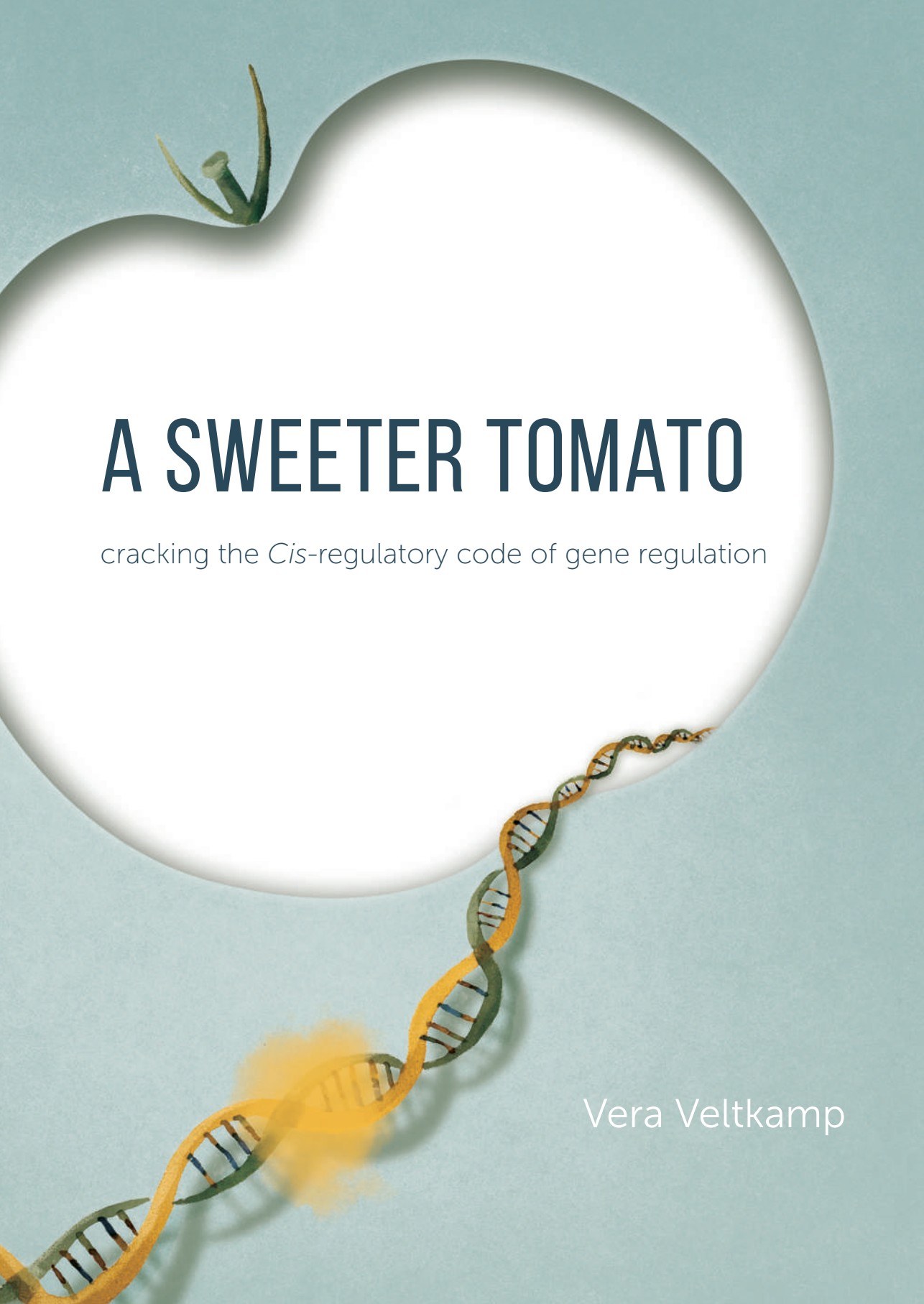


A SWEETER TOMATO

cracking the *Cis*-regulatory code of gene regulation

Vera Veltkamp

PROPOSITIONS

1. The modulation of promoter activity by mutagenesis can result in increased expression of the target gene.
   (this thesis)

2. Accurate prediction of Transcription Factor binding combined with conservation analysis will make promoter-targeting more efficient.
   (this thesis)

3. Publishing with a high societal impact has as much value as publishing with a high scientific impact.

4. The use of study-specific abbreviations needlessly complicates reading of a scientific manuscript.

5. New crop varieties made by targeted mutagenesis should be exempt from the risk assessment as defined in the Genetically Modified Organisms (GMO) Directive 2001/18/EC of the European Union (EU).

6. The uninspired Dutch cuisine has inspired many in the Netherlands to experience the entire world in their menus.

7. Flavorful vegetables will help us save the world.

Proposition belonging to the thesis, entitled:

A sweeter tomato, cracking the *Cis*-regulatory code of gene regulation

Vera Veltkamp

Wageningen, 17 September 2021

A SWEETER TOMATO

cracking the *Cis*-regulatory code of gene regulation

Vera Veltkamp

Thesis committee

Promotor
Prof. Dr Gerco C. Angenent
Personal chair at the Laboratory of Molecular Biology
Wageningen University & Research

Co-promotor
Dr Ruud A. de Maagd
Senior scientist at Bioscience
Wageningen University & Research

Other members
Prof. Dr Leonie Bentsink, Wageningen University & Research
Prof. Dr Sjef C. M. Smeekens, Utrecht University
Dr Walter Verweij, ENZA zaden Research & Development B.V., Enkhuizen
Dr Julian C. Verdonk, Wageningen University & Research

This research was conducted under the auspices of the Graduate School Experimental Plant Sciences

A SWEETER TOMATO

cracking the *Cis*-regulatory code of gene regulation

Vera Veltkamp

Thesis

Submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 17 September 2021
at 4 p.m. in the Aula

Vera Veltkamp

A sweeter tomato, cracking the *Cis*-regulatory code of gene regulation
234 pages

PhD thesis, Wageningen University, Wageningen, The Netherlands (2021)
With references, with summary in English and Dutch

ISBN: 978-94-6395-907-0
DOI: https://doi.org/10.18174/550506

CONTENTS

CHAPTER 1

General introduction

TOMATO

The tomato fruit, the subject of this thesis, is a popular vegetable all over the world, used in many dishes, salads, or as a snack (Oltman et al., 2014; Wang and Seymour, 2017). Total worldwide production in 2018 was 182,256,458 tons. Asia produced more than half of the world's tomatoes, with China being the number one producer by far (**Figure 1**). Europe takes the lead in productivity (measured as the production per area of land used), with the Netherlands in the number one position (**Table 1**). Interestingly, the top 10 countries for yield per area have a suboptimal climate for growing tomatoes. These countries mainly produce fresh-market tomatoes in highly efficient greenhouses, while the top-producing countries are growing tomatoes mainly in the field.

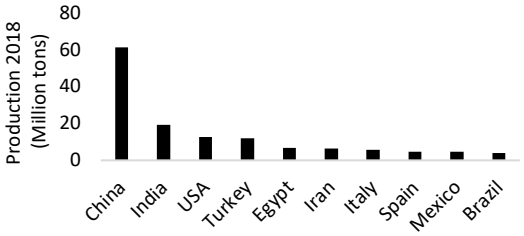


Figure 1: Production of the 10 leading producers in 2018 (source: FAO Statistics; http://www.fao.org/faostat/en/).

There has been some discussion on whether the origin of tomato lies in Mesoamerica or in the Andes (de Candolle, 1886; Jenkins, 1948). The word "*tomato*" does have its roots in Mexico, as the word is derived from the Aztec "*tomatl*", which means "the swelling fruit". Single Nucleotide Polymorphism (SNP)-analysis and population genomics show that pre-domestication began in South America (Peru, Ecuador) from the wild, blueberry-sized, red-fruited *Solanum pimpinellifolium* L. to give rise to the cherry size-fruited *Solanum lycopersicum* L. var. *cerasiforme*. Domestication was completed in Mesoamerica (Mexico) and gave rise to the current *Solanum lycopersicum* L. var. *lycopersicum* (Blanca et al., 2012; Lin et al., 2014; Razifard et al., 2020). Worldwide use and further breeding of tomato was initiated by the Spanish who brought tomato to Europe in the 16th century, from where it turned into one of the highest-value crops.

Solanum lycopersicum is a species in the large *Solanaceae* (nightshades) family, which includes several economically important crops such as potato, eggplant, pepper, and tobacco, as well as the ornamental petunia. The *Solanum* section *Lycopersicon* consists of the cultivated tomato, *S. lycopersicum* and twelve wild species: the green fruited *S. arcanum, S. huaylasense, S. peruvianum, S. corneliomulleri, S. chilense, S. chmielewskii, S. habrochaites, S. pennellii and S. neorickii,* and the orange or red-fruited *S. cheesmaniae, S. galapagense* and *S. pimpinellifolium* (Peralta et al., 2005, 2008;

Bergougnoux, 2014; Knapp and Peralta, 2016). They are easily recognized by their bright yellow flowers. All are diploid, with 2n = 24 chromosomes, and the cultivated tomato genome contains 950 Mb of haploid DNA (Kimura and Sinha, 2008), or 790 Mb according to the latest sequence assembly (Hosmani et al., 2019). The phenotypic variation of cultivated tomato is huge, especially when considering the numerous heirloom (pre-modern breeding varieties) and land-race tomato accessions (Roohanitaziani et al., 2020). However, the genetic variation within modern cultivated tomato is only 5% of the variation found in the wild relatives (Miller and Tanksley, 1990).

Table 1: Tomato yield in different countries and comparison with production and production rank. Data is from 2018 (source: FAO Statistics; http://www.fao.org/faostat/en/).

Country	Yield (hg/ha)	Rank yield	Production (million tons)	Rank production
Netherlands	5,089,485	1	0.910	23
Belgium	4,686,232	2	0.259	57
Sweden	4,557,500	3	0.018	114
Finland	3,892,574	4	0.039	101
Denmark	3,793,226	5	0.012	123
UK	3,635,489	6	0.067	86
Norway	3,377,573	7	0.013	121
Ireland	3,250,000	8	0.004	142
Iceland	3,032,500	9	0.001	151
Austria	2,956,482	10	0.058	89
USA	968,079	17	12.612	3
Spain	849,593	19	4.769	8
Brazil	719,404	30	4.110	10
Turkey	688,658	33	12.150	4
Italy	597,174	40	5.798	7
China	594,023	42	61.523	1
Mexico	504,785	52	4.559	9
Iran	413,679	59	6.577	6
Egypt	409,689	60	6.625	5
India	246,527	90	19.377	2

Early ripening, environmental adaptations, and fruit size, shape and colour were important traits for selection during domestication and improvement. For example, locule number were increased by the Quantitative Trait Loci (QTL) *fasciated (fas)* and *locule number (lc)* (Huang and van der Knaap, 2011; Xu et al., 2015; Somssich et al., 2016; Rodríguez-Leal et al., 2017). All modern cultivars carry the large-fruit allele of the gene underlying the *fresh weight2.2* (*fw2.2*) QTL, *ORFX,* that controls the timing and length of cell division during early fruit development (Alpert et al., 1995; Frary et al., 2000). The shape of tomato is among others controlled by the *SUN*-locus and *OVATE* (Xiao et al., 2008; van der Knaap et al., 2014). A trade-off of selection, mostly for larger fruits, was the loss of flavour (Bai and Lindhout, 2007; Klee and Tieman, 2018). The growth habit of the tomato plant of which there are two main types, indeterminate and determinate, is another important trait for cultivation. After the initial inflorescence, which is preceded by 8 to 12 leaves, the wild-type indeterminate tomato grows in sympodial units of three leaves, a terminating inflorescence, and a new axillary (sympodial) shoot below the inflorescence that grows into the next sympodial unit.

These indeterminate tomatoes are mostly grown in greenhouses and for the fresh tomato market. Determinate tomatoes are homozygous for a recessive allele of *SELF-PRUNING , sp* which causes the sympodial segments to develop progressively fewer nodes, until the shoot is terminated by two inflorescences (Yeager, 1927; Lilac Pnueli et al., 1998).The accompanying outgrowth of lateral shoots results in a bush-like plant where most fruits reach maturity at more or less the same time. These varieties are grown mostly in the field for the processing industry due to the possibility of harvesting an entire field mechanically and in one go.

Apart from being an important vegetable crop, tomato is also a widely used model crop in research (Gebhardt, 2016). Tomato is suitable as a model crop due to its ease of growth, propagation (by self-fertilization), transformation efficiency and simple genome (Bergougnoux, 2014). In 2012, the first complete tomato genome from the cultivar "Heinz 1706" was published (The Tomato Genome Consortium et al., 2012) and since then, the available genome data has been growing steadily. Important tomato resources are available at the *Solanaceae* Genomics Network (SGN) website (https://solgenomics.net/) and from the Tomato Genetic Research Consortium (https://tgrc.ucdavis.edu/). A series of F2 populations, introgression lines, recombinant inbred populations and (advanced) backcross populations/lines were made with several wild tomato relatives to facilitate breeding and identifying QTLs (Azanza et al., 1995; Eshed and Zamir, 1995; Tanksley et al., 1996; Causse et al., 2002; Fulton et al., 2002; Frary and Doğanlar, 2003). The sympodial growth, compound leaves and especially the fleshly fruits, make tomato a valuable alternative model-species besides other models like *Arabidopsis* and rice.

Following opening of the flower (anthesis) fruit development starts with (self) pollination and fertilization (fruit set) after which fertilized ovules develop into seeds and the growing ovary becomes the fruit (Varga, 1976). The fruit functions in the protection and when ripe, dispersal of seeds by attracting frugivores with fleshy textures, colour, aromas, and taste. The outer layer of a fruit, the pericarp derived from the ovary wall, is the main fleshy part (**Figure 2a**). The locular cavities are separated by one or more septa and contain the seeds and gel. The seeds are attached by the placenta to the central parenchymatous axis of the tomato, the columella. Towards the mature green stage, the placenta cells start to increase in size dramatically and liquify, producing the locular gel. Fruit development occurs in three partially overlapping stages after fertilization: cell division, cell expansion and ripening (**Figure 2b**). The plant hormones auxin and Gibberellin (GA) play important roles in fruit set and growth (Gillaspy et al., 1993; Böttcher et al., 2010; Hu et al., 2018). Five to seven days after the beginning of ripening (breaker stage), the fruit is fully red and ripe (Abewoy Fentik, 2017).

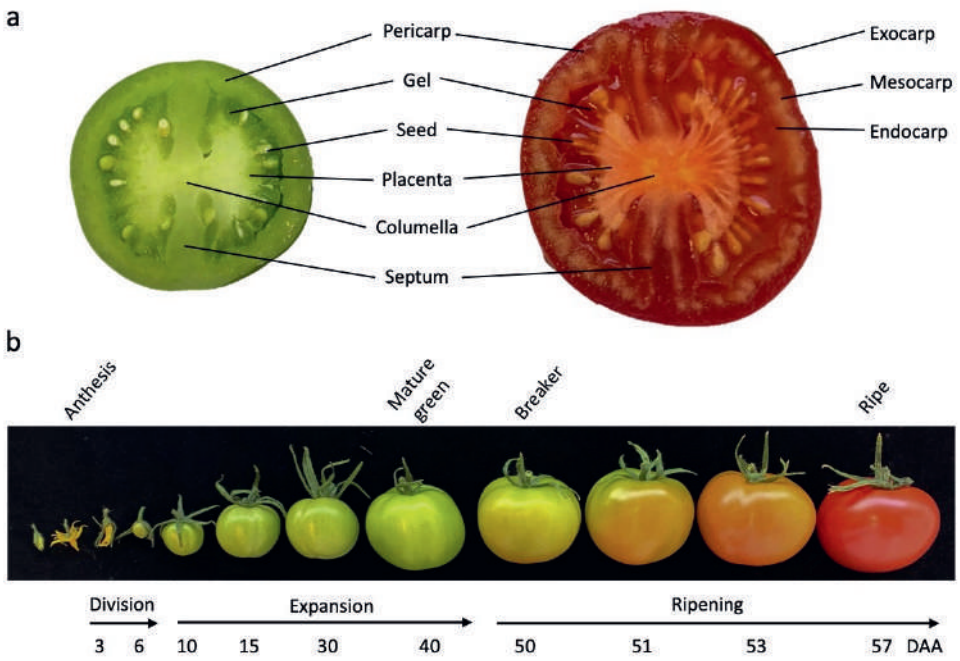


Figure 2: (a) Tomato fruit anatomy of the cross section of a green and ripe stage tomato **(b)** A time series of tomato fruit development from flower to ripe stage. The numbers are an indication of the Days After Anthesis (DAA). Exact time to ripening varies per variety and per season

Tomato is a climacteric fruit, as there is a burst of respiration and ethylene production at the onset of ripening (Gray et al., 1994; Deikman, 1997). In tomato two systems of ethylene production and regulation exist (Lelievre et al., 1997; Barry et al., 2000). The auto-inhibitory System I is functional in developing fruit. In this system, ethylene synthesis is inhibited through negative feedback, resulting in low levels of ethylene. During ripening, the autocatalytic System II is activated, and ETHYLENE RECEPTORS (ETR) perceive and bind ethylene followed by gene expression changes that regulate developmental processes. In tomato, a spontaneous mutant of *ETR3,* also called Never Ripe (NR), is a dominant-negative mutation in the ethylene binding-domain of ETR3 and blocks the ripening process (Lanahan et al., 1994; Wilkinson et al., 1995). Exogenously applied ethylene can induce ripening at the mature green stage, while the inhibitor of ethylene perception, 1-methyl cyclopropane (1-MCP) inhibits ripening (Yokotani et al., 2009). Upon ripening the fruit colour changes as the chloroplasts turn into chromoplasts, where chlorophyll is degraded and the carotenoids lycopene and β-carotene are formed (Shneour and Zabin, 1959; Cunningham and Gantt, 1998). The tomato texture softens as the cell wall polysaccharides pectin, cellulose, and hemicellulose are degraded (Giovannoni et al., 1989; Lashbrook et al., 1994; Hyodo et al., 2013; Uluisik et al., 2016). Delaying this process can extend the important agro-economical trait of shelf-life.

TOMATO FLAVOUR

An important process during tomato development and ripening, is the development of the flavour. The research group of Harry Klee in the U.S. has charted the sugars, acids and volatiles positively and negatively associated with tomato flavour (Tieman et al., 2012, 2017; Klee and Tieman, 2018) (**Table 2**). In Europe a study of flavour has been performed as well, but in that study no specific compounds were measured (Causse et al., 2010). However, they confirmed that diversification of flavour and texture is required to satisfy the wishes of consumers all over Europe, and in extension the world. The flavour of a tomato is a combination of perceived taste, mouthfeel (or texture) and smell. Five classes of receptors on the tongue sense the levels of sweet, sour, salty, bitter and umami. Olfaction, or smell, is essential for flavour perception. In tomato the most important contributors to overall flavour are sugars, acids, salts, minerals, volatiles, and texture. For texture the two major contributing factors are firmness of the fruit pericarp and fruit juiciness (Causse, 2002; Schouten et al., 2019). These attributes can be further broken down in texture components such as firmness, meltiness, mealiness, juiciness and crunchiness and can be measured by sensory analysis, mechanical methods, near infrared spectroscopy or hyper-spectral imaging methods (Harker et al., 2002; Barrett et al., 2010). Fruit softening is a major determining aspect of fruit texture and occurs during ripening and during post-harvest storage and handling (Bertin and Génard, 2018).

VOLATILES

Volatiles are an important determinate of tomato flavour. Volatiles are perceived in two ways by humans. Before ingestion perception occurs through the nostrils by olfaction. After ingestion, volatiles are forced up behind the palate into the nasal cavity leading to retronasal olfaction (Tieman et al., 2012). 400 volatile compounds have been detected in tomato (Petró-Turza, 1986; Rambla et al., 2014). Most flavour-related volatiles in tomato are derived from amino acids, lipids or carotenoids (Buttery and Ling, 1993; Yilmaz, 2001; Rambla et al., 2014). Only about a 30 volatiles have a high enough concentration to be perceived by humans and contribute significantly to flavour (**Table 2**, Yilmaz, 2001; Tieman et al., 2017). An example of an important flavour-contributing volatile is 2-phenylethanol, which contributes to tomato fruit aroma with fruity and floral properties (Baldwin et al., 2000; Rambla et al., 2014). An overabundance of this volatile can lead to a decrease in likability, and thus balance is key (Tadmor et al., 2002; Tieman et al., 2017). Most often, volatiles act in unison. For example, a combination of cis-3-hexenal, cis-3-hexenol, hexanal, 1-penten-3-one, 3-methylbutanal, trans-2-hexenal, 6-methyl-5-hepten-2-one, methyl salicylate, 2- isobutylthiazole, and β-ionone at the proper concentrations produces the aroma of fresh ripe tomato (Buttery and Ling, 1993). When these compounds are diminished, a 'processed' or 'enzymic' flavour is perceived (Kazeniac and Hall, 1970).

Modern cultivars have decreased production of 13 flavour-increasing volatiles compared to heirloom varieties (Tieman et al., 2017). Variability in these heirlooms and in wild tomato relatives has been and will be a valuable source for flavour-related traits to be used in (advanced) backcross populations with information on QTLs and GWAS loci (Sauvage et al., 2014; Tieman et al., 2017; Zhao et al., 2019). The carotenoid-derived 6-methyl-5-hepten-2-one (MHO) is an example of a volatile that has been positively impacted by selections sweeps during breeding, as it associated with a deep-red colour (Lewinsohn et al., 2005; Vogel et al., 2010).

Table 2: Attributes and compounds that contribute to overall liking and/or Tomato Flavour intensity.

Sig.	Attribute	Sig.	Sugars	Sig.	Acids
+	Flavour intensity	+	Brix (soluble solids)	+	citric acid
+	Sweetness	+	fructose	+	glutamic acid
+	Sourness	+	glucose	-	malic acid
+	Saltiness				
+	Umami				

Sig.	Volatiles	Sig.	Volatiles	Sig.	Volatiles
+	1-nitro-2-phenylethane	+	benzaldehyde	+	geranylacetone
+	1-nitro-3-methylbutane	+	benzyl alcohol	+	heptaldehyde
+	1-octen-3-one	+	benzyl cyanide	-	hexyl acetate
+	1-pentanol	+	β-cyclocitral	-	isobutyl acetate
+	1-penten-3-one	+	β-ionone	+	isovaleraldehyde
+	2,5-dimethyl-4-hydroxy-3(2H)-furanone	-	butyl acetate	+	isovaleric acid
+	2-isobutylthiazole	+	*E,E*-2,4-decadienal	+	isovaleronitrile
+	2-methyl-1-butanol	+	*E*-2-heptenal	+	methional
+	2-phenylethanol	+	*E*-2-pentenal	+	nonyl aldehyde
+	3-methyl-1-butanol	+	*E*-3-hexen-1-ol	+	phenylacetaldehyde
+	3-pentanone	-	eugenol	-	prenyl acetate
+	6-methyl-5-hepten-2-ol	+	furaneol	-	salicylaldehyde
+	6-methyl-5-hepten-2-one	+	geranial	+	*Z*-4-decenal

A "+" or "-" in the significance (Sig.). column in front of the compound signifies a significant positive and negative correlation between liking and the compounds concentration, respectively. Data taken from previously published consumer evaluation panels, all done in the U.S. (Tieman et al., 2012, 2017; Klee and Tieman, 2018).

MINERAL, ACIDS, AND SUGARS

A fresh tomato is made up of 90-95% water. Of the residual dry matter, 8% are minerals, 15% are organic acids and free amino acids and 50% are sugars (Yilmaz, 2001; Bertin and Génard, 2018). Potassium and phosphorus are the two major minerals present (Petró-Turza, 1986). Minerals influence the pH and have a buffering capacity. The main organic acids are malic acid and citric acids, of which the latter is the most abundant (Agius et al., 2018). They determine the pH of tomato juice, which usually ranges between 3.9 and 4.9 (Etienne et al., 2013). The positive perception of fruit acidity comes from citric acid, while malic acid content is negatively correlated with liking (**Table 2**). Off the free amino acids, glutamic acid is the most abundant (45% of free amino acids

by weight) and it is an important contributor to flavour responsible for umami taste (Yilmaz, 2001; Jinap and Hajeb, 2010; Sorrequieta et al., 2010; Tieman et al., 2017).

In ripe tomatoes, sugar is mostly present as glucose and fructose, whose amounts can be measured with biochemical techniques, such as with High-Performance Liquid Chromatography (HPLC) (Agius et al., 2018). A faster, cheaper, and widely used method to estimate sugar content, is the measurement of soluble solids, Brix, where one degree Brix (°Brix) corresponds to 1% (weight/weight) sucrose. As tomato contains mostly glucose and fructose, the °Brix is only an approximation of the dissolved sugar content. Fruits that weigh over 15 grams usually do not have a Brix over 5.6 degrees (Roohanitaziani, 2019). Sugar content of tomato was one of the major traits that suffered during domestication, as was demonstrated by Genome-Wide association studies (GWAS) (Tieman et al., 2017; Zhao et al., 2019). In this GWAS two chromosomal regions, on chromosomes 9 and 11, that have significant associations with sugar content have also been identified as regions influenced by domestication and improvement sweeps during the selection of larger fruits (Tieman et al., 2017). Two loci, *Brix9-2-5*, later identified as the tomato cell wall (Lycopersicum) INvertase gene *LIN5*, and *SSC11.*1 were significantly associated with sugar content (Tanksley et al., 1996; Tieman et al., 2017; Kimbara et al., 2018). Almost all modern cultivars contain the reference alleles at these two loci, concurrent with lower sugar content.

BREEDING FOR FLAVOUR

Knowledge on the different aspects contributing to flavour is valuable for breeders and will help to improve tomato flavour in the future. Flavour is one of most challenging traits. Other quality aspects such as external appearance (size, shape, colour), texture (firmness, mealiness and juiciness), nutritional values and shelf-life have received a lot more attention. Flavour is a complex and multigenic trait and it is time-consuming and expensive to phenotype (Tieman et al., 2017). On top of that, improving flavour-aspects can have serious trade-offs, for instance, in yield or shelf-life. The link between yield and the loss of flavour can already be traced back to the early stages of human intervention when the selection for larger fruits was coupled by a decrease in flavour-related components (Powell et al., 2012; Zhu et al., 2018b). Shelf-life has been an important breeding-goal since the 1980s (Bai and Lindhout, 2007). Up till now, delaying ripening, early harvesting at immature stages and post-harvest handling such as cold-storage and chemical treatments have contributed to major flavour loss and resulted in increasing complaints from consumers (Boukobza and Taylor, 2002; Mahajan et al., 2014; Sandarani et al., 2018). Harvesting at an early stage of ripening and cold-storage have especially drastic effects on volatile production (Maul et al., 2000; Bai et al., 2011; Sanchez-Bel et al., 2012; Raffo et al., 2012; Renard et al., 2013; Zhang et al., 2016). Introduction of non-ripening mutations such as *ripening inhibitor* (*rin*) in modern cultivars, have led to slow-ripening fruits that have a longer shelf life, but at the cost of flavour (Kitagawa et al., 2005). Recently, fruit softening has been uncoupled from

ripening, giving some hope that the use of non-ripening mutations can be avoided in the future (Uluisik et al., 2016). Although breaking the negative correlation between yield and flavour would be preferable, consumers have indicated that they are willing to pay more for a better tasting product (Bruhn et al., 1991; Oltman et al., 2014).

Since the 1990s, flavour has been back in focus in a response to complaints about "Wasserbomben" – big watery, tasteless orbs - and the popularity of heirloom varieties (Terhorst, 2006; Tieman et al., 2017; Bauchet et al., 2017). The public perception is that heirlooms have superior taste compared to modern cultivars, although this is definitely not always the case (Tieman et al., 2012). Recent work has shown that since the "Wasserbomben-crisis", the diversity in fruit types and flavours has increased considerably (Schouten et al., 2019). However, consumers are still far from satisfied as became clear from a Swedish consumer survey where more than half of the respondents indicated they were dissatisfied with tomato taste, while they expect good flavour delivered at affordable prices (Fernqvist and Hunter, 2012). The whole tomato production chain – from breeder to retailer – requires a more consumer preference-driven selection approach. In this thesis we have researched one aspect of tomato flavour: the sugar content.

TWO IMPORTANT GENES IN FRUIT SUGAR CONTENT

Sugars are the product of photosynthesis and sucrose is the main sugar that travels from the source, e.g. leaves, to the sinks, e.g. fruit. Sucrose is loaded into the phloem at the source and is unloaded in fruit via the symplastic or the apoplastic route (see chapter 2). Before import into the cell or within the cell itself, sucrose is cleaved into glucose and fructose. Developing fruit accumulates these hexoses and starch, a sugar polymer. Starch is stored in the columella, placenta and septum and its amount peaks around 10-25 days after anthesis (Schaffer and Petreikov, 1997). At the onset of ripening, starch is completely hydrolysed and contributes significantly to final soluble sugar (Brix) content (Bertin and Génard, 2018). A more detailed overview of the sugar pathways in tomato is described in **Chapter 2**. In this general introduction, we have highlighted the two genes that have taken the central role in this thesis: *LIN5* and *ADP-Glucose Pyrophosphorylase Large Subunit 1 (AGPL1)* (**Figure 3**).

LIN5 was discovered as the underlying gene for the *Brix9-2-5* QTL in a tomato line with a *S. pennellii* introgression (Fridman et al., 2002; Baxter et al., 2005b, 2005a). *LIN5* encodes one of the tomato Cell Wall Invertases (CWIN). Invertases cleave the O-C bond in sucrose, breaking it into the monosaccharides glucose and fructose to drive sugar import into the cell (**Figure 3**). The importance of *LIN5* in tomato Brix was further demonstrated by knocking down its expression with RNA interference (RNAi) and by silencing the posttranslational inhibitor of LIN5, *Cell-wall Inhibitor of β-fructosidase* (*CIF1* or *INVINH1*). This led to a lower and higher sugar content in fruits, respectively (Zanor et al., 2009; Jin et al., 2009b). As mentioned before, *LIN5* is associated with an

improvement sweep for fruit size, where a high expression and activity of *LIN5* negatively correlated with size (Tieman et al., 2017). In this *Brix9-2-5* QTL, three amino acid substitutions (positions 348, 366 and 373) in the *S. pennellii LIN5* allele compared to *S. lycopersicum* were linked to increased enzyme-activity and to the increased Brix phenotype (Fridman et al., 2000, 2004). It was hypothesized that the Glu to Asp change at position 348 would be the most important, as it is closest to the active site of the invertase. The Asn to Asp change at position 366 was also correlated with an increase in sugar when constitutively expressed (Tieman et al., 2017).

The other topic of this thesis is *AGPL1*. It encodes the large subunit of the ADP-Glucose Pyrophosphorylase (AGPase). AGPase catalyses the rate-limiting reaction between ATP and glucose 1-phosphate (G1P) to produce ADP-glucose and inorganic pyrophosphate in plastids (**Figure 3,** Ballicora et al., 2004b; Beckles et al., 2001). This is the first dedicated step in the production of starch. At the onset of ripening, stored starch is hydrolysed to form glucose, maltose and G1P, significantly influencing the final sugar content of tomato. AGPase is a tetramer, consisting of two stabilizing large subunits and two catalysing small subunits. Of these subunits, the large subunit *AGPL1* was correlated with high AGPase activity and high starch levels in the immature fruit and high soluble solids content in the mature fruit (Schaffer et al., 2000). *AGPL1* is expressed in immature fruit and is downregulated before the onset of ripening. In an *S. habrochaites* Introgression line, it was found that plants harbouring the *S. habrochaites AGPL1* allele (*AGPL1*^{H}) had higher starch accumulation and increased sugar content, probably caused by the extended temporal expression of *AGPL1* (Petreikov et al., 2006).

Both *LIN5* and *AGPL1* play an important role in sugar accumulation and thus in the perceived tomato quality. However, apart from the described QTLs, variation is not present in modern cultivars (Tieman et al., 2017; Zhao et al., 2019). Thus, it would be interesting to explore how genetic techniques can (re-)create variation and boost LIN5 and AGPL1 activity in developing fruit. In this thesis, we have pursued this by studying and modifying transcription regulation of *LIN5* and *AGPL1*.

TRANSCRIPTIONAL REGULATION OF GENES

The regulation of expression of genes through modulating transcription is a vital process for any organism. Precise patterns of gene activity result in differentiated cell types, organs, structures. These patterns enable an organism to respond to environmental changes. In essence, there are two core components of transcriptional regulation: (1) Transcription factors (TFs) that bind DNA on (2) *Cis*-Regulatory Elements (CRE), specific sequence motifs that affect transcription of a gene (Bulger and Groudine, 2011; Wittkopp and Kalay, 2012; Yáñez-Cuna et al., 2013). Upon binding to a specific DNA motif, TFs can activate or repress transcription of a nearby or distal gene (Latchman, 1997) (**Figure 4**). Through the interaction with TFs, CREs define the

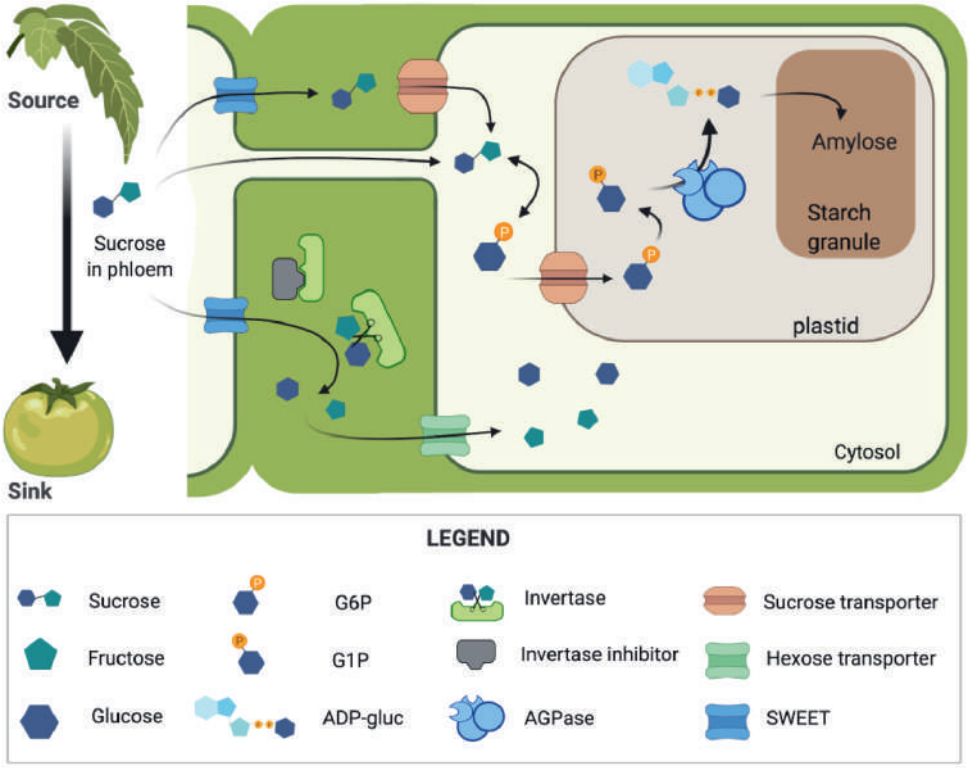


Figure 3: The role of cell wall invertase (LIN5) and AGPase. A simplified representation of sugar transport and storage in a sink parenchyma cell. Sugars, enzymes, and transporters are represented by figures, see the legend for their names. Only the invertase (*LIN5*), AGPase (consisting of *AGPL1* and *AGPS)* and transporters are depicted in this simplified representation. Created with BioRender.com.

spatiotemporal expression patterns of their corresponding genes. Understanding the separate elements and their interactions within their native environments enables researchers and breeders to understand and tweak gene regulatory networks, design specific TFs and predict and design spatiotemporal expression patterns by using CREs.

TRANSCRIPTION FACTORS

The next layer in transcriptional regulation comes from TFs. By combining and balancing pioneering, activating, and repressing activities, specific expression patterns emerge. Positive and negative feedback loops, flip-flop devices and feed forward loops are all methods of fine-tuning expression and make a cell able to adjust expression levels based on new signals. It is common that a certain TF can regulate a multitude of genes. TFs can have either a repressing or activating effect on gene expression. Activators bind and attract general TFs to accelerate their assembly, attract the transcription initiation complex or aid polymerase binding and release, directly or indirectly through interaction with other proteins. In addition, TFs can have

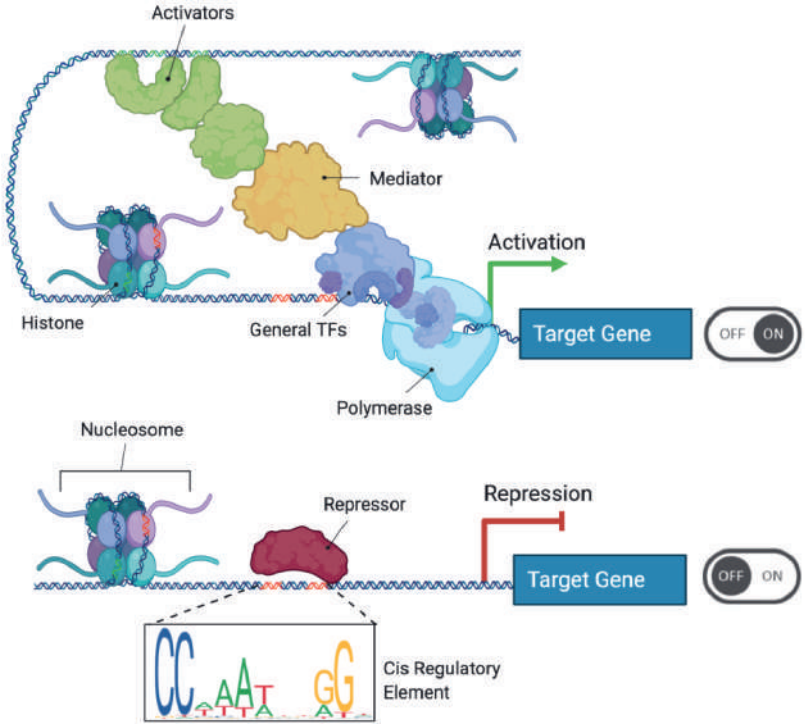


Figure 4: Regulation of genes by *Cis*-Regulatory Elements (CREs) and Transcription factors. Transcription factors can act as activators or as repressors upon binding to CREs in the promoter of a gene. TF can act in a combinatorial way. CREs wrapped in nucleosomes, a complex of histones and DNA, are less accessible to TFs. Created with BioRender.com.

a structural activation domain that can accelerate the rate of transcription initiation. Repressor TF activity can occur when the bound TF blocks or masks activator binding, blocks the general TFs or polymerase or if it recruits chromatin remodelling complexes, histone deacetylases and histone methyl transferases. One TF can act both as a repressor and as activator: the regulatory function depends on the binding location in the promoter, the context, and on the complexes it forms with other proteins and co-factors under different conditions (Bauer et al., 2010; Maher et al., 2018). In all cases, specific DNA recognition and binding is essential.

THE *CIS*-REGULATORY CODE

To correctly regulate genes, TF binding must be specific and timely. TF binding with the DNA is achieved through highly conserved DNA binding domains in the amino-acid chain of the TF protein (Jolma et al., 2013). The outside of the doubled helix DNA strand presents a distinctive pattern of hydrogen bonds and hydrophobic patches for a DNA binding protein to recognize in a specific and strong manner. As the differences between the different base-pairs is more marked in the outer groove, the majority of

TF specific interactions occurs there. Though the specifics differ for each class and specific members of TFs families, two general types of DNA binding domains are found that bind to the major groove: the β-sheets and the α-helixes. Specificity comes from the amino acids and angle. The angle and positioning relative to the DNA is influenced by the number of β-sheets and α-helixes and the rest of the TF polypeptide chain (Sayou et al., 2016). This chain also makes other contacts with the DNA, fine-tuning the interaction. Common types of DNA binding motifs include the helix-turn-helix, the helix-loop-helix, β-hairpin/ribbons and zinc fingers (Luscombe et al., 2000). Many TFs form homodimers, heterodimers and need co-factors to strengthen and specify DNA binding (Jolma et al., 2013).

Binding preferences are often similar for closely related TFs (Weirauch et al., 2014). TFs are grouped into TF families according to their conserved domains, often the DNA binding domain. Examples in plants are the AP2/ERF, bHLH, MYB, WRKY and the MADS TF family. The MADS domain TFs for example bind to a CarG box ($CC(A/T)_6GG$) (West et al., 1997). However, *in vitro* and *in vivo* binding specificity may differ for each family member, even though the DNA binding domain is conserved. One reason is the spatiotemporal expression patterns of a TF and the binding to coregulators (Tao et al., 2012; Völkel et al., 2015). Because of this, including co-expression analysis with target genes can greatly increase the prediction of strength of TF binding and functionality. Even related proteins expressed in the same cell can bind differently *in vivo* to targets as specificity may be co-determined by specific protein-protein interactions, such as with additional co-factors, and due to differences in bindings sites with low and high affinity (Tanay, 2006; Tao et al., 2012; Völkel et al., 2015; Bemer et al., 2017). In addition, not all binding events lead to a transcriptional response.

CREs and TFs function in a synergistic complex network to achieve precise patterns of gene expression (Kaufmann et al., 2010; Karlova et al., 2014; Wang et al., 2019a). Influenced by the spatiotemporal expression of the TFs and the manner of protein-protein interaction, interaction can lead to cooperative, additive and independent binding (Wilczyski and Furlong, 2010; Spitz and Furlong, 2012). The importance of TFs in whole networks is exemplified when they are disrupted. Good examples are the tomato ripening regulators MADS-RIN, NAC-NOR, SPL-CNR, AP2a and FUL1/FUL2. Disruptions in these genes by natural variation, RNAi or CRISPR/Cas9 mutagenesis, can cause severe ripening defects (Robinson and Tomes, 1968; Tigchelaar et al., 1973; Thompson et al., 1999; Karlova et al., 2011; Bemer et al., 2012; Wang et al., 2019b).

The 1-2Kb region upstream the transcription start site, the proximal promoter, is often enriched in TF binding sites (Yamamoto et al., 2009; Yu et al., 2016). In plants it has been shown that broadly expressed genes have conserved, GC-rich promoters, while tissue-specific genes have less stable or AT-rich promoters (Das and Bansal, 2019). Accessible chromatin, increased GC-content and conservation are indicators of putative CRE-harbouring regions (Ricci et al., 2019). TF binding at the CREs in the

promoter is most likely facilitated by the sequence of the promoters' motifs and by the sequence environment (Dror et al., 2016).

CREs and groups of CREs in distal regions are called enhancers (Banerji et al., 1981). In maize, 32.5% of putative enhancers, accessible chromatin regions, were >2kb distal from their nearest gene (Ricci et al., 2019). Enhancers are separated by spacers and insulators. These can vary enormously in size, from a few to over 20Kb (Ricci et al., 2019). Spacers and insulators provide flexibility, chromatin packaging, aid correct looping, prevent crosstalk between regulatory regions and limit the spread of heterochromatin (Gaszner and Felsenfeld, 2006). Enhancers can both stimulate and repress gene activity (Gisselbrecht et al., 2020). However, repressing-enhancers are hard to detect in the much-used reporter-gene assays. A single enhancer can form loops with multiple genes, as was the case with 34% of the loops in the maize genome (Ricci et al., 2019).

Enhancers function through direct interaction with their target promoter through TF and Mediator aided DNA-looping. In maize, loops between accessible chromatin and genes spanning over 20kb were a common feature (Ricci et al., 2019). TF binding can bring enhancers and the promoter close together by forming loops through protein complex formation. Additionally, a connecting and stabilizing layer, or bridge, between the TFs and the transcription initiation complex is formed by the large protein complex Mediator. The molecular structure of Mediator is conserved with a head, middle, tail and CDK8 part. Specific TFs haven been shown to interact with specific plant Mediator components, giving rise to tissue-specific interactions (Buendía-Monreal and Gillmor, 2016). Mediator physically connects enhancers and the core promoters through the formation of loops (Kagey et al., 2010). Extrapolating from this, mediator, cohesion, and polymerase II localization can be used to predict DNA looping and enhancer sites. Other hallmarks of enhancers are specific histones, such as H2A.Z, on the flanks and an open chromatin confirmation (Ricci et al., 2019).

EPIGENETIC REGULATION

Different signals converge on a promoter. Several to hundreds of TFs, co-factors, and structural cues such as DNA looping, nucleosome packing, chromatin condensation, and DNA methylation are integrated to form a specific spatiotemporal and environment- dependent expression pattern. As discussed above, looping of DNA is primarily achieved by higher order protein-DNA interaction, bringing enhancers, promoters, and the core promoter close together. In addition, the local shape of a binding site (helix twist, minor groove width, propeller twist and roll) is influenced by surrounding nucleotides and plays a role in achieving specificity for TF binding (Gordân et al., 2013). This mechanistic role of the native surrounding of a CRE can explain a TF's *in vivo* preference for just a subset of all possible target sites. Using DNA shape features improves TF binding predictions (Mathelier et al., 2016).

DNA methylation is the covalent addition of a methyl group to cytosine (C) in CG, CpG, GHG and CHH (where H corresponds to A, T, or C) sequences (Suzuki and Bird, 2008). DNA methylation can be inherited stably but can also undergo dynamic changes during development (Klose and Bird, 2006). In most cases, DNA methylation is correlated with silenced genes (Holliday and Pugh, 1975; Riggs, 1975; Hsieh, 1994). The methyl groups are thought to physically obstruct TF-DNA binding (Wade, 2001; Hendrich and Tweedie, 2003; Maurano et al., 2015; Domcke et al., 2015), and alter the DNA shape (Tippin et al., 1997; Buck-Koehntop et al., 2012; Lazarovici et al., 2013). Another theory is that methylated DNA attract specific TFs, blocking access to others (Zhang et al., 2010, 2018). As DNA methylation affects the DNA shape and accessibility, DNA methylation contributes to TF-DNA interactions. In a study in *Arabidopsis*, the DNA binding of 76% of all TFs were sensitive to DNA methylation status (O'Malley et al., 2016).

A specific part of the DNA landscape is DNA packaging. DNA is packed into higher order forms of chromatin structure around histone proteins forming nucleosomes. Packing reduces the accessibility of DNA by TFs and is correlated with lower transcriptional activity. The location and interaction strengths of nucleosomes can be influenced by histone-tail modifications. There are more than a hundred modifications known, making DNA accessibility a dynamic and complex process (Tan et al., 2011). Modifications of histone tails include methylation, acetylation, phosphorylation, ubiquitination, SUMOylation, citrullination and ADP-ribosylation (Narlikar et al., 2002; Cano-Rodriguez et al., 2016; Ricci et al., 2019). As the transcription initiation complex seems unable to bind condensed chromatin, specific activator TFs called pioneering factors have an important role in chromatin opening through their ability to bind to specific motifs on a nucleosome, in contrast to non-pioneering TFs that can only bind non-nucleosome-bound DNA (Soufi et al., 2015, Magnani et al., 2011). Chromatin opening is achieved by covalent histone modifications leading to nucleosome remodelling, removal, and replacement. Activators attract histone modification enzymes, chromatin remodelling complexes and histone chaperones. Specific histone marks are then recognized by subsequent proteins, continuing the cascade (He et al., 2011). Other TFs can stabilize the nucleosome, repressing gene expression (Zhu et al., 2018a), while nucleosome depleted regions are characteristic of active core promoters (Mavrich et al., 2008; Jiang and Pugh, 2009; Jin et al., 2009a; Mueller et al., 2017). In plants, it has been found that the majority of open chromatin regions lies 3kb upstream of transcription start site and that open regions cluster together (Maher et al., 2018). In many plant species, especially with large genomes, distal CREs have been created and separated from the transcription start site by Transposable Element proliferation (Lu et al., 2019).

The complex network of integrated signals determines the amount, conditions, time, and space of expression. The whole interplay of CREs, TFs, epigenetics and post-transcriptional effects such as RNA processing splicing, transport selection, localization and degradation result in a complete picture of gene expression as part of a gene

regulatory network (Chen and Rajewsky, 2007; Mejia-Guerra et al., 2012). Changes in gene expression have been shown to underlie more of life's variation than the acquisition of new genes (Wittkopp and Kalay, 2012). Many crop QTLs affecting quality and yield, have been the result of expression differences of underlying genes rather than of differences in the gene products themselves. This observation holds true for many domestication-associated traits (Doebley et al., 2006; Hufford et al., 2012; Ye et al., 2017). Variation in CREs has been a particular driving force in in evolution and breeding (Hufford et al., 2012; Meyer and Purugganan, 2013; Swinnen et al., 2016). Recently, the study of the pangenome of structural variants in tomato revealed that that in addition to SNPs, structural variations in promoters are associated with changes in expression that occurred during domestication or subsequent improvement by breeding (Alonge et al., 2020). An example of a CRE mutation leading to interesting phenotypes in tomato comes from a Genome-Wide Association Study (GWAS) studying malate content. It was found that a 3 bp insertion in the promoter of *Al-Activated Malate Transporter 9* (*SlALMT9*) disrupted a W-box binding site, disrupting binding of a WRKY TF, *WRK42* and increasing *SlAMT9* expression. Increased *SlAMT9* expression led to an increased malate accumulation and enhanced aluminium tolerance (Ye et al., 2017). Two QTLs for locule number are *fas*, an allele with a decreased expression of *CLAVATA 3* (*SlCLV3*), and *lc*, a weak gain-of-function allele of *WUSCHEL* (*SlWUS*) (Huang and van der Knaap, 2011; Xu et al., 2015; Somssich et al., 2016). Both locule number QTLs are most likely caused by changes in the promoters of the two underlying genes. In the *SlCLV3* locus the promoter is inverted and a deletion in the 3' region of *SlWUS* has a putative effect on the binding of a repressor. A recent study used Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR- associated Protein (Cas9) - (CRISPR/Cas) mutagenesis on the promoters of *SlCVL3-SlWUS* to increase locule number and size (Rodríguez-Leal et al., 2017). The possibility to alter CREs in a way we desire, creates new opportunities for research and breeding with fewer detrimental pleiotropic effects than would be caused by entirely knocking out gene function or ectopically expressing it (Swinnen et al., 2016).

STUDY OF PROMOTER FUNCTION

Better understanding transcription regulation and the impact on gene regulatory networks would create these new opportunities for quality breeding in plants. A central question in transcriptional regulation research is to define the regulatory region of a gene and which TFs and CREs influence transcription. Promoter studies with reporter genes have been a widely used method to assess the regulatory region directly upstream of a particular gene. Next Generation Sequencing methods have been a recent widely applied approach in high-throughput and genome wide transcriptional regulation studies, with techniques as RNA-sequencing, DNAse I hypersensitive site sequencing (DNAseI-seq), Chromatin immune Precipitation sequencing (ChIP-seq), Chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) and bisulphite-

sequencing (Meyer et al., 2012). Applications as DNA affinity purification sequencing (DAP-seq), Yeast-one-Hybrid assays (Y1H), ChIP-seq and Electrophoretic Mobility Shift assays (EMSAs) indicate physical interactions between regions of DNA and a TF. Validating gene regulation can be done with a promoter reporter assay and transient expression systems (Bargmann et al., 2013; Kaufmann and Mueller-Roeber, 2018).

CRIPSR/CAS-MEDIATED MUTAGENESIS

CRISPR/Cas mutagenesis of CREs could be used to determine if disruption of a putative CRE leads to expression changes and an altered phenotype. In this thesis, we tested this method by performing a comprehensive study on the promoters of *LIN5* and *AGPL1*. Since variations in *LIN5* and *AGPL1* have been proven to influence sugar content, it makes them interesting targets for experimenting with New Breeding Technologies (NBTs). With NBTs, such as CRISPR/Cas mutagenesis, it is possible to make changes in any tomato variety without the need of introducing a gene from a with wild tomato variety by a lengthy introgression process. In addition, completely new genotypes and phenotypes can be created that are not possible by introgression.

Cas9 is an RNA-Guided double-strand DNA endonuclease that targets a specific genomic sequence to create a double stranded break (Belhaj et al., 2015). The CRISPR/Cas system was initially discovered as an immune system in bacteria and Archaea (Bhaya et al., 2011). However, the potential for gene editing was quickly realized and the system was adapted and developed rapidly for gene editing in eukaryotic cells (Jinek et al., 2012; Cong et al., 2013; Li et al., 2013; Shan et al., 2013; Mali et al., 2013a). CRISPR/Cas9-mutagenesis was soon successfully applied in tomato as well (Brooks et al., 2014; Ron et al., 2014).

The ability to use a customizable small noncoding RNA**,** single guide RNA (sgRNA), has made CRISPR/Cas9 ideal for genome editing purposes (Jinek et al., 2012; Cong et al., 2013; Belhaj et al., 2013; Voytas and Gao, 2014). The first step of target recognition depends on the presence of a protospacer adjacent motif (PAM), typically NGG for the nuclease SpCas9, adjacent to the 3' end of the 20 bp target (Jinek et al., 2012). Cas9 unwinds the target DNA sequence using sgRNA-DNA base pairing and cuts both DNA strands 3 bp upstream of the PAM site. This creates a blunt-ended double-stranded break **(Figure 5a)**. In addition, Cas9 can induce a 1-bp staggered break, often leading to 1 nt insertions (Shi et al., 2019). A Double Stranded Break (DSB) in the DNA is repaired by either Non-Homologous End Joining (NHEJ) or Homology Directed Repair (HDR) (Symington and Gautier, 2011). During classical NHEJ the loose ends are re-joined in an error-prone mechanism, creating small insertions or deletions. Repair through the alternative end joining or microhomology mediated end-joining (MMEJ) pathway frequently results in in larger deletion between microhomology sites of 4-25 bp near the DSB (Deriano and Roth, 2013; van Overbeek et al., 2016; Tan et al., 2020).

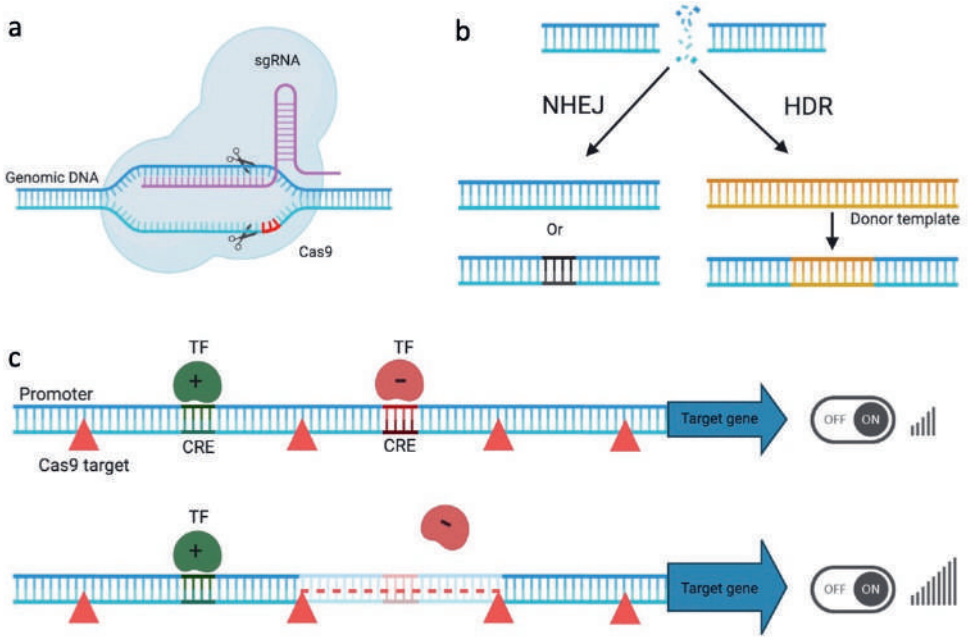


Figure 5: CRISPR/Cas9 as a genome editing tool (**a)** CRISPR/Cas9-DNA/RNA complex. A double stranded break is made by Cas9 3 bp upstream of the PAM site (Red) **(b)** A double stranded break is repaired through either NHEJ or HDR. NHEJ can lead to perfect repair or creates small Insertions or Deletions (black region). HDR uses a template (yellow) for repairing the double stranded break, thereby replacing the original sequence **(c)** Bigger deletions are achieved by multiplexing. Red triangles represent CRISPR/Cas9-induced double stranded breaks. A generated deletion is represented by the dashed red line, disrupting a repressing TF binding site (red coloured DNA strand and protein).

HDR uses an available homologous DNA donor template to repair the break, creating an insertion or replacement of the original with the donor. The template can be the same locus on the sister chromatid or an exogenous piece of DNA (**Figure 5b**) (Symington and Gautier, 2011). In plant cells double stranded breaks are primarily repaired through NHEJ that introduce frameshift mutations or deletions (Shukla et al., 2009; Knoll et al., 2014; Schiml et al., 2014). Larger deletions can be created by the use of multiplexed-genome editing wherein multiple sgRNAs are being used (Cong et al., 2013; Mali et al., 2013b). Simultaneous DSB from two or more sgRNAs can then result in a deletion of the DNA between the two cuts.

OUTLINE OF THIS THESIS

In this thesis, CRISPR/Cas9 multiplex gene targeting was used to create an allelic series of mutations for each of the proposed target genes' promoters (**Figure 5c**). Bigger deletions were achieved by multiplexed targeting of the promoter of our genes of interest. By generating deletions in promoter regions, we intended to study the regulatory function of the targeted CREs. When a mutation occurs, a CRE can be disrupted, disrupting the binding of a Transcription Factor (complex). If the disrupted site was the binding site of a repressor, the expression of the target gene should increase. By increasing expression, we aimed at increasing sugar content and hopefully, tomato flavour.

Chapter 2 reviews current knowledge and explores the sugar accumulation pathway in tomato. In this process, potential breeding targets to increase the sugar content are highlighted. Two of these targets are *AGPL1* and *LIN5*.

In **Chapter 3** (*AGPL1)* and **Chapter 4** (*LIN5*), we set out to gain insight into expression regulation of the two target genes during fruit development. The first step was to identify their *Cis*-regulatory elements and their interacting TFs. Subsequently and simultaneously, we implemented CRISPR/Cas9 targeted mutagenesis to create a range of systematic deletions in the promoters of both target genes. The hypothesis was that this would lead to a clear *in situ* picture of the CRE functions. With the addition of proven methods for CRE-TF interaction application, we worked towards a comprehensive model of the regulatory network of the two genes. This work provided leads for fruit quality improvement and gained us some insight into the potential of genome editing for altering gene expression (up- and down-regulation) in crops by targeting candidate gene CREs.

Chapter 5 explores the role of a post-translational inhibitor of LIN5*, CIF1*. We generated knockout mutants and studied the effect on final soluble solids content (Brix°) and invertase activity. **Chapter 6** focuses on the implementation of a new CRISPR/Cas9 tool: Gene Targeting. A major goal of plant gene editing is the use of homologous recombination to make an exact desired change. Our goal was to modify the three amino acids in LIN5 that were underlying the *Brix9-2-5* QTL. We tested Gene Targeting in tomato with a geminiviral-based replicon donor delivery system in protoplasts and in stable transformation.

