. Plants, 10(8). https://doi.org/10.3390/plants10081698

Awata, L. A. O., Tongoona, P., Danquah, E., Efie, B. E., & Marchelo-Dragga, P. W. (2018). Common Mating Designs in Agricultural Research and Their Reliability in Estimation of Genetic Parameters. 11(7), 16–36. https://doi.org/10.9790/2380-1107021636

Ayenan, M. A. T., Danquah, A., Agre, P. A., Hanson, P., Asante, I. K., & Danquah, E. Y. (2021). Genomic and phenotypic diversity of cultivated and wild tomatoes with varying levels of heat tolerance. Genes, 12(4). https://doi.org/10.3390/genes12040503

Ayenan, M. A. T., Danquah, A., Hanson, P., Asante, I. K., & Danquah, E. Y. (2022). Combining abilities and heterotic patterns for heat tolerant traits in tomatoes (Solanum lycopersicum). Plant Breeding, 141(4), 585–597. https://doi.org/10.1111/pbr.13037

Bai, Y., & Lindhout, P. (2007b). 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

Bahjat, N. M., Yıldız, M., Nadeem, M. A., Morales, A., Wohlfeiler, J., Baloch, F. S., Tunçtürk, M., Koçak, M., Chung, Y. S., Grzebelus, D., Sadik, G., Kuzğun, C., & Cavagnaro, P. F. (2025). Population structure, genetic diversity, and GWAS analyses with GBS-derived SNPs and silicoDArT markers unveil genetic potential for breeding and candidate genes for agronomic and root quality traits in an international sugar beet germplasm collection. BMC Plant Biology, 25(1), 523. https://doi.org/10.1186/s12870-025-06525-7

Banerjee, M. K., & Kalloo, M. K. (1987). Sources and inheritance of resistance to leaf curl virus in Lycopersicon. Theoretical and Applied Genetics, 73(5), 707–710. https://doi.org/10.1007/BF00260780

Barone, A., Chiusano, M. L., Ercolano, M. R., Giuliano, G., Grandillo, S., & Frusciante, L. (2008). Structural and Functional Genomics of Tomato. International Journal of Plant Genomics, 12. https://doi.org/10.1155/2008/820274

Barth, E., de Resende, J. T. V., Mariguele, K. H., de Resende, M. D. V., da Silva, A. L. B. R., & Ru, S. (2022). Multivariate analysis methods improve the selection of strawberry genotypes with low cold requirement. Scientific Reports, 12(1), 1–12. https://doi.org/10.1038/s41598-022-15688-4

Basté, E. M., Liberatti, D. R., Mahuad, S. L., Rodriguez, G. R., Pratta, G. R., Zorzoli, R., & Picardi, L. A. (2010). Diallel analysis for fruit traits among tomato recombinant inbred lines derived from an interspecific cross Solanum lycopersicum x S. Pimpinellifolium. Journal of Applied Horticulture, 12(1), 21–25. https://doi.org/10.37855/jah.2010.v12i01.04

Batley, J., & Edwards, D. (2007). SNP Applications in Plants. Association Mapping in Plants, 95–102. https://doi.org/10.1007/978-0-387-36011-9_6

Bekele, A., & Bekele, E. (2014). Overview: morphological and molecular markers role in crop improvement programs. International Journal of Current Research in Life Sciences, 3(3), 35–42.

Belete, A., Ayalew, G., & Azerefegne, F. (2024). Field efficacy of selected chemical pesticides against two-spotted spider mites, Tetranychus urticae Koch (Acari: Tetranychidae), on

tomato in the Central Rift Valley of Ethiopia. Berhan International Research Journal of Science and Humanities, 8(1), 106–122. https://doi.org/10.61593/DBU.BIRJSH.01.08.06

Benor, S., Zhang, M., Wang, Z., & Zhang, H. (2008). Assessment of genetic variation in tomato (Solanum lycopersicum L .) inbred lines using SSR molecular markers. 35, 373–379.

Bertin, N. (2018). Fruit quality. In Tomatoes. https://doi.org/10.1079/9781780641935.0137

Bertin, N., Causse, M., Brunel, B., Tricon, D., & Génard, M. (2009). Identification of growth processes involved in QTLs for tomato fruit size and composition. Journal of Experimental Botany, 60(1), 237–248. https://doi.org/10.1093/jxb/ern281

Beyene, Y., Botha, A. M., & Myburg, A. A. (2005). A comparative study of molecular and morphological methods of describing genetic relationships in traditional Ethiopian highland maize. African Journal of Biotechnology, 4(7), 586–595. https://doi.org/10.5897/ajb2005.000-3107

Bhanu, A. N. (2017). Assessment of Genetic Diversity in Crop Plants - An Overview. Advances in Plants & Agriculture Research, 7(3), 279–286. https://doi.org/10.15406/apar.2017.07.00255

Bhattarai, U., Sharma, A., Das, R., & Talukdar, P. (2016). Genetic Analysis of Yield and Yield-Attributing Traits for High Temperature Resistance in Tomato. International Journal of Vegetable Science, 22(6), 585–597. https://doi.org/10.1080/19315260.2015.1118421

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. https://doi.org/10.1186/s12864-015-1444-1

Brown, W. L. (1983). Genetic diversity and genetic vulnerability-an appraisal. Economic Botany, 37(1), 4–12. https://doi.org/10.1007/BF02859301

Browning, B. L., Tian, X., Zhou, Y., & Browning, S. R. (2021). Fast two-stage phasing of large-scale sequence data. American Journal of Human Genetics, 108(10), 1880–1890. https://doi.org/10.1016/j.ajhg.2021.08.005

Broz, A. K., Miller, C. M., Baek, Y. S., Tovar-Méndez, A., Acosta-Quezada, P. G., Riofrío-Cuenca, T. E., Rusch, D. B., & Bedinger, P. A. (2021). S-RNase alleles associated with self-compatibility in the tomato clade: Structure, origins, and expression plasticity. Frontiers in Genetics, 12, 780793. https://doi.org/10.3389/fgene.2021.780793

Buckler, E., Casstevens, T., Bradbury, P., & Zhang, Z. (2011). User Manual for TASSEL version 3. 73. www.maizegenetics.net/tassel

Burton, G. W., & DeVane, E. H. (1953). Estimating Heritability in Tall Fescue (Festzjcu Av-undinuceu) from Replicated Clonal Material. Agronomy Journal, 478–481.

Capel, C., Fernández del Carmen, A., Alba, J. M., Lima-Silva, V., Hernández-Gras, F., Salinas, M., Boronat, A., Angosto, T., Botella, M. A., Fernández-Muñoz, R., Granell, A., Capel, J., & Lozano, R. (2015). Wide-genome QTL mapping of fruit quality traits in a tomato RIL population derived from the wild-relative species Solanum pimpinellifolium L. Theoretical and Applied Genetics, 128(10), 2019–2035. https://doi.org/10.1007/s00122-015-2563-4

Cebolla-Cornejo, J., Roselló, S., & Nuez, F. (2013). Phenotypic and genetic diversity of Spanish tomato landraces. Scientia Horticulturae, 162, 150–164. https://doi.org/10.1016/j.scienta.2013.07.044

Celik, I., Gurbuz, N., Uncu, A. T., Frary, A., & Doganlar, S. (2017). Genome-wide SNP discovery and QTL mapping for fruit quality traits in inbred backcross lines (IBLs) of solanum pimpinellifolium using genotyping by sequencing. BMC Genomics, 18(1), 1–10. https://doi.org/10.1186/s12864-016-3406-7

Chaerani, R., Smulders, M. J. M., Van Der Linden, C. G., Vosman, B., Stam, P., & Voorrips, R. E. (2007). QTL identification for early blight resistance (Alternaria solani) in a Solanum lycopersicum × S. arcanum cross. Theoretical and Applied Genetics, 114(3), 439–450. https://doi.org/10.1007/s00122-006-0442-8

Chakrabarti, M., Zhang, N., Sauvage, C., Muños, S., Blanca, J., Cañizares, J., Diez, M. J., Schneider, R., Mazourek, M., McClead, J., Causse, M., & Van Der Knaap, E. (2013). A cytochrome P450 regulates a domestication trait in cultivated tomato. Proceedings of the National Academy of Sciences of the United States of America, 110(42), 17125–17130. https://doi.org/10.1073/pnas.1307313110

Charcosset, A., & Essioux, L. (1994). The effect of population structure on the relationship between heterosis and heterozygosity at marker loci. Theoretical and Applied Genetics, 89(2–3), 336–343. https://doi.org/10.1007/BF00225164

Charcosset, A., Lefort-Buson, M., & Gallais, A. (1991). Relationship between heterosis and heterozygosity at marker loci: a theoretical computation. Theoretical and Applied Genetics, 81(5), 571–575. https://doi.org/10.1007/BF00226720

Chatara, T., Musvosvi, C., Houdegbe, A. C., & Sibiya, J. (2023). Variance Components, Correlation and Path Coefficient Analysis of Morpho-Physiological and Yield Related Traits in Spider Plant (Gynandropsis gynandra (L.) Briq.) under Water-Stress Conditions. Agronomy, 13(3). https://doi.org/10.3390/agronomy13030752

Chaudhary, J., Alisha, A., Bhatt, V., Chandanshive, S., Kumar, N., Sonah, H., Deshmukh, R., Mir, Z., Kumar, A., Yadav, S., & Shivaraj, S. M. S. (2019). Mutation breeding in tomato: Advances, applicability and challenges. Plants, 8(5). https://doi.org/10.3390/plants8050128

Cheniclet, C., Rong, W. Y., Causse, M., Frangne, N., Bolling, L., Carde, J. P., & Renaudin, J. P. (2005). Cell expansion and endoreduplication show a large genetic variability in pericarp and contribute strongly to tomato fruit growth. Plant Physiology, 139(4), 1984–1994. https://doi.org/10.1104/pp.105.068767

Chen, Z., & Zhang, Y. (2024). Research on plant genomics and breeding: 2023. PMC. https://pmc.ncbi.nlm.nih.gov/articles/PMC11548484/

CSA (Central Statistical Agency). (2022). The Federal Democratic Republic of Ethiopia Central Statistical Agency Report on Area and Production of Major Crops. Statistical Bulletin, I, 132.

