Genetics and genomics approaches to improve tomato genotypes for tolerance to high temperature conditions

FABRIZIO OLIVIERI

University of Naples «Federico II»
Department of Agricultural Sciences

Ph.D. Agricultural and Food Science

Cycle XXXII

Tutor: Amalia Barone

Coordinator: Prof .Guido D'urso

2017-2020

Table of contents

Summary

Global warming and the growing food demand due to the world population growth are two critical threats affecting food availability in the next future. The predicted dangerous scenario is the raise of temperatures that could lead to increases between 1.8 and 4.0 °C higher than present in the year 2100. In crop plants, the high temperature (HT) can affect all the reproductive stages starting from pollen viability to final crop yield. Tomato (*S. lycopersicum*) is an important crop distributed worldwide. Notwithstanding its South American origin, the raising of temperatures up to 35°C during the reproductive phases could negatively affect the number of flowers, the fruit set and, consequently, the final yield. For this reason, the selection of tolerant genotypes is one of the main efforts to face tomato yield reduction under the increasing temperatures. A group of tomato genotypes was evaluated for five yield-related traits under high temperature conditions, such as the number of flowers *per* inflorescence (NFL), fruit set (FS), number of fruit *per* plant (TNF), fruit weight (FW) and yield *per* plant (YP) collecting data in three experimental trials located in Campania and Puglia in two years. The statistical analysis revealed that for all traits the genotype by environment interaction resulted statistically significant. Moreover, correlation analyses evidenced that TNF resulted the major yield component in the trials, contributing positively to the final YP. A selection index (SI) was also calculated considering all traits showing a positive correlation to YP and it allowed to select seven genotypes and two hybrid controls as the best performing in three different environments. A stability analysis revealed that one genotype (E42) showed high- and stable-yield trait and a YP comparable to those of two F_1 hybrids selected for high yield under HT.

In order to investigate the genetic variability of the assayed genotypes in the present thesis, two different approaches were used to identify Candidate Genes (CGs) and QTLs putatively involved in the response to HT. The first consisted of a GWAS analysis carried out on a larger population using a high-throughput genomic platform that allowed the identification of 14 marker-associations to the yield-related traits. In particular, 11 genes and a QTL associated to YP were identified that could be involved in the tolerance response. The second approach was the use of a genotyping-by-sequencing (GBS) technique performed on 21 genotypes in order to study the genetic variability that could be used for identification of CGs putatively involved in the tolerance to HT and applied in breeding strategies. A SNP effect analysis was performed to predict how SNPs/InDels could affect the protein effectiveness. Fifty-three SNPs and 55 InDels showed a disruptive effect on protein functions. In the evaluation of the whole genome variability, two genotypes (E36 and E48) showed a very low number of private SNPs/InDels (78 and 97, respectively), whereas the genotypes E42 (9,146 SNPs/InDels) and PDVIT (9,421 SNPs/InDels) showed the highest percentage of private mutations. The high genetic variability recorded in these two genotypes was explored by BLAST analysis carried out on 82 markers, evidencing the presence of wild introgressed regions in both genotypes. Moreover, the calculation of Identity-by-State (IBS) analysis was performed in order to calculate the genetic distances among all genotypes. The analysis confirmed that E42 and PDVIT were genetically different from the other genotypes, with IBS values ranging from 0.60 and 0.71.

Combining the genotypic information with YP and quality traits recorded in previous studies, 19 F_1 hybrid combinations were obtained. In 2018, the hybrids were evaluated for three yield-related traits (TNF, FW and YP) and four hybrids were selected for their high percentages of heterosis (%Het) for YP: 17H14 (%Het=45.45), 17H36 (%Het=25.63), 17H37 (%Het=498.64) and 17H39 (%Het=25.96). These four hybrids were evaluated for a second year, also recording two qualitative traits (Total

Soluble Solid Content, TSSC and Titratable Acidity, TA). The hybrids 17H37 and 17H39 confirmed their high percentage of heterosis for both yield and fruit quality traits.
Finally, a genome editing approach was used to investigate the role of two HSPs in the tolerance to high temperatures. Genetic transformation of a heat tolerant introgressed line (IL4-4) was performed starting from the study of gene expression analysis of two HSP90 acting during the reproductive stages. The analysis revealed that the heat tolerant genotype IL4-4 showed a significant increase of gene expression in these two genes in ovary tissues, with peaks between the 4^{th} and the 7^{th} day post-anthesis, suggesting that these genes might play a key-role in the embryo development. In this way, up till now, 7 transformant plants have been obtained and are under evaluation. Comprehensively, the present study showed that the combination of a rigorous phenotyping activity carried out in multi-environmental conditions with the acquisition of genotypic data represents a successful choice to enhance innovative breeding programs for the selection of genotypes and hybrids under harsh temperature conditions.

1. Introduction

1.1 Climate change and plant adaptation

Global warming and the growing food demand due to the world population growth are two critical threats to the quality of human life. Indeed, the raising temperatures predicted in the next few years (Ainsworth and Ort, 2010; Gourdji et al., 2013) could affect the food availability due to the harvest losses. The predicted dangerous scenario is the raise of temperatures of 0.2°C per decade and could lead to increase temperatures between 1.8 and 4.0°C higher than present in 2100 (Sato et al., 2004). The raising of temperatures leads to critical changes in natural ecosystem inducing Heat Stress (HS) affecting the life cycles of the organisms. In particular, plants have to strike a new balance in response to high temperatures. Notwithstanding the temperatures are not constant during the different phenological phases of crop plants, extreme variations during the crop growing season can affect the different molecular interactions. Indeed, the HS can affect all the reproductive stages such as pollen viability and fruit setting phases (Figure 1.1). Indeed, the flowering period results the most critical phase, since even only a short heat stress period can affect directly the floral buds causing stigma exertion (Figure 1.2), flower abortion and resulting in drop of the final crop yield (Hedhly et al., 2009; Thakur et al., 2010). Temperatures higher than 32°C can already affect negatively the fertility of pollen leading to a reduced fertilization, translating in a substantially reduction of the final yield in different crop species.

Figure 1.1 Effect of high temperatures on crop plants (Hasanuzzaman et al., 2013).

For these reasons, plants have developed different mechanisms in order to mitigate the negative effect of the abiotic stresses as molecular mechanisms of avoidance ones. HS can also influence physiological aspects such as photosynthetic activity, which affects negatively the plant growth and alters the cellular membrane structures. Moreover, plants show adaptation mechanisms correlated to the response to HS modifying the patterns of gene expression (Yang et al., 2006a), inducing the expression of different genes such as the heat-shock proteins (HSPs), and down-regulating many other genes (Yost and Lindquist, 1986). Seven major members of HSP family protein can be distinguished in plants based on the molecular weight (Table 1.1): HSP100, HSP90, HSP70, HSP60 and HSP20 (or well-known small heat-shock proteins, sHSPs), (Swindell et al., 2007); these can be expressed in specific developmental stages, such as embryogenesis, seed germination, pollen development and fruit ripening (Prasinos et al., 2005).

Table 1.1 Major members of HSPs family proteins and basic function in plant system for the tolerance to high temperatures, (Hasanuzzaman et al., 2013).

Major classes of heat-shock proteins	Functions
HSP100	ATP-dependent dissociation and degradation of aggregate protein
HSP90	Co-regulation of heat stress linked signal transduction complexes and management of protein folding. It requires ATP for its function
HSP70, HSP40	Primary stabilization of newly formed proteins, ATP-dependent and binding release
HSP60, HSP10	ATP-dependent specialized folding machinery
HSP20 or small HSP (sHSP)	Formation of high molecular weight oligomeric complexes which serve as cellular matrix for stabilization of unfolded proteins. HSP100, HSP70 and HSP40 are needed for its release

In particular, HSP60 and HSP70 are the most conserved proteins in nature, underlying a consistent role in the response to high temperature stress. In the same way, the sHSPs are the most abundant and are encoded by six nuclear gene families, and their function is associated to distinct cellular compartments, as reported in immune-localization studies (Nieto-Sotelo et al., 2002; Yang et al., 2006b). Moreover, other HSPs show peculiar cellular localization. In tomato, HSP68 is localized in mitochondria, is constitutively expressed and its gene expression is upregulated under heat stress condition (Neumann et al., 1993). Similarly, also the sHSPs are involved in protection, aggregating into a granular structure with the function of preserving the biosynthesis machinery and native configuration of proteins for the heat stress (Miroshnichenko et al., 2005; Wahid et al., 2007). Probably, the peculiar conformational dynamism of sHSPs may have a key-role in the mitigation of

plant cells (Iba, 2002; Schöffl et al., 1999). Moreover, the HSPs are involved also in the expression of genes involved in the signal transduction. In this way, the protein kinases (MAPKK) result upregulated in response to signal molecules linked to heat stress or heat shock proteins, that play a role in prevention of the degradation of intracellular proteins. Moreover, other HSP family members have functions in the protection of biochemical machineries. Indeed, in tomato the HSP21 has the function to protect the photosystem II (PSII) from oxidative stresses deriving from heat stress response (Neta-Sharir et al., 2005). The sHSPs show similar functions, localizing in chloroplast membrane to mitigate the adverse effect of the high temperatures and plays a key-role in photosynthetic electron transport (Barua et al., 2003). The knowledge of these physical and molecular response, which can be explored studying gene expression and different transduction pathways, could be useful for the genetic improvement of crop species.

1.2 Tomato as model species

Tomato (*S. lycopersicum*) is an important crop distributed worldwide. Moreover, for its economic value and for the beneficial nutritional properties of its fruits (Rao and Agarwal, 2000), tomato is largely used in the Mediterranean diet. Moreover, this species is an important model system for plant, for the fleshy fruit biology and from a genetic point of view due to the nature of different domestication events during its history (Blanca et al., 2012; Blanca et al., 2015; Meissner et al., 1997). On one hand, the domestication events permitted to select tomato for its yield and adaptation, on the other hand this selection activity lead to the erosion of genetic variability and then to lose variability sources. Therefore, high genetic diversity can be retrieved from wild ancestors (Víquez-Zamora et al., 2013) as good source for obtaining heat tolerant genotypes. The majority of commercial varieties of tomato are not heat tolerant. Indeed, notwithstanding the south-American origin, the raising of temperatures up to 35°C during the reproductive period, starting from pollen formation and fertility to fruit set (Sato et al., 2004), induces yield decreases. Indeed, it is documented that the prolonged exposition of tomato to a temperature regime of 34°C/19°C affects negatively the number of flowers and the fruit set (with loss up to 71%) (Hazra et al., 2008). These results can be explained by the fact that the fertility of pollen grains results the most critical phase for its vulnerability to the temperature stress, resulting in a drop of tomato yield (Annex, 2012; Bokszczanin et al., 2013; Paupière et al., 2017; Warland et al., 2006). The genotypes showing heat tolerance trait have developed strategies to maintain the pollen viability producing more polysaccharides than heat sensitive genotypes (Firon et al., 2006), resulting in abundance of fruit production compared to heat susceptible ones (Dane et al., 1991). To contrast this serious problem different studies are reported in literature, using controlled conditions of high temperatures (Figure 1.2) that helped to select the best tomato genotypes in

restricted controlled conditions of heat stress (Paupière et al., 2017; Peet et al., 1998; Xu et al., 2017a) and to study all the reproductive phenomena involved in heat stress response.

Figure 1.2 Tomato under HT-controlled conditions. a) flower with stigma exertion; b) fruit setting of tomato plants.

Notwithstanding the importance to characterize the capability of tomato reproduction in controlled condition, experimental trials under high temperature in open field, which represents the natural condition of tomato growing, are important to evaluate the final crop yield and its stability in different stressed environments and are not widely reported. Indeed, in open field fluctuations of other meteorological parameters and other related stress linked to high temperatures can occur, evidencing a strong Genotype by Environment interaction. One model for yield stability calculation is reported in bibliography (Eberhart and Russell, 1966) and it is useful to select the genotypes that show less differences in terms of an analyzed trait in different environments. For these reasons, phenotyping trials allow to identify genotypes that show a good resilience in different environmental conditions.

1.3 Breeding for yield improvement under high temperatures

The traditional breeding can be a solution to face mutating climatic conditions, in order to contrast the predicted harvest losses due to the temperature increases. The narrow genetic variability due to the domestication events has to encourage the researchers to exploit the genetic diversity to guarantee

new genotypes. Generally, the genetic improvement focuses the attention on phenotyping many genotypes based on interesting agronomical traits, such as yield performances and quality traits, in order to select new genotypes combining different favourable alleles. The use of breeding programs for stress tolerance is a new approach. Indeed, in most cases, the genetic improvement focused on the hybrid production to improve yield in optimal conditions, succeeding to improve the yield per unit area (Warren, 1998). The application of breeding programs to maintain yield under extreme temperatures requires knowledge on physiological and genetic mechanisms controlling the stress tolerance. Indeed, the heterosis phenomenon that occurs in cross combinations allows to obtain hybrid progenies with superior traits in terms of growth and productivity, compared to the parental lines (Darwin, 1868), and this occurs also in tomato (Bai and Lindhout, 2007; Bhatt et al., 2001). Two different models can explain the heterosis phenomena. The first is the so-called dominance model, where the deleterious recessive alleles for different loci are genome-wide complemented in the hybrid progenies, whereas the so-called over-dominance model mostly focuses on the interaction of different alleles of many loci in the hybrid, thus explaining the vigour increase. An important example of the last model is described by the Single Flower Truss gene (*sft*), which combined to the recessive Self-Pruning allele (*sp*), drives heterosis for yield in tomato (Jiang et al., 2013; Krieger et al., 2010). In this case, the possibility to combine physiological and genetic information allows to formulate the hypothesis for the improvement of the yield. This goal could be more efficiently pursued using Next-generation sequencing (NGS) technologies for the identification of QTLs and/or genes in order to exploit the genetic variability available.

1.4 Next generation Sequencing (NGS) as support for phenotyping

The heat tolerance is a trait characterized by high complexity, controlled by different genes and their interactions (Howarth, 2005). The molecular biology and the develop molecular markers are useful for the identification of QTLs and/or genes directly involved in the heat tolerance and to dissect the complexity of this trait (Grilli et al., 2007; Xu et al., 2017a). Moreover, the natural phenotypic variation present in different cultivars can be explored using sequencing technologies that could be used in association mapping approaches, consisting in a powerfool tool for the identification of chromosomal regions controlling the heat tolerance (Bergelson and Roux, 2010). In this way, the Next-generation sequencing (NGS) is a new possibility to obtain many genomic dataset to exploit QTLs loci and genes controlling the response to heat stress. Some of NGS technologies, as different previous markers based on restriction enzyme reactions (Restriction Fragment Lenght Polymorphism, RFLP and Amplified Fragment Lenght Polymorphism, AFLP), depend from the restriction enzymes in order to reduct the genomic representation. A good choice for the plant breeding can be the use of

new genotyping strategies sequence-based that allow to link genotype and phenotype information. In this way, the Restriction-site Associated DNA sequencing (RAD-seq) allows the identification of thousands of markers in many individuals. This method has different steps. Starting from the digestion of genomic DNA enzyme-mediated, the fragments are ligated to adapter containing the forward primer for the amplification and a unique barcode for the demultiplexing step (Baird et al., 2008; Craig et al., 2008; Cronn et al., 2008). The restricted fragments were collected and randomly cut to reduce fragments to similar size and then ligated to a second adapter that favors the amplification only of the fragments amplified from the forward primer. Up till now, variants of RAD-seq have been developed, as reported in Peterson et al. (2012), known as double-digest RAD-seq (ddRAD-seq) that differs for two aspects: the elimination of random shearing phase converted in an end repair of genomic DNA and the size-based selection of genomic fragments (Peterson et al., 2012). The principal step of this approach is the use of two endonucleases that leads to decrease the representation of the genome and increase the quality of the sequencing data, by producing reads that fall in the selection window of 50-100 bp. The window size offers two advantages: one is the selection of a small number of fragments within this range, avoiding duplicates of the same fragment, and second, the shared bias in region representation, which favors fragment selection closest to the mean size (Figure 1.3).

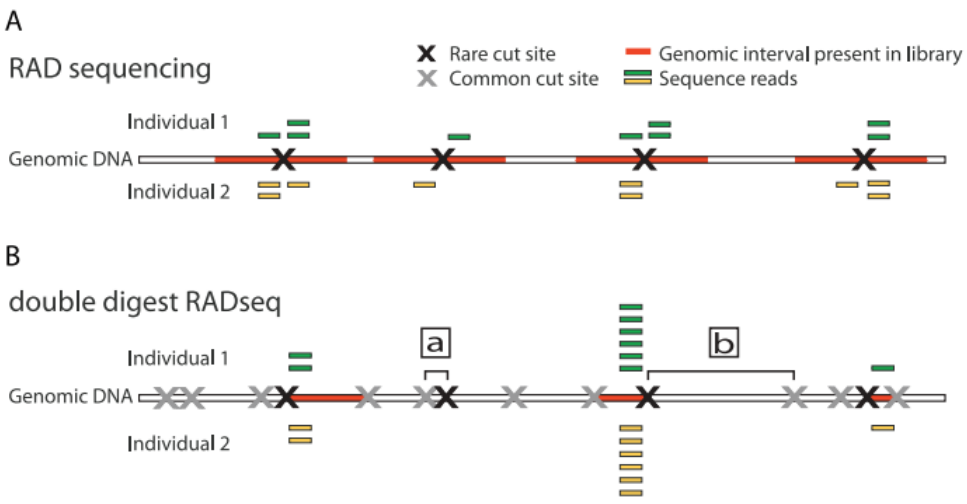


Figure 1.3 Differences between the RAD-seq (A) and ddRAD-seq (B) approaches. (Peterson et al., 2012).

All the polymorphisms (SNPs and InDels) detected in the experiment can be used as genetic markers. This approach can be used in genotyping applications as the development of linkage map using SSR and SNP markers (Xiao et al., 2015; Yagi et al., 2017), QTL analysis (Wang et al., 2017; Wu et al., 2014b). In particular, in tomato different applications of this technique were already reported to discover candidate genes involved in the pathogen resistance (Arafa et al., 2017). Another genotyping

platform that can be used in tomato is the Infinium SolCAP SNP array based on transcriptomic data recorded on six different landraces (Hamilton et al., 2012). This platform was used for exploiting genetic differences and for identifying markers associated to fruit quality traits (Ruggieri et al., 2014; Sacco et al., 2015), for evaluating the SNP effects on gene functions or for exploring the breeding-dependent genomic variation (Sim et al., 2012). The genotyping platforms allow the identification of many markers that could be used in the study of wild ancestors of tomato and to carry out a Genome-Wide Association Study (GWAS) useful to evaluate the association between markers and phenotypic data in a large collection of genotypes (Korte and Farlow, 2013). This approach requires the knowledge of molecular variation depending on breeding population, the stratification and the population structure, and the pattern of linkage disequilibrium-decay (LD-decay) (Bauchet et al., 2017).

1.5 Genome editing in breeding approach

The genetic engineering help to understand the function of many genes, including heat related genes. Indeed, in many plant species the thermotolerance resulted strictly dependent on the HSP expression (Gurley, 2000), involved in translocation, folding of denatured proteins, tolerance to oxidative stresses (Gorantla et al., 2007; Queitsch et al., 2000). The use of biotechnological approaches of genetic transformation allows to investigate the Heat Shock Protein function inducing a frameshift facing in the knockout of the gene. In this way, using the innovative technologies of genome editing is possible to study functionally candidate genes involved in the heat stress response. This technique applies the introduction of a Double-Strand Breaks (DSB), using a nuclease, which induces the DNA repairing machinery. The repair mechanisms of DSB are two: a non-homologous end-joining (NHEJ) mechanism, which introduces insertions or deletions in the damaged region, facing in frameshift mutation, whereas the homologous recombination (HR) mechanism (Figure 1.4) is capable to induce a precise repair of the DSB.

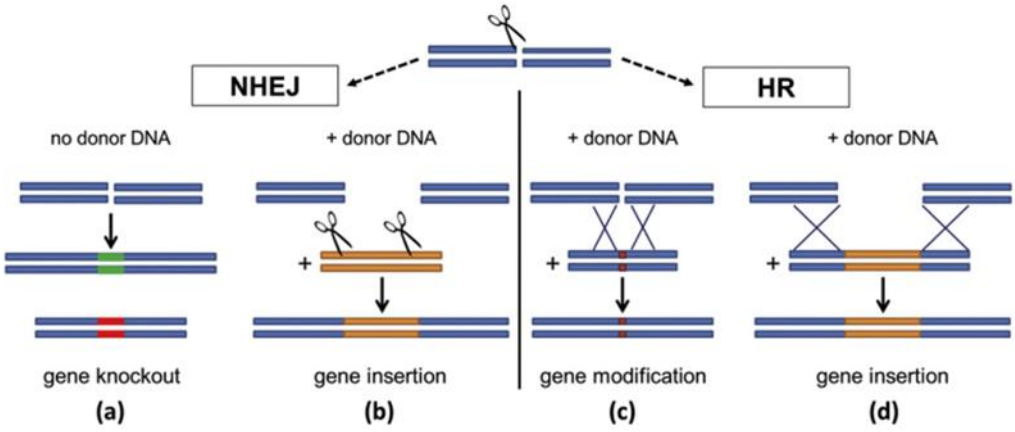


Figure 1.4 Repair mechanism in Genome Editing using two different approaches: NHEJ, Non-Homologous End-Joining and HR, Homolog Recombination (Bortesi and Fischer, 2015).

In the first case the mutation induces the knock-out, useful to study functionally a gene of interest, whereas the HR permits the modification also of a single base, in a technology called Base-Editing. Different genome editing tools are available as the Zinc-finger nuclease (Kim et al., 1996) and transcription activator-like effector nucleases (TALENs) (Christian et al., 2010), which are engineered proteins fused to a nuclease domain of *Fok*I restriction enzyme (Jankele and Svoboda, 2014; Palpant and Dudzinski, 2013). The newest approach is the CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats associated to Cas9) that mimes an RNA interfering mechanism designing RNA guides, which target the gene/portion of interest, inducing insertion or deletion in the genome (D'Ambrosio et al., 2018; Li et al., 2018; Ueta et al., 2017). The CRISPR RNA (crRNA) combined with transactivating CRISPR RNA (trascRNA) guides and activates the Cas9 nuclease (Barrangou et al., 2007). This combined structure is so-called synthetic guide RNA complex (sgRNA) that load a guide RNA (gRNA) designed on target gene proximally to the Protospacer adjacent motif (PAM), a peculiar sequence of three bases "NGG". The Cas9 endonuclease recognizes the PAM sequence in the genome and applies a DSB following a repair mechanism mentioned above (Figure 1.5).

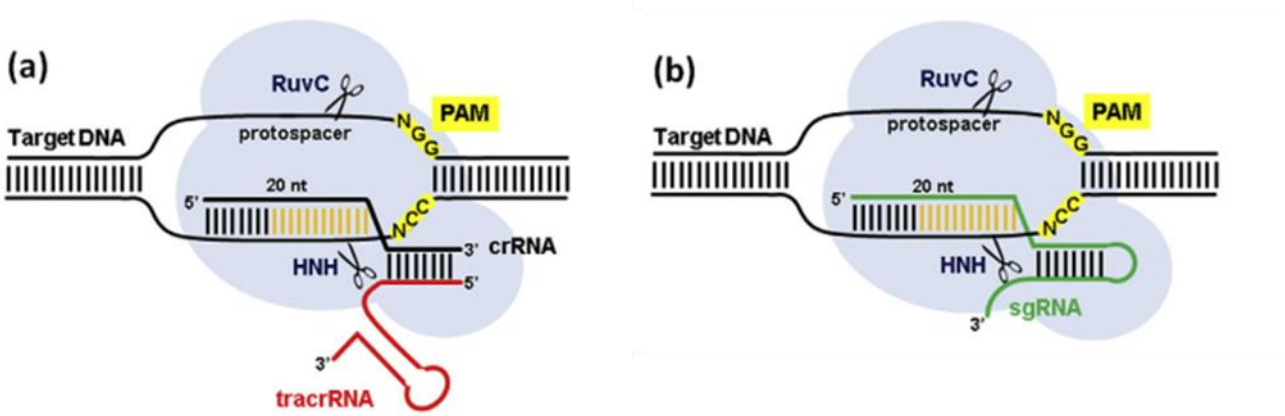


Figure 1.5 gRNA guides cleavage by Cas9. The RNA guide is an oligomer of 20-nt bases responsible to guide the Cas9 on the site that must be modified. When RNA guide matches on the complementary sequences (a), the molecular machinery recognized the Protospacer adjacent motif (b) and Cas9 makes a cleavage (Bortesi and Fischer, 2015).

1.6 Aim of the thesis

The present work will address issues related to heat stress tolerance with the final aim of designing new strategies to maintain high yield of tomato fruit produced at harsh temperature conditions. Indeed, in plant production, high temperature stress often causes yield reduction. One of the most important and economic direction to overcome the negative effects of heat stress is to identify and develop heat-tolerant genotypes that could maintain good yield performances and stability in different adverse environmental conditions. The main goal of the present thesis is to set up breeding strategies in order to select new varieties and heat tolerant hybrids by combining traditional and innovative approaches.

In particular, specific aims of the present thesis are:

1) Selection of tomato superior genotypes in terms of yield-stability when grown under high temperatures (HT), throughout a phenotyping work carried out in multi-environmental conditions.

2) Identification of QTLs/genes associated with good production under high temperature conditions using high-throughput genotyping platforms.

3) Selection of the best F_1 hybrids combining favorable alleles for tolerance to high temperatures in terms of yield and fruit quality.

4) Functional study of two heat-shock proteins in an introgression line characterized by segregation distortion exploiting a genome editing approach.

2. Phenotyping analysis of tomato genotypes

2.1 Material and methods

2.1.1 Plant material

Plant material includes 32 genotypes (coded E) available at the University of Naples, Department of Agricultural Sciences and listed in Supplementary Table S1 (detailed information about materials are hosted at LabArchive repository http://dx.doi.org/10.6070/H4TT4NXN). Nineteen genotypes are Italian, nine derive from other countries and two are hybrids provided by MONSANTO Agricoltura Italia (Latina, Italy), selected for their heat-tolerance and high-yielding traits (from MONSANTO, unpublished results). These two hybrids were considered as positive controls, whereas the cultivar Moneymaker and M82 were used as negative controls. The experimental trials in 2017 were carried out in two locations of south Italy (Campania at 40°56'56.9"N 14°23'20.3"E and Puglia at 41°15'16.2"N 16°17'26.6"E), whereas in 2018 only in Campania. The choice of these two locations depends on the high temperatures usually recorded during the tomato growing season. In order to favour the exposition of plants under high temperatures during the reproductive phases, tomato plants were transplanted with one-month delay respect to the standard agronomical practices of the locations. The planned experimental design consisted of a randomized complete block of 10 plants for each genotype and three replicates for genotype.

2.1.2 Morpho-agronomic analysis in the fields

During the growing seasons, data of five different phenotypic traits were collected: number of flowers *per* inflorescence (NFL), fruit set percentage (FS), number of fruit *per* plant (TNF), fresh fruit weight (FW) and total yield *per* plant (YP). NFL and FS were evaluated *per* each inflorescence from the second to the fifth truss of three plants, whereas TNF, FW and YP were evaluated at red ripe fruit stage on all plants. Finally, a Selection Index (SI) was calculated by assigning to three traits (FS, TNF, YP) an arbitrary scale. For FS and TNF, 0 = score for trait value varying 0 to 10, 1 = score for trait value varying 11 to 20, 2 = score for trait value varying 21 to 30, 3 = score for trait value varying 31 to 40 etc.; for YP, score obtained by multiplying the YP value expressed in kg/plant value by 10.

2.1.3 Morphological evaluation under controlled conditions

During the year 2019, the genotype E42 was evaluated under controlled temperatures. The seeds were disinfected using a chemical and thermal treatment. Briefly, tomato seeds were sterilized using phosphate dodecahydrate ($Na_3PO_4.12H_2O$) in agitation for three hours and then rinsed with sterile milli-ρ distilled water for five times. In the next step, the seeds were placed in 30 % sodium hypochlorite in agitation for one hour and then rinsed with milli-ρ distilled water for ten times. After

the chemical treatment, seeds were dried on a filter paper for 24 h and then put in a silica gel box for the same period. Then, seeds were subject to thermal treatment at 80±2 °C for 24 h and sown in plateau, after 40 days they were transplanted in greenhouse at Paiporta, Valencia (39°25′40.08″N 0°25′05.88″W). In order to study the effect of high temperatures in E42, three different temperature regimes were assayed and each one was fixed for three weeks. The experiment started during the flowering period on the first truss, setting the first temperature (T1) to 25/20 °C day/night. After three weeks the temperatures were raised up to 30/25 °C (T2) and the finally increased up to 35/30 °C (T3). The experimental design consisted of a full randomized block, 5 replicates and 10 plants *per* replicate. The data were recorded considering the same three plants in the full period. During the experiment, the number of flowers, the number of fruit and the fruit weight in different temperatures regimes were monitored.

2.1.4 Statistical analysis

Data collected were analyzed using IBM SPSS v. 23 (IBM Corp.). In order to better understand the genotype by location interactions for all phenotypic traits, ANOVA analyses were carried out, where the genotype and location were considered fixed factors. A Dunnett's test was calculated to compare the genotypes and controls for each phenotypic trait. Moreover, Pearson's correlation was estimated among different trait analyzed.

2.1.5 Yield stability

In order to validate the selection index (SI) calculated on the group of genotypes, the evaluation of yield *per* plant stability performance of selected genotypes was carried out by using a linear regression model (Eberhart and Russell, 1966). The estimation of the model was carried out considering three years of phenotypic data collection in two locations. In this regression model, the coefficient of regression (b_i^v) was obtained from regression analysis of each genotype, plotting on vertical axis yield recorded in each field and on the horizontal axis the environment index (*EI*), as reported in Eberhart and Russell (1966).The b_i^v is an index describing correlation between the yield of a genotype and the corresponding environments. In order to establish the stable-yielding trait, the $b_i^v = 1$ was regarded as the optimum value describing the stability trait. In order to establish the threshold borders of accuracy, the standard deviation of the b_i^v values was calculated on the total genotypes assayed. In order to test the accuracy of linear regression analysis of each genotype, the average deviation from linearity (ADL) was calculated considering the quadratic model as the best-fit non-linear model (Liu et al., 2009). When the analysis of ADL value calculated for each linear regression resulted statistically different from zero, the b_i^v was not reliable for the selection of high stable-yielding genotypes. Finally, the coefficient of determination (r^2) of the linear regression represented the

robustness of the model. Arbitrarily, a r^2 higher than 0.50 was requested to consider accurate the regression model. Summarizing, the genotypes with YP mean (the mean of yield recorded in different fields) higher than the grand-mean (the mean of whole yield values recorded in different fields and in all genotypes) , the $b_i^v \leq 1$, ADL not statistically different from 0 and the $r^2 > 0.50$, were considered as high stable-yielding genotypes .

2.2 Results

2.2.1 Phenotyping in the year 2017

The 32 genotypes were evaluated for productivity estimating five yield-related traits, in two locations of south Italy usually characterized by high temperatures.

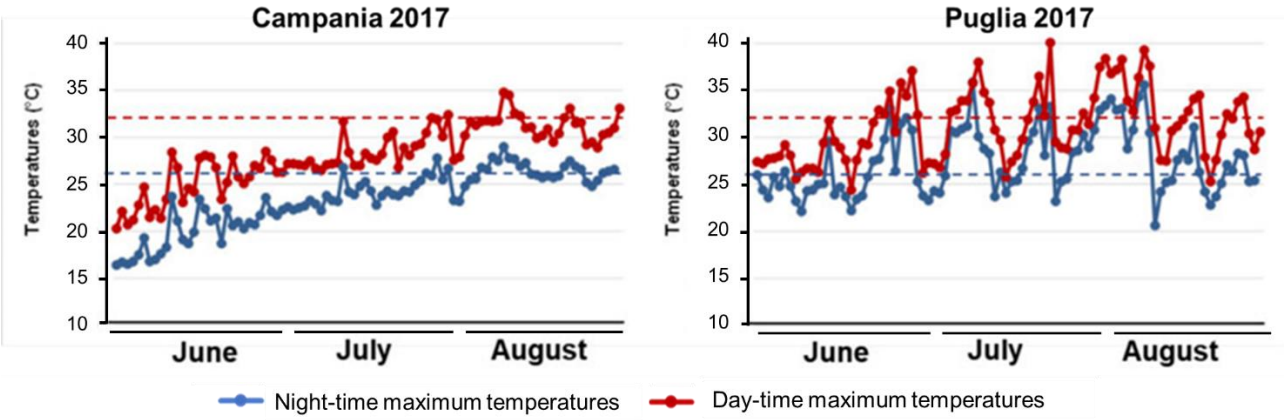
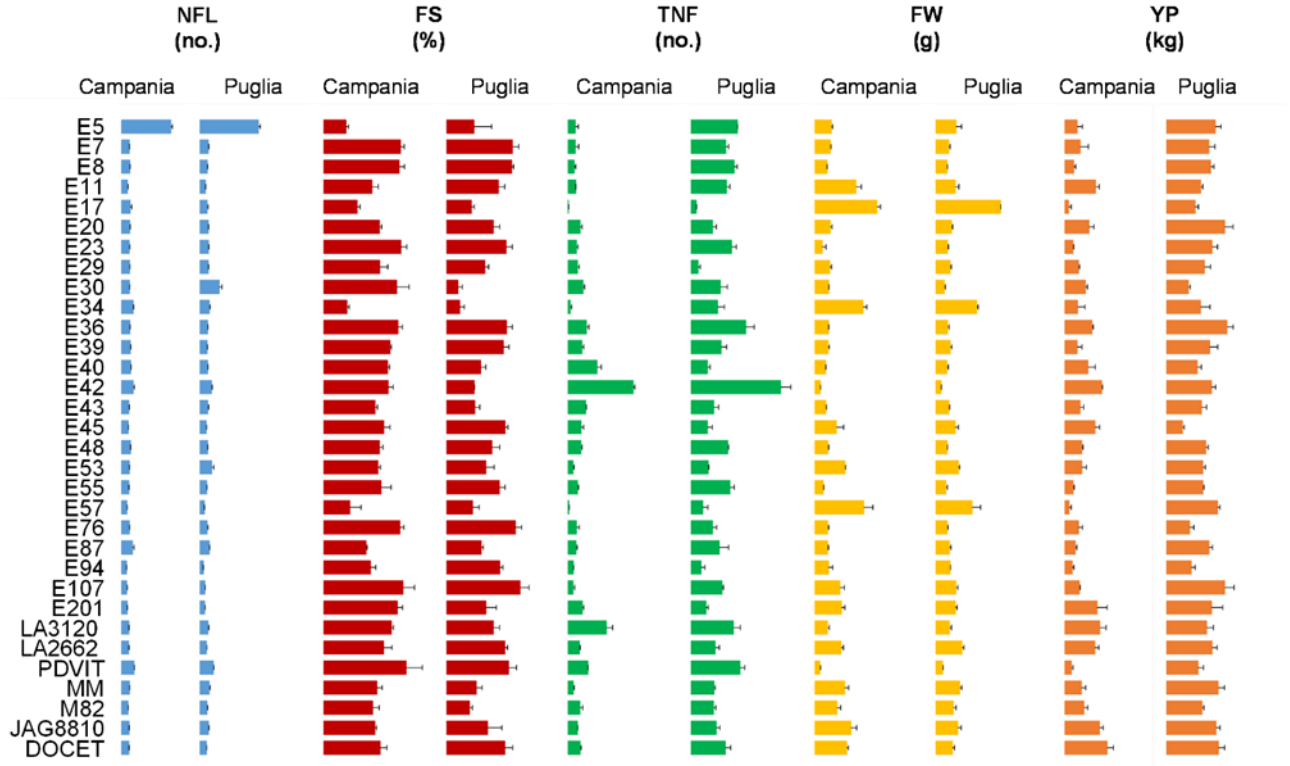


Figure 2.1 Day-time and night-time maximum temperatures recorded in Campania and Puglia in the year 2017. In red the day-time maximum temperature and in blue the night-time maximum temperatures. The red dashed line represents the day critical threshold of 32°C and the blue dashed line the night critical threshold of 26°C.

In order to verify the climate conditions, day/night maximum temperatures were monitored during the growing season (Figure 2.1). In the Campania field the maximum temperatures were generally lower than those recorded in Puglia. As a whole, in Campania the maximum temperature reached 35°C in two days, whereas in Puglia it was over 32 °C for approximately half of the growing season and peaked over 35 °C for 16 days, reaching 38-39 °C. In Puglia field the maximum night temperatures recorded were higher than in Campania with 48 nights exceeding 26 °C in the first region, while in Campania 21 nights reached this temperature. The phenotypic performances evaluated in the two experimental trials were compared to those of two tomato hybrids that are highly productive under high temperature conditions A high variability of data was observed in different phenotypic traits (Figure 2.2). In Campania, the NFL varied from 4.52 (E94) to 45.70 (E5) with a mean value of 8.59. In Puglia, this value ranged from 3.33 flower *per* inflorescence (E94) to 54.53 (E5) with a mean value of 9.26. The coefficients of variation corresponded to 83.92 in Campania field and 97.04 in Puglia field. As for FS, in Campania the value ranged from 21.27 (E5) to 77.15 (PDVIT) with a mean value of 54.60. In Puglia the FS value varied from 10.69 (E30) to 68.50 (E107). Regarding TNF, a wide range of values were recorded in the two fields. Indeed, in Campania the values varied from 3.05 (E17) to 304.18 (E42), with a mean of 62.09. In Puglia, values ranged from 24.55 (E17) to 417.25 (E42). The TNF mean was higher in the latter (138.19 fruit/plant) field than in

Campania (62.09 fruit/plant). As for FW, in Campania the range of FW was between 9.9 g (PDVIT) and 116.28 g (E17). In the same way, in Puglia PDVIT (10.20 g) and E17 (121.19 g) showed the lowest and the highest value of fruit weight, respectively. The coefficients of variation recorded in Campania and in Puglia were similar (66.55 in Campania and 64.03 in Puglia). Finally, a different behaviour was recorded in the two fields for YP. In Campania this value varied from a minimum of 0.34 kg/plant (E17) to 3.43 kg/plant (DOCET) while in Puglia it ranged from 1.53 kg/plant (E45) to a maximum of 5.67 kg/plant (E36) (Supplementary Table S2). The mean values recorded for this trait were 1.62 and 3.74 kg/plant in Campania and Puglia, respectively. The coefficients of variation resulted different between the two fields, with a percentage of ~55% in Campania and ~27% in Puglia.