A summary synthesis of the main outcomes presented in the preceding chapters is given in **Chapter 7** with emphasis on the implementation in breeding and research, societal implications and potential follow up research.

REFERENCES

Abewoy Fentik, D. (2017). Review on Genetics and Breeding of Tomato (*Lycopersicon esculentum* Mill). Adv. Crop Sci. Technol. **05**: 1–6.

Agius, C., von Tucher, S., Poppenberger, B., and Rozhon, W. (2018). Quantification of sugars and organic acids in tomato fruits. MethodsX **5**: 537–550.

Alonge, M. et al. (2020). Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. Cell **182**: 1–17.

Alpert, K.B., Grandillo, S., and Tanksley, S.D. (1995). *fw 2.2*:a major QTL controlling fruit weight is common to both red- and green-fruited tomato species. Theor. Appl. Genet. **91**: 994–1000.

Azanza, F., Kim, D., Tanksley, S.D., and Juvik, J.A. (1995). Genes from *Lycopersicon chmielewskii* affecting tomato quality during fruit ripening. Theor. Appl. Genet. **91**: 495–504.

Bai, J., Baldwin, E.A., Imahori, Y., Kostenyuk, I., Burns, J., and Brecht, J.K. (2011). Chilling and heating may regulate C6 volatile aroma production by different mechanisms in tomato (*Solanum lycopersicum*) fruit. Postharvest Biol. Technol. **60**: 111–120.

Bai, Y. and Lindhout, P. (2007). Domestication and breeding of tomatoes: What have we gained and what can we gain in the future? Ann. Bot. **100**: 1085–1094.

Baldwin, E.A., Scott, J.W., Shewmaker, C.K., and Schuch, W. (2000). Flavour trivia and tomato aroma: Biochemistry and possible mechanisms for control of important aroma components. HortScience **35**: 1013–1022.

Ballicora, M.A., Iglesias, A.A., and Preiss, J. (2004). ADP-glucose pyrophosphorylase: A regulatory enzyme for plant starch synthesis. Photosynth. Res. **79**: 1–24.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell **27**: 299–308.

Bargmann, B.O.R., Marshall-Colon, A., Efroni, I., Ruffel, S., Birnbaum, K.D., Coruzzi, G.M., and Krouk, G. (2013). TARGET: A Transient Transformation System for Genome-Wide Transcription Factor Target Discovery. Mol. Plant **6**: 978–980.

Barrett, D.M., Beaulieu, J.C., and Shewfelt, R. (2010). Colour, flavour, texture, and nutritional quality of fresh-cut fruits and vegetables: Desirable levels, instrumental and sensory measurement, and the effects of processing. Crit. Rev. Food Sci. Nutr. **50**: 369–389.

Barry, C.S., Llop-Tous, M.I., and Grierson, D. (2000). The regulation of 1-aminocyclopropane-1-carboxylic acid synthase gene expression during the transition from system-1 to system-2 ethylene synthesis in tomato. Plant Physiol. **123**: 979–986.

Bauchet, G. et al. (2017). Identification of major loci and genomic regions controlling acid and volatile content in tomato fruit: implications for flavour improvement. New Phytol. **215**: 624–641.

Bauer, D.C., Buske, F.A., and Bailey, T.L. (2010). Dual-functioning transcription factors in the developmental gene network of *Drosophila melanogaster*. BMC Bioinformatics **11**: 1–14.

Baxter, C.J., Carrari, F., Bauke, A., Overy, S., Hill, S.A., Quick, P.W., Fernie, A.R., and Sweetlove, L.J. (2005a). Fruit carbohydrate metabolism in an introgression line of tomato with increased fruit soluble solids. Plant Cell Physiol. **46**: 425–437.

Baxter, C.J., Sabar, M., Quick, W.P., and Sweetlove, L.J. (2005b). Comparison of changes in fruit gene expression in tomato introgression lines provides evidence of genome-wide transcriptional changes and reveals links to mapped QTLs and described traits. J. Exp. Bot. **56**: 1591–1604.

Beckles, D.M., Craig, J., and Smith, A.M. (2001). ADP-glucose pyrophosphorylase is located in the plastid in developing tomato fruit. Plant Physiol. **126**: 261–266.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., and Nekrasov, V. (2013). Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant Methods **9**: 39.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N.J., and Nekrasov, V. (2015). Editing plant genomes with CRISPR/Cas9. Curr. Opin. Biotechnol. **32**: 76–84.

Bemer, M., van Dijk, A.D.J., Immink, R.G.H., and Angenent, G.C. (2017). Cross-Family Transcription Factor Interactions: An Additional Layer of Gene Regulation. Trends Plant Sci. **22**: 66–80.

Bemer, M., Karlova, R., Ballester, A.R., Tikunov, Y.M., Bovy, A.G., Wolters-Arts, M., Rossetto, P. de B., Angenent, G.C., and de Maagd, R.A. (2012). The tomato FRUITFULL homologs TDR4/FUL1 and MBP7/FUL2 regulate ethylene-independent aspects of fruit ripening. Plant Cell **24**: 4437–51.

Bergougnoux, V. (2014). The history of tomato: From domestication to biopharming. Biotechnol. Adv. **32**: 170–189.

Bertin, N. and Génard, M. (2018). Tomato quality as influenced by preharvest factors. Sci. Hortic. (Amsterdam). **233**: 264–276.

Bhaya, D., Davison, M., and Barrangou, R. (2011). CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. Annu. Rev. Genet. **45**: 273–297.

Blanca, J., Cañizares, 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**: e48198.
Böttcher, C., Keyzers, R.A., Boss, P.K., and Davies, C. (2010). Sequestration of auxin by the indole-3-acetic acid-amido synthetase GH3-1 in grape berry (*Vitis vinifera* L.) and the proposed role of auxin conjugation during ripening. J. Exp. Bot. **61**: 3615–3625.
Boukobza, F. and Taylor, A.J. (2002). Effect of postharvest treatment on flavour volatiles of tomatoes. Postharvest Biol. Technol. **25**: 321–331.
Brooks, C., Nekrasov, V., Lippman, Z.B., and Van Eck, J. (2014). Efficient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-Associated9 System. Plant Physiol. **166**: 1292–1297.
Bruhn, C.M., Feldman, N., Garlitz, C., Harwood, J., Ivans, E., Marshall, M., Riley, A., Thurber, D., and Williamson, E. (1991). Consumer perception of quality: apricots, cantaloupes, peaches, pears, strawberries, and tomatoes. J. Food Qual. **14**: 187–195.
Buck-Koehntop, B.A., Stanfield, R.L., Ekiert, D.C., Martinez-Yamout, M.A., Dyson, H.J., Wilson, I.A., and Wright, P.E. (2012). Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. Proc. Natl. Acad. Sci. U. S. A. **109**: 15229–15234.
Buendía-Monreal, M. and Gillmor, C.S. (2016). Mediator: A key regulator of plant development. Dev. Biol. **419**: 7–18.
Bulger, M. and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell **144**: 327–339.
Burke, T.W. and Kadonaga, J.T. (1996). Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes Dev. **10**: 711–724.
Burke, T.W. and Kadonaga, J.T. (1997). The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAF(II)60 of *Drosophila*. Genes Dev. **11**: 3020–3031.
Buttery, R.G. and Ling, L.C. (1993). Volatile Components of Tomato Fruit and Plant Parts.
de Candolle, A. (1886). The origin of cultivated plants (Cambridge University Press).
Cano-Rodriguez, D., Gjaltema, R.A.F., Jilderda, L.J., Jellema, P., Dokter-Fokkens, J., Ruiters, M.H.J., and Rots, M.G. (2016). Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. Nat. Commun. **7**: 1–11.
Causse, M. (2002). QTL analysis of fruit quality in fresh market tomato: a few chromosome regions control the variation of sensory and instrumental traits. J. Exp. Bot. **53**: 2089–2098.
Causse, M., Friguet, C., Coiret, C., LéPicier, M., Navez, B., Lee, M., Holthuysen, N., Sinesio, F., Moneta, E., and Grandillo, S. (2010). Consumer Preferences for Fresh Tomato at the European Scale: A Common Segmentation on Taste and Firmness. J. Food Sci. **75**: S531–S541.
Causse, M., Saliba-Colombani, V., Lecomte, L., Duffé, P., Rousselle, P., and Buret, M. (2002). QTL analysis of fruit quality in fresh market tomato: A few chromosome regions control the variation of sensory and instrumental traits. J. Exp. Bot. **53**: 2089–2098.
Chalkley, G.E. and Verrijzer, C.P. (1999). DNA binding site selection by RNA polymerase II TAFs: A TAF(II)250-TAF(II)150 complex recognizes the initiator. EMBO J. **18**: 4835–4845.
Chen, K. and Rajewsky, N. (2007). The evolution of gene regulation by transcription factors and microRNAs. Nat. Rev. Genet. **8**: 93–103.
Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science (80-.). **339**: 819–823.
Cunningham, F.X. and Gantt, E. (1998). Genes and enzymes of carotenoid biosyntehsis in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **49**: 557–583.
Das, S. and Bansal, M. (2019). Variation of gene expression in plants is influenced by gene architecture and structural properties of promoters. PLoS One **14**: 1–31.
Deikman, J. (1997). Molecular mechanisms of ethylene regulation of gene transcription. Physiol. Plant. **100**: 561–566.
Deng, W. and Roberts, S.G.E. (2005). A core promoter element downstream of the TATA box that is recognized by TFIIB. Genes Dev. **19**: 2418–2423.
Deriano, L. and Roth, D.B. (2013). Modernizing the nonhomologous end-joining repertoire: Alternative and classical NHEJ share the stage. Annu. Rev. Genet. **47**: 433–455.
Doebley, J.F., Gaut, B.S., and Smith, B.D. (2006). The Molecular Genetics of Crop Domestication. Cell **127**: 1309–1321.
Domcke, S., Bardet, A.F., Adrian Ginno, P., Hartl, D., Burger, L., and Schübeler, D. (2015). Competition between DNA methylation and transcription factors determines binding of NRF1. Nature **528**: 575–579.
Dror, I., Rohs, R., and Mandel-Gutfreund, Y. (2016). How motif environment influences transcription factor search dynamics: Finding a needle in a haystack. BioEssays **38**: 605–612.
Eshed, Y. and Zamir, D. (1995). An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield- associated QTL. Genetics **141**: 1147–1162.
Etienne, A., Génard, M., Lobit, P., Mbeguié-A-Mbéguié, D., and Bugaud, C. (2013). What controls fleshy fruit

acidity? A review of malate and citrate accumulation in fruit cells. J. Exp. Bot. **64**: 1451–1469.

Fernqvist, F. and Hunter, E. (2012). Who's to blame for tasteless tomatoes? the effect of tomato chilling on consumers' taste perceptions. Eur. J. Hortic. Sci. **77**: 193–198.

Frary, A. and Doğanlar, S. (2003). Comparative genetics of crop plant domestication and evolution. Turkish J. Agric. For. **27**: 59–69.

Frary, A., Nesbitt, T.C., Frary, A., Grandillo, S., Van Der Knaap, E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K.B., and Tanksley, S.D. (2000). *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. Science (80-.). **289**: 85–88.

Fridman, E., Carrari, F., Liu, Y.-S., Fernie, A.R., and Zamir, D. (2004). Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science **305**: 1786–1789.

Fridman, E., Liu, Y.S., Carmel-Goren, L., Gur, A., Shoresh, M., Pleban, T., Eshed, Y., and Zamir, D. (2002). Two tightly linked QTLs modify tomato sugar content via different physiological pathways. Mol. Genet. Genomics **266**: 821–826.

Fridman, E., Pleban, T., and Zamir, D. (2000). A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc. Natl. Acad. Sci. U. S. A. **97**: 4718–4723.

Fulton, T.M., Bucheli, P., Voirol, E., López, J., Pétiard, V., and Tanksley, S.D. (2002). Quantitative trait loci (QTL) affecting sugars, organic acids and other biochemical properties possibly contributing to flavour, identified in four advanced backcross populations of tomato. Euphytica **127**: 163–177.

Gaszner, M. and Felsenfeld, G. (2006). Insulators: Exploiting transcriptional and epigenetic mechanisms. Nat. Rev. Genet. **7**: 703–713.

Gebhardt, C. (2016). The historical role of species from the *Solanaceae* plant family in genetic research. Theor. Appl. Genet. **129**: 2281–2294.

Gillaspy, G., Ben-David, H., and Gruissem, W. (1993). Fruits: A developmental perspective. Plant Cell **5**: 1439–1451.

Giovannoni, J.J., DellaPenna, D., Bennett, A.B., and Fischer, R.L. (1989). Expression of a chimeric polygalacturonase gene in transgenic *rin* (ripening inhibitor) tomato fruit results in polyuronide degradation but not fruit softening. Plant Cell **1**: 53.

Gisselbrecht, S.S., Palagi, A., Kurland, J. V., Rogers, J.M., Ozadam, H., Zhan, Y., Dekker, J., and Bulyk, M.L. (2020). Transcriptional Silencers in Drosophila Serve a Dual Role as Transcriptional Enhancers in Alternate Cellular Contexts. Mol. Cell **77**: 324-337.e8.

Gordân, R., Shen, N., Dror, I., Zhou, T., Horton, J., Rohs, R., and Bulyk, M.L. (2013). Genomic Regions Flanking E-Box Binding Sites Influence DNA Binding Specificity of bHLH Transcription Factors through DNA Shape. Cell Rep. **3**: 1093–1104.

Gray, J.E., Picton, S., Giovannoni, J.J., and Grierson, D. (1994). The use of transgenic and naturally occurring mutants to understand and manipulate tomato fruit ripening. Plant, Cell Environ. **17**: 557–571.

Gu, W., Lee, H.C., Chaves, D., Youngman, E.M., Pazour, G.J., Conte, D., and Mello, C.C. (2012). CapSeq and CIP-TAP identify pol ii start sites and reveal capped small RNAs as C. elegans piRNA precursors. Cell **151**: 1488–1500.

Hampsey, M. (1998). Molecular Genetics of the RNA Polymerase II General Transcriptional Machinery. Microbiol. Mol. Biol. Rev. **62**: 465–503.

Harker, F.R., Maindonald, J., Murray, S.H., Gunson, F.A., Hallett, I.C., and Walker, S.B. (2002). Sensory interpretation of instrumental measurements 1: Texture of apple fruit. Postharvest Biol. Technol. **24**: 225–239.

He, G., Elling, A.A., and Deng, X.W. (2011). The epigenome and plant development. Annu. Rev. Plant Biol. **62**: 411–435.

Hendrich, B. and Tweedie, S. (2003). The methyl-CpG binding domain and the evolving role of DNA methylation in animals. Trends Genet. **19**: 269–277.

Holliday, R. and Pugh, J. (1975). DNA modification mechanisms and gene activity during development. Science (80-.). **187**: 226–232.

Hosmani, P.S. et al. (2019). An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. bioRxiv **2012**: 767764.

Hsieh, C.L. (1994). Dependence of transcriptional repression on CpG methylation density. Mol. Cell. Biol. **14**: 5487–5494.

Hu, J., Israeli, A., Ori, N., and Sun, T.P. (2018). The interaction between DELLA and ARF/IAA mediates crosstalk between gibberellin and auxin signaling to control fruit initiation in tomato. Plant Cell **30**: 1710–1728.

Huang, Z. and van der Knaap, E. (2011). Tomato fruit weight 11.3 maps close to fasciated on the bottom of chromosome 11. Theor. Appl. Genet. **123**: 465–474.

Hufford, M.B. et al. (2012). Comparative population genomics of maize domestication and improvement. Nat. Genet. **44**: 808–811.

Hyodo, H., Terao, A., Furukawa, J., Sakamoto, N., Yurimoto, H., Satoh, S., and Iwai, H. (2013). Tissue specific localization of pectin-Ca2+ cross-linkages and pectin methyl-esterification during fruit ripening in

tomato (Solanum lycopersicum). PLoS One **8**: e78949.
Jenkins, J.A. (1948). The origin of the cultivated tomato. Econ. Bot. **2**: 379–392.
Jiang, C. and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: Advances through genomics. Nat. Rev. Genet. **10**: 161–172.
Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009a). H3.3/H2A.Z double variant-containing nucleosomes mark "nucleosome-free regions" of active promoters and other regulatory regions. Nat. Genet. **41**: 941–945.
Jin, Y., Ni, D.-A., and Ruan, Y.-L. (2009b). Posttranslational elevation of cell wall invertase activity by silencing its inhibitor in tomato delays leaf senescence and increases seed weight and fruit hexose level. Plant Cell **21**: 2072–2089.
Jinap, S. and Hajeb, P. (2010). Glutamate. Its applications in food and contribution to health. Appetite **55**: 1–10.
Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science (80-.). **337**: 816–821.
Jolma, A. et al. (2013). DNA-binding specificities of human transcription factors. Cell **152**: 327–339.
Kagey, M.H. et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature **467**: 430–435.
Karlova, R., Chapman, N., David, K., Angenent, G.C., Seymour, G.B., and De Maagd, R.A. (2014). Transcriptional control of fleshy fruit development and ripening. J. Exp. Bot. **65**: 4527–4541.
Karlova, R., Rosin, F.M., Busscher-Lange, J., Parapunova, V., Do, P.T., Fernie, A.R., Fraser, P.D., Baxter, C., Angenent, G.C., and de Maagd, R.A. (2011). Transcriptome and metabolite profiling show that APETALA2a is a major regulator of tomato fruit ripening. Plant Cell **23**: 923–41.
Kaufmann, K. and Mueller-Roeber, B. (2018). Plant Gene Regulatory Networks Methods and Protocols Methods in Molecular Biology K. Kaufmann and B. Mueller-Roeber, eds (Humana Press).
Kaufmann, K., Pajoro, A., and Angenent, G.C. (2010). Regulation of transcription in plants: Mechanisms controlling developmental switches. Nat. Rev. Genet. **11**: 830–842.
Kazeniac, S.J. and Hall, R.M. (1970). Flavour chemistry of tomato volatiles. J. Food Sci. **35**: 519–530.
Kim, J.L., Nikolov, D.B., and Burley, S.K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature **365**: 520–527.
Kimbara, J., Ohyama, A., Chikano, H., Ito, H., Hosoi, K., Negoro, S., Miyatake, K., Yamaguchi, H., Nunome, T., Fukuoka, H., and Hayashi, T. (2018). QTL mapping of fruit nutritional and flavour components in tomato (*Solanum lycopersicum*) using genome-wide SSR markers and recombinant inbred lines (RILs) from an intra-specific cross. Euphytica **214**: 1–12.
Kimura, S. and Sinha, N. (2008). Tomato (*Solanum lycopersicum*): A model fruit-bearing crop. Cold Spring Harb. Protoc. **3**: 1–10.
Kitagawa, M., Ito, H., Shiina, T., Nakamura, N., Inakuma, T., Kasumi, T., Ishiguro, Y., Yabe, K., and Ito, Y. (2005). Characterization of tomato fruit ripening and analysis of gene expression in F_1 hybrids of the *ripening inhibitor* (*rin*) mutant. Physiol. Plant. **123**: 331–338.
Klee, H.J. and Tieman, D.M. (2018). The genetics of fruit flavour preferences. Nat. Rev. Genet. **19**: 347–356.
Klose, R.J. and Bird, A.P. (2006). Genomic DNA methylation: The mark and its mediators. Trends Biochem. Sci. **31**: 89–97.
van der Knaap, E., Chakrabarti, M., Chu, Y.H., Clevenger, J.P., Illa-Berenguer, E., Huang, Z., Keyhaninejad, N., Mu, Q., Sun, L., Wang, Y., and Wu, S. (2014). What lies beyond the eye: The molecular mechanisms regulating tomato fruit weight and shape. Front. Plant Sci. **5**: 1–13.
Knapp, S. and Peralta, I.E. (2016). The Tomato (*Solanum lycopersicum* L., Solanaceae) and Its Botanical Relatives (Springer, Berlin, Heidelberg).
Knoll, A., Fauser, F., and Puchta, H. (2014). DNA recombination in somatic plant cells: Mechanisms and evolutionary consequences. Chromosom. Res. **22**: 191–201.
Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., and Ebright, R.H. (1998). New core promoter element in RNA polymerase II-dependent transcription: Sequence-specific DNA binding by transcription factor IIB. Genes Dev. **12**: 34–44.
Lanahan, M.B., Yen Hsiao Ching, Giovannoni, J.J., and Klee, H.J. (1994). The *Never ripe* mutation blocks ethylene perception in tomato. Plant Cell **6**: 521–530.
Lashbrook, C.C., Gonzalez-Bosch, C., and Bennett, A.B. (1994). Two divergent endo-b-1,4-glucanase genes exhibit overlapping expression in ripening fruit and abscising flowers. Plant Cell **6**: 1485.
Latchman, D.S. (1997). Transcription factors: An overview. Int. J. Biochem. Cell Biol. **29**: 1305–1312.
Lazarovici, A., Zhou, T., Shafer, A., MacHado, A.C.D., Riley, T.R., Sandstrom, R., Sabo, P.J., Lu, Y., Rohs, R., Stamatoyannopoulos, J.A., and Bussemaker, H.J. (2013). Probing DNA shape and methylation state on a genomic scale with DNase i. Proc. Natl. Acad. Sci. U. S. A. **110**: 6376–6381.
Lelievre, J.M., Latche, A., Jones, B., Bouzayen, M., and Pech, J.-C. (1997). Ethylene and fuit ripening. Physiol. Plant. **101**: 727–739.
Lewinsohn, E., Sitrit, Y., Bar, E., Azulay, Y., Meir, A., Zamir, D., and Tadmor, Y. (2005). Carotenoid

pigmentation affects the volatile composition of tomato and watermelon fruits, as revealed by comparative genetic analyses. J. Agric. Food Chem. **53**: 3142–3148.

Li, J.-F., Norville, J.E., Aach, J., McCormack, M., Zhang, D., Bush, J., Church, G.M., and Sheen, J. (2013). Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nat. Biotechnol. **31**: 688–91.

Lilac Pnueli, Lea Carmel-Goren, Dana Hareven, Tamar Gutfinger, John Alvarez, Martin Ganal, Daniel Zamir, and Eliezer Lifschitz (1998). The SELF-PRUNING gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of *CEN* and *TFL1*. Development **125**: 1979–1989.

Lin, T. et al. (2014). Genomic analyses provide insights into the history of tomato breeding. Nat. Genet. **46**: 1220–1226.

Lu, Z., Marand, A.P., Ricci, W.A., Ethridge, C.L., Zhang, X., and Schmitz, R.J. (2019). The prevalence, evolution and chromatin signatures of plant regulatory elements. Nat. Plants **5**: 1250–1259.

Luscombe, N.M., Austin, S.E., Berman, H.M., and Thornton, J.M. (2000). An overview of the structures of protein-DNA complexes. Genome Biol. **1**: 1–37.

Magnani, L., Eeckhoute, J., and Lupien, M. (2011). Pioneer factors: Directing transcriptional regulators within the chromatin environment. Trends Genet. **27**: 465–474.

Mahajan, P. V., Caleb, O.J., Singh, Z., Watkins, C.B., and Geyer, M. (2014). Postharvest treatments of fresh produce. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. **372**: 20130309.

Maher, K.A. et al. (2018). Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules. Plant Cell **30**: 15–36.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013a). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. **31**: 833–838.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013b). RNA-guided human genome engineering via Cas9. Science (80-.). **339**: 823–826.

Mathelier, A., Xin, B., Chiu, T.P., Yang, L., Rohs, R., and Wasserman, W.W. (2016). DNA Shape Features Improve Transcription Factor Binding Site Predictions In Vivo. Cell Syst. **3**: 278-286.e4.

Maul, F., Sargent, S.A., Sims, C.A., Baldwin, E.A., Balaban, M.O., and Huber, D.J. (2000). Tomato flavour and aroma quality as affected by storage temperature. J. Food Sci. **65**: 1228–1237.

Maurano, M.T., Wang, H., John, S., Shafer, A., Canfield, T., Lee, K., and Stamatoyannopoulos, J.A. (2015). Role of DNA Methylation in Modulating Transcription Factor Occupancy. Cell Rep. **12**: 1184–1195.

Mavrich, T.N. et al. (2008). Nucleosome organization in the *Drosophila* genome. Nature **453**: 358–362.

Mejia-Guerra, M.K., Pomeranz, M., Morohashi, K., and Grotewold, E. (2012). From plant gene regulatory grids to network dynamics. Biochim. Biophys. Acta - Gene Regul. Mech. **1819**: 454–465.

Meyer, C.A., Tang, Q., and Liu, X.S. (2012). Minireview: Applications of next-generation sequencing on studies of nuclear receptor regulation and function. Mol. Endocrinol. **26**: 1651–1659.

Meyer, R.S. and Purugganan, M.D. (2013). Evolution of crop species: Genetics of domestication and diversification. Nat. Rev. Genet. **14**: 840–852.

Miller, J.C. and Tanksley, S.D. (1990). RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor. Appl. Genet. **80**: 437–448.

Mueller, B., Mieczkowski, J., Kundu, S., Wang, P., Sadreyev, R., Tolstorukov, M.Y., and Kingston, R.E. (2017). Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction. Genes Dev. **31**: 451–462.

Narlikar, G.J., Fan, H.Y., and Kingston, R.E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell **108**: 475–487.

Ni, T., Corcoran, D.L., Rach, E.A., Song, S., Spana, E.P., Gao, Y., Ohler, U., and Zhu, J. (2010). A paired-end sequencing strategy to map the complex landscape of transcription initiation. Nat. Methods **7**: 521–527.

O'Malley, R.C., Huang, S.S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell **165**: 1280–1292.

Oltman, A.E., Jervis, S.M., and Drake, M.A. (2014). Consumer attitudes and preferences for fresh market tomatoes. J. Food Sci. **79**: S2091–S2097.

Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. Genes Dev. **10**: 2657–2683.

van Overbeek, M. et al. (2016). DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. Mol. Cell **63**: 633–646.

Patikoglou, G.A., Kim, J.L., Sun, L., Yang, S.H., Kodadek, T., and Burley, S.K. (1999). TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. Genes Dev. **13**: 3217–3230.

Peralta, I.E., Knapp, S., and Spooner, D.M. (2005). New species of wild tomatoes (*Solanum Section Lycopersicon: Solanaceae*) from Northern Peru. Syst. Bot. **30**: 424–434.

Peralta, I.E., Spooner, D.M., and Knapp, S. (2008). Taxonomy of wild tomatoes and their relatives (*Solanum sect. Lycopersicoides, sect. Juglandifolia, sect. Lycopersicon; Solanaceae*). Syst. Bot. Monogr. **84**: 1–

186.
Petreikov, M., Shen, S., Yeselson, Y., Levin, I., Bar, M., and Schaffer, A.A. (2006). Temporally extended gene expression of the *ADP-Glc pyrophosphorylase large subunit (AgpL1)* leads to increased enzyme activity in developing tomato fruit. Planta **224**: 1465–1479.
Petró-Turza, M. (1986). Flavour of tomato and tomato products. Food Rev. Int. **2**: 309–351.
Powell, A.L.T. et al. (2012). Uniform ripening encodes a Golden 2-like transcription factor regulating tomato fruit chloroplast development. Science (80-.). **336**: 1711–1715.
Raffo, A., Nicoli, S., Nardo, N., Baiamonte, I., Daloise, A., and Paoletti, F. (2012). Impact of different distribution scenarios and recommended storage conditions on flavour related quality attributes in ripening fresh tomatoes. J. Agric. Food Chem. **60**: 10445–10455.
Rambla, J.L., Tikunov, Y.M., Monforte, A.J., Bovy, A.G., and Granell, A. (2014). The expanded tomato fruit volatile landscape. J. Exp. Bot. **65**: 4613–4623.
Razifard, H., Ramos, A., Della Valle, A.L., Bodary, C., Goetz, E., Manser, E.J., Li, X., Zhang, L., Visa, S., Tieman, D., van der Knaap, E., and Caicedo, A.L. (2020). Genomic Evidence for Complex Domestication History of the Cultivated Tomato in Latin America. Mol. Biol. Evol. **37**: 1118–1132.
Renard, C.M.G.C., Ginies, C., Gouble, B., Bureau, S., and Causse, M. (2013). Home conservation strategies for tomato (*Solanum lycopersicum*): Storage temperature vs. duration-Is there a compromise for better aroma preservation? Food Chem. **139**: 825–836.
Ricci, W.A. et al. (2019). Widespread long-range cis-regulatory elements in the maize genome. Nat. Plants **5**: 1237–1249.
Riggs, A.D. (1975). Inactivation, differentiation, and DNA methylation. Cytogenet. Genome Res. **14**: 9–25.
Robinson, R.W. and Tomes, M.L. (1968). Ripening inhibitor: a gene with multiple effect on ripening. Tomato Genet. Coop. **18**: 36-37.
Rodríguez-Leal, D., Lemmon, Z.H., Man, J., Bartlett, M.E., and Lippman, Z.B. (2017). Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing. Cell **171**: 470-480.e8.
Ron, M. et al. (2014). Hairy root transformation using Agrobacterium rhizogenes as a tool for exploring cell type-specific gene expression and function using tomato as a model. Plant Physiol. **166**: 455–69.
Roohanitaziani, R. (2019). Genetic analysis of fruit quality in tomato (Wageningen University, Doctoral dissertation).
Roohanitaziani, R., de Maagd, R.A., Lammers, M., Molthoff, J., Meijer-Dekens, F., Kaauwen, M.P.W. van, Finkers, R., Tikunov, Y., Visser, R.G.F., and Bovy, A.G. (2020). Exploration of a Resequenced Tomato Core Collection for Phenotypic and Genotypic Variation in Plant Growth and Fruit Quality Traits. Genes (Basel). **11**: 1278.
Sanchez-Bel, P., Egea, I., Sanchez-Ballesta, M.T., Sevillano, L., Del Carmen Bolarin, M., and Flores, F.B. (2012). Proteome changes in tomato fruits prior to visible symptoms of chilling injury are linked to defensive mechanisms, uncoupling of photosynthetic processes and protein degradation machinery. Plant Cell Physiol. **53**: 470–484.
Sandarani, M., Dasanayaka, D., and Jayasinghe, C. (2018). Strategies Used to Prolong the Shelf Life of Fresh Commodities. J. Agric. Sci. Food Res. **9**: 1–6.
Sauvage, C., Segura, V., Bauchet, G., Stevens, R., Do, P.T., Nikoloski, Z., Fernie, A.R., and Causse, M. (2014). Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. Plant Physiol. **165**: 1120–1132.
Sayou, C. et al. (2016). A SAM oligomerization domain shapes the genomic binding landscape of the LEAFY transcription factor. Nat. Commun. **7**: 1–12.
Schaffer, a. a. and Petreikov, M. (1997). Sucrose-to-Starch Metabolism in Tomato Fruit Undergoing Transient Starch Accumulation. Plant Physiol. **113**: 739–746.
Schaffer, A.A., Levin, I., Oguz, I., Petreikov, M., Cincarevsky, F., Yeselson, Y., Shen, S., Gilboa, N., and Bar, M. (2000). ADPglucose pyrophosphorylase activity and starch accumulation in immature tomato fruit: The effect of a Lycopersicon hirsutum-derived introgression encoding for the large subunit. Plant Sci. **152**: 135–144.
Schiml, S., Fauser, F., and Puchta, H. (2014). The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in Arabidopsis resulting in heritable progeny. Plant J. **80**: 1139–1150.
Schouten, H.J., Tikunov, Y., Verkerke, W., Finkers, R., Bovy, A., Bai, Y., and Visser, R.G.F. (2019). Breeding Has Increased the Diversity of Cultivated Tomato in The Netherlands. Front. Plant Sci. **10**: 1606.
Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., Qiu, J.L., and Gao, C. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. Nat. Biotechnol. **31**: 686–688.
Shi, X., Shou, J., Mehryar, M.M., Li, J., Wang, L., Zhang, M., Huang, H., Sun, X., and Wu, Q. (2019). Cas9 has no exonuclease activity resulting in staggered cleavage with overhangs and predictable di- and tri-nucleotide CRISPR insertions without template donor. Cell Discov. **5**: 53.
Shneour, E.A. and Zabin, I. (1959). The biosynthesis of lycopene in tomato homogenates. J. Biol. Chem. **234**:

770–773.
Shukla, V.K. et al. (2009). Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature **459**: 437–441.
Smale, S.T. and Baltimore, D. (1989). The "initiator" as a transcription control element. Cell **57**: 103–113.
Somssich, M., Je, B. Il, Simon, R., and Jackson, D. (2016). CLAVATA-WUSCHEL signaling in the shoot meristem. Dev. **143**: 3238–3248.
Sorrequieta, A., Ferraro, G., Boggio, S.B., and Valle, E.M. (2010). Free amino acid production during tomato fruit ripening: A focus on L-glutamate. Amino Acids **38**: 1523–1532.
Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and Zaret, K.S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell **161**: 555–568.
Spitz, F. and Furlong, E.E.M. (2012). Transcription factors: From enhancer binding to developmental control. Nat. Rev. Genet. **13**: 613–626.
Suzuki, M.M. and Bird, A. (2008). DNA methylation landscapes: Provocative insights from epigenomics. Nat. Rev. Genet. **9**: 465–476.
Swinnen, G., Goossens, A., and Pauwels, L. (2016). Lessons from Domestication: Targeting Cis-Regulatory Elements for Crop Improvement. Trends Plant Sci. **21**: 506–515.
Symington, L.S. and Gautier, J. (2011). Double-strand break end resection and repair pathway choice. Annu. Rev. Genet. **45**: 247–271.
Tadmor, Y., Fridman, E., Gur, A., Larkov, O., Lastochkin, E., Ravid, U., Zamir, D., and Lewinsohn, E. (2002). Identification of malodorous, a wild species allele affecting tomato aroma that was selected against during domestication. J. Agric. Food Chem. **50**: 2005–2009.
Tan, J. et al. (2020). Efficient CRISPR/Cas9-based plant genomic fragment deletions by microhomology-mediated end joining. Plant Biotechnol. J.: 1–3.
Tan, M. et al. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell **146**: 1016–1028.
Tanay, A. (2006). Extensive low-affinity transcriptional interactions in the yeast genome. Genome Res. **16**: 962–972.
Tanksley, S.D., Grandillo, S., Fulton Zamir, T.D., Eshed, Y., Petiard -J Lopez Beck-Bunn, V.T., Zamir, D., Eshed, T., and Lopez Hispareco, J.S. (1996). Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium* (Springer-Verlag).
Tao, Z., Shen, L., Liu, C., Liu, L., Yan, Y., and Yu, H. (2012). Genome-wide identification of SOC1 and SVP targets during the floral transition in Arabidopsis. Plant J. **70**: 549–561.
Terhorst, P. (2006). The scaling of the Dutch vegetables-under-glass cluster: Sweet peppers, tomatoes and cucumbers. Tijdschr. voor Econ. en Soc. Geogr. **97**: 434–442.
The Tomato Genome Consortium et al. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**: 635–41.
Thompson, A.J., Tor, M., Barry, C.S., Vrebalov, J., Orfila, C., Jarvis, M.C., Giovannoni, J.J., Grierson, D., and Seymour, G.B. (1999). Molecular and genetic characterization of a novel pleiotropic tomato-ripening mutant. Plant Physiol. **120**: 383–389.
Tieman, D. et al. (2012). The chemical interactions underlying tomato flavour preferences. Curr. Biol. **22**: 1035–1039.
Tieman, D. et al. (2017). A chemical genetic roadmap to improved tomato flavour. Science (80-.). **355**: 391–394.
Tigchelaar, E., Tomes, M., Kerr, E., and Barman, R. (1973). A new fruit ripening mutant, non-ripening (nor). Rep Tomato Genet Coop **23**: 33–34.
Tippin, D.B., Ramakrishnan, B., and Sundaralingam, M. (1997). Methylation of the Z-DNA decamer d(GC)5 potentiates the formation of A-DNA: Crystal structure of d(Gm5CGm5CGCGCGC). J. Mol. Biol. **270**: 247–258.
Uluisik, S. et al. (2016). Genetic improvement of tomato by targeted control of fruit softening. Nat. Biotechnol. **34**: 950–952.
Varga, A. (1976). Roles of Seeds and Auxins in Tomato Fruit Growth. Zeitschrift für Pflanzenphysiologie **80**: 95–104.
Vogel, J.T., Tieman, D.M., Sims, C.A., Odabasi, A.Z., Clark, D.G., and Klee, H.J. (2010). Carotenoid content impacts flavour acceptability in tomato (*Solanum lycopersicum*). J. Sci. Food Agric. **90**: 2233–2240.
Völkel, S., Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., and Suske, G. (2015). Zinc Finger Independent Genome-Wide Binding of Sp2 Potentiates Recruitment of Histone-Fold Protein Nf-y Distinguishing It from Sp1 and Sp3. PLOS Genet. **11**: e1005102.
Voytas, D.F. and Gao, C. (2014). Precision genome engineering and agriculture: opportunities and regulatory challenges. PLoS Biol. **12**: e1001877.
Wade, P.A. (2001). Methyl CpG-binding proteins and transcriptional repression. BioEssays **23**: 1131–1137.
Wang, D. and Seymour, G.B. (2017). Tomato Flavour: Lost and Found? Mol. Plant **10**: 782–784.
Wang, L., Koppitch, K., Cutting, A., Dong, P., Kudtarkar, P., Zheng, J., Cameron, R.A., and Davidson, E.H.

(2019a). Developmental effector gene regulation: Multiplexed strategies for functional analysis. Dev. Biol. **445**: 68–79.

Wang, R., Tavano, E.C. da R., Lammers, M., Martinelli, A.P., Angenent, G.C., and de Maagd, R.A. (2019b). Re-evaluation of transcription factor function in tomato fruit development and ripening with CRISPR/Cas9-mutagenesis. Sci. Rep. **9**: 1696.

Weirauch, M.T. et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell **158**: 1431–1443.

West, A.G., Shore, P., and Sharrocks, A.D. (1997). DNA binding by MADS-box transcription factors: a molecular mechanism for differential DNA bending. Mol. Cell. Biol. **17**: 2876–2887.

Wilczyski, B. and Furlong, E.E.M. (2010). Dynamic CRM occupancy reflects a temporal map of developmental progression. Mol. Syst. Biol. **6**: 383.

Wilkinson, J.Q., Lanahan, M.B., Yen, H.C., Giovannoni, J.J., and Klee, H.J. (1995). An ethylene-inducible component of signal transduction encoded by <I>Never-ripe</i>. Science (80-.). **270**: 1807.

Wittkopp, P.J. and Kalay, G. (2012). Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. Nat. Rev. Genet. **13**: 59–69.

Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J., and Van Der Knaap, E. (2008). A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science (80-.). **319**: 1527–1530.

Xu, C. et al. (2015). A cascade of arabinosyltransferases controls shoot meristem size in tomato. Nat. Genet. **47**: 784–792.

Yamamoto, Y., Ichida, H., Abe, T., Suzuki, Y., Sugano, S., and Obokata, J. (2007). Differentiation of core promoter architecture between plants and mammals revealed by LDSS analysis. Nucleic Acids Res. **35**: 6219–6226.

Yamamoto, Y.Y., Yoshitsugu, T., Sakurai, T., Seki, M., Shinozaki, K., and Obokata, J. (2009). Heterogeneity of *Arabidopsis* core promoters revealed by high-density TSS analysis. Plant J. **60**: 350–362.

Yáñez-Cuna, J.O., Kvon, E.Z., and Stark, A. (2013). Deciphering the transcriptional cis-regulatory code. Trends Genet. **29**: 11–22.

Ye, J., Wang, X., Hu, T. xu, Zhang, F. xia, Wang, B., Li, C. xin, Yang, T. xia, Li, H. xia, Lu, Y. en, Giovannoni, J.J., Zhang, Y., and Ye, Z. (2017). An InDel in the Promoter of *Al-ACTIVATED MALATE TRANSPORTER9* Selected during Tomato Domestication Determines Fruit Malate Contents and Aluminum Tolerance. Plant Cell **29**: tpc.00211.2017.

Yeager, A.F. (1927). Determinate growth in the tomato. J. Hered. **18**: 263–265.

Yilmaz, E. (2001). Chemistry and Evaluation of Tomato Flavour. Turkish J. Agric. For. **25**: 149–155.

Yokotani, N., Nakano, R., Imanishi, S., Nagata, M., Inaba, A., and Kubo, Y. (2009). Ripening-associated ethylene biosynthesis in tomato fruit is autocatalytically and developmentally regulated. J. Exp. Bot. **60**: 3433–3442.

Yu, C.P., Lin, J.J., and Li, W.H. (2016). Positional distribution of transcription factor binding sites in *Arabidopsis thaliana*. Sci. Rep. **6**.

Zanor, M.I. et al. (2009). RNA interference of *LIN5* in tomato confirms its role in controlling brix content, uncovers the influence of sugars on the levels of fruit hormones, and demonstrates the importance of sucrose cleavage for normal fruit development and fertility. Plant Physiol. **150**: 1204–1218.

Zhang, B., Tieman, D.M., Jiao, C., Xu, Y., Chen, K., Fe, Z., Giovannoni, J.J., and Klee, H.J. (2016). Chilling-induced tomato flavour loss is associated with altered volatile synthesis and transient changes in DNA methylation. Proc. Natl. Acad. Sci. **113**: 12580–12585.

Zhang, H., Lang, Z., and Zhu, J.K. (2018). Dynamics and function of DNA methylation in plants. Nat. Rev. Mol. Cell Biol. **19**: 489–506.

Zhang, M., Kimatu, J.N., Xu, K., and Liu, B. (2010). DNA cytosine methylation in plant development. J. Genet. Genomics **37**: 1–12.

Zhao, J., Sauvage, C., Zhao, J., Bitton, F., Bauchet, G., Liu, D., Huang, S., Tieman, D.M., Klee, H.J., and Causse, M. (2019). Meta-analysis of genome-wide association studies provides insights into genetic control of tomato flavour. Nat. Commun. **10**: 1–12.

Zhu, F., Farnung, L., Kaasinen, E., Sahu, B., Yin, Y., Wei, B., Dodonova, S.O., Nitta, K.R., Morgunova, E., Taipale, M., Cramer, P., and Taipale, J. (2018a). The interaction landscape between transcription factors and the nucleosome. Nature **562**: 76–81.

Zhu, G. et al. (2018b). Rewiring of the Fruit Metabolome in Tomato Breeding. Cell **172**: 249-261.e12.

CHAPTER 2

Gene editing targets for increasing sugar content in tomato

Vera Veltkamp[1], Ruud A. de Maagd[2]

[1]Laboratory of Molecular Biology, Wageningen University & Research

[2]Bioscience, Cluster Plant Developmental Systems, Wageningen University & Research

ABSTRACT

Tomato (*Solanum lycopersicum*) is a worldwide high-production and high-value vegetable crop. The tomato flavour is an important trait determined by different factors: sugars, organic acids, minerals, volatiles, and their interactions. Identifying and understanding the role of the genes underlying the production and distribution of these factors is crucial for crop breeding. It is even more critical because of the rapid development of gene editing technologies that enable precise modification of these genes. This review discusses our current knowledge of the genes determining one of the essential factors contributing to tomato flavour: the sugar content. A combination of its synthesis, transport, metabolism, and storage determines the amount of sugar in tomato. Quantitative Trait Loci (QTLs) for sugar content and their underlying genes are outlined. Moreover, we discuss possible new targets for breeding that have emerged recently. Some will need further research and development because not all putative targets have been validated for an overall positive effect (in tomato), especially when negative trade-offs have to be balanced. Combining of QTL trait discovery and molecular analysis of the sugar pathway(s) has yielded a wide array of either proven or promising targets to increase the tomato fruits' sugar content.

INTRODUCTION

Tomato is a popular and economically important fresh or processed vegetable crop (http://www.fao.org/faostat/en/#home)**.** Its flavour requires continuous attention and improvement, as many consumers are not satisfied or seek perfection. The flavour of cheaper fresh-market tomatoes, in particular, has been a significant source of consumer dissatisfaction (Fernqvist and Hunter, 2012; Tieman *et al.*, 2017*a*). Improving the flavour of these tomatoes would meet the demand for affordable and tasty tomatoes. Flavour is clearly an important aspect of quality breeding in tomato, but it is also one of the more challenging ones. A significant difficulty is that flavour is a complex and cumulative trait, breeding is time-consuming and phenotyping expensive, and consumers' demands vary widely (Causse *et al.*, 2010; Tieman *et al.*, 2017*a*). Additionally, improving flavour can have negative trade-offs, such as in yield or shelf-life. Since the 1990s, flavour has been back in focus in response to complaints about watery tomatoes and the growing popularity of heirloom varieties (Terhorst, 2006; Tieman *et al.*, 2017*a*; Bauchet *et al.*, 2017). For instance, major contributions to knowledge on aroma pathways and aroma-QTLs have been made in the last decade (Tadmor *et al.*, 2002; Rambla *et al.*, 2014; Tieman *et al.*, 2017*a*; Martina *et al.*, 2021). Another major contributor to tomato flavour is the sugar content (Causse *et al.*, 2010; Tieman *et al.*, 2017*a*).

Here we review current knowledge and particularly, recent advances in understanding of the genetic determinants of tomato fruit sugar content, with the purpose of

identifying putative targets for improvement that can be implemented in breeding programs by modern breeding tools such as "genome editing". These potential targets may have been identified by genetic analysis, by the identification of genes underlying Quantitative Trait Loci (QTLs) for sugar, or probably will be identified in the future using more advanced genetic analysis such as Multiparent Advanced Generation Intercross (MAGIC) populations, expression Quantitative Trait Loci (eQTL), meta QTL (MQTL) analysis, and Recombinant Inbred Lines (RILs) (Shan *et al.*, 2013; Sander and Joung, 2014; Brooks *et al.*, 2014; Pascual *et al.*, 2015; Pan *et al.*, 2016; Ranjan *et al.*, 2016; Van Eck, 2017; Li *et al.*, 2020), or by the study in genetically modified tomato plants. We limit ourselves herein mostly to genes or their alleles that have been shown to have a positive effect on sugar content when modified in tomato.

SUGAR AND TOMATO FLAVOUR

The tomato flavour is determined by different compounds that influence taste and smell: sugars, organic acids, minerals, volatiles and their interactions (Oltman *et al.*, 2014; Tieman *et al.*, 2017*a*). A set of 33 compounds has been highly correlated with likability by the American consumer (Tieman *et al.*, 2012, 2017*a*)**.** All these compounds warrant attention in the quest for improved flavour. Sugars play a major role in a satisfying sour-sweetness flavour of tomato and are the focus of this review (Baldwin *et al.*, 2000; Tieman *et al.*, 2012, 2017*a*). Glucose and fructose play the major role in tomato sweetness perception. They are formed by hydrolysis of the disaccharide sucrose, of which usually trace amounts are present in cultivated tomato. Glucose and fructose are mostly present in equimolar amounts, with some varieties having slightly higher fructose content (Petró-Turza, 1986). The combined sugars constitute approximately half of the dry fruit weight (Winsor, 1966). The perceived sweetness of sugars on a relative scale is 100 for sucrose (set as reference), 50-80 for glucose and 110-180 for fructose (Tunaley et al., 1987)

The final sugar content in fruit is determined by synthesis, transport from the site of synthesis (source) to the fruit (sink), metabolism, and storage. Sugar is produced in the source tissues, such as leaves or immature green fruit, by photosynthesis and translocated as sucrose via mass transport through the phloem to sink tissues. In tomato, sucrose is the transported sugar. In other species, this may be raffinose and stachyose (Cucurbits, like melon and cucumber), or sugar alcohols like sorbitol (apple, pear) and mannitol (olive) (Yamaki, 2010). In the fruit, cell-to-cell transport achieves the unloading of sucrose with the help of sugar transporters and invertases (Yadav *et al.*, 2015). Sugars are stored in the fruit parenchyma-cell vacuole as hexoses or in plastids as starch (Stein and Granot, 2018). Accumulation in vacuoles creates osmotic pressure leading to water influx, which is believed to provide turgor pressure leading to cell enlargement and fruit growth.

The amount of sugar can be precisely measured with chromatography systems, yet in practice, a faster and simpler measurement of soluble solid contents (°Brix) by refractometry is used as a proxy for sugar content. One-degree Brix (°Brix) equals 1 % w/w sucrose in solution. °Brix of cultivated tomatoes varies from 3 to 10, where fruits weighing more than 15 grams do not go higher than approximately 6 °Brix (Tieman *et al.*, 2017*a*; Roohanitaziani, 2019). This makes tomatoes relatively unsweet compared to other fruits and vegetables like cabbage (5.5 - 6.5), sweet pepper (6.4-8.5), carrot (8.6-12.3), melon (>10), cherry (12-17) or pineapple (>14) (Guthrie *et al.*, 1998; Lu, 2001; Gong *et al.*, 2010; Rashidi, 2011; Ferreira *et al.*, 2012). °Brix is not a measure of the glucose/fructose ratio (Zanor *et al.*, 2009*a*). As fructose is perceived as sweeter, this ratio can be essential for a grower or breeder for determining the final flavour profile. Thus, (high-throughput and) precise analysis of sugars as a standard in breeding for quality is preferable over Brix.

BREEDING FOR BRIX

Current breeding goals increasingly include quality traits (Schouten *et al.*, 2019; Bakir *et al.*, 2020). If it is possible to identify and introgress or newly create superior alleles for quality aspects such as sugar content, these could be used in breeding programs, and this will result in better tasting varieties.

Starting with traditional breeding by consecutive crossing, a significant effort was made in the 1990's to increase genetic variation in cultivated tomato. A series of F2 populations (progeny of a cross between parental lines differing in a trait), introgression lines, recombinant inbred populations and (advanced) backcross populations lines were made with a wild tomato relative as one of the parents. From these populations, 56 chromosomal regions together harbouring 95 QTLs for sugar content related traits (°Brix, fructose, glucose) were identified. Twenty-eight of these regions were linked to QTLs in multiple populations and may represent the same QTL (Fulton *et al.*, 2002; Labate *et al.*, 2007). In most (>85%) of the cases, the wild allele caused an increase in sugar content, emphasising the potential for improvement. Similar results were obtained in a recent QTL analysis from a population backcrossed from *S. pimpinellifolium* (Çolak *et al.*, 2020). From a (meta-) Genome-Wide Association Studies (GWAS), which confirmed already known QTL's as well as identified new ones, it has become increasingly clear that domestication and selection have resulted in a consistent loss of favourable (wild) alleles (Zhao *et al.*, 2019). Other alleles may never have been part of the tomato lineage as the wild relatives harbouring them were not the modern tomato's ancestor. Throughout this review, we will highlight QTLs, genes and other traits that have been associated with an increased fruit sugar content (**Table 1**). QTL names often start with "Brix" or "SSC" (Soluble Solid Content), underlining the contrasting phenotype.

So far, for only a handful of sugar QTLs the underlying genes have been identified (Tanksley *et al.*, 1996; Lin *et al.*, 2014). Recent developments in high throughput sequencing and marker analysis have led to more accurate locus positions, such as by GWAS (Sauvage *et al.*, 2014; Gao *et al.*, 2019; Zhao *et al.*, 2019). Apart from previously identified QTLs *LIN5* and *SSC11.1*, this GWAS delivered several attractive new candidates such as *α-Fuc'ase S1-1* (Solyc03g006980), *Glucosyltransferase* (Solyc05g053400), *Fatty acid elongase 3-ketoacyl-CoA synthase* (Solyc05g009280) and *Glyceraldehyde-3-phosphate dehydrogenase* (Solyc10g005510) (Tieman *et al.*, 2017*a*; Zhao *et al.*, 2019). For most of these candidate genes, modern cultivars, heirlooms, and transitional accessions lost the favourable allele. However, except for *LIN5* and *SSC11*, they seem to have not been lost by domestication or improvement sweeps, but rather reflecting a lack of interest in selection for sugar content (Tieman *et al.*, 2017*a*; Zhao *et al.*, 2019).

Table 1: Overview of QTLs, mutants, genes, and other variations affecting sugar content.

Gene nr	Gene	QTL	Origin	Effect	References
			Heirloom cultivars with rounder leaves	Higher SSC and yield	Rowland et al., 2020
	AtORANGE		Ectopic expression	Higher SSC; LIN5 and *LIN6* expression up, *INVINH1*, expression down; early chromoplast formation	Yazdani et al., 2019
	MdSWEET17		Ectopic expression	Increased fructose content	Lu et al., 2019a
	MdHT2.2		Ectopic expression	Higher SSC, fruit size and increased *LIN5* activity	Wang et al., 2019
	MdERDL6-1		Ectopic expression	Higher SSC	Zhu et al., 2021
	PbSUT2		Ectopic expression	Increased sucrose	Wang et al., 2016
	ZmSPS		Ectopic expression	Higher sucrose unloading in fruit; increased SUS activity	Nguyen Quoc et al., 1999
		IL8-3	*S. pennellii* allele	Higher SSC; high hexose content; increased *AGPL1, LIN6* and *TOMSSF* activity or expression	Ikeda et al., 2013, 2016
Solyc01g100510	*TKN4*	uniform grey-green (*ug*)	wild type allele (*UG*)	Higer SSC; increased chlorophyll levels; non-uniform fruit appearance	Bohk and Scott, 1945; Nadakuduti et al., 2014
Solyc01g109790	*AGPL1*	*AGPL1*^{H}	*S. habrochaites* allele	Increased starch accumulation in young fruit; increased expression; increased stability of AGPase;	Schaffer et al., 2000
Solyc02g081120	*TKN2*	Curl (Cu)	Gain of function	Increased chlorophyll levels in the entire fruit	Parnis et al., 1997
Solyc02g088180	*SlORE1*		RNAi	Higher SSC; delayed leaf senescence; increased yield	Lira et al., 2017
Solyc03g083910	*TIV1 (VIN)*	*sucrose accumulator (sucr)*	Retrotransposon insert in promoter	Increased hexose; expression of TIV1 in developing fruit.	Moy et al., 2007

Table 1 continued

Gene nr	Gene	QTL†	Origin	Effect	References
Solyc04g 005610	*SlNAP2*		RNAi	Higher °Brix; delayed leaf senescence; increased fruit number	Ma et al., 2018
Solyc04g 064610	SlSWEET1a	Fructose to glucose ratio (Fgr)	*S. habrochaites* allele	Increasing fructose-to-glucose ratio in young leaves and flowers; higher expression	Levin et al., 2000; Shammai et al., 2018
Solyc07g 055920	*SlTAGL1*	green striped (gs)	Methylated isoform of *TAGL1* promoter	Higher °Brix; striped fruit	Liu et al., 2020
Solyc08g 065420	SlBEL4		RNAi	Higher SSC; increased chlorophyll accumulation and chloroplast formation	Yan et al., 2020
Solyc08g 065790	*SlVPE1*		RNAi	Higher SSC; increased *TIV1* activity.	Ariizumi et al., 2011
Solyc09g 010080	LIN5	Brix9-2-5	S. pennelli allele	Higher °Brix; increased activity	Fridman et al., 2002; Baxter et al., 2005
Solyc09g 011290	*SlVIF*		RNAi	Highly increased hexoses; delayed ripening; increased *TIV1* activity.	Qin et al., 2016
Solyc10g 008160	*SlGLK2*	uniform (u)	ectopic expression or wt allele (U)	Higher °Brix; increased chlorophyll levels or green shoulder	Powell et al., 2012; Nguyen et al., 2014
Solyc10g 079050	*SlbHLH95*		Ectopic expression	Higher SSC, increased ethylene sensitivity	Zhang et al., 2020
	ARF10?	SSC11.1	*S. pimpinellifolium* RIL	QTL associated with higher °Brix	Zhao et al., 2019; Tieman et al., 2017
Solyc11g0 69190	*ARF4*		Antisense	More starch in early fruit development; increased AGPL1 activity and expression; enhanced chlorophyll content	Jones et al., 2002; Sagar et al., 2013
Solyc11g0 69500	*SlARF10*		Ectopic expression	Higher SSC; increased chlorophyll accumulation	Yuan et al., 2018a
Solyc12g 006340	*SlARF6a*		Ectopic expression	Higher SSC; increased chlorophyll accumulation	Yuan et al., 2019
Solyc12g 095910	SlVPE5		RNAi	Higher SSc; increased *TIV1* activity.	Ariizumi et al., 2011
Solyc12g 099200	INVINH1/ CIF1		RNAi	Increased activity of *LIN5*.	Zanor et al., 2009b; Jin et al., 2009

Abbreviations used: RNAi: RNA interference; SSC: Soluble Solids Content
†QTL, locus, mutant allele or introgression lines

PATHWAYS TO SUGAR IN TOMATO

Beyond trait-based breeding and marker-assisted breeding, New Breeding Techniques (NBT) or "gene editing" hold tremendous potential by combining QTLs, GWAS loci, and the growing knowledge of molecular and physiological mechanisms. The rest of this

review aims at giving an overview of the genes that have been shown to provide the potential for improving tomato sugar content. Different key players in the sugar pathway are highlighted, with a focus on potential breeding targets comprising genes that play a role in (1) synthesis, (2) transport or (3) storage (Osorio *et al.*, 2014). Increasing tomato sugar content most likely be achieved by combining several alleles and strengthened by appropriate cultivation and post-harvest practices.

It is important to note that two major types of tomato cultivation exist based on two contrasting plant growth habits. In the indeterminate growth habit of wild-type tomatoes and current greenhouse and tunnel cultivars, fruit trusses are produced at regular intervals, equally interspersed by three leaves (the sympodial unit) harvested by hand throughout the season. This contrasts with open-field cultivation-type varieties for processing tomatoes that carry a mutation (*self-pruning, sp*) in an inhibitor of flowering (*SP*). This mutation causes a determinate growth habit with a bushy plant phenotype and multiple trusses of fruits that are ripe at more or less the same time. This and the open field cultivation, allow for mechanical harvesting. It is essential to realize that many of the studies presented here report genes or QTLs identified and characterised in only one of these two types. This is perhaps best illustrated by *IL-9-2-5*, an introgression of part of *S. pennellii* chromosome 9 in the determinate variety M82. The increase in °Brix of this line is due to two QTL's: *Brix9-2-5* affecting LIN5 activity, a fruit-based trait (see below) and *PW9-2-5*, causing a semi-determinate habit with more vegetative biomass and plant architecture-based increase in °Brix (Fridman *et al.*, 2002). The first QTL's effect is apparent in greenhouse conditions, but the second is not (Baxter *et al.*, 2005). Thus, conclusions may be valid for only one of the two growth habits, or for both. Additionally, the results obtained by modifying gene expression under laboratory conditions, sometimes in model varieties like cv. MicroTom, gives valuable insights, but many of these insights await validation in modern commercial varieties and regular cultivation practices. The importance of cultivation methods is exemplified in a recent study that added several sugar-pathway related transgenes to an indeterminate variety. Several transformants had a higher °Brix when grown in a polytunnel, but not when grown in a glasshouse (Vallarino *et al.*, 2020).