Dabholkar, A. R. (1999). Elements of biometrical genetics (Rev. and enl. ed.). Concept Publishing Company.

Da Silva, M. J., Pastina, M. M., De Souza, V. F., Schaffert, R. E., Carneiro, P. C. S., Noda, R. W., Carneiro, J. E. de S., Damasceno, C. M. B., & Parrella, R. A. da C. (2017). Phenotypic and molecular characterization of sweet sorghum accessions for bioenergy production. PLoS ONE, 12(8), 1–19. https://doi.org/10.1371/journal.pone.0183504

Daniela Ganeva, D. K. A. N. I. T. S. G. (2020). Tomato Phenotypic Diversity Determined by. Plants, 9(197).

Das, S., Das, S. S., Chakraborty, I., Roy, N., Nath, M. K., & Sarma, D. (2017). Principal component analysis in plant breeding. Biomolecule Reports, August 2018, 1–3. https://www.researchgate.net/publication/327040954

Desalegne, L. (2002). T o m a t o e s Research Experiences and Production Prospects. Ethiopian Agricultural Research Organization.

Dhaliwal, M. S., & Chahal, D. S. (2005). Inheritance of important quality attributes of tomato under low temperature field conditions. Advances in Horticultural Science, 19(3), 158–162. https://doi.org/10.1400/14422

Du, M., Sun, C., Deng, L., Zhou, M., Li, J., Du, Y., Ye, Z., Huang, S., Li, T., Yu, J., Li, C.-B., & Li, C. (2025). Molecular breeding of tomato: Advances and challenges. Journal of Integrative Plant Biology, 67(3), 669–721. https://doi.org/10.1111/jipb.13879

Duvick, D. N. (1999). Chapter 3 Heterosis : Feeding People and Protecting Natural Resources. about 1965, 19–29.

Earl, D. (2012). Structure Harvester: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources, 4(2), 359–361.

Ebert, A. W., & Chou, Y. Y. (2015). The tomato collection maintained by AVRDC - The World Vegetable center: Composition, germplasm dissemination and use in breeding. Acta Horticulturae, 1101(November), 169–176. https://doi.org/10.17660/ActaHortic.2015.1101.26

EBSCO Research Starters. (2023). Genetic diversity. https://www.ebsco.com/research-starters/science/genetic-diversity

Eider, J., Vargas, E., Aguirre, N. C., & Coronado, Y. M. (2020). Study of the genetic diversity of tomato (Solanum spp .). 199–206.

Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. Nature Reviews Genetics, 17(7), 422–433. https://doi.org/10.1038/nrg.2016.58

Emami, S., Nemati, S. H., Azizi, M., & Mobli, M. (2018). Combining ability and gene action of some tomato genotypes under low light condition. Advances in Horticultural Science, 32(4), 459–470. https://doi.org/10.13128/ahs-22479

Ene, C. O., Abtew, W. G., Oselebe, H. O., Ozi, F. U., & Ikeogu, U. N. (2022). Genetic characterization and quantitative trait relationship using multivariate techniques reveal diversity among tomato germplasms. Food Science and Nutrition, 10(7), 2426–2442. https://doi.org/10.1002/fsn3.2850

Ene, C. O., Abtew, W. G., Oselebe, H. O., Ozi, F. U., Ogah, O., Okechukwu, E. C., & Chukwudi, U. P. (2023). Hybrid Vigor and Heritability Estimates in Tomato Crosses Involving Solanum lycopersicum × S. pimpinellifolium under Cool Tropical Monsoon Climate. International Journal of Agronomy, 2023. https://doi.org/10.1155/2023/3003355

Esposito, S., Cardi, T., Campanelli, G., Sestili, S., Díez, M. J., Soler, S., Prohens, J., & Tripodi, P. (2020). ddRAD sequencing-based genotyping for population structure analysis in cultivated tomato provides new insights into the genomic diversity of Mediterranean 'da serbo' type long shelf-life germplasm. Horticulture Research, 7(1). https://doi.org/10.1038/s41438-020-00353-6

Evanno, G., Regnaut, S., & Goudet, J. . (2005). Detecting the number of clusters of individuals using the software STRUCTURE : A simulation study. Molecular Ecology, 14, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

FAOSTAT. (2022). World Food and Agriculture – Statistical Yearbook 2022. In World Food and Agriculture – Statistical Yearbook 2022. https://doi.org/10.4060/cc2211en

Farinon, B., Picarella, M. E., Siligato, F., Rea, R., Taviani, P., & Mazzucato, A. (2022). Phenotypic and Genotypic Diversity of the Tomato Germplasm From the Lazio Region in Central Italy , With a Focus on Landrace Distinctiveness. 13(July), 1–16. https://doi.org/10.3389/fpls.2022.931233

Fasahat, P. (2016). Principles and Utilization of Combining Ability in Plant Breeding. Biometrics & Biostatistics International Journal, 4(1), 1–22. https://doi.org/10.15406/bbij.2016.04.00085

Figàs, M. R., Prohens, J., Raigón, M. D., Fernández-de-Córdova, P., Fita, A., & Soler, S. (2015). Characterization of a collection of local varieties of tomato (Solanum lycopersicum L.) using conventional descriptors and the high-throughput phenomics tool Tomato Analyzer. Genetic Resources and Crop Evolution, 62(2), 189–204. https://doi.org/10.1007/s10722-014-0142-1

Fisseha, Z., Kyallo, M., Tesfaye, K., Harvey, J., Dagne, K., Opyio, S., & Gepts, P. (2018). Integrating phenotypic evaluations with a molecular diversity assessment of an Ethiopian collection of common bean landraces. African Crop Science Journal, 26(2), 315. https://doi.org/10.4314/acsj.v26i2.12

Flint-Garcia, S. A., Thuillet, A. C., Yu, J., Pressoir, G., Romero, S. M., Mitchell, S. E., Doebley, J., Kresovich, S., Goodman, M. M., & Buckler, E. S. (2005). Maize association population: A high-resolution platform for quantitative trait locus dissection. Plant Journal, 44(6), 1054–1064. https://doi.org/10.1111/j.1365-313X.2005.02591.x

FOGLE, H. W., & CURRENCE, T. M. (1950). Inheritance of fruit weight and earliness in a tomato cross. Genetics, 35(3), 363–380. https://doi.org/10.1093/genetics/35.3.363

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

Foolad, M. R., & Panthee, D. R. (2012). Marker-Assisted Selection in Tomato Breeding. Critical Reviews in Plant Sciences, 31(2), 93–123. https://doi.org/10.1080/07352689.2011.616057

Fortuny, A. P., Bueno, R. A., Pereira Da Costa, J. H., Zanor, M. I., & Rodríguez, G. R. (2021). Tomato fruit quality traits and metabolite content are affected by reciprocal crosses and heterosis. Journal of Experimental Botany, 72(15), 5407–5425. https://doi.org/10.1093/jxb/erab222

Franco, J., Crossa, J., Ribaut, J. M., Bertran, J., Warburton, M. L., & Khairallah, M. (2001). A method for combining molecular markers and phenotypic attributes for classifying plant genotypes. Theoretical and Applied Genetics, 103(6–7), 944–952. https://doi.org/10.1007/s001220100641

Frary, A., Nesbitt, T. C., Frary, A., Grandillo, S., Van Der Knaap, E., 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

Fujimoto, R., Uezono, K., Ishikura, W., Osabe, K., Peacock, W. J., & Dennis, E. S. (2018). Recent research on the mechanism of heterosis is important for crop and vegetable breeding systems. Breeding Science, 68(2), 145–158. https://doi.org/10.1270/jsbbs.17155

Gaballah, M. M., Attia, K. A., Ghoneim, A. M., Khan, N., El-Ezz, A. F., Yang, B., Xiao, L., Ibrahim, E. I., & Al-Doss, A. A. (2022). Assessment of Genetic Parameters and Gene Action Associated with Heterosis for Enhancing Yield Characters in Novel Hybrid Rice Parental Lines. Plants, 11(3). https://doi.org/10.3390/plants11030266

Garg, N., Cheema, D. S., & Chawla, N. (2013). Manifestation of heterosis for fruit yield, quality and shelf-life in tomato (Solanum lycopersicum L.) hybrids incorporating rin, nor or alc alleles in main- and late-seasons of north Indian plains. Vegetable Science, *40*(1), 28-33.

Ganal, M. W., Altmann, T., & Röder, M. S. (2009). SNP identification in crop plants. Current Opinion in Plant Biology, 12(2), 211–217. https://doi.org/10.1016/j.pbi.2008.12.009

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., Giovannoni, J. 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

Geng, X., Qu, Y., Jia, Y., He, S., Pan, Z., Wang, L., & Du, X. (2021). Assessment of heterosis based on parental genetic distance estimated with SSR and SNP markers in upland cotton (Gossypium hirsutum L.). BMC Genomics, 22(1), 1–11. https://doi.org/10.1186/s12864-021-07431-6

Getachew Tabor, Yosef Alemu, Selamawit Ketema, M. Y. and G. A. (2021). Vegetable Crops Research in Ethiopia : Achievements and Future Prospects. October, 123–132.