	NFL (no.)		FS (%)		TNF (no.)		FW (g)		YP (kg)	
	Campania	Puglia	Campania	Puglia	Campania	Puglia	Campania	Puglia	Campania	Puglia
MIN	4.52	3.33	21.27	10.69	3.05	24.55	9.90	10.20	0.34	1.53
MEAN	8.59	9.26	54.60	41.00	62.09	138.19	38.71	34.21	1.62	3.74
MAX	45.70	54.53	77.15	68.50	304.18	417.25	116.28	121.19	3.43	5.67
%C.V.	83.92	97.04	28.51	38.69	94.50	57.26	66.55	64.03	54.87	27.27

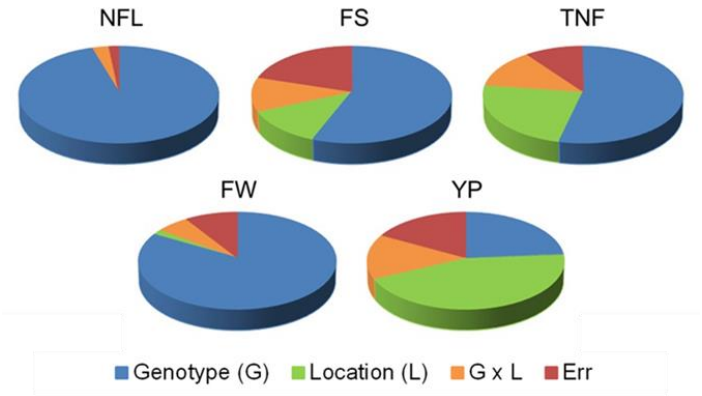
Figure 2.2 Phenotypic variation of five traits observed in two experimental fields on 28 genotypes, two heat tolerant controls (DOCET and JAG8810) and two heat susceptible ones (MM and M82). The trend of variation of traits was measured from flowering to fruit ripening. Each bar represents the mean of values of number of flowers *per* inflorescence (NFL), percentage of fruit set (FS), total number of fruit *per* plant (TNF), fruit weight (FW) and yield *per* plant (YP). Descriptive statistics of phenotypic evaluation in two experimental fields are also reported.

A Dunnett's test was used to estimate statistical differences between the genotypes and the four controls, in two different locations for each trait (Table 2.1).

Table 2.2 Statistical differences of five phenotypic traits between the landraces (coded E) and the four controls DOCET (D) and JAG8810 (J) as heat tolerant and M82 (M) and Moneymaker (MM) as heat susceptible. Dunnett's test (*: $p<0.05$) was perfomed to evaluate the statistical significance. Red and black asterisks indicate the statistical reduction or increase of the trait respect to the control (NFL= No. of flowers *per* inflorescence; FS= Fruit set; TNF=No. of fruit *per* plant; FW=Fruit weight; YP=Yield *per* plant.

Trait (Field)	Control	E5	E7	E8	E11	E17	E20	E23	E29	E30	E34	E36	E39	E40	E42	E43	E45	E48	E53	E55	E57	E76	E87	E94	E107	E201	LA2662	LA3120	PDVIT	D	J	M	MM
NFL (Campania)	D	*									*				*								*					*				
	J	*									*				*								*					*				
	M	*									*				*								*					*				
	MM	*									*				*								*	*				*				
NFL (Puglia)	D	*								*					*				*									*				
	J	*								*											*			*				*				
	M	*								*					*				*									*				
	MM	*								*											*			*	*	*		*				
FS (Campania)	D	*									*										*											
	J	*	*	*				*			*											*			*			*				
	M	*	*	*				*			*	*										*			*	*		*				
	MM	*									*										*							*				
FS (Puglia)	D					*				*	*										*									*		
	J																								*								
	M		*	*				*				*	*				*					*			*		*		*	*			
	MM		*	*																		*			*				*				
TNF (Campania)	D					*					*			*	*						*							*				
	J											*		*	*													*	*				
	M					*					*			*	*						*							*					
	MM									*		*		*	*	*												*	*				
TNF (Puglia)	D					*			*						*						*			*								
	J											*			*														*				
	M											*			*														*				
	MM											*			*														*				
FW (Campania)	D	*	*	*		*	*	*	*	*	*	*	*	*	*	*		*		*	*	*	*	*				*	*				
	J	*	*	*		*	*	*	*	*		*	*	*	*	*	*	*		*		*	*	*				*	*			*	
	M				*	*					*				*					*	*								*		*			
	MM		*	*		*	*	*	*	*	*	*	*	*	*	*		*		*	*	*	*	*				*	*				
FW (Puglia)	D					*					*				*						*							*				
	J			*		*		*		*	*	*		*	*			*		*	*	*							*				
	M					*					*				*						*								*				
	MM		*	*		*		*	*	*	*	*		*	*	*		*		*	*	*	*						*				
YP (Campania)	D	*	*	*		*		*	*	*	*		*			*		*	*	*	*	*	*	*	*				*			*	*
	J	*		*		*		*			*		*							*	*		*	*					*				
	M																													*			
	MM																													*			
YP (Puglia)	D									*							*					*										
	J									*							*					*										
	M											*																					
	MM									*							*					*										

Regarding NFL, two genotypes (E5 and PDVIT) showed a significantly higher number of flowers *per* inflorescence than controls in both locations, whereas other four (E30, E34, E42 and E87) showed NFL significantly higher than controls only in one field. The trend of variation was quite different for FS between the two fields, with some genotypes exhibiting higher values respect to the negative controls in Puglia, whereas in Campania only E5 and E34 showed values lower than all the controls. Regarding TNF, E42 and PDVIT showed a significant increase of the number of fruit *per* plant compared to all controls in both fields while E40 and LA3120 showed values significantly higher than controls only in one location. As for FW, the higher or lower levels respect to the controls were independent of the field, but mostly genotype-specific. As for YP, most genotypes showed values lower than the positive controls in Campania, whereas only a few in Puglia, which also performed worse than the negative control Moneymaker. In all other cases, the genotypes produced YP comparable to the positive controls. Moreover, in order to evaluate the environmental effect on genotypes, a two-way ANOVA was performed to estimate the genotype by locations interaction. As reported in Figure 2.3, the ANOVA analysis revealed that for all traits statistically significant differences were found for each source of variation. The calculation of total sum of square percentage (TSS%) showed that the highest contribution in the total variance was the Genotype (G) in four traits (NFL=95.39%; FS=55.99%; TNF=53.64%; FW=83.31%) whereas for YP, the TSS% of Location resulted higher (44.02%) than that of Genotype (24.00%). Moreover, the variance contribution of Genotype by Location (G x L) for YP resulted the highest (14.94%) compared to the interactions observed for other traits.



Source of variation	NFL		FS		TNF		FW		YP	
	TSS %	Sign.	TSS %	Sign.	TSS %	Sign.	TSS %	Sign.	TSS %	Sign.
Genotype (G)	95.39	***	55.99	***	53.64	***	83.31	***	24.00	***
Location (L)	0.17	***	12.47	***	23.62	***	1.84	***	44.02	***
G x L	2.76	***	11.41	***	12.75	***	5.78	***	14.93	***
Error	1.68		20.13		9.99		9.08		17.06	

Figure 2.3 ANOVA analysis to evaluate the effect on phenotypic traits of the environment recorded on 32 genotypes grown in two different locations. Pie charts describing the contribution in terms of variance for each source of variation are reported. Genotype and Location were considered as fixed factors for the analysis. TSS%, total sum of square percentage; NFL, no. flowers per inflorescence; FS, fruit set; TNF, no. fruit *per* plant; FW, fruit weight; YP, yield *per* plant.

The Pearson's correlation was also calculated among the phenotypic traits recorded for each location (Figure 2.4). In this analysis, FS resulted significantly negative correlated to NFL and to FW in Campania (r=-0.385 and r=-0.594, respectively). As a whole, an expected significant negative correlation was always detected between FW and TNF with the r-value of -0.487 and -0.485 in Campania and Puglia, respectively. TNF was also significantly positive correlated to YP in Campania (r=0.522).

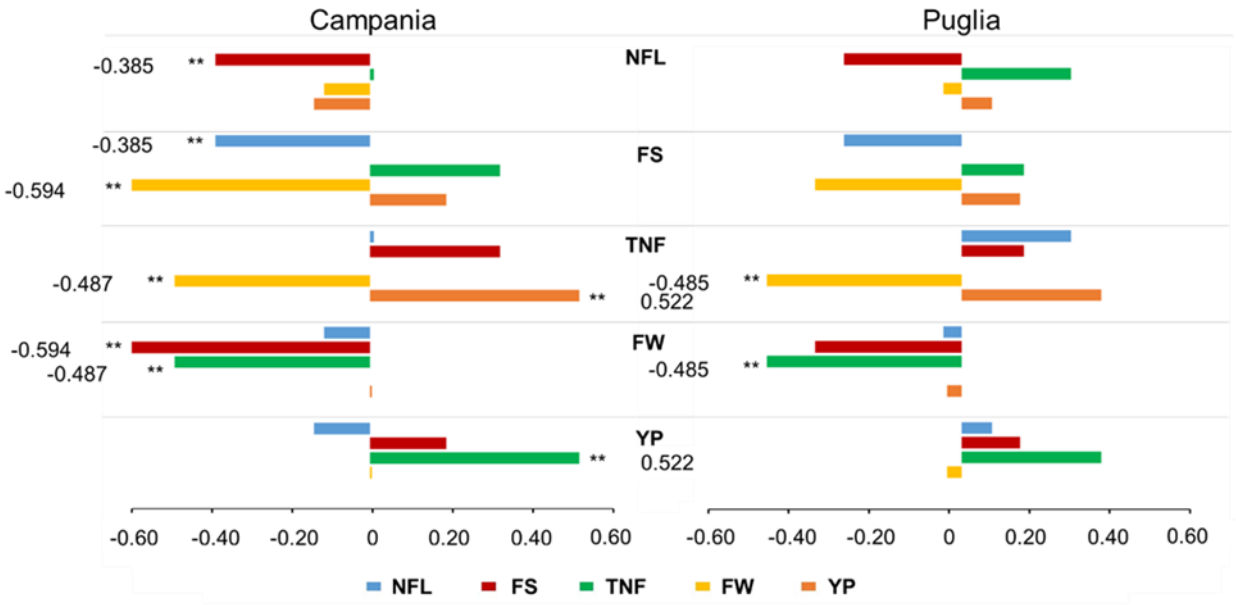


Figure 2.4 Pearson's correlation calculated among the five traits evaluated in two experimental fields. Each correlation block was normalized for each trait. Significance at *p<0.05; **p<0.01. NFL=no. flowers *per* inflorescence; FS=fruit set; TNF=No. fruit *per* plant; FW=fruit weight; YP=yield *per* plant.

Then, the selection index (SI) was calculated considering the traits showing a positive correlation with YP in both fields. As shown in Figure 2.4, two different traits (FS and TNF) showed a positive trend respect to YP in both conditions and were chosen for the determination of the SI. The selection index calculated for each field was scattered on Cartesian axes: on x axis the index of Campania field and on y axis the values for Puglia one (Figure 2.5). Based on SI values the genotypes were assigned to three different groups: in grey the genotypes with worst performances in both fields; in blue the genotypes with good performances only in a single field and in red the ones showing the best performances in both locations. The dashed lines represent the SI mean thresholds for each field. A wide range of SI values were recorded in Campania ranging from 4.03 (E17) to 69.30 (E42) with a

mean of 27.03, whereas the SI in Puglia were generally higher, with values ranging from 27.29 (E45) to 89.10 (E42) and a mean value of 54.17.

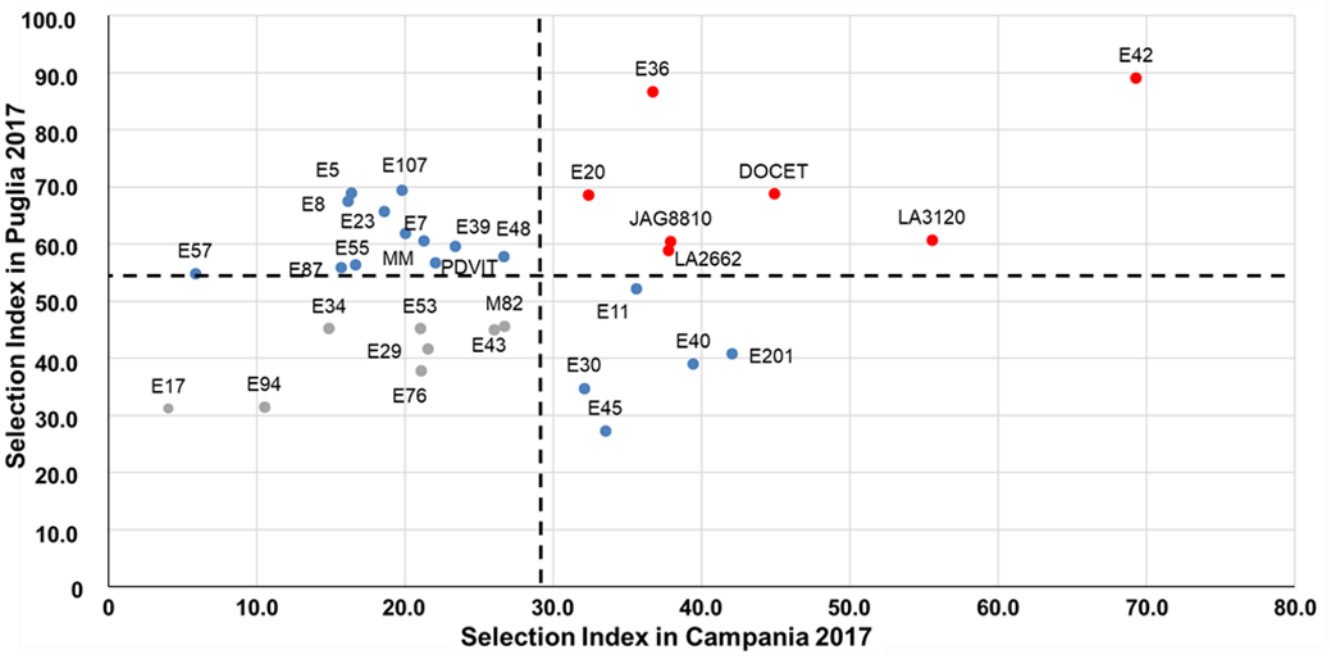


Figure 2.5 Selection Index (SI) estimated in the fields of Campania 2017 (x axis) and Puglia 2017 (y axis) on 28 genotypes and four controls. The SI was calculated on FS, TNF and YP data collected in two fields.

2.2.2 Phenotyping in the year 2018

During the year 2018, 19 genotypes including those showing the best SI (E20, E36, E42, LA2662 and LA3120), which fall in the same red group of the positive heat tolerant controls DOCET and JAG8810, and 12 genotypes showing a high SI only in one location were grown in Campania and the phenotypic analysis were carried out. The negative controls were not evaluated for the second year because they did not clearly show a susceptible phenotype during the year 2017 (Supplementary Figure S3). Climatic data were recorded for the second year as reported in Figure 2.6. In 39 days, the temperatures were higher than 32°C, while in 17 nights the maximum temperatures overcame 26°C. In order to validate data recorded in the previous year, the phenotypic analysis on three yield-related traits (TNF, FW and YP) was carried out in the 2018 in Campania and, then, the data were compared to those of the year 2017.

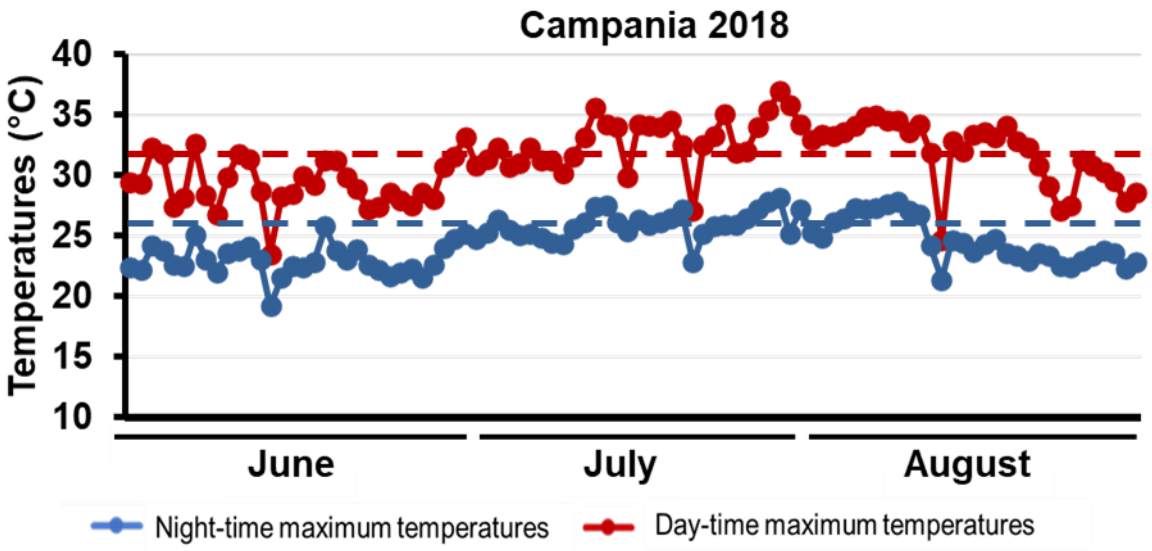


Figure 2.6 Day-time and night-time maximum temperatures recorded in Campania in the year 2018. In red the daily maximum temperature and in blue the night-time maximum temperatures. The blue dashed line represents the night critical threshold of 26°C and the red dashed line the day critical threshold of 32°C.

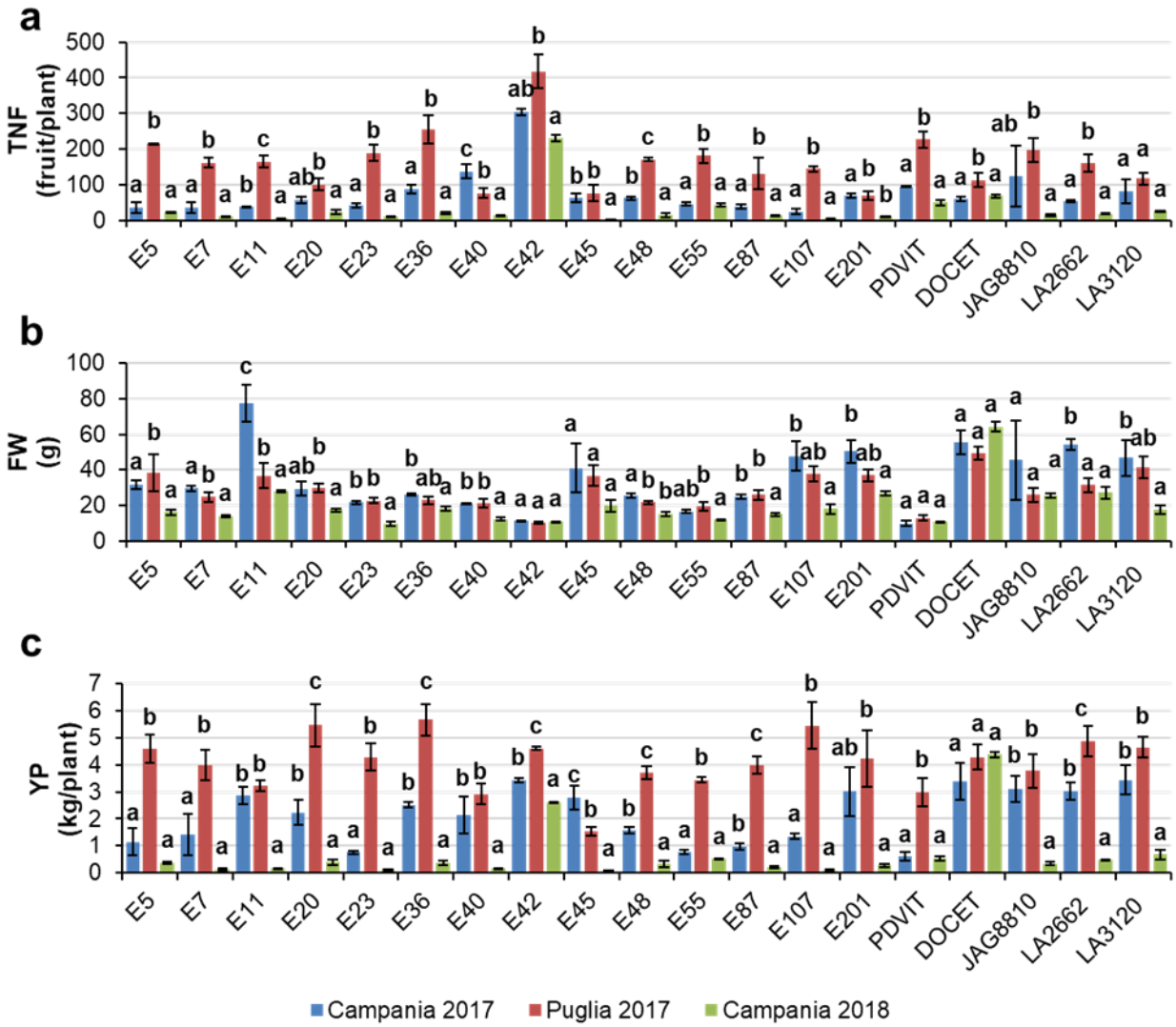
In figure 2.7 the results of Duncan's test carried out in the three experimental trials are reported.

As for TNF (figure 2.7a), higher TNF values were recorded in the experimental field in Puglia in the year 2017, whereas the lowest values were collected in Campania 2018. Most genotypes (14 out of 19 genotypes) showed values statistically similar comparing the Campania fields, while only LA3120 showed comparable values among all fields. Moreover, E42 showed the highest number of fruit *per* plant in all experimental trials.

As for FW (figure 2.7b), five genotypes (E42, E45, PDVIT, DOCET and JAG8810) showed FW values not statistically significant among different locations, whereas lower FW were generally recorded in Campania 2018.

Regarding YP (figure 2.7c), a general lowest yield *per* plant was recorded in Campania in the year 2018. Only the heat tolerant control DOCET showed yield *per* plant comparable among the different fields. Moreover, notwithstanding E42 showed statistical different values among the three fields, the production in Campania 2018 resulted higher than the mean population (0.64 kg/plant), with a YP value of 2.63 kg/plant, resulting the best genotype in Campania 2018, together with the control DOCET (Supplementary table S4).

Moreover, the two-way ANOVA analysis confirmed that for all traits statistically significant differences were found for each source of variation. Indeed, the TSS% showed that the highest contribution in the total variance was the Genotype (G) in for TNF and YP (42.28 % and 40.82 %, respectively) whereas for YP, the TSS% of Location resulted the highest (60.34 %).



Source of variation	TNF		FW		YP	
	TSS%	Sign.	TSS%	Sign.	TSS%	Sign.
Genotype	42.28	***	52.97	***	15.93	***
Location (L)	40.82	***	14.32	***	60.34	***
G x L	10.55	***	19.53	***	14.37	***
Error	6.36		13.18		9.36	

Figure 2.7 Two-way ANOVA carried out on three yield-related traits and Duncan's test on 17 genotypes and two controls. a) no. fruit *per* plant in three different locations, b) fruit weight in three different locations, c) yield per plant in three different locations.

At the end of the growing season the SI for the year 2018 was calculated and revealed that four genotypes and the heat tolerant control DOCET showed SI index above the mean of 12.72. The genotype E42 and the control DOCET showed the highest SI values (Figure 2.8).

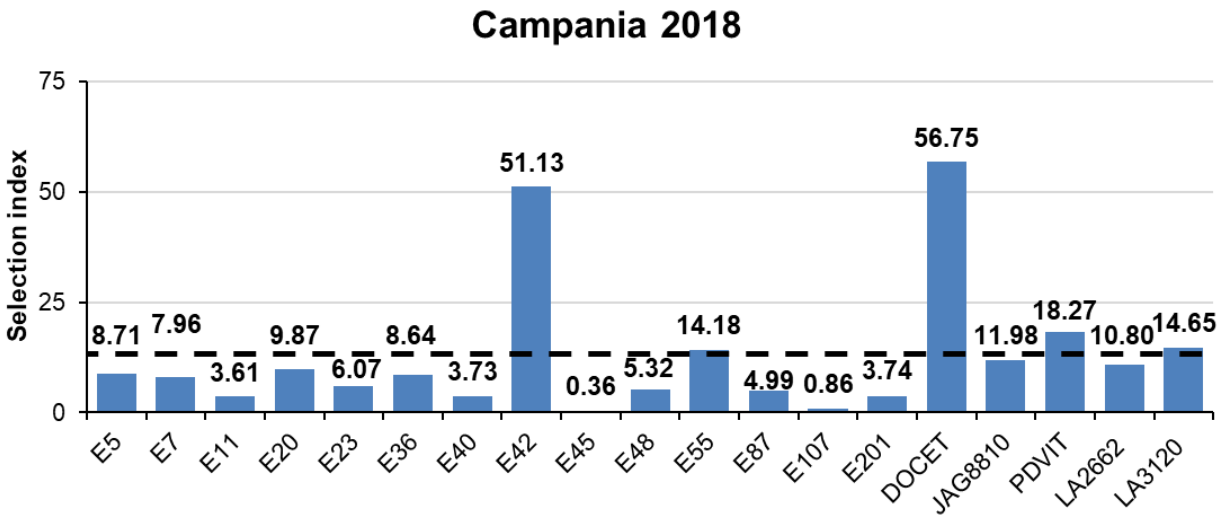


Figure 2.8 Selection Index (SI) estimated in Campania 2018 on 17 genotypes and two heat tolerant controls. The SI was calculated on TNF and YP.

2.2.3 Yield stability

The yield stability trait was evaluated by performing a linear regression analysis for the determination of the coefficient of regression (b_i^v) on a group of nine genotypes selected based on the results of the Selection Index (SI) in three trials. The analysis revealed that the coefficients of regression (b_i^v) ranged from 0.67 (PDVIT) to 1.34 (E36) with a mean value of 0.94 and a standard error of 0.23 (Table 2.3). Six out of nine genotypes showed a b_i^v ranging between the standard error thresholds values of 0.71 and 1.17 around the mean of 0.94; the optimum $b_i^v \simeq 1$ was reached in the genotypes E42 (b_i^v=0.85), LA2662 (b_i^v=0.90) and in the heat tolerant hybrids DOCET (b_i^v=0.89) and JAG8810 (b_i^v=1.03).

Table 2.3 Stability parameters of the linear regression analysis carried out on nine selected genotypes. b_i^v, coefficient of regression; ADL, average deviation from linearity; r^2, coefficient of determination of linear regression model, μ^v, yield *per* plant pooled mean, μ_i^v, yield *per* plant grand-mean.

Yield stability			E20		E36		E42		E55		PDVIT		LA2662		LA3120		DOCET		JAG8810	
Yield *per* plant (kg/plant)	μ_i^v	St. Err.	μ^v	St. Err.	μ^v	St. Err.	μ^v	St. Err.	μ^v	St. Err.	μ^v	St. Err.	μ^v	St. Err.	μ^v	St. Err.	μ^v	St. Err.	μ^v	St. Err.
	2.69	0.23	2.91	1.07	3.15	1.16	3.42	0.69	1.75	0.71	1.71	0.74	2.35	0.80	2.63	0.68	3.67	0.65	2.58	0.89
Stability parameters	b_i^v		1.26		1.34		0.85		0.76		0.67		0.90		0.76		0.89		1.03	
	ADL		0.01		0.00		0.08		0.04		0.03		0.00		0.05		0.05		0.00	
	r^2		0.97		0.93		0.57		0.81		0.59		0.89		0.88		0.71		0.94	

No genotypes showed an average deviation from the linearity (ADL) significantly higher than zero, confirming the consistency of b_i^v values. Moreover, also the coefficient of determination (r^2) showed values higher than 0.50, confirming the accuracy of the linear regression model carried out in the nine genotypes. The highest YP pooled mean (μ^v) among the genotypes was recorded in E42 (3.42

kg/plant) whereas DOCET showed the highest μ^v among all (3.67 kg/plant). The yield stability trait was estimated by plotting on a scatter plot the b_i^v value on y axis and the YP mean (μ_v) of each genotype on x axis (figure 2.9).

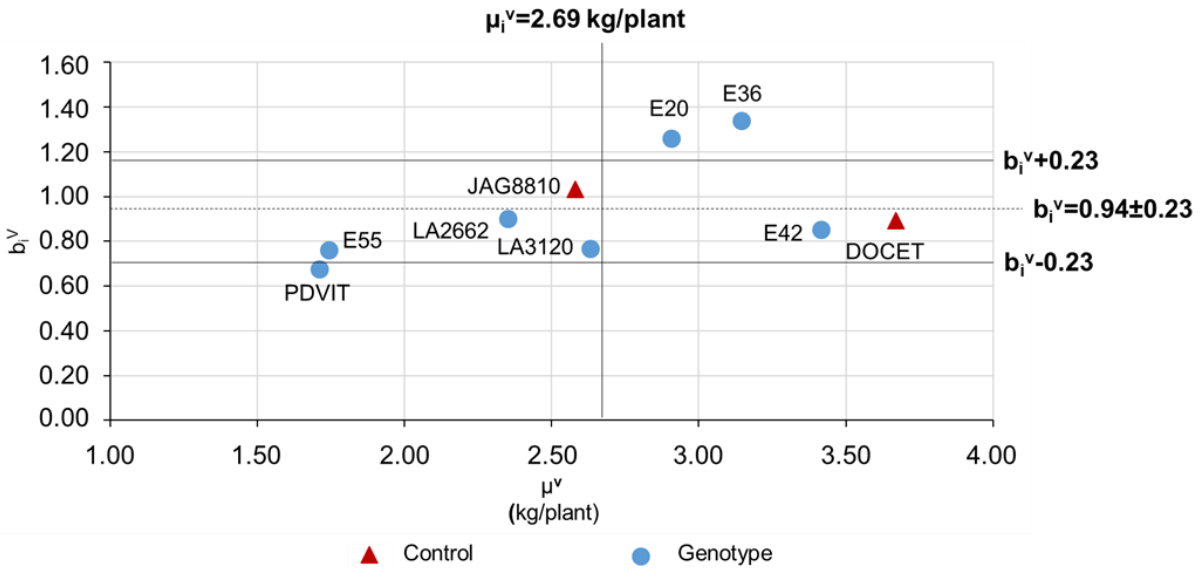


Figure 2.9 Diagram showing the stable-yielding index evaluated in the nine genotypes analyzed. Genotypes were plotted as their YP pooled mean values (μ^v) with the stability parameters, calculated as reported in Table 2.4. The horizontal dashed line represents the mean of b_i^v and the vertical one the mean of the YP grand mean (μ_i^v), considering all the values recorded in all trials. The continous lanes represent the range of b_i^v standard deviation.

The genotype showing at the same time a b_i^v value falling in the standard error range of coefficient of regression and the μ^v higher than the grand-mean of the population (μ_i^v) were regarded as high-stable yielding genotype. In this way, only E42 (3.42 kg/plant, b_i^v=0.85) and (DOCET (3.67 kg/plant; b_i^v=0.89) satisfied these conditions and were selected as high-stable yielding genotypes. Indeed, notwithstanding E20 (2.91 kg/plant, b_i^v=1.26) and E36 (3.15 kg/plant, b_i^v=1.34) showed a μ^v higher than μ_i^v, their coefficient of regression b_i^v was higher than 1 and fall out the standard deviation range of b_i^v, suggesting that these genotypes achieved high yield only in favorable environments. Moreover, E55 (1.75 kg/plant, b_i^v=0.76), LA2662 (2.35 kg/plant, b_i^v=0.90), LA3120 (2.63 kg/plant, b_i^v=0.76) and JAG8810 (2.58 kg/plant, b_i^v=1.03), showed a satisfying b_i^v value but a μ^v lower than the μ_i^v. For this reason, they were not selected as stable high- and stable-yielding genotypes.

2.2.4 Phenotyping of E42 under controlled high temperature conditions

From the analysis of linear regression model, E42 was selected for its good performances under high temperatures. In order to confirm its tolerance level this genotype was evaluated under controlled conditions at three different temperatures. During the experiment NFL, FS and TNF were measured, to evaluate the tolerance threshold of this genotype. NFL remained statistically comparable between

T1 (12.43) and T2 (10.88) whereas showed a significant decrease in T3 (8.47) respect to T1 (Figure 2.10). The value recorded in T3 is comparable with data collected in the intermediate temperature regime.

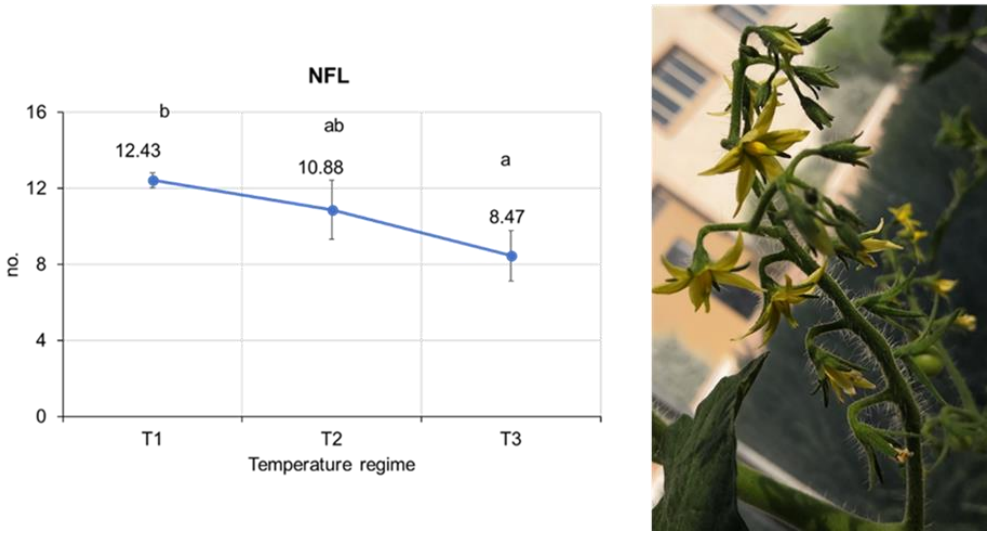


Figure 2.10 Trend of NFL values observed in three different temperature regimes. On the right, inflorescence of the genotype during the T2 treatment. T1=25°C/20°C day/night; T2=30°C/25°C day/night; T3=35°C/30°C day/night.

As for FS (Figure 2.11a), the decrease recorded in T2 (67.67) resulted not statistically significant compared to T1 (85.88) and the drop recorded in T3 (11.05) was statistically significant respect to previous temperature regimes. The TNF (Figure 2.11b) showed a significant decrease both in T2, and in T3, with a loss of ~42 % and ~93 % respectively.

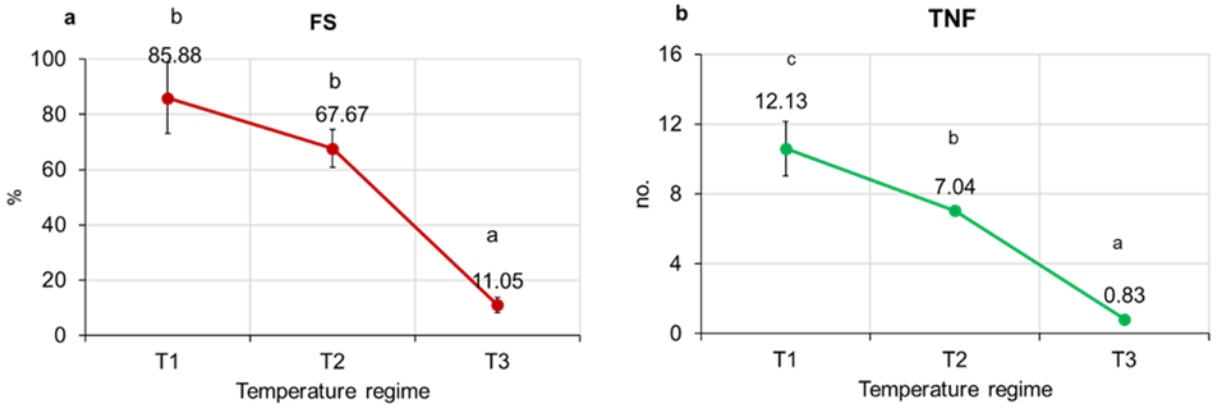


Figure 2.11 Trend of FS (a) and TNF (b) values observed in three different temperature regimes. a) fruit set (FS); b) no. fruit *per* plant (TNF). T1=25°C/20°C day/night; T2=30°C/25°C day/night; T3=35°C/30°C day/night.