PHOTOSYNTHESIS

Photosynthesis is the ultimate source of all sugars in plants. A large body of research aims at increasing photosynthetic capacity, not necessarily for improving sugar content, but mostly to increase yield (Long *et al.*, 2015). However, with the proper approach, one might divert some or all this increased capacity to the tomato fruits' sugar content. Strategies focus on various ways to increase the source capacity: by improving the light reaction, carboxylation or sucrose synthesis, or by influencing sugar signalling (Sonnewald and Fernie, 2018). Photosynthesis is naturally limited by day length, and growing plants under continuous light would theoretically increase production or raise fruit sugar content with production remaining equal. However, in

contrast to many other species, continuous light in tomato causes leaf damage and is potentially lethal. A dominant allele of the type III light-harvesting chlorophyll a/b binding protein 13 gene (*CAB-13*) from *S. pimpinellifolium* confers tolerance to continuous light, leading to up to 20% yield increase. Fruit sugar content was not reported, so it remains to be determined whether this approach could also work for that (Velez-Ramirez *et al.*, 2014).

LEAF NUMBER, SHAPE, AND SENESCENCE

Leaves are the most important photosynthetic source tissues. By model simulations and leaf-pruning experiments, Xiao et al. have shown that in indeterminate cultivars, a sympodial unit of two leaves instead of three does not reduce yield and identified growing leaves as a significant sink, potentially competing with fruits (Xiao *et al.*, 2004). This competition was confirmed by breeding introgression lines carrying an *S. pennellii* allele of *SP3D*, the major activator of flowering in tomato, conferring a sympodial unit of 2 with the expected yield increase (Heldens *et al.*, 2009). These results show that the amount of carbohydrates produced at the source is more than sufficient to sustain production and imply that if the surplus carbohydrate produced in tomatoes were more efficiently allocated to fruits, the result would increase the fruit sugar content.

Leaf shape also affects fruit sugar content. A study comparing natural variation of field grown-tomato varieties found that round(er) leaves, as seen in varieties with the *potato leaf* (*c*) mutation, had a strong positive impact on both tomato yield as well as on sugar content (Rowland *et al.*, 2020). The mechanism through which leaf shape acts on these parameters is yet unknown. Another approach to increasing sugar production is delaying leaf senescence, as it extends the period of productive photosynthesis output of a leaf. A delay in leaf senescence can be achieved by a knock-down of the transcription factor *ORESARA1* (*SlORE1*), resulting a significant increase of sugars in fruits and yield (Lira *et al.*, 2017).

Similar results were obtained by knocking down the expression of another positive regulator of leaf senescence, the *NAC-like TF gene ACTIVATED BY APETALA3/PISTIALLATA 2* (*SlNAP2*) in an indeterminate background (cv. Moneymaker) (Ma *et al.*, 2018). Also, here, both yield (fruit number) as well as °Brix were significantly increased. However, it may be difficult to extrapolate these results to commercial indeterminate-tomato production. The lower leaves in older plants (>3m high) receive little light and are routinely removed. In the open field with determinate processing tomato varieties, lower leaves may also be shaded due to the bushy phenotype. The crop cycle is much shorter than for greenhouse tomatoes (3 months vs. almost a year), which begs the question of whether leaf senescence is a significant factor in the field.

PHOTOSYNTHESIS IN FRUITS

Photosynthesis in immature green tomato fruits can account for 15% of a sympodial unit's photosynthetic activity (Hetherington *et al.*, 1998). Wild-type tomato fruits, including older cultivars and heirlooms, display a 'green shoulder', a darker green area near the stem end. This green shoulder causes uneven ripening of the fruit, which was deemed undesirable to the consumer. Consequently, in the early 20th century, the *uniform ripening* (*u*) mutation that lost the green shoulder was incorporated in virtually all modern commercial cultivars. Strikingly, this may have been responsible for a drop of 0.5 °Brix in ripe fruit (Powell *et al.*, 2012). The gene underlying this mutation was identified as *GOLDEN2-LIKE 2* (*SlGLK2/UNIFORM)*. The *u* allele encodes a truncated loss-of-function version of a transcription factor (TF) that controls fruit chloroplast numbers and chlorophyll levels (Powell *et al.*, 2012). Expression of *SlGLK2* displays a latitudinal gradient in fruit, causing the green-shouldered, non-uniform colour (Nguyen *et al.*, 2014). A *SlGLK2* homolog, *GOLDEN2-LIKE 1* (*SlGLK1*) expressed mostly in leaves plays a role in chloroplast formation and chlorophyll accumulation (Waters *et al.*, 2008). Overexpression of either gene produces dark-green fruit with increased chlorophyll accumulation and chloroplast development, concurrent with up to 1 °Brix increase (Powell *et al.*, 2012; Lupi *et al.*, 2019).

The observation that enhanced fruit chloroplast production and chlorophyll content led to higher fruit sugar content was made in many studies following that on *uniform ripening* (Figure 1). A Class I KNOTTED1-LIKE HOMEOBOX (KNOX) TF gene *TKN4* is mutated in the *uniform grey-green* (*ug*) allele (Bohk and Scott, 1945; Nadakuduti *et al.*, 2014). *TKN4* and the related *TKN2* act upstream of *SlGLK2* and the related gene *ARABIDOPSIS PSEUDORESPONSEREGULATOR2-LIKE (SlAPRR2-LIKE*) in the formation of their expression gradient in fruit (Nadakuduti *et al.*, 2014). Interestingly, a gain-of-function allele of *TKN2*, *Curl* (*Cu*), shows a fruit-specific increase of chloroplasts numbers similar to that seen in *SlGLK2* overexpression (Parnis *et al.*, 1997). This indicates that the use of gain-of-function mutations can substitute for transgene expression, although the *Cu* mutation also affects leaf shape in this case. Ectopic expression of the *Auxin Response Factor (ARF)* genes, *SlARF6a* or *SlARF10,* also increased chlorophyll content (Yuan *et al.*, 2018, 2019). Both *SlGLK1* as well as *SlGLK2* expression are positively regulated by *SlARF6a* and *SlARF10*, while *SlARF4/DR12* represses *SlGLK1.* Indeed, *ARF4* knock-down lines show increased chlorophyll content (Jones *et al.*, 2002; Sagar *et al.*, 2013; Le Roy *et al.*, 2013; Yuan *et al.*, 2018, 2019). Interestingly, the SNP marker for the abovementioned Brix QTL *SSC11.1* (Tieman *et al.*, 2017*b*) is located just downstream of *ARF10,* making this gene (or its regulation) an attractive candidate for underlying the QTL. Other TFs that influence chloroplasts and sugar content are *SlBEL11* and *SlBEL4* (Meng *et al.*, 2018; Yan *et al.*, 2020). Pleiotropic effects observed in RNAi studies predict that knock-out mutations in some of these may have harmful effects on the plant, making them less suitable as a

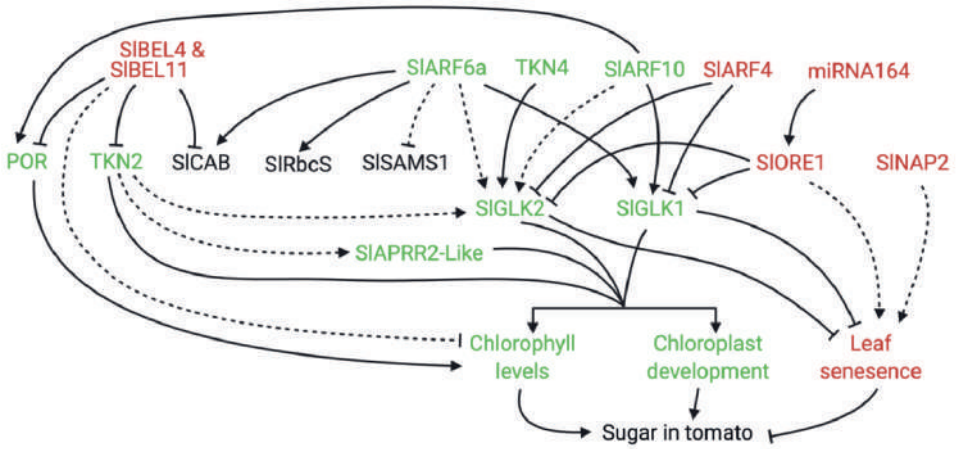


Figure 1: Potential targets in tomato that can increase source capacity. For references, see text. Green text indicates a positive influence on the sugar content of the fruit, while red text indicates a negative influence. A dashed line represents indirect or unproved regulation of the subsequent gene or process. Effects on processes other than related to sugar content are not depicted in this model.

target. The more downstream a gene operates, the fewer pleiotropic effects are expected to occur from modifications (Liu *et al.*, 2014).

Other seemingly unrelated developmental pathways can also increase the photosynthetic capacity of developing fruit. In the spontaneous epigenetic mutant *green stripe* (*gs*), a methylated isoform of *Tomato AGAMOUS-LIKE1* (*TAGL1*), leads to differential chloroplast development and carotenoid production (Liu *et al.*, 2020). *gs* fruits have gained 0.8 °Brix in ripe fruit compared to the wild type. Concurrently, the darker green stripes of the immature fruit developed into lighter red sectors due to a decreased lycopene production.

From the results described above, increasing the photosynthetic capacity of the fruit by raising chloroplast number, chlorophyll content, or both appears to be an attractive approach to improving sugar content. However, wild-type (*U*) fruits are more sensitive to cracking and yellow shoulder, an alternative explanation for the loss of *U* in breeding, and a possible manifestation of photooxidative stress from high light intensity (Cocaliadis *et al.*, 2014). Thus, increasing fruit photosynthetic capacity may only be useful for application in low light seasons or climates.

TRANSPORT

Source and sink activities are tightly linked. If the source sugar production is increased, more sink capacity has to be created as well. Increasing both transport and storage can achieve this capacity. Source to sink transport of sugars is a well-studied field, and there are several excellent reviews on this topic (Fernie et al. 2020; C. Zhang and Turgeon 2018; Yadav, Ayre, and Bush 2015, summarised in **Figure 2**).

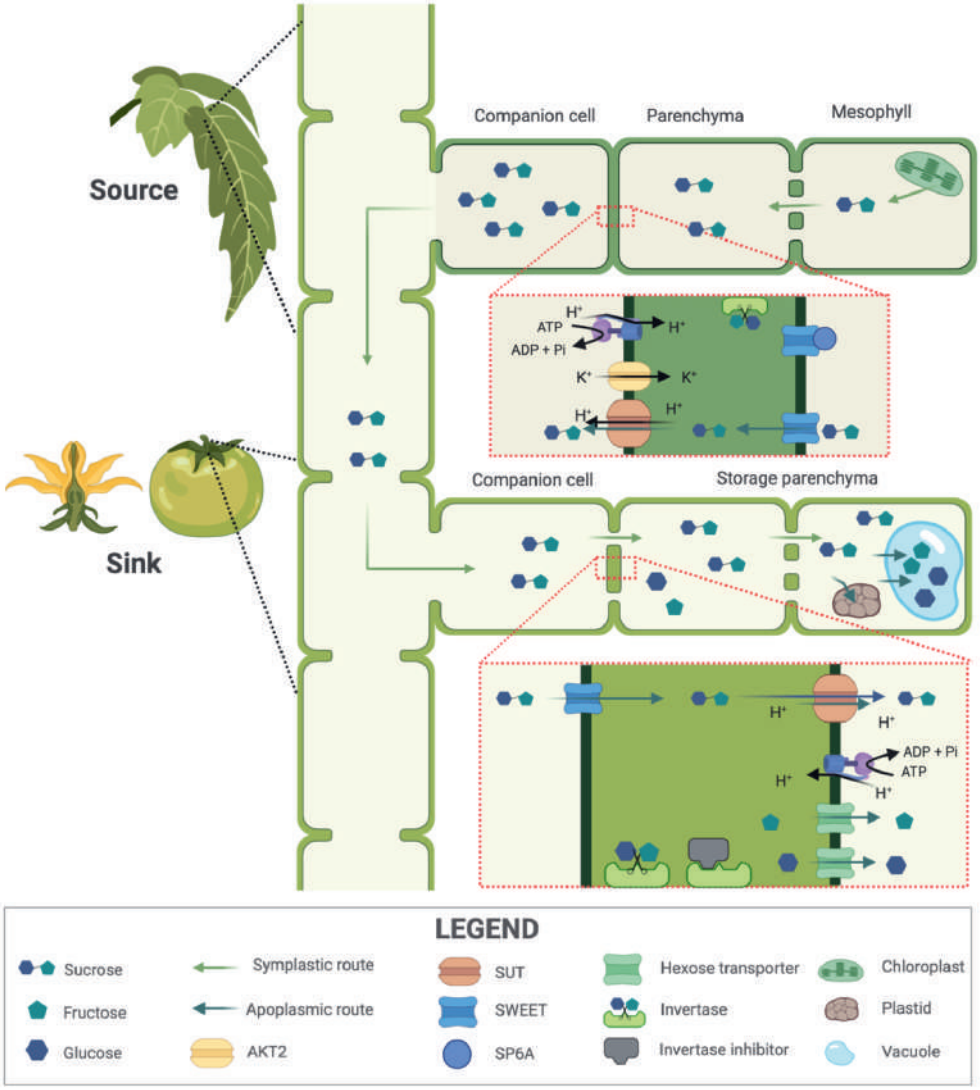


Figure 2: Sugar transport in a tomato plant. Cartoons represent sugars, enzymes, and transporters, see the legend for the protein(complex) names. Created with BioRender.com.

Sucrose mostly starts at the source in mesophyll cells of leaves or to some extent, fruit. This sucrose is translocated via the phloem to developing fruits, the sink. Bulk translocation is achieved by a concentration gradient between the source and the sink. This gradient is generated by concentrating sucrose in the source-phloem via phloem loading, the "push" effect. On the other side, at the sink, a "pull" is generated by storing translocated sugars in vacuoles and plastids. Phloem loading occurs via the plasmodesmata, the symplastic route, to the phloem parenchyma-companion cell-border. There, to create a concentration gradient, sucrose loaded from the symplast to the apoplast (the plant's collective cell wall and intracellular space) via SUGARS WILL

EVENTUALLY be EXPORTED TRANSPORTERS (SWEET) proteins. Sucrose is transported against the gradient to companion cells of the phloem by sucrose/proton symporters (Sucrose Uptake Transporters, SUT; or Sucrose Uptake Carriers, SUC). Bulk transport then occurs via sieve elements connecting the source and the sink tissue. Bulk transport is mostly affected by factors that play a role in the source-sink balance, such as plant growth, defoliation, number of bearing fruits, light intensity and temperature (Yoshihiro, 1986). Unloading at the sink occurs via the symplastic route, through plasmodesmata, or through the apoplastic route. For sucrose unloading, SWEETs transport sucrose into the apoplast, from where the sucrose is transported into the sink parenchyma cells via sucrose transporters or alternatively the sucrose is first hydrolysed to hexoses by cell wall-bound invertases and then is transported by hexose transporters. For maintaining the concentration gradient, sucrose and hexoses are stored in vacuoles or converted into starch in plastids. During phloem loading and unloading, several key genes are known or are predicted to play a role. These are interesting targets for improving the system.

SWEET PROTEINS

One of the key gene (family) involved in sugar transport is the *SWEET* family. *SWEETs* are a recently discovered sugar transporter family of proteins that consist of seven transmembrane domains. These domains are predicted to form a pore across the membrane, through which specific types of sugar can travel (Chen *et al.*, 2010). SWEETs fall into four clades and their clade position seems to be correlated to the selectivity toward hexoses (Clade I and Clade II) or sucrose (clade III) (Eom *et al.*, 2015). In *S. lycopersicum*, twenty-nine *SlSWEET* genes were found (Feng *et al.*, 2015). Most *SlSWEETs* have not yet been functionally characterized, but expression patterns can give clues for their relevance for sugar transport in different aspects of fruit development. Potential candidates would be *SlSWEET1a, 11a, 12a* expressed in leaves and SlWEET1b, *1c, 2a, 7a* and *14* in young fruits (Feng *et al.*, 2015; Shen *et al.*, 2019; Ho *et al.*, 2019). A high expression of *SlSWEET1a,* the underlying gene of the *Fgr*-locus, in an introgression line with wild tomato species *Solanum habrochaites,* was responsible for an increased ratio of fructose-to-glucose (Levin *et al.*, 2000; Shammai *et al.*, 2018). As fructose has a sweeter taste, this results in a sweeter fruit.

To prevent an excess of sucrose leakage into the source cells, sugar transporters need to be tightly regulated (Fernie *et al.*, 2020). For instance, StSP6A physically interacts with StSWEET11 in potato and blocks leakage of sucrose to the apoplast, stimulating symplastic flow (Abelenda *et al.*, 2019). It remains to be determined if similar mechanisms act in tomato fruit growth, but this suggests that modification of SWEETs activity may be a target for enhancing sugar transport.

SUCROSE TRANSPORTERS

Fifty-two putative sugar transporter proteins in tomato, including the membrane proton symporters Sucrose Transporters (SUTs or SUCs) and Monosaccharide Transporters (MSTs), function in phloem loading and unloading (Leggewie *et al.*, 2003; Reuscher *et al.*, 2014). There are three known SUT members in tomato (Kühn and Grof, 2010). SlSUT1 is the main phloem loader, while SlSUT2 is the main unloader (Barker *et al.*, 2000; Hackel *et al.*, 2006). SlSUT4 most likely also functions in unloading in fruits, as its expression is highest in ovaries (Weise *et al.*, 2000; Hackel *et al.*, 2006). A reduction in *SlSUT2* expression leads to reduced pollen viability, reduced sugar and starch content in young fruit, and a lower fruit yield (Hackel *et al.*, 2006). The different roles of *SlSUT1* and *SlSUT2* highlight that *SUTs* are essential in both the "push" and "pull" mechanisms of sucrose transport. In *Arabidopsis*, both *SUT2* and photosynthesis are downregulated when there is abundant apoplastic glucose, which may signal low sink demand or impaired assimilate allocation (Wingenter *et al.*, 2010). Interference with this feedback-regulation by apoplastic sugars or increasing the expression of *SUTs* may be an approach towards higher fruit sugar content. Ectopic expression of a pear (*Pyrus bretschneideri*) homolog of *SlSUT4*, *PbSUT2* led to increased sucrose content in tomato fruit (Wang *et al.*, 2016). Similar results from expression of heterologous *SUTs* were obtained in potato tubers and in pea (Leggewie *et al.*, 2003; Sun *et al.*, 2011; Lu *et al.*, 2019*a*). However, ectopic overexpression of *SUTs* does not always give the desired effects (Rosche *et al.*, 2002; Srivastava *et al.*, 2009).

CELL WALL INVERTASES

Invertases cleave the O-C bond in the disaccharide sucrose, breaking it into glucose and fructose. The presence of these monosaccharides drives import into the fruit. Due to the combined activity of the acid cell wall-bound, vacuolar invertases, and the neutral cytoplasmic invertases (CWIN, VIN and CIN respectively) in fruit tissues, ripe tomatoes contain mostly glucose and fructose at equimolar concentrations and little sucrose (Klee and Giovannoni, 2011). Tomato has four different CWINs, of which *Lycopersicum Invertase 5 (LIN5)* and *LIN7* are expressed in reproductive organs (Zhang *et al.*, 2013). *LIN5* expression and CWIN activity increase in the ovary two days after pollination simultaneously with *SlSUT1* expression (Shen *et al.*, 2019). The earlier mentioned QTL IL9-2-5 was found to cause an increase in CWIN activity resulting in high °Brix, without a yield decrease in the field (Fridman *et al.*, 2002; Baxter *et al.*, 2005) Three amino acid changes in the *LIN5* coding region near the catalytic site of the invertase were shown to be linked to higher catalytic activity of LIN5 (Fridman *et al.*, 2000, 2004). The importance of *LIN5* was confirmed in knock-down lines, giving lower °Brix (Zanor *et al.*, 2009*b*). In addition, silencing the gene encoding an *inhibitor for LIN5, INVINH1* (also known as *Cell-wall Inhibitor of β-Fructosidase, SlCIF1*), led to higher sugar levels in tomato fruit (Jin *et al.*, 2009). Higher activity of LIN5 in IL9-2-5 also resulted in higher expression of sugar translocation protein genes *Hexose Transporter 3* (*SlHT3*)

and *SlSUT4* (Baxter *et al.*, 2005). The same effect was shown in a study where ectopic expression of an apple Hexose Transporter, *MdHT2.2* increased the expression on *LIN5* and hexose content in ripening fruit (Wang *et al.*, 2019). Conversely, increasing the activity of cell wall invertase might result in increased expression of tomato *HT2*, and of *SWEET12C* (Ru *et al.*, 2020). Thus, modification of expression or activity of single sugar import machinery components may have a cascading effect on some of the other components, allowing an increase of the final sugar content of the ripe fruit.

HEXOSE TRANSPORTERS

The cleavage products of sucrose produced by cell wall invertases are translocated into the sink cells by hexose transporters (HT). In fruit parenchyma cells in particular, hexoses accumulate to high levels. Three hexose transporters, *SLHT1, SLHT2* and *SLHT3* are localized in the plasma membranes of these cells. Their role in sugar unloading is shown by RNAi knock-down lines of the *SlHTs* that showed a 55% decrease in hexose accumulation in fruits while source leaves and phloem transport capacity to fruit were unaffected (McCurdy *et al.*, 2010). Although no direct upregulation of *HT* genes in tomato has been reported so far, it has been shown that when *HTs* are upregulated as an effect from other genes, sugar content is increased. This suggests that a higher *HT* expression is beneficial.

The entire pathway of sugar translocation has been subject of intensive studies, both in tomato and in many other species. Yet, surprisingly few modifications of targets that can increase sugar in tomato are known. In general, it seems that overexpressing *SWEETs, SUTs, Invertases* and *HTs* can increase fruit sugar content, but this has never been tried with endogenous tomato genes. Making mutations in the transporter protein genes themselves will most likely not increase tomato sugar content. Another option could be to target the promoters of the genes by, for example, CRISPR/Cas9 to remove negative Cis-regulatory elements. Alternatively, one could disrupt repressive Transcription Factors involved in the pathways. To achieve up-regulation or increased activity in a non-genetic modification (GM) way remains challenging.

STORAGE

The imported sugars are used for metabolism, the maintenance of turgor or for the biosynthesis of starch in immature fruits (**Figure 3**) (N' tchobo *et al.*, 1999). Cytosolic and vacuolar invertases cleave sucrose into glucose and fructose, while Sucrose Synthase (SUS) cleaves imported sucrose into UDP-Glucose and fructose (Chua *et al.*, 2008). Free hexoses are then phosphorylated by hexokinase (mainly glucose and some fructose) or fructokinase (fructose) to Glucose-6-Phosphate (G6P) and Frucose-6-Phosphate (F6P) respectively (Granot *et al.*, 2013; Stein and Granot, 2018). G6Ps are isomerized by phosphoglucomutase to G1P and transported to the plastid, were ADP-glucose pyrophosphorylase (AGPase) converts it to starch. The bulk of starch storage

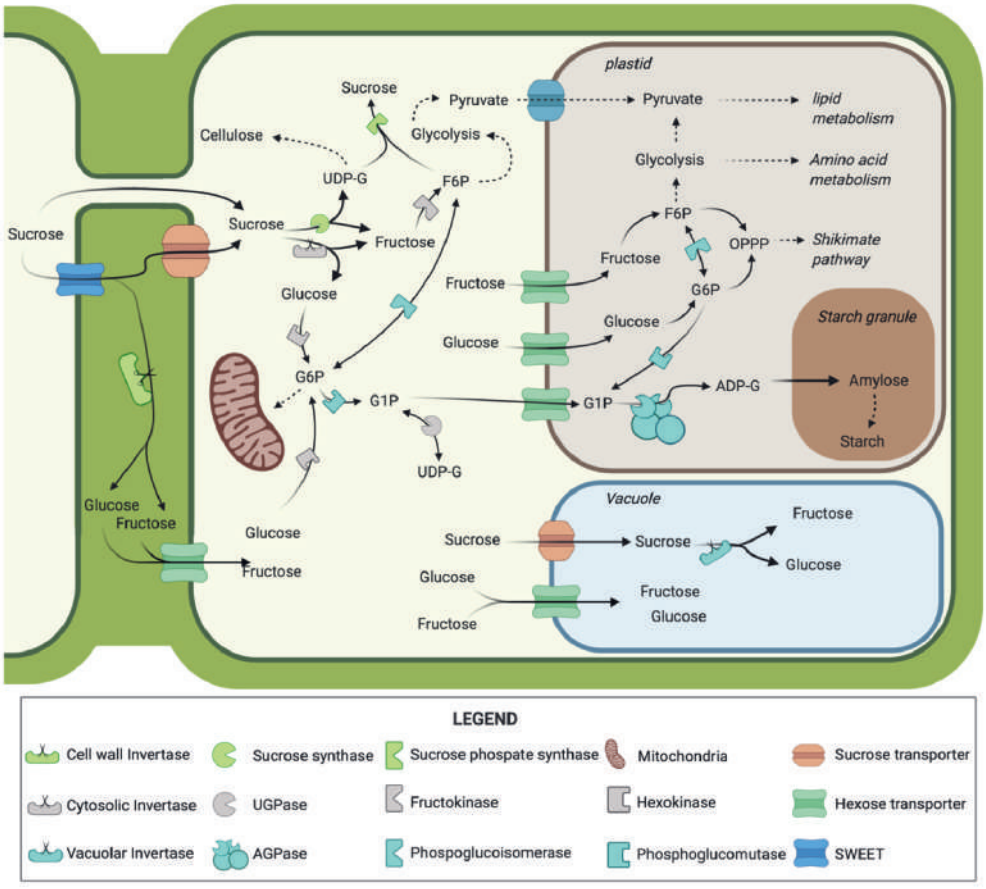


Figure 3: Sugar storage and primary metabolism in a tomato storage parenchyma cell. Cartoons represent enzymes and transporters, see the legend for the protein(complex) names. Note that repressors of the enzymes and transporters are not depicted in this figure. Dashed arrows depict multiple steps or undepicted proteins. Created with BioRender.com.

is located in the pericarp and columella (Schaffer and Petreikov, 1997). Upon ripening, the stored starch is hydrolysed, increasing the hexoses' (Carrari *et al.*, 2006).

To generate a negative sugar gradient from source to sink in the phloem, the sugars unloaded in sinks are stored in the vacuoles and plastids (Stein and Granot, 2018; Fernie *et al.*, 2020). Sugars and other compounds, like organic acids and inorganic ions, are transported to the vacuole by vacuolar/tonoplast transporters, carriers and pumps such as Tonoplast Sugar Transporters (TSTs) and vacuolar SUTs (Aluri and Buttner, 2007; Wingenter *et al.*, 2010; Schneider *et al.*, 2012). Overexpression of a melon (*Cucumis melo*) *Tonoplast Sugar Transporter* 2 (*CmTST2*) in strawberry and cucumber increased sugar accumulation in these fruits (Cheng *et al.*, 2018). In watermelon (*Citrullus lanatus*), an increase in *ciTST2* gene expression has been a major molecular event in domestication (Kyriacou *et al.*, 2018). Expressing an apple (*Malus domestica*) tonoplast

H+/glucose symporter, *MdERDL6-1*, in tomato increased fruit sugar content (Zhu *et al.*, 2021). Increasing native TST expression in tomato could yield similar effects, but this remains to be validated. Once transported into the vacuole, the sugars are hydrolysed by the soluble invertases (Koch, 2004).

VACUOLAR AND CYTOSOLIC INVERTASES

Soluble invertases such as acidic Vacuolar Invertase (VIN) and neutral Cytosolic Invertase (CIN) play a role in cellular metabolism and storage. VIN is responsible for hexose (as opposed to sucrose) accumulation in red-fruited tomatoes. The importance of VIN was first discovered through a backcross line with *S. chmielewskii*. This revealed the *sucr* (*sucrose accumulator*) allele, with the underlying gene identified as *Tonoplast Invertase 1* (*TIV1* or *SlVIN1*) (Chetelat *et al.*, 1993; Klann *et al.*, 1996). Red-fruited hexose-accumulating tomatoes contain a retrotransposon upstream of *TIV1*, resulting in increased expression throughout fruit development compared to the green-fruited sucrose accumulators (Moy *et al.*, 2007). Further evidence for the role of *TIV1* came from a *TIV1* knock-down line that reverted tomato to a sucrose accumulator (Klann *et al.*, 1996). In addition, fruit size was decreased by 30%, in this line, which may reflect the lower osmotic pressure exerted by one equivalent of sucrose versus two equivalents of its products, glucose and fructose, driving water uptake and fruit growth. Similar to CWIN and CIF, TIV1 is post-translationally regulated by *Vacuolar Invertase Inhibitor* (*SlVIF*) (Tauzin *et al.*, 2014). Knock-down of *SlVIF* expression by RNAi delayed ripening and increased hexose levels while overexpression of *SlVIF* accelerated ripening and reduced hexose levels by 40% (Qin *et al.*, 2016). Vacuolar processing enzymes (VPEs), of which there are five in tomato, are involved in proteolysis of VIN. The suppression of *SlVPE1* and *SlVPE5* by RNAi increased VIN activity leading to increased hexose and sucrose accumulation (Ariizumi *et al.*, 2011).

Compared to VIN and CWIN, the physiological roles of cytoplasmic CINs are least understood. Following from work in other plant species, CINs are thought to be important for the maintenance of cytosolic sugar homeostasis, starch accumulation, cellular function and they are essential for normal growth by their involvement in the uridine diphosphate glucose (UDP-G) pathway (Welham *et al.*, 2009; Malinova *et al.*, 2014; Samac *et al.*, 2015; Decker, 2017; Barnes and Anderson, 2018). Their function in tomato and their possible role as a target for increasing sugar content still remains to be determined (Ruan, 2014; Wan *et al.*, 2018).

SUCROSE SYNTHASES

In tomato, there are six *SUS* genes, and their different expression patterns imply a role both in sugar loading (source) as well as in unloading (sink) (Dinh *et al.*, 2018). Increased activity of SUS has been linked to sink strength, which increases fruit growth (Stein and Granot, 2019). Down-regulation of *SlSUS1, SlSUS3* and *SlSUS4* expression by RNAi in

tomato did affect sugar content (Goren *et al.*, 2017). For the genetic variation in *SUS* genes, wild accessions would have to be studied. For instance, a four amino acid difference compared to cultivated tomato in an accession of *S. arcanum* did show a higher activity of SUS3, especially at low temperatures. It has been speculated that this was the cause of a higher ratio of sucrose to hexose in fruits (Dinh *et al.*, 2018, 2019). SUS works in concert with Fructokinases (FRKs) (Pego and Smeekens, 2000; Davies *et al.*, 2005; Poór *et al.*, 2015), but down-regulation of *SlFRK* did not affect ripe fruits (Dai *et al.*, 2002; Odanaka *et al.*, 2002).

Sucrose-6-phosphate synthase (SPS) synthesizes sucrose through a reaction from F6P and UDP-glucose. Overexpressing the *SPS* encoding gene can improve plant growth by increasing carbon assimilation (Singer *et al.*, 2020). In tomato, ectopic expression of a maize *ZmSPS* led to an enhancement of SUS activity in tomato fruit, increased sink strength and sucrose accumulation (Nguyen Quoc *et al.*, 1999).

A different strategy to increase sugar content was based on overcoming a negative sucrose-induced feedback loop in *bZIP1*. *Arabidopsis AtbZIP11* and its tobacco ortholog *tbz17* regulate transcription of *asparagine synthase* (*ASN*), *proline dehydrogenase* (*ProDH*), *sucrose phosphate phosphatase* (*SPP*) and *SPS* (Satoh *et al.*, 2004; Hanson *et al.*, 2008; Thalor *et al.*, 2012). In turn, regulation of these *bZIPs* occurs through an upstream Open Reading Frame (uORF) that causes Sucrose-Induced Repression of Translation (SIRT) (Wiese *et al.*, 2004; Calvo *et al.*, 2009; Hanson and Smeekens, 2009; Von Arnim *et al.*, 2014). Overexpression of *tbz17* removed the sucrose mediated repression and increased *ASN, ProDH, SPP* and *SPS* expression and increased sucrose content in leaves (Thalor *et al.*, 2012). In tomato, this concept was utilized by replacing the uORF of *SLbZIP1* with a fruit specific promoter, causing increased *SPP* and *SPS* expression as well as a 50% increased sugar content without an effect on fruit size (Sagor *et al.*, 2016). Similar results were recently obtained by base-editing the SIRT-uORF of a strawberry *bZIP1*. This mutant had a higher sucrose content without a yield penalty as well (Xing *et al.*, 2020). These studies highlight the value of finding and modifying uORFs to fine-tune the expression of genes.

AGPase

Sugars in starch are stored in plastids of developing fruit to maintain a high sink strength. In the plastids, the tetrameric complex ADP-glucose pyrophosphorylase (AGPase) catalyses the first rate-limiting reaction between ATP and G1P to produce ADP-glucose, which is further converted to starch (Chen and Janes, 1997; Beckles *et al.*, 2001; Ballicora *et al.*, 2004). Starch accumulated in young tomato fruit is depolymerized during ripening to form glucose, maltose or G1P. Glucose and maltose are exported to the cytoplasm, where maltose is converted to glucose. The small catalytic subunit of the AGPase complex is encoded by *SlAGPS* (also called *AgpB*) (Goto *et al.*, 2013). Three genes encode the large, stabilizing subunits: *AGPL1*, *AGPL2* and *AGPL3* (Also called

AgpS1, S2, S3), of which *AGPL1* is the predominant one in tomato fruit (Chen *et al.*, 1998; Xing *et al.*, 2005). An allele of *AGPL1* originating from an introgression with *S. habrochaites, AGPL1*^{H}, had increased AGPase activity, starch levels in young fruit and sugars in ripe fruit (Schaffer *et al.*, 2000). The increased activity of AGPase was due to the prolonged expression of the *AGPL1*^{H} allele and increased stability of the complex, but not to altered intrinsic enzyme activity (Petreikov *et al.*, 2006). The *AGPL1*^{H} allele appears to be already introgressed in several modern processing tomato varieties, as demonstrated in a whole genome-resequencing study (Lin *et al.*, 2014).

AGPase activity and *AGPL1* expression may be increased in a trans-acting fashion such as in *Auxin Response Factor 4* (*SlARF4*) knock-down lines (Jones *et al.*, 2002; Sagar *et al.*, 2013) and in the *S. pennellii* Introgression Line, IL8-3, with higher °Brix as a result (Ikeda *et al.*, 2013, 2016). From the above, AGPase is co-regulated with other sugar-related proteins. Thus, both cis- and trans-regulatory effects leading to an increase in AGPase activity can result in higher sugar content, making it an interesting target.

To summarize, the storage of sugars in sink tissues is vital to keep the sink strength high. Sequestration of sugars prevents the sucrose-mediated downregulation of sugar processing enzymes like fructokinase, SUS, UGPase and AGPase (Du Jardin *et al.*, 1997; Schaffer and Petreikov, 1997; Li *et al.*, 2002; Decker, 2017). Upon ripening, the stored sugars and starch are important to determine the final sugar levels and composition.

THE 'SUGAR VERSUS YIELD' CONFLICT

Favouring fruit size during breeding or domestication while disregarding sugar content, may have led to the loss of the high sugar content-associated alleles of these two QTL's (Tanksley *et al.*, 1996; Kimbara *et al.*, 2018; Zhao *et al.*, 2019). Well-known examples are the loss of wild alleles of *LIN5* and *SSC11.1* (Causse, 2002; Causse *et al.*, 2003; Tieman *et al.*, 2017*a*). Tieman et al. found a negative correlation between sugar and yield, when the Brix9-2-5-derived *LIN5* allele was ectopically expressed (Tieman *et al.*, 2017*a*). However, this is in contrast with the original studies of Brix9-2-5, where no decrease in yield was found in field-grown tomatoes (Fridman *et al.*, 2002; Baxter *et al.*, 2005). In a study by Prudent et al., 15 QTLs for sugar content were detected and in only one case did the QTL not co-localize with a QTL for fruit size (Prudent *et al.*, 2009). These studies strengthen the hypothesis that for most sugar-related QTLs, there is a negative correlation with size. However, it seems possible that there are QTLs or application methods that do not necessarily fit this correlation. For one such QTL found in *AGPL1,* in several both determinate as well as indeterminate backgrounds, the *AGPL1*^{H} allele caused higher sugar content without an adverse effect and actually often a positive effect on yield (Petreikov *et al.*, 2009). Perhaps it is, therefore, not surprising that this allele has already been introgressed in some modern processing lines (Lin *et al.*, 2014). Thus, although there may be a practical upper limit to the yield*Total Soluble Solid in fruit parameter, this is not yet obtained under most reported conditions. In addition,

certain markets allow a sacrifice in yield without decreasing revenue for tomato growers and retailers as many consumers prefer smaller tomatoes with superior taste and are increasingly willing to pay for quality (Oltman *et al.*, 2014).

A useful breeding approach that could be used to increase sugar content while maintaining high yields is Genomic Selection (GS) (Meuwissen *et al.*, 2001; Jannink *et al.*, 2010; Nakaya and Isobe, 2012; Desta and Ortiz, 2014). GS is the selection of candidate parents based on the predicted genomic potential, or genomic estimated breeding values (GEBVs). GEBVs are calculated based on associations between SNPs and phenotypes. GS is especially useful when the target trait is controlled by many genes, as is the case for sugar content. Computer simulations were performed for tomato to predict GEBVs and design a breeding method targeting yield and flavour simultaneously (Yamamoto *et al.*, 2016, 2017). The model found that recurrent GS was needed to improve both fruit yield and soluble solids content (Higashide *et al.*, 2012; Klee and Tieman, 2013). In addition, by looking at the whole genome the model predicted that increasing both yield and soluble solids would lead to an increased height to the first truss without an increase in the number of leaves to the first truss. This suggest that more space is needed between successive leaves to decrease the light extinction coefficient. Increased photosynthetic capability could provide the necessary sugars to facilitate the desired yield and soluble solids content. Another recent promising modelling experiment showed that uncoupling of the size-sweet trade-off is possible as well (Chen *et al.*, 2021). In this model, sucrose transport across the plasma membrane, tonoplast sucrose transport and a larger pH difference between the cytoplasm and vacuole (**Δ**pH) were indispensable to obtain a bigger and sweeter fruit. Simulating higher activity of tonoplast sucrose transporters or SPS led to increased weight and sugar concentration. Downregulation of fructokinase or glucokinase achieved the same. Implementing these models in breeding strategies could yield valuable phenotypes. When writing this review, it became apparent that many studies involving modification using molecular techniques only reported sugar content (or °Brix), or merely fruit size (without the number of fruits) and not total yield or, *vice versa*. This means that many promising approaches for the improvement of sugar content remain promising, awaiting a thorough analysis of yield effect before final conclusions can be drawn.

SUGAR SIGNALLING

Apart from fuelling growth and the synthesis of essential compounds, sugars act as signals to regulate expression of mRNA, TFs and other genes for hormonal, oxidative and defence signalling (Ruan, 2014; Yu *et al.*, 2015; Abelenda *et al.*, 2019). This is achieved by the generation of sugar signalling molecules such as sucrose, glucose, and fructose themselves, and trehalose-6-phosphate (T6P). T6P is synthesized from G6P and UDPG by T6P synthase (TPS; (Paul *et al.*, 2008). T6P is present at very low levels (Lunn *et al.*, 2006). In *Arabidopsis,* T6P is a sensor of sucrose and an indicator of high

sucrose levels (Lunn *et al.*, 2006; Nunes *et al.*, 2013*b*; Wahl *et al.*, 2013). T6P functions as a signalling molecule with a role in plant development, growth, flowering and abiotic stress responses (Schluepmann *et al.*, 2004; Chary *et al.*, 2008; Iordachescu and Imai, 2008; Gómez *et al.*, 2010; Wingler *et al.*, 2012; Wang and Ruan, 2013; Nunes *et al.*, 2013*a*). Under drought conditions, increased T6P led to increased sucrose and total soluble sugars, conveying drought-tolerance (Avonce *et al.*, 2004; Lin *et al.*, 2019). Furthermore, T6P increased starch synthesis and AGPase activation in *Arabidopsis* leaves and in potato tubers (Kolbe *et al.*, 2005; Debast *et al.*, 2011; O'Hara *et al.*, 2013). Thus, T6P levels could directly influence the final sugar content. In tomato, a salt and heat treatment led to a significant activation of trehalose synthesis through T6P-synthase (Rivero *et al.*, 2014). In these plants, glucose, sucrose and starch concentrations increased in leaves (Lyu *et al.*, 2018). The effect on tomato sugar content has not been examined yet. Several other studies in tomato show a clear link between abiotic stress and the sugar metabolism in tomato. Heat stress in tomato led to increased invertases, SUS and Hexokinases expression in stamens, leading to temperature induced stigma exertion (Pan *et al.*, 2019). A higher cell wall invertase activity through *LIN7* expression was correlated to heat tolerance in tomato (Li *et al.*, 2011). The ectopic expression of the apple *MsSWEET17* conferred higher drought tolerance and increased fructose levels (Lu *et al.*, 2019*b*). Overexpression of Na+/H+ antiporter SlNHX4 and/or the serine/threonine protein kinase SlSOS2 resulted in fruits with increased size and sugar content when grown under salt stress (Maach *et al.*, 2020, 2021).

CONCLUDING REMARKS AND FUTURE PERSPECTIVES

The analysis of the pathways associated with important quality traits like sugar for tomato flavour, can uncover attractive targets for breeding. In this review, we have discussed a variety of targets for breeders in sugar metabolism: improving production at the source or increasing transport and storage. Advanced Backcross and Introgression Lines have been the most valuable tool for finding and using variations resulting in sweeter tomatoes. QTLs from these introgression lines have been a key source of genetic variation leading to a higher sugar content, as we have seen for *LIN5*, *AGPL1* and *TIV1*. Many more QTLs remain relatively unexplored. In recent years several groups have sequenced a wide variety of tomato cultivars unlocking the potential of GWAS analysis and PAN genomes (Sauvage *et al.*, 2014; Gao *et al.*, 2019; Zhao *et al.*, 2019). Another interesting source retaining some of the original natural variation are the heirloom cultivars (Rowland *et al.*, 2020).

For the future, a molecular approach to increase sugar content holds great promise. Most of the QTLs have not been characterized extensively, and the underlying genes are not known. Research on these QTLs can increase our knowledge of the sugar pathways and identify new targets. In addition, the many other potential candidate

genes have not been validated for an actual role in tomato sugar accumulation, such as the *SlSWEET* genes. In general, our observations from reviewing the literature are that modifications more often lead to a detrimental effect than positive. It seems that more subtle changes are needed: small changes in protein activity or in expression. The latter could be achieved if the regulation of the respective genes is better understood (Sagor *et al.*, 2016; Li *et al.*, 2017; Xing *et al.*, 2020; Fernie *et al.*, 2020). Regulatory regions in promoters are the drivers of spatial and temporal expression (Chen and Rajewsky, 2007; Mejia-Guerra *et al.*, 2012). When the function of specific regulatory regions is known, for example, as an activator, it can be removed by targeted mutagenesis for fine-tuning expression levels or patterns (Swinnen *et al.*, 2016).

With the rapid developments in targeted mutagenesis, for instance, by CRISPR/Cas, breeders now have breeding opportunities that would have taken years with classical crossing and selection breeding or that were not even possible due to linkage-drag**.** It offers the promise of relatively easy, cheap, and especially fast breeding. When the breeders' goals are to change the expression of a gene while avoiding introducing foreign DNA, they can target the promoter of the gene of interest. A major caveat for the application of targeted mutagenesis is not the technical difficulty, but the regulatory restrictions on its products, for instance, in Europe. Perhaps the public could be swayed in favour of using targeted mutagenesis if more desirable traits were available to the public, for example, a sweeter tomato.

Although improving quality of tomato and other crops is possible using knowledge of the trait and available powerful and fast breeding methods, there is still a major obstacle for breeders and growers to develop this sweeter tomato, an issue that we did not discuss in this review: difficulty in phenotyping. Phenotyping for flavour can be expensive, time consuming as measurements can only be done once the fruit has ripened. A solution that can save time and costs for breeders and lead to higher quality tomatoes is knowledge on the causality of different metabolic compounds and appreciation by the consumer, linked to genetic markers. Valuable work in this area is being done (Zhu *et al.*, 2018; Klee and Tieman, 2018), but a large amount of research in this area remains necessary. An important note here is that to extrapolate scientific results in practice it is important to use representative varieties and cultivation methods. Finally, the genetic makeup for sugar content is just one of the aspects determining the flavour of tomato. Poor growing and post-harvest practices can undercut the breeding work. In the end scientists, breeders, growers, retailers, and the final consumers all play a role in making the tomato we eat taste a little better.

ACKNOWLEDGEMENTS

We thank Gerco Angenent for critically reviewing the manuscript.

REFERENCES

Abewoy Fentik, D. (2017). Review on Genetics and Breeding of Tomato (*Lycopersicon esculentum* Mill). Adv. Crop Sci. Technol. **05**: 1–6.

Agius, C., von Tucher, S., Poppenberger, B., and Rozhon, W. (2018). Quantification of sugars and organic acids in tomato fruits. MethodsX **5**: 537–550.

Alonge, M. et al. (2020). Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. Cell **182**: 1–17.

Alpert, K.B., Grandillo, S., and Tanksley, S.D. (1995). *fw 2.2*:a major QTL controlling fruit weight is common to both red- and green-fruited tomato species. Theor. Appl. Genet. **91**: 994–1000.

Azanza, F., Kim, D., Tanksley, S.D., and Juvik, J.A. (1995). Genes from *Lycopersicon chmielewskii* affecting tomato quality during fruit ripening. Theor. Appl. Genet. **91**: 495–504.

Bai, J., Baldwin, E.A., Imahori, Y., Kostenyuk, I., Burns, J., and Brecht, J.K. (2011). Chilling and heating may regulate C6 volatile aroma production by different mechanisms in tomato (*Solanum lycopersicum*) fruit. Postharvest Biol. Technol. **60**: 111–120.

Bai, Y. and Lindhout, P. (2007). Domestication and breeding of tomatoes: What have we gained and what can we gain in the future? Ann. Bot. **100**: 1085–1094.

Baldwin, E.A., Scott, J.W., Shewmaker, C.K., and Schuch, W. (2000). Flavour trivia and tomato aroma: Biochemistry and possible mechanisms for control of important aroma components. HortScience **35**: 1013–1022.

Ballicora, M.A., Iglesias, A.A., and Preiss, J. (2004). ADP-glucose pyrophosphorylase: A regulatory enzyme for plant starch synthesis. Photosynth. Res. **79**: 1–24.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell **27**: 299–308.

Bargmann, B.O.R., Marshall-Colon, A., Efroni, I., Ruffel, S., Birnbaum, K.D., Coruzzi, G.M., and Krouk, G. (2013). TARGET: A Transient Transformation System for Genome-Wide Transcription Factor Target Discovery. Mol. Plant **6**: 978–980.

Barrett, D.M., Beaulieu, J.C., and Shewfelt, R. (2010). Colour, flavour, texture, and nutritional quality of fresh-cut fruits and vegetables: Desirable levels, instrumental and sensory measurement, and the effects of processing. Crit. Rev. Food Sci. Nutr. **50**: 369–389.

Barry, C.S., Llop-Tous, M.I., and Grierson, D. (2000). The regulation of 1-aminocyclopropane-1-carboxylic acid synthase gene expression during the transition from system-1 to system-2 ethylene synthesis in tomato. Plant Physiol. **123**: 979–986.

Bauchet, G. et al. (2017). Identification of major loci and genomic regions controlling acid and volatile content in tomato fruit: implications for flavour improvement. New Phytol. **215**: 624–641.

Bauer, D.C., Buske, F.A., and Bailey, T.L. (2010). Dual-functioning transcription factors in the developmental gene network of *Drosophila melanogaster*. BMC Bioinformatics **11**: 1–14.

Baxter, C.J., Carrari, F., Bauke, A., Overy, S., Hill, S.A., Quick, P.W., Fernie, A.R., and Sweetlove, L.J. (2005a). Fruit carbohydrate metabolism in an introgression line of tomato with increased fruit soluble solids. Plant Cell Physiol. **46**: 425–437.

Baxter, C.J., Sabar, M., Quick, W.P., and Sweetlove, L.J. (2005b). Comparison of changes in fruit gene expression in tomato introgression lines provides evidence of genome-wide transcriptional changes and reveals links to mapped QTLs and described traits. J. Exp. Bot. **56**: 1591–1604.

Beckles, D.M., Craig, J., and Smith, A.M. (2001). ADP-glucose pyrophosphorylase is located in the plastid in developing tomato fruit. Plant Physiol. **126**: 261–266.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., and Nekrasov, V. (2013). Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant Methods **9**: 39.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N.J., and Nekrasov, V. (2015). Editing plant genomes with CRISPR/Cas9. Curr. Opin. Biotechnol. **32**: 76–84.

Bemer, M., van Dijk, A.D.J., Immink, R.G.H., and Angenent, G.C. (2017). Cross-Family Transcription Factor Interactions: An Additional Layer of Gene Regulation. Trends Plant Sci. **22**: 66–80.

Bemer, M., Karlova, R., Ballester, A.R., Tikunov, Y.M., Bovy, A.G., Wolters-Arts, M., Rossetto, P. de B., Angenent, G.C., and de Maagd, R.A. (2012). The tomato FRUITFULL homologs TDR4/FUL1 and MBP7/FUL2 regulate ethylene-independent aspects of fruit ripening. Plant Cell **24**: 4437–51.

Bergougnoux, V. (2014). The history of tomato: From domestication to biopharming. Biotechnol. Adv. **32**: 170–189.

Bertin, N. and Génard, M. (2018). Tomato quality as influenced by preharvest factors. Sci. Hortic. (Amsterdam). **233**: 264–276.

Bhaya, D., Davison, M., and Barrangou, R. (2011). CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. Annu. Rev. Genet. **45**: 273–297.

Blanca, J., Cañizares, 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**: e48198.
Böttcher, C., Keyzers, R.A., Boss, P.K., and Davies, C. (2010). Sequestration of auxin by the indole-3-acetic acid-amido synthetase GH3-1 in grape berry (*Vitis vinifera* L.) and the proposed role of auxin conjugation during ripening. J. Exp. Bot. **61**: 3615–3625.
Boukobza, F. and Taylor, A.J. (2002). Effect of postharvest treatment on flavour volatiles of tomatoes. Postharvest Biol. Technol. **25**: 321–331.
Brooks, C., Nekrasov, V., Lippman, Z.B., and Van Eck, J. (2014). Efficient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-Associated9 System. Plant Physiol. **166**: 1292–1297.
Bruhn, C.M., Feldman, N., Garlitz, C., Harwood, J., Ivans, E., Marshall, M., Riley, A., Thurber, D., and Williamson, E. (1991). Consumer perception of quality: apricots, cantaloupes, peaches, pears, strawberries, and tomatoes. J. Food Qual. **14**: 187–195.
Buck-Koehntop, B.A., Stanfield, R.L., Ekiert, D.C., Martinez-Yamout, M.A., Dyson, H.J., Wilson, I.A., and Wright, P.E. (2012). Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. Proc. Natl. Acad. Sci. U. S. A. **109**: 15229–15234.
Buendía-Monreal, M. and Gillmor, C.S. (2016). Mediator: A key regulator of plant development. Dev. Biol. **419**: 7–18.
Bulger, M. and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell **144**: 327–339.
Burke, T.W. and Kadonaga, J.T. (1996). Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes Dev. **10**: 711–724.
Burke, T.W. and Kadonaga, J.T. (1997). The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAF(II)60 of *Drosophila*. Genes Dev. **11**: 3020–3031.
Buttery, R.G. and Ling, L.C. (1993). Volatile Components of Tomato Fruit and Plant Parts.
de Candolle, A. (1886). The origin of cultivated plants (Cambridge University Press).
Cano-Rodriguez, D., Gjaltema, R.A.F., Jilderda, L.J., Jellema, P., Dokter-Fokkens, J., Ruiters, M.H.J., and Rots, M.G. (2016). Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. Nat. Commun. **7**: 1–11.
Causse, M. (2002). QTL analysis of fruit quality in fresh market tomato: a few chromosome regions control the variation of sensory and instrumental traits. J. Exp. Bot. **53**: 2089–2098.
Causse, M., Friguet, C., Coiret, C., LéPicier, M., Navez, B., Lee, M., Holthuysen, N., Sinesio, F., Moneta, E., and Grandillo, S. (2010). Consumer Preferences for Fresh Tomato at the European Scale: A Common Segmentation on Taste and Firmness. J. Food Sci. **75**: S531–S541.
Causse, M., Saliba-Colombani, V., Lecomte, L., Duffé, P., Rousselle, P., and Buret, M. (2002). QTL analysis of fruit quality in fresh market tomato: A few chromosome regions control the variation of sensory and instrumental traits. J. Exp. Bot. **53**: 2089–2098.
Chalkley, G.E. and Verrijzer, C.P. (1999). DNA binding site selection by RNA polymerase II TAFs: A TAF(II)250-TAF(II)150 complex recognizes the initiator. EMBO J. **18**: 4835–4845.
Chen, K. and Rajewsky, N. (2007). The evolution of gene regulation by transcription factors and microRNAs. Nat. Rev. Genet. **8**: 93–103.
Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science (80-.). **339**: 819–823.
Cunningham, F.X. and Gantt, E. (1998). Genes and enzymes of carotenoid biosyntehsis in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **49**: 557–583.
Das, S. and Bansal, M. (2019). Variation of gene expression in plants is influenced by gene architecture and structural properties of promoters. PLoS One **14**: 1–31.
Deikman, J. (1997). Molecular mechanisms of ethylene regulation of gene transcription. Physiol. Plant. **100**: 561–566.
Deng, W. and Roberts, S.G.E. (2005). A core promoter element downstream of the TATA box that is recognized by TFIIB. Genes Dev. **19**: 2418–2423.
Deriano, L. and Roth, D.B. (2013). Modernizing the nonhomologous end-joining repertoire: Alternative and classical NHEJ share the stage. Annu. Rev. Genet. **47**: 433–455.
Doebley, J.F., Gaut, B.S., and Smith, B.D. (2006). The Molecular Genetics of Crop Domestication. Cell **127**: 1309–1321.
Domcke, S., Bardet, A.F., Adrian Ginno, P., Hartl, D., Burger, L., and Schübeler, D. (2015). Competition between DNA methylation and transcription factors determines binding of NRF1. Nature **528**: 575–579.
Dror, I., Rohs, R., and Mandel-Gutfreund, Y. (2016). How motif environment influences transcription factor search dynamics: Finding a needle in a haystack. BioEssays **38**: 605–612.
Eshed, Y. and Zamir, D. (1995). An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield- associated QTL. Genetics **141**: 1147–1162.
Etienne, A., Génard, M., Lobit, P., Mbeguié-A-Mbéguié, D., and Bugaud, C. (2013). What controls fleshy fruit

acidity? A review of malate and citrate accumulation in fruit cells. J. Exp. Bot. **64**: 1451–1469.

Fernqvist, F. and Hunter, E. (2012). Who's to blame for tasteless tomatoes? the effect of tomato chilling on consumers' taste perceptions. Eur. J. Hortic. Sci. **77**: 193–198.

Frary, A. and Doğanlar, S. (2003). Comparative genetics of crop plant domestication and evolution. Turkish J. Agric. For. **27**: 59–69.

Frary, A., Nesbitt, T.C., Frary, A., Grandillo, S., Van Der Knaap, E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K.B., and Tanksley, S.D. (2000). *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. Science (80-.). **289**: 85–88.

Fridman, E., Carrari, F., Liu, Y.-S., Fernie, A.R., and Zamir, D. (2004). Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science **305**: 1786–1789.

Fridman, E., Liu, Y.S., Carmel-Goren, L., Gur, A., Shoresh, M., Pleban, T., Eshed, Y., and Zamir, D. (2002). Two tightly linked QTLs modify tomato sugar content via different physiological pathways. Mol. Genet. Genomics **266**: 821–826.

Fridman, E., Pleban, T., and Zamir, D. (2000). A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc. Natl. Acad. Sci. U. S. A. **97**: 4718–4723.

Fulton, T.M., Bucheli, P., Voirol, E., López, J., Pétiard, V., and Tanksley, S.D. (2002). Quantitative trait loci (QTL) affecting sugars, organic acids and other biochemical properties possibly contributing to flavour, identified in four advanced backcross populations of tomato. Euphytica **127**: 163–177.

Gaszner, M. and Felsenfeld, G. (2006). Insulators: Exploiting transcriptional and epigenetic mechanisms. Nat. Rev. Genet. **7**: 703–713.

Gebhardt, C. (2016). The historical role of species from the *Solanaceae* plant family in genetic research. Theor. Appl. Genet. **129**: 2281–2294.

Gillaspy, G., Ben-David, H., and Gruissem, W. (1993). Fruits: A developmental perspective. Plant Cell **5**: 1439–1451.

Giovannoni, J.J., DellaPenna, D., Bennett, A.B., and Fischer, R.L. (1989). Expression of a chimeric polygalacturonase gene in transgenic *rin* (ripening inhibitor) tomato fruit results in polyuronide degradation but not fruit softening. Plant Cell **1**: 53.

Gisselbrecht, S.S., Palagi, A., Kurland, J. V., Rogers, J.M., Ozadam, H., Zhan, Y., Dekker, J., and Bulyk, M.L. (2020). Transcriptional Silencers in Drosophila Serve a Dual Role as Transcriptional Enhancers in Alternate Cellular Contexts. Mol. Cell **77**: 324-337.e8.

Gordân, R., Shen, N., Dror, I., Zhou, T., Horton, J., Rohs, R., and Bulyk, M.L. (2013). Genomic Regions Flanking E-Box Binding Sites Influence DNA Binding Specificity of bHLH Transcription Factors through DNA Shape. Cell Rep. **3**: 1093–1104.

Gray, J.E., Picton, S., Giovannoni, J.J., and Grierson, D. (1994). The use of transgenic and naturally occurring mutants to understand and manipulate tomato fruit ripening. Plant, Cell Environ. **17**: 557–571.

Gu, W., Lee, H.C., Chaves, D., Youngman, E.M., Pazour, G.J., Conte, D., and Mello, C.C. (2012). CapSeq and CIP-TAP identify pol ii start sites and reveal capped small RNAs as C. elegans piRNA precursors. Cell **151**: 1488–1500.

Hampsey, M. (1998). Molecular Genetics of the RNA Polymerase II General Transcriptional Machinery. Microbiol. Mol. Biol. Rev. **62**: 465–503.

Harker, F.R., Maindonald, J., Murray, S.H., Gunson, F.A., Hallett, I.C., and Walker, S.B. (2002). Sensory interpretation of instrumental measurements 1: Texture of apple fruit. Postharvest Biol. Technol. **24**: 225–239.