Getahun, A., Alemu, A., Nida, H., et al. (2025). Multi-locus genome-wide association mapping for major agronomic and yield-related traits in sorghum (Sorghum bicolor (L.) Moench) landraces. BMC Genomics, 26, 304. https://doi.org/10.1186/s12864-025-11458-4

Gharsallah, C., Abdelkrim, A. Ben, Fakhfakh, H., Salhi-Hannachi, A., & Gorsane, F. (2016). SSR marker-assisted screening of commercial tomato genotypes under salt stress. Breeding Science, 66(5), 823–830. https://doi.org/10.1270/jsbbs.16112

Ghosh, K. ., Islam, A. ., Mian, M. A. ., & Hossain, M. . (2010). Variability and Character Association in F2 Segregating Population of Different Commercial Hybrids of Tomato (Solanum lycopersicum L.). Journal of Applied Sciences and Environmental Management, 14(2), 7–11. https://doi.org/10.4314/jasem.v14i2.57871

Goffar, M., Ahmed, A., & Halim, G. (2016). Inheritance mechanism of yield and yield components in tomato. Bangladesh Journal of Agricultural Research, 41(2), 335–344. https://doi.org/10.3329/bjar.v41i2.28235

Govindaraj, M., Vetriventhan, M., & Srinivasan, M. (2015). Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its analytical perspectives. Genetics Research International, 2015(Figure 1). https://doi.org/10.1155/2015/431487

Gowda, M., Kling, C., Würschum, T., Liu, W., Maurer, H. P., Hahn, V., & Reif, J. C. (2010). Hybrid breeding in durum wheat: Heterosis and combining ability. Crop Science, 50(6), 2224–2230. https://doi.org/10.2135/cropsci2009.10.0637

Grandillo, S., Ku, H. M., & Tanksley, S. D. (1999). Identifying the loci responsible for natural variation in fruit size and shape in tomato. Theoretical and Applied Genetics, 99(6), 978–987. https://doi.org/10.1007/s001220051405

Grozeva, S., Nankar, A. N., Ganeva, D., Tringovska, I., Pasev, G., & Kostova, D. (2021a). Characterization of tomato accessions for morphological, agronomic, fruit quality, and virus resistance traits. Canadian Journal of Plant Science, 101(4), 476–489. https://doi.org/10.1139/cjps-2020-0030

Grozeva, S., Nankar, A. N., Ganeva, D., Tringovska, I., Pasev, G., & Kostova, D. (2021b). Characterization of tomato accessions for morphological, agronomic, fruit quality, and virus resistance traits. Canadian Journal of Plant Science, 101(4), 476–489. https://doi.org/10.1139/cjps-2020-0030

Gudero, G., Hussien, T., Dejene, M., & Biazin, B. (2018). Integrated management of tomato late blight [Phytophthora infestans (Mont.) de Bary] through host plant resistance and reduced frequency of fungicide in Arbaminch areas, Southern Ethiopia. Journal of Biology, Agriculture and Healthcare, 8(9), 94–103.

Guo, Y., Long, J., He, J., Li, C. I., Cai, Q., Shu, X. O., Zheng, W., & Li, C. (2012). Exome sequencing generates high quality data in non-target regions. BMC Genomics, 13(1). https://doi.org/10.1186/1471-2164-13-194

Guo, Y., Ye, F., Sheng, Q., Clark, T., & Samuels, D. C. (2013). Three-stage quality control strategies for DNA re-sequencing data. Briefings in Bioinformatics, 15(6), 879–889. https://doi.org/10.1093/bib/bbt069

Hanotte, O., & Jianlin, H. (2005). Genetic characterization of livestock populations and its use in conservation decision making. The Role of Biotechnology in Exploring and Protecting Genetic Resources, 131–136.

Harlan, H. V. (1928). The origin of cultivated plants. Journal of Heredity, 19(4), 167–168. https://doi.org/10.1093/oxfordjournals.jhered.a102976

Hassan, Z., Ul-Allah, S., Khan, A. A., Shahzad, U., Khurshid, M., Bakhsh, A., Amin, H., Jahan, M. S., Rehim, A., & Manzoor, Z. (2021). Phenotypic characterization of exotic tomato germplasm: An excellent breeding resource. PLoS ONE, 16(6 June), 1–12. https://doi.org/10.1371/journal.pone.0253557

Hasan, N. (2021). Marker assisted selection approach for improvement of quality traits in tomato. Just Agriculture, 3(9), 154–156.

Hedges, L. J., & Lister, C. E. (2005). Nutritional attributes of tomatoes. In Crop and Food Research (Issue 1391). https://www.researchgate.net/publication/268516162_Nutritional_attributes_of_tomatoes

Henareh, M., Dursun, A., & Mandoulakani, B. A. (2015). Genetic diversity in tomato landraces collected from Turkey and Iran revealed by morphological characters. Acta Scientiarum Polonorum, Hortorum Cultus, 14(2), 87–96.

Henderson, I. R., & Salt, D. E. (2017). Natural genetic variation and hybridization in plants. Journal of Experimental Botany, 68(20), 5415–5417. https://doi.org/10.1093/jxb/erx377

Hernández-Bautista, A., Lobato-Ortiz, R., Cruz-Izquierdo, S., García-Zavala, J. J., Chávez-Servia, J. L., Hernández-Leal, E., & Bonilla-Barrientos, O. (2015). Fruit size QTLs affect in a major proportion the yield in tomato. Chilean Journal of Agricultural Research, 75(4), 402. https://doi.org/10.4067/S0718-58392015000500004

Hochholdinger, F., & Zhang, Y. (2025). Molecular concepts to explain heterosis in crops. Trends in Plant Science, 30(2), 123–136. https://doi.org/10.1016/j.tplants.2024.07.018

Hogenboom, N. G. (1972). Breaking breeding barriers in Lycopersicon. 1. The genus Lycopersicon, its breeding barriers and the importance of breaking these barriers. Euphytica, 21(2), 221–227. https://doi.org/10.1007/BF00036762

Hosmani, P. S., Flores-Gonzalez, M., Geest, H. van de, Maumus, F., Bakker, L. V, Schijlen, E., Haarst, J. van, Cordewener, J., Sanchez-Perez, G., Peters, S., Fei, Z., Giovannoni, J. J., Mueller, L. A., & Saha, S. (2019). An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. BioRxiv, 2012, 767764. https://doi.org/10.1101/767764

Hu, Z., & Xu, S. (2008). A simple method for calculating the statistical power for detecting a QTL located in a marker interval. Heredity, 101(1), 48–52. https://doi.org/10.1038/hdy.2008.25

Huang, M., Liu, X., Zhou, Y., Summers, R. M., & Zhang, Z. (2019). BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. GigaScience, 8(2), 1–12. https://doi.org/10.1093/gigascience/giy154

Huang, Z., & van der Knaap, E. (2011). Tomato fruit weight 11.3 maps close to fasciated on the bottom of chromosome 11. Theoretical and Applied Genetics, 123(3), 465–474. https://doi.org/10.1007/s00122-011-1599-3

Hussain, E., Cheng, C.-Y., Huang, I.-M., et al. (2025). Molecular screening of wild and cultivated tomato germplasm reveals potential materials for multi-locus disease resistance breeding. Genetic Resources and Crop Evolution, 72, 8619–8632. https://doi.org/10.1007/s10722-025-02457-1

Hussain, I., Khan, S. A., Ali, S., Farid, A., Ali, N., Ali, S., Masaud, S., Hussain, I., Azeem, K., & Raza, H. (2018). Genetic diversity among tomato accessions based on agro-morphological traits. Sains Malaysiana, 47(11), 2637–2645. https://doi.org/10.17576/jsm-2018-4711-06

Ibarbia, E. A., & Lambeth, V. N. (1969). Inheritance of Tomato Fruit W eight 1. 10, 498–500.

Ibitoye, D. O., & Akin-Idowu, P. E. (2011). Marker-assisted-selection (MAS): A fast track to increase genetic gain in horticultural crop breeding. African Journal of Biotechnology, 10(55), 11333–11339. https://doi.org/10.5897/AJB10.302

Ichihashi, Y., & Sinha, N. R. (2014). From genome to phenome and back in tomato. Current Opinion in Plant Biology, 18(1), 9–15. https://doi.org/10.1016/j.pbi.2013.12.004

Ikram, S., & Chardon, F. (2010). Plant Quantitative Traits. Encyclopedia of Life Sciences, December. https://doi.org/10.1002/9780470015902.a0002021.pub2

IPGRI. (1992). Descriptors for Tomato.

Izzo, A., Khojah, H., & Murie, A.-M. (2022). Combining ability and heterosis for yield and some fruit traits of tomato. DYSONA - Applied Science, 3(1), 15–23. https://doi.org/10.30493/DAS.2021.295501

Jahani, M., Mohammadi-Nejad, G., Nakhoda, B., & Rieseberg, L. H. (2019). Genetic dissection of epistatic and QTL by environment interaction effects in three bread wheat genetic backgrounds for yield-related traits under saline conditions. Euphytica, 215(6), 1–25. https://doi.org/10.1007/s10681-019-2426-1

James, F. (2022). Genetic Diversity : Importance , Conversation and its Factors. 11(1000225), 10–11. https://doi.org/10.35248/2161-1041.22.11.225

Jáquez-Gutiérrez, M., Atarés, A., Pineda, B., Angarita, P., Ribelles, C., García-Sogo, B., Sánchez-López, J., Capel, C., Yuste-Lisbona, F. J., Lozano, R., & Moreno, V. (2019). Phenotypic and genetic characterization of tomato mutants provides new insights into leaf development and its relationship to agronomic traits. BMC Plant Biology, 19(1), 1–24. https://doi.org/10.1186/s12870-019-1735-9

Jia, M., Ashrafi, H., & Foolad, M. R. (2025). Identification of late blight resistance QTLs in an interspecific RIL population of tomato via genotyping-by-sequencing. Molecular Breeding, 45, 43. https://doi.org/10.1007/s11032-025-01560-6