3. Genotyping approach for the identification of molecular markers associated to genes/QTLs involved in tolerance to high temperatures

3.1 Material and methods

3.1.1 GWAS experiment

Starting from genotypic data used in previous studies performed at the Department of Agricultural Sciences (Ruggieri et al., 2014; Sacco et al., 2015) and phenotypic data collected before the beginning of the present work in the year 2016 in two regions of south Italy (Campania and Puglia), a GWAS approach was undertaken to identify genes/QTLs involved in traits affected by high temperatures (HT). Six different phenotypic traits were considered: flower earliness (FRL) number of flowers *per* inflorescence (NFL), fruit set (FS), number of fruit *per* plant (TNF) and yield *per* plant (YP). Genotyping data derived from 7.7K Infinium SolCAP single nucleotide polymorphism (SNP) array are available at http://dx.doi.org/10.6070/H4TT4NXN. The starting 7,720 SNPs were reduced to a final dataset of 1,800, applying filtering conditions, as the minimum allele frequency (MAF)>10%, and removing missing data more than 10%. Considering the expected homozygous condition of tomato, heterozygous variants were deleted from the analysis (Ruggieri et al., 2017). All these steps were carried out using the bioinformatic tools VcfTools v.0.1.13 (http://vcftools.sourceforge.net) and Tassel v. 5 (Bradbury et al., 2007; Danecek et al., 2011). The steps concerning the GWAS were reported in Ruggeri et al. (2019). In order to investigate also the genetic variability of genes putatively involved in the response to high temperatures, different genome databases were consulted for genes controlling flowering development and heat stress-related genes. In this way, the FLOR-ID database (Bouché et al., 2016) and the heat-stress section of the STIFDB V2.0 database (Naika et al., 2013) were considered for the exploration of candidate genes. Then, carrying out a BLAST analysis, the homologs in tomato were found by performing a global alignment, taking only matches with *p*-value $<1^{-10}$, the query coverage > 60% and the similarity > 60%. Moreover, genes related to fruit weight were retrieved from the literature (Hernández-Bautista et al., 2015; Ruggieri et al., 2014; Sacco et al., 2015). Finally, IntersectBed (Quinlan and Hall, 2010) was used to identify candidate genes of tomatoes that fell in the range of ±500 kbp from the associated markers.

3.1.2 Genotyping by sequencing experiment

The genomic DNA extraction of 21 genotypes was performed using DNeasy plant mini kit (Qiagen) following manufacturer's protocol. Young leaf tissue was grounded in order to obtain a fine powder. A quantity of 100 mg of powdered tissue was used for the extraction. The estimation of the concentration was performed using Qubit fluorometer (Invitrogen) whereas the quality parameters in

terms of 260/280 and 260/230 λ ratios were verified using the Nanodrop Spettrophotometer (Thermo Fischer). For DNA sequencing, 1 µg of extracted DNA was used for the constitution of ddRAD libraries with minor modifications (Peterson et al., 2012). The enzymes *MboI* and *SphI* were used for the double-digestion reaction. The fragments were sequenced using HiSeq2500 instrument (Illumina) at IgaTECH Technologies Inc..

Demultiplexing and manipulation of the reads were carried out by Stacks v. 2.0 (Catchen et al., 2013). The alignment of the reads was performed against *Solanum lycopersicum* version SL3.0 following BWA-MEM protocol (Li and Durbin, 2009), selecting only the mapped reads showing a quality score higher than 4. Finally, three different VCF files (one with SNP dataset, one with InDels and one merging two datasets) were generated using VcfTools v. 0.1.13 (http://vcftools.sourceforge.net). On each file a filtering step was performed, setting the max-missing values to 0.50 and imposing the minimum Depth of Coverage (min-mean DP) to 5. All the markers found on chromosome 0 were discarded. Then, the merged dataset was represented graphically performing a CircosWEB analysis (http://212.150.245.226/~tools/CircosVCF/) to visualize the genetic variability of assayed genotypes. For this assay the genotypes were grouped on *per*-private SNPs/InDels basis, obtaining three different graphs. Then, using SnpEff tool (http://snpeff.sourceforge.net/) the prediction of the magnitude of mutations was performed. In this way, four different tags categorizing the impact of mutation were established: “high”, if the mutation had a disruptive effect on predicted protein, “moderate” if the effect on protein was non-disruptive with a significant impact on the protein effectiveness, “low” if the mutation was a synonymous variant with low effect on protein functionality and “modifier” if mutation fell in non-coding regions. In order to understand the genetic differences among the assayed genotypes and selecting genotypes for breeding programs, the calculation of the Identity-by-State (IBS) sharing allele values was performed using the software Plink v. 1.8 (http://www.cog-genomics.org/plink2/) and plotted on a heat matrix. Then, in order to explore the genetic variability of the most polymorphic genotypes identified, a global alignment using tomato wild species databases (available at ftp://ftp.solgenomics.net/genomes/) was performed using the software Blast Genome Workbench v. 3.1.0 (https://www.ncbi.nlm.nih.gov/tools/gbench/). The selection of the InDels for this assay was carried out by choosing only those showing an insertion or deletion of at least five nucleotides. Finally, a sequence of about 100 bp, including the InDel mutation, was used as query for the analysis. The InDels were considered putative wild species-derived loci when showed the 100.00% of identities on the InDel region and more than 90% in the whole 100 bp sequence.

3.2 Results

3.2.1 Identification of QTLs by GWAS

Combining genotypic and phenotypic data previously collected on 79 genotypes, the identification of SNPs associated to the six phenotypic traits evaluated in two locations was performed. The two-way ANOVA analysis carried out on the six phenotypic traits revealed a significant genotype by location interaction. For this reason, two different GWAS analyses were carried out and common marker-associations were identified. Out of 35 marker-associations identified, nine were peculiar of flower earliness (FRL) and fruit weight (FW) and eight for number of fruit *per* plant (TNF). Moreover, in Campania specific marker were found to be associated to YP. In the same way in Puglia, no specific associations were observed for FS and NFL (Supplementary Table S5)

Table 3.1 Summary of the markers associated to phenotypic traits evaluated in two locations. The number of associations in each field and common to them and the chromosomes are reported.

Trait	ASSOCIATED MARKERS (No.)				Chr.
	Total	Common	Campania-specific	Puglia-specific	
FRL	9	0	6	3	1,3
NFL	3	1	2	0	3,4,8
FS	3	2	2	0	2,8
TNF	8	3	2	4	1,2,5,6,8
FW	9	8	0	1	3,5,6,12
YP	3	1	0	2	3,5,7

The total marker-trait associations identified mapped into 17 genes and 5 intergenic regions. Of these, 14 resulted to be common to both locations (Table 3.2), targeting 11 genes and three QTL regions. In particular, the marker CS11 associated to FW maps in the coding region of *Solyc01g008350*, which encodes for a MAPKK2 kinase. Since it is reported that the maximum average distance of LD decay in tomato is 500 kbp, (Sim et al., 2012), other candidate genes mapping within the distance of $\pm$ 500 kbp from each marker were explored. As for associations to NFL, marker CS1 showed a position in the genome close to the gene *Solyc08g067050* an homologue to *At1g04870*, which is involved in floral development in *A. thaliana*. Another interesting marker-trait association was found for FS and TNF, closely positioned to the gene *style2.1* (17.4 kbp), a transcription factor controlling the flower style length. Regarding FW, six markers (from CS6 to CS12) on chromosome 3 co-localize with two

major QTLs (fw3.2 and fw3.3), which control the fruit weight. The marker CS13 was co-localized with *Solyc12g009580* (homologue to YABBY) and *Solyc12g014130* (coding for a SUN protein). Finally, the marker CS14 co-localized with a minor QTL (fd7.1 and ty7.1) for yield on chromosome 7.

Table 3.2 Markers associated to six evaluated phenotypic traits detected in both experimental fields. For each marker an arbitrary code, the associated trait, SolCAP marker identity, the no. of chromosome, the marker position in SL3.0 genome version, the co-localized QTL and/or Candidate Genes (CGs) and the predicted function are reported.

Code	Trait	Common marker	Chr.	Marker Position	Associated QTL	Associated CGs	CG/QTL distance from marker	CG predicted function
CS1	NFL	solcap_snp_sl_18195	8	56,317,554		*Solyc08g067050*	285 Kbp	protein arginine methyltransferase
CS2	FS	solcap_snp_sl_10567	2	40,605,496		*Solyc02g070430*	121 Kbp	Gibberellic acid oxidase
CS3	FS/TNF	solcap_snp_sl_20344	2	38,133,120		*Solyc02g067380*	17.4 Kbp	Style2.1
CS4	TNF	solcap_snp_sl_8655	1	2,381,257				
CS5		solcap_snp_sl_8656	1	2,381,153		*Solyc01g008350*	22 Kbp	DNAJ-heat shock
CS6	FW	solcap_snp_sl_33830	3	71,244,368	*fw3.2; fw3.3*	*Solyc03g114940*	5 Mbp	Cytochrome P450
CS7		solcap_snp_sl_33829	3	71,244,962				
CS8		solcap_snp_sl_33828	3	71,245,034		*Solyc03g120190*	1.1 Mbp	SlOFP7 member of a gene family involed in fruit shape
CS9		solcap_snp_sl_33827	3	71,245,238		*Solyc03g120790*	697 Kbp	SlOFP8 member of a gene family involed in fruit shape
CS10		solcap_snp_sl_33822	3	71,293,803		*Solyc03g121760*	23 Kbp	SlSUN11 is part of the 34 member SUN-like gene family
CS11		solcap_snp_sl_36532	3	72,029,476		*Solyc03g123800*	0 Kbp	SlMAPKK2
CS12		solcap_snp_sl_25253	6	1,453,932				
CS13		solcap_snp_sl_59728	12	6,426,351		*Solyc12g009580*	3.6 Mbp	SlYABBY5b is a part of the YABBY gene family
						Solyc12g014130	1.5 Mbp	SlSUN33 is part of the 34 member SUN-like gene family
CS14	YP	solcap_snp_sl_70992	7	65,761,751	*fd7.1/ty7.1*			

3.2.2 Genotyping by sequencing analysis

In order to genotype 19 landraces and two hybrids, a GBS analysis was performed using ddRAD Illumina technology. The SNP calling generated a dataset of 147,409 unfiltered SNP/InDel markers. Transitions (Ts) were more frequent (57.1%) than transversion (Tv) (42.9%) with a Ts/Tv ratio of 1.3 (Figure 3.1). Then, mean depth of coverage of the sequencing was also estimated. In three cases (E36, LA3120 and PDVIT) the SNPs/Mean depth ratio was lower than 1K (0.85, 0.90 and 0.69, respectively), guaranteeing a good coverage of data recorded. Moreover, most of SNPs/InDels detected were homozygous (86.0%) whereas low values were recorded for SNPs/InDels in heterozygous condition (7.2%), and the 6.8% of missing data were observed. Applying the filtering settings described previously, the number of markers was reduced to 58,184 markers (50,036 SNPs and 8,148 InDels).

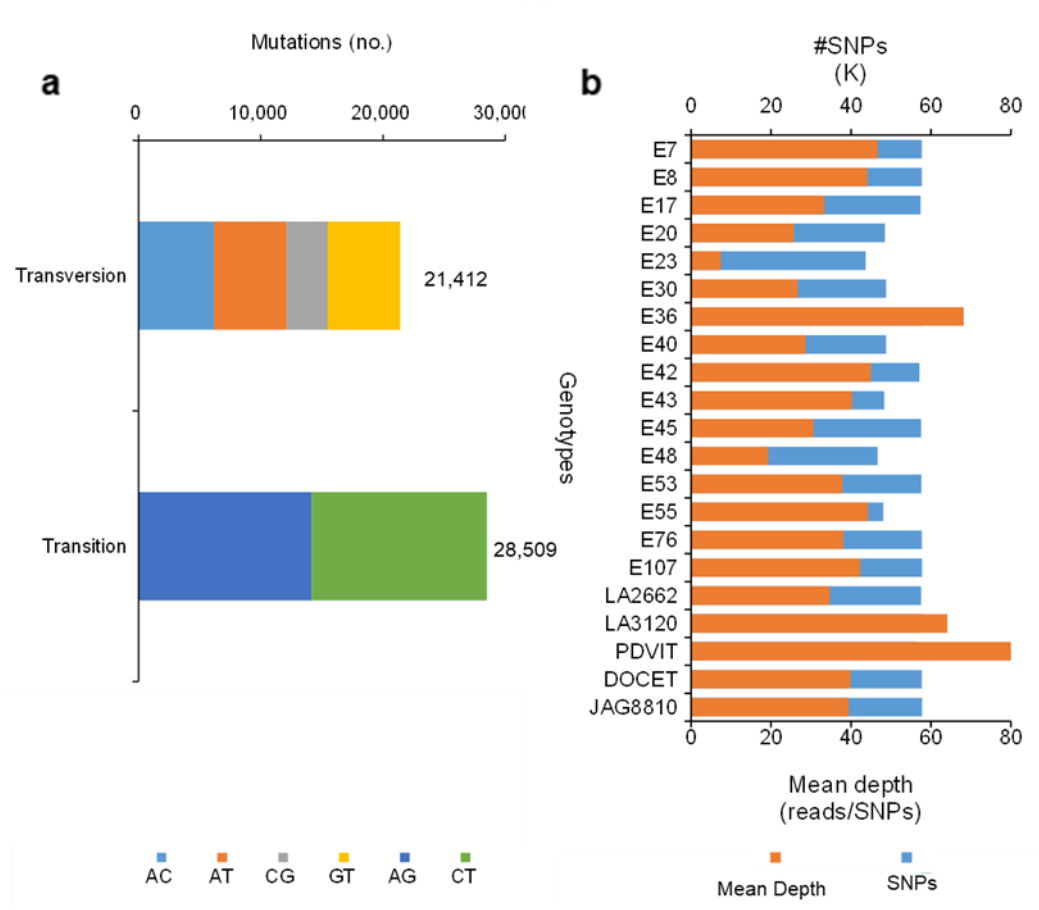


Figure 3.1 Number of Transition (Ts) and Transversion (Tv) detected in filtered SNPs/InDels dataset and evaluation of mean depth of coverage expressed as number of reads per SNP

As expected, the two hybrids DOCET and JAG8810 showed the highest percentage of heterozygous SNPs (24.2 and 34.2%, respectively) (Table 3.3). Observing the filtered 58,184 SNPs/InDels, which mapped throughout the genome, the stable-yielding genotype E42 showed the highest SNP density on chromosomes 1, 4, 7 and 12, whereas both the two hybrids exhibited higher SNP density on chromosomes 4 and 5 (Figure 3.2). In particular, E42 showed 8,018 private SNPs and 1,128 private InDels with the highest density on chromosomes 1 (3,501 SNPs and 416 InDels), and 7 (3,535 SNPs and 497 InDels), corresponding to about ~88% and ~81% of total private SNPs and InDels for this genotype, respectively (Supplementary table S6). In the same way, the genotype PDVIT showed 9,421 (7,980 SNPs and 1,441 InDels) private alleles and the highest SNP density was recorded on chromosome 5 (3,440 SNPs), 9 (1,794 SNPs) and 11 (1,541 SNPs). Moreover, a similar trend was recorded for the InDels density on the same chromosomes (526, 466 and 182 InDels, respectively) (Supplementary table S6). These two genotypes showed the highest number of private SNPs/InDels. In most cases private SNPs and InDels were distributed on the 12 chromosomes. The minimum number of private SNPs was recorded in E36 (78) whereas the minimum number of private InDels in E76 (13).

Table 3.3 Summary of total private SNPs/InDels in 19 genotypes and two hybrid controls. The percentage of homozygous, heterozygous and missing values for each genotype are also reported.

Mutations	E7	E8	E17	E20	E23	E30	E36	E40	E42	E43	E45	E48	E53	E55	E76	E107	LA2662	LA3120	PDVIT	DOCET	JAG8810	Total
Total SNPs	222	104	323	75	153	106	63	94	8,018	166	218	76	135	2,195	105	917	227	329	7,980	109	127	21,742
Total InDels	28	15	43	15	32	16	15	16	1,128	37	44	21	31	307	13	82	35	63	1,441	26	55	3,463
Total private	250	119	366	90	185	122	78	110	9,146	203	262	97	166	2,502	118	999	262	392	9,421	135	182	25,205
% homozygous	93.9	93.6	93.1	77.2	69.8	77.9	93.9	77.7	91.4	77.8	93.4	74.5	93.5	76.1	93.2	93.4	93.1	93.6	89.7	84.7	74.7	86.0
% heterozygous	5.3	5.6	5.7	6.2	5.2	6.0	5.6	6.2	6.6	5.3	5.5	5.9	5.5	6.5	6.0	6.0	5.8	5.7	6.8	14.5	24.6	7.2
% missing	0.8	0.8	1.2	16.6	24.9	16.1	0.5	16.1	1.9	16.9	1.1	19.6	1.0	17.4	0.7	0.7	1.1	0.7	3.6	0.8	0.6	6.8

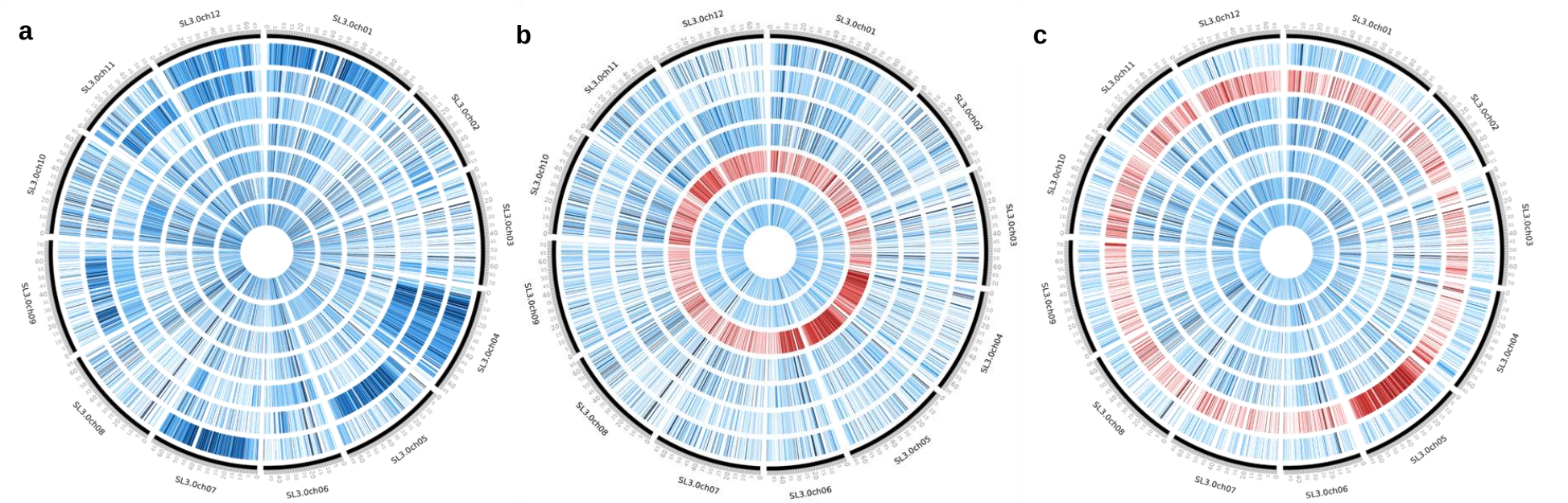


Figure 3.2 Genetic variability evaluated using CircosWEB tools. Three different outputs were generated based on the number of private SNPs/InDels detected. From the outer to inner ring: a) E42, PDVIT, E55, E107, LA3120, E17 and E45. b) LA2662, E7, E43, E23, JAG8810, E53, E30; c) E8, DOCET, E76, E40, E48, E20, E36. Data related to JAG8810 and DOCET are evidenced in red.

Moreover, in order to combine the information deriving from GWAS and the GBS sequencing data, regions containing polymorphisms detected by the SolCAP platform and associated to yield-related traits were explored to search for polymorphisms in the GBS dataset. For this analysis genomic portions of 500 kbp upstream and downstream of SolCAP markers were considered. In particular, in the region containing the markers CS4 and CS5, associated to TNF, 109 SNP/InDels were identified and 49 of these showed allelic variants private of E42. Similarly, in the region containing the marker CS14 associated to YP, 49 SNPs/InDels were found and 41 of these showed allelic variants private of E42 (Table 3.4).

Table 3.4 Number of SNPs/InDels detected by GBS analysis in the region flanking each SolCAP marker associated to yield-related traits. The number of genes, the number of SNPs/InDels and the E42's private SNPs/InDels found in the considered region are also reported.

Marker	SolCAP Marker	GBS dataset		
		Genes (no.)	SNPs/InDels (no.)	E42 (no.)
CS1	solcap_snp_sl_18195	10	14	0
CS2	solcap_snp_sl_10567	25	79	0
CS3	solcap_snp_sl_20344	19	40	1
CS4	solcap_snp_sl_8655	23	109	49
CS5	solcap_snp_sl_8656	23	109	49
CS6	solcap_snp_sl_33830	7	10	6
CS7	solcap_snp_sl_33829	7	10	6
CS8	solcap_snp_sl_33828	7	10	6
CS9	solcap_snp_sl_33827	7	10	6
CS10	solcap_snp_sl_33822	7	10	6
CS11	solcap_snp_sl_36532	10	49	11
CS12	solcap_snp_sl_25253	19	64	0
CS13	solcap_snp_sl_59728	31	121	0
CS14	solcap_snp_sl_70992	25	49	41

Then, the SnpEff analysis carried out on the filtered dataset allowed the prediction of mutations impact on biological activity of the corresponding protein (Table 3.5). Out of 50,036 SNPs, allelic variants with a low impact corresponded to approx. 1.2%, whereas those with modifier and moderate effect were 96.9% and 1.8 %, respectively. In the same way, out of 8,148 InDels, mutations with low impact were 0.8%, whereas the 96.7% and 9.2% corresponded to mutations with modifier and moderate effect, respectively. Regarding the mutations with high impact, these covered only the 0.2% (0.1 % for SNPs and InDels).

Table 3.5 Prediction of the SNPs/InDels mutation effect where the impact of the mutation was categorized with four different tags: high, moderate, low and modifier effect. The total number of genes affected by one or more mutations is also reported.

Type of mutation	Total mutations	Effect on protein				Affected genes
		High	Moderate	Modifier	Low	
SNPs	50,036	52	899	48,471	614	7,545
InDels	8,148	55	149	7879	65	3,498
Total	58,184	107	1,048	3,756	56,350	4,814

The genotyping analysis revealed that all mutations targeted ~4,800 genes, and 107 mutations with effect on protein function were found (52 SNPs and 55 InDels variants), which affect 88 genes (Table 3.6) distributed on the twelve chromosomes. Moreover, some of them could play an important role in the responses to abiotic stress. Two of these (*Solyc02g087680* and *Solyc02g087690)* code for FACT complex subunit SSRP1, whereas *Solyc12g044645* codes for a member of AP2/B3 transcription factor family protein. Interestingly, three out of 52 SNPs and two out of 55 InDels resulted private markers of E42, whereas ten (6 SNPs and 4 InDels) were private of PDVIT. The mutations with moderate impact were 1,048 (899 SNPs and 149 InDels): they affected 503 genes, and ~8% and ~8.5 % of these were private variants of E42 and PDVIT, respectively.

Table 3.6 List of genes affected by SNP/InDel mutations with high impact. For each gene, the reference and alternative alleles are reported, where the reference is Heinz allele.

Chrom.	SNP/InDel position	Reference allele (5→3)	Alternative allele (5→3)	Gene	Mutation type	Protein function
SL3.0ch01	2,490,155	A	T	*Solyc01g008471*	stop_gained	histone-lysine N-methyltransferase SUVR5
SL3.0ch01	23,808,388	T	G	*Solyc01g017200*	stop_lost; splice_region_variant	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
SL3.0ch01	23,860,038	TAATGGAG	TAATGGAGAG	*Solyc01g017370*	frameshift_variant	DNA-directed RNA polymerase subunit beta; contains RNA polymerase Rpb1, domain 1
SL3.0ch01	39,016,299	AGAC	GGAT,GGAC	*Solyc01g028980*	stop_gained	Gamma-tubulin complex component
SL3.0ch01	97,226,004	C	T	*Solyc01g110688*	splice_acceptor_variant; intron_variant	CTP synthase
SL3.0ch02	3,958,213	T	C	*Solyc02g049106*	stop_lost; splice_region_variant	Leucine-rich repeat protein kinase family protein
SL3.0ch02	20,416,788	C	T	*Solyc02g014877*	stop_gained	lectin receptor kinase a4.1
SL3.0ch02	25,963,019	G	T	*Solyc02g030300*	stop_gained	Serine/threonine-protein kinase receptor
SL3.0ch02	34,319,452	AGA	AA	*Solyc02g062190*	frameshift_variant	Allinase
SL3.0ch02	38,310,985	A	G	*Solyc02g067590*	stop_lost; splice_region_variant	DNA-DIRECTED RNA POLYMERASE SUBUNIT BETA
SL3.0ch02	39,312,995	CAA	CA	*Solyc02g068830*	frameshift_variant	Receptor kinase-like protein contains Serine/threonine protein kinase
SL3.0ch02	46,825,407	T	A	*Solyc02g082630*	splice_donor_variant; intron_variant	PHOSPHATIDYLINOSITIDE PHOSPHATASE SAC1
SL3.0ch02	49,909,435	G	A	*Solyc02g086610*	splice_acceptor_variant; intron_variant	Isocitrate dehydrogenase-like protein; containsIsocitrate dehydrogenase NADP-dependent
SL3.0ch02	50,622,904	TGG	TGGG	*Solyc02g087620*	splice_donor_variant; intron_variant	Inositol hexakisphosphate and diphosphoinositol-pentakisphosphate kinase 2; contains Histidine acid phosphatase
SL3.0ch02	50,664,745	CGT	CT	*Solyc02g087680*	frameshift_variant	FACT complex subunit SSRP1
SL3.0ch02	50,665,049	A	T	*Solyc02g087690*	stop_lost; splice_region_variant	FACT complex subunit SSRP1
SL3.0ch03	1,944,628	GTT	GTTT	*Solyc03g007400*	frameshift_variant	Galactose oxidase/kelch repeat superfamily protein
SL3.0ch03	18,084,656	T	A	*Solyc03g046596*	splice_acceptor_variant; intron_variant	BCL-2-associated athanogene 6
	18,087,189	A	G		splice_donor_variant; intron_variant	
SL3.0ch03	18,418,311	AGG	AG	*Solyc03g070435*	frameshift_variant	Alpha-mannosidase
SL3.0ch03	33,847,933	TGG	TTG,TG	*Solyc03g061655*	frameshift_variant	Ribosomal protein S12
SL3.0ch03	64,961,947	CGG	CG	*Solyc03g113310*	frameshift_variant	Pseudouridine synthase family protein
SL3.0ch03	66,978,174	TCC	TC	*Solyc03g115910*	frameshift_variant	MADS-box transcription factor 1; contains Transcription factor, MADS-box
SL3.0ch04	1,873,263	TCG	TG	*Solyc04g008200*	frameshift_variant	LES457048 hero resistance protein 1 homologue
SL3.0ch04	12,538,449	A	C	*Solyc04g039830*	stop_lost; splice_region_variant	Acetyl-CoA carboxylase beta subunit
SL3.0ch04	39,905,377	ATGCG	GTGCA	*Solyc04g049170*	splice_acceptor_variant; splice_region_variant; intron_variant	Dehydrin 9
SL3.0ch04	48,811,041	T	A	*Solyc04g050890*	stop_gained	DNA-DIRECTED RNA POLYMERASE SUBUNIT ALPHA
SL3.0ch05	6,223,628	T	A	*Solyc05g013120*	stop_gained	Ninja-family protein AFP1
SL3.0ch05	11,438,471	A	G	*Solyc05g015570*	start_lost	UDP-glucose 6-dehydrogenase 1
SL3.0ch05	17,404,131	CTGTGTGT	CTGTGT	*Solyc05g016525*	frameshift_variant	Core-2/I-branching beta-1
SL3.0ch05	25,821,654	A	C	*Solyc05g020010*	stop_lost; splice_region_variant	Photosystem I P700 chlorophyll a apoprotein A2
SL3.0ch05	27,336,098	G	T	*Solyc05g021410*	stop_gained	histone-lysine N-methyltransferase SUVR5
	27,336,101	A	T		stop_gained	
	27,336,154	G	A		stop_gained	
	27,336,462	A	T		stop_gained	
	27,336,522	CAA	CA		frameshift_variant	
	27,336,692	C	A		stop_gained	
	27,336,697	G	T		stop_gained	
SL3.0ch05	33,154,148	GAC	GC	*Solyc05g025530*	frameshift_variant	DNA-directed RNA polymerase subunit beta
SL3.0ch05	33,160,167	A	T	*Solyc05g025540*	stop_gained	Molybdenum cofactor sulfurase
SL3.0ch05	36,720,012	GTTTTTTTTTC	GTTTTTTTTTTC	*Solyc05g025780*	frameshift_variant	receptor-like protein kinase 3
SL3.0ch05	54,250,703	C	A	*Solyc05g041700*	stop_gained	Auxin efflux carrier component
SL3.0ch05	58,428,505	G	A	*Solyc05g045790*	splice_donor_variant; intron_variant	CYTOCHROME C OXIDASE SUBUNIT 2

Table 3.6 Continued

Chrom.	SNP/InDel position	Reference allele (5→3)	Alternative allele (5→3)	Gene	Mutation type	Protein function
SL3.0ch06	919,488	AG	AGAACG	*Solyc06g005930*	frameshift_variant	Sensitivity to red light reduced protein 1
SL3.0ch06	3,336,382	G	A	*Solyc06g009415*	stop_gained	MYB transcription factor
SL3.0ch06	11,052,798	G	A	*Solyc06g011663*	stop_gained	beta glucosidase 25
	11,052,789	C	A		stop_gained	
	11,054,142	TA	TGA		frameshift_variant	
	11,054,328	C	T		stop_gained	
	11,054,818	A	T		stop_gained	
SL3.0ch06	11,106,015	CTGTGTGTGC	CTGTGTGC	*Solyc06g025415*	frameshift_variant	Mediator of RNA polymerase II transcription subunit 14
SL3.0ch06	11,249,266	A	G	*Solyc06g024380*	splice_acceptor_variant; intron_variant	DNA polymerase epsilon catalytic subunit
SL3.0ch06	11,254,310	TGGGGGGC	TGGGGGGGC	*Solyc06g024386*	frameshift_variant	Alpha-mannosidase
SL3.0ch06	11,650,561	G	A	*Solyc06g024200*	stop_gained	BSD domain-containing protein
SL3.0ch06	12,303,996	AGC	AGGC	*Solyc06g024203*	frameshift_variant; splice_region_variant	BSD domain-containing protein
SL3.0ch06	30,374,560	AG	ATG	*Solyc06g043038*	frameshift_variant	Ycf68
SL3.0ch06	49,741,555	CA	CTA	*Solyc06g084626*	frameshift_variant	1-deoxy-D-xylulose 5-phosphate reductoisomerase
	49,741,558	AA	ATA	*Solyc06g084626*		
	49,741,771	G	A	*Solyc06g084626*		
SL3.0ch07	4,478	TG	GT	*Solyc07g004993*	start_lost	Phosphatidylinositol N-acetyglucosaminlytransferase subunit P-like protein
	4,524	C	G		stop_gained	
	5,004	G	C		splice_acceptor_variant; intron_variant	
	5,025	AA	TG		stop_gained	
SL3.0ch07	809,567	C	T	*Solyc07g005970*	stop_gained	Autophagy-related protein 11
SL3.0ch07	1,795,652	TGT	TT	*Solyc07g007000*	frameshift_variant	Protein trichome birefringence
SL3.0ch07	7,582,232	ACC	ACCC	*Solyc07g017575*	frameshift_variant	Flavin-containing monooxygenase
SL3.0ch07	17,487,354	C	G	*Solyc07g021370*	stop_lost; splice_region_variant	DNA-directed DNA polymerase
SL3.0ch07	58,909,896	C	T	*Solyc07g045613*	stop_gained	Ggamma-subunit 1
SL3.0ch07	62,107,933	CTT	CTTT	*Solyc07g053565*	frameshift_variant	NAD
SL3.0ch07	64,634,848	CTTTTTTTTTTG	CTTTTTTTTTTTTG	*Solyc07g056700*	frameshift_variant	Armadillo/beta-catenin-like repeat family protein
SL3.0ch08	934,013	T	C	*Solyc08g006252*	stop_lost; splice_region_variant	BR enhanced expression 1
SL3.0ch08	7,149,964	AGA	AA	*Solyc08g016275*	frameshift_variant	Ulp1 protease family protein
SL3.0ch08	39,213,014	T	G	*Solyc08g028915*	stop_lost	DNA-directed RNA polymerase subunit beta"
SL3.0ch08	57,492,141	CA	CATATA	*Solyc08g068330*	frameshift_variant	Aspartate aminotransferase
SL3.0ch09	5,231,913	TCT	TT	*Solyc09g011937*	frameshift_variant	Ulp1 protease family protein
SL3.0ch09	27,883,419	A	G	*Solyc09g031830*	stop_lost; splice_region_variant	Protein Ycf2
SL3.0ch09	34,670,992	C	T	*Solyc09g048970*	stop_gained	DNA-directed RNA polymerase II subunit RPB7
SL3.0ch10	1,696,465	CAAT	TAAG	*Solyc10g007300*	stop_gained	succinate dehydrogenase 2-1
SL3.0ch10	6,143,924	ATGC	ACGA	*Solyc10g017890*	start_lost	Photosystem I P700 chlorophyll a apoprotein A1
SL3.0ch10	6,148,795	GAAAAAT	GAAAAT	*Solyc10g017920*	frameshift_variant	DNA-directed RNA polymerase subunit beta
SL3.0ch10	26,313,660	CTATC	CTG	*Solyc10g044430*	frameshift_variant; synonymous_variant	FMN-linked oxidoreductases superfamily protein
SL3.0ch10	40,732,132	GTT	GTTT	*Solyc10g047410*	frameshift_variant	PHOTOSYSTEM II CP43 REACTION CENTER PROTEIN
SL3.0ch10	40,737,931	TCT	TT	*Solyc10g047430*	frameshift_variant	RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN, CATALYTIC DOMAIN-CONTAINING PROTEIN
SL3.0ch10	51,775,257	AGA	AA	*Solyc10g051273*	frameshift_variant	Ulp1 protease family
SL3.0ch10	56,116,455	A	T	*Solyc10g054967*	stop_gained	Mediator of RNA polymerase II transcription subunit 20-like protein
SL3.0ch10	63,076,669	C	A	*Solyc10g082055*	stop_gained	Beige/BEACH and WD40 domain-containing protein
SL3.0ch10	63,080,118	C	T	*Solyc10g082065*	stop_gained	Nuclear transport factor 2 family protein with RNA binding domain isoform 2
	63,080,121	GCC	GC		frameshift_variant	
	63,080,127	A	T		stop_gained	

Table 3.6 Continued

Chrom.	SNP/InDel position	Reference allele (5→3)	Alternative allele (5→3)	Gene	Mutation type	Protein function
SL3.0ch11	7,886,588	G	A	*Solyc11g017090*	stop_gained	Serine/threonine protein phosphatase 7 long form isogeny
SL3.0ch11	9,702,107	TCA	TA	*Solyc11g018853*	frameshift_variant	Adenylate isopentenyltransferase
SL3.0ch11	9,725,431	GCG	ACAA	*Solyc11g019880*	frameshift_variant; missense_variant	Adenylate isopentenyltransferase
SL3.0ch11	10,916,212	G	A	*Solyc11g020345*	splice_acceptor_variant; intron_variant	Small nuclear ribonucleoprotein family protein
SL3.0ch11	11,099,900	CTT	TTG,TTT	*Solyc11g020358*	stop_gained	Adenylate isopentenyltransferase
SL3.0ch11	13,390,388	T	G	*Solyc11g021130*	splice_acceptor_variant; intron_variant	28S RIBOSOMAL PROTEIN S7, MITOCHONDRIAL
SL3.0ch11	13,404,678	A	T	*Solyc11g021190*	stop_lost; splice_region_variant	NAD
SL3.0ch11	13,408,356	TGG	TG	*Solyc11g021210*	frameshift_variant	CYTOCHROME B/B6 PROTEIN-RELATED
SL3.0ch11	23,584,770	A	G	*Solyc11g030910*	splice_donor_variant; intron_variant	Leucine-rich repeat protein kinase family protein
SL3.0ch11	31,528,406	A	T	*Solyc11g045260*	splice_donor_variant; intron_variant	Photosystem II CP43 reaction center protein
SL3.0ch11	45,625,653	C	A	*Solyc11g056340*	stop_lost; splice_region_variant	PHOTOSYSTEM II PROTEIN D1
SL3.0ch11	48,875,863	CAG	TAA	*Solyc11g061818*	stop_gained	Adenylate isopentenyltransferase
SL3.0ch11	49,163,841	A	C	*Solyc11g062050*	stop_lost; splice_region_variant	Protein Ycf2
SL3.0ch12	22,655,957	ATT	AT	*Solyc12g076357*	frameshift_variant	50S ribosomal protein L22
SL3.0ch12	51,497,675	TATGTCTCTAGGTATATTCTCTCT TCGGCAAAAGACTCTTCGGTGA	TA	*Solyc12g038540*	frameshift_variant; splice_acceptor_variant; splice_region_variant; intron_variant	Transducin/WD40 repeat-like superfamily protein
SL3.0ch12	52,525,425	GTAATT	GTAATTAATT	*Solyc12g038970*	frameshift_variant	EMB1873 protein
SL3.0ch12	60,656,260	GAAAAAAAAAAG	GAAAAAAAAAAAAG	*Solyc12g044645*	frameshift_variant	AP2/B3 transcription factor family protein
	60,656,273	T	C		stop_lost; splice_region_variant	

The relationships among the 21 genotypes were investigated calculating Identity-by-State (IBS) allele-sharing distance matrix for all the pairwise comparisons using the filtered dataset of 58,184 SNP/InDels. As reported in figure 3.3, the distribution of the pairwise comparisons revealed that most of the paired genotypes showed an IBS ranging from 0.91 and 0.96 (116 comparisons out of 210) and, interestingly, 36 pairs of genotypes showed an IBS value ranging from 0.60 and 0.70.