He, G., Elling, A.A., and Deng, X.W. (2011). The epigenome and plant development. Annu. Rev. Plant Biol. **62**: 411–435.

Hendrich, B. and Tweedie, S. (2003). The methyl-CpG binding domain and the evolving role of DNA methylation in animals. Trends Genet. **19**: 269–277.

Holliday, R. and Pugh, J. (1975). DNA modification mechanisms and gene activity during development. Science (80-.). **187**: 226–232.

Hosmani, P.S. et al. (2019). An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. bioRxiv **2012**: 767764.

Hsieh, C.L. (1994). Dependence of transcriptional repression on CpG methylation density. Mol. Cell. Biol. **14**: 5487–5494.

Hu, J., Israeli, A., Ori, N., and Sun, T.P. (2018). The interaction between DELLA and ARF/IAA mediates crosstalk between gibberellin and auxin signaling to control fruit initiation in tomato. Plant Cell **30**: 1710–1728.

Huang, Z. and van der Knaap, E. (2011). Tomato fruit weight 11.3 maps close to fasciated on the bottom of chromosome 11. Theor. Appl. Genet. **123**: 465–474.

Hufford, M.B. et al. (2012). Comparative population genomics of maize domestication and improvement. Nat. Genet. **44**: 808–811.

Hyodo, H., Terao, A., Furukawa, J., Sakamoto, N., Yurimoto, H., Satoh, S., and Iwai, H. (2013). Tissue specific localization of pectin-Ca2+ cross-linkages and pectin methyl-esterification during fruit ripening in

tomato (Solanum lycopersicum). PLoS One **8**: e78949.
Jenkins, J.A. (1948). The origin of the cultivated tomato. Econ. Bot. **2**: 379–392.
Jiang, C. and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: Advances through genomics. Nat. Rev. Genet. **10**: 161–172.
Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009a). H3.3/H2A.Z double variant-containing nucleosomes mark "nucleosome-free regions" of active promoters and other regulatory regions. Nat. Genet. **41**: 941–945.
Jin, Y., Ni, D.-A., and Ruan, Y.-L. (2009b). Posttranslational elevation of cell wall invertase activity by silencing its inhibitor in tomato delays leaf senescence and increases seed weight and fruit hexose level. Plant Cell **21**: 2072–2089.
Jinap, S. and Hajeb, P. (2010). Glutamate. Its applications in food and contribution to health. Appetite **55**: 1–10.
Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science (80-.). **337**: 816–821.
Jolma, A. et al. (2013). DNA-binding specificities of human transcription factors. Cell **152**: 327–339.
Kagey, M.H. et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature **467**: 430–435.
Karlova, R., Chapman, N., David, K., Angenent, G.C., Seymour, G.B., and De Maagd, R.A. (2014). Transcriptional control of fleshy fruit development and ripening. J. Exp. Bot. **65**: 4527–4541.
Karlova, R., Rosin, F.M., Busscher-Lange, J., Parapunova, V., Do, P.T., Fernie, A.R., Fraser, P.D., Baxter, C., Angenent, G.C., and de Maagd, R.A. (2011). Transcriptome and metabolite profiling show that APETALA2a is a major regulator of tomato fruit ripening. Plant Cell **23**: 923–41.
Kaufmann, K. and Mueller-Roeber, B. (2018). Plant Gene Regulatory Networks Methods and Protocols Methods in Molecular Biology K. Kaufmann and B. Mueller-Roeber, eds (Humana Press).
Kaufmann, K., Pajoro, A., and Angenent, G.C. (2010). Regulation of transcription in plants: Mechanisms controlling developmental switches. Nat. Rev. Genet. **11**: 830–842.
Kazeniac, S.J. and Hall, R.M. (1970). Flavour chemistry of tomato volatiles. J. Food Sci. **35**: 519–530.
Kim, J.L., Nikolov, D.B., and Burley, S.K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature **365**: 520–527.
Kimbara, J., Ohyama, A., Chikano, H., Ito, H., Hosoi, K., Negoro, S., Miyatake, K., Yamaguchi, H., Nunome, T., Fukuoka, H., and Hayashi, T. (2018). QTL mapping of fruit nutritional and flavour components in tomato (*Solanum lycopersicum*) using genome-wide SSR markers and recombinant inbred lines (RILs) from an intra-specific cross. Euphytica **214**: 1–12.
Kimura, S. and Sinha, N. (2008). Tomato (*Solanum lycopersicum*): A model fruit-bearing crop. Cold Spring Harb. Protoc. **3**: 1–10.
Kitagawa, M., Ito, H., Shiina, T., Nakamura, N., Inakuma, T., Kasumi, T., Ishiguro, Y., Yabe, K., and Ito, Y. (2005). Characterization of tomato fruit ripening and analysis of gene expression in F_1 hybrids of the *ripening inhibitor* (*rin*) mutant. Physiol. Plant. **123**: 331–338.
Klee, H.J. and Tieman, D.M. (2018). The genetics of fruit flavour preferences. Nat. Rev. Genet. **19**: 347–356.
Klose, R.J. and Bird, A.P. (2006). Genomic DNA methylation: The mark and its mediators. Trends Biochem. Sci. **31**: 89–97.
van der Knaap, E., Chakrabarti, M., Chu, Y.H., Clevenger, J.P., Illa-Berenguer, E., Huang, Z., Keyhaninejad, N., Mu, Q., Sun, L., Wang, Y., and Wu, S. (2014). What lies beyond the eye: The molecular mechanisms regulating tomato fruit weight and shape. Front. Plant Sci. **5**: 1–13.
Knapp, S. and Peralta, I.E. (2016). The Tomato (*Solanum lycopersicum* L., Solanaceae) and Its Botanical Relatives (Springer, Berlin, Heidelberg).
Knoll, A., Fauser, F., and Puchta, H. (2014). DNA recombination in somatic plant cells: Mechanisms and evolutionary consequences. Chromosom. Res. **22**: 191–201.
Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., and Ebright, R.H. (1998). New core promoter element in RNA polymerase II-dependent transcription: Sequence-specific DNA binding by transcription factor IIB. Genes Dev. **12**: 34–44.
Lanahan, M.B., Yen Hsiao Ching, Giovannoni, J.J., and Klee, H.J. (1994). The *Never ripe* mutation blocks ethylene perception in tomato. Plant Cell **6**: 521–530.
Lashbrook, C.C., Gonzalez-Bosch, C., and Bennett, A.B. (1994). Two divergent endo-b-1,4-glucanase genes exhibit overlapping expression in ripening fruit and abscising flowers. Plant Cell **6**: 1485.
Latchman, D.S. (1997). Transcription factors: An overview. Int. J. Biochem. Cell Biol. **29**: 1305–1312.
Lazarovici, A., Zhou, T., Shafer, A., MacHado, A.C.D., Riley, T.R., Sandstrom, R., Sabo, P.J., Lu, Y., Rohs, R., Stamatoyannopoulos, J.A., and Bussemaker, H.J. (2013). Probing DNA shape and methylation state on a genomic scale with DNase i. Proc. Natl. Acad. Sci. U. S. A. **110**: 6376–6381.
Lelievre, J.M., Latche, A., Jones, B., Bouzayen, M., and Pech, J.-C. (1997). Ethylene and fuit ripening. Physiol. Plant. **101**: 727–739.
Lewinsohn, E., Sitrit, Y., Bar, E., Azulay, Y., Meir, A., Zamir, D., and Tadmor, Y. (2005). Carotenoid

pigmentation affects the volatile composition of tomato and watermelon fruits, as revealed by comparative genetic analyses. J. Agric. Food Chem. **53**: 3142–3148.

Li, J.-F., Norville, J.E., Aach, J., McCormack, M., Zhang, D., Bush, J., Church, G.M., and Sheen, J. (2013). Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nat. Biotechnol. **31**: 688–91.

Lilac Pnueli, Lea Carmel-Goren, Dana Hareven, Tamar Gutfinger, John Alvarez, Martin Ganal, Daniel Zamir, and Eliezer Lifschitz (1998). The SELF-PRUNING gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of *CEN* and *TFL1*. Development **125**: 1979–1989.

Lin, T. et al. (2014). Genomic analyses provide insights into the history of tomato breeding. Nat. Genet. **46**: 1220–1226.

Lu, Z., Marand, A.P., Ricci, W.A., Ethridge, C.L., Zhang, X., and Schmitz, R.J. (2019). The prevalence, evolution and chromatin signatures of plant regulatory elements. Nat. Plants **5**: 1250–1259.

Luscombe, N.M., Austin, S.E., Berman, H.M., and Thornton, J.M. (2000). An overview of the structures of protein-DNA complexes. Genome Biol. **1**: 1–37.

Magnani, L., Eeckhoute, J., and Lupien, M. (2011). Pioneer factors: Directing transcriptional regulators within the chromatin environment. Trends Genet. **27**: 465–474.

Mahajan, P. V., Caleb, O.J., Singh, Z., Watkins, C.B., and Geyer, M. (2014). Postharvest treatments of fresh produce. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. **372**: 20130309.

Maher, K.A. et al. (2018). Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules. Plant Cell **30**: 15–36.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013a). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. **31**: 833–838.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013b). RNA-guided human genome engineering via Cas9. Science (80-.). **339**: 823–826.

Mathelier, A., Xin, B., Chiu, T.P., Yang, L., Rohs, R., and Wasserman, W.W. (2016). DNA Shape Features Improve Transcription Factor Binding Site Predictions In Vivo. Cell Syst. **3**: 278-286.e4.

Maul, F., Sargent, S.A., Sims, C.A., Baldwin, E.A., Balaban, M.O., and Huber, D.J. (2000). Tomato flavour and aroma quality as affected by storage temperature. J. Food Sci. **65**: 1228–1237.

Maurano, M.T., Wang, H., John, S., Shafer, A., Canfield, T., Lee, K., and Stamatoyannopoulos, J.A. (2015). Role of DNA Methylation in Modulating Transcription Factor Occupancy. Cell Rep. **12**: 1184–1195.

Mavrich, T.N. et al. (2008). Nucleosome organization in the *Drosophila* genome. Nature **453**: 358–362.

Mejia-Guerra, M.K., Pomeranz, M., Morohashi, K., and Grotewold, E. (2012). From plant gene regulatory grids to network dynamics. Biochim. Biophys. Acta - Gene Regul. Mech. **1819**: 454–465.

Meyer, C.A., Tang, Q., and Liu, X.S. (2012). Minireview: Applications of next-generation sequencing on studies of nuclear receptor regulation and function. Mol. Endocrinol. **26**: 1651–1659.

Meyer, R.S. and Purugganan, M.D. (2013). Evolution of crop species: Genetics of domestication and diversification. Nat. Rev. Genet. **14**: 840–852.

Miller, J.C. and Tanksley, S.D. (1990). RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor. Appl. Genet. **80**: 437–448.

Mueller, B., Mieczkowski, J., Kundu, S., Wang, P., Sadreyev, R., Tolstorukov, M.Y., and Kingston, R.E. (2017). Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction. Genes Dev. **31**: 451–462.

Narlikar, G.J., Fan, H.Y., and Kingston, R.E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell **108**: 475–487.

Ni, T., Corcoran, D.L., Rach, E.A., Song, S., Spana, E.P., Gao, Y., Ohler, U., and Zhu, J. (2010). A paired-end sequencing strategy to map the complex landscape of transcription initiation. Nat. Methods **7**: 521–527.

O'Malley, R.C., Huang, S.S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell **165**: 1280–1292.

Oltman, A.E., Jervis, S.M., and Drake, M.A. (2014). Consumer attitudes and preferences for fresh market tomatoes. J. Food Sci. **79**: S2091–S2097.

Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. Genes Dev. **10**: 2657–2683.

van Overbeek, M. et al. (2016). DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. Mol. Cell **63**: 633–646.

Patikoglou, G.A., Kim, J.L., Sun, L., Yang, S.H., Kodadek, T., and Burley, S.K. (1999). TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. Genes Dev. **13**: 3217–3230.

Peralta, I.E., Knapp, S., and Spooner, D.M. (2005). New species of wild tomatoes (*Solanum Section Lycopersicon: Solanaceae*) from Northern Peru. Syst. Bot. **30**: 424–434.

Peralta, I.E., Spooner, D.M., and Knapp, S. (2008). Taxonomy of wild tomatoes and their relatives (*Solanum sect. Lycopersicoides, sect. Juglandifolia, sect. Lycopersicon; Solanaceae*). Syst. Bot. Monogr. **84**: 1–

186.
Petreikov, M., Shen, S., Yeselson, Y., Levin, I., Bar, M., and Schaffer, A.A. (2006). Temporally extended gene expression of the *ADP-Glc pyrophosphorylase large subunit (AgpL1)* leads to increased enzyme activity in developing tomato fruit. Planta **224**: 1465–1479.
Petró-Turza, M. (1986). Flavour of tomato and tomato products. Food Rev. Int. **2**: 309–351.
Powell, A.L.T. et al. (2012). Uniform ripening encodes a Golden 2-like transcription factor regulating tomato fruit chloroplast development. Science (80-.). **336**: 1711–1715.
Raffo, A., Nicoli, S., Nardo, N., Baiamonte, I., Daloise, A., and Paoletti, F. (2012). Impact of different distribution scenarios and recommended storage conditions on flavour related quality attributes in ripening fresh tomatoes. J. Agric. Food Chem. **60**: 10445–10455.
Rambla, J.L., Tikunov, Y.M., Monforte, A.J., Bovy, A.G., and Granell, A. (2014). The expanded tomato fruit volatile landscape. J. Exp. Bot. **65**: 4613–4623.
Razifard, H., Ramos, A., Della Valle, A.L., Bodary, C., Goetz, E., Manser, E.J., Li, X., Zhang, L., Visa, S., Tieman, D., van der Knaap, E., and Caicedo, A.L. (2020). Genomic Evidence for Complex Domestication History of the Cultivated Tomato in Latin America. Mol. Biol. Evol. **37**: 1118–1132.
Renard, C.M.G.C., Ginies, C., Gouble, B., Bureau, S., and Causse, M. (2013). Home conservation strategies for tomato (*Solanum lycopersicum*): Storage temperature vs. duration-Is there a compromise for better aroma preservation? Food Chem. **139**: 825–836.
Ricci, W.A. et al. (2019). Widespread long-range cis-regulatory elements in the maize genome. Nat. Plants **5**: 1237–1249.
Riggs, A.D. (1975). Inactivation, differentiation, and DNA methylation. Cytogenet. Genome Res. **14**: 9–25.
Robinson, R.W. and Tomes, M.L. (1968). Ripening inhibitor: a gene with multiple effect on ripening. Tomato Genet. Coop. **18**: 36-37.
Rodríguez-Leal, D., Lemmon, Z.H., Man, J., Bartlett, M.E., and Lippman, Z.B. (2017). Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing. Cell **171**: 470-480.e8.
Ron, M. et al. (2014). Hairy root transformation using Agrobacterium rhizogenes as a tool for exploring cell type-specific gene expression and function using tomato as a model. Plant Physiol. **166**: 455–69.
Roohanitaziani, R. (2019). Genetic analysis of fruit quality in tomato (Wageningen University, Doctoral dissertation).
Roohanitaziani, R., de Maagd, R.A., Lammers, M., Molthoff, J., Meijer-Dekens, F., Kaauwen, M.P.W. van, Finkers, R., Tikunov, Y., Visser, R.G.F., and Bovy, A.G. (2020). Exploration of a Resequenced Tomato Core Collection for Phenotypic and Genotypic Variation in Plant Growth and Fruit Quality Traits. Genes (Basel). **11**: 1278.
Sanchez-Bel, P., Egea, I., Sanchez-Ballesta, M.T., Sevillano, L., Del Carmen Bolarin, M., and Flores, F.B. (2012). Proteome changes in tomato fruits prior to visible symptoms of chilling injury are linked to defensive mechanisms, uncoupling of photosynthetic processes and protein degradation machinery. Plant Cell Physiol. **53**: 470–484.
Sandarani, M., Dasanayaka, D., and Jayasinghe, C. (2018). Strategies Used to Prolong the Shelf Life of Fresh Commodities. J. Agric. Sci. Food Res. **9**: 1–6.
Sauvage, C., Segura, V., Bauchet, G., Stevens, R., Do, P.T., Nikoloski, Z., Fernie, A.R., and Causse, M. (2014). Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. Plant Physiol. **165**: 1120–1132.
Sayou, C. et al. (2016). A SAM oligomerization domain shapes the genomic binding landscape of the LEAFY transcription factor. Nat. Commun. **7**: 1–12.
Schaffer, a. a. and Petreikov, M. (1997). Sucrose-to-Starch Metabolism in Tomato Fruit Undergoing Transient Starch Accumulation. Plant Physiol. **113**: 739–746.
Schaffer, A.A., Levin, I., Oguz, I., Petreikov, M., Cincarevsky, F., Yeselson, Y., Shen, S., Gilboa, N., and Bar, M. (2000). ADPglucose pyrophosphorylase activity and starch accumulation in immature tomato fruit: The effect of a Lycopersicon hirsutum-derived introgression encoding for the large subunit. Plant Sci. **152**: 135–144.
Schiml, S., Fauser, F., and Puchta, H. (2014). The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in Arabidopsis resulting in heritable progeny. Plant J. **80**: 1139–1150.
Schouten, H.J., Tikunov, Y., Verkerke, W., Finkers, R., Bovy, A., Bai, Y., and Visser, R.G.F. (2019). Breeding Has Increased the Diversity of Cultivated Tomato in The Netherlands. Front. Plant Sci. **10**: 1606.
Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., Qiu, J.L., and Gao, C. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. Nat. Biotechnol. **31**: 686–688.
Shi, X., Shou, J., Mehryar, M.M., Li, J., Wang, L., Zhang, M., Huang, H., Sun, X., and Wu, Q. (2019). Cas9 has no exonuclease activity resulting in staggered cleavage with overhangs and predictable di- and tri-nucleotide CRISPR insertions without template donor. Cell Discov. **5**: 53.
Shneour, E.A. and Zabin, I. (1959). The biosynthesis of lycopene in tomato homogenates. J. Biol. Chem. **234**:

770–773.
Shukla, V.K. et al. (2009). Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature **459**: 437–441.
Smale, S.T. and Baltimore, D. (1989). The "initiator" as a transcription control element. Cell **57**: 103–113.
Somssich, M., Je, B. Il, Simon, R., and Jackson, D. (2016). CLAVATA-WUSCHEL signaling in the shoot meristem. Dev. **143**: 3238–3248.
Sorrequieta, A., Ferraro, G., Boggio, S.B., and Valle, E.M. (2010). Free amino acid production during tomato fruit ripening: A focus on L-glutamate. Amino Acids **38**: 1523–1532.
Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and Zaret, K.S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell **161**: 555–568.
Spitz, F. and Furlong, E.E.M. (2012). Transcription factors: From enhancer binding to developmental control. Nat. Rev. Genet. **13**: 613–626.
Suzuki, M.M. and Bird, A. (2008). DNA methylation landscapes: Provocative insights from epigenomics. Nat. Rev. Genet. **9**: 465–476.
Swinnen, G., Goossens, A., and Pauwels, L. (2016). Lessons from Domestication: Targeting Cis-Regulatory Elements for Crop Improvement. Trends Plant Sci. **21**: 506–515.
Symington, L.S. and Gautier, J. (2011). Double-strand break end resection and repair pathway choice. Annu. Rev. Genet. **45**: 247–271.
Tadmor, Y., Fridman, E., Gur, A., Larkov, O., Lastochkin, E., Ravid, U., Zamir, D., and Lewinsohn, E. (2002). Identification of malodorous, a wild species allele affecting tomato aroma that was selected against during domestication. J. Agric. Food Chem. **50**: 2005–2009.
Tan, J. et al. (2020). Efficient CRISPR/Cas9-based plant genomic fragment deletions by microhomology-mediated end joining. Plant Biotechnol. J.: 1–3.
Tan, M. et al. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell **146**: 1016–1028.
Tanay, A. (2006). Extensive low-affinity transcriptional interactions in the yeast genome. Genome Res. **16**: 962–972.
Tanksley, S.D., Grandillo, S., Fulton Zamir, T.D., Eshed, Y., Petiard -J Lopez Beck-Bunn, V.T., Zamir, D., Eshed, T., and Lopez Hispareco, J.S. (1996). Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium* (Springer-Verlag).
Tao, Z., Shen, L., Liu, C., Liu, L., Yan, Y., and Yu, H. (2012). Genome-wide identification of SOC1 and SVP targets during the floral transition in Arabidopsis. Plant J. **70**: 549–561.
Terhorst, P. (2006). The scaling of the Dutch vegetables-under-glass cluster: Sweet peppers, tomatoes and cucumbers. Tijdschr. voor Econ. en Soc. Geogr. **97**: 434–442.
The Tomato Genome Consortium et al. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**: 635–41.
Thompson, A.J., Tor, M., Barry, C.S., Vrebalov, J., Orfila, C., Jarvis, M.C., Giovannoni, J.J., Grierson, D., and Seymour, G.B. (1999). Molecular and genetic characterization of a novel pleiotropic tomato-ripening mutant. Plant Physiol. **120**: 383–389.
Tieman, D. et al. (2012). The chemical interactions underlying tomato flavour preferences. Curr. Biol. **22**: 1035–1039.
Tieman, D. et al. (2017). A chemical genetic roadmap to improved tomato flavour. Science (80-.). **355**: 391–394.
Tigchelaar, E., Tomes, M., Kerr, E., and Barman, R. (1973). A new fruit ripening mutant, non-ripening (nor). Rep Tomato Genet Coop **23**: 33–34.
Tippin, D.B., Ramakrishnan, B., and Sundaralingam, M. (1997). Methylation of the Z-DNA decamer d(GC)5 potentiates the formation of A-DNA: Crystal structure of d(Gm5CGm5CGCGCGC). J. Mol. Biol. **270**: 247–258.
Uluisik, S. et al. (2016). Genetic improvement of tomato by targeted control of fruit softening. Nat. Biotechnol. **34**: 950–952.
Varga, A. (1976). Roles of Seeds and Auxins in Tomato Fruit Growth. Zeitschrift für Pflanzenphysiologie **80**: 95–104.
Vogel, J.T., Tieman, D.M., Sims, C.A., Odabasi, A.Z., Clark, D.G., and Klee, H.J. (2010). Carotenoid content impacts flavour acceptability in tomato (*Solanum lycopersicum*). J. Sci. Food Agric. **90**: 2233–2240.
Völkel, S., Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., and Suske, G. (2015). Zinc Finger Independent Genome-Wide Binding of Sp2 Potentiates Recruitment of Histone-Fold Protein Nf-y Distinguishing It from Sp1 and Sp3. PLOS Genet. **11**: e1005102.
Voytas, D.F. and Gao, C. (2014). Precision genome engineering and agriculture: opportunities and regulatory challenges. PLoS Biol. **12**: e1001877.
Wade, P.A. (2001). Methyl CpG-binding proteins and transcriptional repression. BioEssays **23**: 1131–1137.
Wang, D. and Seymour, G.B. (2017). Tomato Flavour: Lost and Found? Mol. Plant **10**: 782–784.
Wang, L., Koppitch, K., Cutting, A., Dong, P., Kudtarkar, P., Zheng, J., Cameron, R.A., and Davidson, E.H.

(2019a). Developmental effector gene regulation: Multiplexed strategies for functional analysis. Dev. Biol. **445**: 68–79.

Wang, R., Tavano, E.C. da R., Lammers, M., Martinelli, A.P., Angenent, G.C., and de Maagd, R.A. (2019b). Re-evaluation of transcription factor function in tomato fruit development and ripening with CRISPR/Cas9-mutagenesis. Sci. Rep. **9**: 1696.

Weirauch, M.T. et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell **158**: 1431–1443.

West, A.G., Shore, P., and Sharrocks, A.D. (1997). DNA binding by MADS-box transcription factors: a molecular mechanism for differential DNA bending. Mol. Cell. Biol. **17**: 2876–2887.

Wilczyski, B. and Furlong, E.E.M. (2010). Dynamic CRM occupancy reflects a temporal map of developmental progression. Mol. Syst. Biol. **6**: 383.

Wilkinson, J.Q., Lanahan, M.B., Yen, H.C., Giovannoni, J.J., and Klee, H.J. (1995). An ethylene-inducible component of signal transduction encoded by <I>Never-ripe</i>. Science (80-.). **270**: 1807.

Wittkopp, P.J. and Kalay, G. (2012). Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. Nat. Rev. Genet. **13**: 59–69.

Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J., and Van Der Knaap, E. (2008). A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science (80-.). **319**: 1527–1530.

Xu, C. et al. (2015). A cascade of arabinosyltransferases controls shoot meristem size in tomato. Nat. Genet. **47**: 784–792.

Yamamoto, Y., Ichida, H., Abe, T., Suzuki, Y., Sugano, S., and Obokata, J. (2007). Differentiation of core promoter architecture between plants and mammals revealed by LDSS analysis. Nucleic Acids Res. **35**: 6219–6226.

Yamamoto, Y.Y., Yoshitsugu, T., Sakurai, T., Seki, M., Shinozaki, K., and Obokata, J. (2009). Heterogeneity of *Arabidopsis* core promoters revealed by high-density TSS analysis. Plant J. **60**: 350–362.

Yáñez-Cuna, J.O., Kvon, E.Z., and Stark, A. (2013). Deciphering the transcriptional cis-regulatory code. Trends Genet. **29**: 11–22.

Ye, J., Wang, X., Hu, T. xu, Zhang, F. xia, Wang, B., Li, C. xin, Yang, T. xia, Li, H. xia, Lu, Y. en, Giovannoni, J.J., Zhang, Y., and Ye, Z. (2017). An InDel in the Promoter of *Al-ACTIVATED MALATE TRANSPORTER9* Selected during Tomato Domestication Determines Fruit Malate Contents and Aluminum Tolerance. Plant Cell **29**: tpc.00211.2017.

Yeager, A.F. (1927). Determinate growth in the tomato. J. Hered. **18**: 263–265.

Yilmaz, E. (2001). Chemistry and Evaluation of Tomato Flavour. Turkish J. Agric. For. **25**: 149–155.

Yokotani, N., Nakano, R., Imanishi, S., Nagata, M., Inaba, A., and Kubo, Y. (2009). Ripening-associated ethylene biosynthesis in tomato fruit is autocatalytically and developmentally regulated. J. Exp. Bot. **60**: 3433–3442.

Yu, C.P., Lin, J.J., and Li, W.H. (2016). Positional distribution of transcription factor binding sites in *Arabidopsis thaliana*. Sci. Rep. **6**.

Zanor, M.I. et al. (2009). RNA interference of *LIN5* in tomato confirms its role in controlling brix content, uncovers the influence of sugars on the levels of fruit hormones, and demonstrates the importance of sucrose cleavage for normal fruit development and fertility. Plant Physiol. **150**: 1204–1218.

Zhang, B., Tieman, D.M., Jiao, C., Xu, Y., Chen, K., Fe, Z., Giovannoni, J.J., and Klee, H.J. (2016). Chilling-induced tomato flavour loss is associated with altered volatile synthesis and transient changes in DNA methylation. Proc. Natl. Acad. Sci. **113**: 12580–12585.

Zhang, H., Lang, Z., and Zhu, J.K. (2018). Dynamics and function of DNA methylation in plants. Nat. Rev. Mol. Cell Biol. **19**: 489–506.

Zhang, M., Kimatu, J.N., Xu, K., and Liu, B. (2010). DNA cytosine methylation in plant development. J. Genet. Genomics **37**: 1–12.

Zhao, J., Sauvage, C., Zhao, J., Bitton, F., Bauchet, G., Liu, D., Huang, S., Tieman, D.M., Klee, H.J., and Causse, M. (2019). Meta-analysis of genome-wide association studies provides insights into genetic control of tomato flavour. Nat. Commun. **10**: 1–12.

Zhu, F., Farnung, L., Kaasinen, E., Sahu, B., Yin, Y., Wei, B., Dodonova, S.O., Nitta, K.R., Morgunova, E., Taipale, M., Cramer, P., and Taipale, J. (2018a). The interaction landscape between transcription factors and the nucleosome. Nature **562**: 76–81.

Zhu, G. et al. (2018b). Rewiring of the Fruit Metabolome in Tomato Breeding. Cell **172**: 249-261.e12.

CHAPTER 3

Increased *AGPL1* expression and Brix in tomato fruit by targeting the promoter with CRISPR/Cas9-mutagenesis

Vera Veltkamp[1,2], Douwe Zantinge, Julia Ruiz-Capella, Moniek Schippers, Marrit Alderkamp, Christine de Vries, Gerco C. Angenent[1,2] and Ruud A. de Maagd[2]

[1]Laboratory of Molecular Biology, Wageningen University, The Netherlands

[2]Bioscience, Wageningen Plant Research, The Netherlands

ABSTRACT

Through the study of an introgression from a wild relative into cultivated tomato, ADP-glucose pyrophorylase (AGPase) activity has been shown to be a rate-limiting factor in starch accumulation and final sugar content of tomato fruit. Increasing AGPase activity by raising expression of the large AGPase subunit (*AGPL1*) gene could result in higher sugar content, an important trait that is appreciated by consumers and tomato processors. The regulation of *AGPL1* expression was studied to understand the regulatory mechanism and to increase expression. A first *in silico* analysis resulted in the identification of conserved and accessible promoter-regions. Interacting Transcription Factors (TFs) were identified with Yeast-one-hybrid (Y1H) assays and validated with Promoter Reporter Assays. Two potential repressors as well as several activators were found. To modulate expression *in vivo* and create plants with a higher sugar content, multiplexed mutagenesis with CRISPR/Cas9 was used on the *AGPL1* promoter in tomato cv. Moneyberg aiming at the removal of repressing Cis-regulatory Elements (CREs). In several of the obtained mutant lines, soluble solids content ($^{\circ}$Brix) was increased, however always correlated with decreased fruit weight. For two out of three lines with the highest Brix, an increase of *AGPL1* expression was demonstrated.

INTRODUCTION

Sugar content is an essential aspect of tomato quality. Sugars, the product of photosynthesis, are transported as sucrose from source tissues, such as leaves, to sink tissues, such as developing fruit. Sucrose is converted and stored as glucose and fructose in the vacuole of fruit parenchyma cells or as starch in the plastids, thus maintaining sink strength (Stein and Granot, 2018).

Transient starch storage is an important factor in determining tomato sugar content. Immature tomatoes have a starch content of approximately 20% of their dry weight (Ho et al., 1982) and the stored starch is mainly located in the pericarp, septa and columella (Schaffer and Petreikov, 1997a). While the fruits mature, starch content increases and subsequently, during ripening, starch is converted to glucose and maltose. In the ripe fruit, starch is virtually absent. Glucose and maltose are exported to the cytoplasm where maltose is converted to glucose, increasing the tomato sweetness-perception (Ho et al., 1982; Wang et al., 1993; Schaffer and Petreikov, 1997a; Luo et al., 2020). Thus, an interesting breeding goal is to increase the transient starch storage with the expectation that it will lead to higher sugar content in ripe fruit.

Several enzymes are involved in the metabolism of sucrose to starch, such as sucrose synthase, starch synthase and ADP-glucose pyrosphorylase (AGPase) (Schaffer and Petreikov, 1997a, 1997b). AGPase activity was shown to be the rate-limiting factor in starch accumulation (Schaffer et al., 2000; Ikeda et al., 2016). Thus, AGPase is a prime

candidate for manipulation to increase transient starch accumulation and final sugar content. In the plastids, the tetrameric complex AGPase catalyses the reaction between ATP and Glucose-1-Phosphate to produce ADP-glucose, which is converted to starch by downstream enzymes, and inorganic pyrophosphate (PPi) (Chen and Janes, 1997; Beckles et al., 2001; Ballicora et al., 2004). The small catalytic subunit of the AGPase complex is encoded by *SlAGPS* (also called *AgpB*) (Crevillén et al., 2003; Goto et al., 2013), and three genes, *AGPL1, AGPL2* and *AGPL3* (Also called *AgpS1, S2, S3*), encode the large, stabilising subunit. Of these three, *AGPL1* is the main one expressed in tomato fruit (Chen et al., 1998; Park and Chung, 1998; Xing et al., 2005; Petreikov et al., 2010). Additionally, *AGPL1* is expressed in roots (Chen et al., 1998). An *S. habrochaites* allele of *AGPL1* (*AGPL1^{H}*) present in an introgression line causes increased AGPase activity and starch levels in young fruit, and higher sugar content in ripe fruit (Schaffer et al., 2000). The increased activity of AGPase in plants with the *AGPL1^{H}* allele was due to the prolonged expression of *AGPL1* during immature fruit development and increased stability of the complex, not to altered intrinsic enzyme activity (Petreikov et al., 2006, 2009). There were multiple sequence differences in the *AGPL1^{H}* promoter compared to the *S. lycopersicum* promoter, which are leads for understanding and eventually modifying the transcriptional regulation of *AGPL1*. Increasing or prolonging *AGPL1* expression and concurrently AGPase activity would increase starch content and with it the final sugar content, an agronomically valuable phenotype. To understand the regulation of *AGPL1* transcriptional activity and harness this knowledge for increasing expression, we set out to study the regulation of AGPase activity and *AGPL1* expression.

Sequestering of sugars prevents the sucrose-mediated downregulation of sugar processing enzymes like fructokinase, sucrose synthase, UDP-glucose pyrophosphorylase (UGPase) and AGPase (Du Jardin et al., 1997; Schaffer and Petreikov, 1997a; Li et al., 2002; Decker, 2017). Known regulators of the transcription of the *AGPase* genes are sucrose, phosphate and nitrate (Du Jardin et al., 1997; Li et al., 2002). In watermelon (*Citrullus lanatus*), the *AGPL1* promoter contains the TCCAAAA-element that results in repression of *AGPL1* in leaves (Yin et al., 2009). In *S. lycopersicum*, this element is present in the 5' UTR. On the other hand, in *Arabidopsis,* sweet potato (*Ipomoea batatas*) and carrot (*Daucus carota L.*) storage tissues, *AGPL1* expression is positively correlated with endogenous sucrose levels, suggesting a positive feedback loop (Müller-Röber et al., 1990; Geigenberger et al., 1994; Takeda et al., 1994; Nakata and Okita, 1995; Li et al., 2002; Crevillén et al., 2005; Kwak et al., 2006). In tomato, it was shown using truncated versions of the promoter that a region 1.3–3.0 kb upstream of the transcription start site is responsible for sucrose sensitivity (Li et al., 2002). AGPase activity is negatively correlated with malate levels. The malate influenced the redox state of the plastid, causing alterations in the activation of AGPase (Centeno et al., 2011). A lack of starch in young fruits resulted in a decreased accumulation of soluble sugars, elevated water loss, increased wrinkling and decreased resistance to post-harvest pathogens (Centeno et al., 2011).

This study aimed to elucidate the regulation of *AGPL1* further by studying the promoter using a variety of methods. The goal was to increase expression and create a tomato with a higher Brix.

RESULTS

TWO PROMOTER REGIONS HAVE POTENTIAL FOR HARBOURING *CIS*-REGULATORY ELEMENTS

The upstream region of *ADP-glucose pyrophosphorylase Large Subunit 1* (*AGPL1*) was profiled to identify areas that may be relevant for controlling gene expression. The first neighbouring gene is 21.2 kb upstream of *AGPL1*, in the opposite orientation. Thus, this 21.2 kb non-coding region most likely contains the promoters of both genes. To narrow the scope of our investigation, we initially looked at the 10 kb upstream of *AGPL1*. The FruitEncode project (Lu et al., 2017) allowed us to determine Open Chromatin Regions (OCRs) from published DNase I hypersensitive sites sequencing (DNAse-seq) data, identifying DNAse I-hypersensitive areas. OCRs represents regions that are accessible to Transcription Factors (TFs) and may contain active *Cis*-Regulatory Elements (CREs) at a given developmental stage (Boyle et al., 2008; Zhang et al., 2012, 2016; Jiang, 2015). Four significant OCRs were found in the 10 kb upstream region of *AGPL1* (OCR1-4, **Figure S1**). OCR1 (closest to the *AGPL1,* approximately 100 bp upstream of the transcription start site), had the highest significance and was the widest. Both OCR1 and OCR2 (±2.5 kb upstream of the transcription start site) had the highest DNAse I hypersensitivity in fruit development compared to leaf, indicating that they might be important for CRE-activity during fruit development. The DNAse I hypersensitive sites were alternating with highly methylated regions (**Figure S1**).

Next, sequence conservation analysis was performed to determine which regions are evolutionary conserved. Upstream regions of *AGPL1* orthologs from wild tomato relatives, potato, pepper and tobacco were subjected to mVISTA analysis (**Table S1**) (Frazer et al., 2004). One highly conserved region corresponded with OCR1 (**Figure 1**). OCR2 aligned with a region that was conserved up to pepper and less in the *Nicotiana* species. OCR3 is conserved in tomato and potato, and OCR4 only in tomato. Taken together, the results suggest that OCR1, and to a lesser extend OCR2 harbour conserved CREs. Other highly conserved promoter regions were also present in the *Solanum* spp. Although not part of an OCR, these could still be relevant, but were not further investigated here.

FIVE INTERACTING TFS WERE IDENTIFIED

Having defined OCR1 as the most likely CRE-harbouring region in the *AGPL1* promoter, we then identified potentially interacting TFs. A Yeast-one-hybrid assay (Y1H) was performed to identify candidates. Two different Y1H methods were applied to screen large numbers of TF candidates. First, a matrix-based TF library screen was used with a

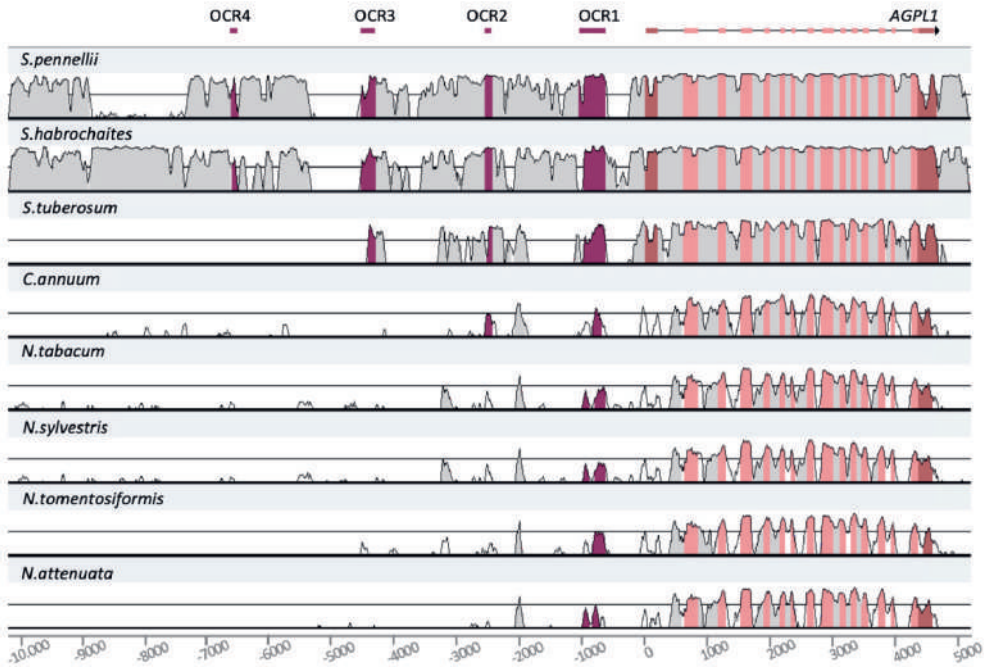


Figure 1: Conservation analysis of the AGPL1 promoter with mVISTA. The upper row represents the gene model of *AGPL1*, which is used as the reference point for the non-coding adjacent regions. Rows were ordered by decreasing homology. Conservation was determined by LAGAN, a global multiple alignment tool, with a RankVISTA cut-off probability value of 0.05 and a window size of 100 bp. If a particular region is above the pre-determined cut-off value, then the region is represented by the presence of colour: pink and dark pink regions represent exons and UTRs, grey represents non-coding regions and purple regions are the previously determined OCRs **(Figure S1).** Position numbering is relative to the Transcription Start Site.

collection containing more than 2000 *Arabidopsis* TFs fused to the GAL4-activation domain (Castrillo et al., 2011). A promoter-reporter construct was made of a 340 bp region (-1234 to -895 bp from the Transcription Start Site) from the *AGPL1* promoter. This specific region was chosen as it displayed the largest difference between the *S. lycopersicum promoter (pAGPL1^{M})* and *S. habrochaites* promoter (*pAGPL1^{H}*) in the first 2 kb upstream of the transcription start site (**Figure S2a**). Compared to the pAGPL1^{M}, in that region, a stretch of 77 bp is missing in *pAGPL1^{H}*, and several Single Nucleotide Polymorphisms (SNPs) occur. Our hypothesis is that since the *AGPL1^{H}* allele was shown to be expressed at higher levels and more prolonged during fruit development (Schaffer et al., 2000), the 77 bp in *pAGPL1^{M}* could harbour repressing CREs. However, an effect of other differences cannot be excluded. In addition, this region is partially overlapping with OCR1.

A Y1H screen was performed with the TF library, and positive hits were re-evaluated twice. "Sticky" false positive TFs were excluded by comparing our screen with an independent screen done with a completely different promoter and excluding those TFs based on a Welch's T-test (Ouwerkerk and Meijer, 2001; Mitsuda et al., 2010). Eight

interacting TFs were found for *pAGPL1*^{M} and four for *pAGPL1*^{H}, of which two were found for both (**Table 1**). For all hits, the tomato ortholog was found by best reciprocal BLAST hits search (**Table 1**). Based on expression in fruit and co-expression with *AGPL1*, a selection of six tomato TFs was made for a second Y1H experiment (**Figure 2a**). CONSTANS-like 1 (CO1) and FRUITFULL 1 (FUL1) did not show interaction with either *pAGPL1* promoters. In contrast, the homolog of the *Arabidopsis* B-box zinc-finger TF, BBX19, FRUITFULL 2 (FUL2) and JASMONIC ACID 1 (JA1) showed interaction in both promoter allelic variants. Interestingly, the CCCH-type Zinc Finger C3H13, interacted with *pAGPL1*^{M} and not with *pAGPL1*^{H}. For JA1, a potential binding motif was available, based on the *Arabidopsis* ortholog *At1G69780* (**Figure S2b**) (Jin et al., 2017). A match for this motif was present in the tested region of both *pAGPL1*^{M} and *pAGPL1*^{H} (CAAATAATTTA ($p<1.10^{-4}$), Grant et al., 2011).

Table 1: Y1H with the A. thaliana TF library results and subsequent S. lycopersicum ortholog selection

Y1H with *A. thaliana* TF library							TF selection for a Y1H with *S. lycopersicum* TFs			
Transcription Factor			Promoter reporter				Potential candidates			
A. thaliana			*S. lycopersicum*		*S. habrochaites*		*S. lycopersicum*			
Gene ID	alias	TF family	Sig.	175 225	Sig.	175 225	Gene ID	alias	Pearson	Av. expression
AT5G53210	SPCH	bHLH			*		Solyc03g007410	bHLH	*<0.65*	*0.1*
AT4G11250	AGL52	MADS	*		*		Solyc11g020320	MADS	*<0.65*	*0*
AT3G30260	AGL79	MADS	*				Solyc06g069430	FUL1	*<0.65*	*143.5*
							Solyc03g114830	FUL2	*<0.65*	*151.7*
AT5G26660	MYB86	MYB	*				Solyc01g102340	R2R3MYB61	*0.73*	*10..2*
AT3G01220	HB20	HB	*				Solyc05g007180	JA1	*0.81*	*33.6*
AT5G15850	COL1	C2C2-CO-like			*		Solyc02g089540	CO1	*0.66*	*1*
AT3G10590	-	MYB					Solyc10g076820	MYB	*<0.65*	*0.1*
AT1G48000	MYB112	MYB					Solyc05g053330	MYB	*<0.65*	*0.2*
AT5G26749	C2H2	C3H	*				Solyc01g100990	C3H13	*<0.65*	*7.8*
AT5G50820	NAC97	NAC	*		*		Solyc03g097650	NAC	*<0.65*	*1.6*
AT3G10580	-	MYB-related					Solyc05g055240	mybi	*<0.65*	*22.3*
AT5G15480	MAZ1	C2H2	*				-			
AT1G71030	MYBL2	MYB-related					Solyc05g008250	MYB76	*<0.65*	*0.1*
AT4G38960	BBX19	Orphans	*				Solyc01g110370	BBX19	*<0.65*	*9.1*
Water										

Bold *S. lycopersicum* TFs were used in a subsequent Y1H screen. Black-coloured *S. lycopersicum* TFs were confirmed with Y1Hseq in the S. lycopersicum promoter, while grey coloured TFs were not. Pearson correlation (r) of co-expression is shown relative to *AGPL1* (solyc01g109790). Average expression in fruit (RPM) was obtained from the Tomato Expression Atlas (http://tea.solgenomics.net/).

To determine if the TFs (BBX19, FUL2, JA1 and C3H13) that interacted with the promoter in the Y1H assay were also able to activate or repress *AGPL1* promoter fragment-induced transcription, promoter reporter assays were performed. Coding sequences of all interacting TFs were recombined into expression cassettes and transformed into *Agrobacterium* (**Figure 2b**). *Agrobacterium* strains containing the TF expression cassette, a vector that contained the 1897-bp upstream of the start codon of either

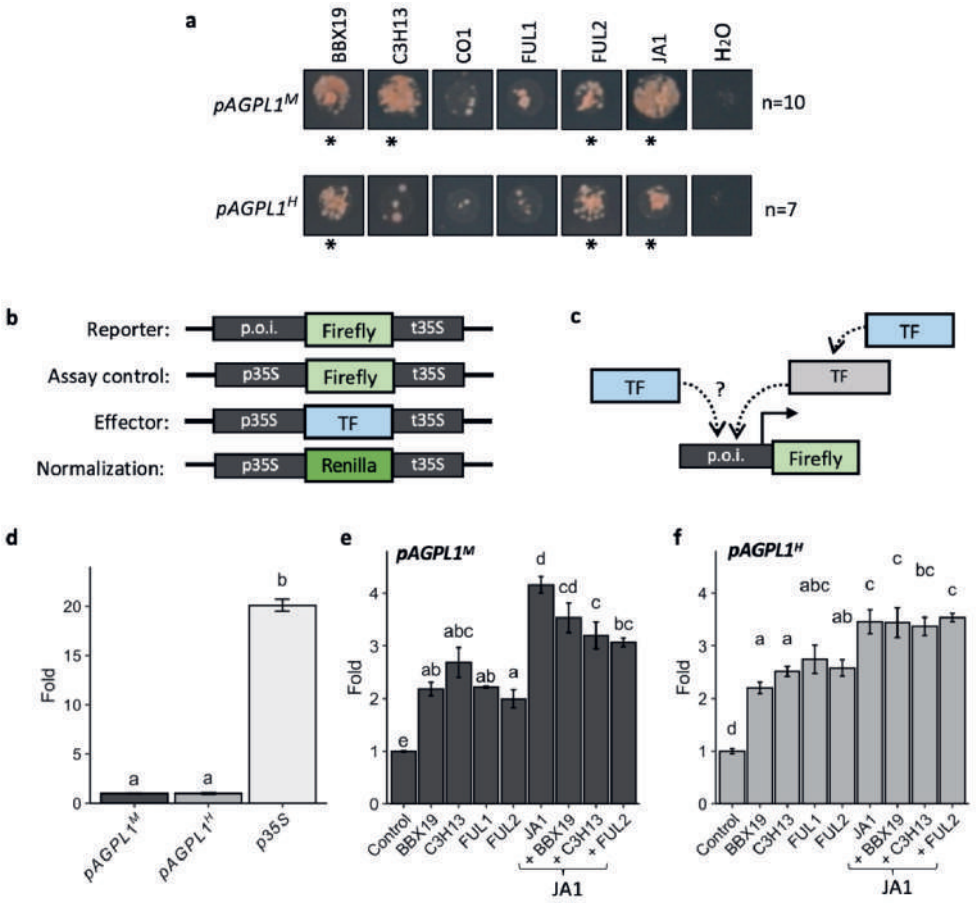


Figure 2: (a) Representative colonies from the Y1H assay results obtained from the analysis of the interaction between the *AGPL1* promoters from *S. lycopersicum* cv. Moneyberg (*pAGPL1^{M})* and *S. habrochaites (pAGPL1^{H})* and tomato TFs BBX19, C3H13, CO1, FUL1, and JA1. As a negative control, reporter-strains were grown in the absence of an effector strain (H_2O). The assay was performed in selection medium with 150 ng/ml of Aureobasidin, the concentration at which the analysed strains were not autoactivating. "*" indicates significant interaction, as described in the methods section. "n" indicates the number of reporter-strains used. **(b)** Schematic diagrams of promoter reporter (promoter of interest (p.o.i.)) and TF effector constructs used in the promoter reporter assays. The 2xpCAMV 35S (p35S) promoter was used in the assay control, effector, and normalization strains. All parts were inserted into separate Golden Gate pL1 expression vectors and transformed to *A. tumefaciens* C58c1. **(c)** Schematic representation of the promoter reporter assays performed in tobacco (*Nicotiana benthamiana*) to examine interaction between different TFs and two *AGPL1* promoters, *S. lycopersicum* cv. Moneyberg (*pAGPL1^{M})* and *S. habrochaites (pAGPL1^{H}).* **(d-f)** The Firefly:Renilla ratio of the reporter without a TF (control) was set to 1 and the fold change of the promoters co-infiltrated with TFs are represented in the figure. Each value represents the mean ± SE of three biological replicates. Significant differences are represented by unique letters.

pAGPL1^{H} or *pAGPL1^{M}* fused to a Firefly luciferase reporter, as well as a normalization vector containing Renilla luciferase were co-infiltrated in *N. benthamiana* leaves. Each co-infiltration was done in three individual plants. TF interaction with the promoter, either directly or indirectly, leads to a change in the Firefly luciferase signal (**Figure 2c**).

There was low basal activity of the promoter without an effector (control). As a positive control for the luminescence assay, a 2xpCAMV35S::Firefly::t35S reporter was used (**Figure 2d**). The combination of *pAGPL1*^{H} or *pAGPL1*^{M} with BBX19, C3H13, FUL1, or FUL2 expression resulted in approximately a 2-fold increased activity of the promoters (**Figure 2e, f**). The addition of JA1 resulted in the strongest activation, 3.5-fold for *pAGPL1*^{H} and 3.6-fold for *pAGPL1*^{M}. As we were interested in finding repressing TFs, we combined the strongest activator, JA1, with the other TFs. In *pAGPL1*^{M} the addition of C3H13 or FUL2 decreased the activation by JA1, while no such effect was observed in *pAGPL1*^{H}. These results indicate that JA1 is a strong activator, while BBX19, C3H13, FUL1 and FUL2 only mildly increase activation. C3H13 and FUL2 could repress *pAGPL1*^{M}, at least in combination with JA1 and possibly as competitor of JA1.

HIGH-THROUGHPUT YEAST-1-HYBRID ASSAYS GENERATED MORE POTENTIALLY INTERACTING TFS

A High Throughput yeast-one-hybrid tomato cDNA library screen followed by Next Generation Sequencing (Y1Hseq) was the secondly applied Y1H method (Erffelinck et al., 2018). The screen was done by using a cDNA library from different stages of developing tomato fruit fused to the Gal4-Activation Domain screened on a selected region of the *AGPL1* promoter. The chosen region was a combination of the part that contained OCR1 (-1080 to -622 bp from the transcription start site) and the immediate upstream region of the *AGPL1* transcription start site (-322 to-30 bp). The screen resulted in 11,194 sequenced fragments, which were aligned to gene models, of which 493 were annotated as a TF (**Figure 3a**). bHLH and bZIP TFs were overrepresented in the TFs present. To help us define the most relevant TFs amongst these hits, we looked for a high relative representation in the pool, measured as absolute Fragments Per Kilobase of Exon Per Million Fragments Mapped (FPKM). A second filter-step was to determine the Signal to Noise (SNR) ratio, which is the number of reads compared to the number of reads for the same gene in a negative control (empty promoter-vector). However, none of the TFs with a FPKM in the 95th percentile (>159,759) had a SNR in the 95th percentile (>7.71) (**Table S2**). Plotting these two parameters against each other, revealed the candidates that had the best combinations of high SNR and FPKM (**Figure 3b, Figure S3)**. The TFs, which we identified in the previous Y1H were represented in this Y1Hseq data (**Figure 3b**).

A next selecting step for biological relevance was to look for expression in fruits. Four candidates were selected and validated in a promoter reporter assay: Nuclear Factor-Y subunit A10 (NF-YA10), the homolog of *Arabidopsis* Telomeric Repeat Binding Factor-Like 3 (TRFL) and WRKY24. (**Figure 3c**). Additionally, JA1, was also used in the promoter reporter assay. JA1 was already confirmed as an activator following the Y1H experiments (**Figure 2d,f**), and JA1 was also found in the Y1Hseq hits. All five selected TFs resulted in an increased activation of the *pAGPL1*^{M}*-firefly* construct in the promoter

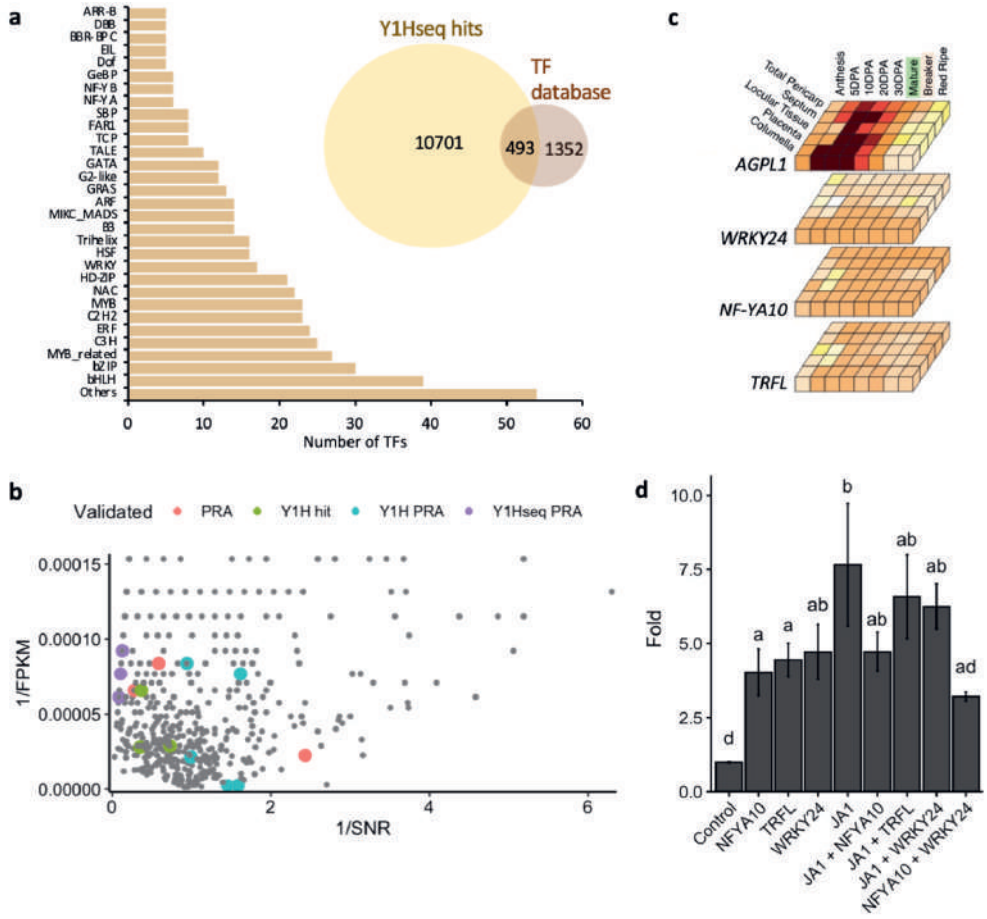


Figure 3: (a) Graph that represents the type and number of TFs identified in the Y1Hseq screen with the *AGPL1* promoter fragments as bait. The Venn-diagram shows the total number of hits, and which are TFs. **(b)** The inverse FPKM (1/FPKM) and the inverse SNR (1/SNR) of each TF-hit from the Y1H seq. Values close to 0 have a high FPKM and high SNR. Coloured hits represent TFs validated in the promoter reporter assays (red, blue, and purple), or were overlapping with orthologs from hits we obtained in the *Arabidopsis*-based Y1H screen (Y1H hit, green, **Table 1**). **(c)** Expression data of selected candidate TF hits, obtained from http://tea.solgenomics.net/ **(d)** Transient expression promoter report assay results performed in *N. benthamiana* to examine interaction between different TFs and the *AGPL1* promoters from *S. lycopersicum* cv. Moneyberg (*pAGPL1*^{M}). The Firefly Luciferase/Renilla luciferase-ratio of the reporter without a TF (control) was set to 1 and the fold-change of the promoters co-infiltrated with TFs are represented in the figure. Each value represents the mean and SE of three biological replicates. Significant differences are represented by unique letters.

reporter assay (**Figure 3d**). A *WRKY24-binding* motif, "AAATGTCAAAT" from the Plant TF Database (p<1E-4) was present in the 77 bp *pAGPL1*^{m} region that was missing in *pAGPL1*^{H} (**Figure S2b,** Grant et al., 2011; Jin et al., 2017). When we combined JA1 and the other four candidates all TFs decreased the activation established by JA1 sightly, but not significantly. This could indicate competition for the same region or negative interaction between the TFs. In conclusion, the results from the selected hits from Y1Hseq did not yield a clear candidate for a repressing TF.

PROMOTER MUTATIONS INCREASED *AGPL 1* EXPRESSION AND BRIX

In parallel with the study of the promoter through predictive analysis with Y1H and promoter reporter assay assays, a random promoter deletion approach was applied. CRISPR/Cas9-mutagenesis was applied to systematically create allelic variation in the promoter of *AGPL1*. We hypothesized that by studying the effect of different promoter mutations on expression and phenotype, regulatory regions could be identified and a mutant with a higher *AGPL1* expression and increased Soluble Solids content ($^{\circ}$Brix) might be identified. To obtain systematic allelic variation, twenty single guide RNA (sgRNA), targeting each of four different regions (R1, R2, R3, R4) with three sgRNAs. A fifth region (R5) was defined by combining eight sgRNAs (**Figure 4a**) spanning regions R1, 2 and 3. Primary transformants (T_0) with these vectors targeting the five respective regions were genotyped to identify mutations in the promoter. In total, 66 mutants out of a 100 transformants were obtained of which 28 were selected for propagation by self-fertilization. Selections were mainly based on having unique mutations.