Jiang, K., Liberatore, K. L., Park, S. J., Alvarez, J. P., & Lippman, Z. B. (2013). Tomato yield heterosis is triggered by a dosage sensitivity of the florigen pathway that fine-tunes shoot architecture. PLoS Genetics, 9(12), e1004043.https://doi.org/10.1371/journal.pgen.1004043

John F. Doebley, Brandon S. Gaut, & Smith, and B. D. (2006). The Molecular Genetics of Crop Domestication. Cell, 127. https://doi.org/10.1080/03081079.2015.1106730

Johnson, H. W., Robinson, H. F., & Comstock, R. E. (1955a). Estimates of Genetic and Environmental Variability in Soybeans 1. Agronomy Journal, 47(7), 314–318. https://doi.org/10.2134/agronj1955.00021962004700070009x

Joshi, B. K., Davis, J., & Panthee, D. R. (2011). How we Improve and Develop Tomato Cultivars How we Improve and Develop Tomato Cultivars. June. https://doi.org/10.13140/RG.2.2.12728.96001

Kaeppler, S. (2012). Heterosis: Many Genes, Many Mechanisms—End the Search for an Undiscovered Unifying Theory. ISRN Botany, 2012, 1–12. https://doi.org/10.5402/2012/682824

Kaler, A. S., Gillman, J. D., Beissinger, T., & Purcell, L. C. (2020). Comparing Different Statistical Models and Multiple Testing Corrections for Association Mapping in Soybean and Maize. Frontiers in Plant Science, 10(February), 1–13. https://doi.org/10.3389/fpls.2019.01794

Kandel, D. R., Bedre, R. H., Mandadi, K. K., Crosby, K., & Avila, C. A. (2019). Genetic Diversity and Population Structure of Tomato (<i>Solanum lycopersicum</i>) Germplasm Developed by Texas A&M Breeding Programs. American Journal of Plant Sciences, 10(07), 1154–1180. https://doi.org/10.4236/ajps.2019.107083

Kassambara, A. (2017). Practical Guide To Cluster Analysis in R (real version). Sthda, 187. https://www.datanovia.com/en/wp-content/uploads/dn-tutorials/book-preview/clustering_en_preview.pdf

Kaul, M. L. H. (1991). Reproductive Biology in Tomato. 39–50. https://doi.org/10.1007/978-3-642-84275-7_4

Kaunda, J. S., & Zhang, Y. J. (2019). The Genus Solanum: An Ethnopharmacological, Phytochemical and Biological Properties Review. In Natural Products and Bioprospecting (Vol. 9, Issue 2). Springer Singapore. https://doi.org/10.1007/s13659-019-0201-6

Kaushal, A., Singh, A., & Jeena, A. S. (2017). Genetic diversity in tomato (Solanum lycopersicum L.) genotypes revealed by simple sequence repeats (SSR) markers. Journal of Applied and Natural Science, 9(2), 966–973. https://doi.org/10.31018/jans.v9i2.1305

Kawicha, P., Tongyoo, P., Wongpakdee, S., Rattanapolsan, L., Duangjit, J., Chunwongse, J., Suwor, P., Sangdee, A., & Thanyasiriwat, T. (2023). Genome-wide association study revealed genetic loci for resistance to fusarium wilt in tomato germplasm. Crop Breeding and Applied Biotechnology, 23(1), 1–9. https://doi.org/10.1590/1984-70332023v23n1a1

Kearsey, M. J. (1998). The principles of QTL analysis (a minimal mathematics approach). Journal of Experimental Botany, 49(327), 1619–1623. https://doi.org/10.1093/jxb/49.327.1619

Kesho, A., & Tadesse, Y. (2023). Fruit yield loss assessment of tomato (Solanum lycopersicum Mill.) due to late blight (Phytophthora infestans) disease in central parts of Ethiopia. Middle East Research Journal of Biological Sciences, 3(1). https://doi.org/10.36348/merjbs.2023.v03i01.003

Ketema, S., Geleto, J., Alemu, Y., Wondimu, G., & Hinsermu, M. (2017). Yield Stability and Quality Performance of Processing Tomato (Lycopersicon Esculentum Mill) Varieties in the Central Rift Valley of Ethiopia. 4(12), 11–15.

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., Caig, V., Heller-Uszynska, K., Jaccoud, D., Hopper, C., Aschenbrenner-Kilian, M., Evers, M., Peng, K., Cayla, C., Hok, P., & Uszynski, G. (2012). Diversity arrays technology: A generic genome profiling technology on open platforms. Methods in Molecular Biology, 888, 67–89. https://doi.org/10.1007/978-1-61779-870-2_5

Kim, M., Nguyen, T. T. P., Ahn, J. H., Kim, G. J., & Sim, S. C. (2021). Genome-wide association study identifies QTL for eight fruit traits in cultivated tomato (Solanum lycopersicum L.). Horticulture Research, 8(1). https://doi.org/10.1038/s41438-021-00638-4

Kıymacı, G., Uncu, A. Ö., & Türkmen, Ö. (2024). Description of the phenotypic characteristics of some tomato genotypes. Selcuk Journal of Agriculture and Food Sciences, 38(1), 9–26. https://doi.org/10.15316/SJAFS.2024.002

Korir, N. K., Diao, W., Tao, R., Li, X., Kayesh, E., Li, A., Zhen, W., & Wang, S. (2014). Genetic diversity and relationships among different tomato varieties revealed by EST-SSR markers. Genetics and Molecular Research, 13(1), 43–53. https://doi.org/10.4238/2014.January.8.3

Kovach, K. E., & Egertsdotter, E. M. U. (2009). Assessment of Genetic Variation of.

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. https://doi.org/10.1038/ng.550

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), 1547–1549. https://doi.org/10.1093/molbev/msy096

Kwon, Y. S., Park, S. G., & Yi, S. I. (2009). Assessment of genetic variation among commercial tomato (Solanum lycopersicum L) varieties using SSR Markers and morphological characteristics. Genes and Genomics, 31(1), 1–10. https://doi.org/10.1007/BF03191132

Labate, J. A. (2021). DNA variation in a diversity panel of tomato genetic resources. Journal of the American Society for Horticultural Science, 146(5), 339–345. https://doi.org/10.21273/JASHS05066-21

Le Van Trong, L., Le Thi Lam, L., & Lai Thi Thanh, L. (2024). Research on nutritional composition of tomatoes fruit (Solanum lycopersicum L.) grown in Vietnam. Middle East Research Journal of Agriculture and Food Science, 4(2). https://doi.org/10.36348/merjafs.2024.v04i02.004

Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q., Zheng, Z., Zhang, Z., Lun, Y., Li, S., Wang, X., Huang, Z., Li, J., Zhang, C., Wang, T., Zhang, Y., Wang, A., Zhang, Y., Lin, K., Li, C., ... Huang, S. (2014). Genomic analyses provide insights into the history of tomato breeding. Nature Genetics, 46(11), 1220–1226. https://doi.org/10.1038/ng.3117

Liu, K., & Muse, S. V. (2005). PowerMaker: An integrated analysis environment for genetic maker analysis. Bioinformatics, 21(9), 2128–2129. https://doi.org/10.1093/bioinformatics/bti282

Liu, X., Geng, X., Zhang, H., Shen, H., & Yang, W. (2017). Association and genetic identification of loci for four fruit traits in tomato using InDel markers. Frontiers in Plant Science, 8(July), 1–14. https://doi.org/10.3389/fpls.2017.01269

Liu, Z., Jiang, J., & Li, J. (2023). GWAS Characterization of the Genetic Regions Associated with Nine Important Agronomic Traits in Tomato (Solanum lycopersicum L.). Agronomy, 13(5), 1–18. https://doi.org/10.3390/agronomy13051191

Liu, Z., Jiang, J., Ren, A., Xu, X., Zhang, H., Zhao, T., Jiang, X., Sun, Y., Li, J., & Yang, H. (2021). Heterosis and combining ability analysis of fruit yield, early maturity, and quality in tomato. Agronomy, 11(4). https://doi.org/10.3390/AGRONOMY11040807

Lippman, Z., & Tanksley, S. D. (2001). Dissecting the genetic pathway to extreme fruit size in tomato using a cross between the small-fruited wild species Lycopersicon pimpinellifolium and L. esculentum var. Giant Heirloom. Genetics, 158(1), 413–422. https://doi.org/10.1093/genetics/158.1.413

Luisa, M., Sardaro, S., Marmiroli, M., Maestri, E., & Marmiroli, N. (2012). Genetic characterization of Italian tomato varieties and their traceability in tomato food products. https://doi.org/10.1002/fsn3.8

Lule, Z., Yeshitila, M., Bayissa, M., & Mitiku, F. (2024). Tomato value chain analysis in Seka Chekorsa District, Jimma Zone, Oromia region, Ethiopia. Cogent Food & Agriculture, 10(1). https://doi.org/10.1080/23311932.2024.2398859

Macherey-Nagel. (2014). Genomic DNA from plant - User manual (NucleoSpin® Plant II, -Midi, -Maxi). July, 36.

Marefatzadeh-Khameneh, M., Fabriki-Ourang, S., Sorkhilalehloo, B., Abbasi-Kohpalekani, J., & Ahmadi, J. (2021). Genetic diversity in tomato (Solanum lycopersicum L.) germplasm using fruit variation implemented by tomato analyzer software based on high throughput phenotyping. Genetic Resources and Crop Evolution, 68(6), 2611–2625. https://doi.org/10.1007/s10722-021-01153-0

Mata-Nicolás, E., Montero-Pau, J., Gimeno-Paez, E., Garcia-Carpintero, V., Ziarsolo, P., Menda, N., Mueller, L. A., Blanca, J., Cañizares, J., van der Knaap, E., & Díez, M. J. (2020a). Exploiting the diversity of tomato: the development of a phenotypically and genetically detailed germplasm collection. Horticulture Research, 7(1). https://doi.org/10.1038/s41438-020-0291-7

McVean, G. (2009). A genealogical interpretation of principal components analysis. PLoS Genetics, 5(10). https://doi.org/10.1371/journal.pgen.1000686

Melchinger, A. E., Utz, H. F., Piepho, H.-P., Zeng, Z.-B., & Schön, C. C. (2007). The role of epistasis in the manifestation of heterosis: A systems-oriented approach. Genetics, 177(3), 1815–1825. https://doi.org/10.1534/genetics.107.077537

Mersha, D., & Sime, M. (2022). Analysis of Tomato Value Chain in Central Rift Valley of Ethiopia (Issue October).