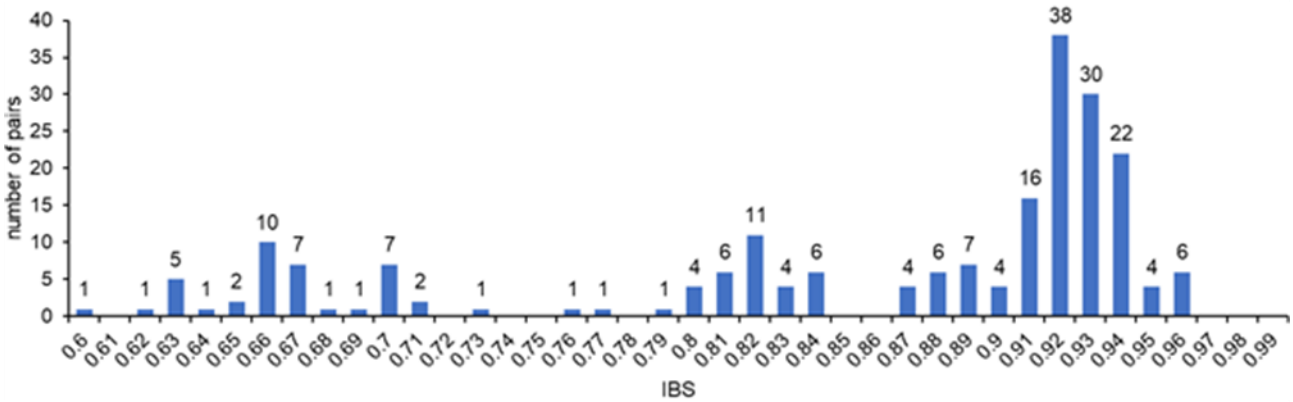


Figure 3.3 Distribution of Identity-by-state sharing values among 21 tomato genotypes determined by analyzing the 58,184 Snp/InDels dataset.

In order to exploit the all pairwise comparisons, the IBS values were plotted in a heat matrix (Figure 3.4). The analysis revealed that in all comparisons involving the genotypes E42, E55, PDVIT and the hybrid controls DOCET and JAG8810, the IBS value did not exceed 0.89, whereas among all other comparisons this value was higher than 0.90. Moreover, among the three genotypes and two hybrids, E55, DOCET and JAG8810 showed intermediate values, ranging from 0.65 and 0.89. Interestingly, the comparisons involving the genotypes E42 and PDVIT showed the lowest IBS values, corresponding to the 36 pairs of genotypes with lowest IBS values. Then, 62 pair comparisons resulted intermediate ranging from 0.71 to 0.80 (10) and from 0.81 to 0.90 (52) while most of the comparisons fell between 0.91 and 1.00 (106). These results evidenced that there is a high genetic distance comparing E42 and PDVIT genotypes against the others. Moreover, the lowest IBS value was recorded comparing E42 and PDVIT (0.60), evidencing a high genetic distance also among them.

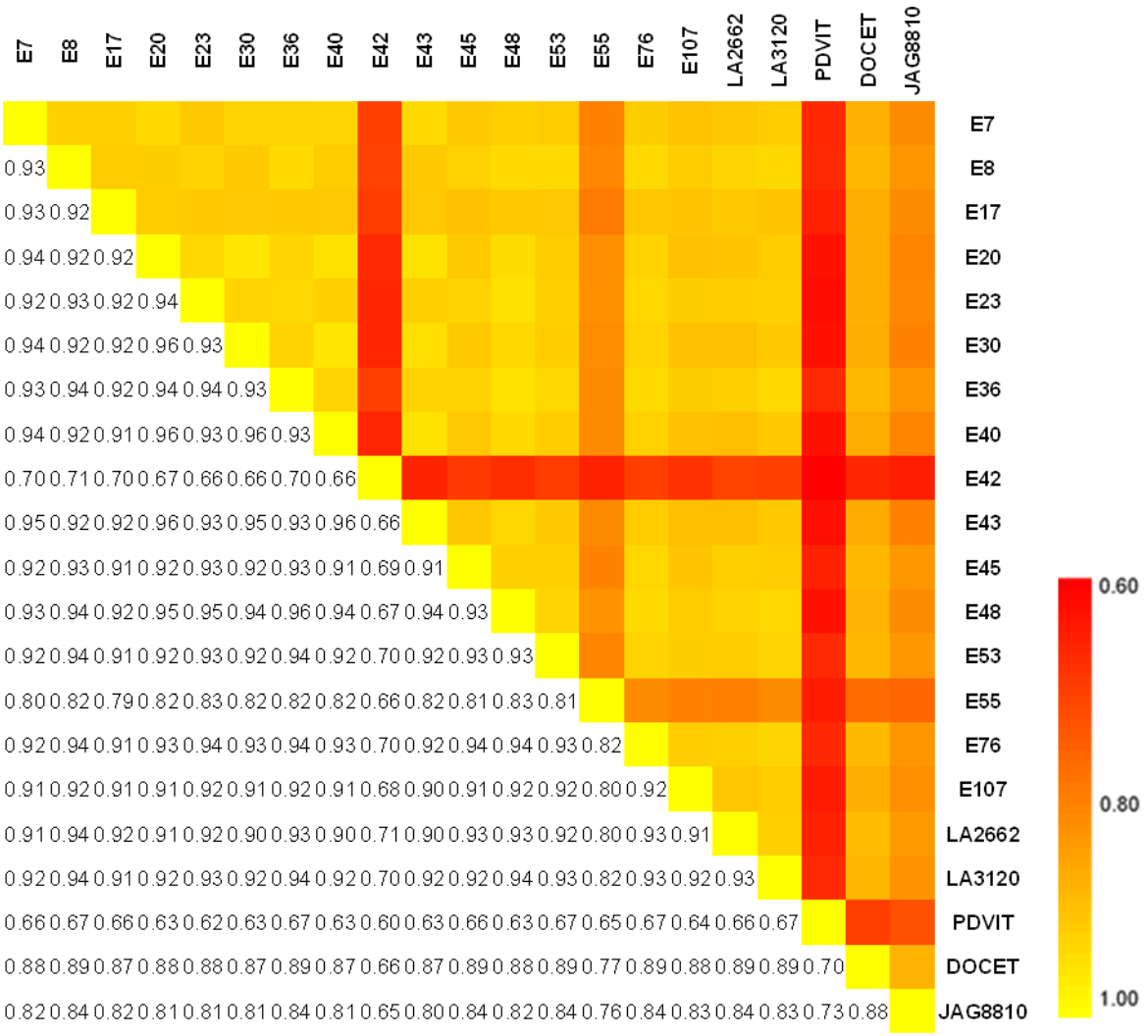


Figure 3.4 Heat matrix describing the Identity-by-state (IBS) pair-wise comparisons among the 21 assayed genotypes. In the color scale the red represents the lower IBS value (more genetic distance) whereas the yellow represents the higher IBS value (less genetic distance).

Indeed, the number of SNPs/InDels common between these two genotypes corresponds to ~4% of the total dataset of 58,184 SNPs/InDels and, only on chromosome 12, 1,811 common SNPs/InDels were identified (Table 3.7). In this way, in order to better understand the high genetic variability detected in E42 (Figure 3.5a) and PDVIT (Figure 3.5b), InDel mutations that resulted private for E42 or PDVIT and common between them were selected.

Table 3.7 Summary of the common markers identified comparing E42 and PDVIT for each chromosome.

Private	Ch01	Ch02	Ch03	Ch04	Ch05	Ch06	Ch07	Ch08	Ch09	Ch10	Ch11	Ch12	Total
E42	3,917	70	47	440	55	38	4,032	35	81	122	119	190	9,146
Common	35	74	6	152	15	7	43	7	7	18	8	1,811	2,183
PDVIT	257	346	77	212	3,966	252	23	87	2,260	72	1,723	146	9,421

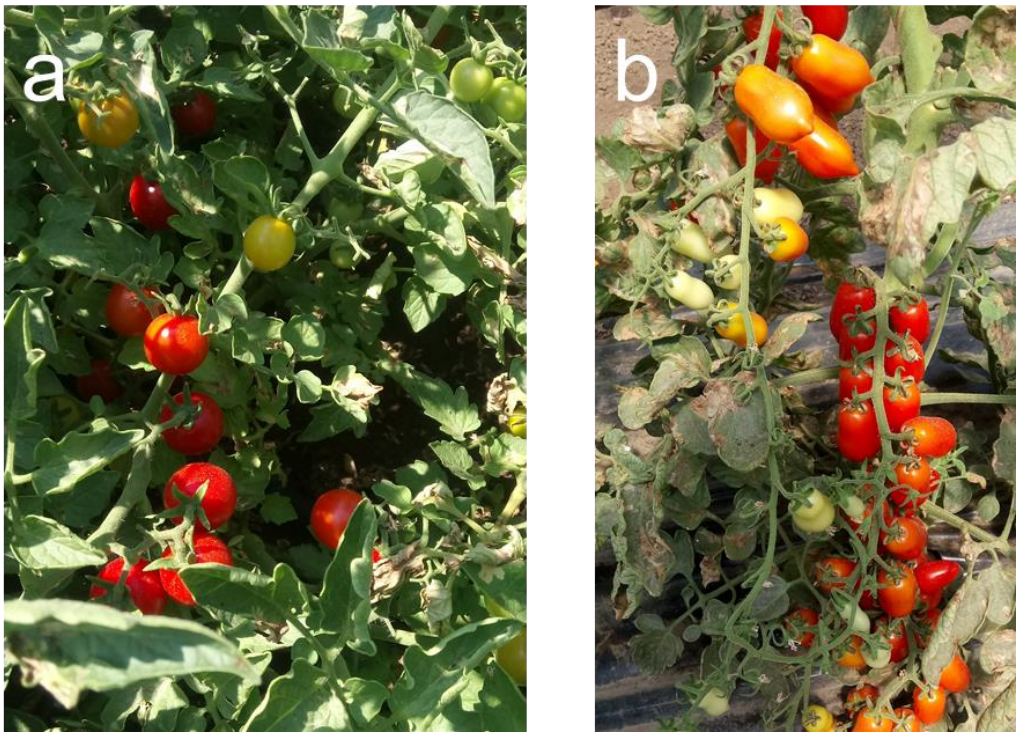


Figure 3.5 Fruit morphology of the two highest polymorphic genotypes. a) E42; b) PDVIT

Blast analysis was performed aligning the selected InDels and the genomic sequences of five wild species (*S. chilense, S. galapagense, S. pennellii, S. peruvianum* and *S. pimpinellifolium*). At this purpose, 82 InDels were chosen and for each of them an approximately 100 bp sequence was selected, including the InDel, to carry out the BLAST analysis (Table Supplementary S7). For each sequence, both the reference and the alternative alleles were aligned against the five tomato wild genomes. The analysis revealed a total of 186 matching sequences (Figure 3.6). Twenty-four assayed markers showed higher identity scores with different tomato wild genome only for the reference sequence, 25 only with the alternative one and for six markers no matched sequences were found. Totally, ~53% of the identified sequences showed the best match with the alternative allele sequences, while ~47% with the reference one. The tomato wild genome with the highest number of matching with the reference sequences was *S. galapagense* (48 out of 68 identified markers). Considering the alternative allele sequences, 23 out of 40 matched against *S. chilense* whereas, 13 out of 28 against *S. pennellii*. Only two sequences matched against *S. peruvianum* (M19 for alternative variant and M31 for reference one). These two markers confirmed that M31 showed the best fit with reference sequence as in other wild genomes, while M19 showed the best fit with the alternative one. Interestingly, 37 out of 48 sequences identified in *S. pimpinellifolium* showed higher identity scores than the reference sequences and 8 out of 13 markers common between PDVIT and E42 were peculiar of this wild species. Moreover, 11 (three peculiar sequences of E42, five of PDVIT and three common to both genotypes) assayed markers, which matched sequences on chromosome 2, showed an identity score

of 100 % and in one of the ~98 %. Totally, the 90 % of the sequences common for both genotypes, the 80 % for E42 and 70 % for PDVIT showed high similarity with *S. pimpinellifolium*.

Figure 3.6 Results of the alignment against tomato wild genomes. The red color represents the best fit with the reference sequence, whereas the green represents the best fit with the PDVIT/E42 variants.

4. Phenotyping of F_1 hybrids

4.1 Material and methods

4.1.1 Plant material

During the years 2017 and 2018, 19 F_1 hybrid combinations were obtained (Table 4.3) using 19 parental genotypes, selected for quality and yield traits. The F_1 plants were evaluated during the following growing seasons 2018 and 2019. Moreover, in order to compare the yield *per* plant of the hybrids grown under high temperatures (HT), the two determinate heat-tolerant hybrids DOCET and JAG8810 and the indeterminate hybrid PAIPAI (ENZA ZADEN, Spain) were used as heat-tolerant controls. The two experimental trials were carried out in Campania (40°56'57.7"N, 14°23'20.3"E), transplanting plants with one month delay respect to the standard agronomical practices in order to favour the exposition of the plants to HT during reproductive stages. The experimental design consisted of a randomized complete block of 10 plants for each genotype and three replicates for genotype.

4.1.2 Phenotyping of F_1 hybrids

During the growing seasons data of three different phenotypic traits were collected: number of fruit *per* plant (TNF), fruit weight (FW) and yield *per* plant (YP). Moreover, in 2019 two qualitative traits were evaluated in F_1 hybrids and their parents: total soluble solid content (TSSC) and titratable acidity (TA). The red ripe fruits from three replicates were collected and were stored at -20 °C until the analysis. The TSSC was measured using the refractometer (HANNA instruments) and pH using the digital pHmeter Fiveasy F20 (Mettler Toledo). In order to evaluate the TA, fruits were triturated and centrifuged for 20 min at 6,000 rpm. Then, 2.5 g of extract were used for the analysis adding 30 ml of double-distilled water milliQ. Using a pre-charged buret, a solution of NaOH 0.1 N was dropped until reaching the endpoint of pH = 8. Data collected were used to calculate the acidity as follows:

$$Acidity\ (mmolH^+/100\ g) = \frac{[\mathrm{NaOH}]\ (\mathrm{N})\ \mathrm{x}\ \mathrm{Vol}_{NaOH}(\mathrm{ml})}{m_{sample}\ (\mathrm{g})}\ x\ \frac{1000}{10}\ =\ G1$$

where Vol_{NaOH} corresponds to the final volume added for the titration, m_{sample} the grams of starting matrix. Moreover, in order to calculate the titratable acidity in g/100g of fresh weight, G1 was multiplied with a fixed factor of 0.070 corresponding to the concentration of citric acid monohydrate.

4.1.3 Statistical analysis

Data collected were analyzed using IBM SPSS v. 23 (IBM Corp.). The calculation of heterosis percentage (%Het) on three yield-related and qualitative traits was carried out by using the formula %Het = [(F1 – PM) / PM] × 100 considering the mean of parents (PM) and of the hybrid (F_1) values in the calculation.

4.1.4 Design of molecular markers

To select the best parental lines and then to verify the F_1 hybrids, molecular markers were designed using two different approaches. In the first, some SNPs associated to yield-related traits (Chapter 3) including a restriction site were selected to design CAPS markers (Cleaved Amplified Polymorphic Sequence), using the CAPS designer tool (www.solgenomics.net/tools/caps_designer) (Table 4.1). Then, a primer pair targeting the region including the polymorphism was designed using the Primer3 webtool (http://primer3.ut.ee/), and used for the PCR reaction. Following the PCR, a restriction enzyme digestion of the amplified fragment was performed and analysed by standard agarose-gel electrophoresis. The second approach was the design of derived-Cleaved Amplified Polymorphic Sequence (dCAPS) markers. In this way, on SNP loci which do not include any restriction sites a primer that introduces/destroys restriction sites around the SNP was designed and coupled with another one to target the region including the mutated site. Following the PCR, the same steps followed for CAPS markers were carried out. The PCR reaction and cycles for these markers were set up as reported in Table 4.2.

Table 4.1 Design of PCR markers from SNPs identified in GWAS. For each marker, the arbitrary code, solcap SNP ID, type of designed marker, primer sequences, length of oligomers, melting temperatures and the expected amplicon size are reported.

Code	SNP marker	Marker	Sequence (5'->3')	Length (bp)	Melting temperatures (°C)	Amplicon size (bp)	Restriction enzyme
HT1	solcap_snp_sl_20344	CAPS	CTCCCCTATTTCCCTACCCC	20	64	608	*BclI*
			TCCCTTCTTCTAAGCTCCGC	20	62		
HT2	solcap_snp_sl_59728	CAPS	GGCAACGATGGTCACAAAGA	20	60	420	*HinfI*
			TGTGTTTGGTCAGCTGCATC	20	60		
HT3	solcap_snp_sl_9382	dCAPS	TGTTGACGGAGATGGATGGA	20	60	244	*ScaI*
			GCACCACTAAAACCATGTGAGT	22	62		
HT4	solcap_snp_sl_24679	CAPS	GTGTGTATGGGTTGGTTCGG	20	62	787	*MspAI*
			GCGGCTGAGCTTAGTAGAGA	20	62		
HT5	solcap_snp_sl_5428	CAPS	GACTGTGGGAGATTGAGGCT	20	62	136	*TaqI*
			GTGGTATAGTCTGGGCAGCA	20	60		

A standard agarose-gel electrophoresis analysis was performed for the evaluation of the PCR products. After the PCR analysis, the restriction reaction was set up as follows: 10 µl of PCR product, 5 µl of Enzyme Buffer 10X and 1µl of 10 U/µl enzyme (varying among different markers) and

double-distilled milliQ water up to the final volume of 50 µl. Another standard agarose-gel electrophoresis analysis was conducted to evaluate the digestion pattern. The concentration of agarose-gel varied from 1.2% for CAPS markers to 3% for dCAPS.

Table 4.2 PCR reactions and cycles set up for the isolation of expected amplicons used for the restriction enzyme reaction. The PCR mix, times and temperatures of PCR cycles were reported.

PCR reaction			PCR cycle			
Reagent	Reagent conc.	Volume	Step	Time	Temperature	
DNA	50 ng/µl	1 µl	Denaturation	5 min	95°C	1 cycle
Primer Forward	2.5 µM	2.5 µl	Denaturation	30 sec	95°C	
Primer Reverse	2.5 µM	2.5 µl	Annealing	30 sec	Primer pair dependent	30 cycles
Enzyme buffer	5X	10 µl				
dNTPs	10 µM	1 µl	Extension	30 sec	72°C	1 cycle
Taq DNA polymerase	5 U/µl	0.2 µl	Final extension	7 min	72°C	
H_2O		up to 50 µl	Storage	∞	4°C	

4.2 Results

4.2.1 Selection of parentals genotypes

The parental genotypes were selected based on their heat tolerance and/or their qualitative traits and genetic variability. Nineteen genotypes were selected in order to constitute different hybrid combinations. As reported in Table 4.3, 19 combinations were obtained in two years, combining the genotypes based on their genotypic and phenotypic traits (Table 4.3).

Table 4.3 Nineteen cross combination constituted in 2017 and 2018 and evaluated in the growing seasons 2018 and 2019. HT=heat tolerance

Cross year	Hybrid code	Growing season	♀		♂	Trait
2017	17H10	2018	E37	x	E33	HT, Quality
	17H11	2018	E37	x	STIZ	HT, Quality
	17H12	2018	STIZ	x	E33	Quality
	17H13	2018	STIZ	x	E37	HT, Quality
	17H14	2018/2019	PDVIT	x	E7	HT, Genetic variability, Quality
	17H16	2018	PDVIT	x	E64	HT, Genetic variability, Quality
	17H25	2018	E111	x	E42	HT, Genetic variability, Quality
	17H32	2018	E42	x	E20	HT, Genetic variability
	17H36	2018/2019	E103	x	E42	HT, Genetic variability, Quality
	17H37	2018/2019	E103	x	E48	HT, Quality
	17H39	2018/2019	E103	x	PDLUC	Quality
	17H56	2019	E103	x	E55	Genetic variability, Quality
	17H57	2019	E103	x	E111	Quality
2018	18H13	2019	E20	x	E42	HT, Genetic variability
	18H17	2019	E36	x	E42	HT, Genetic variability, Quality
	18H48	2019	E109	x	E45	Quality
	18H56	2019	E11	x	LA2662	HT, Genetic variability, Quality
	18H57	2019	E11	x	LA3120	HT, Genetic variability, Quality
	18H59	2019	E11	x	E42	HT, Genetic variability, Quality

Indeed, data collected from GBS analysis experiment (reported in Chapter 3) revealed that the most polymorphic genotypes were E42, E55 and PDVIT (Figure 3.2). Moreover, using the dataset obtained

from the Infinium SolCAP array, E11 was selected for its high genetic variability respect to the other genotypes, estimating the percentage of private SNPs (38.4 %) on a total of 1,535 private polymorphic loci (Figure 4.1).

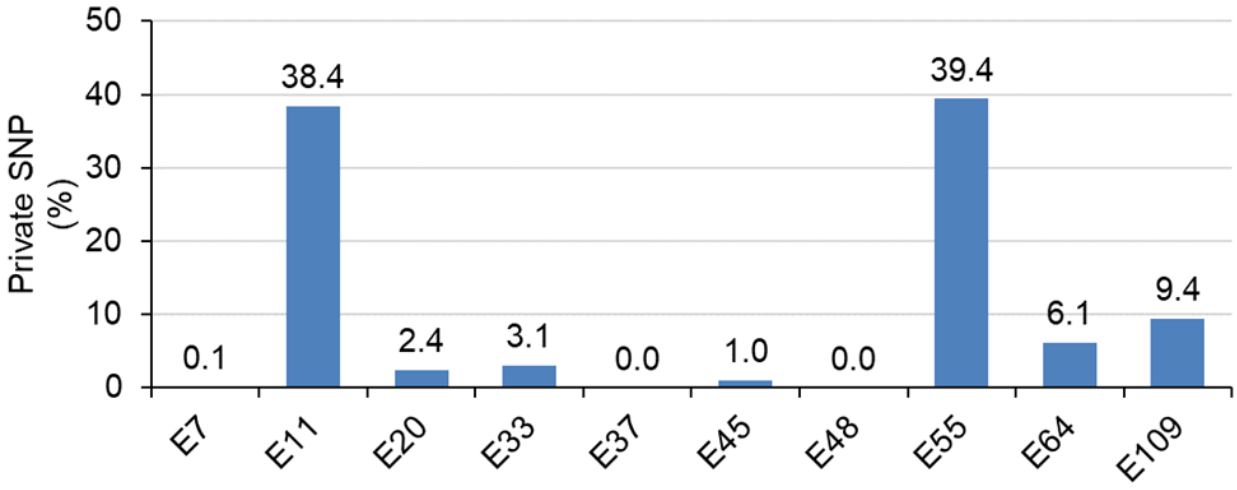


Figure 4.1 Percentage of private SNPs detected in 10 genotypes involved in cross combinations. The dataset was obtained from Infinium SolCAP array platform.

From the phenotypic analysis carried out in the present thesis (reported in Chapter 2) seven genotypes (E7, E20, E36, E37, E42, LA2662, LA3120) were selected for their good yield performances under HT. Finally, other parents were selected based on previous studies (Ruggeri et al., 2014; Sacco et al., 2015), mostly for their good fruit quality traits. The aim was to combine genotypes that showed good yield under HT and other traits of interest as fruit quality traits. In particular, seven combinations (17H14, 17H16, 17H25, 17H36, 18H17 and 18H59), including HT, genetic variability and quality traits, were obtained. Six other hybrid combinations (17H10, 17H11, 17H13, 17H37, 18H56 and 18H57) included good yield performances and quality traits. Among these hybrids the genotypes E37 and E48 were selected for the good yield performances in other studies (unpublished data). Moreover, also LA2662 and LA3120 were used as parents for good yield performances under heat stress (Abdul-Baki, 1991; Chetelat and Petersen, 2005; Dane et al., 1991; Opeña et al., 1992; Rudich et al., 1977). Finally, other four hybrids (17H12, 17H39, 17H57 18H48) involving genotypes previously selected for quality traits were also constituted (Ruggieri et al., 2014; Sacco et al., 2015).

4.2.2 Validation of F_1 hybrids using molecular markers

A group of SNPs associated to different phenotypic traits was transformed into PCR markers (Table 4.1) for the validation of F_1 hybrids. Totally, five SNP markers were transformed in five CAPS markers and one dCAPS marker and assayed on tomato parental genotypes and F_1 hybrids. Before the harvest, the validation of F_1 hybrids and their parents was carried out. In all cases the expected

amplicons for each molecular marker were observed and the restriction enzyme-mediated reactions were performed. All hybrids and parents were verified and corresponded to the expected pattern (Figure 4.2).

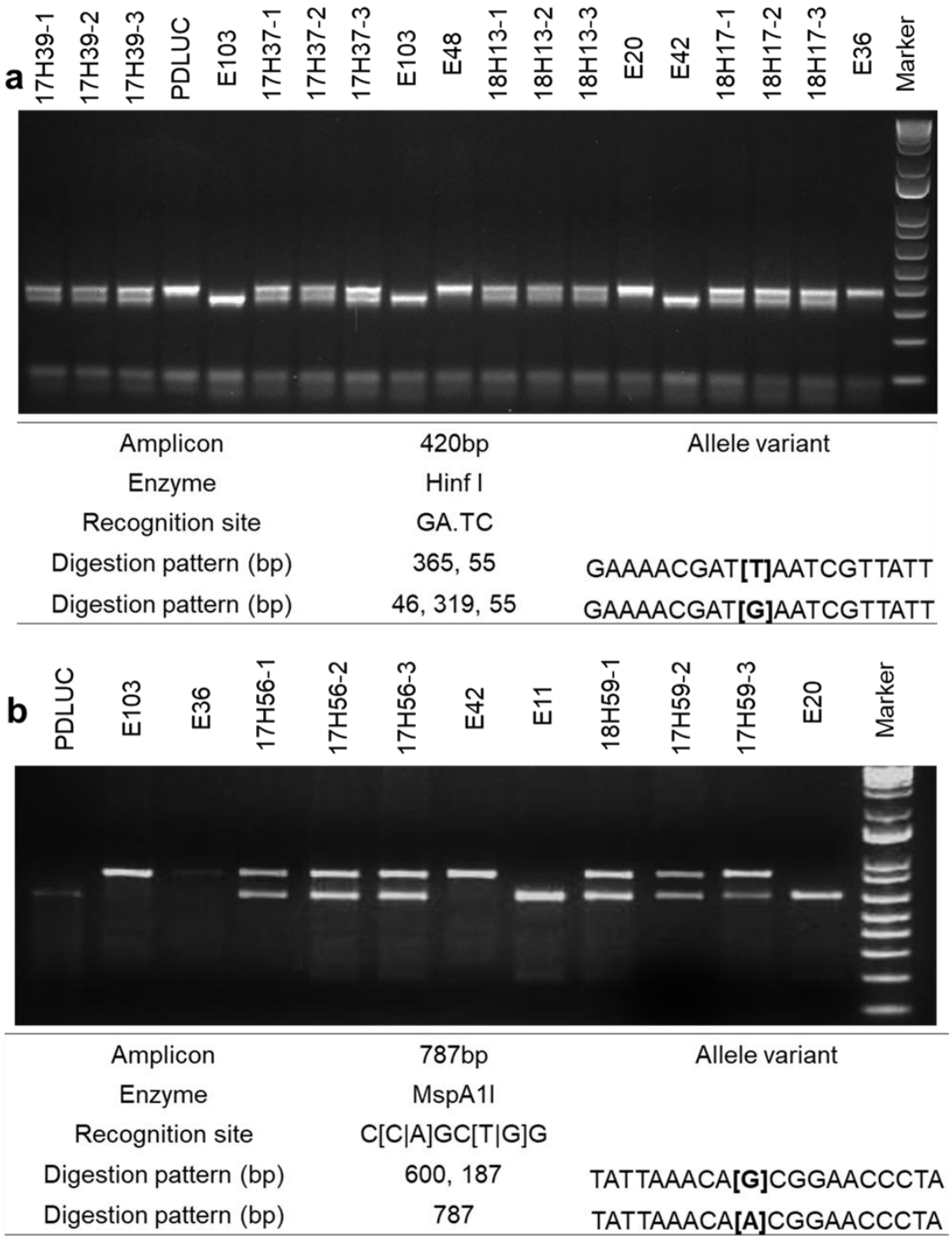


Figure 4.2 Electrophoresis analysis of restriction enzyme-mediated reaction of two CAPS markers. a) digestion pattern of HT2; b) digestion pattern of HT4 (for details on markers HT2 and HT4 see Table 4.1).

4.2.3 Phenotyping of hybrids in 2018

In the year 2018, 11 hybrids and 12 respective parents were grown in open field evaluating three yield-related traits (Supplementary Table S8): total number of fruit *per* plant (TNF), fruit weight (FW) and yield *per* plant (YP). The Fischer's LSD-test was performed to evaluate differences among the hybrids, by comparing with the controls (Figure 4.3). The YP recorded in DOCET (4.38 kg/plant) and PAIPAI (2.50 kg/plant) (Supplementary Table S9) resulted in all cases significantly higher than the other hybrids. The hybrids 17H10 (1.20 kg/plant), 17H11 (1.29 kg/plant), 17H12 (1.35 kg/plant), 17H13 (1.05 kg/plant), 17H36 (1.66 kg/plant) and 17H37 (1.08 kg/plant) showed a YP statistically higher than JAG8810 (0.40 kg/plant).

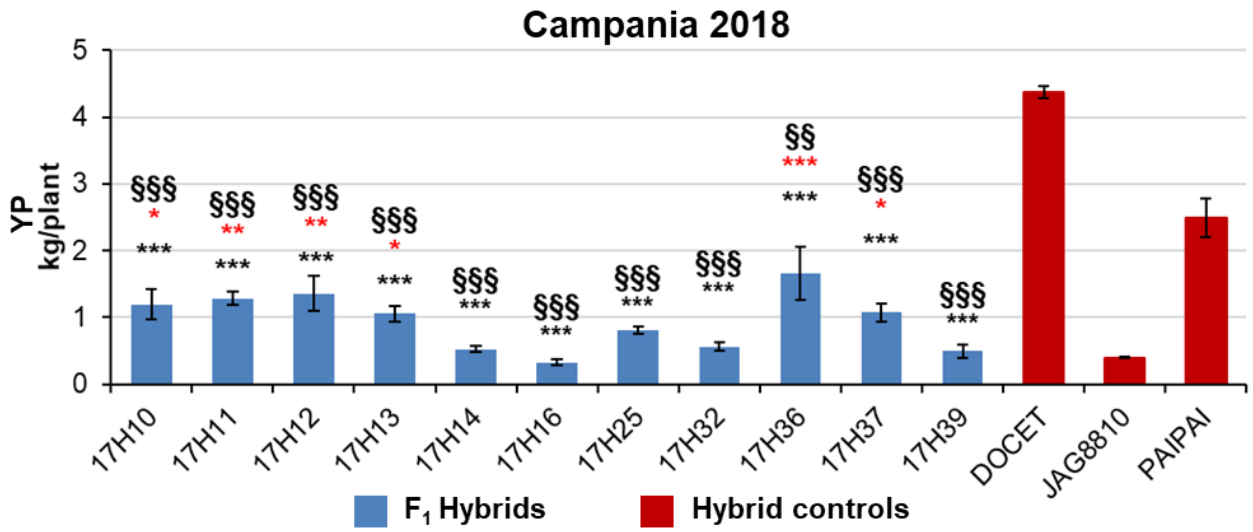


Figure 4.3 Phenotypic evaluation of yield *per* plant (YP) of 11 F_1 hybrids compared to positive controls in 2018. Fischer's LSD-test is reported. Black asterisk indicates the significance respect to DOCET, the red one indicates the significance respect to JAG8810, the symbol § indicates the significance respect to PAIPAI. */§, significance at 0.05; **/§§, significance at 0.01; ***/§§§, significance at 0.001.

Moreover, a Duncan's test was carried out in order to evidence differences between the hybrids and their parents. As for TNF (Figure 4.4), only 17H37 showed a TNF value (74.98 fruit/plant) significantly higher than both parents E103 (13.57 fruit/plant) and E48 (14.65 fruit/plant). The hybrids 17H14 (52.73 fruit/plant), 17H25 (73.04 fruit/plant) and 17H36 (126.96 fruit/plant) showed TNF values intermediate between the parents.

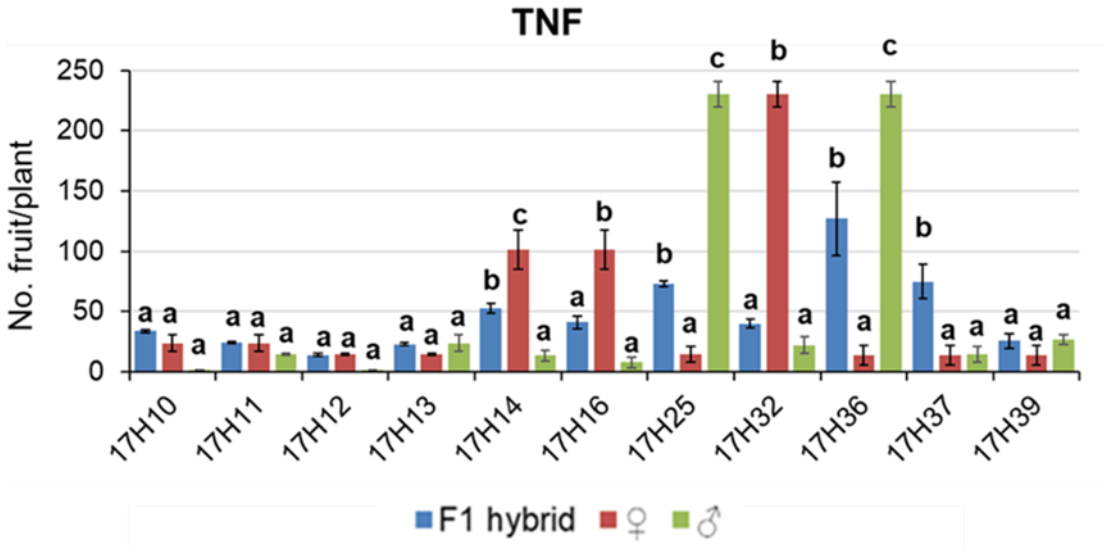


Figure 4.4 Number of fruit *per* plant (TNF) detected in F*1* hybrids and their parents. Duncan's test is also reported, in order to evaluate the significance of difference between F_1 hybrids and parents.

As for FW (figure 4.5), no hybrids showed a significant increase respect to both parents, with 17H14 (10.04 g), 17H32 (13.81 g) and 17H39 (19.80 g) showing a fruit weight intermediate between the parents.

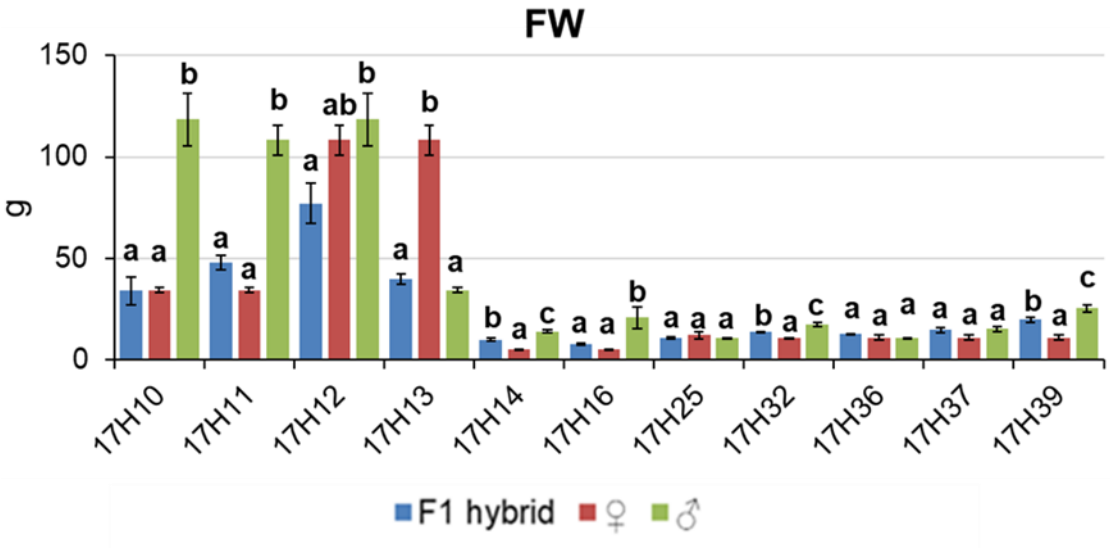


Figure 4.5 Fruit weight (FW) detected in F_1 hybrids and their parents. Duncan's test is also reported, in order to evaluate the significance of differences between F_1 hybrids and parents.

Finally, for YP (figure 4.6), only 17H37 (1.08 kg/plant) showed a significant increase compared to both parents E103 (0.13 kg/plant) and E48 (0.23 kg/plant). The YP for 17H14 (0.53 kg/plant), 17H25 (0.81 kg/plant), 17H36 (1.66 kg/plant), 17H10 (1.20 kg/plant) and 17H12 (1.35 kg/plant) resulted intermediate between their parents.

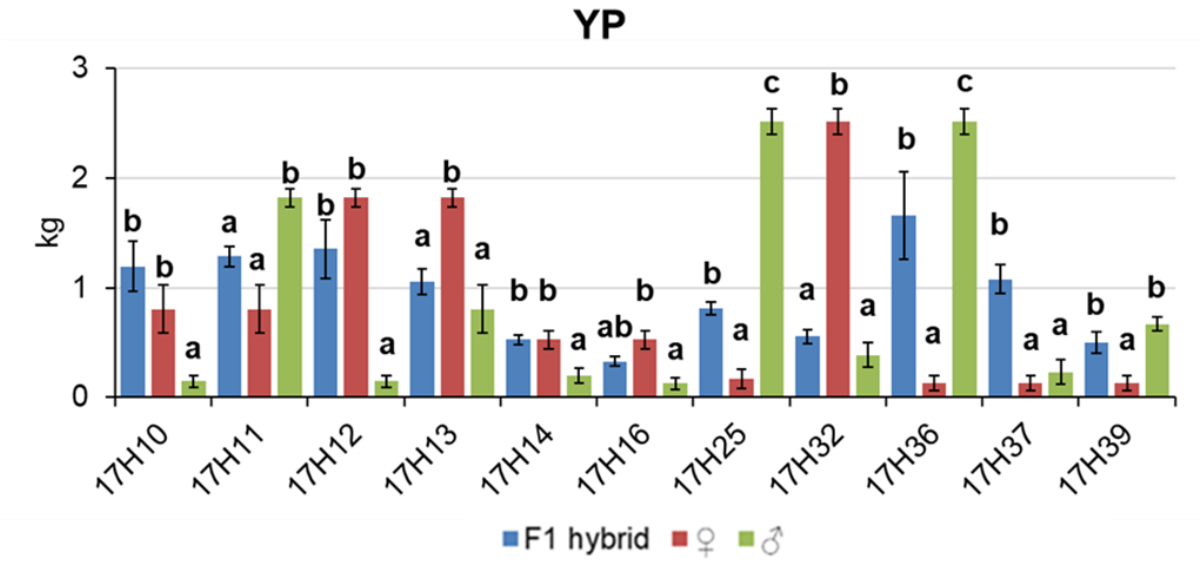


Figure 4.6 Yield *per* plant (YP) detected in F_1 hybrids and their parents. Duncan's test is also reported, in order to evaluate the significance of differences between F_1 hybrids and parents.