Progeny (T_1) were genotyped for segregation of homozygous mutations, while lacking the CRISPR/Cas9 T-DNA insertion. Twenty-six alleles were selected (named *pAGPL1-cr-rx-y*, where "*rx*" stands for the region and, and "*y*" for the allele number, **Figure S4**). The transformants that were targeted by a vector with three sgRNAs (R1-R4) had small mutations surrounding the targeted cut-site, while the transformants that had been targeted by eight sgRNAs (R5) showed additional guide-to guide deletions, insertions, or inversions. In the T_1 generation, one or two plants homozygous for each mutant allele were phenotyped for Brix and fruit weight and compared to a wild-type control grown at the same time, which had a Brix ranging between 4 and 5 depending on the season (**Figure 4b, Figure S5**) as a preliminary screen for interesting phenotypes. Only *pAGPL1-cr-r1-02* was phenotyped in a heterozygous state. The obtained promoter mutations resulted in a variation in Brix, ranging between a decrease of 13% (*pAGPL1-cr-r5-06*) to an increase of almost 20% (*pAGPL1-cr-r2-03*) compared to wild-type. However, in most cases where Brix was increased, fruit weight was lower.

The overall location of the mutation seemed to correlate with the effect on Brix and Weight. Mutations in region 1 (closest to the transcription start site) decreased the Brix and had little effect on fruit weight or had even increased fruit weight (*pAGPL1-cr-r1-01*). The heterozygous mutant, *pAGPL1-cr-r1-02*, had no effect on Brix or weight. In region 2, all alleles resulted in a relatively high Brix increase, correlated with the highest reduction in weight. Interestingly, this is the region containing OCR1 and the deletion in *pAGPL1^{H}*. Mutations in region 3 and region 5 showed less consistency between the location of the mutation and the effect on Brix or weight. Only one plant was phenotyped for region 4, *pAGPL1-cr-r4-01,* and this plant's fruits showed no difference from a wild-type phenotype. As the mutant alleles in region 2 showed the highest Brix increase, they were phenotyped more extensively in the T_2 generation.

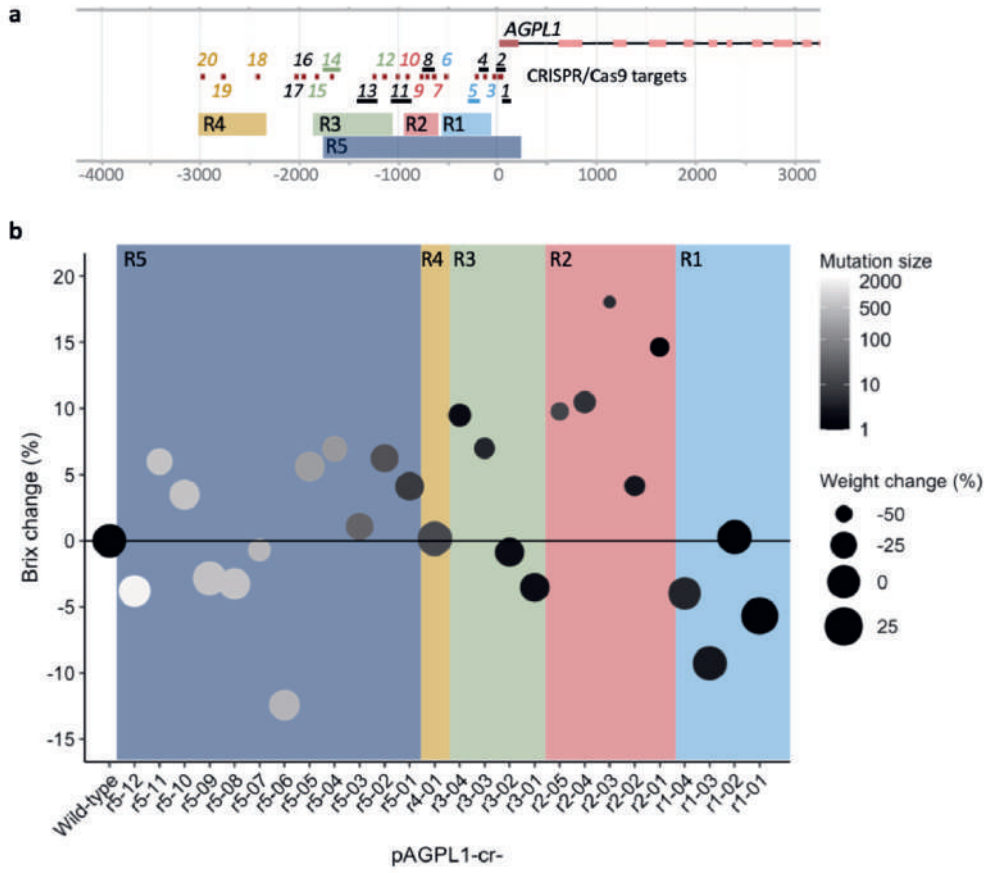


Figure 4: (a) In the top row the gene model is shown in pink (exons) and dark pink (UTRs). The CRISPR/Cas9-targets are depicted in dark red. The colours of the target numbers correspond to the colour of the targeted block (R1-R4) and underlined targets represent combinations of targets that were multiplexed in one vector targeting R5 and. The different groups are referred to as region 1 (R1, Target 3, 5, 6), region 2 (R2, target 7, 9, 10), region 3 (R3, target 12, 14, 15), region 4 (R4, target 18, 19, 20) and region 5 (R5, target 1, 2, 4, 5, 8, 11, 13, 14). The scale is in bp, relative to the transcription start site. **(b)** Phenotype of mutant T_1 fruits**.** The positions of the circles along the Y-axis represents the average change in Brix (by %) compared to a wild-type plant. The circle size represents the change in weight (by %). The colour (greyscale) indicates the size of the permutations in the promoter (**Figure S4**). The mutants are ordered by the different targets that were mutated, divided into the different regions. Three trusses with each six fruits from one or two plants were used for measuring Brix and weight.

To obtain more statistically sound data, three alleles of region 2, *pAGPL1-cr-r2-03*, *-04* and *-05*, were used for phenotyping of T_2 fruits (**Figure 5a**). All contained a five or six-bp deletion in sgRNA target 9, while *pAGPL1-cr-r2-05* contained an additional 7-bp deletion in target 7. For phenotyping, five homozygous plants per allele were grown. Per plant, two trusses with six fruits each were used for measuring Brix and weight. In all three lines, a significant increase in Brix compared to wild-type was found, as in the T_1 generation (**Figure 5b**). Fruit weight decreased in *pAGPL1-cr-r2-03* and *-05*, but not in *pAGPL1-cr-r2-04* (**Figure 5c**). *AGPL1* expression analysis was performed at 7, 14 and

21 Days After Anthesis (DAA) with a leaf sample as control (**Figure 5d**). At 7 DAA, *pAGPL1-cr-r2-03* and *-04* had a significantly increased *AGPL1* expression compared to wild-type. At 14 DAA the *AGPL1* expression of all three promoter alleles was slightly, but not significantly, higher. At 21 DAA no differences were observed. In the leaf, expression of *AGPL1* was generally low, but *pAGPL1-cr-r2-04* showed a small but significant increase. Our data show that *AGPL1* expression peaks earlier with *pAGPL1-cr-r2-03* and *-04* alleles, which could be causing the increased Brix. We did not see correlation between expression at the measured timepoints and Brix in *pAGPL1-cr-r2-05*, even though it did show the highest increase in Brix (and decrease in size).

Apart from finding a mutation that would lead to increased *AGPL1* expression and increased Brix, the second purpose of making promoter deletions was to assess if CREs could be identified using promoter mutagenesis. We hypothesized that the region 2 alleles that led to increased Brix, could harbour CREs overlapping with target 7 or target 9. Motif finding was used to predict which TF (families) could bind on or near the mutated sites. With the Multiple Expectation Maximization for Motif Elicitation (MEME)-suite (Bailey and Elkan, 1994) we sought conserved 50-bp motifs around the targets using the same orthologs from the earlier conservation analysis, *S. pennellii, S. habrochaites, S. tuberosum, C. annuum, N. tabacum, N. sylvestris, N. tomentosiformis* and *N. attenuata*. Target 7 and target 9 were part of highly conserved motifs designated motif 1 and motif 2 respectively (**Figure S6a**), indicating that their mutation could affect the function of a CRE. The MEME-suite motif comparison tool (TOMTOM) was used to compare the two motifs against the JASPAR plant core TF database (Gupta et al., 2007; Khan et al., 2018). Motif 1, the region surrounding target 7 and the -7 bp deletion found in *pAGPL1-cr-r2-05*, gave the GATA-motif binding TF *AtGATA4* as a potential interacting candidate (**Figure S6b**). The mutation sites in target 9 of motif 2 gave two Abscisic Acid Responsive Elements-Binding Factors (*AtABF/AREB)* and two WRKY-motif containing TFs (*AtWRKYs)* candidates (**Figure S6b**). For these candidates, tomato-orthologs were selected based on best-reciprocal hits search. As both the *WRKYs* and *GATAs* are part of large TF families, a narrower selection was done for TFs that were co-expressed with *AGPL1*. From this analysis, *SlAREB1, SlGATA9, SlWRKY41* and *SlWRKY81* were selected for validation with a promoter reporter assay comparing the wild-type promoter (*pAGPL1*) to the different mutant promoters (*pAGPL1-cr-r2-03, -04, -05*, **Figure 5e-h**). *WRKY41* was a strong activator for all promoters, while *AREB1* and *WRKY81* only gave a 2-2.5-fold activation. *GATA9* significantly increased activation of only *pAGPL1-cr-r-03,* although the effect was marginal. The *AREB* motif was present in target 9 (TAAAAAGACAGGTGTTTG, p=0.0009) (Grant et al., 2011; Jin et al., 2017). A WRKY motif was found in target 9 as well (GTGTTTGGCCTGT, p=0.008). To assess potential repressing properties of the TFs, they were combined with *JA1* and the wild-type promoter (**Figure 5e**).

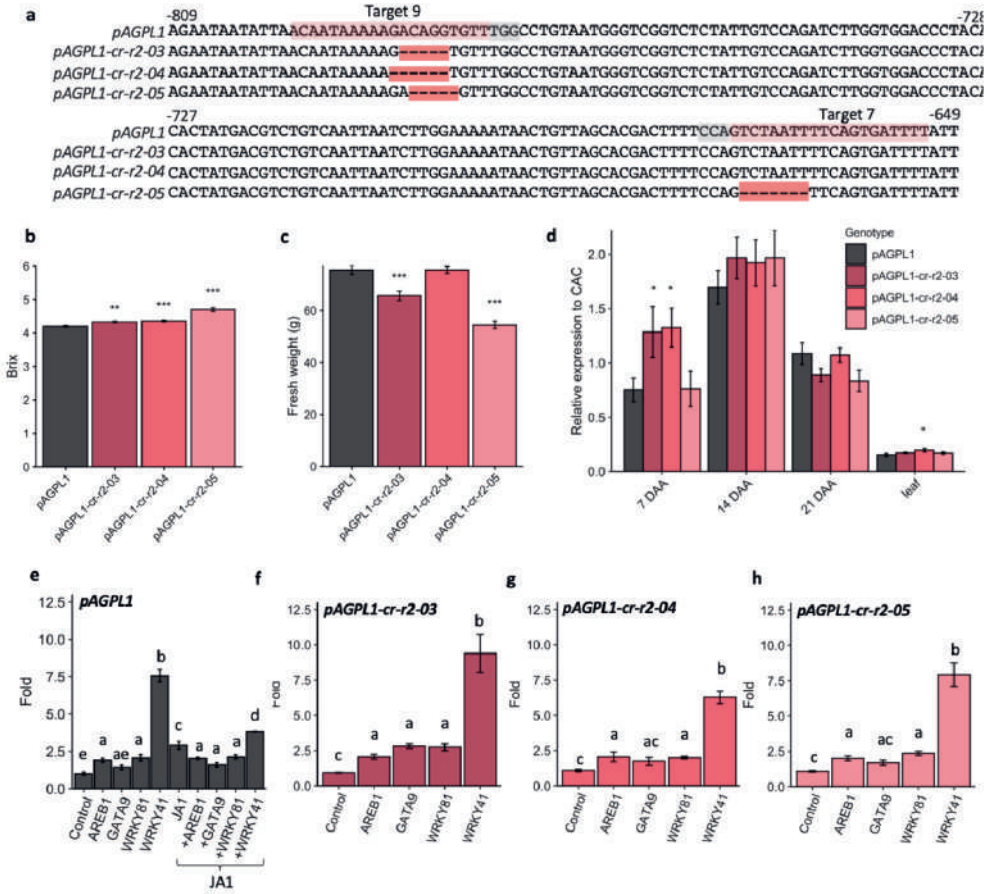


Figure 5: (a) Alignment of the wild-type and mutant promoters. Numbers are relative to the transcription start of *AGPL1*. sgRNA target regions are highlighted pink with a grey PAM. Mutations are highlighted in red. **(b)** Brix and **(c)** weight of T_2 fruits from the third and fifth truss of five plants per genotype. Trusses were pruned to six fruits and harvested at breaker +7 days. **(d)** The second, fourth and sixth truss were used for gene expression analysis. Per timepoint, a pool of three fruits was collected per plant. Statistically significant differences are represented by asterisks (* p<0.05, ** p<0.01, *** p<0.001). **(e-h)** Transient expression promoter report assay results performed in *Nicotiana benthamiana* to examine interaction between different TFs and the wild-type and mutant *AGPL1* promoters from *S. lycopersicum* Var. Moneyberg (*pAGPL1*). The Firefly/Renilla luciferase ratio of the reporter without a TF (control) was set to 1 and the fold change of the promoters co-infiltrated with TFs are represented in the figure. Each value represents the mean ± SE of three biological replicates. Significant differences are represented by unique letters.

The combination of *JA1* with *AREB1*, *GATA9* or *WRKY81* significantly decreased the expression observed with only *JA1*. *WRKY41* and *JA1* combined resulted in expression that was intermediate between *JA1* and *WRKY41* alone. In conclusion, we have identified several candidate TFs (and their tomato homologs) that potentially regulate *AGPL1* expression during fruit development, yet no TF that likely differentially regulates the mutants could be identified.

DISCUSSION

This study aimed at improving the final sugar content of tomato, an important quality aspect (Baldwin et al., 2000; Tieman et al., 2012, 2017; Oltman et al., 2014). Transient starch storage contributes greatly to the final sugar content (Ho et al., 1982; Wang et al., 1993; Schaffer and Petreikov, 1997a; Luo et al., 2020). A rate-limiting factor in the accumulation of starch is ADP-glucose pyrosphorylase (AGPase) activity (Chen and Janes, 1997; Beckles et al., 2001; Ballicora et al., 2004). It was reported that increased expression of the large subunit encoded by *AGPL1,* from a *Solanum habrochaites* allele (*AGPL1*^{H}) led to an increased sugar content (Petreikov et al., 2006). This *AGPL1*^{H} allele was already introgressed in several modern processing tomato varieties, as was found in a whole genome-resequencing study (Lin et al., 2014). Based on this observation, we hypothesized that if we could find and remove repressing *Cis*-Regulatory Elements (CREs) in the *AGPL1* promoter, it would be possible to replicate this phenotype in any cultivar, including indeterminate varieties. A first *in silico* analysis resulted in the identification of conserved and accessible promoter-regions. Putative interacting Transcription Factors (TFs) were identified with Yeast-one-hybrid (Y1H) assays and their effect on transcription validated with Promoter Reporter Assays. In parallel, multiplexed mutagenesis with CRISPR/Cas9 was used to modulate *AGPL1* expression *in situ*. **Figure 6** gives an overview of the techniques, which were used to study the *APL1* promoter.

To modify the expression of *AGPL1*, a thorough understanding of the promoter and regulatory network controlling expression is required (Li et al., 2020). Our first step was to use conservation analysis to find regions likely harbouring CREs. In general, the core promoter of a gene is the 1-2 kb upstream the transcription start site (Yamamoto et al., 2009; Yu et al., 2016), while distal enhancers are further than 2 kb from the nearest gene (Priest et al., 2009; Ricci et al., 2019). Non-coding regions that are conserved, have been under evolutionary pressure to remain unchanged, and have a regulatory function (Ganley and Kobayashi, 2007). The latest available databases, especially FruitEncode, have proven to be a valuable starting point (Lu et al., 2017). The DNAse-seq data at different developmental stages made it possible to determine which regions were most accessible during fruit development and are most likely interacting with Transcription Factors (TF) (Mavrich et al., 2008; Jiang and Pugh, 2009; Jin et al., 2009; Mueller et al., 2017). Four Open Chromatin Regions (OCR) were identified, which were alternating with methylation-rich areas. Methylation is associated with closely-packed DNA, harder to access by TFs and thus thought to not harbour (active) CREs (Zhang et al., 2018; Gallego-Bartolomé, 2020). From our analysis, we hypothesize that the core regulatory region for *AGPL1* spans up to 3000 bp upstream of the transcription start site. This in line with earlier promoter-GUS studies, where a 3200-bp fragment upstream of the *AGPL1* transcription start site demonstrated the expected expression pattern of *AGPL1* in roots and developing fruits (Xing et al., 2005).

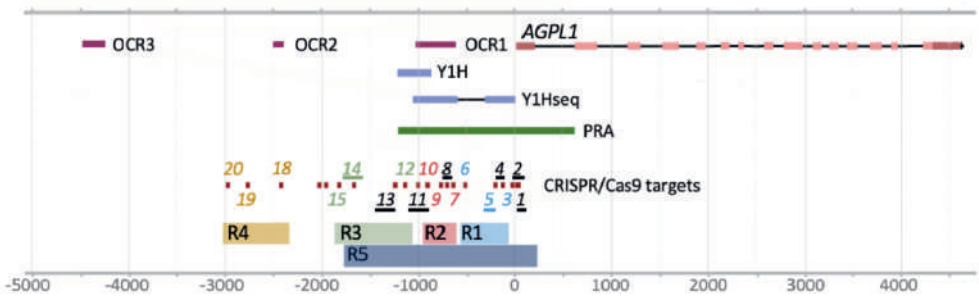


Figure 6: Overview of all experiments performed on the promoter of *AGPL1*. In the top row the gene model is shown in pink (exons) and dark pink (UTRs) combined with the open chromatin regions (OCR) defined in **Figure S1**. Below in blue are the regions used in a Yeast-one-Hybrid assay (Y1H) and Y1Hseq. The green regions have been used for promoter reporter assays. In the bottom row, in dark red, the CRISPR/Cas9 targets are shown. The coloured and underlined targets represent combinations of targets that were multiplexed in one vector and subsequent transformation. The different groups are referred to as region 1 (R1, Target 3, 5, 6), region 2 (R2, target 7, 9, 10), region 3 (R3, target 12, 14, 15), region 4 (R4, target 18, 19, 20) and region 5 (R5, target 1, 2, 4, 5, 8, 11, 13, 14). The scale is in bp, relative to the transcription start.

INTERACTING TFS PLAY A ROLE IN FRUIT DEVELOPMENT AND (A)BIOTIC STRESS RESPONSE

We identified JASMONIC ACID 1 (JA1) as a strong activator of *AGPL1* expression with a Yeast-one-Hybrid (Y1H) comparing a part of the *S. lycopersicum* cv. Moneyberg (*pAGPL*^{M}) and the *S. habrochaites* promoter (*pAGPL*^{H}). JA1 is a homeodomain leucine zipper class I (HD-ZIP I) TF. In both *pAGPL*^{H} and *pAGPL*^{M} a potential JA1 binding-motif was found. Our observation suggests that *JA1* plays a role in regulating the sugar-signalling and storage in tomato. This is in line with earlier research in *Arabidopsis* where *AtJA1* (*ATHB13*) was identified as a component of the sucrose-signalling pathway as well (Hanson et al., 2001). In addition, *ATHB13* was linked to abiotic and biotic stress resistance pathways (Chi et al., 2013). The homolog of the *Arabidopsis* B-box zinc-finger TF BBX19 activated both promoters. *BBX19* is expressed during fruit ripening and is regulated by light and chloroplast maturation (Lira et al., 2020). The CCCH-type Zinc Finger C3H13 interacted in the Y1H assay with *pAGPL*^{M} and resulted in a 2.5-fold activation of both promoters in a promoter reporter assay. When combined with *JA1*, only the cv. Moneyberg promoter activation decreased. Similar results were obtained for FRUITFULL 2 (FUL2). FUL2 is a well-known regulator of fruit ripening (Bemer et al., 2012). The repressing effects of C3H13 and FUL2 when combined with JA1 were not seen in in *pAGPL1*^{H}. In a *ful2* knockout mutant analysis, Brix increased (Jiang et al., 2020). Consequently, it is highly likely that the sugar pathway was affected in these mutants and *AGPL1* could have been a target. Our work has shown that both FUL2 and C3H13 could be putative repressors, or competitors of JA1, while JA1 and BBX19 are putative activators of *AGPL1*. This awaits further confirmation, for example by study of mutations in the respective genes and their effect on *AGPL1* expression and sugar accumulation.

As alternative approach to identify *AGPL1*-promoter-interacting TFs expressed in fruit, we applied a High Throughput yeast-one-hybrid library screening using Next Generation Sequencing (Y1Hseq) with yeast containing the integrated promoter-reporter fusion as "bait" (Erffelinck et al., 2018). In our case an advantage of this method was that a cDNA library from developing tomato could be used instead of an *Arabidopsis* TF library. Because of the cost of NGS only one replicate was performed, precluding an estimation of the variability. What is clear from our results though, is that the assay yields an abundant amount of hits and that most of the obtained hits were not TFs (**Figure 3a**). Here, as well as in the matrix-based screen with Arabidopsis TFs, yeast-one-hybrid assays suffer from strong variation and poor reproducibility, bringing into question the overall usefulness for reliable identification of interacting TFs. Our selection criteria to screen several TFs with a promoter reporter assay, were based on a high Fragments Per Kilobase of Exon Per Million Fragments Mapped (FPKM), high signal to noise ratio (SNR), and co-expression. All selected hits realised an at least 4-fold increase transcriptional activation in the promoter reporter assays, demonstrating the merits of these selection criteria. A Nuclear Factor Y TF, NF-YA10, was the mildest activator. When NF-YA10 was combined with JA1, there was sharp decrease in activation, potentially indicating a repressive or JA1- competing function. It has been suggested that NF-YA10 plays a role in the suppression of ripening (Yang et al., 2016; Li et al., 2016), which could affect the AGPase mediated transition of sucrose to starch as well. The WRKY TF, WRKY24 behaved as an activator of *AGPL1*. The WRKY TF family is a large group with many vital roles (Chi et al., 2013; Karkute et al., 2018). WRKY binding sites, the W-box motif (TTGAC/T) has been found frequently in (drought)stress related genes (Yamasaki et al., 2013; Chi et al., 2013; Karkute et al., 2018). In general, two types of interacting TFs have been found, those involved in fruit development (BBX19, FUL2, NF-YA10 and JA1) and those involved in (a)biotic stress responses (JA1 and WRKY24). The link between starch metabolism controlled by *AGPL1* and fruit development/ripening and stress response is unclear and needs further investigation.

To ultimately increase *AGPL1* expression, we were looking for a repressing TF. The promoter reporter assay however is not an ideal method to find a repressor. The method relies on comparing luciferase signal of the promoter combined with a TF to the promoter without an added TF. The basal luciferase signal of *AGPL1* was relatively low, making it challenging to find a (significant) decrease in activation. Combining TFs with JA1 did lead to a decreased activation in some cases (C3H13 and FUL2). If this decrease means that the added TF is a repressor, is not evident. It could also be that there is competition between JA1 and the other TF for the same or nearby CRE. We took FUL1 as a negative control, as it did not interact in the Y1H. However, FUL1 also increased the promoter activity (**Figure 2e**). FUL1 needs to form heterodimers to interact with a CRE, which it cannot in the yeast screen without an additional MADS-box protein partner, but it might find that partner in the *in planta* system, explaining the signal (Fujisawa et al., 2014).

PROMOTER DELETIONS CAN BE USED TO MODULATE EXPRESSION AND CREATE DESIRED PHENOTYPES

In parallel to unravelling the regulatory TFs of *AGPL1*, a shot-gun approach was used to directly influence expression by promoter mutagenesis. Random systematic allelic variation was created in the promoter by multiplexed CRISPR/Cas9-mutagenesis. We hypothesized that by removing CREs, expression could be modified in a tissue- or stage-specific manner without severe pleiotropic effects. Rather than targeting the entire promoter with a large number of guide RNAs, we choose an approach where the promoter was divided in smaller regions, each targeted by 3 single guide RNAs (sgRNAs). This was done to avoid producing only very large deletions between targets of active sgRNAs at the extremes of the promoter, and thus losing smaller deletions in between those. Contrary to our expectations, we found relatively few target-to-target deletions in the selected regions, but rather small indels as products of individual gRNA activity. This implies that decreasing the number of simultaneously active gRNAs to three, reduces the chances of gRNAs acting together to produce target-to-target deletions, at least when there are no two guides being active simultaneously. Increasing the number of gRNAs in one construct, as used for "Region 5", apparently increased the occurrence of two highly active gRNAs and hence the frequency of target-to-target deletions.

25 mutant alleles were selected and analysed in the T_1 generation as homozygous, transgene-free plants. In several of these mutant lines, Brix was increased. However, an increased Brix was always correlated with lower fruit weight. In contrast, the use of *AGPL1*^{H} resulted in an increased expression of *AGPL1*, a higher sugar content and increased size (Petreikov et al., 2006, 2009). Either the use of a different variety or growing system could be the cause of this difference. Another reason could be that not only the expression of *AGPL1* was affected in the introgression line with *S. habrochaites*. For only two alleles, *pAGPL1-cr-r2-03* and *pAGPL1-cr-r2-04*, increased *AGPL1* expression was found at 7 Days After Anthesis (DAA). Since *pAGPL1-cr-r2-05* did also result in increased Brix an expression difference was expected. The chosen time-points could have been insufficient to measure a difference, as between 5 and 20 DAA *AGPL1* expression fluctuates significantly. Even increased expression at 30 DAA or later could be causal for increased Brix (Petreikov et al., 2006). Overall, we demonstrated that targeting the promoter with CRISPR/Cas9 led to a desired phenotype (increased Brix) and in two lines we could demonstrate the underlying increased *AGPL1* expression.

Concluding, in this study we used a bioinformatics-guided and a (semi-)random approach to study the *AGPL1* promoter in parallel. This led to the identification of two potential repressors for *AGPL1* (FUL2 and C3H13) and several activators (AREB1, BBX19, GATA9, JA1, NFY-A1, TFRL, WRKY24, WRKY41 and WRKY81). In addition, we demonstrated that a semi-random promoter-mutagenesis approach could result in

increased *AGPL1* expression and an elevated Brix. However, as CRISPR/Cas9-mutagenesis, transformation, and regeneration, and especially phenotyping all obtained mutants are space- and labour-intensive processes, narrowing the area targeting for deletions might prove to be more effective. Overall, the promoter-deletion studies were in line with the general picture provided from the DNAse-Seq and conservation analysis. Mutations in OCR1 resulted in the largest effect on Brix and influenced *AGPL1* expression. Combined, the DNAse-Seq, conservation analysis and promoter deletions studies revealed that OCR1 plays a dominant role in the regulation of *AGPL1* expression. Combining Y1H and promoter reporter assays gave interesting leads. Which of these are biologically relevant, remains to be resolved. Further validation could be achieved with Electrophoretic Mobility Shift Assay (EMSA) or DNA affinity TF pulldown assays that could be a valuable additional method for TF interaction discovery (Gaudinier and Brady, 2016). The ultimate test for a role of those TFs *in planta* should come from promoter deletions and study of mutants affected in the respective TF functions.

METHODS

OPEN CHROMATIN

Dnase I Hypersensitive Sites Sequencing (DNAse-seq) data was retrieved for the 25 Kb upstream and 25 kb downstream of *AGPL1* from fruitENCODE (http://www.epigenome.cuhk.edu.hk/encode.html, SI2.50 tomato genome assembly) of several stages of fruit development, respectively during high expression of *AGPL1* (7 Days After Anthesis; DAA), reduced expression (17 DAA), no expression (47 DAA) and leaf profile (control) (Petreikov et al., 2006; Lu et al., 2017). The data was used to determine open chromatin regions in the promoter (10 kb upstream) and in the open reading frame of *AGPL1*, via a simplified model of the peak calling algorithm "Hotspot" (Koohy et al., 2014). For every measured site, a short reading sequence of approx. 250 bp was calculated, which gives the score n. The score for the measured site was compared against the local background of 50 kb (Koohy et al., 2014). A Z-score was calculated and compared to a cut-off value for the probability of the Z score for that site (in this case $p \leq 0.5$). Sites that had a score above the cut-off value were considered open chromatin region (Koohy et al., 2014).

SEQUENCE CONSERVATION ANALYSIS

To perform conservation analysis, the coding sequence, together with the 10 kb upstream of *AGPL1* (Solyc01g109790) was retrieved from the SolGenomics database version 2.5. In total, the upstream region of 9 orthologs from different species were found by using the NCBI blast tool (Query: Solyc01g109780 genomic sequence, BLAST tool: Megablast) (**Table S1**). The retrieved sequences were used as input for conservation analysis with mVISTA (Fasta format, alignment program: AVID (global pair-

wise alignment for finished & draft sequences, cut-off ranking mVISTA value p=0.05, window size of 100 bp) (Mayor et al., 2000; Brudno et al., 2003; Frazer et al., 2004).

YEAST-ONE-HYBRID ASSAY (Y1H)

A region of the *AGPL1* promoter was amplified both from *S. lycopersicum* cv. Moneyberg as well as from *S. habrochaites* genomic DNA with specific primers (**Table S3, Figure 2**). The fragments were recombined into a Gateway® compatible reporter plasmid, version of the pAbAi vector containing the reporter gene *Aur1-c*, which confers resistance to aureobasidin (http://www.clontech.com, Danisman et al. 2012). The two-resulting promoter-reporter constructs were linearized by adding 30 µl vector to 5 µl buffer G Thermo Scientific, 2 µl *Bbs*I and 13 µl MQ and incubating them at 37 °C for 1 h while shaking at 450 rpm. The mixes were inactivated for 20 min at 65 °C. The linearized fragment was transformed into yeast strain PJ69-4 α. A culture was grown overnight in 10 ml SD-complete (6.7 g/L Yeast nitrogen base, 20 g/L Dextrose for 9x SD medium supplemented with the correct 1x dropout. Complete dropout solution: 300 mg/L Isoleucine, 1500 mg/L L-Valine, 200 mg/L L-Adenine hemisulfate salt, 200 mg/L L-Arginine HCL, 200 mg/L L-Histidine HCl Monohydrate, 1000 mg/L L-Leucine, 300 mg/L L-Lysine HCl, 200 mg/L L-Methionine, 500 mg/L L-Phenylalanine, 2000 mg/L L-Threonine, 200 mg/L L-Tryptophan, 300 mg/L L-Tyrosine, 200 mg/L L-Uracil). The culture was diluted to 50 ml and re-grown till an OD_{600} between 0.4 and 0.6 was centrifuged for 5 min at 5000 rpm. The pellet was resuspended in 25 ml 100 mM Lithium Acetate (LiAc) and centrifuged again for 5 minutes at 5000 rpm. A mix of 2.4 ml 50% (w/v) polyethylene glycol (PEG) 3350. 360 µl 1 M LiAc, 50 µl salmon sperm DNA (10 mg/ml, boiled), 20 µl purified linearized plasmid and 680 µl MQ. After vortexing, the mix was incubated for 20 min at 42 °C and centrifuged again (5 minutes, 5000 rpm). The pellet was then washed with 0.5 ml MQ and spun down. Finally, the pellet was resuspended in 0.3 ml MQ and 200 µl was plated on selective medium, SD agar minus Uracil. After three days of growth at 30 °C, transformant colonies were visible. Auto-activation tests were performed at a range of Aureobasidin A (AbA) concentrations (0, 50, 100, 150, 200, 500 ng/µl AbA) to determine background expression of the reporter bait constructs. Reporter constructs were selected for growth at 50 or 100 ng/µl AbA and absence of growth from 150 ng/µl AbA.

We used an expanded REGIA Transcription Factor (TF) prey collection containing more than 2000 *Arabidopsis* TFs in the PJ68-4A yeast strain, as previously reported (Castrillo et al., 2011). A screen with this library was performed twice at 200 ng/ml AbA, as described previously (Danisman et al., 2012). Putative hits were scored on a scale of 0 (no growth) to 3 (full growth). The resulting 186 unique positive hits with a score of at least 1 with either promoter-reporter were re-screened at 175 and 225 ng/ml. Significant positive hits ($p<0.05$) were determined by using a Welch's t-test for analysis of false positives compared to a random independent Y1H screen done with the same library as described before (Ouwerkerk and Meijer, 2001; Mitsuda et al., 2010).

The selected significant positive hits from the REGIA library were used in a second Y1H by with the tomato orthologs of the identified *Arabidopsis* hits. Orthologs were identified d by best reciprocal BLAST hit for the amino acid sequences. Candidates were further selected for (co-)expression in tomato with data from the Tomato Expression Atlas (http://tea.solgenomics.net/). The selected TFs were amplified from cDNA of either 5- or 10-day old fruitlets depending on the TFs expression. The amplified TFs were cloned into Prey expression constructs containing an GAL4 Activation Domain (pDEST22) with Gateway® recombination and were transformed into the PJ69-4A yeast strain. All used strains were confirmed by sequencing. The screen was performed with 10 independent non-auto activating reporter constructs for the *S. lycopersicum* promoter and 7 for the *S. habrochaites* promoter at 150 ng/ul AbA. A score was given for growth (0-10, where a 10 is a full colony) and multiplied with the score for the used promoter-strain in the autoactivation test (0-10, where a 10 is a no colony) to give an interaction score. If the average interaction score was higher than 40, we determined it a significant interaction.

HIGH THROUGHPUT YEAST-ONE-HYBRID LIBRARY SCREENING USING NGS (Y1HSEQ)

Two promoter regions were amplified from genomic DNA with specific primers (**Table S3, Figure 6**) and combined into one region with overlap extension PCR. The bait-construct was cloned via Gateway® recombination in the pDONR P4-P1R vector and subsequently in the pMW#2 vector in combination with the reporter gene. This was followed by linearization and integration into the YM4271 yeast strain. The bait-construct was then used in a high throughput yeast-one-hybrid cDNA library screen using NGS as described before (Erffelinck et al., 2018). As prey, a tomato fruit cDNA library from different stages of fruit development fused to a Gal4-activation domain was used. In short, a screen was performed with the prey-cDNA library. The resulting yeast plates were scraped of all colonies, pooled and plasmid was isolated. The cDNA inserts of the prey plasmids were amplified by PCR and sequenced by Next Generation Sequencing (Illumina HiSeq 2000 sequencing, 125-bp paired-end reads). An empty-bait reporter was screened as control to eliminate false-positive interactions and to correct for the abundance of each prey represented by the cDNA library. A hit was reported as Fragments Per Kilobase of Exon Per Million Fragments Mapped (FPKM) in both the sample and a negative control (an empty bait-vector). The Signal to Noise (SNR) ratio for each gene was calculated as the sample FPKM value divided by the negative control FPKM value. A putative hit can be considered when it has a high SNR and a was above a predefined expression level (high FPKM).

PROMOTER REPORTER ASSAY

The Golden Gate Molecular Cloning (MoClo) toolkit was used to assemble the constructs with the Golden Gate cloning strategy (Engler et al., 2008; Weber et al., 2011; Werner et al., 2012a). The full-length open reading frames of selected TFs were

amplified from cDNA with gene-specific primers (**Table S3).** Where necessary, a nested PCR was performed, and internal Golden Gate restriction sites were removed with overlap extension PCR (primers not listed). The products were ligated into pL0-CDS (pICH41308) and subsequently into pL1-F2 expression vectors (pICH47742) with a 2x CaMV 35S promoter (pICH51288) and 35S terminator (pICH41414) to form the effector pL2 constructs. The promoter was amplified with specific primers (**Table S3)** from wild-type, mutant and *S. habrochaites* genomic DNA, totalling about 1880-bp including the 5'UTR. Again, internal Golden Gate restriction sites were removed. Amplified parts were cut-ligated into pL0-pro+5'UTR (pICH41295). The pL0-promoter was then cut-ligated with a Golden Gate-compatible firefly luciferase coding vector(pICSL50006) and a 35S terminator (pICH41414) to form the reporter construct in a pL1-F1 expression vector (pICH47732). A positive control vector was constructed by ligating a 2x35S promoter with the firefly luciferase coding sequence and 35S terminal. To normalize luciferase activity, a third pL1 vector was used: 2x35Sp::Renilla, which was made Golden Gate compatible from pGreenII:0800-LUC (Hellens et al., 2005).

Correct plasmids were transformed to *A. tumefaciens* C58c1. Transformed cells were grown from an overnight culture in LB with 10mM 2-(N-morpholine)-ethanesulfonic acid (MES) and 40 µM Acetosyringone to an OD_{600} between 0.4 and 1. The cells were concentrated in 10 mM $MgCL_2$ with 200 µM Acetosyringone at and OD_{600} of 1 for the reporter and Renilla construct, and an OD_{600} of 2 for the effector. Different combinations were made with the following ratios: 2.5 Reporter: 62.5 Effector: 1 Renilla. Al mixes were equalled at 20 ml and left to stand for 2-3 hours. For each combination, transfections were performed on three different *Nicotiana benthamiana* plants (biological replicates) of about five weeks old (before flowering). The infiltrated tobacco plants were grown for three days in a greenhouse. Samples were taken by closing a 2-ml tube on a transfected leaf-area while avoiding large veins. From each transfected leaf, three technical replicates were taken. Samples were snap-frozen in liquid Nitrogen. Sample preparation and measurements were done with the Dual-Luciferase Reporter Assay kit (Promega, cat. no. E1910), as described in (Sherf et al., 1996). 200 µl Passive Lysis Buffer mix was used to lysate a cryo-ground leaf disk for 15 min on ice, followed by 3 min centrifuge. Measurements were done in a GlowMax® Navigator Microplate Luminometer with 50 µl supernatant and 25 µl of both the LARII and Stop&Glow reagents. The Firefly/Renilla ratio of each sample was calculated relative to the sample with just the promoter and Renilla construct. Statistical analysis was performed with R. For all data, normal distribution was confirmed, and ANOVAs with post-hoc Tukey pairwise analysis were used to test for significant differences.

CRISPR/CAS9 DESIGN AND ASSEMBLY

The Rgenome Cas-designer tool (http://www.rgenome.net/cas-designer/, spCas9, target: *Solanum lycopersicum*, allowed 2 nucleotide bulge as off target) was used to design effective single guide RNAs (sgRNAs) (Bae et al., 2014; Concordet and Haeussler,

2018). SgRNAs were filtered out if they contained a stretch of 4*T (stop codon) and if they had an off-target with one or more nucleotides mismatch. Twelve targets for sgRNA were chosen targeting the 3000 bp upstream of AGPL1 (**Table S3**). The selected targets were divided in five different regions (**Figure 2**): region 1 (Target 3, 5, 6), region 2 (target 7, 9, 10), region 3 (target 12, 14, 15), region 4 (target 18, 19, 20) and region 5 (target 1, 2, 4, 5, 8, 11, 13, 14). These different groups of guides were all assembled in separate vectors.

The MoCLo toolkit and Golden Gate cut and ligation were used to assemble the vectors (Engler et al., 2009; Weber et al., 2011). In short, each sgRNA was fused to an *Arabidopsis* U6 promoter as AtpU6:sgRNA:TTTT and cut-ligated to a level 1 vector. Level 1 constructs pICH47732-pL1-pNOS::NPTII::tOCS, pICH47742-pL2-p35S::hCas9::tNOS, pICH47751-pL3-p2x35S::tGFP::t35S, pICH47761-pL1-F4-pU6::sgRNA1, pICH47772-pL1-F5-pU6::sgRNA2, pICH47781-pL1-F6p-pU6::sgRNA3 and pICH41822-pLE6E were cut-ligated into the level 2 vector pICSL4723 (Werner et al., 2012b). For the construct with eight guides, a two-step Golden Gate cut-ligation was performed. The NPTII, Cas9, GFP and the two first guides were cut-ligated with pICH49299-pELB5 into the level 2 vector to create an intermediate level 2 (pL2i-1). This plasmid was then used to construct the final level 2 (pL2-2) by cut-ligating it with pICH47781-pL1-F6-pU6::sgRNA3, pL1-F7-pU6::sgRNA4, pICH47732-pL1-F1-pU6::sgRNA5, pICH47742-pL1-F2-pU6::sgRNA6, pICH47751-pL1-F3-pU6::sgRNA7, pICH47761-pL1-F4-pU6::sgRNA8 and pICH41780-pELE4. All constructed vectors were checked by sequencing. All pL2 and the pL2-2 constructs were transformed to *Agrobacterium tumefaciens* C58C1 and grown under rifampicin, gentamycin, and kanamycin selection. In *Agrobacterium*, the presence of the correct construct was validated with restriction analysis (BamH1, SAL1).

TRANSFORMATION

Tomato transformation was done with *Agrobacterium tumefaciens* strain C58C1 containing the appropriate vector as previously described, but with media B supplemented with 1 mg/L 2,4D and with 200 cotyledon explants as starting material (Van Roekel et al., 1993). *Solanum lycopersicum* L. cv. 'Moneyberg' was used in this study. Tissue culture was done in a growth chamber with 16 h light and 8 h dark cycle at 25 °C. Once shoots were formed, GFP-positive shoots were selected and rooted on Rooting Inducing Medium (Van Roekel et al. 1993). Rooted shoots were placed on rockwool and moved to a growth chamber (16 h light and 8 h dark at 25 °C). A ploidy test was done on transformed plant leaf samples at Iribov Analytical Services B.V. Diploid shoots were genotyped for presence of the transgene and for mutations in the target region.

GENOTYPING

Genomic DNA from young leaves was isolated using the CTAB as described (Porebski et al., 1997). Alternatively, a PCR was done directly on sampled leaf tissue by using the Phire Plant Direct PCR kit (Thermo Scientific, Catalog number F130WH). Detection of transgenes in each generation was done by a PCR on Cas9 and/or NPTII. The target region of transformed plants was amplified, sequenced, and aligned in Benchling (https://benchling.com, for primers used see **Table S3**). Heterozygous and bi-allelic Transformed (T_0) plants for the region of interest were selfed. T_1 plants were segregated for the presence of a homozygous or bi-allelic mutation in the region of interest, while lacking the T-DNA insertion. In the T_2 generation only homozygous plants were grown.

GROWTH CONDITIONS

Shoots grown from callus were transferred to rockwool once they had roots. Alternatively, T_1 and T_2 seeds were germinated on filter paper and transferred to cubes of rockwool after a week. They were grown in a growth chamber (16 h light and 8 h dark at 25 °C). Five to Eight weeks later seedlings were transplanted into a greenhouse (Unifarm, Wageningen 51.57° N, 5.31° E, The Netherlands) on rockwool slabs at a density of 2.5 plants/m^2. Nutrients were provided by fertigation (EC 4.5, pH 5.6). Climatic conditions in the greenhouse were at ambient temperature (> 20 °C) under a 16h light/8h dark cycle (0.6–28.4 MJ m–2 day–1 natural light supplemented with artificial light using high pressure sodium lamps (SON-T Agro 600 Watt, Philips, Eindhoven, The Netherlands). Side shoots were removed once a week. Flowers were pollinated by vibrating each flower/truss three times a week with an electric toothbrush.

PHENOTYPING FRUITS

T_0 plants were only used for genotyping and seed collection. In the T_1 generation, one or two plants per genotype were placed randomly on a row in the greenhouse together. As it was not possible to put all the T_1 plants in the greenhouse at the same time, phenotyping was done in batches. For every batch of T_1 mutant plants, a wild-type control was taken along. Six flowers per truss were vibrated at anthesis. Excess flowers were removed. The first truss was used for seed collection and the subsequent three trusses were used for phenotyping. Individual fruits were harvested at breaker+7 ($\pm$1 day) and Fresh weight (in grams) and Brix measurements were done. Brix measurements were done in duplo per fruit with an Atago PR-32α digital refractometer.

In the T_2 generation, five plants per genotype were randomly placed on a row in the greenhouse. Six flowers per truss were vibrated with an electric toothbrush-holder at anthesis. Excess flowers were removed. The first truss was used for seed collection and the three subsequent trusses were used for phenotyping. The third and fifth truss fruits were harvested at breaker+7 ($\pm$1 day) for phenotyping. Of these fruits, width, weight,

and Brix were measured. Of the second and fourth and sixth truss, samples were taken for gene expression analysis.

GENE EXPRESSION ANALYSIS

Gene expression analysis was done in the T_2 generation with five plants per genotype. Samples were taken from each plant as a pool of three fruits 7, 14 and 21 DAA with a leaf sample as control. RNA was isolated from cryo-ground whole fruit samples by using the MaqMax™-96 total RNA isolation kit with Plant RNA isolation aid (Thermofisher) with a KingFisher 96 Magnetic Particle Processor. cDNA was synthesized with the iScript cDNA synthesis kit (Bio-Rad). qRT-PCR was performed with iQ SYBR Green Supermix (Bio-Rad) on a Bio-Rad CFX6 qRT-PCR instrument. The following condition was applied for PCR amplification: 3 min 95˚C, 40*[15 s 95˚C, 60 s 60˚C], followed by a melt-curve. *CAC* was used as a refence (**Table S3**). Relative expression changes were calculated according to the $2^{-\Delta CT}$ method as described (Livak and Schmittgen, 2001). Statistical analysis was performed with R, package version 3.5. For all data, normal distribution was confirmed, and ANOVAs were used to test for significant differences between mutants and wild-type.

ACKNOWLEDGMENTS

We thank Avanish Raia and Alain Goosens (University of Gent) for the help with the Y1Hseq and running the Y1H. pDEST22:FUL1 and pDEST22:FUL2 were gifts from Marian Bemer. All MoClo kit plasmids were a gift from Sylvestre Marillonnet. pICSL50006 was a gift from Nicola Patron. The pGreenII:0800-LUC was a gift from Kieran Elborough. The luciferase vectors were made Golden Gate-compatible by Lena Maas. We thank Renze Heidstra for his assistance with the luciferase measurements, and Geurt Versteeg, Teus van den Brink and Sean Geurts for their expertise and care for the tomato plants.

SUPPLEMENTARY FIGURES

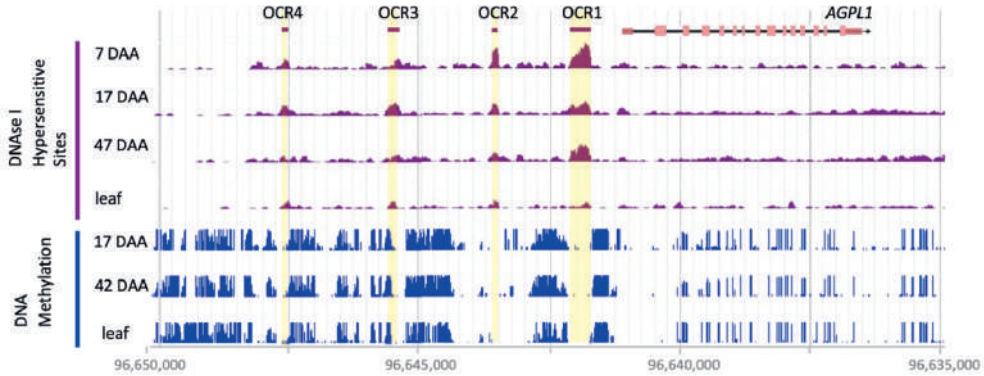


Figure S1: Local DNAse I hypersensitive sites and DNA methylation profiles of AGPL1. The upper row shows the gene model of *AGPL1* (Solyc01g109790) in pink. The four rows below (purple) show the DHS profiles at 7, 17 and 47 DAA as well as leaf tissue on a scale of 0 to 3. The yellow highlights represent significant Open Chromatin Regions (OCRs). These OCRs were determined by peak calling (p<.05, background comparison: 50kb, local region: 250 bp). The last three rows in blue show the DNA methylation profiles at respectively 17 and 42 DAA as well as the leaf tissue profile on a scale of 0 to 1. In the last row, the nucleotide position on chromosome 1 is shown. Note that the reverse-complement orientation is shown, as *AGPL1* resides on the antisense-strand. All data is retrieved from FruitENCODE (http://www.epigenome.cuhk.edu.hk/encode.html, SI2.50 tomato genome assembly) (Lu et al., 2017).

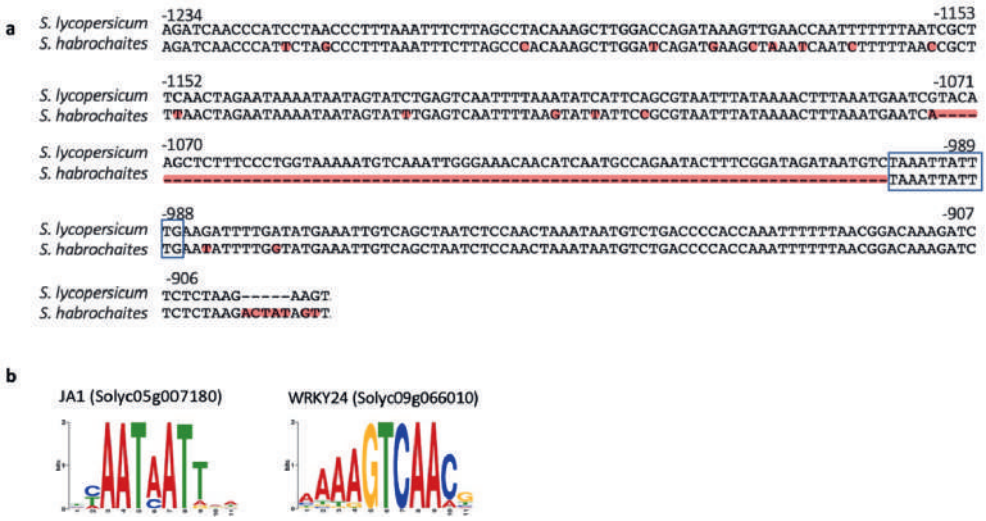


Figure S2: (a) Alignment of *S. lycopersicum* and *S. habrochaites* sequences at the 5' end of OCR1 in the *AGPL1* promoter. Note the 77 bp deletion in *S. habrochaites*. Location numbers are relative to the transcription start of *AGPL1*. This region was used in the Y1H screen. **(b)** Position Weight Matrix Motifs obtained from the Plant TF Database (http://planttfdb.gao-lab.org/).