Mohammadi, S. A., & Prasanna, B. M. (2003). Analysis of genetic diversity in crop plants - Salient statistical tools and considerations. Crop Science, 43(4), 1235–1248. https://doi.org/10.2135/cropsci2003.1235

Molecular Ecology & Evolution. (2024). Genetic diversity. Wheaton College Open Press. https://openpress.wheatoncollege.edu/molecularecologyv1/chapter/genetic-diversity/

Moll, R. H., Lonnquist, J. H., & Fortuno, J. W. L. E. Z. (1965). THE RELATIONSHIP O F HETEROSIS AND GENETIC DIVERGENCE IN MAIZE. 1908, 139–144.

Monamodi, E. L., Lungu, D. M., & Fite, G. L. (2013). Analysis of fruit yield and its components in determinate tomato (Lycopersicon lycopersci) using correlation and path coefficient. Bots. J. Agric. Appl. Sci, 9(1), 29–40. http://www.ub.bw/ojs/index.php/bojaas/article/viewFile/185/65

Mondini, L., Noorani, A., & Pagnotta, M. A. (2009). Assessing plant genetic diversity by molecular tools. Diversity, 1(1), 19–35. https://doi.org/10.3390/d1010019

Mu, Q., Huang, Z., Chakrabarti, M., Illa-Berenguer, E., Liu, X., Wang, Y., Ramos, A., & van der Knaap, E. (2017). Fruit weight is controlled by Cell Size Regulator encoding a novel protein that is expressed in maturing tomato fruits. PLoS Genetics, 13(8), 1–26. https://doi.org/10.1371/journal.pgen.1006930

Mungai, G. W., Owino, W., Ambuko, J., Giovannoni, J. J., Nyende, A. B., & Michuki, G. (2021). Genetic diversity within African tomato using next generation sequencing. 16(2018), 296–305. https://doi.org/10.1017/S1479262117000314

Muños, S., Ranc, N., Botton, E., Bérard, A., Rolland, S., Duffé, P., Carretero, Y., Paslier, M. C. Le, Delalande, C., Bouzayen, M., Brunel, D., & Causse, M. (2011). Increase in tomato locule number is controlled by two single-nucleotide polymorphisms located near WUSCHEL. Plant Physiology, 156(4), 2244–2254. https://doi.org/10.1104/pp.111.173997

Muthoni, J., & Shimelis, H. (2020). Mating designs commonly used in plant breeding: A review. Australian Journal of Crop Science, 14(12), 1855–1969. https://doi.org/10.21475/ajcs.20.14.12.p2588

Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., Hatipoğlu, R., Ahmad, F., Alsaleh, A., Labhane, N., Özkan, H., Chung, G., & Baloch, F. S. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnology and Biotechnological Equipment, 32(2), 261–285. https://doi.org/10.1080/13102818.2017.1400401

Neyhart, J. L., Lorenz, A. J., & Smith, K. P. (2019). Multi-trait Improvement by Predicting Genetic Correlations in Breeding Crosses. G3: Genes, Genomes, Genetics, 9(10), 3153–3165. https://doi.org/10.1534/g3.119.400406

Nyaga, C., Gowda, M., Beyene, Y., Murithi, W. T., Burgueno, J., Toledo, F., Makumbi, D., Olsen, M. S., Das, B., Suresh, L. M., Bright, J. M., & Prasanna, B. M. (2020). Combining Ability , and Hybrid Prediction. Plants, 9(468), 1–15.

Oakey, H., Verbyla, A. P., Cullis, B. R., Wei, X., & Pitchford, W. S. (2007). Joint modeling of additive and non-additive (genetic line) effects in multi-environment trials. Theoretical and Applied Genetics, 114(8), 1319–1332. https://doi.org/10.1007/s00122-007-0515-3

Osei, M., Prempeh, R., Adjebeng-Danquah, J., A. Opoku, J., Danquah, A., Danquah, E., Blay, E., & Adu-Dapaah, H. (2019). Marker-Assisted Selection (MAS): A Fast-Track Tool in Tomato Breeding. Recent Advances in Tomato Breeding and Production. https://doi.org/10.5772/intechopen.76007

Opeña, R. T., Chen, J. T., Kalb, T., & Hanson, P. (2001). Hybrid Seed Production in Tomato. AVRDC. International Cooperator´s Guide, 1(527), 1–8.

Pailles, Y., Ho, S., Pires, I. S., Tester, M., Negrão, S., & Schmöcke, S. M. (2017). Genetic diversity and population structure of two tomato species from the galapagos islands. Frontiers in Plant Science, 8(February), 1–11. https://doi.org/10.3389/fpls.2017.00138

Paril, J. (2024). Heterosis in crop improvement. The Plant Journal. https://doi.org/10.1111/tpj.16488

Pavan, M. P., Gangaprasad, S., Dushyanthakumar, B. M., Adivappar, N., & Shashikumara, P. (2022). Heterosis and combining ability studies by line × tester analysis for fruit biochemical, morpho-physiological, and yield traits governing shelf life in tomato (Solanum lycopersicum L.). Euphytica, 218(7). https://doi.org/10.1007/s10681-022-03038-4

Peakall, R., & Smouse, P. E. (2012). GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics, 28(19), 2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Peralta, I., & Spooner, D. (2000). Classification of wild tomatoes: a review. Kurtziana, 28(I), 45–54. http://agris.fao.org/agris-search/search.do?recordID=US201300799648

Perrier, X., & Jacquemoud-Collet, J. P. (2020). DARwin software: Dissimilarity analysis and representation for windows. Website Http://Darwin. Cirad. Fr/Darwin [Accessed 1 March 2013], 115. https://darwin.cirad.fr/

Pessoa, H. P., Dariva, F. D., Copati, M. G. F., de Paula, R. G., Dias, F. O., & Gomes, C. N. (2023). Uncovering tomato candidate genes associated with drought tolerance using Solanum pennellii introgression lines. PLoS ONE, 18(6), e0287178. https://doi.org/10.1371/journal.pone.0287178

Phan, N. T., Trinh, L. T., Rho, M. Y., Park, T. S., Kim, O. R., Zhao, J., Kim, H. M., & Sim, S. C. (2019a). Identification of loci associated with fruit traits using genome-wide single nucleotide polymorphisms in a core collection of tomato (Solanum lycopersicum L.). Scientia Horticulturae, 243(September 2018), 567–574. https://doi.org/10.1016/j.scienta.2018.09.003

Phan, N. T., Trinh, L. T., Rho, M. Y., Park, T. S., Kim, O. R., Zhao, J., Kim, H. M., & Sim, S. C. (2019b). Identification of loci associated with fruit traits using genome-wide single nucleotide polymorphisms in a core collection of tomato (Solanum lycopersicum L.).

Scientia Horticulturae, 243(May 2018), 567–574. https://doi.org/10.1016/j.scienta.2018.09.003

Piepho, H. P. (2000). Optimal marker density for interval mapping in a backcross population. Heredity, 84(4), 437–440. https://doi.org/10.1046/j.1365-2540.2000.00678.x

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., & Lareu, M. V. (2013). An overview of STRUCTURE: Applications, parameter settings, and supporting software. Frontiers in Genetics, 4(MAY), 1–13. https://doi.org/10.3389/fgene.2013.00098

Price, A. H. (2006). Believe it or not, QTLs are accurate! Trends in Plant Science, 11(5), 213–216. https://doi.org/10.1016/j.tplants.2006.03.006

Quinet, M., Angosto, T., Yuste-Lisbona, F. J., Blanchard-Gros, R., Bigot, S., Martinez, J. P., & Lutts, S. (2019). Tomato Fruit Development and Metabolism. Frontiers in Plant Science, 10(November), 1–23. https://doi.org/10.3389/fpls.2019.01554

Raghavan. (2012). Effect of small mapping population sizes on reliability of quantitative trait locus (QTL) mapping. African Journal of Biotechnology, 4(47), 10661–10674. https://doi.org/10.5897/ajb11.2032

Rambla, J. L., Medina, A., Fernández-Del-Carmen, A., Barrantes, W., Grandillo, S., Cammareri, M., López-Casado, G., Rodrigo, G., Alonso, A., García-Martínez, S., Primo, J., Ruiz, J. J., Fernández-Muñoz, R., Monforte, A. J., & Granell, A. (2017). Identification, introgression, and validation of fruit volatile QTLs from a red-fruited wild tomato species. Journal of Experimental Botany, 68(3), 429–442. https://doi.org/10.1093/jxb/erw455

Ramírez-Ojeda, G., Peralta, I. E., Rodríguez-Guzmán, E., Sahagún-Castellanos, J., Chávez-Servia, J. L., Medina-Hinostroza, T. C., Rijalba-Vela, J. R., Vásquez-Núñez, L. P., & Rodríguez-Pérez, J. E. (2021). Edaphoclimatic Descriptors of Wild Tomato Species (Solanum Sect. Lycopersicon) and Closely Related Species (Solanum Sect. Juglandifolia and Sect. Lycopersicoides) in South America. Frontiers in Genetics, 12(November), 1–16. https://doi.org/10.3389/fgene.2021.748979

Rasheed, A., Ilyas, M., Khan, T. N., Mahmood, A., Riaz, U., Chattha, M. B., Al Kashgry, N. A. T., Binothman, N., Hassan, M. U., Wu, Z., & Qari, S. H. (2023). Study of genetic variability, heritability, and genetic advance for yield-related traits in tomato (Solanum lycopersicon MILL.). Frontiers in Genetics, 13(January), 1–13. https://doi.org/10.3389/fgene.2022.1030309

Rasheed1, A. (2017). Genetic Association and path coefficient analysis among yield and yield related traits in Tomato (Solanum lycopersicon MILL.). International Journal of Biosciences (IJB), 11(5), 21–26. https://doi.org/10.12692/ijb/11.5.21-26

Rashidi, M., Gholami, M., & Khabbaz, B. G. (2009). Response of yield and yield components of tomato (Lycopersicon esculentum) to different tillage methods. International Journal of Agriculture and Biology, 11(5), 626–630.