The percentage of heterosis (%Het) was also calculated (Table 4.4). The percantage varied from -68.27 (17H32) to 431.49 (17H37) for TNF, from -55.47 (17H10) to 19.83 (17H36) for the FW and from -61.72 (17H32) and 498.64 (17H37) for YP. Particularly interesting, 17H10 and 17H37 showed very high values of %Het for TNF and YP. A group of hybrids showing positive heterotic effects in the yield-related traits was evaluated for a second year in the same location.

Table 4.4 Heterosis values (%Het) in three yield-related traits evaluated in 2018. TNF, number of fruit *per* plant, FW, fruit weight, YP, yield *per* plant. In bold, the positive heterosis values.

Trait	17H10	17H11	17H12	17H13	17H14	17H16	17H25	17H32	17H36	17H37	17H39
TNF	**168.21**	**27.20**	**76.18**	**19.62**	-8.16	-24.27	-40.30	-68.27	**4.07**	**431.49**	**28.71**
FW	-55.47	-33.04	-31.95	-44.19	**3.54**	-39.91	-4.16	-2.53	**19.83**	**12.96**	**9.03**
YP	**150.72**	-2.12	**37.63**	-19.76	**45.45**	-0.13	-39.64	-61.72	**25.63**	**498.64**	**25.96**

4.2.4 Phenotyping of hybrids in 2019

During the year 2019, 12 combinations and 15 respective parents were evaluated for three yield-related (Supplementary Table S10) and two quality traits (Supplementary Table S11). The 12 combinations included four F_1 hybrids already evaluated in the year 2018 (17H14, 17H36, 17H37 and 17H39) and eight new combinations evaluated for the first time (17H56, 17H57, 18H13, 18H17, 18H48, 18H56, 18H57 and 18H59). As reported in Figure 4.7, the hybrids were compared to the positive controls for YP under HT.

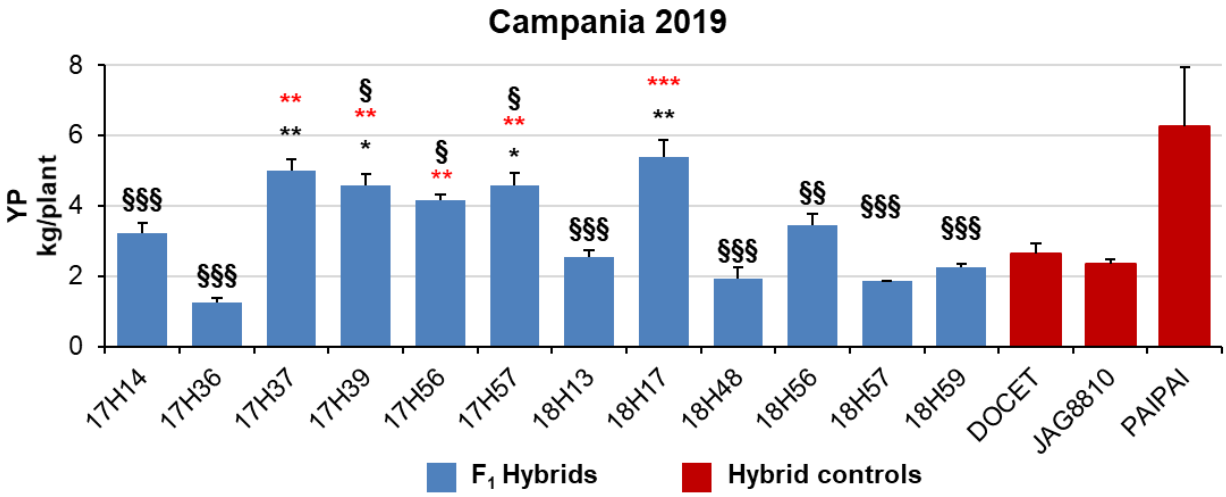


Figure 4.7 Phenotypic evaluation of yield *per* plant (YP) of 12 F_1 hybrids compared to positive controls in 2019. Fischer's LSD-test is reported. Black asterisk indicates the significance respect to DOCET, the red one indicates the significance respect to JAG8810, the symbol § indicates the significance respect to PAIPAI. */§, significance at 0.05; **/§§, significance at 0.01; ***/§§§, significance at 0.001.

Four hybrids (17H37, 17H39, 17H57 and 18H17) showed a significant higher yield respect to DOCET (2.65 kg/plant) and JAG8810 (2.35 kg/plant), whereas, 17H56 (4.15 kg/plant) only higher than JAG8810. Moreover, only 17H37 (4.98 kg/plant) and 18H17 (5.39 kg/plant), showed YP comparable to that of PAIPAI (6.24 kg/plant), that showed the highest YP among the controls. Duncan's test was carried out to evidence the statistical differences among the hybrids and their parents for different traits. As for TNF (figure 4.8), the hybrids 17H39, 17H56, 17H57 and 18H17 showed a significant increase of the number of fruit compared to both parental genotypes.

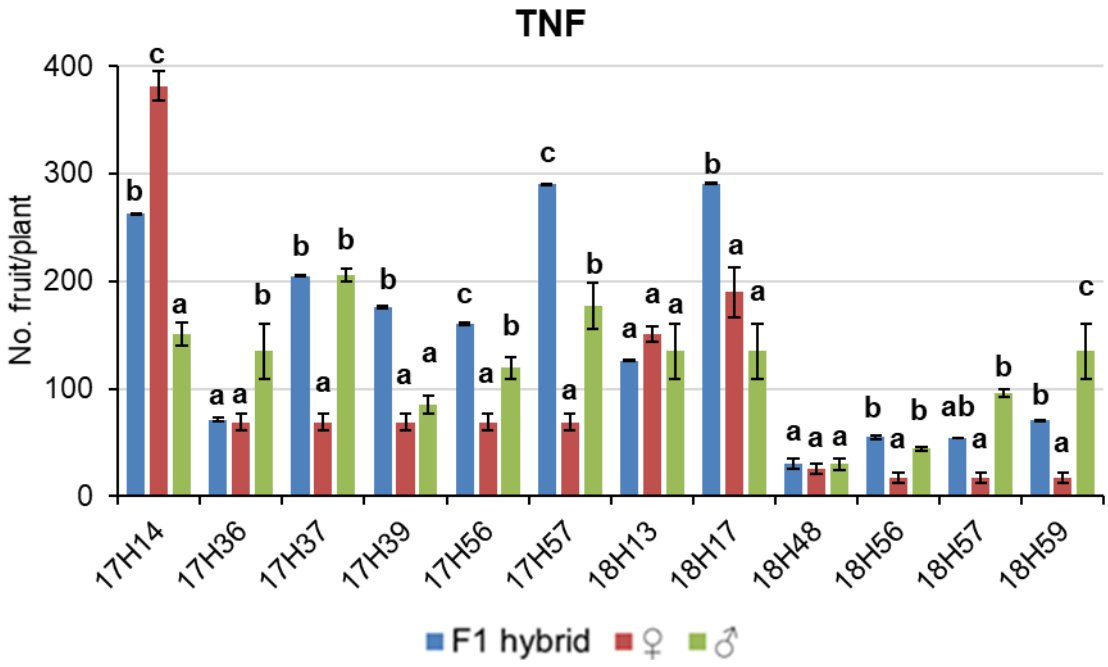


Figure 4.8 Number of fruit *per* plant (TNF) detected in F_1 hybrids and their parents in 2019. Duncan's test is also reported, in order to evaluate the significance of differences between F_1 hybrids and parents.

As for FW (Figure 4.9), the hybrids 17H37, 17H39 and 17H56 showed a FW significantly higher than both parents. Others six hybrids (17H14, 17H57, 18H13, 18H17, 18H56 and 18H59) showed FW intermediate between the parents.

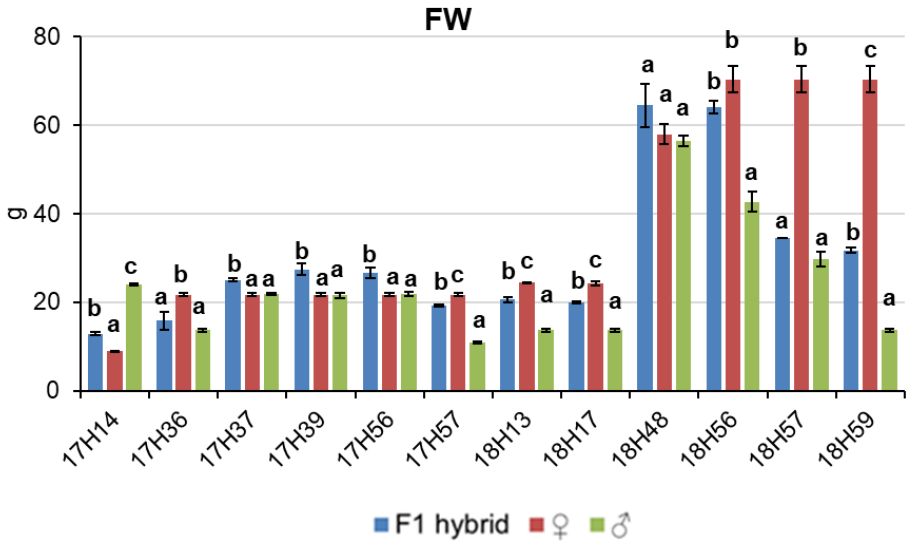


Figure 4.9 Fruit weight (FW) detected in F_1 hybrids and their parents in 2019. Duncan's test is also reported, in order to evaluate the significance of differences between F_1 hybrids and parents.

Finally, for YP (figure 4.10) the hybrids 17H37 (4.98 kg/plant), 17H39 (4.57 kg/plant), 17H56 (4.15 kg/plant) and 18H56 (3.45 kg/plant) showed a final yield *per* plant significantly higher than those of corresponding parental genotypes. The hybrids 18H17 (5.39 kg/plant) and 18H59 (2.24 kg/plant) showed YP values comparable to the best parental genotypes E36 (4.28 kg/plant) and E42 (2.21 kg/plant), respectively. Then, 17H14 (3.22 kg/plant) and 18H48 (1.92 kg/plant) resulted statistically comparable to the parents.

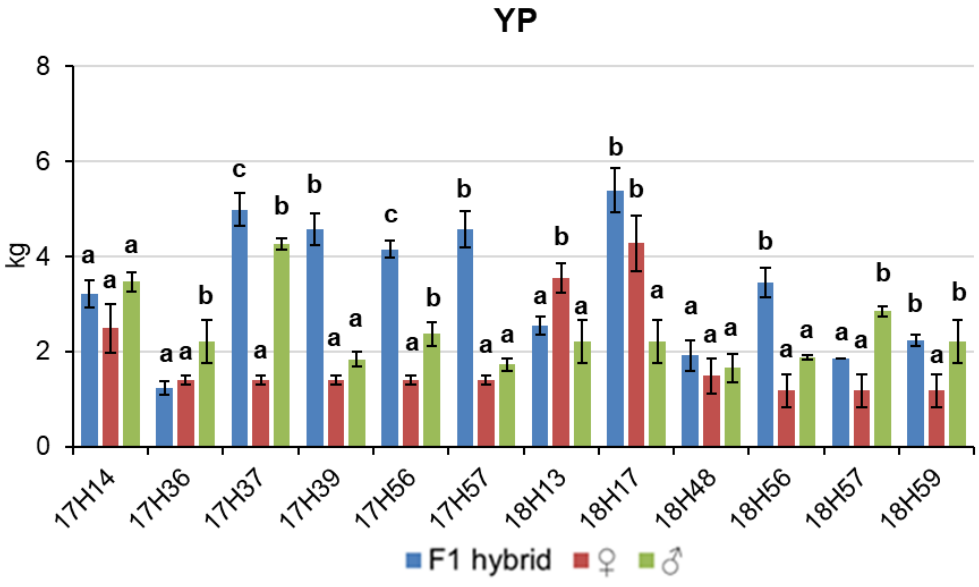


Figure 4.10 Yield *per* plant (YP) detected in F_1 hybrids and their parents in 2019. Duncan's test is also reported, in order to evaluate the significance of differences between F_1 hybrids and parents.

During the year 2019, qualitative data were also collected in terms of titratable acidity content in terms of citric acid monohydrate (TA) and total soluble solid content (TSSC). Generally, the hybrids showed TA comparable to their parents (Figure 4.11a). Moreover, the hybrids 18H13 (0.51 g/100 g) and 18H17 (0.52 g/100 g) showed a TA significantly higher than E42 (0.36 g/100 g), one of their parents. The other parents E20 (0.54 g/100 g) and E36 (0.47 g/100 g) resulted statistically comparable to the hybrids. The hybrid 18H48 (0.31 g/100 g) showed an intermediate value between the two parents E109 (0.35 g/100 g) and E45 (0.26 g/100 g). As for the TSSC, globally the hybrids and parents showed a certain homogeneity for this trait (Figure 4.11b). The hybrids 17H14, 17H36, 18H17 and 18H59 showed a significant increase of TSSC compared to at least one parent. In the case, the value of 17H14 (6.98) resulted intermediate between the two parental genotypes PDVIT (8.08) and E7 (4.95). The heterotic effects for different yield-related traits and of Total Soluble Solid Content/Titratable Acidity ratio (TSSC/TA) were calculated (Table 4.5). As for TNF, the hybrids 17H37 (49.28), 17H39 (128.27), 17H56 (69.83), 17H57 (136.32), 18H17 (79.27) and 18H56 (80.34) showed high positive heterosis. Regarding FW, 17H14 (-21.44), 17H36 (-10.65), 18H57 (-31.29) and 18H59 (-24.50) showed negative values of heterosis respect to the other hybrids which showed positive values. As for YP, interestingly 9 out of 12 different hybrids showed a positive heterotic effect, some with values exceeding 100 % heterosis. For the TSSC/TA ratio 7 out of 12 hybrids showed a positive heterotic effect. Moreover, the hybrids 17H37, 17H39, 17H56, 17H57, and 18H48 combined high heterotic effects for both yield related and quality traits and were selected as the best performers in 2019.

Table 4.5 Heterosis values (% Het) in traits evaluated in the year 2019 on 12 F_1 hybrids. TNF, number of fruit per plant, FW, fruit weight, YP, yield per plant, TSSC/TA, Total soluble solid content/Titratable acidity ratio. In bold, the positive heterosis values

Trait	17H14	17H36	17H37	17H39	17H56	17H57	18H13	18H17	18H48	18H56	18H57	18H59
TNF	-1.37	-30.35	**49.28**	**128.27**	**69.83**	**136.32**	-11.71	**79.27**	**9.60**	**80.34**	-4.79	-6.97
FW	-21.44	-10.65	**15.08**	**27.05**	**22.42**	**18.49**	**8.12**	**4.99**	**12.70**	**13.33**	-31.29	-24.50
YP	**8.00**	-31.85	**87.93**	**183.09**	**120.76**	**192.65**	-11.48	**65.92**	**22.04**	**126.21**	-8.24	**31.93**
TSSC/TA	-6.45	-6.30	**9.44**	**11.20**	**9.51**	**8.33**	-8.81	-10.22	**3.04**	-3.39	**9.61**	**6.45**

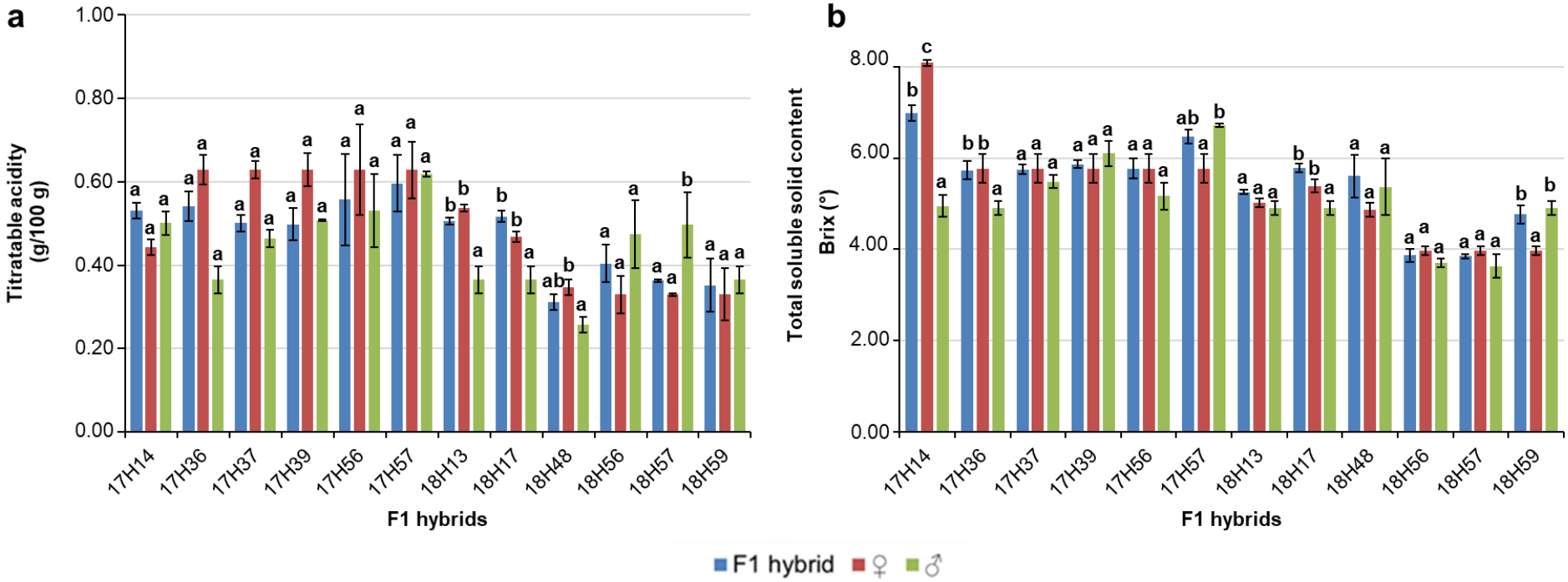


Figure 4.11 Qualitative analysis carried out in F_1 hybrids and their parents in 2019. Duncan's test is also reported, in order to evaluate the significance of differences between F_1 hybrids and parents. a) Titratable acidity, b) Total soluble solid content.

4.2.5 Comparison of hybrids grown in two years

Four different F_1 hybrids combinations were therefore evaluated for two years (2018 and 2019) under HT conditions. As reported in Figure 4.12a, different temperature trends were recorded in the two years in Campania and a comparable number of days with maximum temperatures higher than 32°C. In Campania 2018, globally the temperatures recorded were variable with temperatures that exceeded in most cases 30°C, whereas data of Campania 2019 (Figure 4.12b) showed two different temperature windows, one with temperatures that did not exceed the 25°C (from 16.7 to 24.3°C) and another with maximum temperatures that exceeded 32°C in most cases.

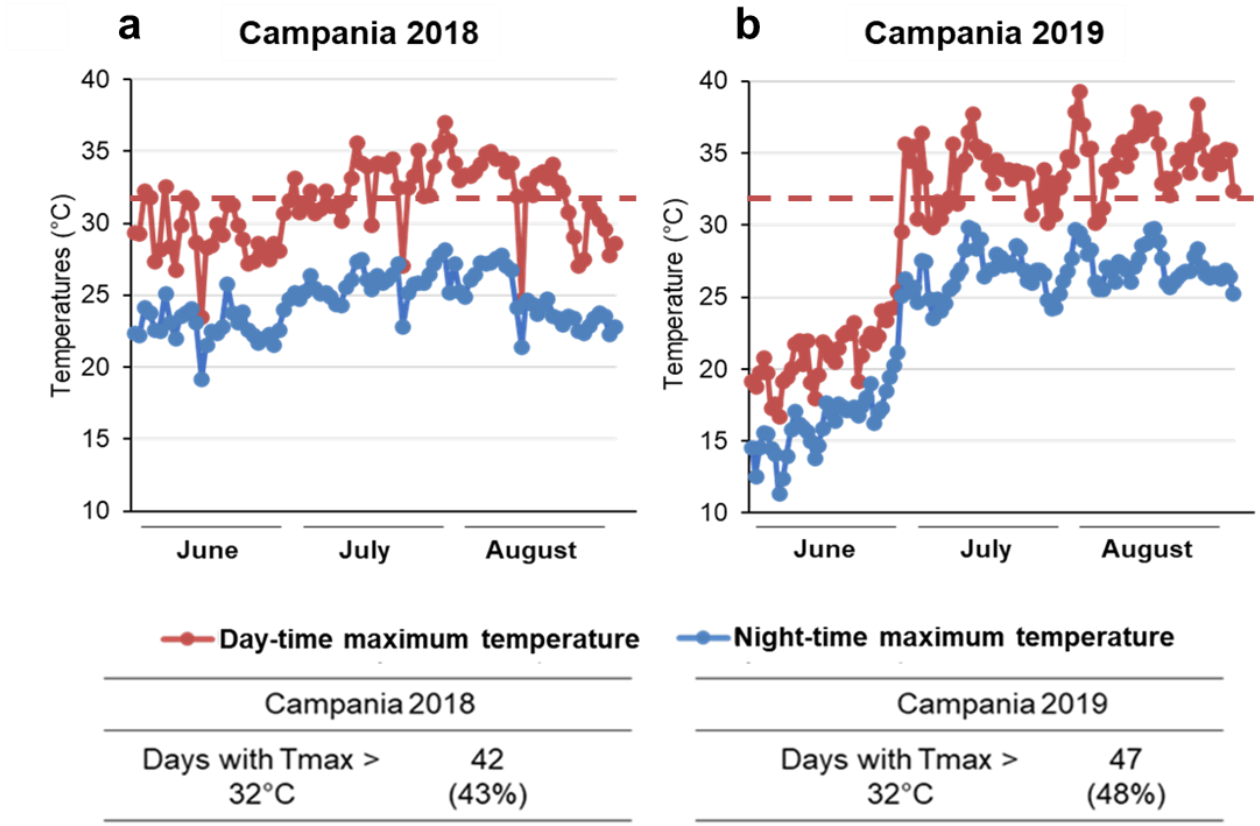


Figure 4.12 Day-time and night-time maximum temperatures recorded in Campania in the year 2018 (a) and 2019 (b). In red, the daily maximum temperature and in blue, the night-time maximum temperatures.

The validation was carried out by comparing the performances in two years and observing if heterotic effects on three yield-related traits were observed. As for TNF (Figure 4.13a), only 17H37 (431.49 and 49.28 in 2018 and 2019, respectively) and 17H39 (28.71 and 128.27, in the 2018 and 2019, respectively) confirmed positive heterotic effects on the TNF. As for FW (Figure 4.13b), positive values for 17H37 and 17H39 were recorded in both years. Finally, for YP (Figure 4.13c) the four hybrids showing the positive heterotic effect in the year 2018 with %Het particularly high in 17H37. In the year 2019, three hybrids confirmed positive %Het values with 17H37 and 17H39 that showed the highest %Het values (87.93 and 183.09, respectively). For these reasons and considering the

qualitative analysis reported in the previous paragraph, 17H37 and 17H39 were selected as best performers hybrids evaluated in two years.

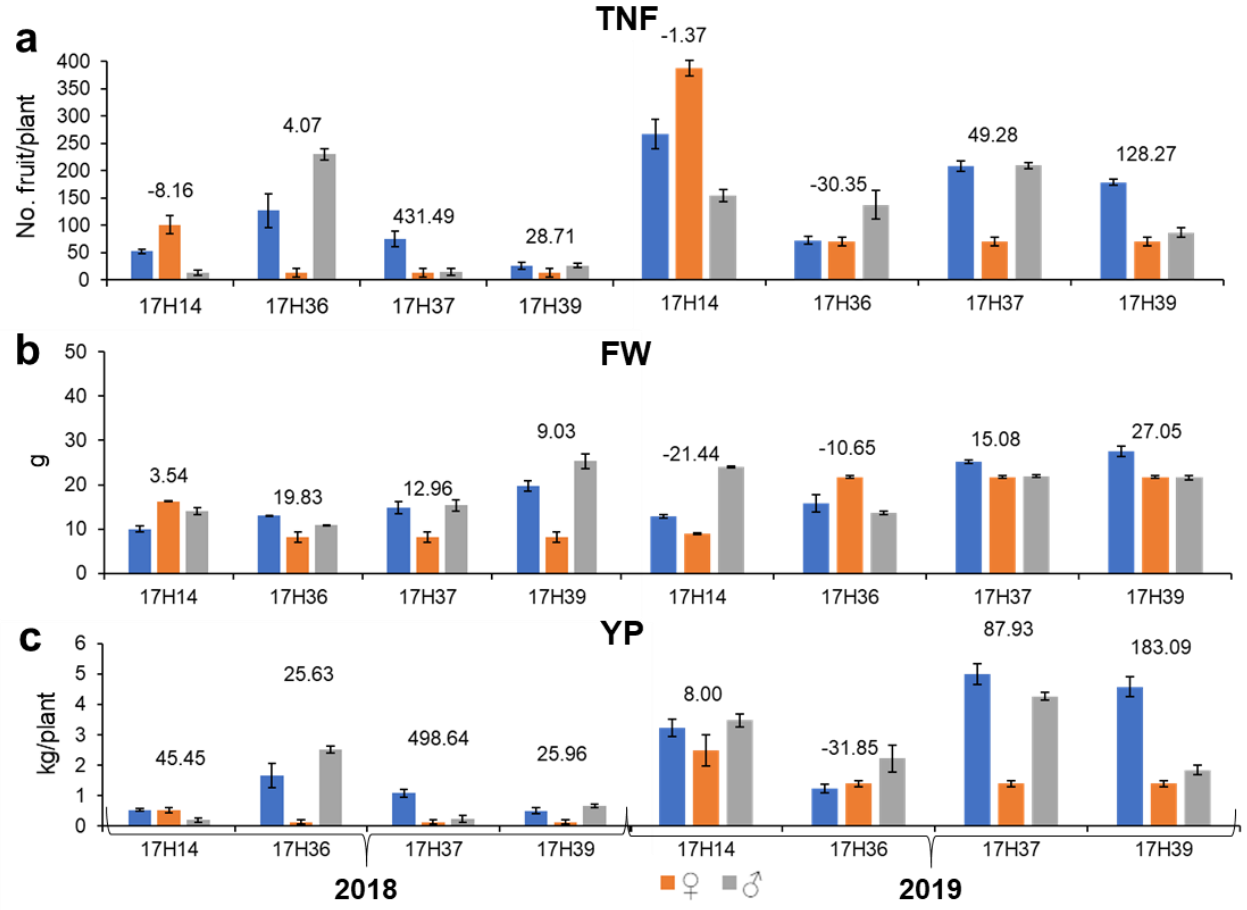


Figure 4.13 Comparison of four hybrids assayed in two year in Campania. The %Het values were reported above the bars. TNF, number of fruit *per* plant (a), FW, fruit weight (b), YP, yield *per* plant.

5. Genome editing on candidate genes involved in heat stress response

5.1 Materials and methods

5.1.1 Plant materials

The heat tolerant introgression line (IL4-4), derived by crossing *S. pimpinellifolium* accession TO-937 and *S. lycopersicum* cv Moneymaker, was used for the experiments to perform expression analysis and genetic transformation. tomato seeds were sterilized using phosphate dodecahydrate ($Na_3PO_4.12H_2O$) in agitation for three hours and then rinsed with sterile milli-ρ distilled water for five times. In the next step, the seeds were placed in 30 % sodium hypochlorite in agitation for one hour and then rinsed with milli-ρ distilled water for ten times. After the chemical treatment, seeds were dried on a filter paper for 24 h and then put in a silica gel box for the same period. Then, seeds were subject to thermal treatment at 80 ± 2 °C for 24 h and sown in plateau.

5.1.2 Expression analysis

In order to investigate two HSPs (*Solyc04g081630* and *Solyc04g081640*) mapping in the introgressed region of IL4-4, expression analysis was carried out using as plant tissues the ovaries, fecundated ovaries and anthers, as reported in the literature (Buzgo et al., 2004; Xiao et al., 2009). For each sample five flowers *per* plant and three replicates at different developmental stage were collected, considering as reference time the Days Post-Anthesis (DPA). Moreover, some samples derived by ovaries and anthers were collected one day before the anthesis (-1 DPA) and on the first day of anthesis (0 DPA). Ovaries were also collected at 4, 7, 10 and 13 after the fecundation. Totally, eight different floral stages were considered for the expression analysis: anthers collected at -1 DPA (ANT-1) and 0 DPA (ANT0), ovaries collected at -1 DPA (OVA-1), 0 DPA (OVA0), 4 DPA (OVA4), 7 DPA (OVA7), 10 DPA (OVA10), and 13 DPA (OVA13). In order to carry out the RNA extraction, flowers were dissected and divided in ovaries and anthers and stored in liquid nitrogen until the extraction. The tissue was grinded with sterile and RNAse-free mortar and pestle using liquid nitrogen. Starting from 100 mg of powdered tissue, RNA extraction was performed using RNeasy Plant mini Kit (Qiagen) following the manufacturer's protocol. The RNA concentration was verified with Nanodrop Spettrophotometer (Thermo Fischer) whereas the quality was verified using standard agarose-gel electrophoresis analysis. The cDNA synthesis was performed by the SuperScript III Reverse Transcriptase kit (Thermo Fischer) following the manufacturer's instructions. In order to check the synthesis of cDNA, all samples were verified performing a PCR reaction (Table 5.1) using primer pairs that amplified a portion of constitutively expressed Elongation Factor 1α gene (EF-1α) (Table 5.2).

Table 5.1 PCR reaction conditions to check cDNA synthesis. The PCR mix, times, temperatures, and cycles are reported.

PCR reaction			PCR cycle			
Reagent	Reagent conc.	volume	Step	Time	Temperature	
cDNA		1 µl	Denaturation	5 min	94 °C	1 cycle
Primer Forward	2.5 µM	1.25 µl	Denaturation	30 sec	94 °C	
Primer Reverse	2.5 µM	1.25 µl	Annealing	30 sec	53 °C	30 cycles
Enzyme buffer	5X	5 µl				
dNTPs	10 µM	0,5 µl	Extension	30 sec	72 °C	
Taq DNA polymerase	5 U/µl	0.2 µl	Final extension	7 min	72 °C	1 cycle
H2O		up to 25 µl	Storage	∞	4 °C	

All the PCR reactions were performed in Veriti Thermal Cycler (Applied Biosystem). The correct amplification was verified performing a standard agarose-gel electrophoresis analysis. Primer for the quantitative Real-Time-PCR (qRT-PCR) were designed using the bioinformatic tool Primer3plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and verified possible secondary structure on Primer-BLAST tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) (Table 5.2).

Table 5.2 Primers designed for qRT-PCR. The sequence, the oligo length, the melting temperatures and the expected amplicons are reported

Primer	Sequence (5'→3')	Oligo length (bp)	Melting (°C)	Amplicon (bp)
qRT-Solyc04g081630_F	GGTATGGACACACCAGAATCAA	22	64	100
qRT-Solyc04g081630_R	TGAACTACAACCTGCCACAC	20	60	
qRT-Solyc04g081640_F	TCGACTGTTATGGGGTTTGTAA	22	62	156
qRT-Solyc04g081640_R	CATGATTTTTCTGCTAGTAAGTTC	24	64	
EF-1α_F	CTCCATTGGGTCGTTTTGCT	20	60	101
EF-1α_R	GTCACCTTGGCACCAGTTG	19	60	

The qRT-PCR and expression analysis were carried using Rotor Gene 6000 thermal cycler (Corbett Research). Reactions were prepared as reported in Table 5.3.

Table 5.3 qRT-PCR reaction for the gene expression analysis. The qRT-PCR mix, times, temperatures and cycles are reported.

PCR reaction			PCR cycle			
Reagent	Reagent conc.	volume	Step	Time	Temperature	
cDNA		1 µl	Denaturation	10 min	95 °C	1 cycle
Primer Forward	3 µM	1 µl	Denaturation	30 sec	95 °C	
Primer Reverse	3 µM	1 µl	Melting	30 sec	Depending on primers	45 cycles
SYBR Green	5X	5 µl				
H2O		up to 10 µl	Extension	15 sec	72 °C	

Data acquisition and analysis were performed using the Real-Time software Rotor-Gene 6000Software, v. 1.7 (Corbett Research) using the comparative method, based on the 2-ΔΔCt formula considering ΔCt = Ct target gene– Ct endogenous control, ΔΔCt = ΔCt sample – ΔCt calibrator (Livak and Schmittgen, 2001). Then, the endogenous gene used for the normalization and verification of the expression levels was the housekeeping EF-1α gene. A Student's-test was performed in order to verify the statistical significance of the analysis.

5.1.3 CRISPR/Cas9 vector assembly

5.1.3.1 Design of RNA guides

Starting from a previous work performed in IBMCP-Universitat Polìtecnica de Valencia in Granell's laboratory (Barrantes et al., 2014), two genes coding for two HSP90 (*Solyc04g081630* and *Solyc04g081640*) mapping in the introgressed region 4-4, were used as targets to investigate their roles in floral development in IL4-4. At this purpose, two different vectors were constructed for the knock-out. Gene sequences of these two genes available at www.solgenomics.net were loaded on bioinformatic tool Benchling CRISPR guide RNA design tool (available at https://www.benchling.com/crispr) to design guides for the genetic transformation CRISPR/Cas9-mediated. Three different features were considered:

- **Choosing the target region**: RNA guides and Crispr/Cas9 machinery must produce a deletion that provokes an alteration of reading frame and/or interrupts the gene sequence. Moreover, if possible, it should occur at the beginning of the coding sequence.
- **High specificity of target gene**: It is necessary to avoid the presence of off-targets in transformed plants. By default, Benchling software calculates on-target and off-target score to avoid this problem.

- **High efficiency**: A good guide must be designed just before a Protospacer Adjacent Motif (PAM) sequence (5'-NGG-3', canonic form of the). By default, Benchling software selects only the guides which respect this condition.

For *Solyc04g081630* and *Solyc04g081640* a pair and two pairs of complementary oligos were designed in order to produce heterodimer with peculiar compatible overhangs, respectively. In order to constitute the heterodimers, 5 µl of each oligo concentrated at 1 µM (Table 5.4) were mixed with its complementary one and incubate for 30 minutes at room temperature.

Table 5.4. List of oligo pairs used for the constitution of heterodimers as gRNAs of the CRISPR tools. The oligo pair identified in the sequence of *Solyc04g081630* (SL4g30) has *BsaI* compatible overhangs, whereas the pairs identified in *Solyc04g081640* (SL4g40.1/.2) have *BsmBI* compatible overhangs. In bold the peculiar overhangs sites

Oligos	Sequence (5'->3')	Length (bp)	Heterodimer
SL4g30_F	ATTGGCAGGTTGTAGTTCAAAAG	23	**ATTG**GCAGGTTGTAGTTCAAAAG CGTCCAACATCAAGTTTTC**CAAA**
SL4g30_R	AAACCTTTTGAACTACAACCTGC	23	
SL4g40.1_F	GTGCACATTCACAAAATAATTGAAG	25	**GTGCA**CATTCACAAAATAATTGAAG TGTAAGTGTTTTATTAACTTC**CAAA**
SL4g40.1_R	AAACCTTCAATTATTTTGTGAATGT	25	
SL4g40.2_F	GTGCAAGAACCCACAGAGAAAATTG	25	**GTGCA**AGAACCCACAGAGAAAATTG TTATTGGGTGTCTCTTTTAAC**CAAA**
SL4g40.2_R	AAACCAATTTTCTCTGTGGGTTCTT	25	

5.1.3.2 GoldenBraid (GB) domestication

Two different domestication events, one for each heterodimer, were performed for the knock-out of *Solyc04g061740* (Table 5.5). The cycle used was 37°C for 10 minutes, 25 repetitions with 37 °C for 3 minutes, 16 °C for 4 minutes, 50 °C for 10 minutes and a step at 80 °C for 10 minutes. This reaction cycle was the same for all GoldenBraid steps. The correct assembly of heterodimers in pUPD2 was verified by Sanger DNA sequencing performed in IBMCP institute.

Table 5.5 Golden Braid reactions for the domestication of heterodimers in GB parts of the gene Solyc04g081640. The reagent type, quantity, restriction-ligation cycle are also reported.

Two-gRNA guides assembly					
Reagent	Type of reagent	Quantity	Reagent	Type of reagent	Quantity
Heterodimer.1	Oligos	2 ng	Heterodimer.2	Oligos	2 ng
GB1207	GB part	75 ng	GB1208	GB part	75 ng
pUPD2	Domestication Plasmid	75 ng	pUPD2	Domestication Plasmid	75 ng
BsmBI		5 U	*BsmBI*		5 U
T4 ligase		3 U	*T4* ligase		3 U
Ligase buffer		5X	Ligase buffer		5X

5.1.3.3 CRISPR multipartite assemblies in α level

After the validation of the first level, the RNA guides were put under the control of a constitutive promoter. At this purpose, for *Solyc04g061740* a multilevel assembly was carried out (Table 5.6).