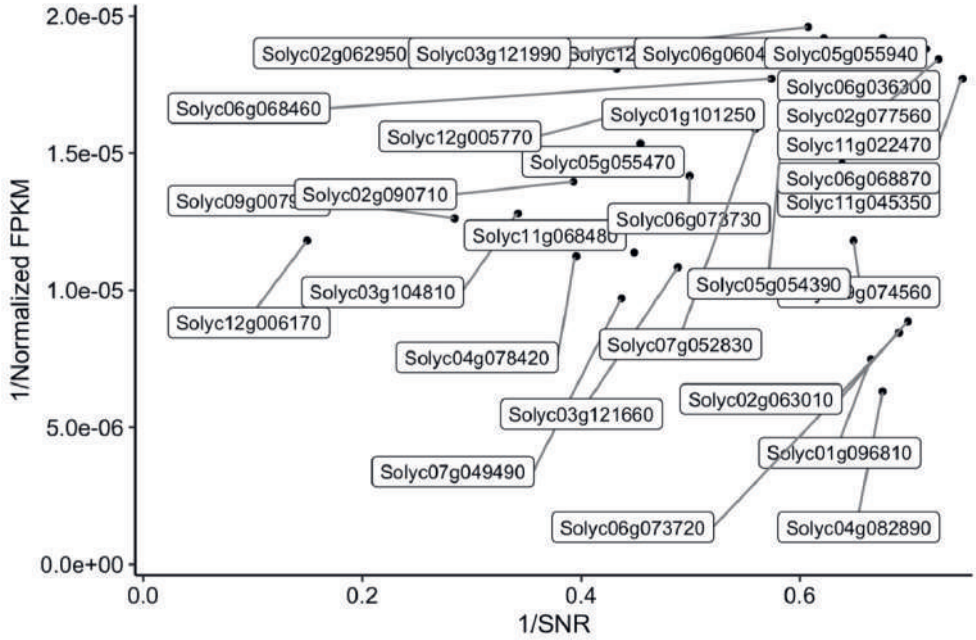


Figure S3: The top 30 TFs with high SNR and high FPKM from the Y1Hseq experiment

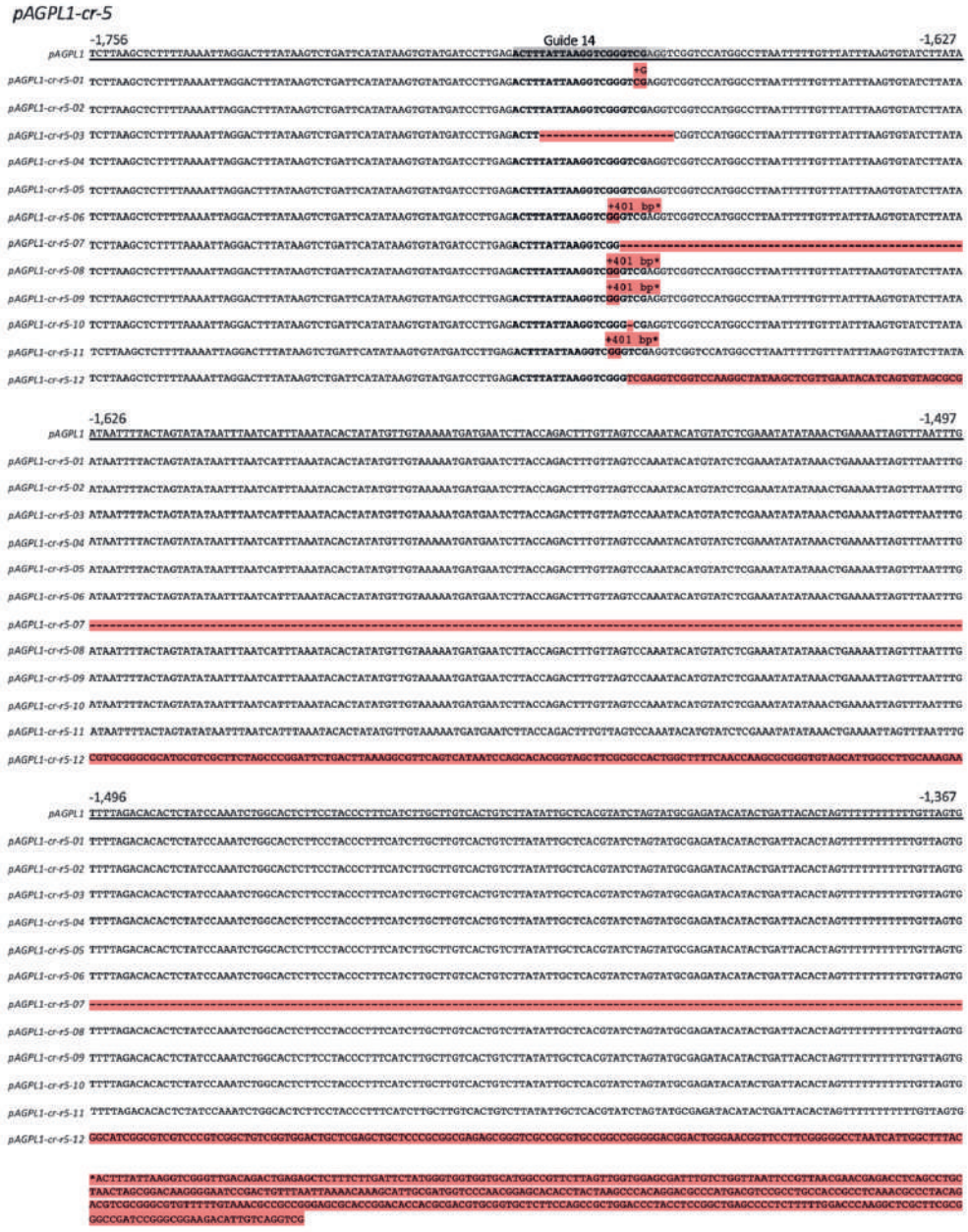


Figure S4: Alignments of the wild-type and mutant promoters. Location numbers are relative to the transcription start site of *AGPL1*. sgRNA target regions are highlighted with a grey PAM site. Mutations are highlighted in red.

pAGPL1-cr-5 (continued)

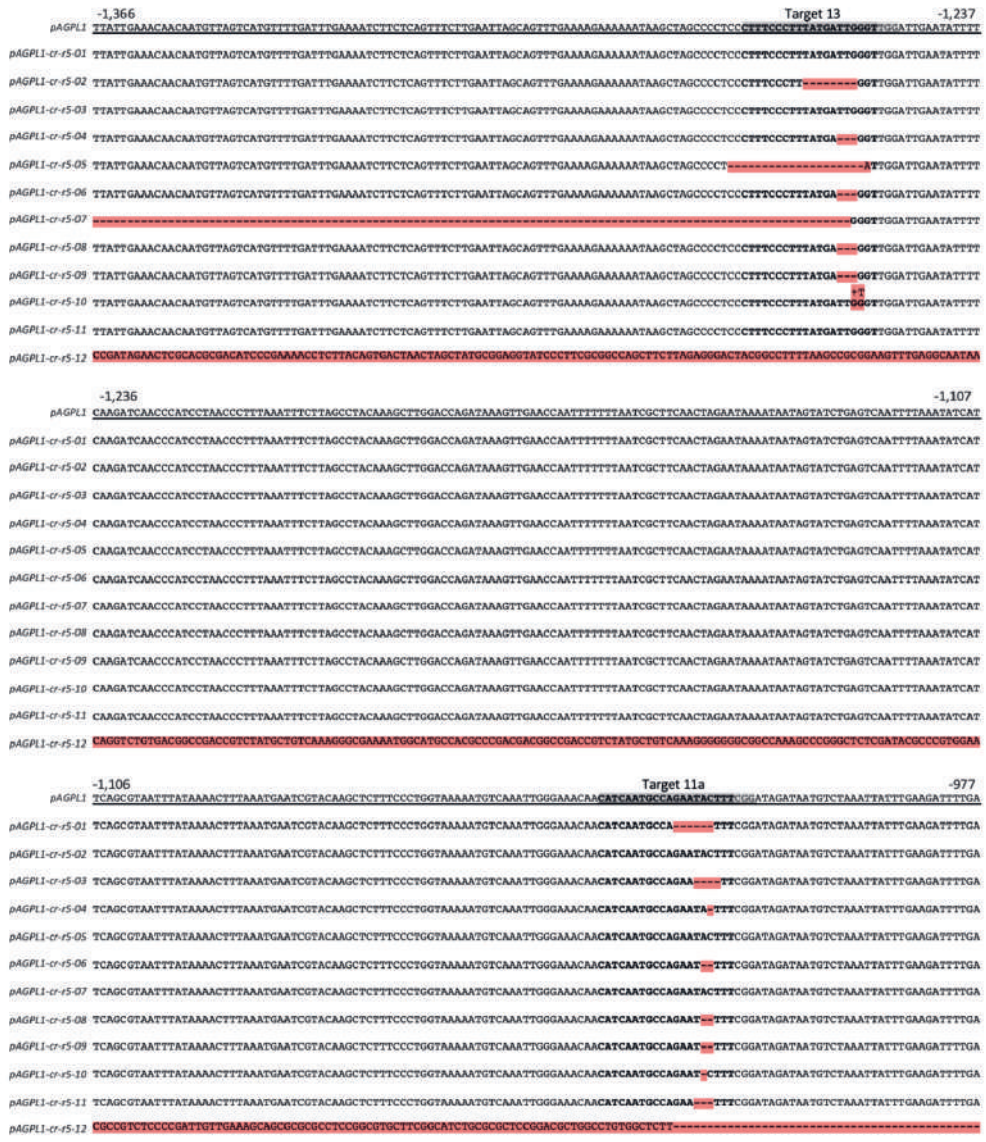


Figure S4 continued

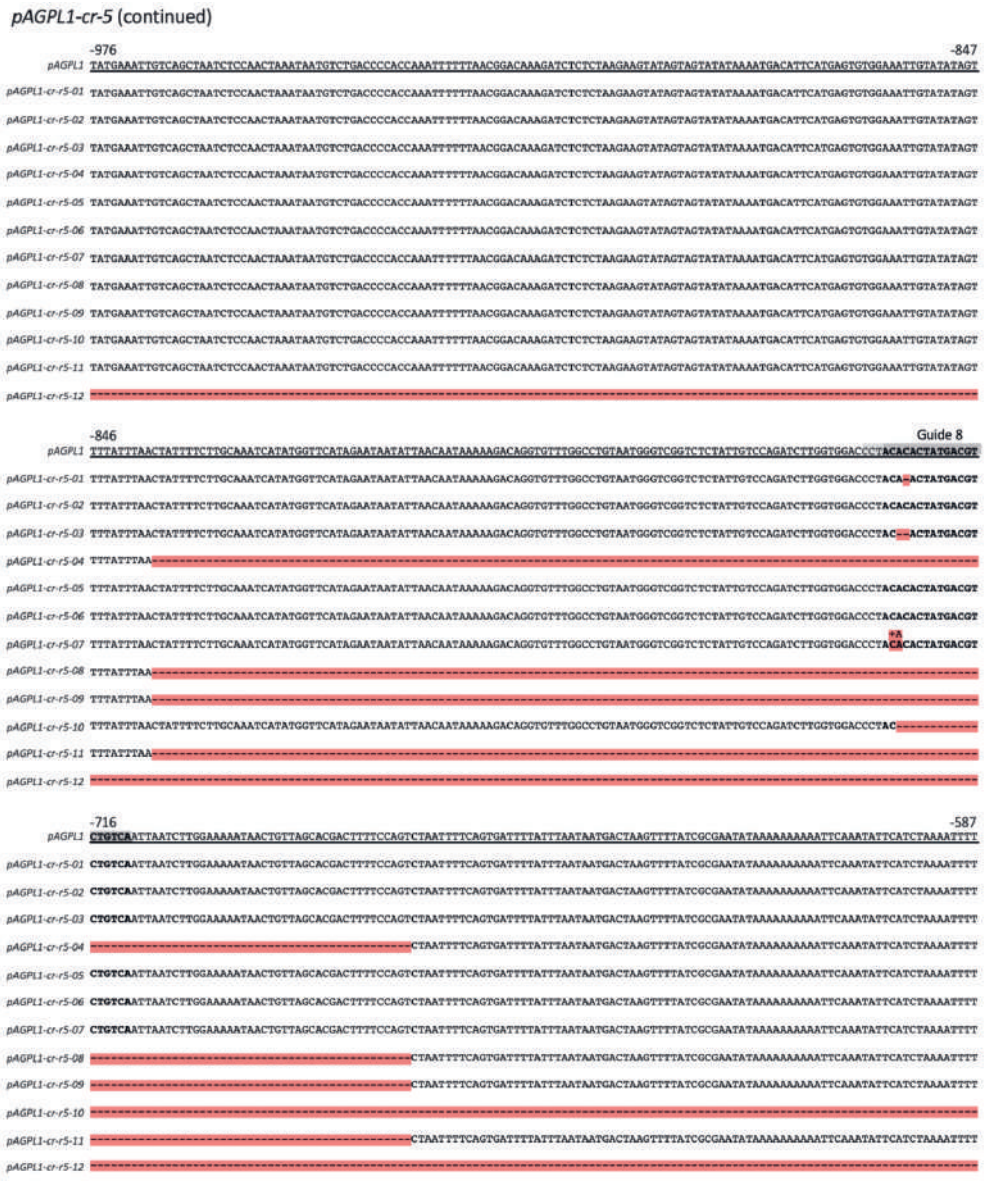


Figure S4 continued

pAGPL1-cr-5 (continued)

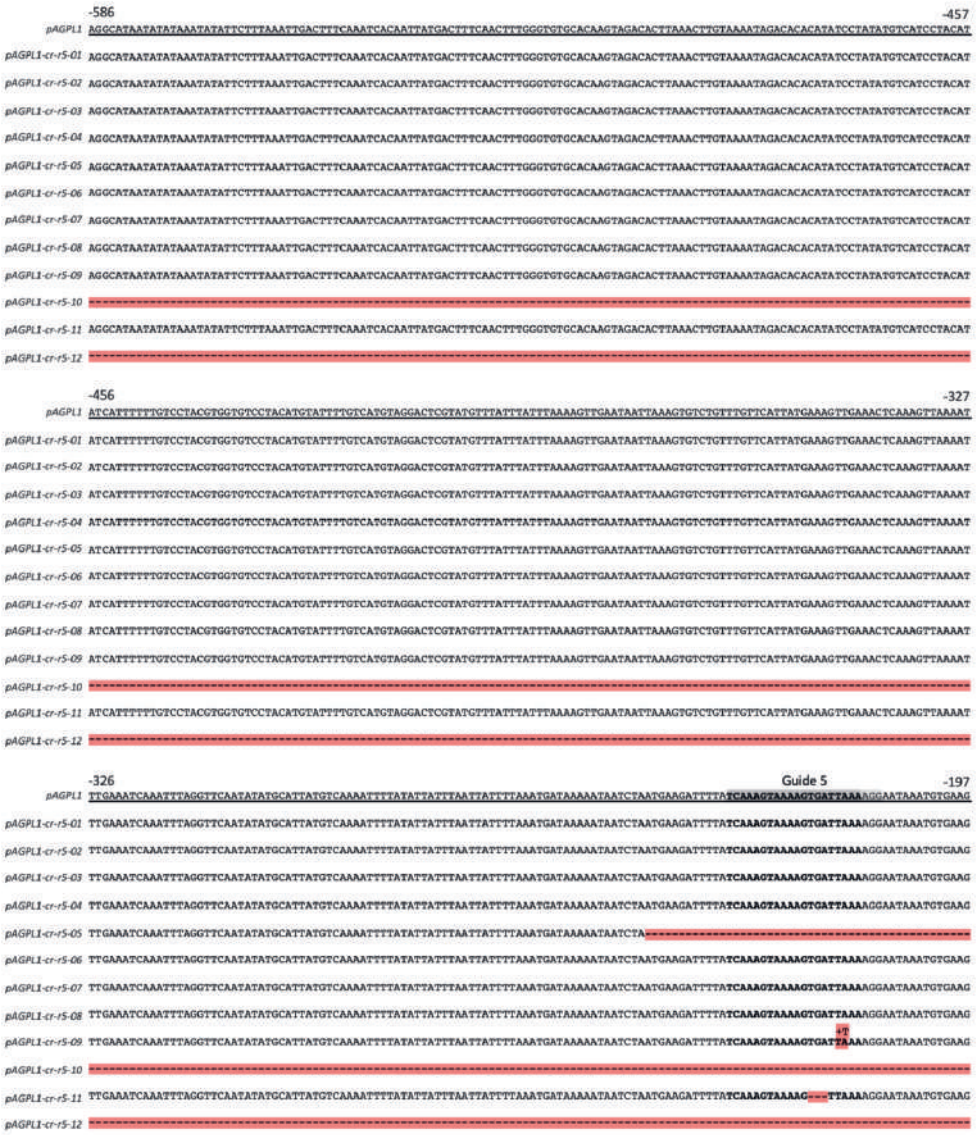


Figure S4 continued

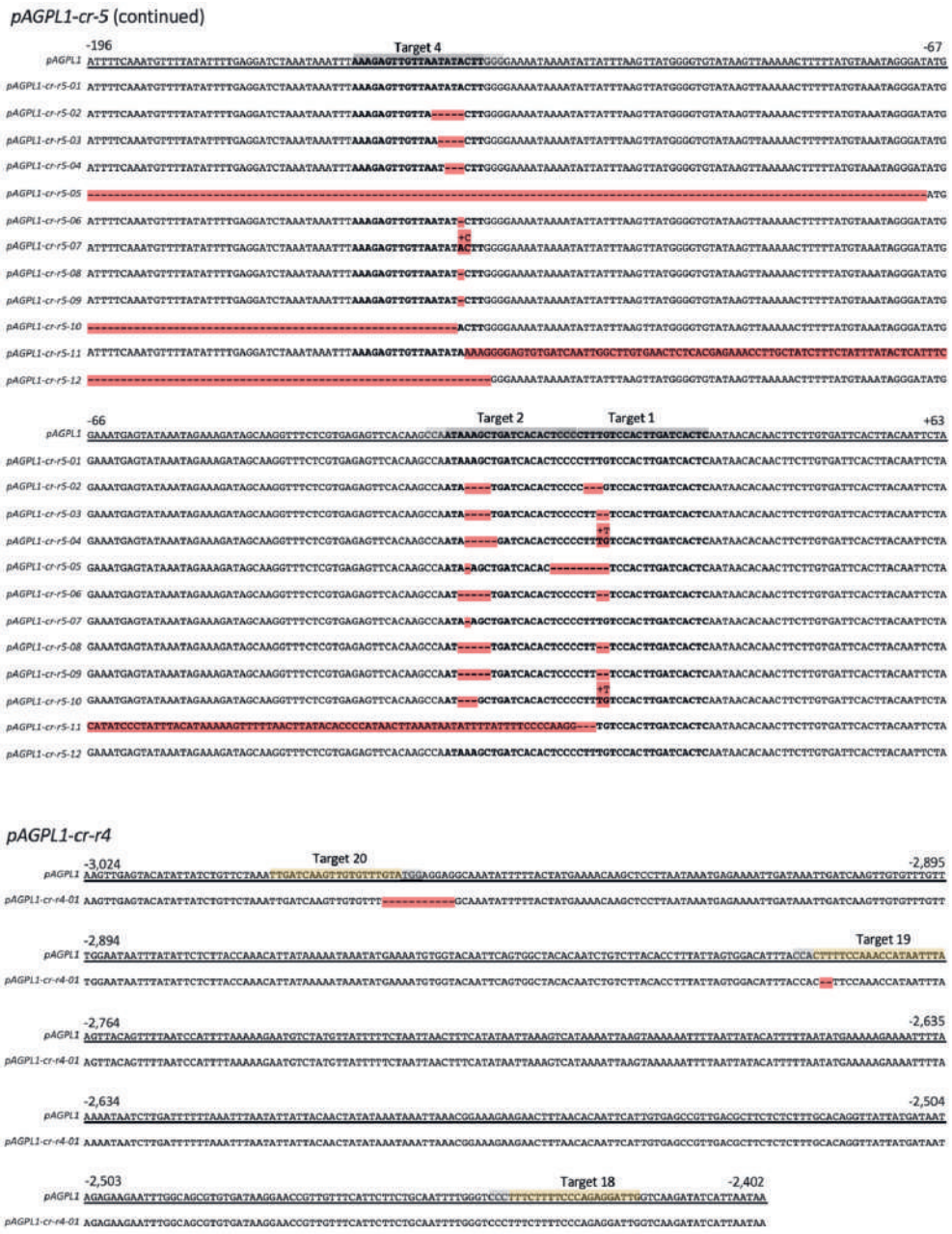


Figure S4 continued

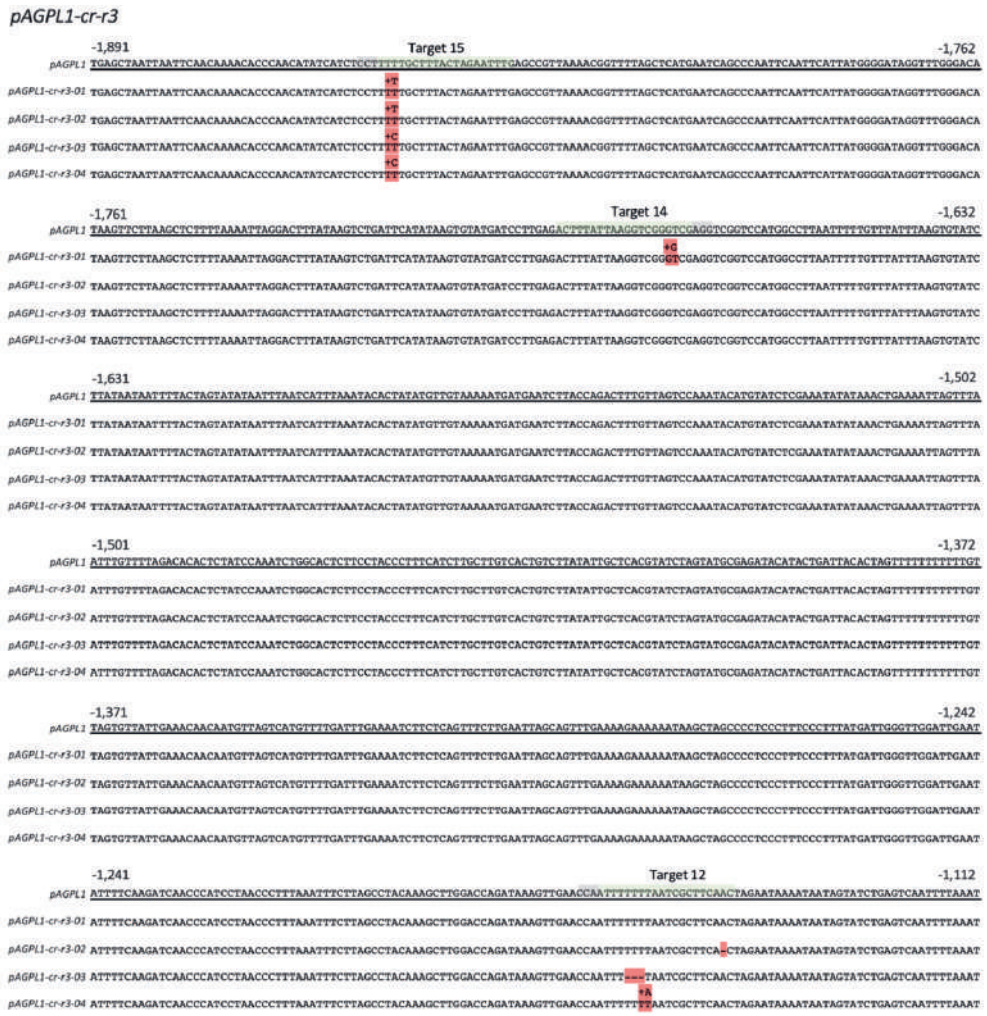


Figure S4 continued

Figure S4 continued

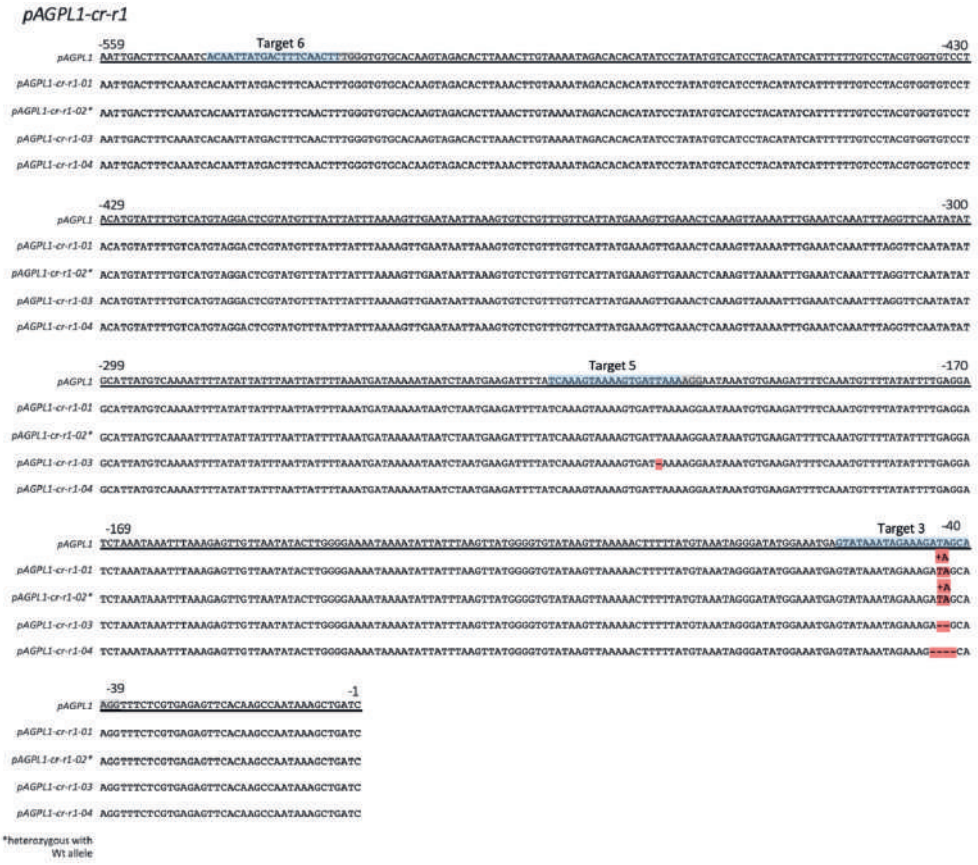


Figure S4 continued

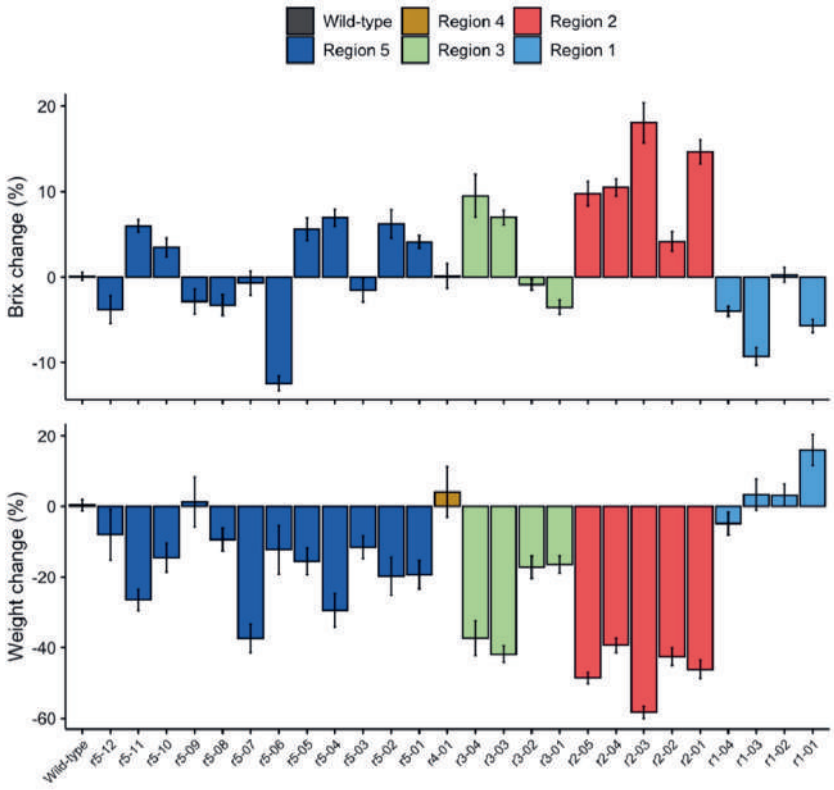


Figure S5: Phenotypes T1. Change in Brix (%) and change in weight (%). compared to a Wild-type plant grown at the same time as the mutant. The mutants are ordered by the different targets they have mutated, divided in the different regions (**Figure 1**). Error bars represent the SE. For most mutants, three trusses with 6 fruits of one plant were measured, for some two plants were measured.

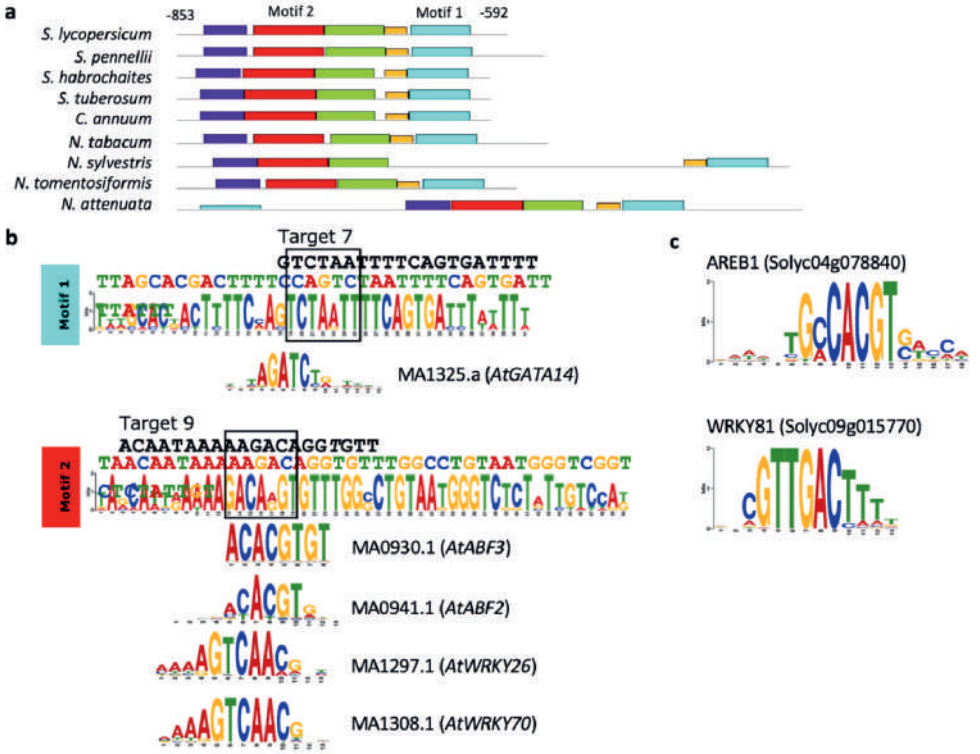


Figure S6: Motif prediction and TF selection. **(a)** MEME motif prediction of region 2 in the *AGPL1* promoter containing mutation leading to a higher brix phenotype. Bp position numbering is relative to the 5' UTR Transcription Start Site of the *S. lycopersicum AGPL1*. The blue and red motifs are designated motif 1 and 2 respectively (http://meme-suite.org/), **(b)** Position Weight Matrix Motifs and alignment with the *S. lycopersicum* genome of motif 1 and motif 2. The mutation sites are indicated by a black box. Underneath the motif are the resulting TF binding motifs obtained from a TOMTOM motif comparison with the JASPAR core plants database (2018) **c.** Position Weight Matrix Motifs obtained of *S. lycopersicum* orthologs of the JASPAR-hits, obtained from the Plant TF Database (http://planttfdb.gao-lab.org/).

SUPPLEMENTARY TABLES

Supplementary Table 1: *AGPL1* orthologs used for promoter sequence conservation analysis

Species	Locus number
Solanum lycopersicum	Solyc01g109780 (AGPL1)
Solanum pennellii	LOC107008126
Solanum habrochaites	Sohab1.0:CBYS010003766.1 (107195..93703)
Solanum tuberosum	LOC102591766
Capsicum annuum	LOC107840421
Nicotiana tabacum	LOC107826565
Nicotiana sylvestris	LOC104215019
Nicotiana attenuata	LOC109227102
Nicotiana tomentosiformis	LOC104113035

Supplementary Table 2: 95th percentile Y1Hseq hits based on SNR (TFs only)

Gene ID	TF family	$FPKM_{empty}$	$FPKM_{pAGPL1}$	SNR	Tested
Solyc12g056510	Trihelix	1408	46669	33.14	
Solyc10g005330	HD-ZIP	1408	23877	16.96	
Solyc07g020710	CPP	1408	22792	16.19	
Solyc03g006910	bHLH	1408	19536	13.87	
Solyc06g071300	M-type_MADS	1408	19536	13.87	
Solyc12g099130	MYB	2816	36901	13.10	
Solyc09g057880	FAR1	1408	17365	12.33	
Solyc04g076360	MYB_related	1408	17365	12.33	
Solyc06g065190	TCP	1408	17365	12.33	
Solyc09g066010	WRKY	1408	16280	11.56	Y1Hseq PRA
Solyc02g067230	Dof	1408	16280	11.56	
Solyc12g038510	MIKC_MADS	1408	15194	10.79	
Solyc04g011670	bZIP	1408	13024	9.25	
Solyc04g005800	HD-ZIP	1408	13024	9.25	
Solyc02g079760	bHLH	1408	13024	9.25	
Solyc01g006930	NF-YA	1408	13024	9.25	Y1Hseq PRA
Solyc01g009070	MYB	4224	37986	8.99	
Solyc01g073950	MYB_related	2816	24962	8.86	
Solyc12g042010	bHLH	1408	11939	8.48	
Solyc03g078120	NAC	1408	11939	8.48	
Solyc08g081140	bHLH	2816	22792	8.09	
Solyc07g005400	bHLH	2816	21706	7.71	
Solyc06g083430	Trihelix	2816	21706	7.71	
Solyc09g007810	ARF	1408	10853	7.71	
Solyc04g005130	bHLH	1408	10853	7.71	
Solyc05g009880	bHLH	1408	10853	7.71	
Solyc10g047040	C3H	1408	10853	7.71	
Solyc07g052760	GeBP	1408	10853	7.71	
Solyc10g080300	MYB_related	1408	10853	7.71	Y1Hseq PRA
Solyc05g018350	Trihelix	1408	10853	7,71	

PRA: Promoter Reporter Assay

Supplementary Table 3: Primers used throughout this chapter

Target description	Name	Sequence (5'-3')
Y1H primers		
promoter of *AGPL1* (Solyc01g109790) from *S. lycopersicum*	Y1H Lyc Fw	ACTTCTTAGAGAGATCTTTGTCCGT
	Y1H Lyc Rv	AGATCAACCCATCCTAACCCTTT
promoter of *AGPL1* (Sohab1.0:CBYS010003766) from *S. habrochaites*	Y1H Hab Fw	ATAGTCTTAGAGAGATCTTTGTCCGT
	Y1H Hab Rv	AGATCAACCCATTCTAGCCCTTT
ORF *BBX19* (Solyc01g110370)	Y1H Fw	GGGGACAAGTTTGTACAAAAAAGC AGGCTTCATGAGAACCCTTTGTGATG
	Y1H Rv	GGGGACCACTTTGTACAAGAAAGC TGGGTCTTAAAGCTTTGACATTGCAT
ORF *C3H3* (Solyc01g100990)	Y1H Fw	GGGGACAAGTTTGTACAAAAAAGC AGGCTTCATGCCGTTGGGTAAATACT
	Y1H Rv	GGGGACCACTTTGTACAAGAAAGC TGGGTCCTATCCCCACTCAACAAATG
ORF *CO1* (Solyc02g089540)	Y1H Fw	GGGGACAAGTTTGTACAAAAAAGC AGGCTTCAAAGTTAGTCCAATGTTGA
	Y1H Rv	GGGGACCACTTTGTACAAGAAAGC TGGGTCTCAGAATGAAGGGACAATTC
ORF JA1 (Solyc05g007180)	Y1H Fw	GGGGACAAGTTTGTACAAAAAAGC AGGCTTCATGGCTTTCTTTCCAACAA
	Y1H Rv	GGGGACCACTTTGTACAAGAAAGC TGGGTCCTAATTAAAATGTTGTTGCT
Y1Hseq Ra	Y1H-seq Fw	CGAGAAACCTTGCTATCTTTCT
	Y1H-seq Rv	TATCGCGAATCAAATTTAGGTTCAATATATGCATTATG
Y1Hseq Rb	Y1H-seq Fw	CCTAAATTTGATTCGCGATAAAACTTAGTCAT
	Y1H-seq Rv	TGAATCGTACAAGCTCTTTCC
Promoter Reporter Assay (PRA) primers (given without Golden Gate compatible ends)		
pAGPL1 Cv. Moneyberg	PRA Fw	AGATCAACCCATCCTAACCCTTT
	PRA Rv	TTCTGCAGCCAATCAACAA
pAGPL1 *S. habrochaites*	PRA Fw	AGATCAACCCATTCTAGCCCT
	PRA Rv	TTCTGCAGCCAATCAACAA
AREB1 (Solyc04g078840)	PRA Fw	ATGGGGAGTAATTATC
	PRA RV	TTACCATGGACCAG
BBX19 (Solyc01g110370)	PRA Fw	ATGAGAACCCTTTGTGATG
	PRA RV	TTAAAGCTTTGACATTGC
C3H13 (Solyc01g100990)	PRA Fw	ATGCCGTTGGGTAAATACTA
	PRA RV	CTATCCCCACTCAACAA
FUL1 (Solyc06g069430)	PRA Fw	ATGGGAAGAGGAAGAGTCC
	PRA RV	TTAATTATTAAGATGACGAAGCATCC
FUL2 (Solyc03g114830)	PRA Fw	ATGGGTAGAGGAAGAGTACA
	PRA RV	TTAACCGTTGAGATGGC
GATA9 (solyc11g069510)	PRA Fw	ATGGATGAAATTCCTACTGGTC
	PRA RV	TCAGTATACATCAAAC
JA1 (Solyc05g007180)	PRA Fw	ATGGCTTTCTTTCCAACAAA
	PRA RV	TAATTAAAATGTTGTTGCTCTAGC

Supplementary Table 3 continued

Description	name	Sequence (5'-3')
NF-YA10 (solyc01g006930)	PRA Fw	ATGAATACTACTATATTTTCCAAAGG
	PRA RV	TCATACTTTGAGGTTGCAAC
TRFL (Solyc10g080300)	PRA Fw	ATGCATGCGGAGGTCA
	PRA RV	TTATGTTTCTTTCTCGTCTGC
WRKY24 (Solyc09g066010)	PRA Fw	ATGGAGGAGATTGAGGAAGCTAACAG
	PRA RV	TCATGCATTTGCCGATTGG
WRKY41 (Solyc05g012770)	PRA Fw	ATGGAGAAAGTTAAAAGTATGGAG
	PRA RV	TTAAATGAAGAATTCTTCAATGTC
WRKY81 (Solyc09g015770)	PRA Fw	ATGGATAACTCATCGTCTG
	PRA RV	CTACACTTGATCAAAGTTCC
sgRNA primers		
Amplification sgRNA 1	pAGPL1-g1	TGTGGTCTCAATTGGAGTGATCAAGTGGACAAAG GTTTTAGAGCTAGAAATAGCAAG
Amplification sgRNA 2	pAGPL1-g2	TGTGGTCTCAATTGGGGAGTGTGATCAGCTTTAT GTTTTAGAGCTAGAAATAGCAAG
Amplification sgRNA 3	pAGPL1-g3	TGTGGTCTCAATTGGTATAAATAGAAAGATAGCA GTTTTAGAGCTAGAAATAGCAAG
Amplification sgRNA 4	pAGPL1-g4	TGTGGTCTCAATTGAAAGAGTTGTTAATATACTT GTTTTAGAGCTAGAAATAGCAAG
Amplification sgRNA 5	pAGPL1-g5	TGTGGTCTCAATTGTCAAAGTAAAAGTGATTAAA GTTTTAGAGCTAGAAATAGCAAG
Amplification sgRNA 6	pAGPL1-g6	TGTGGTCTCAATTGACAATTATGACTTTCAACTT GTTTTAGAGCTAGAAATAGCAAG
Amplification sgRNA 7	pAGPL1-g7	TGTGGTCTCAATTGAAAATCACTGAAAATTAGAC GTTTTAGAGCTAGAAATAGCAAG
Amplification sgRNA 8	pAGPL1-g8	TGTGGTCTCAATTGTGACAGACGTCATAGTGTGT GTTTTAGAGCTAGAAATAGCAAG
Amplification sgRNA 9	pAGPL1-g9	TGTGGTCTCAATTGACAATAAAAAGACAGGTGTT GTTTTAGAGCTAGAAATAGCAAG
Amplification sgRNA 10	pAGPL1-g10	TGTGGTCTCAATTGTGTCCGTTAAAAAATTTGGT GTTTTAGAGCTAGAAATAGCAAG
Amplification sgRNA 11	pAGPL1-g11	TGTGGTCTCAATTGCATCAATGCCAGAATACTTT GTTTTAGAGCTAGAAATAGCAAG
Amplification sgRNA 12	pAGPL1-g12	TGTGGTCTCAATTGGTTGAAGCGATTAAAAAAAT GTTTTAGAGCTAGAAATAGCAAG
Amplification sgRNA 13	pAGPL1-g13	TGTGGTCTCAATTGCTTTCCCTTTATGATTGGGT GTTTTAGAGCTAGAAATAGCAAG
Amplification sgRNA 14	pAGPL1-g14	TGTGGTCTCAATTGACTTTATTAAGGTCGGGTCG GTTTTAGAGCTAGAAATAGCAAG
Amplification sgRNA 15	pAGPL1-g15	TGTGGTCTCAATTGCAAATTCTAGTAAAGCAAAA GTTTTAGAGCTAGAAATAGCAAG
Amplification sgRNA 16	pAGPL1-g16	TGTGGTCTCAATTGGAGATTAGTTAGATAATTGA GTTTTAGAGCTAGAAATAGCAAG
Amplification sgRNA 17	pAGPL1-g17	TGTGGTCTCAATTGAGTGACATACATATAAAGAT GTTTTAGAGCTAGAAATAGCAAG
Amplification sgRNA 18	pAGPL1-g18	TGTGGTCTCAATTGCAATCCTCTGGGAAAAGAAA GTTTTAGAGCTAGAAATAGCAAG

Supplementary Table 3 continued

Description	name	Sequence (5'-3')
Amplification sgRNA 19	pAGPL1-g19	TGTGGTCTCAATTGTAAATTATGGTTTGGAAAAG GTTTTAGAGCTAGAAATAGCAAG
Amplification sgRNA 20	pAGPL1-g20	TGTGGTCTCAATTGTTGATCAAGTTGTGTTTGTA GTTTTAGAGCTAGAAATAGCAAG
Universal amplification sgRNAs	sgRNA_rv	TGTGGTCTCAAGCGTAATGCCAACTTTGTAC
Genotyping primers		
Region 1 Genotyping	pAGPL1 Fw	TCCAGATCTTGGTGGACCCT
(Guide 3, 5, 6)	pAGPL1 Rv	ACCTGGAAAGTGATGATTTTTGGA
Region 2 Genotyping	pAGPL1 Fw	ACACTCTATCCAAATCTGGCACT
(Guide 7, 9, 10)	pAGPL1 Rv	GTAGGACACCACGTAGGACA
Region 3 Genotyping	pAGPL1 Fw	TGCACTGTCTCATTGTCTCCT
(Guide 12, 14, 15)	pAGPL1 Rv	CTGGCATTGATGTTGTTTCCCA
Region 4 Genotyping	pAGPL1 Fw	CTTGGATCCCTGCGCCTATT
(Guide 17, 19, 20)	pAGPL1 Rv	GGGCTGATTCATGAGCTAAAACC
Region 5 Genotyping (Guide 1, 2, 4, 5, 8, 11, 13, 14)	pAGPL1 Fw	AGTGAAAGAAGACAAGCACATCT
	pAGPL1 Fw	TCCAGATCTTGGTGGACCCT
	pAGPL1 Rv	AGTGAAAGAAGACAAGCACATCT
qPCR primers		
CAC	qPCR Fw	CCTCCGTTGTGATGTAACTGG
	qPCR Rv	ATTGGTGGAAAGTAACATCATCG
AGPL1	qPCR Fw	GTTGTGCGGCTATGAAATCGACG
	qPCR Rv	TCTCCATTGTTAAAGCCACCAGTGC

The underlined sequence represents the spacers in the sgRNA primers. Fw: Forward, Rv: Reverse, Lyc: *S.lycopersicum*, Hab: *S.habrochaites;* sgRNA: single guide RNA, g: guide

REFERENCES

Bae, S., Park, J., and Kim, J.-S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics **30**: 1473–1475.

Bailey, T. and and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. AAAI Press: 28–36.

Baldwin, E.A., Scott, J.W., Shewmaker, C.K., and Schuch, W. (2000). Flavour trivia and tomato aroma: Biochemistry and possible mechanisms for control of important aroma components. HortScience **35**: 1013–1022.

Ballicora, M.A., Iglesias, A.A., and Preiss, J. (2004). ADP-glucose pyrophosphorylase: A regulatory enzyme for plant starch synthesis. Photosynth. Res.

Beckles, D.M., Craig, J., and Smith, A.M. (2001). ADP-glucose pyrophosphorylase is located in the plastid in developing tomato fruit. Plant Physiol. **126**: 261–266.

Bemer, M., Karlova, R., Ballester, A.R., Tikunov, Y.M., Bovy, A.G., Wolters-Arts, M., Rossetto, P. de B., Angenent, G.C., and de Maagd, R.A. (2012). The tomato FRUITFULL homologs TDR4/FUL1 and MBP7/FUL2 regulate ethylene-independent aspects of fruit ripening. Plant Cell **24**: 4437–51.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-Resolution Mapping and Characterization of Open Chromatin across the Genome. Cell **132**: 311–322.

Brudno, M., Malde, S., Poliakov, A., Do, C.B., Couronne, O., Dubchak, I., and Batzoglou, S. (2003). Glocal alignment: Finding rearrangements during alignment. In Bioinformatics (Oxford Academic), pp. 54–62.

Castrillo, G., Turck, F., Leveugle, M., Lecharny, A., Carbonero, P., Coupland, G., Paz-Ares, J., and Oñate-Sánchez, L. (2011). Speeding cis-trans regulation discovery by phylogenomic analyses coupled with screenings of an arrayed library of *Arabidopsis* transcription factors. PLoS One **6**: e21524.

Centeno, D.C. et al. (2011). Malate plays a crucial role in starch metabolism, ripening, and soluble solid content of tomato fruit and affects postharvest softening. Plant Cell **23**: 162–84.

Chen, B.Y. and Janes, H.W. (1997). Multiple forms of ADP-glucose pyrophosphorylase from tomato leaf. Plant Physiol. **113**: 235–241.

Chen, B.Y., Janes, H.W., and Gianfagna, T. (1998). PCR cloning and characterization of multiple ADP-glucose pyrophosphorylase cDNAs from tomato. Plant Sci. **136**: 59–67.

Chi, Y., Yang, Y., Zhou, Y., Zhou, J., Fan, B., Yu, J.Q., and Chen, Z. (2013). Protein-protein interactions in the regulation of WRKY transcription factors. Mol. Plant **6**: 287–300.

Concordet, J.P. and Haeussler, M. (2018). CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res. **46**: W242–W245.

Crevillén, P., Ballicora, M.A., Mérida, Á., Preiss, J., and Romero, J.M. (2003). The different large subunit isoforms of *Arabidopsis thaliana* ADP-glucose pyrophosphorylase confer distinct kinetic and regulatory properties to the heterotetrameric enzyme. J. Biol. Chem.

Crevillén, P., Ventriglia, T., Pinto, F., Orea, A., Mérida, Á., and Romero, J.M. (2005). Differential pattern of expression and sugar regulation of Arabidopsis thaliana ADP-glucose pyrophosphorylase-encoding genes. J. Biol. Chem. **280**: 8143–8149.

Danisman, S., van der Wal, F., Dhondt, S., Waites, R., de Folter, S., Bimbo, A., van Dijk, A.J., Muino, J.M., Cutri, L., Dornelas, M.C., Angenent, G.C., and Immink, R.G.H. (2012). *Arabidopsis* class i and class II TCP transcription factors regulate jasmonic acid metabolism and leaf development antagonistically. Plant Physiol. **159**: 1511–1523.

Decker, D.O. (2017). UDP-sugar metabolizing pyrophosphorylase in plants formation of precursors for essential glycosylation-reaction. Fysiol. Bot.

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes. PLoS One **4**: e5553.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS One **3**: 1–7.

Erffelinck, M.L., Ribeiro, B., Perassolo, M., Pauwels, L., Pollier, J., Storme, V., and Goossens, A. (2018). High-throughput yeast two-hybrid library screening using next generation sequencing. bioRxiv: 368704.

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I. (2004). VISTA: Computational tools for comparative genomics. Nucleic Acids Res. **32**: W273–W279.

Gallego-Bartolomé, J. (2020). DNA methylation in plants: mechanisms and tools for targeted manipulation. New Phytol. **227**: 38–44.

Ganley, A.R.D. and Kobayashi, T. (2007). Phylogenetic Footprinting to Find Functional DNA Elements. In Comperative Genomics (Humana Press), pp. 367–379.

Gaudinier, A. and Brady, S.M. (2016). Mapping Transcriptional Networks in Plants: Data-Driven Discovery of Novel Biological Mechanisms. Annu. Rev. Plant Biol. **67**: 575–594.

Geigenberger, P., Merlo, L., Reimholz, R., and Stitt, M. (1994). When growing potato tubers are detached

from their mother plant there is a rapid inhibition of starch synthesis, involving inhibition of ADP-glucose pyrophosphorylase. Planta **193**: 486–493.

Goto, Y., Nonaka, S., Yin, Y.-G., Koiwa, T., Asamizu, E., Ezura, H., and Matsukura, C. (2013). Isolation and characterisation of the ADP-glucose pyrophosphorylase small subunit gene (AgpS1) promoter in tomato (*Solanum lycopersicum* L.). Plant Biotechnol. **30**: 279–286.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: Scanning for occurrences of a given motif. Bioinformatics **27**: 1017–1018.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biol. **8**: R24.

Hanson, J., Johannesson, H., and Engström, P. (2001). Sugar-dependent alterations in cotyledon and leaf development in transgenic plants expressing the HDZhdip gene ATHB13. Plant Mol. Biol. **45**: 247–262.

Hellens, R.P., Allan, A.C., Friel, E.N., Bolitho, K., Grafton, K., Templeton, M.D., Karunairetnam, S., Gleave, A.P., and Laing, W.A. (2005). Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. Plant Methods **1**: 13.

Ho, L.C., Sjut, V., and Hoad, G. V. (1982). The effect of assimilate supply on fruit growth and hormone levels in tomato plants. Plant Growth Regul. **1**: 155–171.

Ikeda, H., Shibuya, T., Imanishi, S., Aso, H., Nishiyama, M., and Kanayama, Y. (2016). Dynamic Metabolic Regulation by a Chromosome Segment from a Wild Relative during Fruit Development in a Tomato Introgression Line, IL8-3. Plant Cell Physiol. **57**: 1257–1270.

Du Jardin, P., Harvengt, L., Kirsch, F., Le, V.Q., NguyenQuoc, B., and Yelle, S. (1997). Sink-cell-specific activity of a potato ADP-glucose pyrophosphorylase B-subunit promoter in transgenic potato and tomato plants. Planta **203**: 133–139.

Jiang, C. and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: Advances through genomics. Nat. Rev. Genet. **10**: 161–172.

Jiang, J. (2015). The "dark matter" in the plant genomes: Non-coding and unannotated DNA sequences associated with open chromatin. Curr. Opin. Plant Biol. **24**: 17–23.

Jiang, X., Lubini, G., Hernandes-Lopes, J., Rijnsburger, K., Veltkamp, V., Maagd, R.A. de, Angenent, G.C., and Bemer, M. (2020). FRUITFULL-like genes regulate flowering time and inflorescence architecture in tomato. bioRxiv **September**: 1–32.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark "nucleosome-free regions" of active promoters and other regulatory regions. Nat. Genet. **41**: 941–945.

Jin, J., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J., and Gao, G. (2017). PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res. **45**: D1040–D1045.

Karkute, S.G., Gujjar, R.S., Rai, A., Akhtar, M., Singh, M., and Singh, B. (2018). Genome wide expression analysis of *WRKY* genes in tomato (*Solanum lycopersicum*) under drought stress. Plant Gene **13**: 8–17.

Khan, A. et al. (2018). JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. **46**: D260–D266.

Koohy, H., Down, T.A., Spivakov, M., and Hubbard, T. (2014). A comparison of peak callers used for DNase-Seq data. PLoS One **9**: 96303.

Kwak, M.S., Noh, S.A., Oh, M.J., Huh, G.H., Kim, K.N., Lee, S.W., Shin, J.S., and Bae, J.M. (2006). Two sweetpotato ADP-glucose pyrophosphorylase isoforms are regulated antagonistically in response to sucrose content in storage roots. Gene **366**: 87–96.

Li, Q., Sapkota, M., and van der Knaap, E. (2020). Perspectives of CRISPR/Cas-mediated cis-engineering in horticulture: unlocking the neglected potential for crop improvement. Hortic. Res. **7**: 1–11.

Li, S., Li, K., Ju, Z., Cao, D., Fu, D., Zhu, H., Zhu, B., and Luo, Y. (2016). Genome-wide analysis of tomato NF-Y factors and their role in fruit ripening. BMC Genomics **17**: 1–16.

Li, X., Xing, J., Gianfagna, T.J., and Janes, H.W. (2002). Sucrose regulation of *ADP-glucose pyrophosphorylase subunit* genes transcript levels in leaves and fruits. Plant Sci. **162**: 239–244.

Lin, T. et al. (2014). Genomic analyses provide insights into the history of tomato breeding. Nat. Genet. **46**: 1220–1226.

Lira, B.S., Oliveira, M.J., Shiose, L., Wu, R.T.A., Rosado, D., Lupi, A.C.D., Freschi, L., and Rossi, M. (2020). Light and ripening-regulated *BBX* protein-encoding genes in *Solanum lycopersicum*. Sci. Rep. **10**: 19235.

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.

Lu, P. et al. (2017). The fruitENCODE project sheds light on the genetic and epigenetic basis of convergent evolution of climacteric fruit ripening. bioRxiv: 231258.

Luo, A., Kang, S., and Chen, J. (2020). SUGAR Model-Assisted Analysis of Carbon Allocation and Transformation in Tomato Fruit Under Different Water Along With Potassium Conditions. Front. Plant Sci. **11**.

Mavrich, T.N. et al. (2008). Nucleosome organization in the *Drosophila* genome. Nature **453**: 358–362.

Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., Pachter, L.S., and Dubchak, I. (2000). VISTA: Visualizing global DNA sequence alignments of arbitrary length. Bioinformatics **16**: 1046–1047.

Mitsuda, N., Miho, I., Takada, S., Takiguchi, Y., Kondou, Y., Yoshizumi, T., Fujita, M., Shinozaki, K., Matsui, M., and Ohme-Takagi, M. (2010). Efficient Yeast one-/two-hybrid screening using a library composed only of transcription factors in *Arabidopsis thaliana*. Plant Cell Physiol. **51**: 2145–2151.

Mueller, B., Mieczkowski, J., Kundu, S., Wang, P., Sadreyev, R., Tolstorukov, M.Y., and Kingston, R.E. (2017). Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction. Genes Dev. **31**: 451–462.

Müller-Röber, B.T., Koßmann, J., Hannah, L.C., Willmitzer, L., and Sonnewald, U. (1990). One of two different *ADP-glucose pyrophosphorylase* genes from potato responds strongly to elevated levels of sucrose. MGG Mol. Gen. Genet. **224**: 136–146.

Nakata, P.A. and Okita, T.W. (1995). Differential regulation of ADP-glucose pyrophosphorylase in the sink and source tissues of potato. Plant Physiol. **108**: 361–368.

Oltman, A.E., Jervis, S.M., and Drake, M.A. (2014). Consumer attitudes and preferences for fresh market tomatoes. J. Food Sci. **79**: S2091–S2097.

Ouwerkerk, P.B.F. and Meijer, A.H. (2001). Yeast One-Hybrid Screening for DNA-Protein Interactions. Curr. Protoc. Mol. Biol. **55**: 12.12.1-12.12.12.

Park, S.W. and Chung, W. Il (1998). Molecular cloning and organ-specific expression of three isoforms of tomato *ADP-glucose pyrophosphorylase* gene. Gene **206**: 215–221.

Petreikov, M., Eisenstein, M., Yeselson, Y., Preiss, J., and Schaffer, A.A. (2010). Characterization of the AGPase large subunit isoforms from tomato indicates that the recombinant L3 subunit is active as a monomer. Biochem. J. **428**: 201–212.

Petreikov, M., Shen, S., Yeselson, Y., Levin, I., Bar, M., and Schaffer, A.A. (2006). Temporally extended gene expression of the *ADP-Glc pyrophosphorylase large subunit (AgpL1)* leads to increased enzyme activity in developing tomato fruit. Planta **224**: 1465–1479.

Petreikov, M., Yeselson, L., Shen, S., Levin, I., Schaffer, A.A., Dagan, B., Efrati, A., Bar, M., and Co, G.S. (2009). Carbohydrate balance and accumulation during development of near-isogenic tomato lines differing in the AGPase-L1 allele. J. Am. Soc. Hortic. Sci. **134**: 134–140.

Porebski, S., Bailey, L.G., and Baum, B.R. (1997). Modification of a CTAB DNA Extraction Protocol for Plants Containing High Polysaccharide and Polyphenol Components. Plant Mol. Biol. Report. **15**: 8–15.

Priest, H.D., Filichkin, S.A., and Mockler, T.C. (2009). cis-Regulatory elements in plant cell signaling. Curr. Opin. Plant Biol. **12**: 643–649.

Ricci, W.A. et al. (2019). Widespread long-range cis-regulatory elements in the maize genome. Nat. Plants **5**: 1237–1249.

Van Roekel, J.S.C., Damm, B., Melchers, L.S., Hoekema, A., and Davey, M.R. (1993). Factors influencing transformation frequency of tomato (*Lycopersicon esculentum*). Plant Cell Rep. **12**: 644–647.

Schaffer, a. a. and Petreikov, M. (1997a). Sucrose-to-Starch Metabolism in Tomato Fruit Undergoing Transient Starch Accumulation. Plant Physiol. **113**: 739–746.

Schaffer, A.A., Levin, I., Oguz, I., Petreikov, M., Cincarevsky, F., Yeselson, Y., Shen, S., Gilboa, N., and Bar, M. (2000). ADPglucose pyrophosphorylase activity and starch accumulation in immature tomato fruit: The effect of a Lycopersicon hirsutum-derived introgression encoding for the large subunit. Plant Sci. **152**: 135–144.

Schaffer, A.A. and Petreikov, M. (1997b). Inhibition of fructokinase and sucrose synthase by cytosolic levels of fructose in young tomato fruit undergoing transient starch synthesis. Physiol. Plant. **101**: 800–806.

Sherf, B.A., Navarro, S.L., Hannah, R.R., and Wood, K. V. (1996). Dual-Luciferase TM Reporter Assay : An Advanced Co-Reporter Technology Integrating Firefly and Renilla Luciferase Assays. Promega Notes Mag. Number **57**: 3–10.

Stein, O. and Granot, D. (2018). Plant Fructokinases: Evolutionary, Developmental, and Metabolic Aspects in Sink Tissues. Front. Plant Sci. **9**: 1–12.

Takeda, S., Mano, S., Ohto a, M., and Nakamura, K. (1994). Inhibitors of protein phosphatases 1 and 2A block the sugar-inducible gene expression in plants. Plant Physiol. **106**: 567–574.

Tieman, D. et al. (2012). The chemical interactions underlying tomato flavour preferences. Curr. Biol. **22**: 1035–1039.

Tieman, D. et al. (2017). A chemical genetic roadmap to improved tomato flavour. Science (80-.). **355**: 391–394.

Wang, F., Sanz, A., Brenner, M.L., and Smith, A. (1993). Sucrose synthase, starch accumulation, and tomato fruit sink strength. Plant Physiol. **101**: 321–327.

Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. PLoS One **6**: e16765.

Werner, S., Engler, C., Weber, E., Gruetzner, R., and Marillonnet, S. (2012a). Fast track assembly of multigene constructs using golden gate cloning and the MoClo system. Bioeng. Bugs **3**: 38–43.

Werner, S., Engler, C., Weber, E., Gruetzner, R., and Marillonnet, S. (2012b). Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. Bioengineered **3**: 38–43.

Xing, J., Li, X., Luo, Y., Gianfagna, T.J., and Janes, H.W. (2005). Isolation and expression analysis of two tomato *ADP-glucose pyrophosphorylase S (large) subunit* gene promoters. Plant Sci. **169**: 882–893.

Yamamoto, Y.Y., Yoshitsugu, T., Sakurai, T., Seki, M., Shinozaki, K., and Obokata, J. (2009). Heterogeneity of *Arabidopsis* core promoters revealed by high-density TSS analysis. Plant J. **60**: 350–362.

Yamasaki, K., Kigawa, T., Seki, M., Shinozaki, K., and Yokoyama, S. (2013). DNA-binding domains of plant-specific transcription factors: Structure, function, and evolution. Trends Plant Sci. **18**: 267–276.

Yang, J., Zhang, N., Zhou, X., Si, H., and Wang, D. (2016). Identification of four novel stu-miR169s and their target genes in *Solanum tuberosum* and expression profiles response to drought stress. Plant Syst. Evol. **302**: 55–66.

Yin, T., Wu, H., Zhang, S., Liu, J., Lu, H., Zhang, L., Xu, Y., and Chen, D. (2009). Two negative cis-regulatory regions involved in fruit-specific promoter activity from watermelon (*Citrullus vulgaris* S.). J. Exp. Bot. **60**: 169–185.

Yu, C.P., Lin, J.J., and Li, W.H. (2016). Positional distribution of transcription factor binding sites in *Arabidopsis thaliana*. Sci. Rep. **6**: 1–7.

Zhang, H., Lang, Z., and Zhu, J.K. (2018). Dynamics and function of DNA methylation in plants. Nat. Rev. Mol. Cell Biol. **19**: 489–506.

Zhang, T., Marand, A.P., and Jiang, J. (2016). PlantDHS: A database for DNase I hypersensitive sites in plants. Nucleic Acids Res. **44**: D1148–D1153.

Zhang, W., Zhang, T., Wu, Y., and Jiang, J. (2012). Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in *Arabidopsis*. Plant Cell **24**: 2719–2731.

CHAPTER 4

Modification of the transcriptional regulation of *LIN5* to increase Brix in tomato

Vera Veltkamp[1,2], Marrit Alderkamp[1], Moniek Schippers[1], Gerco C. Angenent[1,2], and Ruud A. de Maagd[2]

[1]Laboratory of Molecular Biology, Wageningen University, The Netherlands

[2]Bioscience, Wageningen Plant Research, The Netherlands

ABSTRACT

Increased activity of *Solanum lycopersicum Cell Wall Invertase 5 (LIN5)* has been linked to higher Soluble Solids Content (Brix) in ripe fruit. We studied the transcriptional regulation of *LIN5* to utilize this knowledge to increase its expression and to be able to create this desirable phenotype in any tomato cultivar. Yeast-one-hybrid and reporter assays of the *LIN5* promoter revealed that putative homolog of the *Arabidopsis* transcription factor (TFs) GATA9 (Solyc11g069510), the pepper homolog of MYB48 (Solyc06g005310), of MYB54 (Solyc10g083900), of R2R3MYB58 (Solyc11G073120), and a NAC TF (Solyc08g008660) may act as repressors while the C3H-type Zinc Finger TF C3H13 (Solyc01g100990), and the homolog of NAKED PINS IN YUC MUTANTS 1 (NPY1, Solyc08g006870) increased expression of a *LIN5* reporter in tobacco leaves. To modulate expression *in situ* by removing repressing *Cis*-regulatory Elements (CREs) and to try creating plants with a higher Brix, multiplexed systematic mutagenesis with CRISPR/Cas9 was used on the *LIN5* promoter in cv. Moneyberg. Increased Brix was found in several of the resulting mutants and in three lines that were studied further, altered expression was found. In a reporter assay comparing the mutant promoters with the wild-type promoter, the homolog of the *C. baccatum* HAT4 (Solyc08G078300), the homolog of *S. tuberosum* MYB1R1 (Solyc06G071230) and WRKY24 (Solyc09g066010) were identified as weak activators of the mutant promoters.

INTRODUCTION

The *Brix9-2-5* locus from an introgression line of *S. pennellii* in cultivated tomato was discovered to be responsible for high Soluble Solids content (measured as Brix) (Fridman et al., 2000, 2002). The partially dominant *Brix9-2-5* allele increased Brix by 20-35% in both determinate and indeterminate tomato varieties (Fridman et al., 2000), and *S. lycopersicum Invertase 5* (*LIN5*) was identified as the underlying gene (Eshed and Zamir, 1996; Baxter et al., 2005b, 2005a). Three amino acid substitutions in the *S. pennellii LIN5* allele compared to *S. lycopersicum* (Glu348Asp, Asn366Asp and Leu373Val) resulted in an enzymatically more active protein likely leading to the increased Brix, while there was no difference in expression between the *S. lycopersicum* and *S. pennellii* allele (Fridman et al., 2000, 2004). The observations with the *Brix9-2-5* QTL indicate that *LIN5* is an interesting target for quality breeding and increasing agroeconomic value.

LIN5 is a cell wall-bound invertase (or β-Fructosidase), catalysing the hydrolytic cleavage of the sucrose OC-bond resulting in the hexoses glucose and fructose (Nguyen-Quoc and Foyer, 2001; Matsukura, 2016). The combination of acidic cell wall-bound, acidic vacuolar and neutral cytoplasmic invertases drive sugar import and sink-strength, resulting in fruit with equimolar concentrations of glucose and fructose in cultivated tomato (Klann et al., 1992; Klee and Giovannoni, 2011; Matsukura, 2016).

Invertase proteins consist of a five bladed β-propeller and a β-sandwich module. At the bottom of negatively charged funnel-shaped depression of the β-propeller are three carboxylates that are essential for catalysis (Alberto et al., 2004). The β-propeller is composed of the firs ±300 residues and the C-terminal β-sandwich by the next ±130 residues, linked with a 10-residue linker. In the β-propeller domain the conserved catalytic domain "WEC(P/V)DF " is found in all invertase types, where the cell wall invertases have a proline residue and cytosolic or vacuolar invertases have a valine residue (Godt and Roitsch, 1997). In addition, a "NPDN" and a "RDP" motif , both essential for sucrose recognition and stable binding, are found in the β-propeller domain (Alberto et al., 2004; Lammens et al., 2009; Slugina et al., 2019).