Razifard, H., Ramos, A., Della Valle, A. L., Bodary, C., Goetz, E., Manser, E. J., Li, X., Zhang, L., Visa, S., Tieman, D., Van Der Knaap, E., & Caicedo, A. L. (2020a). Genomic evidence

for complex domestication history of the cultivated tomato in Latin America. Molecular Biology and Evolution, 37(4), 1118–1132. https://doi.org/10.1093/molbev/msz297

Reif, J. C., Zhao, Y., Würschum, T., Gowda, M., & Hahn, V. (2013). Genomic prediction of sunflower hybrid performance. Plant Breeding, 132(1), 107–114. https://doi.org/10.1111/pbr.12007

Renquist, A. R., & Reid, J. B. (1998). Quality of processing tomato (Lycoperscion esculentum) fruit from four bloom dates in relation to optimal harvest timing. New Zealand Journal of Crop and Horticultural Science, 26(2), 161–168. https://doi.org/10.1080/01140671.1998.9514052

Rick, C. M., Fobes, J. F., & Holle, M. (1977). Genetic variation in Lycopersicon pimpinellifolium: Evidence of evolutionary change in mating systems. Plant Systematics and Evolution, 127(2–3), 139–170. https://doi.org/10.1007/BF00984147

Riedelsheimer, C., Czedik-Eysenberg, A., Grieder, C., Lisec, J., Technow, F., Sulpice, R., Altmann, T., Stitt, M., Willmitzer, L., & Melchinger, A. E. (2012). Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nature Genetics, 44(2), 217–220. https://doi.org/10.1038/ng.1033

Robbins, M. D., Sim, S. C., Yang, W., Van Deynze, A., Van Der Knaap, E., Joobeur, T., & Francis, D. M. (2011). Mapping and linkage disequilibrium analysis with a genome-wide collection of SNPs that detect polymorphism in cultivated tomato. Journal of Experimental Botany, 62(6), 1831–1845. https://doi.org/10.1093/jxb/erq367

Robinson, H. F., Comstock, R. E., & Harvey, P. H. (1949). Heritability and the Degree of dominance in Corn. Agronomy Journal, 8(41), 353–359.

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

Ruggieri, V., Calafiore, R., Schettini, C., Rigano, M. M., Olivieri, F., Frusciante, L., & Barone, A. (2019). Exploiting genetic and genomic resources to enhance heat-tolerance in tomatoes. Agronomy, 9(1), 1–13. https://doi.org/10.3390/agronomy9010022

Ruggieri, V., Francese, G., Sacco, A., D'Alessandro, A., Rigano, M. M., Parisi, M., Milone, M., Cardi, T., Mennella, G., & Barone, A. (2014). An association mapping approach to identify favourable alleles for tomato fruit quality breeding. BMC Plant Biology, 14(1), 1–15. https://doi.org/10.1186/s12870-014-0337-9

Sacco, A., Ruggieri, V., Parisi, M., Festa, G., Rigano, M. M., Picarella, M. E., Mazzucato, A., & Barone, A. (2015). Exploring a tomato landraces collection for fruit-related traits by the aid of a high-throughput genomic platform. PLoS ONE, 10(9), 1–20. https://doi.org/10.1371/journal.pone.0137139

Sahu, K. K., & Chattopadhyay, D. (2017). Genome-wide sequence variations between wild and cultivated tomato species revisited by whole genome sequence mapping. BMC Genomics, 18(1), 1–10. https://doi.org/10.1186/s12864-017-3822-3

Said, A. A. (2014). Generation mean analysis in wheat (Triticum aestivum L (Vol. 59, pp. 177–184).

Salgotra, R. K., & Chauhan, B. S. (2023). Genetic Diversity, Conservation, and Utilization of Plant Genetic Resources. Genes, 14(1). https://doi.org/10.3390/genes14010174

Salim, M. M. R., Rashid, M. H., Hossain, M. M., & Zakaria, M. (2020). Morphological characterization of tomato (Solanum lycopersicum L.) genotypes. Journal of the Saudi Society of Agricultural Sciences, 19(3), 233–240. https://doi.org/10.1016/j.jssas.2018.11.001

Sandhu, N., Ankush, A. P., Singh, J., et al. (2023). Integrating association mapping, linkage mapping, fine mapping with RNA-seq conferring seedling vigor improvement for successful crop establishment in deep sown direct-seeded rice. Rice, 16, 46. https://doi.org/10.1186/s12284-023-00665-w

Sang, Z., Wang, H., Yang, Y., Zhang, Z., Liu, X., Li, Z., & Xu, Y. (2022). Epistasis activation contributes substantially to heterosis in temperate by tropical maize hybrids. Frontiers in Plant Science, 13, 921608. https://doi.org/10.3389/fpls.2022.921608

Sansaloni, C. P., Petroli, C. D., Carling, J., Hudson, C. J., Steane, D. A., Myburg, A. A., Grattapaglia, D., Vaillancourt, R. E., & Kilian, A. (2010). A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in Eucalyptus. Plant Methods, 6(1), 16. https://doi.org/10.1186/1746-4811-6-16

Sauvage, C., Segura, V., Bauchet, G., Stevens, R., Do, P. T., Nikoloski, Z., Fernie, A. R., & Causse, M. (2014). Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. Plant Physiology, 165(3), 1120–1132. https://doi.org/10.1104/pp.114.241521

Schreinemachers, P., Simmons, E. B., & Wopereis, M. C. S. (2018). Tapping the economic and nutritional power of vegetables. Global Food Security, 16(June 2017), 36–45. https://doi.org/10.1016/j.gfs.2017.09.005

Semagn, K., & Asmund Bjørnstad, Y. (2010). The genetic dissection of quantitative traits in crops. Electronic Journal of Biotechnology, 13(5), 1–45. https://doi.org/10.2225/vol13-issue5-fulltext-21

Senthil Kumar, P., & Bharathi, P. (2009). Studies on relationship between gca and sca effects in maize (Zea mays L.). Electronic Journal of Plant Breeding, 1(1), 24–27.

Shah, L. R., Ahmed, N., Hussain, K., Mansoor, S., Khan, T., Khan, I., Narayan, S., Afroza, B., Murtaza, I., Shikari, A. B., Bhat, B., & Masoodi, K. Z. (2024). Mapping phenotypic performance and novel SNPs for cold tolerance in tomato (Solanum lycopersicum) genotypes through GWAS and population genetics. BMC Genomic Data, 25(1), 1–23. https://doi.org/10.1186/s12863-024-01190-5

Shalaby, T. A. (2013). Mode of gene action, heterosis and inbreeding depression for yield and its components in tomato (Solanum lycopersicum L.). Scientia Horticulturae, 164, 540–543. https://doi.org/10.1016/j.scienta.2013.10.006

Sharma, H. R., Sharma, D., & Thakur, A. K. (2006). Analysis of genetic divergence in tomato (Lycopersicon esculentum Mill.). Journal of Horticultural Sciences, 1(1), 52–54.

Sharma, P., Thakur, S., & Negi, R. (2019). Recent Advances in Breeding of Tomato- A Review. International Journal of Current Microbiology and Applied Sciences, 8(03), 1275–1283. https://doi.org/10.20546/ijcmas.2019.803.151

Shiberu, T., & Getu, E. (2017). Experimental analysis of economic action level of tomato leafminer, Tuta absoluta Meyrick (Lepidoptera: Gelechiidae) on tomato plant under open field. Advances in Crop Science and Technology, 5(6). https://doi.org/10.4172/2329-8863.1000306

Shushay, C., Derbew, B., & Fetien, A. (2014). Genetic diversity studies for quantitative traits of tomato (Solanum lycopersicon L.) genotypes in Western Tigray, Northern Ethiopia. Journal of Plant Breeding and Crop Science, 6(9), 105–113. https://doi.org/10.5897/jpbcs2014.0470

Sim, S. C., Robbins, M. D., Van Deynze, A., Michel, A. P., & Francis, D. M. (2011). Population structure and genetic differentiation associated with breeding history and selection in tomato (Solanum lycopersicum L.). Heredity, 106(6), 927–935. https://doi.org/10.1038/hdy.2010.139

Sim, S. C., van Deynze, A., Stoffel, K., Douches, D. S., Zarka, D., Ganal, M. W., Chetelat, R. T., Hutton, S. F., Scott, J. W., Gardner, R. G., Panthee, D. R., Mutschler, M., Myers, J. R., & Francis, D. M. (2012). High-Density SNP Genotyping of Tomato (Solanum lycopersicum L.) Reveals Patterns of Genetic Variation Due to Breeding. PLoS ONE, 7(9), 1–18. https://doi.org/10.1371/journal.pone.0045520

Simko, I., Haynes, K. G., & Jones, R. W. (2006). Assessment of linkage disequilibrium in potato genome with single nucleotide polymorphism markers. Genetics, 173(4), 2237–2245. https://doi.org/10.1534/genetics.106.060905