Table 5.6 Golden Braid multipartite reactions of the two constructs. The reagents, quantity, and restriction-ligation cycle are also reported

Multipartite assembly		
Reagent	Type of reagent	Quantity
Heterodimer	Oligos	2 ng
GB1001	GB part	75 ng
GB0645	GB part	75 ng
pUPD2_gRNA1	Vector + Guide	75 ng
pUPD2_gRNA2	Vector + Guide	75 ng
pDGB3_ α1	Vector α	75 ng
BsaI		5 U
T4 ligase		3 U
Ligase buffer		5X

For the gene *Solyc04g061730* the domestication of the heterodimer was not necessary. For this reason, the first step was the restriction/ligation step by mixing the same GB parts reported in Table 5.6 (GB1001 and GB0645) with 75 ng of α vector the pDGB3_α1. From this step, each reaction of restriction/ligation was monitored using restriction enzymes cutting specific sites.

5.1.3.4 CRISPR binary assembly in Ω level destination vector

The following step was the same for both vectors. It was necessary to assembly the transcription units (TUs) constituted previously. The reactions of the final level are reported in Table 5.7.

Table 5.7 Assembly of the multipartite assembly level of the two constructs. The cycle used for the restriction-ligation reaction is also reported

Transcription Unit assembly			Golden Braid cycle		
Reagent	Type of reagent	Quantity	Temperatures	Time	
Heterodimer	Oligos	2 ng	37 °C	10 min	X1
GB0639	GB part	75 ng			
pDGB3_ α1	TU	75 ng	37 °C	3 min	
pDGB3_ Ω2	Vector Ω	75 ng	16 °C	4 min	X25
BsmBI		5 U	50 °C	10 min	
T4 ligase		3 U			
Ligase buffer		5X	80 °C	10 min	X1

5.1.3.5 Validation of final constructs

To confirm that the final GB constructs (Figure 5.1) were correctly assembled and the reading frame of the gRNAs was in order, specific fragments of the final vector were amplified and sequenced. In this way, a PCR final reaction was performed and set up as follow: 50 ng of final construct, 10 µM of primer Forward and Reverse (Table 5.8), 5 U/µl of MyTaq polymerase (Bioline), 1X of MyTaq Buffer (Bioline) and MilliQ double distilled water up to 25 µl of final volume. The PCR conditions were 94ºC for 5 minutes for the denaturing step, followed by 30 cycles of amplification (annealing and extension) with 94 ºC for 30 seconds, 58 ºC for 30 seconds and 72 ºC for 30 seconds, and a final extension at 72 ºC for 5 minutes. PCR products were run in a 1 % agarose gel using the Gene Ruler 100 bp Plus DNA Ladder as marker. After confirmation of the presence of a single amplicon, 10 µl of PCR product mixture were added to 4 µl of ExoSAP-IT PCR Product Cleanup (Applied Biosystem). The mixture was incubated in the thermocycler at 37 ºC for 15 minutes to eliminate residual primers and nucleotides and then the temperature was increased to 80 ºC for 15 minutes. After nanodrop quantification, 100 ng/µl of the purified PCR products were sequenced by Sanger DNA sequencing performed by the "Secuenciación de ADN y análisis de la expresión génica" service in IBMCP in order to validate the vectors.

Table 5.8 List of primer pairs used for the sequencing used for the validation of the final constructs. The primer codes, oligo sequences, oligo length, melting temperatures and GC% are also reported.

Primer	Sequence (5'->3')	Length (bp)	Tm (°C)	GC (%)
pDGB3alpha1_1 LB-Tnos F	GACTGATGGGCTGCCTGTAT	20	59.53	55
pDGB3alpha1_1 LB-Tnos R	AGCGCGCAAACTAGGATAAA	20	57.98	45
pDGB3alpha1_2 SF-35s F	TGAACAAAGAACAATAGTGGATGAA	25	57.18	32
pDGB3alpha1_2 SF-35s R	AAAGGAGATCAGCTTGGCTCT	21	59.09	47.6
pDGB3alpha1_3 RB F	AGACGTCAGGTGGCACTTTT	20	59.82	50
pDGB3alpha1_3 RB R	AAACCTTTTCACGCCCTTTT	20	56.64	40

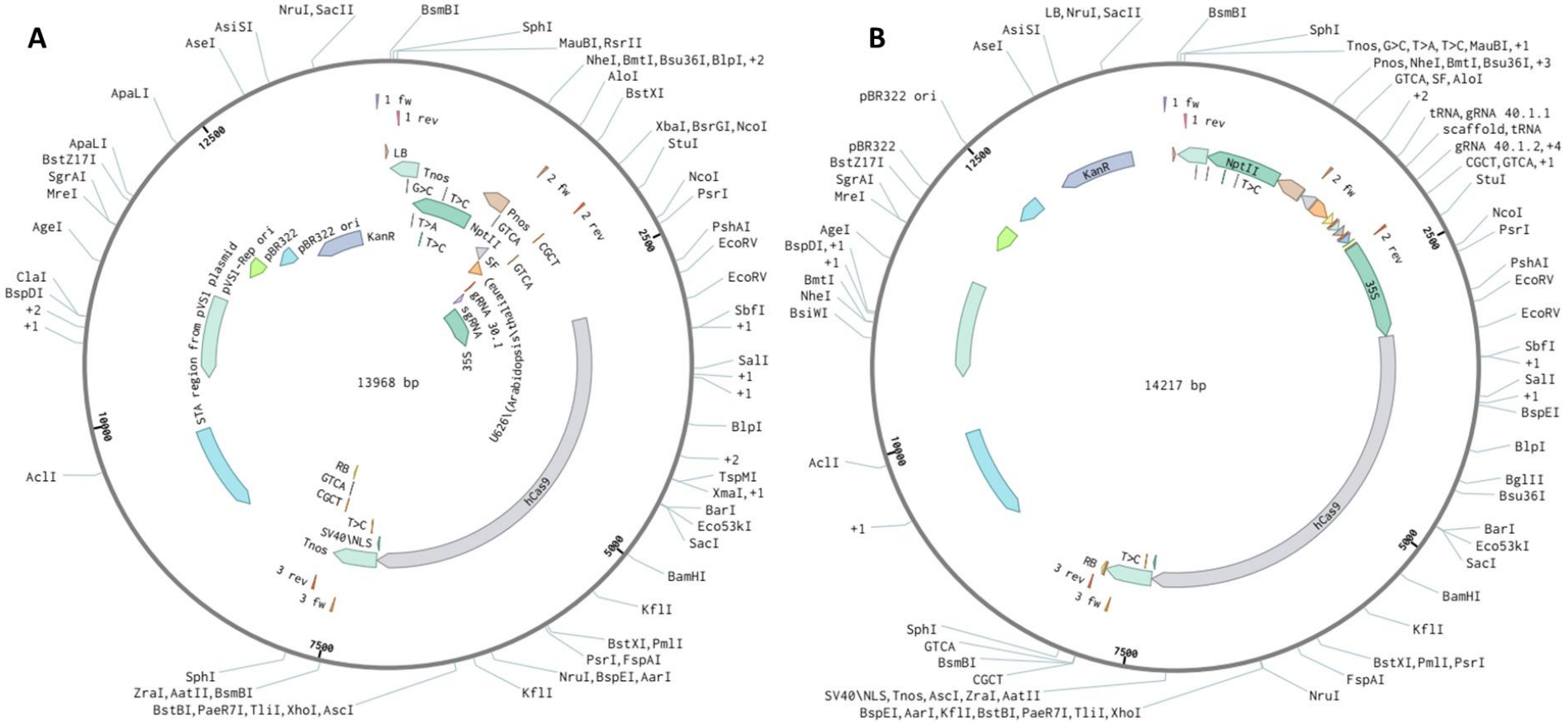


Figure 5.1. Circular map of final plasmids for the genome editing of two genes coding for HSP90, included in introgressed region of IL4-4. A) Final vector for the knock-out of the gene *Solyc04g081630*. B) Final vector for the knock-out of the gene *Solyc04g081640*.

5.1.3.6 Transformation of Escherichia coli TOP-10 cells

After each assembly step, for cloning the constructs aliquots of Top-10 *E. coli* chemically competent cells were thawed in ice until mixing. After adding 1 µl (concentrated 75-100 ng/ µl) of the plasmid DNA to the aliquot, the mixture was gently mixed by pipetting and then incubated for 30 minutes on ice. At the mixture, 300 µl of SOC medium (2 % tryptone, 0.5 % yeast extract, 10 mM sodium chloride, 2.5 mM potassium chloride, 10 mM magnesium chloride, 10 mM magnesium sulphate, and 20 mM glucose) were added and mixed gently. The cell culture was grown at 37 ºC under shaking for 1 hour. After incubation, using a plastic spreader, 50-100 µl of cells were put in LB agar plates (1 % tryptone, 0.5 % yeast extract, 1 % sodium chloride, and 1.5 % agar) containing the selective antibiotics (chloramphenicol at 34 µg/mL, kanamycin, or spectinomycin at 50 µg/mL, based on different assembly steps), 0.5 mM of Isopropyl-β-D-1-thiogalactopyranoside (IPTG) and 20 µg/ml of Xgal. The plates were put for 24 hours at 37 ºC. Finally, the negative colonies (white colour) to IPTG/Xgal assay were regarded as transformed and used for the next step. Single colonies were picked and transferred to sterile LB liquid medium containing the appropriate antibiotic. In order to validate each step of the cloning process, DNA extraction was performed using E.Z.N.A.® HP Plasmid Mini Kit I (Omega Bio-tek, Inc.) according to the manufacturer's instruction.

5.1.4 Agrobacterium transformation

Agrobacterium tumefaciens strain LBA4404 electrocompetent cells were used for the genetic transformation of tomato cotyledon explants. Aliquots of electrocompetent cells were thawed on ice until mixing. After adding 1 µl (concentrated 75-100 ng/µl) of the plasmid DNA, the mixture was gently mixed by pipetting being cautious to not cause air bubbles. The mixture was transferred to an electroporation cuvette and placed on ice. After turning on the electroporator (BTX™-Harvard Apparatus ECM™ 399 electroporator) and setting the power to 2.2 kV, the cuvette was placed in the electroporator holder and the pulse activated. Immediately, 500 µl of LB were added and gently mixed in the cuvette and transferred to a 1.5 ml microcentrifuge tube and incubated for an hour at 37 °C for the recovering. Then, the bacterial solution was spread on Petri dishes containing LB and Kanamycin at 50 µg/ml.

5.1.5 Stable genetic transformation in tomato

In order to work in axenic condition, virus-free seeds of IL4-4 were washed three times with milli-Q double distilled water. Then, seeds were treated with EtOH 70 % for two minutes, followed by a treatment with a commercial sodium hypochlorite at 3 % in agitation for 20 minutes. Next, under flow chamber, the diluted bleach was discarded, and the seeds were washed with milli-Q double

distilled water for five times. One hundred and fifty-five seeds were put in jars containing the germination medium (Table 5.9) in dark condition for 48 hours and, then put in a growth chamber under a photoperiod of 16/8 hours at 24 °C. Next, the cotyledons were cut off with a blade and put on plastic Petri dish in order to dissect the explants by removing the stems and the apexes. The explants were transferred in a new Petri dish containing agrobacterial culture previously pelleted and resuspended to a final 0.2 OD at 600 nm in co-cultivation liquid medium for 30 minutes, as reported in Table 5.9. After the explants were plotted on sterile filter paper and then put onto plates containing solid co-cultivation medium (Table 5.9) for 48 hours in dark condition under controlled temperatures. After two days the explants were transferred in induction medium 1 for ten days and, then, were put in induction medium 2 (Table 5.9) for three weeks. Every three weeks the exhaust medium was changed. Once observed the emergence of buds on the explants, the regenerating material was transferred in elongation medium and then the elongated shoots were put in rooting medium (Table 5.9).

Table 5.9 Composition of the nutritive media used during all the *in vitro* phases to obtain transformed plants.

COMPOUND	Germination Med.	Co-cultivation Med.	Induction 1 Med.	Induction 2 Med.	Elongation Med.	Rooting Med.
MS salt including vitamins	2,5 g/l	4,9 g/l	4,9 g/l	4,9 g/l	4,9 g/l	4,9 g/l
Sucrose	15 g/l					
Glucose		20 g/l	20 g/l	20 g/l	20 g/l	20 g/l
Phyto Agar	10 g/l	10 g/l	10 g/l	10 g/l	10 g/l	10 g/l
IAA		1.0 mg/l	1.0 mg/l	0.1 mg/l		
IBA						0.2 mg/l
Trans-Zeatin		0.75 mg/l	0.75 mg/l	0.75 mg/l	0.1 mg/l	
Acetosyringone		200 μM				
Timentin			150 mg/l	150 mg/l		
Kanamycin			100 mg/l	100 mg/l	100 mg/l	100 mg/l

5.1.6 Validation of transformants

The genomic DNA of putative transgenic plants with two constructs were extracted using DNeasy plant DNA MiniKit (Qiagen) and a PCR reaction was set up in order to confirm the presence of the Cas9 in tomato plants. The primer pairs used for the amplification are reported in Table 5.10.

Table 5.10 Primer pairs used to validate the presence of Cas9 in putative transgenic plants. The primer code, the oligo sequences, the oligo length, the melting temperatures and expected amplicon size are reported

Primer	Sequence 5'-3'	Oligo (bp)	Melting (°C)	Amplicon size
hCas9_FW	CATCGCTAATCTTGCAGGTAG	21	62	2,046
hCas9_RV	GGCAACAGGATTCAATCTTAAG	22	62	

The PCR cycles and reaction conditions are reported in Table 5.11. The putative transgenic plants were evaluated through standard-agarose gel electrophoresis concentrated at 1 %, by observing the expected amplicon length of 2,046 bp.

Table 5.11 PCR reaction and cycles set up to confirm the genetic transformation of plants.

PCR reaction			PCR cycle		
Reagent	Reagent conc.	Volume	Step	Temperature (°C)	Cycles (no.)
DNA	50 ng/µl	1 µl	Denaturation	95	X1
Primer Forward	2.5 µM	2.5 µl	Denaturation	95	X30
Primer Reverse	2.5 µM	2.5 µl	Annealing	58	
Enzyme buffer	5X	5 µl	Extension	72	
dNTPs	10 µM	0,5 µl	Final extension	72	X1
GoTaq (Promega)	5 U/µl	0.2 µl	Storage	4	
H2O		up to 25 µl			

5.2 Results

5.2.1 Expression analysis of two HSP genes

The study on the size of the *S. pimpinellium* introgression region 4-4 carried out in a previous work (Fakhet, 2016) allowed the identification of two HSP genes (*Solyc04g081630* and *Solyc04g081640)* mapping in this region. The expression analysis of these genes was performed on floral tissues and ovaries with the aim of observing their changes when comparing the heat tolerant introgression line IL4-4 and Moneymaker (Figure 5.2). Regarding the gene *Solyc04g081630*, in the anthers (ANT-1 DPA and ANT 0) the expression levels in IL4-4 did not differ statistically from those in Moneymaker, whereas in the ovaries (from -1 DPA to 10 DPA) the expression level increased in IL4-4, varying from 3 to 6-fold respect to Moneymaker. Moreover, an increase of the expression level in ovary at 7 DPA respect to 4 DPA was recorded in IL4-4. After 13 DPA the expression level showed a significant drop in IL4-4 respect to Moneymaker.

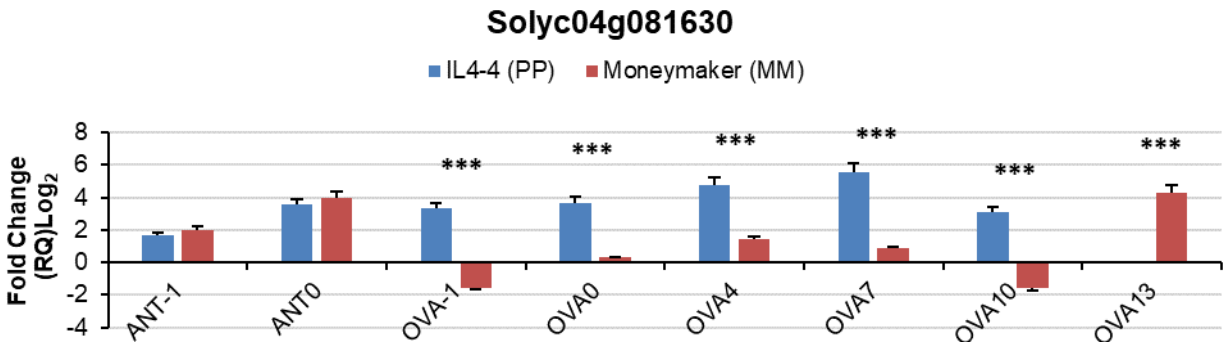


Figure 5.2 Gene expression of the HSP gene *Solyc04g081630* in introgression line IL4-4 and the background genotype Moneymaker. The expression quantification was performed by Real Time RT-PCR in anthers and ovaries of tomato flowers at different Days Post-Anthesis (DPA). Student's t-test denotes statistically significant differences (*, p<0.05; **, p<0.001; ***, p<0.0001). Error bars indicate standard error.

Regarding *Solyc04g081640*, the expression levels resulted significantly higher in all cases in IL4-4 respect to Moneymaker in ovaries, whereas an increase of the gene expression was recorded in Moneymaker after 10 DPA followed by a drop at 13 DPA (Figure 5.3).

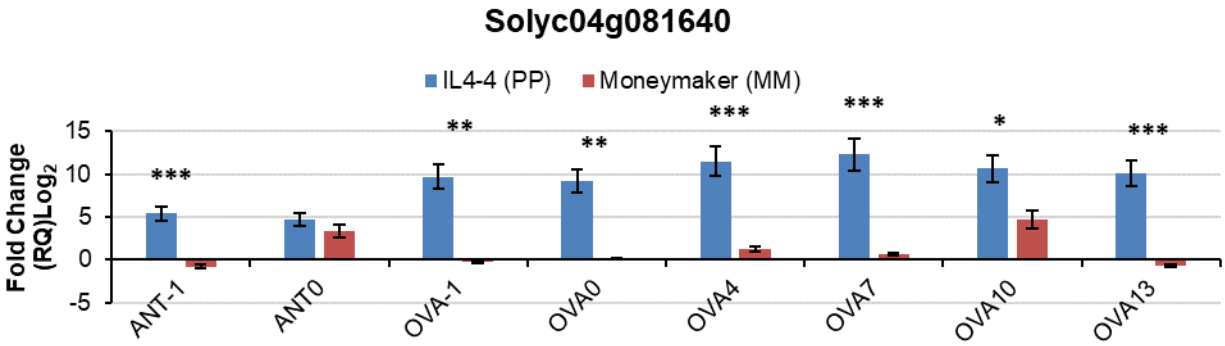


Figure 5.3 Gene expression of the HSP gene *Solyc04g081640* in introgression line IL4-4 and the background genotype Moneymaker. The expression quantification was performed by Real Time RT-PCR in anthers and ovaries of tomato flowers at different Days Post-Anthesis (DPA). Student's t-test denotes statistically significant differences (*, p<0.05; **, p<0.001; ***, p<0.0001). Error bars indicate standard error.

Moreover, the expression level recorded in *Solyc04g081640* showed in ovaries samples the same trends of *Solyc04g081630* with an 9 and 12-fold increase in IL4-4 respect to Moneymaker. From the analysis carried out these two genes could be good candidates to exploit the heat tolerance trait in this introgression line.

5.2.2 Validation of construct

A genetic engineering strategy using CRISPR/Cas9 technologies was undertaken to knock-out these genes and to observe their effect on the progeny. As reported previously, the correct construction of the vector was verified by sequencing three PCR products, targeting different zones of the final construct (Figure 5.4) WHICH include the targeted fragments for each construct. Then, these three amplicons were sequenced using Sanger's platform in order to validate the final constructs.

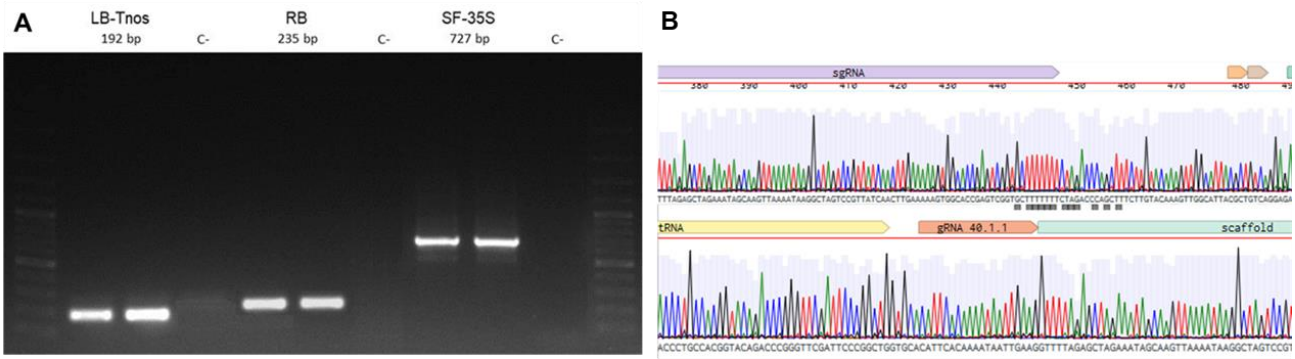


Figure 5.4 PCR carried out on final constructions for the genome editing. A) Amplification of three different regions of the final construct for the knock-out of *Solyc04g081640*. LB-Tnos, amplified region starting from left border to Tnos of the resistance gene. RB, amplified region targeting the right border. SF-35S, amplified region which includes the guides and the Cas9 B) Sanger's sequencing results of the region SF-35S in both the final constructions.

After the validation of the two plasmids, these were transferred in *Agrobacterium tumefaciens* for stable transformation. In this way, in order to obtain cotyledon explants for the genetic transformation, about 150 seeds have been sowed and about 100 seeds germinated. These were separated into two groups of about 50 seeds for the two experiments: from the plants about 60 explants for each genetic transformation were obtained and posed for the regeneration steps. About 30 regenerating explants from both the genetic transformations were obtained (Figure 5.5).

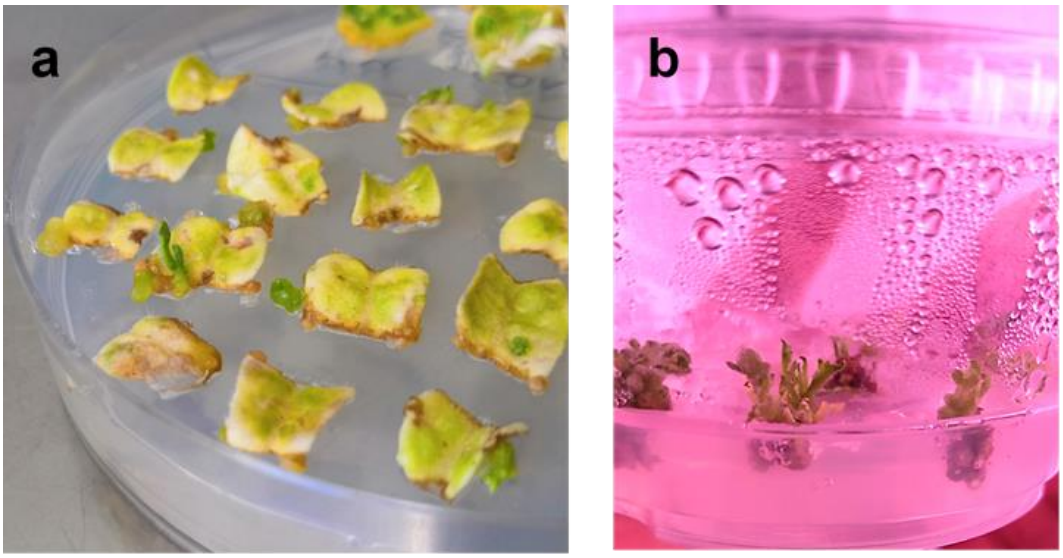


Figure 5.5 *In vitro* culture steps of transformed plant. a) regenerating cotyledon explants, b) plants obtained by regenerated explants

Up till now, four putative transformed plants engineered for *Solyc04g081630* and three for *Solyc04g081640* have been evaluated for the integration of the Cas9 in tomato genome using PCR reaction. Indeed, as shown in Figure 5.6, four transformation events were observed in plants transformed with vector for the editing of the gene *Solyc04g081630* which resulted positive for the amplification of the Cas9 region. Therefore, three different plants showed the Cas9 peculiar amplicon expected after the transformation with the vector for the editing of the gene *Solyc04g081640*.

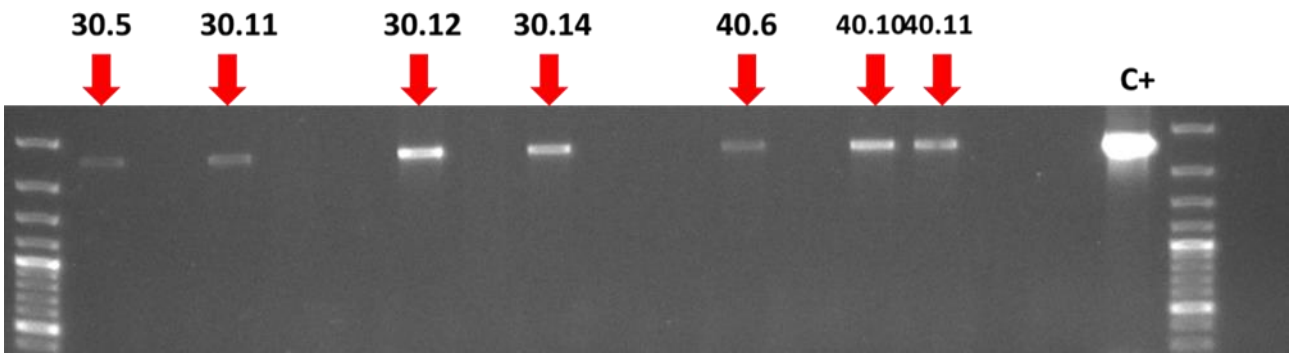


Figure 5.6 PCR amplification carried out on leaves of 16 putative transformation events. The first number is the ID of the gene, (30 for *Solyc04g081630* and 40, *for Solyc04g081640*) and the second one is a progressive number indicating the number of transformant.

Now, the other putative transformant events are in course of validation. Therefore, the validated transformed plants were *in vitro* propagated in order to obtain more clones and were transferred to the greenhouse in order to obtain seeds of the generation T0, for the phenotypic evaluation and gene expression analysis. Up till now about 10 nodal cuttings from each transformed plant were generated. Moreover, plants not yet evaluated are under PCR analysis to validate the transformation events. Then, cloned plants have been put in greenhouse for growth and phenotypic analysis.

6. Discussion

Nowadays, facing the damages of abiotic stresses, such as those deriving from the global warming, is an important aim for the agricultural sciences. The research of innovative tools and the exploration of genetic resources, combined with the use of innovative management practices, could be an effective start point to achieve this goal. As known, heat tolerance is a polygenic trait and the possibility to evaluate genotypes in open fields represent an interesting challenge for improving agronomical traits in different environmental conditions (Hall, 2011). Indeed, studying the heat response under controlled experimental conditions induces specific responses of the genotypes to heat stress thus not allowing the selection under different environmental conditions and the evaluation of interactions between genotype and environment (de Souza et al., 2012). In general, phenotyping is the first step to evaluate the behavior of plants under stress and to select genotypes with desirable traits, even though molecular analyses could favour the selection of the best performing genotypes. In this way, the cost-effectiveness of new sequencing technologies allows the manipulation of large informative dataset of markers that could be useful in identification of genes and/or QTLs involved in heat stress response. Moreover, for the most important crops for which the complete genome is available, such as tomato, the re-sequencing of different varieties can be a good strategy to identify molecular markers that co-segregate with traits affected by high temperatures (HT). The construction of markers dataset could help in the experiment of QTL analysis and association mapping using GWAS approach (Morgante and Salamini, 2003). In this way, combining genotyping data and breeding strategies for the identification of key traits in diverse donor genotypes could help the understanding of the inheritance and molecular genetics of such traits (Bita and Gerats, 2013). In addition, the production of F_1 hybrids could be a successful strategy to improve tomato genotypes and to expand the knowledge on the complex mechanism of the heat tolerance. Finally, other innovative genetic and molecular strategies have been developed in the last years, involving also genetic transformation. In the last decade the genome editing technique had large diffusion and it could be useful for studying key-genes responsible for the control of desirable traits.

Based on these tools, in the present work 32 tomato landraces have been tested, evaluating five different yield-related traits for two years in two different open fields under high temperature conditions. The climatic data recorded during the growing seasons resulted highly variable in the years 2017 and 2018. The evaluation of five yield-related traits highly affected by HT (Sato et al., 2000) allowed to identify genotypes showing desirable traits in terms of high yield performances under high temperatures.

Up till now, the response to heat stress has been usually evaluated by setting fixed conditions in greenhouse or growth chambers (Jegadeesan et al., 2018; Marko et al., 2019; Upadhyay et al., 2019),

applying short or long-term heat stress and then studying reproductive traits (Mesihovic et al., 2016; Müller and Rieu, 2016; Sato et al., 2006; Xu et al., 2017a; Xu et al., 2017b). This approach is not comparable to what happens in open field, where plants are subject to the fluctuation of temperatures and other environmental factors, such as humidity, soil composition, biotic stresses. In the present thesis, among the different evaluated traits, yield was considered the most reliable to estimate heat tolerance, even though changes in yield components as NFL, FS, TNF and FW demonstrated an impact on the yield potential and adaptation of plants to adverse abiotic stresses (Araus et al., 2018; Wessel-Beaver and Scott, 1992). Considering results obtained in the year 2017, FS resulted higher in Campania than in Puglia, even though critical drops of FS were not recorded in the latter region. This could be due to the fluctuation of temperatures during the day, which reached peaks followed by gradual decreases, and this fluctuation probably did not alter FS as usually occurs under fixed high temperatures (Sato et al., 2000). As for YP, the total production recorded in Campania 2018 resulted lower than those of fields evaluated in the year 2017. The principal cause of this strong reduction of YP could be due to pathogen attacks that, combined with the high temperatures recorded during the growing season, caused a dramatic drop in the yield. Since YP is one of the traits used to evaluate the tolerance under HT in order to select the best genotypes, and since it usually shows high fluctuations depending on the environmental conditions (Wessel-Beaver and Scott, 1992). In this way, across the three experimental trials the yield stability analysis revealed that E42 showed YP above the mean of the whole YP values. The selection of the genotype E42 for its high yield stability (Eberhart and Russell, 1966; Kumar et al., 2019) measured across three different trials, is a noteworthy result of. the present work. Indeed, YP of E42 resulted to be comparable to those exhibited by the heat tolerant hybrids DOCET and JAG8810, used as control genotypes. In contrast with data reported in other study, YP exhibited by the evaluated genotypes did not result dependent on both TNF and FW (Ayenan et al., 2019). Indeed, the positive correlation recorded between YP and TNF occurred independently of FW, which varied from very small fruit (E42 and PDVIT) to big one (E17). The selected genotype E42, notwithstanding belongs to a small fruit typology, showed high yield performances, confirming the positive correlation between TNF and YP. Moreover, this genotype is also characterized by the "high NFL" trait that could favour the fruit production, thus contributing positively to the final yield. Therefore, the genotype-dependent trait "high NFL" observed in E42 probably was the most important trait to determine its good and stable performances in multi-environmental high temperature conditions. The exploitation of its "high NFL" trait was evaluated at Instituto de Biologìa molecular y cellular de Plantas (IBMCP) in Valencia also under controlled temperature conditions in greenhouse. The genotype did not result particularly affected by different

temperature regimes in terms of number of flowers, with a significant decrease of about the 32% only at 35°C/30°C day/night temperatures respect to the first thermal condition of 25°C/20°C day/night.
In order to transfer the traits potentially associated to the heat tolerance detected in E42, the breeding strategies could take advantage of the use of cost-effective high-throughput genotyping to study the genetic variability and to carry out association mapping approaches and then the identification of molecular markers/QTLs targeting the traits under study. Therefore, two high-throughput genotyping approaches, the SolCAP array on a wide germplasm collection, and the GBS techniques on a selected group of genotypes, were exploited with two different purposes. The first was the design of markers detecting polymorphisms among genotypes to be used as parental lines of F_1 hybrids, the second was the identification of genes/QTLs associated to yield related traits under high temperature.
In the first case, a group of SNPs identified using an array genotyping platform was transformed in PCR-derived markers (CAPS and dCAPS), useful for the screening of parental lines. Then, retrieving also genotyping data deriving from GBS analysis, the calculation of genetic distances estimated among a group of best performing genotypes allowed to select as parents of F_1 hybrids those that combined good yield under high temperature with good quality traits, and with allelic variability at a higher number of loci. This condition could increase the combination of favourable alleles and therefore the heterosis in the hybrids (Birchler et al., 2010). The progenies of the F_1 cross combinations obtained in this thesis were confirmed by using the same PCR-derived markers and then were tested under high temperatures. For this reason, the YP of the F_1 hybrids were compared against three F_1 hybrids (DOCET, JAG8810 and PAIPAI) that showed high yield performances under high temperatures in the two years. Moreover, four different hybrid combinations were evaluated in two years; however, only three confirmed the positive heterotic effect, in both years and two (17H37 and 17H39) showed high values of heterosis referred to mid-parental values. This behaviour could be the results of the effect of non-additive genes (Solieman et al., 2013). Generally, the breeding programs combine good yield performances and quality traits desirable from the consumers. In some cases, the positive values of heterosis recorded in the F_1 hybrids for the TSSC/TA ratio could reflect various degrees of dominance. Indeed, both dominant and additive gene effect were involved in the inheritance of TSSC (Hannan et al., 2007). In particular, six hybrids combining good yield performances and quality trait were selected in 2019, showing positive heterosis values in YP and quality traits confirming that breeding in terms of constitution of hybrids can be used to improve yield combined to quality in tomato (Hannan et al., 2007).
The second purpose of molecular analyses carried out by the two genotyping platforms was to understand the genetic basis of the better performances exhibited by some genotypes when grown under HT. The yield is a quantitative trait that is affected by many QTLs with minor effects

(Hernández-Bautista et al., 2015; Xu et al., 2017a). For this reason, the GWAS analysis performed on a wide germplasm collection, and reported in Ruggeri *et al*. (2019), allowed to dissect the complex heat tolerance trait that plays an important role in the final yield under HT (Ruggieri et al., 2019). The GWAS carried out on yield-related traits revealed that the highest number of marker-associations was found on chromosomes 1 and 3, in accordance with data reported in previous work that reported the identification of a high number of QTLs associated to heat tolerance were localized on chromosomes 1, 2, and 3 (Xu et al., 2017a). The 14 common markers identified by the GWAS in both fields targeted a number of 11 genes, which were studied in terms of functional annotation (iTAG3.2) and could be involved in the stress response to high temperatures. The gene *Solyc03g123800,* that is close to a SNP associated to FW , codes for a SlMAPPK2 kinase involved in the signal transduction during the abiotic and biotic stresses (Wu et al., 2014a). Moreover, genes included in a region of 500 Kbp upstream and downstream the identified markers were also explored in order to identify candidate genes associated to the analyzed traits. In this way, the *Solyc01g008350,* close to marker CS5, is a DNAJ annotated as heat shock protein and it is reported that plays a function similar to a molecular chaperone, involved in protein folding, translocation and also degradation in stress condition (Park and Seo, 2015). Regarding the marker CS1, it co-localizes with a gene coding for an arginine methyltransferase and resulted associated to NFL. This gene was identified as homologous of AT1G04780, which has a role in floral development as reported in FLOR-ID DB (Bouché et al., 2016). Some developmental stages are affected by temperature fluctuations that occur during anther and pollen development, causing fruit set reduction (Erickson and Markhart, 2002; Giorno et al., 2013) and therefore heat susceptibility. The GWAS study identified a marker (CS3) highly associated with FS and TNF traits. This marker is localized close to the gene *style2.1* (*Solyc02g067380*), a transcription factor controlling the cell elongation, and is a gene responsible of style exertion regulation (Chen et al., 2007). Moreover, this transcription factor is one of the five genes contributing to the complex locus *se2.1* and mapping on chromosome 2 in tomato (Chen and Tanksley, 2004). As for YP, a marker-association (CS14) was found on chromosome 7 close to a minor QTL for yield previously identified in tomato (*fd7.1* or *ty7.1*) (Hernández-Bautista et al., 2015). This marker could explain a small part of YP and could be intriguing to further studying its link to heat tolerance. Interestingly, the search for polymorphisms detected by the GBS analysis in the regions where marker associated to yield-related traits were found using the GWAS approach, revealed that in the region flanking the marker CS14 500 kbp upstream and downstream 49 SNPs/InDels were found and 41 of these resulted private of E42, suggesting a putative role of some of these polymorphisms in the good final yield observed in this genotype. Moreover, in the region targeting the markers CS4 and CS5

associated to TNF, out of 109 SNPs/InDels found, 49 were private of E42, which could explain the high number of TNF recorded in the phenotypic analysis for this genotype.

As a whole, following the high-throughput genotyping characterization of the landraces analysed, E42 and PDVIT showed the highest number of SNP/InDels. Among all mutations, 88 and 1,048 mapped within coding regions and exhibited a high and moderate impact on the corresponding protein, respectively. Searching for genes potentially involved in the response to high temperatures, it was possible to identify mutations in four genes with a disrupting impact on the corresponding protein, three of them showing SNPs/InDels private for the genotype E42. A private mutation that occur in E42 maps on the gene *Solyc07g042660*, coding for a RNA helicase, involved in the regulation of the response to abiotic stresses, as temperatures, salt and light stresses (Mahajan and Tuteja, 2005; Owttrim, 2006). The last two E42's private mutations were found in the gene *Solyc12g044645*, a member of AP2/B3 transcription factor family protein. This gene in *A. thaliana* is homologous to *At1g51120.1* coding for a member of the RAV superfamily that is involved in seed germination, in plant development (Feng et al., 2014) and is involved in different abiotic stresses (Fu et al., 2014). These genes, due to the involvement in the abiotic stress responses could be good candidate genes with the aim to understand the molecular mechanism of the response to high temperatures in tomato. In the GBS dataset, other genes putative linked to tolerance to high temperature. Indeed, two identified genes (*Solyc02g087680* and *Solyc02g087690)* mapping on chromosome 2 code for two members of the FACT subunit complex SSRP1 proteins. In Arabidopsis, the homolog is a regulator of seed dormancy, controls anthocyanin biosynthesis, and is involved in photo-oxidative and UV stresses responses (Pfab et al., 2018).