In tomato, four functional cell wall invertase genes have been identified. These most likely developed from a common ancestor after subsequent tandem and segmental duplications within the tomato genome (Godt and Roitsch, 1997; Fridman and Zamir, 2003). *LIN5* (*SlCWIN1*) and *LIN7* (*SlCWIN3*) are located on chromosome 9 and their expression is specific for flowers and fruits (Goetz et al., 2001; Fridman and Zamir, 2003; Proels et al., 2003). *LIN6* (*SlCWIN2*) and *LIN8* (*SlCWIN4*), on chromosome 10, are expressed specifically in vegetative tissues. In addition, five defective cell wall invertase genes (*SlCWIN5-8* and *SldeCWIN1*) have been identified (Wei et al., 2020). *SlCWIN5-8* lack the "NPDN" motif. *SldeCWIN1* has a D239A substitution and is unable to hydrolyse sucrose (Le Roy et al., 2013). They do all contain the catalytic "WEC(P/V)DF" motif.

LIN5 is expressed in floral organs and in early fruit development (Fridman and Zamir, 2003; Proels et al., 2003). Four hours after pollination, *LIN5* transcript level is increased in the style. Two days after pollination, *LIN5* expression is high in the ovaries (Shen et al., 2019). There is a burst of invertase activity during ovary to fruit transition, probably facilitating phloem unloading through the apoplasmic pathway, stimulating cell division (Palmer et al., 2015). During the early stage of cell division in the fruit, LIN5 is active in the conductive tissues, the placenta and the pericarp (Fridman and Zamir, 2003). RNA interference (RNAi) of *LIN5* resulted in increased numbers of petals and sepals, aberrant pollen morphology, reduction in pollen tube elongation, increased rates of fruit abortion, reduced fruit size, and decreased Brix (Zanor et al., 2009). Combined with the observation of the *Brix9-2-5* phenotype this shows that *LIN5* has an important function in early fruit set and development and through this or other mechanisms, on Brix in the ripe fruit.

The importance of *LIN5* for tomato Brix was further demonstrated by the silencing of the posttranslational inhibitor of LIN5, Cell-wall inhibitor of β-fructosidase (SlCIF1 or INVINH1), leading to a higher sugar content (Jin et al., 2009b). CIF1 and LIN5 are co-localized at the cell-walls of sieve elements and physically interact. Before anthesis, *LIN5* and *CIF1* transcripts are dispersed in the columella, placenta and ovule (Palmer et al., 2015). After anthesis, both are localized in the vasculature of these tissues. *LIN5*

expression after fertilization is localized especially near the vasculature in the columella and placenta and to a lesser extent in the pericarp.

Thus far, most research has focused on understanding LIN5 activity and posttranslational regulation. Less is known about the transcriptional regulation of *LIN5*. It was determined that the first kb upstream of *LIN5* directed pollen and floral organ-specific expression, while the 1.5 to 2.5 kb upstream sequence responds to Gibberellic acid, auxin and abscisic acid (Proels et al., 2003). We aimed at elucidating the regulation of *LIN5* expression further by studying the promoter. Interacting Transcription Factors (TFs) were identified with Yeast-one-hybrid (Y1H) assays and transcriptional regulation was confirmed with Reporter Assays. To modulate expression *in situ*, multiplexed mutagenesis with CRISPR/Cas9 was used on the *LIN5* promoter in cv. Moneyberg. We hypothesized that by removing *Cis*-regulatory Elements (CREs), expression could be modified locally without severe pleiotropic effects. The obtained mutants were phenotyped for soluble solids content ($^{\circ}$Brix) and expression of *LIN5*.

RESULTS

TWO TRANSPOSABLE ELEMENTS IN THE *LIN5* PROMOTER

The upstream region of *S. lycopersicum Invertase 5* (*LIN5*) coding sequence was profiled to identify putatively relevant regions for the control of its expression. The first gene upstream of *LIN5* is 4.5 kb away in the same orientation. Thus, the entire 4.5 kb could contain *Cis*-Regulatory Elements (CREs) for *LIN5*. Publicly available DNAse I hypersensitive sites Sequencing (DNAse-seq) data in the FruitEncode project (Lu et al., 2017) allowed us to identify regions of open chromatin in the selected 4.5 kb (**Figure 1a**). Open chromatin regions represent regions that are accessible to Transcription Factors (TFs) and may contain active CREs at a given developmental stage (Boyle et al., 2008; Zhang et al., 2012, 2016; Jiang, 2015). Significant open chromatin regions were identified by comparing read density of a 250 bp sliding window against a 50 kb background to determine if it fell above a cut-off value (Koohy et al., 2014). However, none were found in the data of the available developmental stages (leaf, fruits 7, 17 and 47 Days After Anthesis (DAA)). *LIN5* expression peaks before 5 DAA (Godt and Roitsch, 1997; Fridman et al., 2002; Fridman and Zamir, 2003; Proels et al., 2003), thus the most relevant data were probably not available in the database. The FruitEncode project also stored DNA methylation data. Methylation is associated with close-packed DNA, which is harder to access by TFs and thus thought less likely to harbour (active) CREs (Zhang et al., 2018a; Gallego-Bartolomé, 2020). Two highly methylated regions between 2000 and 4000 bp upstream of the Transcription Start Site were of particular interest (**Figure 1b**).

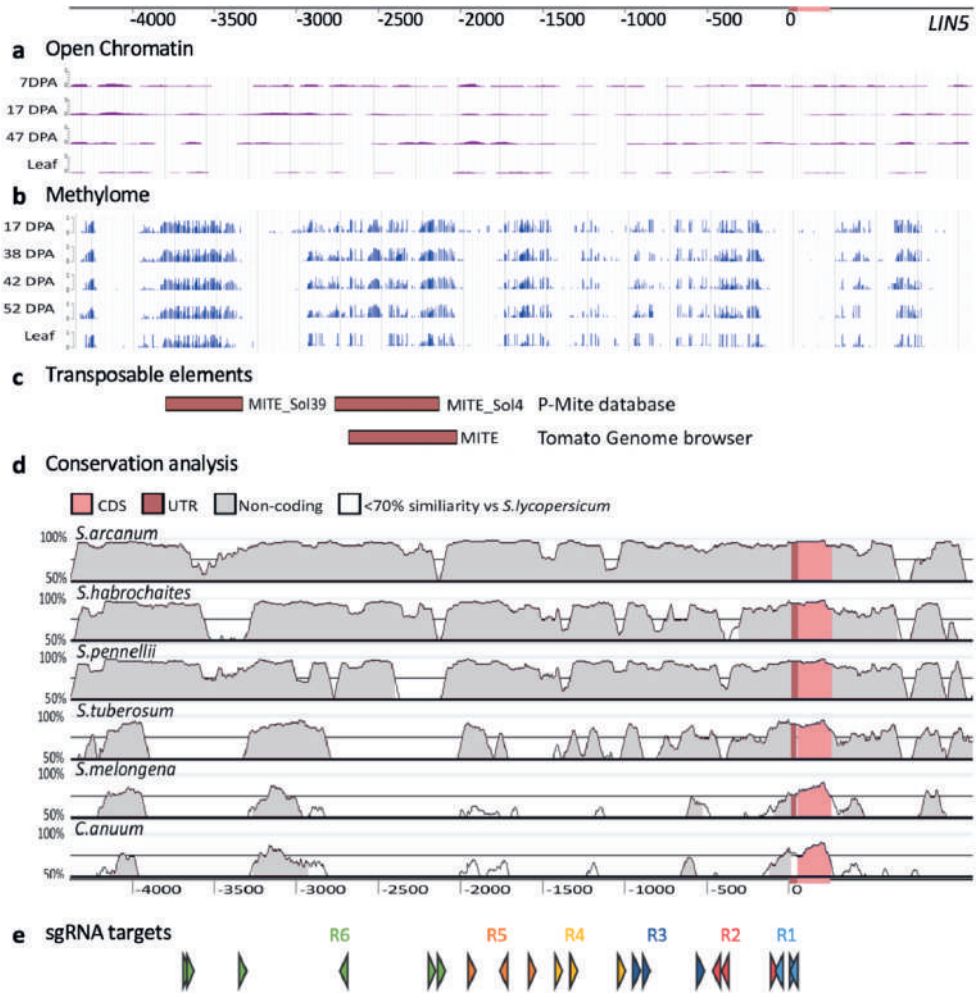


Figure 1: Overview of the *LIN5* promoter. In the top row, the gene model is shown in pink (exons) and dark pink (UTRs). The scale is in bp, relative to the Transcription Start Site. **(a)** Local DNAse I-hypersensitive sites (DHS) profiles of *LIN5*. The four rows (purple) show the DHS profiles of fruit at 7, 17 and 47 DAA as well as of leaf tissue on a scale of 0 to 5. No significant open regions were found (determined by peak calling (sign.<.05, background comparison: 50kb). **(b)** DNA methylation profiles of 17, 42, 52 DAA fruit and leaf on a scale of 0 to 1. Methylation and DHS data was retrieved from FruitENCODE (http://www.epigenome.cuhk.edu.hk/encode.html, SI2.50 tomato genome assembly, Lu et al., 2017). **(c)** Position of Miniature Inverted–Repeat Transposable Elements (MITEs) identified in the Tomato Model Organism System Database Project (GMOD) and P-MITE database. **(d)** Conservation analysis of the *LIN5* promoter with MVISTA. *LIN5* was used as a reference point for the non-coding adjacent regions. The orthologs are from (starting from the upper row): *S. arcanum, S. habrochaites* (the 3' part of the gene was not available for analysis), *S. pennellii, S. tuberosum*, *S. melongena,* and *C. annuum*. The order was determined by the measure of relatedness. Conservation was determined by LAGAN, a global multiple alignment tool, with a RankVISTA cut-off probability value of 0.05 and a window size of 100 bp. If a particular region is above the pre-determined cut-off value (70%), then the region is represented by the presence of colour: pink and dark pink regions represent exons and UTRs respectively, while grey represents non-coding regions. **(e)** The CRISPR/Cas9 targets used in this study. The colours represent combinations of targets that were combined in one vector. The different groups are referred to as region 1 (closest to the UTR, light blue), region 2 (red), region 3 (dark blue), region 4 (yellow), region 5 (orange) and region 6 (green).

These two regions were identified, by homology, as Miniature Inverted–Repeat Transposable Elements (MITEs, **Figure 1c)** (Stein et al., 2002; Chen et al., 2014). From conservation analysis using mVISTA (Frazer et al., 2004), it became apparent that these two MITEs are relatively conserved in the wild tomato varieties *S. arcanum, S. habrochaites* and *S. pennellii*, while completely absent in *S. tuberosum* (potato), *S. melongena* (eggplant) and *C. annuum* (pepper) (**Figure 1d**). Genome-wide expression analysis has shown that genes that are associated with MITEs have significantly lower expression than genes that are not (Lu et al., 2012). Thus, the MITEs represent an interesting region that might play a role in repressing *LIN5* expression. Apart from these MITEs the proximal region upstream of *LIN5* is most conserved and could contain potential CREs.

FOUR TF REPRESSING THE EXPRESSION OF *LIN5 IN VITRO* WERE IDENTIFIED

Since the proximal promoter was highly conserved, we chose to screen 400 bp immediately upstream of *LIN5* (from -369 to +31 relative to the transcription start site) for putative binding transcription factors in a Yeast-one-hybrid assay (Y1H). This region was relatively conserved in all species (**Figure 1d**). A matrix-based TF library screen was done twice with a library containing more than 2000 *Arabidopsis* TFs (Castrillo et al., 2011). The first Y1H screen resulted in 63 binding events (**Figure 2a**). The second screen yielded 35 hits, of which 9 overlapped with the first screen. The combined total 89 TFs were screened a third time. This confirmed 14 hits, of which 8 belonged to the initial 9 overlapping hits (**Figure 2a, Table S2**). From these *Arabidopsis* hits, tomato orthologs were identified via best-reciprocal hit search (**Table S2**). Based on their expression in fruit, four tomato TFs were selected for a promoter reporter assay: homologs of Alfin-Like 6 (AL6, Solyc01g102750), of AL8 (Solyc01g102760), of CONSTANS7 (CO7, Solyc02g093590), and of Zinc Finger Protein 593 (ZFP593, Solyc09g008230). These TFs were cloned into expression cassettes and transformed to *Agrobacterium* (**Figure 2b**). Co-infiltration with a construct that contained 1258 bp of the *LIN5* upstream region, fused with a *Firefly luciferase* cassette and a normalization vector containing *Renilla luciferase* was done in tobacco leaves. If a TF interacts directly or indirectly with the *LIN5* promoter, a Firefly luciferase signal is generated (**Figure 2c**). Expression levels in the presence of TFs were normalized to the basal activity without added TF (set at 1). As a positive control, a p2xCAMV-35S:Firefly luciferase reporter was used (**Figure 2d**). None of the tomato TFs identified in the Y1H changed the basal expression of the *LIN5* promoter (**Figure 2e**). To find potential regulators, an array of different tomato TFs were tested in a second promoter-reporter assay. These were selected in **Chapter 3** for analysis of interaction with the *AGPL1* promoter. Since all the TFs are expressed in early fruit development, they were a logical set of candidates to test on the *LIN5* promoter as well. The promoter reporter assay indeed yielded several potential repressors, the homolog of *Arabidopsis* GATA9 (Solyc11g069510), the MYB TF MYB48 (Solyc06g005310), MYB54 (Solyc10g083900) and a NAC-domain protein (NAC, Solyc08g008660) (**Figure 2f**).

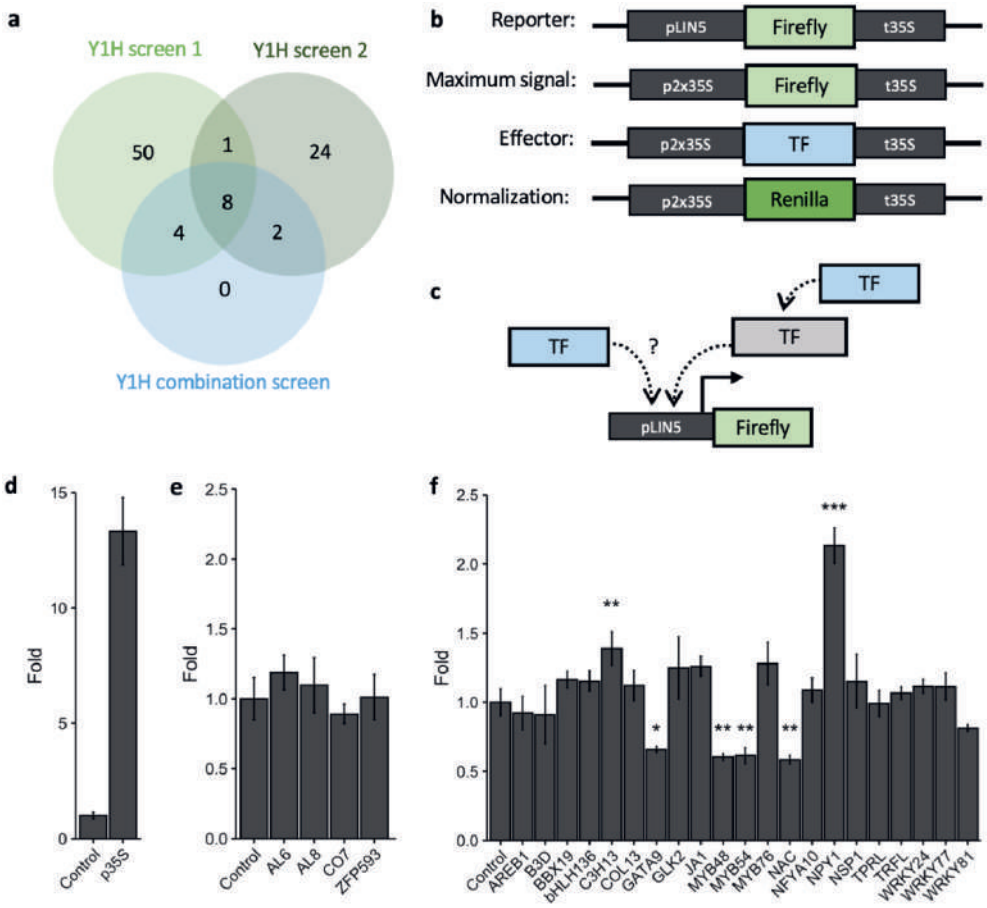


Figure 2: (a) Y1H results VENN diagram for the three reported screens. Y1H results from the combination screen are represented in **Table S1. (b)** Schematic diagrams of promoter-reporter and TF effector expression constructs used in the promoter reporter assays. The p2xCAMV 35S (p2x35S) promoter was used in the assay control, effector, and normalization strains. All parts were inserted into separate Golden Gate pL1 expression vectors and transformed to *A. tumefaciens* C58c1. **(c)** Schematic representation of the promoter reporter assays performed in tobacco (*Nicotiana benthamiana*) to examine interaction between different TFs and the *LIN5* promoter. **(d-f)** Transient expression promoter-reporter assay (PRA) results performed in *Nicotiana benthamiana* to examine interaction between different TFs and the *LIN5* promoter. The Firefly Luciferase:Renilla ratio of the reporter without a TF (control) was set to 1 and the fold change of the promoter co-infiltrated with TFs are represented in the figure. Each value represents the mean ± SE of three biological replicates (three infiltrated leaves, processed separately). Each biological replicate is the average of three technical replicates (tissue harvested from one leaf). Significant differences are represented by an asterisk (* p<0.05, ** p<0.01, *** p<0.001). **(d)** Control PRA experiment where the *LIN5* promoter without effector and the p35S are used. **(e)** The PRA experiment with the TFs selected based on the Y1H screens. **(f)** The PRA experiment with a selection of TFs that are expressed in the early stages of fruit development

For GATA9, MYB48 and MYB54 a significant binding site could be identified in the LIN5 promoter (**Table S3**). In addition, C3H-type Zinc Finger C3H13 and NAKED PINS IN YUC MUTANTS 1 (NPY1) increased activation of *LIN5*.

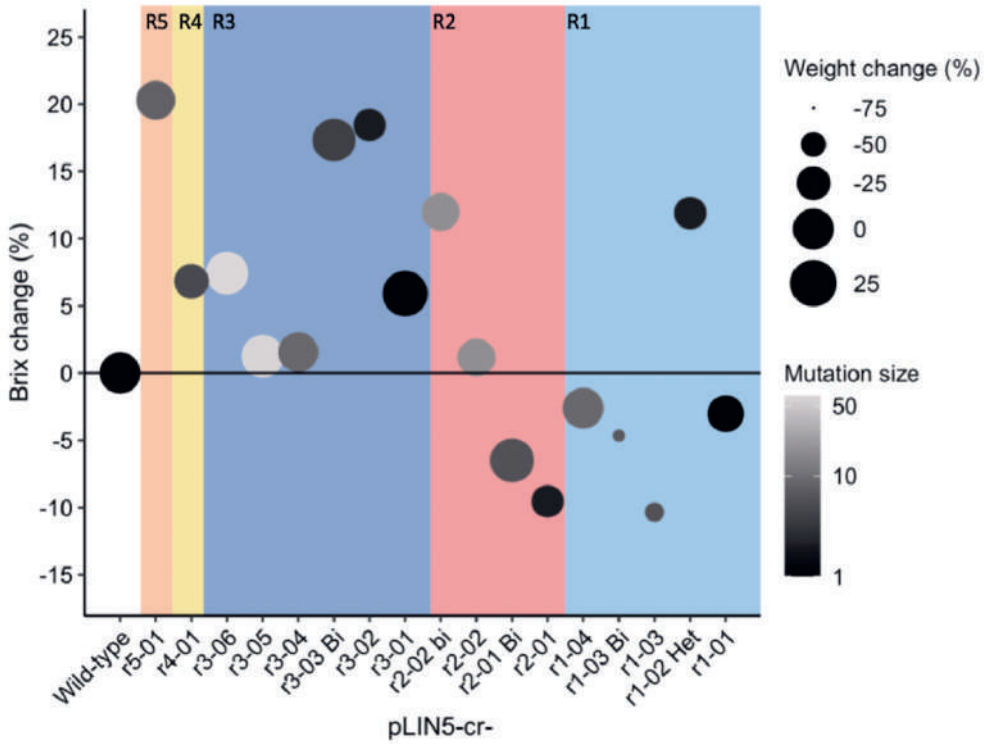


Figure 3: Phenotype of T_1 *LIN5* promoter mutants. The vertical position of a circle represents the average change in Brix (%) compared to a wild-type plant grown at the same time as the mutant. The size of the circle represents the change in weight (%) compared to wild-type fruits. The shading of the circles indicates the size of the mutations in the promoter. The mutants are ranked according to the different mutated targets, distributed over the different regions (**Figure 1**). For each mutant, three trusses with each 6 fruits of one or two plants were measured. Plants were homozygous unless specified otherwise (Bi: bi-allelic, Het: heterozygous).

PROMOTER MUTATIONS INCREASE *LIN5* EXPRESSION AND BRIX

In parallel with understanding the promoter architecture by Y1H and promoter-reporter assays, a random promoter deletion approach was used to understand and modulate expression. CRISPR/Cas9-mutagenesis was applied to create allelic variation in the promoter of *LIN5*. We hypothesized that by studying the effect of different promoter mutations on expression and phenotype, regulatory regions could be identified and a mutant with a higher *LIN5* expression and increased Soluble Solids Content (°Brix) might be identified. To obtain systematic allelic variation, twenty-one single guide RNA (sgRNA) were designed to target six different regions (R1, R2, R3, R4, R5, R6) in the promoter. R1-R5 were targeted by three sgRNAs each and R6 was targeted by six sgRNAs (**Figure 1e**). The R6 targets were selected to remove the identified MITEs. Primary transformants (T_0) containing CRISPR-vectors targeting the six respective regions were genotyped for mutations in the promoter. In total, 125 transformants were obtained. However, none of the primary transformants targeting R6 had a mutation.

In the progeny (T_1), plants were selected for the mutations, and for the absence of the CRISPR/Cas9 T-DNA insertion. In most cases, homozygous plants were used in the T_1, but in some bi-allelic or heterozygous variants were used. All alleles were genotyped and named *pLIN5-cr-rx-y* (where "*rx*" indicates the region and, and "*y*" the allele number, **Figure S1**). In R5, just two mutant alleles were found. *pLIN5-cr-r5-01* had an 8-bp deletion. Another allele, *pLIN5-cr-r5-02,* had a 5-bp deletion in target 5.3, respectively. Two plants homozygous for the *r5-02* allele were grown in the T_1, but both plants stayed small and did not produce fruit. For R4 only allele *pLIN5-cr-r4-01*, with a 5-bp deletion in target 4.2 was phenotyped. Multiple alleles were obtained in R3, R2, and R1 (**Figure S1**). Apart from a target-to-target deletion in *pLIN5-cr-r3-05*, all obtained mutations were small insertions or deletions at the targeted cut-site. A clear difference in efficacy could be observed between sgRNAs. Some resulted in a mutation in virtually all transformants (such as for targets 3.1, 3.2, 2.3, 1.3, **Figure S1**), while others never did. The lack of multiple effective guides in each of the different regions could explain the lack of target-to-target, and thus bigger, deletions.

17 transformants were selected for a first screen of interesting phenotypes in the T_1 generation. One or two plants per mutant allele were phenotyped for Brix and fruit weight (**Figure 3, Figure S2**). The phenotyped promoter mutations resulted in a variation of Brix, ranging between a decrease of 10% (*pLIN5-cr-r1-03*) to an increase of almost 20% (*r5-01*) compared to wild-type. Both lines with a higher Brix (*r4-01* and *r3-02*) and a lower Brix (*r2-01* and *r1-03),* had a decreased fruit weight compared to wild-type fruits. Other alleles with a higher Brix displayed no effect on fruit weight (*r5-01, r3-06, r3-03* (bi-allelic), *r2-02* (bi-allelic) and *r1-02* (heterozygous). The line with *r2-01* (bi-allelic) had a decreased Brix without an effect on weight. The line with *r3-01* had a slightly higher Brix and a higher weight. From this data, no correlation between Brix and weight could be observed. In addition, there was little consistency in phenotype of alleles within a region. Only mutations in R3, which is near a conserved region in the *LIN5* promoter mostly led to increased Brix.

The lines with the highest Brix, *pLIN5-cr-r5-01, r3-03* and *r3-02*, were selected for further analysis in T_2. In addition, lines with alleles *r1-04 and r3-04* were studied (**Figure 4**). In this generation, five homozygous plants per allele were grown for statistical analysis. Two trusses with six fruits were used for measuring Brix and weight. *pLIN5-cr-r1-04* had an increased Brix and decreased weight (**Figure 5a**). *r5-01* had an increased Brix without decreased weight in the T_1, yet this phenotype was not reproduced in the T_2. Instead, the Brix was not different from wild-type and fruit weight decreased substantially. The increased Brix in the T_1 of *r3-02* was not reproduced either. Additionally, a decreased weight was found. *r3-03* was bi-allelic in the T_1. The T_2 segregation resulted in two very similar alleles (**Figure 4**), and both alleles resulted in a higher Brix.

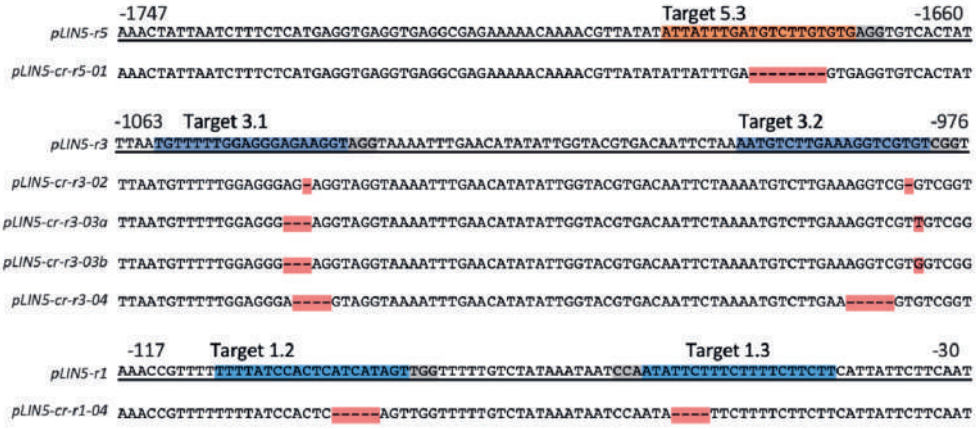


Figure 4: Alignments of the wild-type and mutant promoters. Numbers are relative to the start codon of *LIN5*. sgRNA target regions are highlighted with a grey PAM site. Mutations are highlighted in red. The a and b alleles are given.

pLIN5-cr-r3-03b had a decreased weight, while the weight of *r3-03a* did not differ from wild-type. Both *r3-03a* and *r3-03b* had the same Brix yield index (soluble solids content (Brix) × fruit yield (weight)) as wild-type. *pLIN5-cr-r3-04* had a 4 bp deletion in target 3.1 and a 5 bp deletion in target 3.2. Both Brix and weight decreased for this line.

LIN5 expression in ovaries from 1 day before anthesis until 5 Days After Anthesis (DAA) was analysed for mutant T_2 generation and compared to wild-type (**Figure 5b**). Only the lines that had a higher Brix were studied, as well as *r3-04*. *LIN5* expression increased in *pLIN5-cr3-03a* before anthesis, while expression in *r3-04* increased at 4 and 5 DAA compared to wild-type. At 0 DAA *LIN5* expression in *r1-04* decreased slightly compared to wild-type. From 3 DAA, expression increased slightly compared to wild-type, but not significantly.

LIN5 activity is capped in unfertilized flowers by the post-translational inhibitor Cell-wall inhibitor of β-fructosidase (CIF1 or INVINH1) and a defective cell wall invertase, deCWIN1. deCWIN1 is proposed to act as a catalyst on both the activity of LIN5 and CIF1 (Le Roy et al., 2013). Expression of the *LIN5* post-translational inhibitor *Cell-wall inhibitor of β-fructosidase* (*SlCIF1* or *INVINH1*), was unaltered in the mutants compared to wild-type (**Figure S4**). A significant increased *deCWIN1* expression was observed at 1 and 2 DAA in *r3-03a*, at 4 DAA in *r1-04* and at 4-5 DAA in *r3-04* (**Figure S4**). However, *deCWIN1* expression values were very low overall. Thus, a feedback-loop between *LIN5*, *CIF1* and *deCWIN1* seems unlikely.

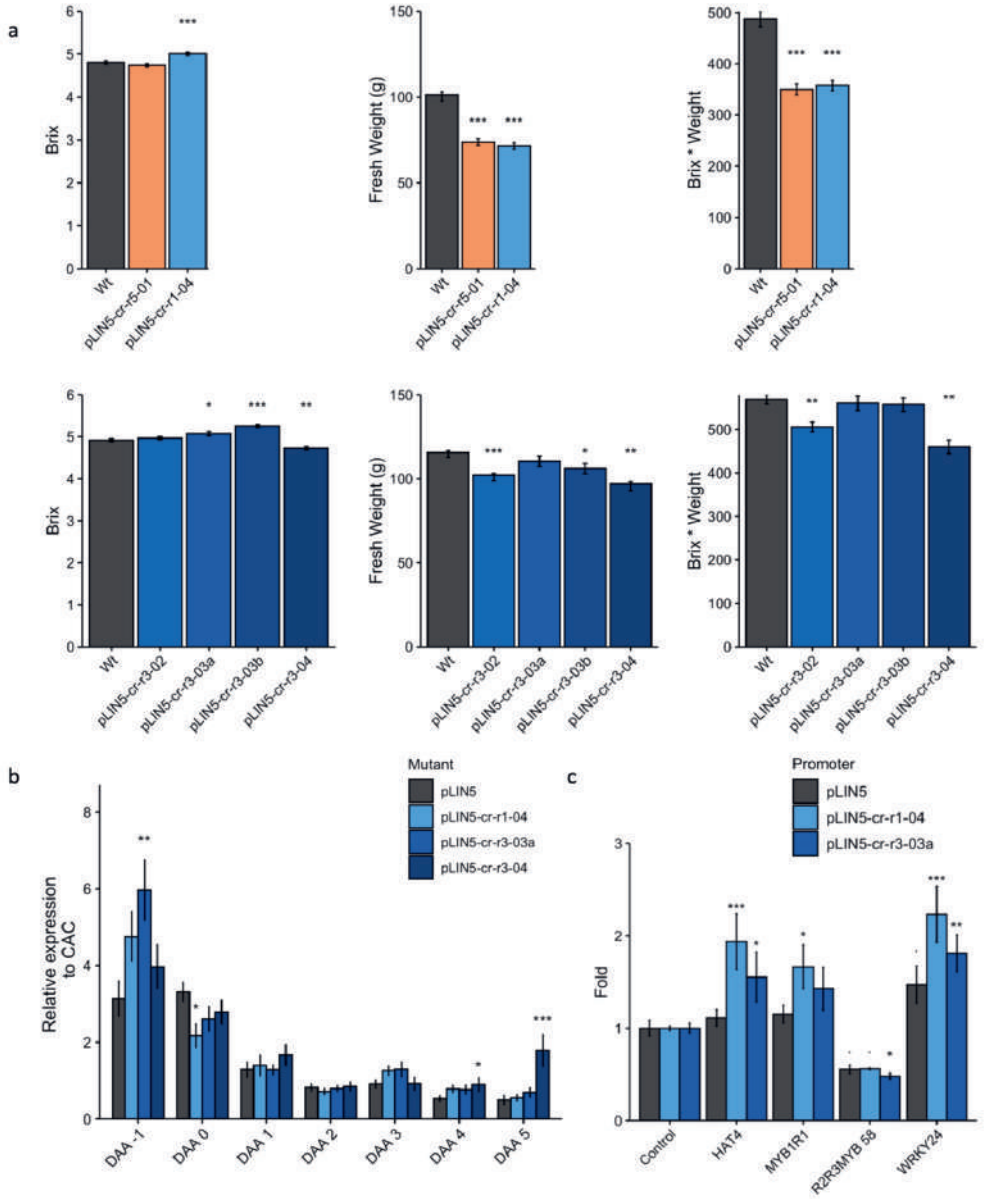


Figure 5: (a) Brix, weight, and Brix*Weight of T_2 fruits from the third and fifth truss of five plants per genotype. Trusses were pruned to six fruits and harvested at breaker +7 days. Wt: wild-type. The soluble solids content (Brix) × fruit yield (weight) gives the Brix yield index **(b)** The second, fourth and sixth trusses were used for gene expression analysis. Per timepoint, a pool of three fruits was collected per plant. **(c)** Transient expression promoter-reporter assay (PRA) results performed in (*Nicotiana benthamiana*) to examine interaction between different TFs and the wild-type and mutant *LIN5* promoters from *S. lycopersicum* cv. Moneyberg (*pLIN5*). The Firefly :Renilla luciferase ratio of the reporter without a TF (control) was set to 1. The fold change of the promoters co-infiltrated with TFs are represented in the figure. Each value represents the mean ± SE of three biological replicates. Statistically significant differences are represented by an asterisk (* p<0.05, ** p<0.01, *** p<0.001).

TF BINDING SITES IN THE *LIN5* PROMOTER

Apart from finding a mutation that would lead to increased *LIN5* expression and increased Brix, the second purpose of this study was to assess if promoter mutagenesis could identify CREs. Motif finding was used to predict which TF (families) could bind to the mutated regions. First, conserved 50-bp motifs present around the targets were identified using conservation analysis with the Multiple Expectation Maximization for Motif Elicitation (MEME)-suite using the same *LIN5* orthologs from *S. arcanum, S. habrochaites, S. pennellii, S. tuberosum, S. melongena* and *C. annuum* (Bailey and Elkan, 1994). Three conserved motives (1-3) were identified that overlapped with targets 3.1, 3.2, and 1.2, respectively (**Figure S3a**). The MEME-suite motif comparison tool (TOMTOM) was used to compare the three motifs against the JASPAR plant core TF database to find matches between TF-binding sites and the three motifs (Gupta et al., 2007; Khan et al., 2018). Motif 1, the region surrounding target 3.1 and the mutations in *pLIN5-cr-r3-03 and r3-04*, contained a putative MYB TF-binding site adjacent to the mutated site (**Figure S3b**). Motif 2, at target 3.2, did not give any binding site close to the mutated region in *pLIN5-cr-r3-04*.

Analysis of the mutation in motif 1, at target 1.2, identified a potential Homeobox Leucine Zipper from Arabidopsis thaliana 1 (AtHAT1) binding site. In addition, 5' of the mutation we identified an AtMYBR1 binding motif, which was annotated as a Sugar Repressive Element in *Arabidopsis* and thus highly interesting in this context (Van Bel et al., 2018). For these candidates, tomato-orthologs were selected based on best-reciprocal hits search. As MYBs are part of a large family of TF, a selection was made for TFs that were co-expressed with *LIN5*. From this analysis, tomato R2R3MYB58 (Solyc11G073120) was chosen for motif 1 (deletion in R3) while the Homeobox-leucine zipper protein HAT4 (Solyc08G078300, based on homolog of *C. baccatum*) and MYB1R1 (Solyc06G071230, based on the homolog of *S. tuberosum*) were chosen for motif 3 (deletion in R1). These three TFs were tested for affecting transcriptional activity of the wild-type and mutant promoters (*pLIN5-cr-r1-04* and *pLIN5-cr-r3-03a*) in a reporter assay (**Figure 5c**). WRKY24 was taken as a non-activating control for the wild-type promoter, based on the analysis shown in **Figure 2f**. R2R3MYB58 repressed activity in all promoters, with the strongest repressing response in *pLIN5-cr-r3-03a*. *HAT4, MYB1R1 and WRKY24* increased the luciferase signal in *pLIN5-cr-r1-04*, while none of these TFs influenced the wild-type promoter. HAT4 and WRKY24 also activated *pLIN5-cr-r3-03a*, but MYB1R1 did not (Fig. 5c).

DISCUSSION

This study aimed at increasing the soluble solids (Brix) in tomato, an important quality aspect (Baldwin et al., 2000; Tieman et al., 2012, 2017; Oltman et al., 2014). The ovary and fruit-specific cell wall invertase *S. lycopersicum Invertase 5* (*LIN5*), the underlying gene of the *Brix9-2-5* Quantitative Trait Locus, is known to play a role in the

accumulation of Brix (sugar content)(Fridman and Zamir, 2003; Proels et al., 2003; Zanor et al., 2009; Shen et al., 2019). We hypothesized that if we could identify and remove repressing *Cis*-Regulatory Elements (CREs) in the *LIN5* promoter, it might enhance expression and create a phenotype with a higher Brix in any desired cultivar.

PROMOTER PROFILING

To modify the expression of *LIN5*, a thorough understanding of the promoter and regulatory network influencing expression is preferable (Li et al., 2020). Our first step was to use conservation analysis and DNase-seq data to find open chromatin regions likely harbouring CREs (Mavrich et al., 2008; Jiang and Pugh, 2009; Jin et al., 2009a; Mueller et al., 2017). The available data from FruitEncode did not yield significantly open chromatin in the LIN5 locus (Lu et al., 2017). However, the earliest time point available was 7 Days After Anthesis (DAA) and *LIN5* expression is high between -1 and 5 DAA. Moreover, as *LIN5* is probably only expressed in a small number of cells around the vasculature (Palmer et al., 2015), a DNase-seq signal corresponding to open chromatin and expression may be too diluted to detect. The conservation analysis in combination with methylation data and screening for transposable elements revealed two Miniature Inverted–Repeat Transposable Elements (MITEs). We designed a CRISPR/Cas9 strategy to remove these MITEs, but the generation of mutants in this region failed. The rest of our CRISPR/Cas9 strategy focussed on the first 2500 bp upstream of the *LIN5* open reading frame. The introns, especially the first, could contain CREs as well. The first intron had some conserved areas and most striking was, its lack of methylation. A CACTA transposon-like insertion was identified in the first intron (Proels and Roitsch, 2006), which could have led to inactivation (Lönnig and Saedler, 1997). Fridman et al. speculated that the third intron could explain the Brix9-2-5 allele, as it is different between *S. lycopersicum* and *S. pennellii* (Fridman et al., 2000). However, there was no difference in *LIN5* expression between the *S. lycopersicum* and *S. pennellii* alleles and the role of the third intron was therefore never tested (Fridman et al., 2000, 2004).

PROMOTER MUTATIONS INFLUENCE BRIX, WITHOUT A CORRELATION WITH FRUIT WEIGHT

Random systematic allelic variation was created in the *LIN5* promoter by using multiplexed CRISPR/Cas9-mutagenesis. We observed mostly small deletion at the effective sgRNAs, while larger, possibly more informative target-to-target deletions were not common outcomes of an approach using 3 sgRNAs for each region. 17 mutant alleles were selected and analysed in the first generation (T_1) as transgene-free segregants. In several of these mutant lines, Brix was increased. Of these, six alleles were studied further in the T_2 generation. Instead of one plant per genotype, five were taken to increase the accuracy of the measurements. This was necessary, as results for several lines that had a higher Brix or weight, could not be replicated in the T_2 generation. In our study the relationship between Brix and weight varied between mutant lines, and overall, there was no correlation between them. For the original *Brix9-*

2-5 allele no negative effect on total yield was reported, however fruit weight did decrease in the introgression line (Eshed and Zamir, 1996; Fridman et al., 2000, 2002). In general, sugar content is an important part of fruit quality, but it is often associated with lower fruit weight. Sugar content of tomato has been one of the major traits that has suffered during domestication, as was demonstrated by Genome-Wide association studies (GWAS) (Tieman et al., 2017; Zhao et al., 2019). In this GWAS, *LIN5* had significant association with sugar content, but was also identified as a region affected by domestication and improvement sweeps during the selection of larger fruits. A negative correlation between the occurrence of the high Brix allele of *LIN5* and fruit size was found in this study (Tieman et al., 2017). Almost all modern cultivars contain the reference alleles at this locus, i.e. the allele that is not associated with increased sugar content. Our study has shown that increasing the Brix through *LIN5* promoter mutations, can result in lines that do not have a weight penalty, such as for *pLIN5-cr-r3-01, r3-03a* and *r3-03b*. However, other lines did have a decreased fruit weight, so the effects must be measured for each different mutant separately. In addition, the total Brix*yield index can depend on the use of determinate or indeterminate varieties.

LIN5 FUNCTIONS IN EARLY FRUIT SET AND GROWTH

LIN5 expression in *pLIN5-cr-r3-03a* (higher Brix, similar weight as wild-type) was almost twice that of wild-type, before anthesis. On the other hand, a line with decreased Brix (and weight), containing allele *r3-04,* had increased *LIN5* expression at 4 and 5 DAA. These results suggest that increased *LIN5* expression needs to occur very early to result in increased Brix. Upon anthesis, there is a burst of cell wall invertase activity, increasing phloem unloading and stimulating cell division (Palmer et al., 2015). Elevated cell wall invertase activity in ovaries enhanced the expression of genes involved in the cell cycle and cell wall synthesis, but reduced the expression of genes associated with photosynthesis and protein degradation (Ru et al., 2017). RNAi of *LIN5* led to more fruit abortion, further supporting its crucial role in fruit set (Zanor et al., 2009). Elevated cell wall invertase activity in early fruit development (10-15 DAA) led to increased expression of *Sugar Will Eventually Be Exported 12C (SWEET12c)* and *Hexose Transporter 2 (HT2),* promoting rapid fruit expansion (Ru et al., 2020). A feedback loop between increased sugar transport and *LIN5* is further demonstrated by overexpression of the apple *MdHT2.2* in tomato, where *LIN5* was highly upregulated (Wang et al., 2019). Combined with the early expression of *LIN5* in fruit development (Shen et al., 2019), we can conclude that cell wall invertase plays a key role in normal seed development, especially from an early stage. Our work confirms the role of LIN5 in early fruit set and growth, as a higher expression at a very early stage, before anthesis, can lead to similar-sized fruits with a higher Brix as was demonstrated by the mutant allele *pLIN5-cr-r3-03a*.

TF INTERACTIONS

In a promoter reporter assay comparing the mutant promoters with a wild-type promoter, the Homeobox-leucine zipper protein *HAT4, MYB1R1* and *WRKY24* were identified as weak activators of the mutant's promoters. Y1H and promoter reporter assays of the wild-type promoter further revealed C3H-type Zinc Finger *C3H13* and the homolog of pepper *NAKED PINS IN YUC MUTANTS 1* (*NPY1)* as activators of *LIN5*. The homolog of the *Arabidopsis* GATA-motif containing TF *GATA9, MYB48, MYB54, MYB58* and a *NAC* TF (*NAM, ATAF1,2,* or *CUC2*) behaved as repressors. As increased *LIN5* expression is correlated with higher fruit quality, the repressors are interesting candidates for further study. In general, the GATA-domain TFs, MYB TFs and NAC TFs are all part of large transcription factor families and are involved in diverse developmental and signalling pathways (Chi et al., 2013; Li et al., 2016; Mohanta et al., 2020). *GATA9* is one of the GATA-TFs that is mostly expressed in flowers and fruits (Yuan et al., 2018). MYB TFs that harbour an ethylene-responsive element binding factor-associated amphiphilic repression motif (EAR-motif) are associated with repressing transcription of their targets (Chen et al., 2019). EAR-motifs have been found to be "LxLxL" or "DLNxxP" (where x can be one of the 20 common amino acids) (Ohta et al., 2001; Kagale and Rozwadowski, 2011). *MYB48* has an EAR-motif as "LILEL" at position 82, while MYB54 has a less conserved "LIIRL*"* at position 71. Both act as repressors in our study. MYB76 also contains an EAR-motif but did not result in repression. This indicates that their different binding affinities may play a role in determining their effect on promoter activity. MYBs regulating cell wall invertases has been found in *Arabidopsis* as well. In *Arabidopsis,* an *in silico* and transgenics approach determined that AtMYB21, together with AtARF6, AtARF8, AtAP3 and AtCRC were likely regulators of the *LIN5* and *LIN7 cell wall invertase* orthologs *AtCWIN2* and *AtCWIN4* (Fridman and Zamir, 2003; Wang and Ruan, 2013; Ruan, 2014; Liao et al., 2020; Li et al., 2021). Another predicted *Arabidopsis* interactor was *AT1G60240*, encoding a NAC family TF, and we demonstrated that in tomato a NAC TF represses *LIN5*. We identified the C3H-type zinc finger, C3H13, another type of zinc-finger TF, as a moderate activator. NPY1 is a BZIP-domain TF that is involved in phototropic responses and auxin-mediated plant development in *Arabidopsis* (Cheng et al., 2007). No functional studies have been done for these TFs in tomato.

Besides analysing the transcriptional regulation of the wild-type promoter, we utilized our mutants to study if the increased Brix phenotype could be correlated to disrupted CREs. We studied *pLIN5-cr-r1-04* and *r3-03a* in detail. At 1 day before anthesis (-1 DAA) *LIN5* expression levels were higher in lines with *r1-04* or *r3-r3a*. At anthesis, *LIN5* expression dropped below wild-type levels for the line with *r1-04*. In both mutants, expression levels were slightly higher from 3 DAA compared to wild-type, but not significantly. Near the mutation in *r1-04* a sugar repressive element was found. This is sugar repressive element contains a putative MYB-binding site (**Figure S3b,** motif 3). We tested MYB1R1 and R2R3MYB58 on the mutant promoters. R2R3MYB58 was expected

to act as a repressor since its expression increased between 0 and 5 DAA, the time when *LIN5* expression decreases. It also contains an EAR-domain motif as "LVLEL". Indeed, it gave a repressing signal. This effect was demonstrated both in the wild-type and tested mutant promoters, so the mutation did not disrupt this promoter-TF interaction. MYB1R1 did not influence expression of the wild-type promoter, but in *r1-04* there was a slight but significant 1.5-fold increase. Allele *r1-04* was activated by HAT4, in contrast to the wild-type allele. The *Arabidopsis* HAT4 homolog, AtHAT1, is involved in the Salicylic acid dependent defence response (Zou et al., 2016), and in the abscisic acid-dependent drought response (Zhang et al., 2018b). The last tested TF, WRKY24, resulted in more active mutant promoters. Our tested mutants were more responsive to the tested TF, indicating that even the small mutations can influence TF binding, gene expression, and final Brix.

CONCLUSION

In this study we used yest-1-hybrid assays, promoter reporter assays and CRISPR/Cas9 mutagenesis to study the *LIN5* promoter. This led to the identification of five repressors of *LIN5*, namely GATA9, MYB48, MYB58, MYB76, and a NAC TF. C3H13, HAT4, MYB1R1, NF-YA10, NPY1, and WRKY24 activated *LIN5* expression. In addition, we demonstrated that an unbiased promoter-mutagenesis approach could result in increased *LIN5* expression and elevated Brix. Elevated Brix values were found in combination with decreased and increased weight changes. The variability in phenotypes demonstrates that *LIN5* is and remains a promising target for quality breeding, however, more information is required on how and when *LIN5* exactly functions and how the gene is regulated.

MATERIALS AND METHODS

OPEN CHROMATIN

Dnase I Hypersensitive Sites (DHS) Sequencing data was retrieved for the 25 Kb upstream and 25 kb downstream of *LIN5* from fruitENCODE (http://www.epigenome.cuhk.edu.hk/encode.html, SI2.50 tomato genome assembly) of several stages of fruit development, 7 Days After Anthesis (DAA), 17 DAA, 47 DAA, and leaf profile (control) (Petreikov et al., 2006; Lu et al., 2017). The data was used to determine open chromatin regions in the promoter (5 kb upstream) and in the open reading frame of *LIN5*, via a simplified model of the peak calling algorithm "Hotspot" (Koohy et al., 2014). For every measured site, a short reading sequence of approx. 250 bp was calculated, which gave the score n. The score for the measured site was compared against the local background of 50 kb (Koohy et al., 2014). A Z-score was calculated and compared to a cut-off value for the probability of the Z score for that site (in this case $p \leq 0.5$). Sites that had a score above the cut-off value were considered open chromatin region (Koohy et al., 2014).

SEQUENCE CONSERVATION ANALYSIS

To perform conservation analysis, the coding sequence, together with the 4.5 kb upstream of *LIN5* (Solyc09g010080) was retrieved from the SolGenomics database version 2.5. In total, the upstream region of six homologs from different species were found by using the NCBI blast tool (Query: Solyc09g010080 genomic sequence, BLAST tool: Megablast). The retrieved sequences were used as input for conservation analysis with mVISTA (**Table S1,** Fasta format, alignment program: AVID (global pair-wise alignment for finished & draft sequences, cut-off ranking mVISTA value p=0.05, window size of 100 bp) (Mayor et al., 2000; Brudno et al., 2003; Frazer et al., 2004).

YEAST-ONE-HYBRID ASSAY (Y1H)

A 400 bp region of the *LIN5* promoter was amplified both from *S. lycopersicum* cv. Moneyberg genomic DNA with specific primers (**Table S4**). The fragment was recombined into a Gateway® compatible reporter plasmid pAbAi vector containing the reporter gene *Aur1-c*, which confers resistance to aureobasidin (http://www.clontech.com, Danisman et al. 2012). The resulting promoter-reporter construct was linearized by adding 30 µl vector to 5 µl buffer G Thermo Scientific, 2 µl *Bbs*I and 13 µl MQ and incubating them at 37 °C for 1 h while shaking at 450 rpm. The mixes were inactivated for 20 min at 65 °C. The linearized fragment was transformed into yeast strain PJ69-4 α. A culture was grown overnight in 10 ml SD-complete (6.7 g/L Yeast nitrogen base, 20 g/L Dextrose for 9x SD medium supplemented with the correct 1x dropout. Complete dropout solution: 300 mg/L Isoleucine, 1500 mg/L L-Valine, 200 mg/L L-Adenine hemisulfate salt, 200 mg/L L-Arginine HCL, 200 mg/L L-Histidine HCl Monohydrate, 1000 mg/L L-Leucine, 300 mg/L L-Lysine HCl, 200 mg/L L-Methionine, 500 mg/L L-Phenylalanine, 2000 mg/L L-Threonine, 200 mg/L L-Tryptophan, 300 mg/L L-Tyrosine, 200 mg/L L-Uracil). The culture was diluted to 50 ml and re-grown till an OD_{600} between 0.4 and 0.6 was centrifuged for 5 min at 5000 rpm. The pellet was resuspended in 25 ml 100 mM Lithium Acetate (LiAc) and centrifuged again for 5 minutes at 5000 rpm. A mix of 2.4 ml 50% (w/v) polyethylene glycol (PEG) 3350. 360 µl 1 M LiAc, 50 µl salmon sperm DNA (10 mg/ml, boiled), 20 µl purified linearized plasmid and 680 µl MQ. After vortexing, the mix was incubated for 20 min at 42 °C and centrifuged again (5 minutes, 5000 rpm). The pellet was then washed with 0.5 ml MQ and spun down. Finally, the pellet was resuspended in 0.3 ml MQ and 200 µl was plated on selective medium, SD agar minus Uracil. After three days of growth at 30 °C, transformant colonies were visible. Auto-activation tests were performed at a range of Aureobasidin A (AbA) concentrations (0, 50, 100, 150, 200, 500 ng/µl AbA) to determine background expression of the reporter bait constructs. Reporter constructs were selected for growth at 50 or 100 ng/µl AbA and absence of growth from 150 ng/µl AbA.

We used an expanded REGIA Transcription Factor (TF) prey collection containing more than 2000 *Arabidopsis* TFs in the PJ68-4A yeast strain, as previously reported (Castrillo et al., 2011). A screen with this library was performed twice at 200 ng/ml AbA, as described previously (Danisman et al., 2012). Putative hits were scored on a scale of 0 (no growth) to 3 (>75% growth). The resulting 89 unique positive hits with a score of at least 1 were re-screened at 175 and 225 ng/ml Aba. Tomato orthologs of the *Arabidopsis* hits were identified by best reciprocal BLAST hit for the amino acid sequences. Candidates were further selected for (co-)expression in tomato with data from the Tomato Expression Atlas ((http://tea.solgenomics.net/).

PROMOTER REPORTER-ASSAYS

The Golden Gate Molecular Cloning (MoClo) toolkit was used to assemble the constructs with the Golden Gate cloning strategy (Engler et al., 2008; Weber et al., 2011; Werner et al., 2012). The full-length open reading frames of selected TFs were amplified from cDNA with gene-specific primers (**Table S4).** Where necessary, a nested PCR was performed. Internal Golden Gate restriction sites were removed with overlap extension PCR (primers not listed). The products were ligated into pL0-CDS (pICH41308) and subsequently into pL1-F2 expression vectors (pICH47742) with a 2xCaMV 35S promoter (pICH51288) and 35s terminator (pICH41414) to form the effector pL2 constructs. The promoter was amplified with specific primers (**Table S4)** from wild-type cv. Moneyberg or mutant genomic DNA, totalling 1258-bp including the 5'UTR. An internal Golden Gate restriction site was removed by amplifying the promoter in two parts with the addition of seamless Golden Gate Overhangs. Correct amplicons were cloned into to pGEM®-T Easy Vectors (Promega) as described in the manual. After a sequence check, the parts were cut-ligated into pL0-pro+5'UTR (pICH41295). The pL0-promoter was then ligated with a Golden Gate compatible firefly luciferase coding vector (pICSL50006) and a 35S terminator (pICH41414) to form the reporter construct in a pL1-F1 expression vector (pICH47732). A positive control vector was constructed by combining a 2x35S promoter with the firefly luciferase coding sequence and 35S terminal. To normalize luciferase activity, a third pL1 vector was used: 2x35Sp::Renilla, which was made Golden Gate compatible from pGreenII:0800-LUC (Hellens et al., 2005).

Correct plasmids were transformed to *A. tumefaciens* C58c1. Transformed cells were grown from an overnight culture in LB with 10mM 2-(N-morpholine)-ethanesulfonic acid (MES) and 40 µM Acetosyringone to an OD_{600} between 0.4 and 1. The cells were concentrated in 10 mM MgCL2 with 200 µM Acetosyringone at and OD_{600} of 1 for the reporter and Renilla construct, and an OD_{600} of 2 for the effector. Different combinations were made with the following ratios: 2.5 Reporter: 62.5 Effector: 1 Renilla. Al mixes were equalled at 20 ml and left to stand for 2-3 hours. For each combination, transfections were performed on three different *Nicotiana benthamiana* plants (biological replicates) of approximately five weeks old (before flowering). The

infiltrated tobacco plants were grown for three days in a greenhouse. Samples were taken by closing a 2-ml tube on a transfected leaf-area while avoiding large veins. From each transfected leaf, three technical replicates were taken. Samples were snap-frozen in liquid Nitrogen. Sample preparation and measurements were done with the Dual-Luciferase Reporter Assay kit (Promega, cat. no. E1910), as described in (Sherf et al., 1996). 200 µl Passive Lysis Buffer mix was used to lysate a cryo-ground leaf disk for 15 min on ice, followed by 3 min centrifuge. Measurements were done in a GlowMax® Navigator Microplate Luminometer with 50 µl supernatant and 25 µl of both the LARII and Stop&Glow reagents. The Firefly/Renilla ratio of each sample was calculated relative to the sample with just the promoter and Renilla construct. Statistical analysis was performed with R. For all data, normal distribution was confirmed, and ANOVAs with post-hoc Tukey pairwise analysis were used to test for significant differences.

CRISPR/CAS9 DESIGN AND ASSEMBLY

The Rgenome Cas-designer tool (http://www.rgenome.net/cas-designer/, spCas9, target: *Solanum lycopersicum*, allowed 2 nucleotide bulge as off target) was used to design effective single guide RNAs (sgRNAs) (Bae et al., 2014; Concordet and Haeussler, 2018). sgRNAs were filtered out if they had an off-target with one or more nucleotides mismatch. Twenty-one targets for sgRNAs were chosen targeting the 4000 bp upstream of *LIN5* (**Table S4**). The selected targets were divided in six different regions (**Figure 1, Figure S1**): region 1 to 5 contained three adjacent guides each and region 6 contained six targets. The different groups of sgRNAs were all assembled in separate vectors.

The MoCLo toolkit and Golden Gate cloning were used to assemble the vectors (Engler et al., 2009; Weber et al., 2011). In short, each sgRNA was fused to an *Arabidopsis* U6 promoter as AtpU6:sgRNA:TTTT and cut-ligated to a level 1 vector. Level 1 constructs pNOS:NPTII:tOCS (pICH47732-pL1), p35S:hCas9:tNOS (pICH47742-pL2), p2x35S:turboGFP:t35S (pICH47751-pL3), pU6:sgRNA1 (pICH47761-pL1-F4), pU6:sgRNA2 (pICH47772-pL1-F5), pU6:sgRNA3 (pICH47781-pL1-F6), and pLE6E (pICH41822) were cut-ligated into the level 2 vector pICSL4723 (pL2) (Werner et al., 2012). For the construct with six sgRNAs, a two-step Golden Gate cut-ligation was performed. The NPTII, Cas9, GFP and the two first sgRNAs were cut-ligated with pELB5 (pICH49299) into the level 2 vector to create an intermediate level 2 (pL2i-1). This plasmid was then used to construct the final level 2 (pL2-2) by ligating it with pU6:sgRNA3 (pICH47781-pL1-F6), pU6:sgRNA4 (pL1-F7), pU6:sgRNA5 (pICH47732-pL1-F1), pU6:sgRNA6 (pICH47742-pL1-F2), and pELE2 (pICH41780). All constructed vectors were checked by sequencing. All pL2 and the pL2-2 constructs were transformed to *Agrobacterium tumefaciens* C58C1 and grown under rifampicin, gentamycin, and kanamycin selection. In *Agrobacterium*, the presence of the correct construct was validated with restriction analysis (BamH1, SAL1).

TRANSFORMATION

Tomato transformation was done with *Agrobacterium tumefaciens* strain C58C1 containing the appropriate vector as previously described, but with media B supplemented with 1 mg/L 2,4D and with 200 cotyledon explants as starting material (Van Roekel et al., 1993). *Solanum lycopersicum* L. cv. 'Moneyberg' was used in this study. Tissue culture was done in a growth chamber with 16 h light and 8 h dark cycle at 25 °C. Once shoots were formed, GFP-positive shoots were selected and rooted on Rooting Inducing Medium (Van Roekel et al. 1993). Rooted shoots were placed on rockwool and moved to a growth chamber (16 h light and 8 h dark at 25 °C). A ploidy test was done on transformed plant leaf samples at Iribov Analytical Services B.V.. Diploid shoots were genotyped for presence of the transgene and for mutations in the target region.

GENOTYPING

Genomic DNA from young leaves was isolated using the CTAB as described (Porebski et al., 1997). Alternatively, a PCR was done directly on sampled leaf tissue by using the Phire Plant Direct PCR kit (Thermo Scientific, Catalog number F130WH). Detection of transgenes in each generation was done by a PCR on Cas9 and/or NPTII. The target region of transformed plants was amplified, sequenced, and aligned in Benchling (https://benchling.com, for primers used see **Table S4**). Heterozygous and bi-allelic **Transformed (T_0)** plants for the region of interest were selfed. T_1 plants were segregated for the presence of a homozygous or bi-allelic mutation in the region of interest, while lacking the T-DNA insertion. In the T_2 generation only homozygous plants were grown.