Sims, W. L. (1980). History of Tomato Production for Industry Around the World. In Acta Horticulturae (Issue 100, pp. 25–26). https://doi.org/10.17660/actahortic.1980.100.1

Singh, Major & Gujjar, Ranjit & Karkute, Suhas & Prasanna, H & Tiwari, S. K. (2016). Biology of Solanum lycopersicum (Tomato). April. https://www.researchgate.net/publication/336319341_Biology_of_Solanum_lycopersicum_Tomato

Smith, O. S. (1986). Covariance between Line per se and Testcross Performance 1 . Crop Science, 26(3), 540–543. https://doi.org/10.2135/cropsci1986.0011183x002600030023x

Solieman, T. H. I., El-Gabry, M. A. H., & Abido, A. I. (2013). Heterosis, potence ratio and correlation of some important characters in tomato (Solanum lycopersicum L.). Scientia Horticulturae, 150, 25–30. https://doi.org/10.1016/j.scienta.2012.10.024

Solomon, A. M., Han, K., Lee, J. H., Lee, H. Y., Jang, S., & Kang, B. C. (2019). Genetic diversity and population structure of Ethiopian Capsicum germplasms. PLoS ONE, 14(5), 1–20. https://doi.org/10.1371/journal.pone.0216886

Song, L., Wang, R., Yang, X., Zhang, A., & Liu, D. (2023). Molecular Markers and Their Applications in Marker-Assisted Selection (MAS) in Bread Wheat (Triticum aestivum L.). Agriculture (Switzerland), 13(3). https://doi.org/10.3390/agriculture13030642

Soriano, J. M. (2020). Molecular marker technology for crop improvement. Agronomy, 10(10). https://doi.org/10.3390/agronomy10101462

Souza, L. M. de, Melo, ;Paulo César T, Luders, ; Reginaldo R, & Melo, ; Arlete MT. (2012). Correlations between yield and fruit quality characteristics of fresh market tomatoes. Horticultura Brasileira, 30(4), 627–631.

Sprague, G. F., & Tatum, L. A. (1942). General vs. Specific Combining Ability in Single Crosses of Corn 1 . Agronomy Journal, 34(10), 923–932. https://doi.org/10.2134/agronj1942.00021962003400100008x

Stange, M., Friedrich Utz, H., Schrag, T. A., Melchinger, A. E., & Würschum, T. (2013). High-density genotyping: An overkill for QTL mapping? lessons learned from a case study in maize and simulations. Theoretical and Applied Genetics, 126(10), 2563–2574. https://doi.org/10.1007/s00122-013-2155-0

Sun, S., Yuan, Y., Xu, M., Liu, Z., Yuan, X., Li, X., Li, R., & Cui, X. (2023). TKN3 affects cell expansion to regulate fruit development in tomato. Horticultural Plant Journal, 9(?), ?–?. https://doi.org/10.1016/j.hpj.2022.11.013

Sushma, K., Saidaiah, P., Sudini, H., Geetha, A., & Reddy, K. R. (2021). Study of qualitative traits of germplasm of tomato (Solanum lycopersicum L .). Journal of Pharmacognosy and Phytochemistry, 10(1), 2050–2054.

Tamisier, L., Szadkowski, M., Nemouchi, G., Lefebvre, V., Szadkowski, E., Duboscq, R., Santoni, S., Sarah, G., Sauvage, C., Palloix, A., & Moury, B. (2020). Genome-wide association mapping of QTLs implied in potato virus Y population sizes in pepper: evidence for widespread resistance QTL pyramiding. Molecular Plant Pathology, 21(1), 3–16. https://doi.org/10.1111/mpp.12874

Tamta, S., & Singh, J. P. (2018). Heterosis in Tomato for Growth and Yield Traits. International Journal of Vegetable Science, 24(2), 169–179. https://doi.org/10.1080/19315260.2017.1407857

Tanksley, S. D. (2004). The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. Plant Cell, 16(SUPPL.), 181–189. https://doi.org/10.1105/tpc.018119

Taranto, F., D'Agostino, N., Greco, B., Cardi, T., & Tripodi, P. (2016). Genome-wide SNP discovery and population structure analysis in pepper (Capsicum annuum) using genotyping by sequencing. BMC Genomics, 17(1), 1–13. https://doi.org/10.1186/s12864-016-3297-7

Taylor, P., & Duvick, D. N. (2008). Crop Breeding in the Twenty- First Century Crop Breeding in the Twenty-First Century. Journal Of Crop Improvement, September 2011, 37–41. https://doi.org/10.1300/J411v14n01

Thornsberry, J. M., Goodman, M. M., Doebley, J., Kresovich, S., Nielsen, D., & Buckler, E. S. (2001). Dwarf8 polymorphisms associate with variation in flowering time. Nature Genetics, 28(3), 286–289. https://doi.org/10.1038/90135

Tieman, D., Zhu, G., Resende, M. F., Jr, Lin, T., Nguyen, C., Bies, D., Rambla, J. L., Beltran, K. S. O., Taylor, M., Zhang, B., Ikeda, H., Liu, Z., Fisher, J., Zemach, I., Monforte, A., Zamir, D., Granell, A., Kirst, M., Huang, S., & Klee, H. (2017). A chemical genetic roadmap to improved tomato flavor. Science, 355(6323), 391–394.https://doi.org/10.1126/science.aal1556

Timberlake, W. E. (2013). Heterosis. Brenner's Encyclopedia of Genetics: Second Edition, 2, 451–453. https://doi.org/10.1016/B978-0-12-374984-0.00705-1

Toni, H. C., Djossa, B. A., TeleAyenan, M. A., & Teka, O. (2020). Tomato (Solanum lycopersicum) pollinators and their effect on fruit set and quality. The Journal of Horticultural Science and Biotechnology. Advance online publication. https://doi.org/10.1080/14620316.2020.1773937

UPOV. (2002). SOLAN_LYC Solan. Variety, 21, 1–26. https://repository.cimmyt.org/bitstream/handle/10883/20997/62776.pdf?sequence=1&isAllowed=y

Vales, M. I., Schön, C. C., Capettini, F., Chen, X. M., Corey, A. E., Mather, D. E., Mundt, C. C., Richardson, K. L., Sandoval-Islas, J. S., Utz, H. F., & Hayes, P. M. (2005). Effect of population size on the estimation of QTL: A test using resistance to barley stripe rust. Theoretical and Applied Genetics, 111(7), 1260–1270. https://doi.org/10.1007/s00122-005-0043-y

Vanlay, M., Samnang, S., Jung, H. J., Choe, P., Kang, K. K., & Nou, I. S. (2022). Interspecific and Intraspecific Hybrid Rootstocks to Improve Horticultural Traits and Soil-Borne Disease Resistance in Tomato. Genes, 13(8). https://doi.org/10.3390/genes13081468

Verma, A., & Singh, Y. (2018). Generation mean analysis of horticultural traits in mid-late cauliflower (Brassica oleracea L. var. botrytis) under sub temperate conditions of Western Himalayas. Plant Breeding, 137(1), 97–108. https://doi.org/10.1111/pbr.12556

Wang, J., & Zhang, Z. (2021). GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. Genomics, Proteomics and Bioinformatics, 19(4), 629–640. https://doi.org/10.1016/j.gpb.2021.08.005

Wang, S., Wu, H., Lu, Z., Liu, W., Wang, X., Fang, Z., & He, X. (2023). Combining Ability Analysis of Yield-Related Traits of Two Elite Rice Restorer Lines in Chinese Hybrid Rice. International Journal of Molecular Sciences, 24(15). https://doi.org/10.3390/ijms241512395

Wang, T., Zhang, Z., Zhu, H., Zhang, Y., Gao, W., Wang, X., Piao, Z., & Zou, Q. (2020). Phenotypic diversity and genome-wide association mapping of earliness-related traits in cultivated tomato (Solanum lycopersicum L.). Scientia Horticulturae, 264(November 2019), 109194. https://doi.org/10.1016/j.scienta.2020.109194

Wang, T., Zou, Q. D., Qi, S. Y., Wang, X. F., Wu, Y. Y., Liu, N., Zhang, Y. M., Zhang, Z. J., & Li, H. T. (2016). Analysis of genetic diversity and population structure in a tomato (Solanum lycopersicum L.) germplasm collection based on single nucleotide polymorphism markers. Genetics and Molecular Research, 15(3), 1–12. https://doi.org/10.4238/gmr.15038209

Wang, X., Gao, L., Jiao, C., Stravoravdis, S., Hosmani, P. S., Saha, S., Zhang, J., Mainiero, S., Strickler, S. R., Catalá, 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, X., Pang, Y., Zhang, J., Zhang, Q., Tao, Y., Feng, B., Zheng, T., Xu, J., & Li, Z. (2014). Genetic background effects on QTL and QTL × environment interaction for yield and its component traits as revealed by reciprocal introgression lines in rice. Crop Journal, 2(6), 345–357. https://doi.org/10.1016/j.cj.2014.06.004

Xia, W., Luo, T., Zhang, W., Mason, A. S., Huang, D., Huang, X., Tang, W., Dou, Y., Zhang, C., & Xiao, Y. (2019). Development of high-density snp markers and their application in evaluating genetic diversity and population structure in elaeis guineensis. Frontiers in Plant Science, 10(February), 1–11. https://doi.org/10.3389/fpls.2019.00130

Xiang, Y., Zhang, H., & Liu, Y. (2025). Integrating QTL mapping and GWAS to decipher the genetic architecture of quantitative traits. Frontiers in Plant Science, 16, 1565329. https://doi.org/10.3389/fpls.2025.1565329

Xiaolei Liu, Meng Huang, Bin Fan, Edward S. Buckler, Z. Z. (2016). Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome- Wide Association Studies. PLOS Genetics, 12(2), 1–24. https://doi.org/10.1371/journal.pgen.1005767