Considering the mutations with moderate effect, five could be putatively involved in the response to HT. The gene *Solyc01g017220* codes for an ATP synthase epsilon chain, involved in drought stress (Eom et al., 2019). The other genes could have an impact on FS: *Solyc02g070280*, coding for an amino acids transporter active during the pedicel abscission (Xu et al., 2015), and *Solyc07g053640*, which codes for an arabinogalactan-protein involved in abscission processes (Perrakis et al., 2019), and both might impact on FS, a yield-related trait highly affected by high temperatures (Sato et al., 2000). Then, a missense variant was found in *Solyc08g079260*, that codes for a Hsp70 with a tetratricopeptide domain. *Solyc12g043090* also showed variations with moderate effect and it codes for a trihelix transcription factor that is enhanced in response to different abiotic stresses, (Yu et al., 2015) The mutations identified in *Solyc07g053640* and in *Solyc12g043090* were private for E42. Altogether, the genotyping analyses here performed demonstrated that E42 and PDVIT have higher genetic variability than the other genotypes here investigated. The reason of this result could depend on breeding events that involved tomato wild-species, as confirmed in the BLAST analysis carried

out using a group of selected polymorphic InDels. Interestingly, out of 82 assayed markers, 37 showed putative introgressed regions deriving from *S. pimpinellifolium*. Most of these mapped in chromosomes 1 and 2 in both genotypes, whereas other small regions on chromosome 7 could derive from *S. galapagense* in E42. It is reported that the wild species usually show higher pollen viability and pollen number under heat stress conditions (Driedonks et al., 2018). In the case of the high-stable yielding genotype E42, it is possible to attribute the good yield performances under high temperatures to the high NFL, and it could be hypothesized that one or more genes influencing the number of flowers in tomato could exhibit mutations in E42. Indeed, different QTLs and genes are known to affect the inflorescence structure and the number of flowers in tomato (Zhang et al., 2018), in particular two QTLs carrying the *S. pimpinellifolium* allele and mapping on chromosomes 2 and 4 increased the NFL.

As a whole, by combining data deriving from the GWAS and GBS analyses, a group of 169 candidate genes has been identified that could be crucial to enhance production under high temperatures, and thus to be transferred or combined in new improved genotypes. In order to better understand the role of these candidate genes, it would be desirable to verify their action by techniques of gene silencing (Bortesi and Fischer, 2015). In this view, in the present thesis, a genome editing strategy has been applied to verify the involvement in the heat tolerance of two heat shock proteins Hsp90 (*Solyc04g081630* and *Solyc04g081640*) identified in a heat tolerant introgression line obtained by crossing *S. pimpinellifolium* and cv Moneymaker, and available at the IBMCP . As reported in the Tomato Expression Database, (http://ted.bti.cornell.edu/cgi-bin/), in *S. pimpinellifolium* accession LA1589 these two proteins resulted to be expressed in ovaries tissues. In this way, the gene expression analysis of these two genes in the introgression line IL4-4 and in Moneymaker was carried out. The genes showed a similar expression profile in the fertilized ovaries, underlying a putative key-role as helper for the embryo formation. Indeed, the increase of the gene expression level between the 4^{th} to 7^{th} DPA indicate a role during the cellular and globular stage embryo, characterized by active cell divisions, cell elongation, and thus play a crucial role in the early seed development. Moreover the HSPs are reported to be involved in the stabilization and folding of peptides, in cell to cell crosstalk and in pollen tube guidance or embryo-sac disruption upon fecundation (Shimizu and Okada, 2000). Moreover, these two genes were found to be interactors to HSP40 that could promote the thermotolerance in tomato (Roth et al., 2018; Wang et al., 2020). Moreover, this family proteins in particular is responsible of the crosstalk with HSP70 during the heat stress response and contribute to control the level of Heat Shock Factors (HSF) (Hahn et al., 2011). The role of these two genes in determining heat tolerance in tomato has been investigating by using genome editing strategies aimed at silencing them. The choice of the genome editing technique is based on its high precision to induce

mutations. Here, we described the use of this technique to target two HSP90 those are not yet explored, obtaining, up till now, totally seven transformant events. These and the other plants in course of validation represent an important genetic material useful to assess the involvement of the target genes in the heat tolerance mechanisms during the reproductive stages, evaluating them under HT conditions. They are very promising for future perspective, representing an exploitable biotechnological tool applied in the complex trait of the heat tolerance.

7. Conclusions

The objective of the present thesis were the selection of genotypes showing good yield under high temperature and the use of genomic resources to identify molecular markers targeting candidate genes and/or QTLs involved in the response to high temperatures.

The phenotyping study conducted in two years and in two locations allowed to select seven genotypes (E20, E36, E42, E55, LA2662, LA3120 and PDVIT) for their best performances in terms of yield in different growing conditions. Moreover, the yield stability analysis revealed that E42 showed the "stable-yielding" trait and the final yield *per* plant similar to that of the heat tolerant commercial F_1 hybrid DOCET. These genotypes were subsequently investigated for their genetic variability and exploited for the production of F_1 hybrids, with the aims of combining different desirable traits for tomato breeding.

The genotyping experiments carried out using data deriving from the platform Infinium SolCAP SNP array and from a Genotyping by Sequencing (GBS) approach resulted useful to identify molecular markers associated to yield-related traits under high temperatures. The GWAS revealed 14 common associations to the six phenotypic traits analyzed, and the GBS approach allowed to explore the genetic variability of a group of genotypes, evidencing in particular that PDVIT and the stable-yielding genotype E42 are the most polymorphic genotypes. Moreover, exploring the genetic variability of specific InDel markers in these two genotypes, the presence of putative wild introgressed regions of *S. pimpinellifolium* was evidenced. The analysis carried out by these two high-throughput genotypic platforms allowed to identify totally 169 candidate genes, among which particularly interesting are the *Solyc12g038540* and *Solyc07g042660* with private mutations of E42, whose involvement in the response to high temperature could be verified in the future, using techniques of genome editing.

Finally, the potential transfer of genes affecting a good response to high temperature was investigated by evaluating a set of F_1 hybrids obtained by crosses involving genotypes selected for good yield performances under heat conditions. A group of four hybrids evaluated for two years under high temperatures evidenced that high heterotic effects could be obtained from different parental combinations, and, in particular, two hybrids (17H37 and 17H39) showed also a high heterotic percentage for both yield and fruit quality traits.

Comprehensively, in the present study the combination of a rigorous phenotypic work in multi-environmental conditions and of the acquisition of genotypic data represented a successful choice to enhance innovative breeding programs for the selection of genotypes and hybrids under harsh temperature conditions.

8. References

1. Abdul-Baki, A. A. (1991). Tolerance of tomato cultivars and selected germplasm to heat stress. *Journal of the American Society for Horticultural Science* **116**, 1113-1116.
2. Ainsworth, E. A., and Ort, D. R. (2010). How do we improve crop production in a warming world? *Plant physiology* **154**, 526-530.
3. Annex, I. (2012). Managing the risks of extreme events and disasters to advance climate change adaptation. *Sciences* **10**, 97-104.
4. Arafa, R. A., Rakha, M. T., Soliman, N. E. K., Moussa, O. M., Kamel, S. M., and Shirasawa, K. (2017). Rapid identification of candidate genes for resistance to tomato late blight disease using next-generation sequencing technologies. *PloS one* **12**.
5. Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., and Cairns, J. E. (2018). Translating high-throughput phenotyping into genetic gain. *Trends in plant science* **23**, 451-466.
6. Ayenan, M. A. T., Danquah, A., Hanson, P., Ampomah-Dwamena, C., Sodedji, F. A. K., Asante, I. K., and Danquah, E. Y. (2019). Accelerating Breeding for Heat Tolerance in Tomato (Solanum lycopersicum L.): An Integrated Approach. *Agronomy* **9**, 720.
7. Bai, Y., and Lindhout, P. (2007). Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? *Annals of botany* **100**, 1085-1094.
8. Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., and Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one* **3**.
9. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712.
10. Barrantes, W., Fernández-del-Carmen, A., López-Casado, G., González-Sánchez, M. Á., Fernández-Muñoz, R., Granell, A., and Monforte, A. J. (2014). Highly efficient genomics-assisted development of a library of introgression lines of Solanum pimpinellifolium. *Molecular breeding* **34**, 1817-1831.
11. Barua, D., Downs, C. A., and Heckathorn, S. A. (2003). Variation in chloroplast small heat-shock protein function is a major determinant of variation in thermotolerance of photosynthetic electron transport among ecotypes of Chenopodium album. *Functional Plant Biology* **30**, 1071-1079.
12. Bauchet, G., Grenier, S., Samson, N., Bonnet, J., Grivet, L., and Causse, M. (2017). Use of modern tomato breeding germplasm for deciphering the genetic control of agronomical traits by Genome Wide Association study. *Theoretical and applied genetics* **130**, 875-889.
13. Bergelson, J., and Roux, F. (2010). Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. *Nature Reviews Genetics* **11**, 867-879.
14. Bhatt, R., Biswas, V., and Kumar, N. (2001). Heterosis, combining ability and genetics for vitamin C, total soluble solids and yield in tomato (Lycopersicon esculentum) at 1700 m altitude. *The Journal of Agricultural Science* **137**, 71-75.
15. Birchler, J. A., Yao, H., Chudalayandi, S., Vaiman, D., and Veitia, R. A. (2010). Heterosis. *The Plant Cell* **22**, 2105-2112.
16. Bita, C., and Gerats, T. (2013). Plant tolerance to high temperature in a changing environment: scientific fundamentals and production of heat stress-tolerant crops. *Frontiers in plant science* **4**, 273.
17. Blanca, J., Canizares, J., Cordero, L., Pascual, L., Diez, M. J., and Nuez, F. (2012). Variation revealed by SNP genotyping and morphology provides insight into the origin of the tomato. *PloS one* **7**.

18. Blanca, J., Montero-Pau, J., Sauvage, C., Bauchet, G., Illa, E., Díez, M. J., Francis, D., Causse, M., Van der Knaap, E., and Cañizares, J. (2015). Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *BMC genomics* **16**, 257.
19. Bokszczanin, K. L., Fragkostefanakis, S., Bostan, H., Bovy, A., Chaturvedi, P., Chiusano, M. L., Firon, N., Iannacone, R., Jegadeesan, S., and Klaczynskid, K. (2013). Perspectives on deciphering mechanisms underlying plant heat stress response and thermotolerance. *Frontiers in plant science* **4**, 315.
20. Bortesi, L., and Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnology advances* **33**, 41-52.
21. Bouché, F., Lobet, G., Tocquin, P., and Périlleux, C. (2016). FLOR-ID: an interactive database of flowering-time gene networks in Arabidopsis thaliana. *Nucleic Acids Research* **44**, D1167-D1171.
22. Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633-2635.
23. Buzgo, M., Soltis, D. E., Soltis, P. S., and Ma, H. (2004). Towards a comprehensive integration of morphological and genetic studies of floral development. *Trends in plant science* **9**, 164-173.
24. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular ecology* **22**, 3124-3140.
25. Chen, K.-Y., Cong, B., Wing, R., Vrebalov, J., and Tanksley, S. D. (2007). Changes in regulation of a transcription factor lead to autogamy in cultivated tomatoes. *Science* **318**, 643-645.
26. Chen, K.-Y., and Tanksley, S. D. (2004). High-resolution mapping and functional analysis of se2. 1: a major stigma exsertion quantitative trait locus associated with the evolution from allogamy to autogamy in the genus Lycopersicon. *Genetics* **168**, 1563-1573.
27. Chetelat, R. T., and Petersen, J. (2005). Revised list of miscellaneous stocks. *Tomato Genet. Cooperative Rep* **55**, 48-69.
28. Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A. J., and Voytas, D. F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757-761.
29. Craig, D. W., Pearson, J. V., Szelinger, S., Sekar, A., Redman, M., Corneveaux, J. J., Pawlowski, T. L., Laub, T., Nunn, G., and Stephan, D. A. (2008). Identification of genetic variants using bar-coded multiplexed sequencing. *Nature methods* **5**, 887.
30. Cronn, R., Liston, A., Parks, M., Gernandt, D. S., Shen, R., and Mockler, T. (2008). Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic acids research* **36**, e122-e122.
31. D'Ambrosio, C., Stigliani, A. L., and Giorio, G. (2018). CRISPR/Cas9 editing of carotenoid genes in tomato. *Transgenic research* **27**, 367-378.
32. Dane, F., Hunter, A. G., and Chambliss, O. L. (1991). Fruit set, pollen fertility, and combining ability of selected tomato genotypes under high-temperature field conditions. *Journal of the American Society for Horticultural Science* **116**, 906-910.
33. Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., and Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158.
34. Darwin, C. (1868). "ANIMALS AND PLANTS," Murray, London.
35. de Souza, M. A., Pimentel, A. J. B., and Ribeiro, G. (2012). Breeding for heat-stress tolerance. *In* "Plant breeding for abiotic stress tolerance", pp. 137-156. Springer.
36. Driedonks, N., Wolters-Arts, M., Huber, H., de Boer, G.-J., Vriezen, W., Mariani, C., and Rieu, I. (2018). Exploring the natural variation for reproductive thermotolerance in wild tomato species. *Euphytica* **214**, 67.

37. Eberhart, S. t., and Russell, W. (1966). Stability parameters for comparing varieties 1. *Crop science* **6**, 36-40.
38. Eom, S. H., Lee, H. J., Lee, J. H., Wi, S. H., Kim, S. K., and Hyun, T. K. (2019). Identification and Functional Prediction of Drought-Responsive Long Non-Coding RNA in Tomato. *Agronomy* **9**, 629.
39. Erickson, A., and Markhart, A. (2002). Flower developmental stage and organ sensitivity of bell pepper (Capsicum annuum L.) to elevated temperature. *Plant, Cell & Environment* **25**, 123-130.
40. Fakhet, D. (2016). High resolution mapping of a locus in chromosome 4 causing distorted segregation in tomato.
41. Feng, C. Z., Chen, Y., Wang, C., Kong, Y. H., Wu, W. H., and Chen, Y. F. (2014). Arabidopsis RAV 1 transcription factor, phosphorylated by S n RK 2 kinases, regulates the expression of ABI 3, ABI 4, and ABI 5 during seed germination and early seedling development. *The Plant Journal* **80**, 654-668.
42. Firon, N., Shaked, R., Peet, M., Pharr, D., Zamski, E., Rosenfeld, K., Althan, L., and Pressman, E. (2006). Pollen grains of heat tolerant tomato cultivars retain higher carbohydrate concentration under heat stress conditions. *Scientia Horticulturae* **109**, 212-217.
43. Fu, M., Kang, H. K., Son, S.-H., Kim, S.-K., and Nam, K. H. (2014). A subset of Arabidopsis RAV transcription factors modulates drought and salt stress responses independent of ABA. *Plant and Cell Physiology* **55**, 1892-1904.
44. Giorno, F., Wolters-Arts, M., Mariani, C., and Rieu, I. (2013). Ensuring reproduction at high temperatures: the heat stress response during anther and pollen development. *Plants* **2**, 489-506.
45. Gorantla, M., Babu, P., Reddy Lachagari, V., Reddy, A., Wusirika, R., Bennetzen, J. L., and Reddy, A. R. (2007). Identification of stress-responsive genes in an indica rice (Oryza sativa L.) using ESTs generated from drought-stressed seedlings. *Journal of experimental botany* **58**, 253-265.
46. Gourdji, S. M., Sibley, A. M., and Lobell, D. B. (2013). Global crop exposure to critical high temperatures in the reproductive period: historical trends and future projections. *Environmental Research Letters* **8**, 024041.
47. Grilli, G. V. G., Braz, L. T., and Lemos, E. G. M. (2007). QTL identification for tolerance to fruit set in tomato by fAFLP markers. *Crop Breeding and Applied Biotechnology*, 234-241.
48. Gurley, W. B. (2000). HSP101: a key component for the acquisition of thermotolerance in plants. Am Soc Plant Biol.
49. Hahn, A., Bublak, D., Schleiff, E., and Scharf, K.-D. (2011). Crosstalk between Hsp90 and Hsp70 chaperones and heat stress transcription factors in tomato. *The Plant Cell* **23**, 741-755.
50. Hall, A. E. (2011). Breeding cowpea for future climates. *Crop adaptation to climate change*, 340-355.
51. Hamilton, J. P., Sim, S.-C., Stoffel, K., Van Deynze, A., Buell, C. R., and Francis, D. M. (2012). Single nucleotide polymorphism discovery in cultivated tomato via sequencing by synthesis. *The Plant Genome* **5**, 17-29.
52. Hannan, M., Ahmed, M., Razvy, M., Karim, R., Khatun, M., Haydar, A., Hossain, M., and Roy, U. (2007). Heterosis and correlation of yield and yield components in tomato (Lycopersicon esulentum Mill.). *American-Eurasian Journal of Scientific Research* **2**, 146-150.
53. Hasanuzzaman, M., Nahar, K., Alam, M., Roychowdhury, R., and Fujita, M. (2013). Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants. *International journal of molecular sciences* **14**, 9643-9684.
54. Hazra, P., Ansary, S., Dutta, A., Balacheva, E., and Atanassova, B. (2008). Breeding tomato tolerant to high temperature stress. *In* "IV Balkan Symposium on Vegetables and Potatoes 830", pp. 241-248.

55. Hedhly, A., Hormaza, J. I., and Herrero, M. (2009). Global warming and sexual plant reproduction. *Trends in plant science* **14**, 30-36.
56. Hernández-Bautista, A., Lobato-Ortiz, R., Cruz-Izquierdo, S., García-Zavala, J. J., Chávez-Servia, J. L., Hernández-Leal, E., and Bonilla-Barrientos, O. (2015). Fruit size QTLs affect in a major proportion the yield in tomato. *Chilean journal of agricultural research* **75**, 402-409.
57. Howarth, C. (2005). Genetic improvements of tolerance to high temperature. In 'Abiotic stresses–plant resistance through breeding and molecular approaches'.(Eds M Ashraf, PJC Harris) pp. 277–300. The Haworth Press: New York.
58. Iba, K. (2002). Acclimative response to temperature stress in higher plants: approaches of gene engineering for temperature tolerance. *Annual Review of Plant Biology* **53**, 225-245.
59. Jankele, R., and Svoboda, P. (2014). TAL effectors: tools for DNA targeting. *Briefings in functional genomics* **13**, 409-419.
60. Jegadeesan, S., Beery, A., Altahan, L., Meir, S., Pressman, E., and Firon, N. (2018). Ethylene production and signaling in tomato (Solanum lycopersicum) pollen grains is responsive to heat stress conditions. *Plant reproduction* **31**, 367-383.
61. Jiang, K., Liberatore, K. L., Park, S. J., Alvarez, J. P., and 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**.
62. Kim, Y.-G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences* **93**, 1156-1160.
63. Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant methods* **9**, 29.
64. Krieger, U., Lippman, Z. B., and Zamir, D. (2010). The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nature genetics* **42**, 459.
65. Kumar, R., Singh, S., and Srivastava, K. (2019). Stability Analysis in Tomato Inbreds and Their F 1 s for Yield and Quality Traits. *Agricultural Research* **8**, 141-147.
66. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics* **25**, 1754-1760.
67. Li, X., Wang, Y., Chen, S., Tian, H., Fu, D., Zhu, B., Luo, Y., and Zhu, H. (2018). Lycopene is enriched in tomato fruit by CRISPR/Cas9-mediated multiplex genome editing. *Frontiers in plant science* **9**, 559.
68. Liu, J. p., Chow, S. C., and Hsieh, T. C. (2009). Deviations from linearity in statistical evaluation of linearity in assay validation. *Journal of Chemometrics: A Journal of the Chemometrics Society* **23**, 487-494.
69. Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *methods* **25**, 402-408.
70. Mahajan, S., and Tuteja, N. (2005). Cold, salinity and drought stresses: an overview. *Archives of biochemistry and biophysics* **444**, 139-158.
71. Marko, D., El-shershaby, A., Carriero, F., Summerer, S., Petrozza, A., Iannacone, R., Schleiff, E., and Fragkostefanakis, S. (2019). Identification and Characterization of a Thermotolerant TILLING Allele of Heat Shock Binding Protein 1 in Tomato. *Genes* **10**, 516.
72. Meissner, R., Jacobson, Y., Melamed, S., Levyatuv, S., Shalev, G., Ashri, A., Elkind, Y., and Levy, A. (1997). A new model system for tomato genetics. *The Plant Journal* **12**, 1465-1472.
73. Mesihovic, A., Iannacone, R., Firon, N., and Fragkostefanakis, S. (2016). Heat stress regimes for the investigation of pollen thermotolerance in crop plants. *Plant reproduction* **29**, 93-105.
74. Miroshnichenko, S., Tripp, J., Nieden, U. z., Neumann, D., Conrad, U., and Manteuffel, R. (2005). Immunomodulation of function of small heat shock proteins prevents their assembly into heat stress granules and results in cell death at sublethal temperatures. *The Plant Journal* **41**, 269-281.

75. Morgante, M., and Salamini, F. (2003). From plant genomics to breeding practice. *Current Opinion in Biotechnology* **14**, 214-219.
76. Müller, F., and Rieu, I. (2016). Acclimation to high temperature during pollen development. *Plant reproduction* **29**, 107-118.
77. Naika, M., Shameer, K., Mathew, O. K., Gowda, R., and Sowdhamini, R. (2013). STIFDB2: an updated version of plant stress-responsive transcription factor database with additional stress signals, stress-responsive transcription factor binding sites and stress-responsive genes in Arabidopsis and rice. *Plant and Cell Physiology* **54**, e8-e8.
78. Neta-Sharir, I., Isaacson, T., Lurie, S., and Weiss, D. (2005). Dual role for tomato heat shock protein 21: protecting photosystem II from oxidative stress and promoting color changes during fruit maturation. *The Plant Cell* **17**, 1829-1838.
79. Neumann, D., Emmermann, M., Thierfelder, J.-M., Zur Nieden, U., Clericus, M., Braun, H.-P., Nover, L., and Schmitz, U. (1993). HSP68—a DnaK-like heat-stress protein of plant mitochondria. *Planta* **190**, 32-43.
80. Nieto-Sotelo, J., Martínez, L. M., Ponce, G., Cassab, G. I., Alagón, A., Meeley, R. B., Ribaut, J.-M., and Yang, R. (2002). Maize HSP101 plays important roles in both induced and basal thermotolerance and primary root growth. *The Plant Cell* **14**, 1621-1633.
81. Opeña, R., Chen, J., Kuo, C., and Chen, H. (1992). Genetic and physiological aspects of tropical adaptation in tomato. *In* "Adaptation of food crops to temperature and water stress: proceedings of an international symposium", pp. 13-18.
82. Owttrim, G. W. (2006). RNA helicases and abiotic stress. *Nucleic Acids Research* **34**, 3220-3230.
83. Palpant, N., and Dudzinski, D. (2013). Zinc finger nucleases: looking toward translation. *Gene therapy* **20**, 121-127.
84. Park, C.-J., and Seo, Y.-S. (2015). Heat shock proteins: a review of the molecular chaperones for plant immunity. *The plant pathology journal* **31**, 323.
85. Paupière, M. J., van Haperen, P., Rieu, I., Visser, R. G., Tikunov, Y. M., and Bovy, A. G. (2017). Screening for pollen tolerance to high temperatures in tomato. *Euphytica* **213**, 130.
86. Peet, M., Sato, S., and Gardner, R. (1998). Comparing heat stress effects on male-fertile and male-sterile tomatoes. *Plant, cell & environment* **21**, 225-231.
87. Perrakis, A., Bita, E.-C., Arhondakis, S., Krokida, A., Mekkaoui, K., Blazakis, K. N., Kaloudas, D., and Kalaitzis, P. (2019). Suppression of a Prolyl 4 hydroxylase results in delayed abscission of overripe tomato fruits. *Frontiers in plant science* **10**, 348.
88. Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one* **7**.
89. Pfab, A., Breindl, M., and Grasser, K. D. (2018). The Arabidopsis histone chaperone FACT is required for stress-induced expression of anthocyanin biosynthetic genes. *Plant molecular biology* **96**, 367-374.
90. Prasinos, C., Krampis, K., Samakovli, D., and Hatzopoulos, P. (2005). Tight regulation of expression of two Arabidopsis cytosolic Hsp90 genes during embryo development. *Journal of experimental botany* **56**, 633-644.
91. Queitsch, C., Hong, S.-W., Vierling, E., and Lindquist, S. (2000). Heat shock protein 101 plays a crucial role in thermotolerance in Arabidopsis. *The Plant Cell* **12**, 479-492.
92. Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842.
93. Rao, A. V., and Agarwal, S. (2000). Role of antioxidant lycopene in cancer and heart disease. *Journal of the American College of Nutrition* **19**, 563-569.
94. Roth, M., Florez-Rueda, A. M., Paris, M., and Städler, T. (2018). Wild tomato endosperm transcriptomes reveal common roles of genomic imprinting in both nuclear and cellular endosperm. *The Plant Journal* **95**, 1084-1101.

95. Rudich, J., Zamski, E., and Regev, Y. (1977). Genotypic variation for sensitivity to high temperature in the tomato: pollination and fruit set. *Botanical Gazette* **138**, 448-452.
96. Ruggieri, V., Anzar, I., Paytuvi, A., Calafiore, R., Cigliano, R. A., Sanseverino, W., and Barone, A. (2017). Exploiting the great potential of Sequence Capture data by a new tool, SUPER-CAP. *Dna Research* **24**, 81-91.
97. Ruggieri, V., Calafiore, R., Schettini, C., Rigano, M. M., Olivieri, F., Frusciante, L., and Barone, A. (2019). Exploiting genetic and genomic resources to enhance heat-tolerance in tomatoes. *Agronomy* **9**, 22.
98. Ruggieri, V., Francese, G., Sacco, A., D'Alessandro, A., Rigano, M. M., Parisi, M., Milone, M., Cardi, T., Mennella, G., and Barone, A. (2014). An association mapping approach to identify favourable alleles for tomato fruit quality breeding. *BMC plant biology* **14**, 337.
99. Sacco, A., Ruggieri, V., Parisi, M., Festa, G., Rigano, M. M., Picarella, M. E., Mazzucato, A., and Barone, A. (2015). Exploring a tomato landraces collection for fruit-related traits by the aid of a high-throughput genomic platform. *PLoS One* **10**.
100. Sato, S., Kamiyama, M., Iwata, T., Makita, N., Furukawa, H., and Ikeda, H. (2006). Moderate increase of mean daily temperature adversely affects fruit set of Lycopersicon esculentum by disrupting specific physiological processes in male reproductive development. *Annals of Botany* **97**, 731-738.
101. Sato, S., Peet, M., and Thomas, J. (2000). Physiological factors limit fruit set of tomato (Lycopersicon esculentum Mill.) under chronic, mild heat stress. *Plant, Cell & Environment* **23**, 719-726.
102. Sato, S., Peet, M. M., and Gardner, R. G. (2004). Altered flower retention and developmental patterns in nine tomato cultivars under elevated temperature. *Scientia Horticulturae* **101**, 95-101.
103. Schöffl, F., Prandl, R., and Reindl, A. (1999). Molecular responses to heat stress. *Molecular responses to cold, drought, heat and salt stress in higher plants* **83**, 93.
104. Shimizu, K. K., and Okada, K. (2000). Attractive and repulsive interactions between female and male gametophytes in Arabidopsis pollen tube guidance. *Development* **127**, 4511-4518.
105. 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., and Gardner, R. G. (2012). High-density SNP genotyping of tomato (Solanum lycopersicum L.) reveals patterns of genetic variation due to breeding. *PloS one* **7**.
106. Solieman, T., El-Gabry, M., and Abido, A. (2013). Heterosis, potence ratio and correlation of some important characters in tomato (Solanum lycopersicum L.). *Scientia Horticulturae* **150**, 25-30.
107. Swindell, W. R., Huebner, M., and Weber, A. P. (2007). Transcriptional profiling of Arabidopsis heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways. *BMC genomics* **8**, 125.
108. Thakur, P., Kumar, S., Malik, J. A., Berger, J. D., and Nayyar, H. (2010). Cold stress effects on reproductive development in grain crops: an overview. *Environmental and Experimental Botany* **67**, 429-443.
109. Ueta, R., Abe, C., Watanabe, T., Sugano, S. S., Ishihara, R., Ezura, H., Osakabe, Y., and Osakabe, K. (2017). Rapid breeding of parthenocarpic tomato plants using CRISPR/Cas9. *Scientific reports* **7**, 1-8.
110. Upadhyay, R. K., Handa, A. K., and Mattoo, A. K. (2019). Transcript Abundance Patterns of 9- and 13-Lipoxygenase Subfamily Gene Members in Response to Abiotic Stresses (Heat, Cold, Drought or Salt) in Tomato (Solanum lycopersicum L.) Highlights Member-Specific Dynamics Relevant to Each Stress. *Genes* **10**, 683.
111. Víquez-Zamora, M., Vosman, B., van de Geest, H., Bovy, A., Visser, R. G., Finkers, R., and Van Heusden, A. W. (2013). Tomato breeding in the genomics era: insights from a SNP array. *BMC genomics* **14**, 354.

112. Wahid, A., Gelani, S., Ashraf, M., and Foolad, M. R. (2007). Heat tolerance in plants: an overview. *Environmental and experimental botany* **61**, 199-223.
113. Wang, J., Wang, Z., Du, X., Yang, H., Han, F., Han, Y., Yuan, F., Zhang, L., Peng, S., and Guo, E. (2017). A high-density genetic map and QTL analysis of agronomic traits in foxtail millet [Setaria italica (L.) P. Beauv.] using RAD-seq. *Plos one* **12**.
114. Wang, X., Zhang, H., Xie, Q., Liu, Y., Lv, H., Bai, R., Ma, R., Li, X., Zhang, X., and Guo, Y.-D. (2020). SlSNAT interacts with HSP40, a molecular chaperone, to regulate melatonin biosynthesis and promote thermotolerance in tomato. *Plant and Cell Physiology*.
115. Warland, J., McKeown, A. W., and McDonald, M. R. (2006). Impact of high air temperatures on Brassicacae crops in southern Ontario. *Canadian journal of plant science* **86**, 1209-1215.
116. Warren, G. (1998). Spectacular increases in crop yields in the United States in the twentieth century. *Weed Technology* **12**, 752-760.
117. Wessel-Beaver, L., and Scott, J. (1992). Genetic variability of fruit set, fruit weight, and yield in a tomato population grown in two high-temperature environments. *Journal of the American Society for Horticultural Science* **117**, 867-870.
118. Wu, J., Wang, J., Pan, C., Guan, X., Wang, Y., Liu, S., He, Y., Chen, J., Chen, L., and Lu, G. (2014a). Genome-wide identification of MAPKK and MAPKKK gene families in tomato and transcriptional profiling analysis during development and stress response. *PloS one* **9**.
119. Wu, K., Liu, H., Yang, M., Tao, Y., Ma, H., Wu, W., Zuo, Y., and Zhao, Y. (2014b). High-density genetic map construction and QTLs analysis of grain yield-related traits in Sesame (Sesamum indicum L.) based on RAD-Seq techonology. *BMC plant biology* **14**, 274.
120. Xiao, B., Tan, Y., Long, N., Chen, X., Tong, Z., Dong, Y., and Li, Y. (2015). SNP-based genetic linkage map of tobacco (Nicotiana tabacum L.) using next-generation RAD sequencing. *Journal of Biological Research-Thessaloniki* **22**, 11.
121. Xiao, H., Radovich, C., Welty, N., Hsu, J., Li, D., Meulia, T., and van der Knaap, E. (2009). Integration of tomato reproductive developmental landmarks and expression profiles, and the effect of SUN on fruit shape. *BMC Plant Biology* **9**, 49.
122. Xu, J., Driedonks, N., Rutten, M. J., Vriezen, W. H., de Boer, G.-J., and Rieu, I. (2017a). Mapping quantitative trait loci for heat tolerance of reproductive traits in tomato (Solanum lycopersicum). *Molecular Breeding* **37**, 58.
123. Xu, J., Wolters-Arts, M., Mariani, C., Huber, H., and Rieu, I. (2017b). Heat stress affects vegetative and reproductive performance and trait correlations in tomato (Solanum lycopersicum). *Euphytica* **213**, 156.
124. Xu, T., Wang, Y., Liu, X., Lv, S., Feng, C., Qi, M., and Li, T. (2015). Small RNA and degradome sequencing reveals microRNAs and their targets involved in tomato pedicel abscission. *Planta* **242**, 963-984.
125. Yagi, M., Shirasawa, K., Waki, T., Kume, T., Isobe, S., Tanase, K., and Yamaguchi, H. (2017). Construction of an SSR and RAD marker-based genetic linkage map for carnation (Dianthus caryophyllus L.). *Plant molecular biology reporter* **35**, 110-117.
126. Yang, J.-D., Yun, J.-Y., Zhang, T.-H., and Zhao, H.-L. (2006a). Presoaking with nitric oxide donor SNP alleviates heat shock damages in mung bean leaf discs. *Bot Stud* **47**, 129-136.
127. Yang, K. A., Lim, C. J., Hong, J. K., Park, C. Y., Cheong, Y. H., Chung, W. S., Lee, K. O., Lee, S. Y., Cho, M. J., and Lim, C. O. (2006b). Identification of cell wall genes modified by a permissive high temperature in Chinese cabbage. *Plant science* **171**, 175-182.
128. Yost, H. J., and Lindquist, S. (1986). RNA splicing is interrupted by heat shock and is rescued by heat shock protein synthesis. *Cell* **45**, 185-193.
129. Yu, C., Cai, X., Ye, Z., and Li, H. (2015). Genome-wide identification and expression profiling analysis of trihelix gene family in tomato. *Biochemical and biophysical research communications* **468**, 653-659.

130. Zhang, S., Yu, H., Wang, K., Zheng, Z., Liu, L., Xu, M., Jiao, Z., Li, R., Liu, X., and Li, J. (2018). Detection of major loci associated with the variation of 18 important agronomic traits between Solanum pimpinellifolium and cultivated tomatoes. *The Plant Journal* **95**, 312-323.

9. Supplementary Material

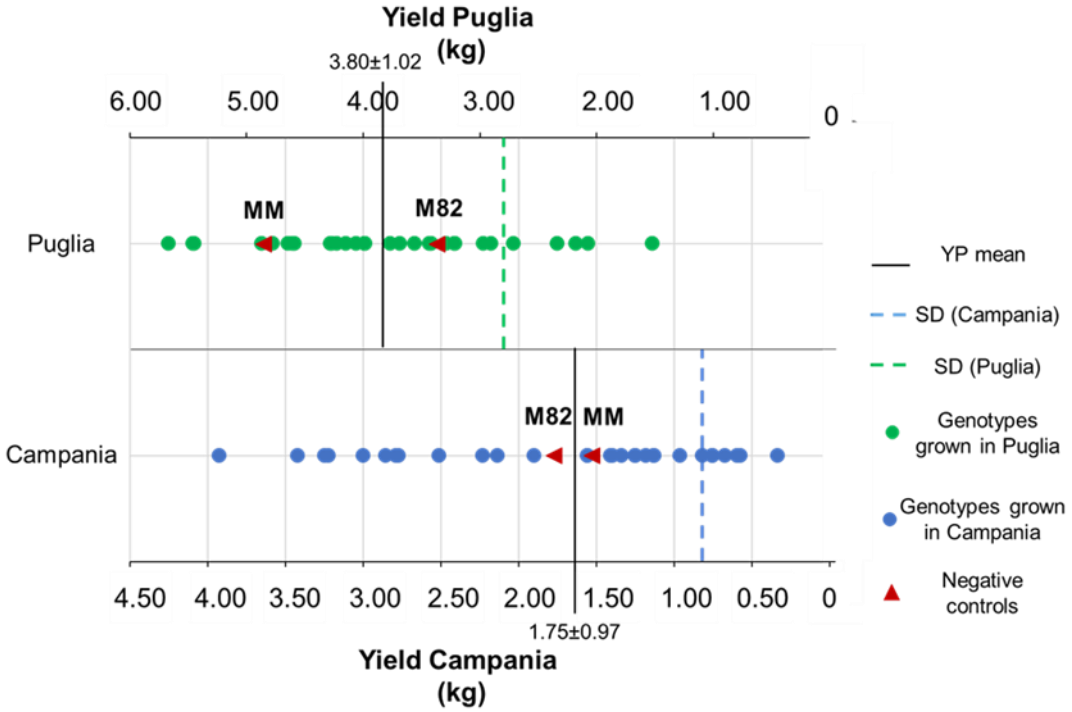
Supplementary Table S1 - List of genetical materials used in the experimental trials. Sources are represented by donor institutions (ARCA2010, Italian Research Society; CRA: Research Center of Agriculture; NPGS: Plant Genetic Resources Unit; Campania Region;TGRC: Tomato Genetics Resource Center; UNINA: University of Naples).