Xu, W., Gong, C., Mai, P., Li, Z., Sun, B., & Li, T. (2024). Genetic diversity and population structure analysis of 418 tomato cultivars based on single nucleotide polymorphism markers. Frontiers in Plant Science, 15, 1445734. https://doi.org/10.3389/fpls.2024.1445734

Xu, Y., Li, P., Yang, Z., & Xu, C. (2017). Genetic mapping of quantitative trait loci in crops. Crop Journal, 5(2), 175–184. https://doi.org/10.1016/j.cj.2016.06.003

Yang, H., Zhang, Z., Zhang, N., Li, T., Wang, J., Zhang, Q., Xue, J., Zhu, W., & Xu, S. (2024). QTL mapping for plant height and ear height using bi-parental immortalized heterozygous populations in maize. Frontiers in Plant Science, 15, 1371394. https://doi.org/10.3389/fpls.2024.1371394

Yan, J., Yang, X., Shah, T., Sánchez-Villeda, H., Li, J., Warburton, M., Zhou, Y., Crouch, J. H., & Xu, Y. (2010). High-throughput SNP genotyping with the Goldengate assay in Maize. Molecular Breeding, 25(3), 441–451. https://doi.org/10.1007/s11032-009-9343-2

YAO, X. yun, WANG, J. yu, LIU, J., WANG, W., YANG, S. long, ZHANG, Y., & XU, Z. jin. (2016). Comparison and analysis of QTLs for grain and hull thickness related traits in two

recombinant inbred line (RIL) populations in rice (Oryza sativa L.). Journal of Integrative Agriculture, 15(11), 2437–2450. https://doi.org/10.1016/S2095-3119(15)61311-9

Ye, J., Wang, X., Wang, W., Yu, H., Ai, G., Li, C., Sun, P., Wang, X., Li, H., Ouyang, B., Zhang, J., Zhang, Y., Han, H., Giovannoni, J. J., Fei, Z., & Ye, Z. (2021). Genome-wide association study reveals the genetic architecture of 27 agronomic traits in tomato. Plant Physiology, 186(4), 2078–2092. https://doi.org/10.1093/plphys/kiab230

Yeon, J., Nguyen, T. T. P., Kim, M., & et al. (2024). Prediction accuracy of genomic estimated breeding values for fruit traits in cultivated tomato (Solanum lycopersicum L.). BMC Plant Biology, 24, 222. https://doi.org/10.1186/s12870-024-04934-8

Yordanov, M. (1983). Heterosis in the Tomato. 6(1912), 189–219. https://doi.org/10.1007/978-3-642-81977-3_7

Yu, D., Gu, X., Zhang, S., Dong, S., Miao, H., Gebretsadik, K., & Bo, K. (2021). Molecular basis of heterosis and related breeding strategies reveal its importance in vegetable breeding. Horticulture Research, 8(1). https://doi.org/10.1038/s41438-021-00552-9

Yu, K., Wang, H., Liu, X., Xu, C., Li, Z., Xu, X., Liu, J., Wang, Z., & Xu, Y. (2020). Large-Scale Analysis of Combining Ability and Heterosis for Development of Hybrid Maize Breeding Strategies Using Diverse Germplasm Resources. Frontiers in Plant Science, 11(June), 1–16. https://doi.org/10.3389/fpls.2020.00660

Zhang, J., Zhao, J., Liang, Y., & Zou, Z. (2016). Genome-wide association-mapping for fruit quality traits in tomato. Euphytica, 207(2), 439–451. https://doi.org/10.1007/s10681-015-1567-0

Zhang, J., Zhao, J., Xu, Y., Liang, J., Chang, P., Yan, F., Li, M., Liang, Y., & Zou, Z. (2015). Genome-wide association mapping for tomato volatiles positively contributing to tomato flavor. Frontiers in Plant Science, 6(NOVEMBER), 1–13. https://doi.org/10.3389/fpls.2015.01042

Zhang, Y., Chen, J., Gao, Z., Wang, H., Liang, D., Guo, Q., Zhang, X., Fan, X., Wu, Y., & Liu, Q. (2024). Identification of heterosis and combining ability in the hybrids of male sterile and restorer sorghum [Sorghum bicolor (L.) Moench] lines. PLoS ONE, 19(1 January), 1–18. https://doi.org/10.1371/journal.pone.0296416

Zhou, H., Wang, X., Huang, Z., Gao, J., Guo, Y., Du, Y., & Hu, H. (2016). Identification of Quantitative Trait Loci for Fruit Weight, Soluble Solids Content, and Plant Morphology Using an Introgression Line Population of Solanum pennellii in a Fresh Market Tomato Inbred Line. Horticultural Plant Journal, 2(1), 26–34. https://doi.org/10.1016/j.hpj.2016.02.007

Zhou, R., Wu, Z., Cao, X., & Jiang, F. L. (2015). Genetic diversity of cultivated and wild tomatoes revealed by morphological traits and SSR markers. Genetics and Molecular Research, 14(4), 13868–13879. https://doi.org/10.4238/2015.October.29.7

Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and Prospects of Association Mapping in Plants. The Plant Genome, 1(1), 5–20. https://doi.org/10.3835/plantgenome2008.02.0089

APPENDICES

Appendix Table 3.1: Description of 15 quantitative traits and methods of measuring the data

Trait	Code	Method of measurement
Days to 50% flowering	DFF	The date when 50% of plants in a plot had one open flower
Plant height (cm)	PH	Taken by measuring the length of five randomly selected plants from the base of the soil to the tip of the

Total number of fruits per plat	TNP	The sum of the number of marketable and unmarketable fruits at each harvest was divided by the number of
Marketable number of fruits per plant	MNP	The sum of the number of marketable fruits at each harvest was divided by the number of plants in a plot.
Marketable yield (g) per plant	MYP	The sum of the weight of marketable fruit at each harvest was divided by the number of plants in a plot.
Total yield (g) per plant	TYP	The sum of the marketable and unmarketable fruit weights at each harvest was divided by the number of
Fruit width (mm)	FWd	The average width of ten randomly selected fruits using a Vernier caliper at the second harvest.
Fruit weight (g)	FWt	The average weight of ten randomly selected fruits was calculated using a sensitive balance at the second
Fruit length (mm)	FL	The average height of ten randomly selected fruits was recorded using a Vernier caliper at the second harvest.
Fruit wall thickness (mm)	PCTk	The average fruit wall thickness of ten randomly selected fruits was recorded at the second harvest. It was determined by cross-sectional cutting of each fruit using a knife and measuring the fruit wall thickness using a
Total soluble solid (%)	TSS	Average fruit TSS of ten randomly selected fruits at the second harvest. It was determined by cutting each fruit into smaller pieces using a knife, followed by the preparation of composite juice and measuring the TSS from a drop of composite juice using a refractometer
Fruit number per cluster	NFtC	Average count of fruits from five clusters per plant of five randomly selected plants in a plot
Number of locules	NL	Average count of the number of locules in ten randomly selected fruits at the second harvest. It was determined by cross-sectional cutting of fruits and counting of the
Fruit shape index	FSHI	Average ration of the value of fruit length to fruit width of ten randomly selected fruits
Juice volume (ml)	JV	The average volume of juice extracted from ten randomly selected fruits at second harvest

Appendix Table 3.2: Flower traits data collection method

Traits	Code	Descriptors	Remark
Inflorescence type	1	Generally multiparous	Whole plot observation at full maturity stage
	2	Both	
	3	Generally uniparous	
Style position	4	Highly exserted	

	3	Slightly exerted	
	2	Same level as stamen	
	1	Inserted	
Style hairiness	0	Absent	
	1	Present	

Appendix Table 3.3: Growth pattern traits data collection method

Traits	Code	Descriptors	Remark
Growth type	1	Dwarf	
	2	Semi-determinate	Observed on the whole plot when ripening of fruits starts
	3	Indeterminate	
	4	Determinate	
Plant size	3	Small	Visual estimation of the whole plot
	7	Large	
	5	Intermediate	
Stem pubescence density	7	Dense	
	3	Sparse	
	5	Intermediate	
Foliage density	7	Dense	Visual estimation of the whole plot
	3	Sparse	
	5	Intermediate	
Leaf Attitude	5	Horizontal	
	7	Drooping	
	3	Semi erect	
Leaf type	5	Pimpinellifolium	
	1	Dwarf	
	6	Hirsutum	
	2	Potato leaf	
	4	Peruvianum	
	7	Other	
	3	Standard	

Appendix Table 3.4: Immature and mature stage fruit trait data collection method

Traits	Cod	Descriptors	Remark
Exterior color of immature fruits	1	Greenish-white	Taken before maturity
	5	Green	
	3	Light green	
	7	Dark green	

	9	Very dark green	
Presence of green shoulder	1	Present	
	0	Absent	
Intensity of green shoulder	3	Slight	
	5	Intermediate	
	7	Strong	
	0	Null	
Extent of green shoulder		Large	
		Small	
		Medium	
		Null	
Fruit firmness	3	Soft	Recorded by pressing the widest part of the fruit side after harvesting in full ripeness stage
	5	Intermediate	
	7	Firm	
Fruit blossom end shape	1	Indented	
	3	Pointed	
	2	Flat	
Exterior color of mature fruit	4	Pink	
	3	Orange	
	5	Red	
	1	Green	
	2	Yellow	
	6	Other	
Flesh color of pericarp	4	Pink	
	3	Orange	
	5	Red	
	1	Green	
	2	Yellow	
	6	Other	
Predominant fruit shape	7	Pyriform (pear	Taken when the fruit turn color at harvesting
	5	Heart shape	
	1	Flattened	
	8	Ellipsoid(plum	
	2	Slightly flattened	
	4	High rounded	
	3	Rounded	
	6	Cylindrical(long	
	9	Other	