Genotype code	Origin	Source	Common Name	Growth habit	Fruit size
E5	Italy	Campania Region	Cento Scocche	Indeterminate	Small
E7	Italy	CRA	Corbarino PC04	Indeterminate	Small
E8	Italy	CRA	Corbarino PC05	Indeterminate	Small
E11	Italy	ARCA2010	Fabrizio	Determinate	Medium
E17	Italy	CRA	Pantano Romanesco	Indeterminate	Big
E20	Italy	CRA	Pizzutello	Indeterminate	Small
E23	Italy	Campania Region	San Marzano 1-38 SMEC	Indeterminate	Small
E29	Italy	CRA	Sel PC03	Indeterminate	Small
E30	Italy	CRA	Sel PC07	Indeterminate	Small
E34	Italy	CRA	Sorrento PS04	Indeterminate	Big
E36	Italy	ARCA2010	Siccagno 36	Indeterminate	Small
E39	Italy	UNINA	Casarbore	Indeterminate	Small
E40	Italy	UNINA	GiaGiù	Indeterminate	Small
E42	Italy	UNINA	Pl15250	Determinate	Small
E43	Italy	ARCA2010	Principe Borghese	Indeterminate	Small
E45	Italy	ARCA2010	San Marzano 246	Determinate	Medium
E48	Italy	UNINA	Vesuvio 2001	Indeterminate	Small
E53 (LA0147)	Latin America	TGRC		Indeterminate	Medium
E55 (LA0358)	Latin America	TGRC		Indeterminate	Small
E57 (LA0404)	Latin America	TGRC		Indeterminate	Small
E76 (LA4449)	URSS	TGRC	Black Plum	Indeterminate	Small
E87	Bolivia	NPGS	28	Indeterminate	Small
E94	Guatemala	NPGS	1404	Indeterminate	Small
E107	Spain	NPGS	E-L-19	Semi-determinate	Medium
E201	Italy	ARCA2010	Siccagno 201	Determinate	Medium
LA2662	USA	TGRC	Saladette	Determinate	Medium
LA3120	USA	TGRC	Malintka 101	Determinate	Medium
PDVIT	Italy	ARCA2010	Cannellino Vitiello	Indeterminate	Small
MONEY-MAKER (LA2706)	Great Britain	TGRC	Moneymaker	Indeterminate	Medium
M82 (LA3475)	USA	TGRC	M82	Determinate	Medium
DOCET	USA	MONSANTO	Docet	Determinate	Medium
JAG8810	USA	MONSANTO	JAG8810	Determinate	Medium

Supplementary Table S2 Phenotypic variation of 5 yield-related traits in Campania and Puglia during the year 2017. NFL, no. flower *per* inflorescence, FS, fruit set, TNF, no. fruit *per* plant, FW, fruit weight, YP, yield *per* plant.

Genotype	Flowers/inflorescence				Fruit set				Fruit/plant				Fruit weight				Yield/plant			
	no.				%				no.				g				kg			
	Campania	SE	Puglia	SE	Campania	SE	Puglia	SE	Campania	SE	Puglia	SE	Campania	SE	Puglia	SE	Campania	SE	Puglia	SE
E5	45.70	1.55	54.53	1.91	21.27	2.51	25.51	16.46	35.73	14.27	215.95	0.35	31.72	2.40	38.18	10.50	1.14	0.50	4.60	0.54
E7	7.05	0.20	7.94	0.56	71.90	3.50	61.21	5.94	35.84	16.06	161.53	13.04	29.51	1.38	24.68	2.68	1.42	0.77	3.99	0.57
E8	7.67	0.33	7.03	0.27	70.61	5.06	60.60	1.80	30.30	6.73	200.35	14.35	23.49	0.09	21.63	0.91	0.82	0.19	4.15	0.31
E11	5.36	0.58	5.32	0.12	45.36	5.90	48.26	5.84	37.13	0.89	164.85	15.95	77.37	10.12	36.83	6.73	2.86	0.32	3.22	0.19
E17	7.91	1.55	6.86	0.90	31.52	2.93	23.17	2.62	3.05	2.25	24.55	1.85	116.28	6.85	121.19	0.48	0.34	0.24	2.72	0.29
E20	7.60	0.43	8.00	0.25	52.33	2.14	43.47	6.00	57.60	9.50	100.58	16.81	29.35	3.83	29.57	2.49	2.24	0.46	5.46	0.77
E23	6.88	0.63	8.11	0.39	72.18	5.59	55.30	5.94	41.40	6.79	188.93	22.11	14.49	7.27	22.68	1.54	0.76	0.06	4.28	0.51
E29	7.38	0.09	7.94	0.20	52.70	7.47	35.65	3.52	46.03	8.44	32.77	10.58	28.44	3.63	26.99	2.29	1.26	0.12	3.56	0.58
E30	7.33	0.14	18.06	2.58	68.31	11.69	10.69	4.18	71.30	7.34	136.52	32.91	26.79	0.45	16.57	1.68	1.91	0.18	2.08	0.19
E34	10.71	0.43	8.86	0.85	22.14	2.01	12.36	4.23	12.53	6.47	123.77	31.54	90.60	6.87	76.60	2.76	1.19	0.68	3.22	0.88
E36	7.69	0.25	7.22	0.29	69.65	3.99	55.72	5.48	87.31	10.93	254.27	39.78	26.07	0.84	22.82	2.24	2.52	0.12	5.67	0.58
E39	8.06	0.39	6.93	0.21	62.17	1.25	52.66	5.43	66.00	9.00	140.30	25.14	25.72	2.28	27.64	2.98	1.13	0.45	4.07	0.75
E40	8.30	0.45	7.25	0.42	59.85	1.80	31.71	4.88	137.30	19.10	76.30	12.47	20.87	0.52	21.26	2.69	2.14	0.68	2.91	0.38
E42	10.72	1.21	10.92	0.72	60.45	4.53	25.84	0.29	304.18	8.07	417.25	47.32	11.27	0.07	10.20	0.67	3.43	0.07	4.61	0.39
E43	6.70	0.41	7.44	0.83	48.29	2.30	26.32	4.56	84.05	3.95	106.07	22.52	21.29	1.21	24.87	1.44	1.40	0.34	3.30	0.48
E45	6.22	0.12	5.83	0.58	56.35	5.84	54.43	2.71	62.27	11.92	76.00	23.24	40.95	13.69	36.70	5.75	2.79	0.44	1.53	0.16
E48	8.19	0.19	7.31	0.12	52.34	3.37	42.21	7.36	61.67	5.22	170.60	4.42	25.45	1.42	21.61	0.97	1.56	0.12	3.69	0.24
E53	7.14	0.12	11.11	2.10	50.97	2.37	36.45	7.86	24.70	5.31	80.37	2.65	56.68	1.43	42.50	2.57	1.56	0.46	3.42	0.25
E55	6.57	0.20	6.56	0.15	53.80	9.36	49.13	5.35	45.76	4.54	181.43	19.57	16.63	0.94	19.51	2.65	0.76	0.10	3.44	0.09
E57	4.95	0.16	4.23	0.20	24.79	10.50	24.27	6.09	6.50	1.10	54.37	23.99	91.94	16.94	68.12	15.89	0.39	0.19	4.78	0.23
E76	7.17	0.22	6.86	0.82	71.24	3.82	63.98	6.10	41.33	13.42	100.47	19.37	24.92	2.07	21.79	1.66	1.25	0.38	2.18	0.41
E87	10.06	1.39	8.84	0.56	39.72	1.19	32.10	2.29	38.81	5.93	130.60	44.12	24.98	1.27	25.99	2.80	0.97	0.14	4.00	0.33
E94	4.52	0.16	3.33	0.00	44.02	5.00	49.58	2.92	26.40	1.90	46.73	17.52	26.21	7.96	27.21	0.81	0.45	0.24	2.34	0.38
E107	4.82	0.36	4.83	0.36	74.23	10.77	68.50	8.30	24.50	8.23	143.92	6.48	47.75	8.20	37.71	4.32	1.35	0.09	5.45	0.87
E201	5.26	0.01	4.56	0.47	69.00	4.77	36.49	9.95	68.44	6.44	69.06	10.90	50.27	6.43	36.90	3.36	3.01	0.92	4.23	1.05
PDVIT	11.56	0.31	12.53	0.57	77.15	15.15	57.67	7.40	94.50	1.44	227.23	22.94	9.90	1.75	12.99	1.40	0.60	0.15	2.98	0.51
LA2662	6.50	0.21	6.25	0.21	56.06	7.94	54.22	2.58	55.46	3.04	112.17	19.99	49.65	4.27	49.29	3.62	2.78	0.38	4.28	0.47
LA3120	6.69	0.03	7.61	0.79	63.47	1.93	43.50	5.78	179.63	29.75	197.12	32.86	24.35	3.54	25.84	3.81	3.25	0.57	3.77	0.64
M82	5.92	0.33	6.81	0.46	46.08	6.09	21.29	2.82	56.94	13.34	104.48	11.23	42.04	7.00	33.24	5.18	1.77	0.35	3.36	0.17
MM	7.25	0.14	8.75	0.90	49.97	4.96	27.62	5.43	26.00	4.41	107.27	5.98	56.47	7.51	45.06	3.84	1.53	0.40	4.85	0.60
DOCET	6.67	0.17	6.28	0.12	53.07	6.08	54.14	7.39	57.40	6.21	159.96	24.57	60.13	3.01	31.41	4.09	3.93	0.58	4.88	0.57
JAG8810	6.80	0.27	8.07	0.49	48.08	1.67	38.30	13.27	44.90	2.78	116.93	16.97	67.80	11.01	41.41	6.33	3.23	0.33	4.65	0.38

Supplementary Figure S3 Dot plot of the YP values recorded in Campania (blue dots) and Puglia (green dots) recorded in 2017. The dashed lines represent the lower range of standard deviation for each field. YP, yield *per* plant; SD, standard deviation.



The yield *per* plant recorded in both fields resulted similar to the mean of the total production in both fields. Indeed, in Campania M82 and MM showed a YP (1.77 kg/plant and 1.53 kg/plant, respectively), similar to the mean of 1.75 kg/plant. Notwithstanding MM showed a YP lower than the total mean production, its YP value falls in the range between the mean and the standard deviation. In the same way, similar results were recorded in Puglia. In*deed*, MM and M82 showed, respectively, values higher (4.85 kg/plant) and lower (3.36 kg/plant) than the mean of the production (3.74 kg/plant).In this case the YP value of M82 falls within the range between the total mean production and the standard deviation. For these reasons the two negative controls were evaluated as not heat susceptible genotype and were not cultivated in the second year.

Supplementary Table S4. Phenotypic variation recorded in Campania 2018 of three yield-related traits. TNF, no. fruit *per* plant, FW, fruit weight, YP, yield *per* plant

Genotype	Campania 2018					
	TNF		FW		YP	
	MEAN	SE	MEAN	SE	MEAN	SE
E5	22.83	1.27	16.28	1.61	0.37	0.04
E7	9.53	2.28	14.04	0.74	0.13	0.04
E11	5.75	0.25	27.94	0.36	0.16	0.00
E20	22.33	6.77	17.42	0.98	0.39	0.11
E23	10.87	0.96	9.66	1.04	0.11	0.02
E36	19.80	3.16	18.09	1.16	0.36	0.08
E40	14.06	0.95	12.39	0.79	0.17	0.00
E42	230.41	10.40	10.91	0.06	2.63	0.01
E45	1.62	0.74	19.85	3.31	0.05	0.02
E48	14.65	6.10	15.35	1.21	0.32	0.12
E55	42.28	3.72	11.85	0.55	0.51	0.01
E87	13.38	1.44	14.91	0.73	0.20	0.02
E107	4.71	0.29	18.00	2.68	0.09	0.02
E201	10.03	2.13	26.80	1.24	0.27	0.07
PDVIT	49.38	8.09	10.68	0.21	0.53	0.09
DOCET	68.43	4.42	64.29	2.72	4.38	0.09
JAG8810	13.41	2.68	25.66	1.31	0.34	0.06
LA2662	18.19	2.23	27.14	3.27	0.48	0.02
LA3120	25.00	1.00	17.73	2.45	0.67	0.19

Supplementary Table S5 Markers associated to six evaluated phenotypic traits specific of each location (Campania and Puglia). For each marker an arbitrary code (SP), the associated trait, the location, SolCAP marker identity, the no. of chromosome, the marker position in SL3.0 genome version, the co-localized QTL and/or Candidate Genes (CGs) and the predicted function are reported.

Code	Trait	Location	SolCAP marker	Chr.	Marker Position	Associated QTL	Associated CGs	CG/QTL distance from marker	CG predicted function
SP1	FRL	Campania	solcap_snp_sl_20420	1	6,798,476		*Solyc01g010820*	635 kbp	Cullin family protein
SP2		Campania	solcap_snp_sl_19066	1	42,998,981		*Solyc01g044270*	0 kbp	DnaJ domain-containing protein
SP3		Campania	solcap_snp_sl_16519	1	53,634,350		*Solyc01g056340*	48.3 kbp	Light-mediated development protein DET1
SP4		Campania	solcap_snp_sl_26417	1	69,396,388		*Solyc01g059870*	501 kbp	phytochrome B1
							Solyc01g060130	219 kbp	Arf-like GTPase family
SP5		Puglia	solcap_snp_sl_9755	1	85,016,512		*Solyc01g093960*	396 kbp	MADS box transcription factor
							Solyc01g093980	412 kbp	Gibberellin 20-oxidase 4
SP6		Puglia	solcap_snp_sl_9756	1	85,136,490				
SP7		Puglia	solcap_snp_sl_34568	1	85,019,392		*Solyc01g093960*	396 kbp	MADS box transcription factor
							Solyc01g093980	412 kbp	Gibberellin 20-oxidase 4
SP8		Puglia	solcap_snp_sl_9385	3	64,278,295		*Solyc03g112410*	21.4 kbp	Hystone deacetylase 6
SP9	FRL/NFL	Campania	solcap_snp_sl_9382	3	64,461,564		*Solyc03g112410*	162 kbp	Hystone deacetylase 6
							Solyc03g113130	425 kbp	2-aminoethanethiol dioxygenase
							Solyc03g113930	1.0 Mbp	Heat-shock protein
							Solyc03g114470	1.5 Mbp	Cell growth defect factor-like
							Solyc03g114820	1.7 Mbp	Adenine nucleotide alpha hydrolases-like
							Solyc03g114830	1.7 Mbp	FRUITFULL-like MADS-box 2
							Solyc03g115140	1.9 Mbp	DnaJ domain-containing protein
							Solyc03g115230	2.0 Mbp	Heat-shock protein
SP10	NFL	Campania	solcap_snp_sl_29313	4	65,697,531				
SP11		Campania	solcap_snp_sl_24679	10	61,164,387		*Solyc10g081250*	1.3 Mbp	DNA polymerase
SP12	FS	Campania	solcap_snp_sl_19783	8	887,485		*Solyc08g066170*	1.5 Mbp	Early nodulin-93
SP13	TNF	Campania	solcap_snp_sl_12248	5	64,057,988		*Solyc05g053090*	6.6 kbp	Transcription factor GRAS
SP14		Puglia	solcap_snp_sl_25253	6	1,453,932				
SP15		Puglia	solcap_snp_sl_34942	8	64,754,343				
SP16	FW	Puglia	solcap_snp_sl_18306	5	2,136,081	*lcn5.1*	*Solyc05g007130*	441 kbp	SlSUN15
SP17	YP	Puglia	solcap_snp_sl_34105	3	68,783,132	*fw3.2*	*Solyc03g114940*	2.5 Mbp	P540 cytochrome
SP18		Puglia	solcap_snp_sl_25859	5	63,590,154	*fw3.3*			

Supplementary Figure S6 Private SNPs/InDels recorded in 19 assayed genotypes. SNPs and inDels grouped for chromosome are also reported.

Genotype	SL3.0ch01		SL3.0ch02		SL3.0ch03		SL3.0ch04		SL3.0ch05		SL3.0ch06		SL3.0ch07		SL3.0ch08		SL3.0ch09		SL3.0ch10		SL3.0ch11		SL3.0ch12		Total private SNPs
	SNPs	InDel	SNPs	InDel	SNPs	InDel	SNPs	InDel	SNPs	InDel	SNPs	InDel	SNPs	InDel	SNPs	InDel	SNPs	InDel	SNPs	InDel	SNPs	InDel	SNPs	InDel	
E7	15	2	37	9	4	1	8	1	18	1	30	5	14	2	12	1	12	2	21	1	46	3	5	0	250
E8	13	1	5	1	3	0	0	1	8	1	28	7	2	0	9	1	8	0	18	2	9	0	1	1	119
E17	8	2	8	1	61	10	13	0	146	18	25	2	3	0	14	1	9	1	2	4	9	2	25	2	366
E20	17	0	4	0	3	0	2	1	5	4	3	1	5	2	12	4	4	1	12	1	2	1	6	0	90
E23	21	4	15	2	9	4	10	2	5	2	17	2	5	1	16	3	8	4	21	2	11	3	15	3	185
E30	6	0	10	0	9	0	4	0	6	1	3	3	10	2	22	7	7	0	18	3	3	0	8	0	122
E36	3	2	2	2	6	0	2	0	9	0	18	1	4	0	9	3	2	2	4	0	2	4	2	1	78
E40	9	0	6	0	10	3	5	3	10	0	15	5	7	0	7	1	2	1	3	0	8	3	12	0	110
E42	3501	416	55	15	39	8	354	86	48	7	33	5	3535	497	31	4	70	11	108	14	94	25	150	40	9146
E43	12	2	5	0	15	4	11	1	3	0	11	6	4	0	25	8	3	2	56	8	11	1	10	5	203
E45	72	14	10	2	20	3	24	5	3	1	13	3	8	2	6	6	26	2	15	0	13	3	8	3	262
E48	8	2	11	1	12	2	1	0	3	0	5	3	3	0	16	5	2	1	3	1	4	2	8	4	97
E53	20	3	2	2	27	8	3	0	2	1	18	3	32	7	11	3	1	0	1	2	9	2	9	0	166
E55	24	11	46	10	86	29	120	23	17	0	606	70	117	26	72	12	865	77	164	30	38	8	40	11	2502
E76	8	1	2	0	5	2	8	0	1	0	13	1	6	0	9	0	5	0	6	0	3	0	39	9	118
E107	10	2	10	4	13	3	17	0	10	2	47	6	25	1	14	7	2	1	745	48	14	6	10	2	999
LA2662	58	12	20	1	26	5	17	0	14	4	8	0	18	4	5	4	2	0	9	0	39	5	11	0	262
LA3120	6	1	7	2	24	8	196	33	21	5	19	4	16	3	10	1	5	0	16	1	6	3	3	2	392
PDVIT	206	51	287	59	58	19	169	43	3440	526	199	53	18	5	78	9	1794	466	63	9	1541	182	127	19	9421
DOCET	9	2	14	3	3	2	15	4	11	0	6	6	21	1	5	2	6	1	2	0	5	4	12	1	135
JAG8810	11	3	4	1	7	5	23	14	11	4	34	15	2	0	9	3	0	0	6	1	13	4	7	5	182

Supplementary Table S7. Blast results of the 82 assayed sequences in different wild species databases. The identity score, the coverage and Subject sequences are also reported.

Query	*S. pimpinellifolium*				*S. galapagense*				*S. chilense*				*S. pennellii*				*S. peruvianum*			
	Variant	Identity	Coverage	Subject	Variant	Identity	Coverage	Subject	Variant	Identity	Coverage	Subject	Variant	Identity	Coverage	Subject	Variant	Identity	Coverage	Subject
M1	Alternative	98,73	95,18	contig 2566819	Reference	100,00	100,00	scaffold119823__7.9												
M2	Alternative	100,00	100,00	contig 5031318																
M3					Reference	100,00	100,00	scaffold119433_13.4	Reference	96,04	100,00	scaffold965								
M4	Reference	100,00	100,00	contig 6617089	Reference	100,00	100,00	scaffold4822_14.2_contig7	Alternative	99,07	100,00	scaffold965	Alternative	99,07	100,00	Spenn-ch01:2733619..2733726				
M5																				
M6	Alternative	100,00	79,65	contig 6488329	Alternative	100,00	100,00	scaffold42564_12.2_contig3												
M7	Alternative	100,00	85,05	contig 6607356	Alternative	99,07	100,00	scaffold94470_13.6_contig1												
M8																				
M9	Alternative	100,00	100,00	contig 6652606	Reference	100,00	100,00	scaffold2817_11.9_contig2												
M10					Reference	99,01	100,00	C9507465_42.0												
M11					Reference	97,87	93,07	scaffold57955__5.1_contig2												
M12	Alternative	100,00	91,30	contig 3876605																
M13					Reference	100,00	100,00	C9302529__5.0												
M14	Alternative	100,00	100,00	contig 6453285	Alternative	100,00	100,00	scaffold99124_10.1	Wild	100,00	100,00	scaffold4568	Alternative	99,03	97,17	Spenn-ch01:91995761..91995863				
M15	Alternative	100,00	100,00	contig 6559546																
M16	Alternative	100,00	100,00	contig 6608601	Reference	100,00	100,00	scaffold83404_11.4_contig1	Alternative	97,87	86,24	scaffold866	Alternative	99,08	100,00	Spenn-ch02:47532230..47532338				
M17	Alternative	100,00	100,00	contig 6709952	Reference	100,00	100,00	scaffold83438_12.2_contig4	Reference	99,01	100,00	scaffold2067								
M18	Alternative	100,00	100,00	contig 6605169	Alternative	100,00	100,00	scaffold100837_10.9_contig4												
M19	Alternative	100,00	100,00	contig 6668767	Alternative	100,00	100,00	scaffold93365_11.3_contig8									Alternative	95,37	97,22	a6818
M20					Reference	98,72	77,23	scaffold112549__9.1_contig1												
M21	Alternative	100,00	100,00	contig 6484118	Reference	100,00	100,00	scaffold48433_11.7_contig4					Alternative	98,20	97,35	Spenn-ch02:57996230..57996340				
M22	Alternative	98,92	98,92	contig 6595391	Reference	100,00	100,00	C9710229__8.0												
M23																				
M24	Alternative	100,00	100,00	contig 1006140	Alternative	100,00	100,00	scaffold124091_14.5					Reference	99,01	100,00	Spenn-ch02:59005059..59005159				
M25	Alternative	100,00	100,00	contig 6728065	Reference	100,00	100,00	scaffold86649_13.8_contig3					Reference	99,01	100,00	Spenn-ch02:59668356..59668456				
M26	Alternative	100,00	100,00	contig 6545172																
M27																				
M28					Alternative	100,00	100,00	C9422939__7.0					Alternative	98,28	99,15	Spenn-ch04:35880823..35880938				
M29	Alternative	100,00	100,00	contig 6462044					Alternative	97,27	100,00	scaffold5451	Alternative	97,27	100,00	Spenn-ch04:63238714..63238823				
M30																				
M31	Reference	100,00	100,00	contig 6449752	Reference	99,01	100,00	scaffold54594_14.3_contig2	Reference	100,00	100,00	scaffold16921	Reference	97,70	86,14	Spenn-ch04:67745822..67745908	Reference	97,98	98,02	a8434
M32	Alternative	100,00	100,00	contig 6432537	Alternative	100,00	100,00	scaffold72808_13.0												
M33					Reference	100,00	100,00	C9375253__5.0												
M34					Reference	100,00	100,00	scaffold48422__9.4_contig2	Reference	97,03	100,00	scaffold42606; scaffold2361	Reference	96,04	100,00	Spenn-ch05:32563026..32563126				
M35	Alternative	100,00	100,00	contig 6460391	Reference	100,00	98,02	C9751563_53.0												
M36					Reference	100,00	100,00	scaffold23333_17.1												
M37	Alternative	100,00	63,55	contig 6537387	Reference	100,00	100,00	scaffold109989_14.2_contig1												
M38					Alternative	100,00	100,00	scaffold51628_12.9_contig3												
M39																				
M40					Reference	100,00	100,00	scaffold82240_11.4_contig3												

Supplementary Table S7. Continued

Query	*S. pimpinellifolium*				*S. galapagense*				*S. chilense*				*S. pennellii*				*S. peruvianum*			
	Variant	Identity	Coverage	Subject	Variant	Identity	Coverage	Subject	Variant	Identity	Coverage	Subject	Variant	Identity	Coverage	Subject	Variant	Identity	Coverage	Subject
M41	Alternative	96.40	98.20	contig 2583451	Alternative	100.00	84.68	scaffold110710__5.1_contig1												
M42					Alternative	100.00	100.00	scaffold26439_12.4_contig2												
M43					Reference	100.00	100.00	scaffold64786_11.5_contig1												
M44					Alternative	100.00	100.00	scaffold1541_15.3_contig2					Alternative	96.43	100.00	Spenn-ch07:21188758..21188869				
M45					Alternative	100.00	100.00	C9741765_11.0												
M46					Alternative	98.15	100.00	scaffold18002_14.3_contig4												
M47	Alternative	100.00	100.00	contig 711842	Alternative	97.20	97.20	scaffold85567_12.5_contig5	Wild	96.26	99.07	scaffold2917	Alternative	98.13	100.00	Spenn-ch07:3385506..3385612				
M48	Alternative	98.17	100.00	contig 6707462	Reference	99.01	100.00	scaffold26422_13.6_contig3												
M49									Reference	95.10	100.00	scaffold1012	Reference	97.03	99.01	Spenn-ch07:44718481..44718580				
M50					Reference	97.03	100.00	scaffold22694_13.5_contig4	Reference	98.02	100.00	scaffold1258	Reference	97.03	100.00	Spenn-ch07:16858785..16858885				
M51					Reference	100.00	100.00	scaffold39889_14.8_contig5												
M52	Reference	98.02	100.00	contig 6711923	Reference	99.01	100.00	scaffold35153_10.2_contig3					Reference	91.09	95.05	Spenn-ch07:61628129..61628224				
M53	Alternative	100.00	100.00	contig 442915	Reference	100.00	100.00	C9750657_15.0	Alternative	100.00	100.00	scaffold4865	Alternative	100.00	100.00	Spenn-ch07:73303880..73303992				
M54	Alternative	100.00	71.17	contig 5929686																
M55					Reference	100.00	100.00	scaffold28440__9.1												
M56					Reference	100.00	100.00	scaffold90596_15.0	Reference	97.98	98.02	scaffold3215	Reference	98.02	100.00	Spenn-ch07:8526890..8526990				
M57	Reference	100.00	83.17	contig 6618754	Reference	100.00	100.00	scaffold66210_12.1_contig1												
M58	Reference	100.00	100.00	contig 6739192	Reference	100.00	100.00	scaffold86153_14.0_contig4	Alternative	99.08	100.00	scaffold3926	Alternative	97.25	98.17	Spenn-ch09:16833489..16833595				
M59					Reference	100.00	100.00	scaffold6505_13.4_contig2					Alternative	92.38	95.28	Spenn-ch09:19158778..19158881				
M60	Alternative	90.79	78.65	contig 4305081																
M61	Reference	100.00	100.00	contig 2401624	Reference	100.00	100.00	C9835341_12.0												
M62	Reference	100.00	100.00	contig 6575130	Reference	97.98	98.02	scaffold22324_13.8_contig2	Alternative	100.00	100.00	scaffold422								
M63																				
M64	Alternative	100.00	100.00	contig 6568390	Alternative	100.00	100.00	scaffold117545_13.1					Alternative	99.11	100.00	Spenn-ch09:3552680..3552791				
M65	Reference	96.12	100.00	contig 6486609	Reference	99.01	100.00	scaffold7696_13.1_contig7	Reference	95.10	100.00	scaffold1243	Reference	99.01	100.00	Spenn-ch09:4215228..4215328				
M66	Reference	100.00	100.00	contig 6443352	Reference	100.00	100.00	scaffold91671_11.3												
M67					Reference	100.00	100.00	scaffold7220_13.7_contig2												
M68	Reference	100.00	100.00	contig 6555491	Reference	100.00	100.00	scaffold6086_13.3_contig1	Reference	98.02	100.00	scaffold5430								
M69	Reference	100.00	100.00	contig 6614832	Reference	99.01	100.00	scaffold45330_14.8_contig1	Reference	98.02	100.00	scaffold8775	Reference	98.02	100.00	Spenn-ch09:69104745..69104845				
M70																				
M71	Alternative	100.00	100.00	contig 6489748	Alternative	100.00	100.00	scaffold50024_12.8_contig6												
M72	Alternative	98.95	84.82	contig 6620295	Reference	98.02	100.00	C9524755__6.0												
M73																				
M74	Alternative	100.00	100.00	contig 6468439	Reference	99.01	100.00	scaffold31357_12.7_contig9					Reference	99.01	100.00	Spenn-ch05:39471484..39471584				
M75					Reference	100.00	100.00	scaffold2586_15.3_contig2												
M76	Alternative	100.00	100.00	contig 6712699	Reference	100.00	100.00	scaffold114604_14.5_contig2												
M77																				
M78					Alternative	100.00	100.00	scaffold47993_15.6_contig1	Alternative	95.24	98.13	scaffold11367	Alternative	98.13	100.00	Spenn-ch12:68324308..68324414				
M79					Alternative	100.00	100.00	scaffold47841_16.4_contig2												
M80																				
M81	Reference	100.00	100.00	contig 6606810	Reference	100.00	100.00	C9557009_11.0	Reference	99.02	100.00	scaffold1880								
M82	Alternative	96.43	99.11	contig 6689634	Reference	98.02	100.00	scaffold22968_12.6_contig2												

Supplementary Table S8. Yield-related traits recorded in F1 hybrids and their parents during 2018. TNF, no fruit *per* plant, FW, Fruit Weight, YP, Yield *per* plant.

Cross	Genotype	TNF		FW		YP	
		MEAN	SE	MEAN	SE	MEAN	SE
F1 hybrid	**17H10**	33.75	1.25	34.00	6.69	1.20	0.23
♀	**E37**	23.88	6.91	34.32	1.21	0.81	0.22
♂	**E33**	1.29	0.52	118.36	13.11	0.15	0.05
F1 hybrid	**17H11**	24.60	0.95	47.76	3.58	1.29	0.09
♀	**E37**	23.88	6.91	34.32	1.21	0.81	0.22
♂	**STIZ**	14.80	0.66	108.33	7.36	1.82	0.08
F1 hybrid	**17H12**	14.17	1.17	77.14	10.08	1.35	0.27
♀	**STIZ**	14.80	0.66	108.33	7.36	1.82	0.08
♂	**E33**	1.29	0.52	118.36	13.11	0.15	0.05
F1 hybrid	**17H13**	23.13	1.57	39.81	2.44	1.05	0.11
♀	**STIZ**	14.80	0.66	108.33	7.36	1.82	0.08
♂	**E37**	23.88	6.91	34.32	1.21	0.81	0.22
F1 hybrid	**17H14**	52.73	3.95	10.04	0.78	0.53	0.04
♀	**PDVIT**	101.24	16.29	5.35	0.11	0.53	0.09
♂	**E7**	13.59	4.27	14.04	0.74	0.20	0.07
F1 hybrid	**17H16**	41.29	5.23	7.89	0.35	0.33	0.04
♀	**PDVIT**	101.24	16.29	5.35	0.11	0.53	0.09
♂	**E64**	7.81	4.39	20.91	5.22	0.13	0.06
F1 hybrid	**17H25**	73.04	2.24	11.06	0.51	0.81	0.06
♀	**E111**	14.27	6.46	12.18	1.67	0.17	0.09
♂	**E42**	230.41	10.40	10.91	0.06	2.51	0.11
F1 hybrid	**17H32**	40.09	3.91	13.81	0.24	0.56	0.06
♀	**E42**	230.41	10.40	10.91	0.06	2.51	0.11
♂	**E20**	22.33	6.77	17.42	0.98	0.39	0.11
F1 hybrid	**17H36**	126.96	30.74	13.08	0.06	1.66	0.40
♀	**E103**	13.57	8.17	10.93	1.25	0.13	0.07
♂	**E42**	230.41	10.40	10.91	0.06	2.51	0.11
F1 hybrid	**17H37**	74.98	14.01	14.84	1.37	1.08	0.13
♀	**E103**	13.57	8.17	10.93	1.25	0.13	0.07
♂	**E48**	14.65	6.10	15.35	1.21	0.23	0.11
F1 hybrid	**17H39**	25.89	6.07	19.80	1.19	0.50	0.10
♀	**E103**	13.57	8.17	10.93	1.25	0.13	0.07
♂	**PDLUC**	26.66	3.83	25.38	1.67	0.67	0.06

Supplementary Table S9. Yield-related traits recorded in heat-tolerant controls in 2018 and 2019. TNF, number fruit *per* plant, FW, Fruit weight, YP, Yield *per* plant.

Control	2018						2019					
	TNF	Error	FW	Error	YP	Error	TNF	Error	FW	Error	YP	Error
DOCET	68.43	4.42	64.29	2.72	4.38	0.09	44.82	6.49	59.73	2.85	2.65	0.29
JAG8810	13.41	2.68	25.66	1.31	0.40	0.01	37.32	3.66	63.73	3.48	2.35	0.11
PAIPAI	27.82	3.40	89.90	1.14	2.50	0.29	80.69	17.55	77.03	3.85	6.24	1.69

Supplementary Table S10. Yield-related traits recorded in F1 hybrids and their parents during 2019. TNF, no. fruit *per* plant, FW, Fruit Weight, YP, Yield *per* plant

Cross	Genotype	TNF		FW		YP	
		MEAN	SE	MEAN	SE	MEAN	SE
F1 hybrid	**17H14**	262.65	26.97	12.89	0.41	3.22	0.29
♀	**PDVIT**	381.49	13.84	8.87	0.21	2.49	0.51
♂	**E7**	151.09	10.93	23.94	0.24	3.47	0.21
F1 hybrid	**17H36**	70.97	6.54	15.79	1.94	1.23	0.14
♀	**E103**	68.79	7.37	21.68	0.31	1.40	0.10
♂	**E42**	134.97	25.67	13.67	0.37	2.21	0.45
F1 hybrid	**17H37**	204.92	9.79	25.07	0.37	4.98	0.35
♀	**E103**	68.79	7.37	21.68	0.31	1.40	0.10
♂	**E48**	205.75	5.77	21.89	0.29	4.26	0.13
F1 hybrid	**17H39**	175.46	5.17	27.44	1.22	4.57	0.34
♀	**E103**	68.79	7.37	21.68	0.31	1.40	0.10
♂	**PDLUC**	84.94	8.44	21.53	0.54	1.83	0.16
F1 hybrid	**17H56**	159.77	12.57	26.67	1.25	4.15	0.17
♀	**E103**	68.79	7.37	21.68	0.31	1.40	0.10
♂	**E55**	119.35	9.88	21.89	0.44	2.37	0.25
F1 hybrid	**17H57**	290.08	33.83	19.27	0.18	4.57	0.37
♀	**E103**	68.79	7.37	21.68	0.31	1.40	0.10
♂	**E111**	176.71	21.27	10.84	0.27	1.73	0.14
F1 hybrid	**18H13**	126.16	9.72	20.57	0.57	2.55	0.19
♀	**E20**	150.83	7.54	24.38	0.22	3.55	0.31
♂	**E42**	134.97	25.67	13.67	0.37	2.21	0.45
F1 hybrid	**18H17**	290.92	32.46	19.89	0.22	5.39	0.46
♀	**E36**	189.58	23.65	24.22	0.40	4.28	0.58
♂	**E42**	134.97	25.67	13.67	0.37	2.21	0.45
F1 hybrid	**18H48**	30.10	5.40	64.47	4.87	1.92	0.32
♀	**E109**	25.30	5.26	58.00	2.23	1.49	0.37
♂	**E45**	29.62	5.80	56.40	1.15	1.66	0.30
F1 hybrid	**18H56**	54.78	3.67	64.07	1.45	3.45	0.31
♀	**E11**	16.71	5.01	70.40	3.06	1.18	0.35
♂	**LA2662**	44.04	1.82	42.67	2.27	1.87	0.05
F1 hybrid	**18H57**	53.63	0.00	34.40	0.00	1.85	0.00
♀	**E11**	16.71	5.01	70.40	3.06	1.18	0.35
♂	**LA3120**	95.95	3.46	29.73	1.76	2.84	0.10
F1 hybrid	**18H59**	70.56	3.47	31.73	0.67	2.24	0.12
♀	**E11**	16.71	5.01	70.40	3.06	1.18	0.35
♂	**E42**	134.97	25.67	13.67	0.37	2.21	0.45

Supplementary Table S11. Quality traits measured recorded in F1 hybrids and their parents during the 2019. .

Cross	Genotype	Titratable Acidity (g/100 g)		Total soluble solid content (Brix°)	
		MEAN	SE	MEAN	SE
F1 hybrid	**17H14**	0.53	0.02	6.98	0.17
♀	**PDVIT**	0.44	0.03	8.08	0.07
♂	**E7**	0.50	0.03	4.95	0.24
F1 hybrid	**17H25**	0.60	0.03	6.22	0.13
♀	**E111**	0.62	0.01	6.75	0.04
♂	**E42**	0.36	0.03	4.90	0.15
F1 hybrid	**17H36**	0.54	0.04	5.73	0.20
♀	**E103**	0.63	0.12	5.77	0.32
♂	**E42**	0.36	0.03	4.90	0.15
F1 hybrid	**17H37**	0.50	0.02	5.75	0.10
♀	**E103**	0.63	0.12	5.77	0.32
♂	**E48**	0.46	0.02	5.48	0.15
F1 hybrid	**17H39**	0.50	0.04	5.87	0.09
♀	**E103**	0.63	0.12	5.77	0.32
♂	**PDLUC**	0.51	0.00	6.10	0.28
F1 hybrid	**17H56**	0.56	0.11	5.78	0.22
♀	**E103**	0.63	0.12	5.77	0.32
♂	**E55**	0.53	0.09	5.17	0.30
F1 hybrid	**17H57**	0.60	0.07	6.53	0.14
♀	**E103**	0.63	0.12	5.77	0.32
♂	**E111**	0.62	0.01	6.75	0.04
F1 hybrid	**18H13**	0.51	0.01	5.25	0.05
♀	**E20**	0.54	0.02	5.02	0.09
♂	**E42**	0.36	0.03	4.90	0.15
F1 hybrid	**18H17**	0.52	0.01	5.78	0.09
♀	**E36**	0.47	0.01	5.53	0.14
♂	**E42**	0.36	0.03	4.90	0.15
F1 hybrid	**18H48**	0.31	0.02	5.60	0.47
♀	**E109**	0.35	0.03	4.87	0.15
♂	**E45**	0.26	0.02	4.75	0.62
F1 hybrid	**18H56**	0.40	0.04	3.87	0.15
♀	**E11**	0.33	0.01	3.97	0.09
♂	**LA2662**	0.47	0.08	3.70	0.10
F1 hybrid	**18H57**	0.36	0.00	3.90	0.05
♀	**E11**	0.33	0.01	3.97	0.09
♂	**LA3120**	0.50	0.08	3.63	0.26
F1	**18H59**	0.35	0.06	4.77	0.20
♀	**E11**	0.33	0.01	3.97	0.09
♂	**E42**	0.36	0.03	4.90	0.15