GROWTH CONDITIONS

Shoots grown from callus were transferred to rockwool once they had roots. Alternatively, T_1 and T_2 seeds were germinated on filter paper and transferred to cubes of rockwool. They were grown in a growth chamber (16 h light and 8 h dark at 25 °C). Five to Eight weeks later seedlings were transplanted into a greenhouse (Unifarm, Wageningen 51.57° N, 5.31° E, The Netherlands) on rockwool slabs at a density of 2.5 plants/m^2. Nutrients were provided by fertigation (EC 4.5, pH 5.6). Climatic conditions in the greenhouse were at ambient temperature (> 20 °C) under a 16h light/8h dark cycle (0.6–28.4 MJ m–2 day–1 natural light supplemented with artificial light using high pressure sodium lamps (SON-T Agro 600 Watt, Philips, Eindhoven, The Netherlands). Side shoots were removed once a week. Flowers were pollinated by vibrating each flower/truss three times a week with an electric toothbrush.

PHENOTYPING FRUITS

T_0 plants were only used for genotyping and seed collection. In the T_1 generation, one or two plants per genotype were placed randomly on a row in the greenhouse together. As it was not possible to put all the T_1 plants in the greenhouse at the same time, phenotyping was done in batches. For every batch of T_1 mutant plants, a wild-type control was taken along. Six flowers per truss were vibrated with an electric toothbrush-holder at anthesis. Excess flowers were removed. The first truss was used for seed collection and the subsequent three trusses were used for phenotyping.

Individual fruits were harvested at breaker+7 ($\pm$1 day) and fresh weight (g) and Brix measurements were done *in duplo* per fruit with an Atago PR-32α digital refractometer.

In the T_2 generation, five plants per genotype were randomly placed on a row in the greenhouse. Six flowers per truss were vibrated at anthesis. Excess flowers were removed. The first truss was used for seed collection and the three subsequent trusses were used for phenotyping. The third and fifth truss fruits were harvested at breaker+7 ($\pm$1 day) for phenotyping. Of these fruits, width, weight, and Brix were measured. Of the second and fourth and sixth truss, samples were taken for gene expression analysis at different time points (-1 to 5 Days After Anthesis, DAA) and immediately snap-frozen in liquid Nitrogen.

GENE EXPRESSION ANALYSIS

Gene expression analysis was done in the T_2 generation with five plants per genotype. At the different time points (-1 to 5 DAA) samples were taken from each plant as a pool of two ($\geq$1 DAA) or three (<1 DAA) fruits with a leaf sample as control. RNA was isolated from cryo-ground whole fruit samples by using the MaqMax™-96 total RNA isolation kit with Plant RNA isolation aid (Thermofisher) with a KingFisher 96 Magnetic Particle Processor. cDNA was synthesized with the iScript cDNA synthesis kit (Bio-Rad). qRT-PCR was performed with iQ SYBR Green Supermix (Bio-Rad) on a Bio-Rad CFX6 qRT-PCR instrument. The following condition was applied for PCR amplification: 3 min 95˚C, 40*[15 s 95˚C, 60 s 60˚C], followed by a melt-curve. *CAC* was used as a refence (**Table S4**). Relative expression changes were calculated according to the $2^{-\Delta CT}$ method as described (Livak and Schmittgen, 2001). Statistical analysis was performed with R, package version 3.5. For all data, normal distribution was confirmed, and ANOVAs were used to test for significant differences between mutants and wild-type.

ACKNOWLEDGMENTS

All MoClo kit plasmids were a gift from Sylvestre Marillonnet. pICSL50006 was a gift from Nicola Patron. The pGreenII:0800-LUC was given to us by Kieran Elborough. The luciferase vectors were made Golden Gate compatible by Lena Maas. We thank Renze Heidstra for his assistance with the luciferase measurements. We thank Geurt Versteeg, Teus van den Brink and Sean Geurts for their expertise and care in the tomato greenhouse cultivation.

SUPPLEMENTARY FIGURES

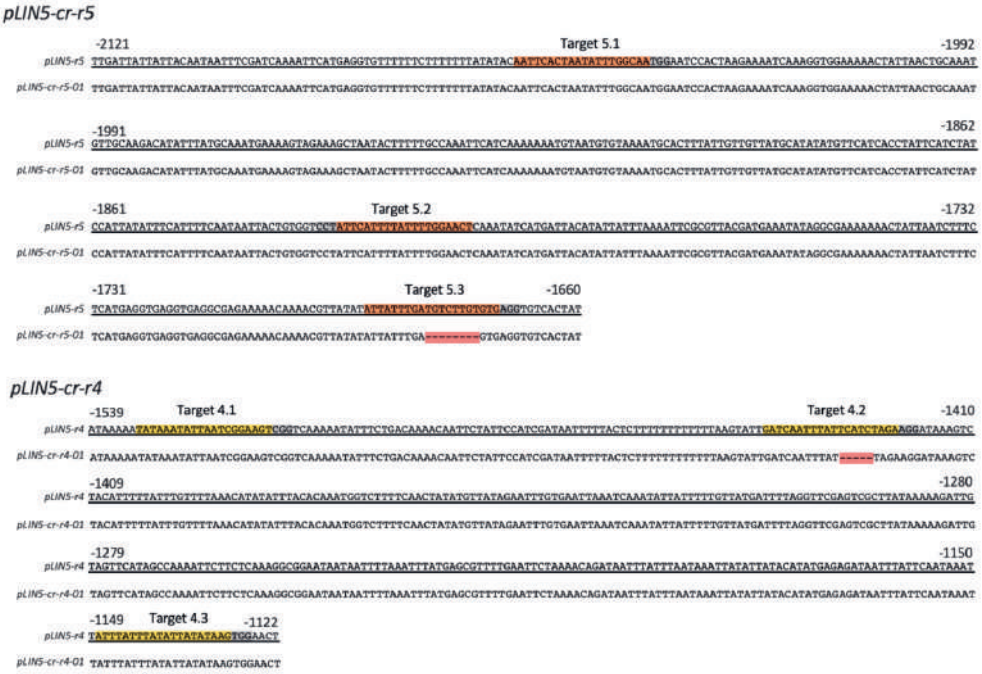


Figure S1: Alignments of the wild-type and mutant promoters. Location numbers are relative to the start codon of the *LIN5* ORF. sgRNA target regions are highlighted with a grey PAM site. Mutations are highlighted in red. Het; heterozygous mutant, only the mutated allele is given. Bi: bi-allelic mutant. The a and b alleles are given.

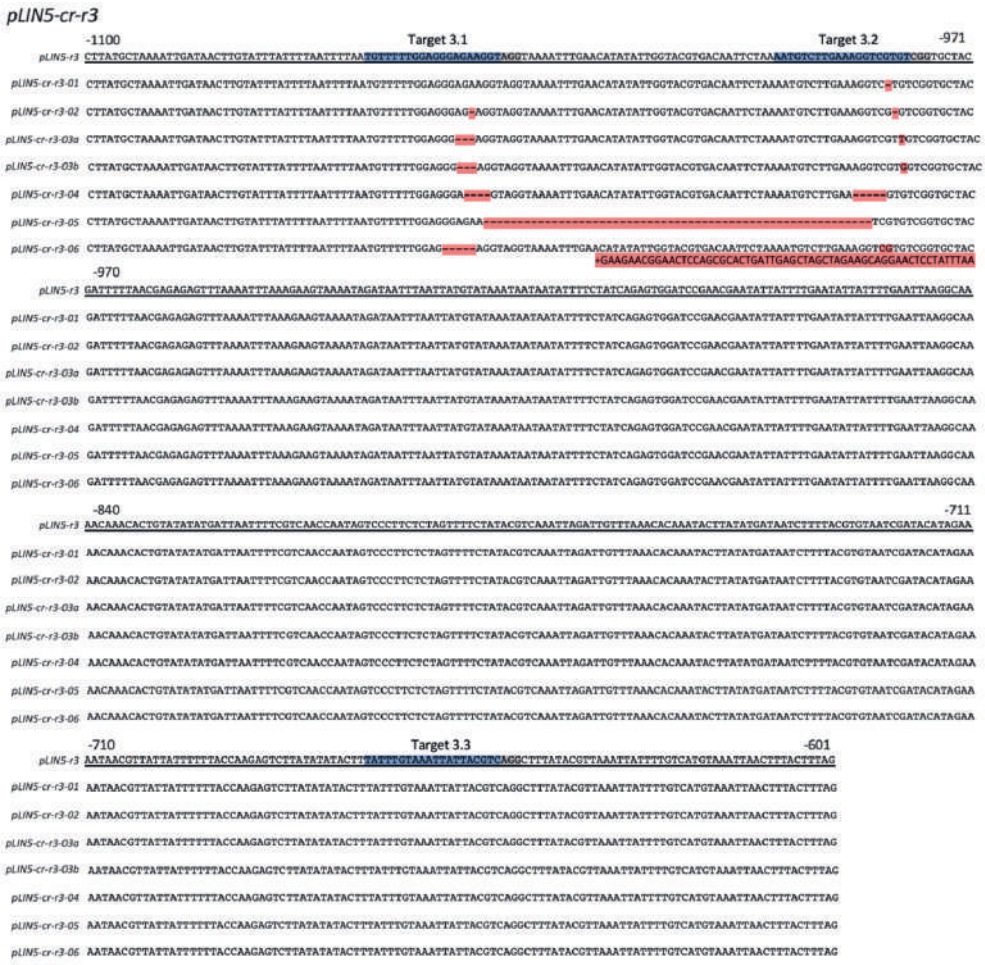


Figure S1 continued

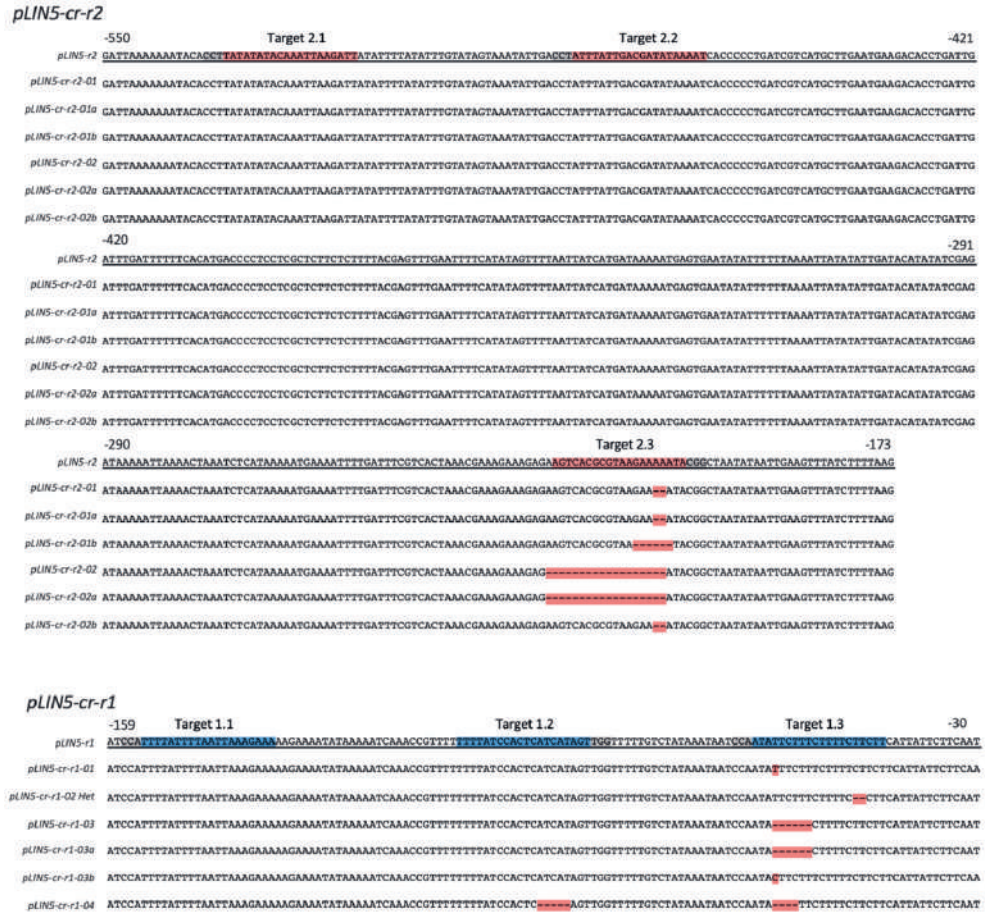


Figure S1 continued

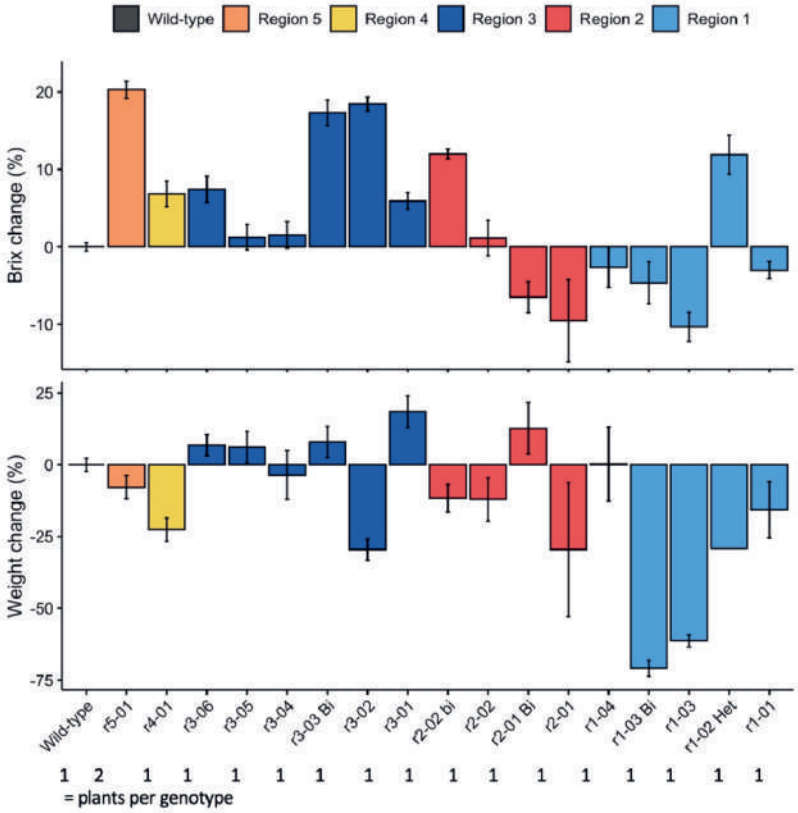


Figure S2: Phenotypes T_1. Change in Brix (%) and change in weight (%). compared to a wild-type plant grown at the same time as the mutant. The mutants are ordered by the different targets they have mutated, divided in the different regions (**Figure 1**). Error bars represent the SE. For each mutant, three trusses with 6 fruits of one or two plants were measured.

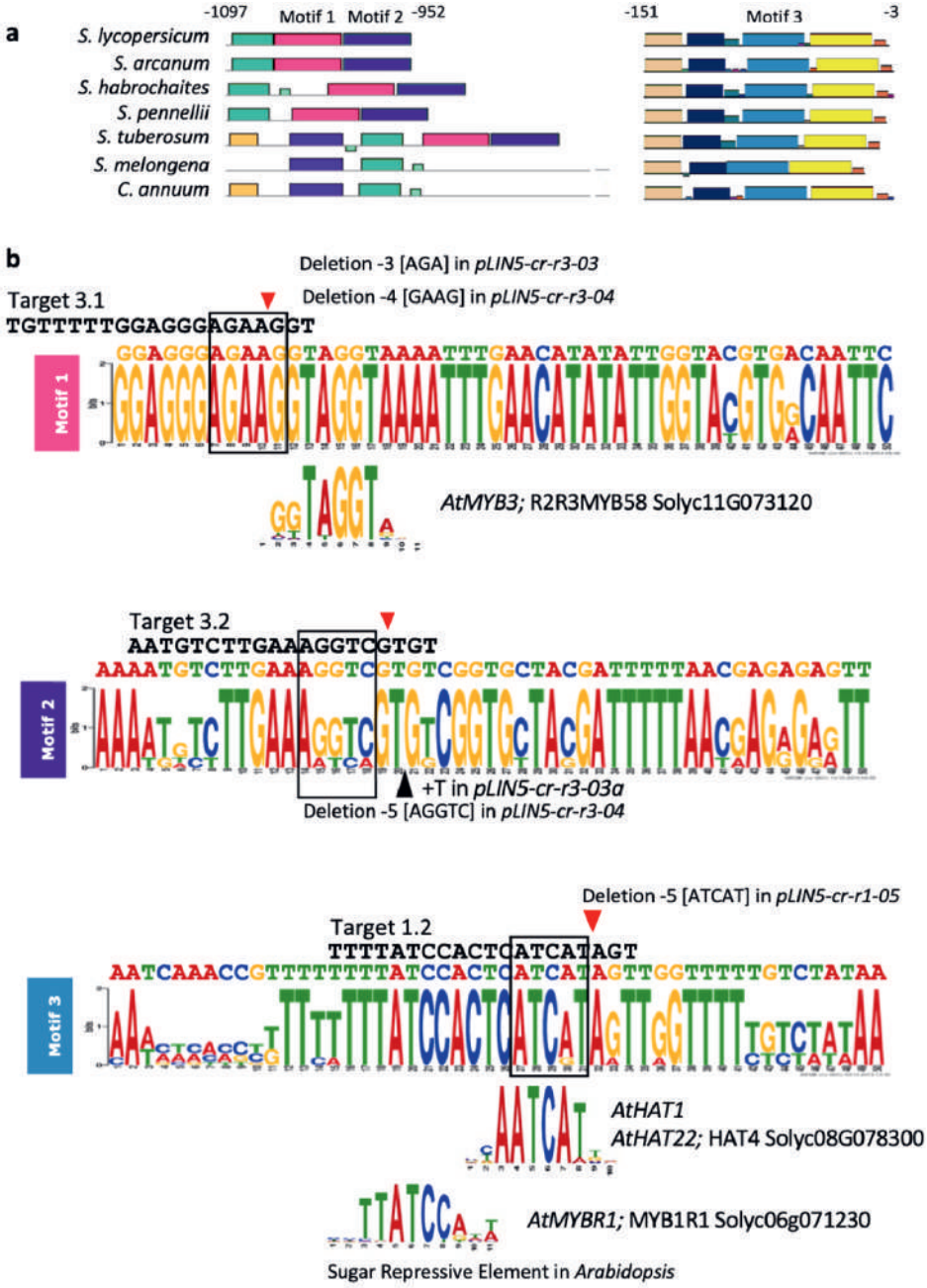


Figure S3: Motif prediction and TF selection. **(a)** MEME motif prediction of the mutation sites in region 3 and region 1 in the *LIN5* promoter leading to a higher Brix phenotype. The same six LIN5 orthologs were used as in the conservation analysis (**Figure 1**). Bp position numbering is relative to the Translation Start Site of the *S. lycopersicum LIN5* ORF. The blue and red motifs are designated motif 1 and 2 respectively (http://meme-suite.org/), **(b)** Motifs and alignment with the *S. lycopersicum* genome. Target sites are indicated, and the mutation sites are indicated by a black box. Underneath the motif are the resulting TF binding motifs from the TOMTOM motif comparison with the JASPAR core plants database (2018)

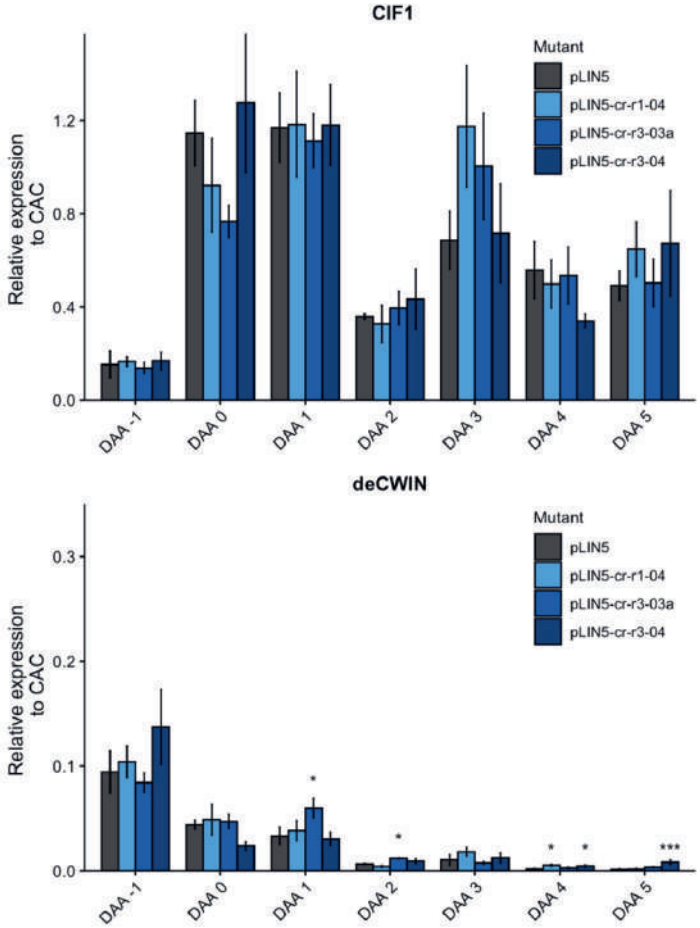


Figure S4: The second, fourth and sixth truss was used for gene expression analysis. Per timepoint, a pool of three fruits was collected per plant. Statistically significant differences are represented by asterisks (* $p<0.05$, ** $p<0.01$, *** $p<0.001$).

SUPPLEMENTARY TABLES

Supplementary table 1: *LIN5* Orthologs used in conservation analysis

Species	Locus number
Solanum lycopersicum	Solyc09g010080
Solanum arcanum	contig_210, contig_211, and contig_212
Solanum habrochaites	contig_8529 and contig_8530
Solanum pennellii	LOC107029797
Solanum tuberosum	LOC102595942
Solanum melongena	Scaffold Sme2.5_01674.1
Capsicum annuum	ch9[238637285..238649862]

Abbreviations: *ADP-glucose pyrophosphorylase (AGPL1), Glucose-1-phosphate adenylyltransferase (AGPS1)*

Supplementary table 2: Y1H combination screen results with *A. thaliana* TFs and subsequent *S. lycopersicum* ortholog selection

Y1H with *A.thaliana* TF library						TF selection for a Y1H with *S. lycopersicum* TFs			
Transcription Factor *A. thaliana*						Potential candidates *S. lycopersicum*			
Gene ID	alias	TF family	175	225	Screen	Gene ID	alias	TF family	Av. expression
AT1G47760	AGL102	MADS			1	Solyc04g056740	AGL61	MADS	0.0
AT5G02320	ATMYB3R5	MYB			1	Solyc09g010820	MYB3R-3	Myb	10.3
AT5G03790	ATHB51; LMI1	HB			1, 2	Solyc09g008810	ATHB-22	HD-ZIP	0.2
AT2G36930	T1J8.11	C2H2			1, 2	**Solyc09g008230**	**ZF593**	**ZF**	**179.3**
AT3G58630	F14P22_220	TRIHELIX			1	Solyc12g010890	ASIL1-like	TRIHELIX	8.9
AT1G78700	BEH4; BES1; BZR1	BES1			1	Solyc07g062260	BES1; BZR1	BES1	5.1
AT3G28920	AtHB34; ZHD9	ZF-HD			1, 2	Solyc02g067320	ZF_HD1	ZF-HD	3.6
AT2G02540	ZHD3	ZF-HD			1, 2	Solyc01g102980	ZHD2	ZF-HD	0.2
AT3G50890	AtHB28; ZHD7	ZF-HD			1, 2	Solyc02g085160	ZHD6	ZF-HD	2.0
AT1G55650	HMGB11	ARID			1, 2	Solyc09g091960	HMGP15	HMGP	1.6
AT5G66700	ATHB53; HB-8	HB			1, 2	Solyc02g062960	HOX14	HD-ZIP	4.7
AT1G14510	AL7	ALFIN-like			1, 2	**Solyc01g102750, Solyc01g102760**	**AL6, AL8**	**ALFIN-like**	**56.6, 99.7**
AT2G30420	ETC2	MYB-related			2	Solyc01g095640	TCL1	MYB-related	0.3
AT1G49130	BBX17; COL8	C2C2-CO like			2	**Solyc02g093590**	**CONSTANS7**	**ZF**	**15.0**
Water									

Bold *S. lycopersicum* TFs were used in a subsequent Y1H. Average expression in fruit (RPM) was obtained from the Tomato Expression Atlas (http://tea.solgenomics.net/).
* ng/ml Aureobasidin

Supplementary table 3: Motif search

Action	TF	*Arabodopsis* orhtolog [1]	Motif [2]	Present in *pLIN5* PRA region [3]
Repressor	GATA9 (Solyc11g069510)	AT1G08010		p = 5.8e-5 AAAACTAAATCTCAT
Repressor	MYB48 (Solyc06g005310)	AT5G59780*		p = 5.2e-5 AGGTAGGTAA
Repressor	MYB54 (Solyc10g083900)	AT3G53200		p = 3.5e-5 GAGAAGGTAGGTAAA p = 7.6e-5 ATTTAATTATGTATA
Repressor	NAC (Solyc08g008660)	AT2G46770		-
Activator	C3H13 (Solyc01g100990)	AT5G26749	-	
Activator	NAF-YA10 (Solyc01g006930)	AT5G06510	-	
Activator Mutants	WRKY24 (Solyc09g066010)	AT2G30590		

[1]Orthologs selected based on the Plant TF database http://planttfdb.gao-lab.org/ and Dicots Plaza4.5.
[2]Motif from the Plant TF database (Jin et al., 2015, 2017; Van Bel et al., 2018; Tian et al., 2020).
[3]Searched for occurrences motif in the promoter with the Motif Alignment and Search Tool (MAST) (Bailey and Gribskov, 1998).
*Non-best hit, but with a Motif in the database

Supplementary table 4: Used primers

Description	Name	Sequence (5'-3')
Y1H primers		
pLIN5	Y1H Fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTTATTTCACATG ACCCCTCCT
	Y1H Rv	GGGGACCACTTTGTACAAGAAAGCTGGGTAGTGATTTTG GAGAAAGAT
promoter reporter assay (PRA) primers (without Golden Gate overhang)		
pLIN5 (Solyc09g010080)	PRA Fw	CCAAAATTCTTCTCAAAGGCGG
	PRA Rv	GTGATTTTGGAGAAAGATATTGAAG
pLIN5 BpiI removal	PRA Fw	GTCGAAGACCGTGCAGACACCTGATTGATTTG
	PRA Rv	GTCGAAGACGCGTCATTCAAGCATGACGATCAGG
AL6 (Solyc01g102750)	PRA Fw	ATGGAAAATTCGGTACCCAGGAC
	PRA Rv	TTAGACTTTAGCTCTCTTGCTACTG
AL8 (Solyc01g102760)	PRA Fw	ATGGAAAATACGGTACCTAGGAC
	PRA Rv	TTAAACTCTAGCTCTCTTGCTACTG
CO7 (Solyc02g093590)	PRA Fw	ATGCATATGTCCGATGCCGGATG
	PRA Rv	TCAATCCCTCTTAACAAATCTCCCC
ZFP593 (Solyc09g008230)	PRA Fw	ATGGGAGGAAAGTGTCCTCA
	PRA Rv	TAACTCATTGACATTAGCTTTGG
AREB1 (Solyc04g078840)	PRA Fw	ATGGGGAGTAATTATC
	PRA RV	TTACCATGGACCAG
B3D (Solyc08g029090)	PRA Fw	ATGCCTCCTCAGTTCTTCAAG
	PRA RV	TCAGGAAACGGGAGTTGTG
BBX19 (Solyc01g110370)	PRA Fw	ATGAGAACCCTTTGTGATG
	PRA RV	TTAAAGCTTTGACATTGC
bHLH136 (Solyc06g062460)	PRA Fw	ATGAATTTTCAGGAATTTGGAG
	PRA RV	TGTAGTAATCACATCAAAATATGATAC
C3H13 (Solyc01g100990)	PRA Fw	ATGCCGTTGGGTAAATACTA
	PRA RV	CTATCCCCACTCAACAA
COL13 (Solyc02g079430)	PRA Fw	ATGTGGTTTGTAGTGTTGTAGG
	PRA RV	CCGATTATCTCCATCTCCAT
GATA9 (Solyc11g069510)	PRA Fw	ATGGATGAAATTCCTACTGGTC
	PRA RV	TCAGTATACATCAAAC
GLK2 (Solyc10g008160)	PRA Fw	ATGGCAACTCATGGCCTC
	PRA RV	TCAAGTTGGGGGTATTTTGGT
JA1/LEJA1 (Solyc05g007180)	PRA Fw	ATGGCTTTCTTTCCAACAAA
	PRA RV	TAATTAAAATGTTGTTGCTCTAGC
MYB48 (Solyc06g005310)	PRA Fw	ATGGCACAAGAAGAAATGAGA
	PRA RV	TTAGCCAGCAAAGAATGTGT
MYB54/MYB27 (Solyc10g083900)	PRA Fw	ATGCAAGAAGAGGAACTACG
	PRA RV	TATGGAGGATAACAAGAACATAGG
MYB76/MYB6 (Solyc05g008250)	PRA Fw	ATGAGAAAGCCTTGTTGTGA
	PRA RV	TATGGAATTAAATTGAGATCAAGCA
NAC (Solyc08g008660)	PRA Fw	ATGAATCTCTCTGTAAATGGTC
	PRA RV	ACTACATCCTATACAGTAGGAG
NFYA10 (Solyc01g006930)	PRA Fw	ATGAATACTACTATATTTTCCAAAGG
	PRA RV	TCATACTTTGAGGTTGCAAC
NPY1 (Solyc08g006870)	PRA Fw	ATGAAGTTCATGAAGTTGGG
	PRA RV	AATTTCAAATCACAAGATTCAAAAATC

Supplementary table 4 continued

Description	Name	Sequence (5'-3')
NSP1 (Solyc12g049320)	PRA Fw	ATGACCATTGATGAAACTACTTC
	PRA RV	TCAACTCCTCGAGCAATGT
TPRL (Solyc06g068130)	PRA Fw	ATGTTATTACGAAGCTCATCATC
	PRA RV	TTACTTCTTCAATTCCTCCTCAG
TRFL (Solyc10g080300)	PRA Fw	ATGCATGCGGAGGTCA
	PRA RV	TTATGTTTCTTTCTCGTCTGC
WRKY24 (Solyc09g066010)	PRA Fw	ATGGAGGAGATTGAGGAAGCTAACAG
	PRA RV	TCATGCATTTGCCGATTGG
WRKY77 (Solyc10g007970)	PRA Fw	ATGGAAGGTGTGCTACAAAAAG
	PRA RV	TCATATACCATCAACGATACGATT
WRKY81 (Solyc09g015770)	PRA Fw	ATGGATAACTCATCGTCTG
	PRA RV	CTACACTTGATCAAAGTTCC
HAT4 (Solyc08G078300)	PRA Fw	ATGATGGTTGAAAAAGAAGATTTGGGG
	PRA RV	TCATGATCTCGGATGTAGTG
MYB1R1 (Solyc06G071230)	PRA Fw	ATGTCGAGCGTTTGCAGTGATAAG
	PRA RV	TCATGCCACACGGATGATGCT
R2R3MYB58 (Solyc11G073120)	PRA Fw	ATGGTGCAAGAAGAAATAATG
	PRA RV	TATCGAAAGTTATTCCAATC
sgRNA primers		
pLIN5-1.1	sgRNA Fw	TGTGGTCTCAATTGTTTCTTTAATTAAAATAAAA
		GTTTTAGAGCTAGAAATAGCAAG
pLIN5-1.2	sgRNA Fw	TGTGGTCTCAATTGTTTTATCCACTCATCATAGT
		GTTTTAGAGCTAGAAATAGCAAG
pLIN5-1.3	sgRNA Fw	TGTGGTCTCAATTGAAGAAGAAAAGAAAGAATAT
		GTTTTAGAGCTAGAAATAGCAAG
pLIN5-2.1	sgRNA Fw	TGTGGTCTCAATTGAATCTTAATTTGTATATATA
		GTTTTAGAGCTAGAAATAGCAAG
pLIN5-2.2	sgRNA Fw	TGTGGTCTCAATTGATTTTATATCGTCAATAAAT
		GTTTTAGAGCTAGAAATAGCAAG
pLIN5-2.3	sgRNA Fw	TGTGGTCTCAATTGAGTCACGCGTAAGAAAAATA
		GTTTTAGAGCTAGAAATAGCAAG
pLIN5-3.1	sgRNA Fw	TGTGGTCTCAATTGTGTTTTTGGAGGGAGAAGGT
		GTTTTAGAGCTAGAAATAGCAAG
pLIN5-3.2	sgRNA Fw	TGTGGTCTCAATTGAATGTCTTGAAAGGTCGTGT
		GTTTTAGAGCTAGAAATAGCAAG
pLIN5-3.3	sgRNA Fw	TGTGGTCTCAATTGTATTTGTAAATTATTACGTC
		GTTTTAGAGCTAGAAATAGCAAG
pLIN5-4.1	sgRNA Fw	TGTGGTCTCAATTGTATAAATATTAATCGGAAGT
		GTTTTAGAGCTAGAAATAGCAAG
pLIN5-4.2	sgRNA Fw	TGTGGTCTCAATTGGATCAATTTATTCATCTAGA
		GTTTTAGAGCTAGAAATAGCAAG
pLIN5-4.3	sgRNA Fw	TGTGGTCTCAATTGATTTATTTATATTATATAAG
		GTTTTAGAGCTAGAAATAGCAAG
pLIN5-5.1	sgRNA Fw	TGTGGTCTCAATTGAATTCACTAATATTTGGCAA
		GTTTTAGAGCTAGAAATAGCAAG
pLIN5-5.2	sgRNA Fw	TGTGGTCTCAATTGAGTTCCAAAATAAAATGAAT
		GTTTTAGAGCTAGAAATAGCAAG
pLIN5-5.3	sgRNA Fw	TGTGGTCTCAATTGATTATTTGATGTCTTGTGTG
		GTTTTAGAGCTAGAAATAGCAAG

Supplementary table 4 continued

Description	Name	Sequence (5'-3')
pLIN5-6.1	sgRNA Fw	TGTGGTCTCAATTGCTTTCAAACGGATACATTAA
		GTTTTAGAGCTAGAAATAGCAAG
pLIN5-6.2	sgRNA Fw	TGTGGTCTCAATTGAAATTTGTAAAACTTATAAG
		GTTTTAGAGCTAGAAATAGCAAG
pLIN5-6.3	sgRNA Fw	TGTGGTCTCAATTGTATGTTTTAAGTGTGTTTAT
		GTTTTAGAGCTAGAAATAGCAAG
pLIN5-6.4	sgRNA Fw	TGTGGTCTCAATTGATTTACAAAAATCTTTCAAA
		GTTTTAGAGCTAGAAATAGCAAG
pLIN5-6.5	sgRNA Fw	TGTGGTCTCAATTGTTTGTGATGTCGGTGATGGG
		GTTTTAGAGCTAGAAATAGCAAG
pLIN5-6.6	sgRNA Fw	TGTGGTCTCAATTGTTCACGTTAGCAGCGAGTGT
		GTTTTAGAGCTAGAAATAGCAAG
Universal sgRNA	sgRNA Rv	TGTGGTCTCAAGCGTAATGCCAACTTTGTAC
Genotyping primers		
Cas9	Cas9 Fw	CTGGCCAATGGAGAGATTCG
	Cas9 Rv	GCTCCCTGATGGGCTTATCC
NPT2	NPT2 Fw	AGACAATCGGCTGCTCTGAT
	NPT2 Rv	AGCCAACGCTATGTCCTGAT
Region 1	pLIN5 Fw	GGAGGTTGAAAATGAAAACGAGTTC
	pLIN5 Rv	CCTCCTCGCTCTTCTCTTTTACGA
Region 2	pLIN5 Fw	GAAGCAAATGCCCTATTAATGTTTG
	pLIN5 Rv	TCGTCAACCAATAGTCCCTTCTCT
Region 3 outer	pLIN5 Fw	CCAAAATTCTTCTCAAAGGCGGAAT
	pLIN5 Rv	GAGCGAGGAGGGGTCATGTG
Region 3 inner	pLIN5 Fw	TCTTCATTCAAGCATGACGATCAG
	pLIN5 Rv	AAATTCTTCTCAAAGGCGGAATAAT
Region 4	pLIN5 Fw	TGAGGTGAGGCGAGAAAAACA
	pLIN5 Rv	AATCGTAGCACCGACACGAC
Region 5	pLIN5 Fw	CAGAAATATTTTTGACCGACTTCCG
	pLIN5 Rv	CAGCGAGTGTAGGACATTCTCTCTT
Region 6	pLIN5 Fw	CCAATTAATTTTTGTGGCCACTACA
	pLIN5 Rv	CGCCTATATTTCATCGTAACGCG
Region 6 inner	pLIN5 Fw	GGATTCCATTGCCAAATATTAGTGA
	pLIN5 Rv	TGCTGTTGATAAGAAAACGAACCTT
qPCR primers		
CAC	qPCR Fw	CCTCCGTTGTGATGTAACTGG
	qPCR Rv	ATTGGTGGAAAGTAACATCATCG
LIN5	qPCR Fw	TGGGGTTGGTCAAATGAATCCG
	qPCR Rv	GAATACCTTGAATTCCAGCCCATCC
CIF1	qPCR Fw	GTTGGTAGAGCCATTGTAAGAAATTT
	qPCR Rv	TGATCATAATGTGACGAATCGAAT
deCWIN	qPCR Fw	AGTATGAGAGGATGGGCTGG
	qPCR Rv	TCTGCCTGTGTAGCATTGAC

Fw: Forward, Rv: Reverse; the underlined sequence represents the spacers in the sgRNA primers

REFERENCES

Alberto, F., Bignon, C., Sulzenbacher, G., Henrissat, B., and Czjzek, M. (2004). The Three-dimensional Structure of Invertase (β-Fructosidase) from *Thermotoga maritima* Reveals a Bimodular Arrangement and an Evolutionary Relationship between Retaining and Inverting Glycosidases. J. Biol. Chem. **279**: 18903–18910.

Bae, S., Park, J., and Kim, J.-S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics **30**: 1473–1475.

Bailey, T. and and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. AAAI Press: 28–36.

Bailey, T.L. and Gribskov, M. (1998). Combining evidence using p-values: Application to sequence homology searches. Bioinformatics **14**: 48–54.

Baldwin, E.A., Scott, J.W., Shewmaker, C.K., and Schuch, W. (2000). Flavour trivia and tomato aroma: Biochemistry and possible mechanisms for control of important aroma components. HortScience **35**: 1013–1022.

Baxter, C.J., Carrari, F., Bauke, A., Overy, S., Hill, S.A., Quick, P.W., Fernie, A.R., and Sweetlove, L.J. (2005a). Fruit carbohydrate metabolism in an introgression line of tomato with increased fruit soluble solids. Plant Cell Physiol. **46**: 425–437.

Baxter, C.J., Sabar, M., Quick, W.P., and Sweetlove, L.J. (2005b). Comparison of changes in fruit gene expression in tomato introgression lines provides evidence of genome-wide transcriptional changes and reveals links to mapped QTLs and described traits. J. Exp. Bot. **56**: 1591–1604.

Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van De Peer, Y., Coppens, F., and Vandepoele, K. (2018). PLAZA 4.0: An integrative resource for functional, evolutionary and comparative plant genomics. Nucleic Acids Res. **46**: D1190–D1196.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-Resolution Mapping and Characterization of Open Chromatin across the Genome. Cell **132**: 311–322.

Brudno, M., Malde, S., Poliakov, A., Do, C.B., Couronne, O., Dubchak, I., and Batzoglou, S. (2003). Glocal alignment: Finding rearrangements during alignment. In Bioinformatics (Oxford Academic), pp. 54–62.

Castrillo, G., Turck, F., Leveugle, M., Lecharny, A., Carbonero, P., Coupland, G., Paz-Ares, J., and Oñate-Sánchez, L. (2011). Speeding cis-trans regulation discovery by phylogenomic analyses coupled with screenings of an arrayed library of *Arabidopsis* transcription factors. PLoS One **6**: e21524.

Chen, C., Zhang, K., Khurshid, M., Li, J., He, M., Georgiev, M.I., Zhang, X., and Zhou, M. (2019). MYB Transcription Repressors Regulate Plant Secondary Metabolism. CRC. Crit. Rev. Plant Sci. **38**: 159–170.

Chen, J., Hu, Q., Zhang, Y., Lu, C., and Kuang, H. (2014). P-MITE: A database for plant miniature inverted-repeat transposable elements. Nucleic Acids Res. **42**: D1176–D1181.

Cheng, Y., Qin, G., Dai, X., and Zhao, Y. (2007). NPY1, a BTB-NPH3-like protein, plays a critical role in auxin-regulated organogenesis in *Arabidopsis*. Proc. Natl. Acad. Sci. U. S. A. **104**: 18825–18829.

Chi, Y., Yang, Y., Zhou, Y., Zhou, J., Fan, B., Yu, J.Q., and Chen, Z. (2013). Protein-protein interactions in the regulation of WRKY transcription factors. Mol. Plant **6**: 287–300.

Concordet, J.P. and Haeussler, M. (2018). CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res. **46**: W242–W245.

Danisman, S., van der Wal, F., Dhondt, S., Waites, R., de Folter, S., Bimbo, A., van Dijk, A.J., Muino, J.M., Cutri, L., Dornelas, M.C., Angenent, G.C., and Immink, R.G.H. (2012). *Arabidopsis* class i and class II TCP transcription factors regulate jasmonic acid metabolism and leaf development antagonistically. Plant Physiol. **159**: 1511–1523.

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes. PLoS One **4**: e5553.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS One **3**: 1–7.

Eshed, Y. and Zamir, D. (1996). Less-than-additive epistatic interactions of quantitative trait loci in tomato. Genetics **143**: 1807–1817.

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I. (2004). VISTA: Computational tools for comparative genomics. Nucleic Acids Res. **32**: W273–W279.

Fridman, E., Carrari, F., Liu, Y.-S., Fernie, A.R., and Zamir, D. (2004). Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science **305**: 1786–1789.

Fridman, E., Liu, Y.S., Carmel-Goren, L., Gur, A., Shoresh, M., Pleban, T., Eshed, Y., and Zamir, D. (2002). Two tightly linked QTLs modify tomato sugar content via different physiological pathways. Mol. Genet. Genomics **266**: 821–826.

Fridman, E., Pleban, T., and Zamir, D. (2000). A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc. Natl. Acad. Sci. U. S. A.

97: 4718–4723.

Fridman, E. and Zamir, D. (2003). Functional divergence of a syntenic invertase gene family in tomato, potato, and *Arabidopsis*. Plant Physiol. **131**: 603–9.

Gallego-Bartolomé, J. (2020). DNA methylation in plants: mechanisms and tools for targeted manipulation. New Phytol. **227**: 38–44.

Godt, D.E. and Roitsch, T. (1997). Regulation and tissue-specific distribution of mRNAs for three extracellular invertase isoenzymes of tomato suggests an important function in establishing and maintaining sink metabolism. Plant Physiol. **115**: 273–282.

Goetz, M., Godt, D.E., Guivarc'h, A., Kahmann, U., Chriqui, D., and Roitsch, T. (2001). Induction of male sterility in plants by metabolic engineering of the carbohydrate supply. Proc. Natl. Acad. Sci. U. S. A. **98**: 6522–6527.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biol. **8**: R24.

Hellens, R.P., Allan, A.C., Friel, E.N., Bolitho, K., Grafton, K., Templeton, M.D., Karunairetnam, S., Gleave, A.P., and Laing, W.A. (2005). Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. Plant Methods **1**: 13.

Jiang, C. and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: Advances through genomics. Nat. Rev. Genet. **10**: 161–172.

Jiang, J. (2015). The "dark matter" in the plant genomes: Non-coding and unannotated DNA sequences associated with open chromatin. Curr. Opin. Plant Biol. **24**: 17–23.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009a). H3.3/H2A.Z double variant-containing nucleosomes mark "nucleosome-free regions" of active promoters and other regulatory regions. Nat. Genet. **41**: 941–945.

Jin, J., He, K., Tang, X., Li, Z., Lv, L., Zhao, Y., Luo, J., and Gao, G. (2015). An *Arabidopsis* transcriptional regulatory map reveals distinct functional and evolutionary features of novel transcription factors. Mol. Biol. Evol. **32**: 1767–1773.

Jin, J., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J., and Gao, G. (2017). PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res. **45**: D1040–D1045.

Jin, Y., Ni, D.-A., and Ruan, Y.-L. (2009b). Posttranslational elevation of cell wall invertase activity by silencing its inhibitor in tomato delays leaf senescence and increases seed weight and fruit hexose level. Plant Cell **21**: 2072–2089.

Kagale, S. and Rozwadowski, K. (2011). EAR motif-mediated transcriptional repression in plants: An underlying mechanism for epigenetic regulation of gene expression. Epigenetics **6**: 141–146.

Khan, A. et al. (2018). JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. **46**: D260–D266.

Klann, E., Yelle, S., and Bennett, A.B. (1992). Tomato Fruit Acid Invertase Complementary DNA : Nucleotide and Deduced Amino Acid Sequences. PLANT Physiol. **99**: 351–353.

Klee, H.J. and Giovannoni, J.J. (2011). Genetics and Control of Tomato Fruit Ripening and Quality Attributes. Annu. Rev. Genet. **45**: 41–59.

Koohy, H., Down, T.A., Spivakov, M., and Hubbard, T. (2014). A comparison of peak callers used for DNase-Seq data. PLoS One **9**: 96303.

Lammens, W., Le Roy, K., Schroeven, L., Van Laere, A., Rabijns, A., and Van Den Ende, W. (2009). Structural insights into glycoside hydrolase family 32 and 68 enzymes: Functional implications. J. Exp. Bot. **60**: 727–740.

Li, J., Foster, R., Ma, S., Liao, S., Bliss, S., Kartika, D., Wang, L., Wu, L., Eamens, A.L., and Ruan, Y. (2021). Identification of Transcription Factors Controlling Cell Wall Invertase Gene Expression for Reproductive Development via Bioinformatic and Transgenic analyses. Plant J.: tpj.15218.

Li, Q., Sapkota, M., and van der Knaap, E. (2020). Perspectives of CRISPR/Cas-mediated cis-engineering in horticulture: unlocking the neglected potential for crop improvement. Hortic. Res. **7**: 1–11.

Li, Z., Palmer, W.M., Martin, A.P., Wang, R., Rainsford, F., Jin, Y., Patrick, J.W., Yang, Y., and Ruan, Y.L. (2011). High invertase activity in tomato reproductive organs correlates with enhanced sucrose import into, and heat tolerance of, young fruit. J. Exp. Bot. **63**: 1155–1166.

Li, Z., Peng, R., Tian, Y., Han, H., Xu, J., and Yao, Q. (2016). Genome-wide identification and analysis of the MYB transcription factor superfamily in solanum lycopersicum. Plant Cell Physiol. **57**: 1657–1677.

Liao, S., Wang, L., Li, J., and Ruan, Y.L. (2020). Cell wall invertase is essential for ovule development through sugar signaling rather than provision of carbon nutrients. Plant Physiol. **183**: 1126–1144.

Liu, Y.H., Offler, C.E., and Ruan, Y.L. (2016). Cell wall invertase promotes fruit set under heat stress by suppressing ROS-independent cell death. Plant Physiol. **172**: 163–180.

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.

Lönnig, W.E. and Saedler, H. (1997). Plant transposons: Contributors to evolution? In Gene (Elsevier), pp. 245–253.

Lu, C., Chen, J., Zhang, Y., Hu, Q., Su, W., and Kuang, H. (2012). Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa*. Mol. Biol. Evol. **29**: 1005–1017.

Lu, P. et al. (2017). The fruitENCODE project sheds light on the genetic and epigenetic basis of convergent evolution of climacteric fruit ripening. bioRxiv: 231258.

Matsukura, C. (2016). Sugar accumulation in tomato fruit and its modification using molecular breeding techniques. In Biotechnology in Agriculture and Forestry (Springer International Publishing), pp. 141–154.

Mavrich, T.N. et al. (2008). Nucleosome organization in the *Drosophila* genome. Nature **453**: 358–362.

Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., Pachter, L.S., and Dubchak, I. (2000). VISTA: Visualizing global DNA sequence alignments of arbitrary length. Bioinformatics **16**: 1046–1047.

Mohanta, T.K., Yadav, D., Khan, A., Hashem, A., Tabassum, B., Khan, A.L., Allah, E.F.A., and Al-Harrasi, A. (2020). Genomics, molecular and evolutionary perspective of NAC transcription factors. PLoS One **15**: e0231425.

Mueller, B., Mieczkowski, J., Kundu, S., Wang, P., Sadreyev, R., Tolstorukov, M.Y., and Kingston, R.E. (2017). Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction. Genes Dev. **31**: 451–462.

Nguyen-Quoc, B. and Foyer, C.H. (2001). A role for "futile cycles" involving invertase and sucrose synthase in sucrose metabolism of tomato fruit. J. Exp. Bot. **52**: 881–889.

Ohta, M., Matsui, K., Hiratsu, K., Shinshi, H., and Ohme-Takagi, M. (2001). Repression Domains of Class II ERF Transcriptional Repressors Share an Essential Motif for Active Repression. Plant Cell **13**: 1959–1968.

Oltman, A.E., Jervis, S.M., and Drake, M.A. (2014). Consumer attitudes and preferences for fresh market tomatoes. J. Food Sci. **79**: S2091–S2097.

Palmer, W.M.M., Ru, L., Jin, Y., Patrick, J.W.W., and Ruan, Y.-L. (2015). Tomato Ovary-to-Fruit Transition is Characterized by a Spatial Shift of mRNAs for Cell Wall Invertase and its Inhibitor with the Encoded Proteins Localized to Sieve Elements. Mol. Plant **8**: 315–328.

Petreikov, M., Shen, S., Yeselson, Y., Levin, I., Bar, M., and Schaffer, A.A. (2006). Temporally extended gene expression of the *ADP-Glc pyrophosphorylase large subunit (AgpL1)* leads to increased enzyme activity in developing tomato fruit. Planta **224**: 1465–1479.

Porebski, S., Bailey, L.G., and Baum, B.R. (1997). Modification of a CTAB DNA Extraction Protocol for Plants Containing High Polysaccharide and Polyphenol Components. Plant Mol. Biol. Report. **15**: 8–15.

Proels, R.K., Hause, B., Berger, S., and Roitsch, T. (2003). Novel mode of hormone induction of tandem tomato invertase genes in floral tissues. Plant Mol. Biol. **52**: 191–201.

Proels, R.K. and Roitsch, T. (2006). Cloning of a CACTA transposon-like insertion in intron I of *tomato invertase Lin5* gene and identification of transposase-like sequences of *Solanaceae* species. J. Plant Physiol. **163**: 562–569.

Renau-Morata, B., Carrillo, L., Cebolla-Cornejo, J., Molina, R. V., Martí, R., Domínguez-Figueroa, J., Vicente-Carbajosa, J., Medina, J., and Nebauer, S.G. (2020). The targeted overexpression of *SlCDF4* in the fruit enhances tomato size and yield involving gibberellin signalling. Sci. Rep. **10**: 1–14.

Van Roekel, J.S.C., Damm, B., Melchers, L.S., Hoekema, A., and Davey, M.R. (1993). Factors influencing transformation frequency of tomato (*Lycopersicon esculentum*). Plant Cell Rep. **12**: 644–647.

Le Roy, K., Vergauwen, R., Struyf, T., Yuan, S., Lammens, W., Mátrai, J., De Maeyer, M., and Van den Ende, W. (2013). Understanding the role of defective invertases in plants: Tobacco nin88 fails to degrade sucrose. Plant Physiol. **161**: 1670–1681.

Ru, L., He, Y., Zhu, Z., Patrick, J.W., and Ruan, Y.-L. (2020). Integrating Sugar Metabolism With Transport: Elevation of Endogenous Cell Wall Invertase Activity Up-Regulates *SlHT2* and *SlSWEET12c* Expression for Early Fruit Development in Tomato. Front. Genet. **11**.

Ru, L., Osorio, S., Wang, L., Fernie, A.R., Patrick, J.W., and Ruan, Y.-L. (2017). Transcriptomic and metabolomics responses to elevated cell wall invertase activity during tomato fruit set. J. Exp. Bot. **68**: 4263–4279.

Ruan, Y.L. (2014). Sucrose metabolism: Gateway to diverse carbon use and sugar signaling. Annu. Rev. Plant Biol. **65**: 33–67.

Shen, S., Ma, S., Liu, Y., Liao, S., Li, J., Wu, L., Kartika, D., Mock, H.P., and Ruan, Y.L. (2019). Cell wall invertase and sugar transporters are differentially activated in tomato styles and ovaries during pollination and fertilization. Front. Plant Sci. **10**: 506.

Sherf, B.A., Navarro, S.L., Hannah, R.R., and Wood, K. V. (1996). Dual-Luciferase TM Reporter Assay : An Advanced Co-Reporter Technology Integrating Firefly and Renilla Luciferase Assays. Promega Notes Mag. Number **57**: 3–10.

Slugina, M.A., Dzhos, E.A., Shchennikova, A. V, and Kochieva, E.Z. (2019). Up-regulation of the vacuolar invertase *TA1* gene may contribute to the accumulation of carotenoids in tomato fruits. Curr. Challenges Plant Genet. Genomics, Bioinformatics, Biotechnol. **24**: 146–149.

Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W., Arva, A., and Lewis, S. (2002). The generic genome browser: A building block for a model organism system database. Genome Res. **12**: 1599–1610.

Tian, F., Yang, D.C., Meng, Y.Q., Jin, J., and Gao, G. (2020). PlantRegMap: Charting functional regulatory maps in plants. Nucleic Acids Res. **48**: D1104–D1113.

Tieman, D. et al. (2012). The chemical interactions underlying tomato flavour preferences. Curr. Biol. **22**: 1035–1039.

Tieman, D. et al. (2017). A chemical genetic roadmap to improved tomato flavour. Science (80-.). **355**: 391–394.

Wang, L. and Ruan, Y.L. (2013). Regulation of cell division and expansion by sugar and auxin signaling. Front. Plant Sci. **4**: 163.

Wang, Z., Wei, X., Yang, J., Li, H., Ma, B., Zhang, K., Zhang, Y., Cheng, L., Ma, F., and Li, M. (2019). Heterologous expression of the apple hexose transporter *MdHT 2.2* altered sugar concentration with increasing cell wall invertase activity in tomato fruit. Plant Biotechnol. J.: 1–13.

Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. PLoS One **6**: e16765.

Wei, H., Chai, S., Ru, L., Pan, L., Cheng, Y., Ruan, M., Ye, Q., Wang, R., Yao, Z., Zhou, G., Chen, Y., and Wan, H. (2020). New insights into the evolution and expression dynamics of invertase gene family in *Solanum lycopersicum*. Plant Growth Regul.: 1–13.

Werner, S., Engler, C., Weber, E., Gruetzner, R., and Marillonnet, S. (2012). Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. Bioengineered **3**: 38–43.

Wingler, A., Juvany, M., Cuthbert, C., and Munné-Bosch, S. (2015). Adaptation to altitude affects the senescence response to chilling in the perennial plant *Arabis alpina*. J. Exp. Bot. **66**: 355–367.

Yuan, Q., Zhang, C., Zhao, T., Yao, M., and Xu, X. (2018). A genome-wide analysis of GATA transcription factor family in tomato and analysis of expression patterns. Int. J. Agric. Biol. **20**: 1274–1282.

Zanor, M.I. et al. (2009). RNA interference of *LIN5* in tomato confirms its role in controlling brix content, uncovers the influence of sugars on the levels of fruit hormones, and demonstrates the importance of sucrose cleavage for normal fruit development and fertility. Plant Physiol. **150**: 1204–1218.

Zhang, H., Lang, Z., and Zhu, J.K. (2018a). Dynamics and function of DNA methylation in plants. Nat. Rev. Mol. Cell Biol. **19**: 489–506.

Zhang, N., Shi, J., Zhao, H., and Jiang, J. (2018b). Activation of small heat shock protein (SlHSP17.7) gene by cell wall invertase inhibitor (SlCIF1) gene involved in sugar metabolism in tomato. Gene **679**: 90–99.

Zhang, T., Marand, A.P., and Jiang, J. (2016). PlantDHS: A database for DNase I hypersensitive sites in plants. Nucleic Acids Res. **44**: D1148–D1153.

Zhang, W., Zhang, T., Wu, Y., and Jiang, J. (2012). Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in *Arabidopsis*. Plant Cell **24**: 2719–2731.

Zhao, J., Sauvage, C., Zhao, J., Bitton, F., Bauchet, G., Liu, D., Huang, S., Tieman, D.M., Klee, H.J., and Causse, M. (2019). Meta-analysis of genome-wide association studies provides insights into genetic control of tomato flavour. Nat. Commun. **10**: 1–12.

Zou, L.-J., Deng, X.-G., Han, X.-Y., Tan, W.-R., Zhu, L.-J., Xi, D.-H., Zhang, D.-W., and Lin, H.-H. (2016). Role of Transcription Factor HAT1 in Modulating *Arabidopsis thaliana* Response to *Cucumber mosaic virus*. Plant Cell Physiol. **57**: 1879–1889.

CHAPTER 5

Re-analysis of tomato Cell wall Invertase Inhibitor 1 (CIF1) function with CRISPR/Cas9

Vera Veltkamp[1], Jeroen Busscher[2], Marrit Alderkamp[3], Gerco C. Angenent[1,4] and Ruud A. de Maagd[4]

[1]Laboratory of Molecular Biology, Wageningen University, The Netherlands

[2]Laboratory of Plant Physiology, Wageningen University, The Netherlands

[3] Plant-Microbe Interactions, Utrecht University, The Netherlands

[4]Bioscience, Wageningen Plant Research, The Netherlands

ABSTRACT

Sugar content is an important determinant of tomato flavour. During sugar import in the fruit, Cell Wall Invertase cleaves sucrose that was transported from photosynthetic source tissues, into the hexoses, fructose, and glucose. Hexoses constitute the majority of sugars in tomato fruit. Increasing the final hexose content and perceived sweetness has been a major breeding goal. Previous work has identified an allele of cell wall invertase *LIN5* with higher enzymatic activity, leading to higher soluble solids content (Brix) in fruit. A promising alternative strategy is to increase invertase activity in fruits by knocking down expression of the post-translational repressor of fruit-specific invertases, Cell-wall Inhibitor of β-fructosidase (CIF1 or INVINH1). In this study CRISPR/Cas9 was used to knockout *CIF1* in an indeterminate tomato variety, cv. Moneyberg. Four mutant alleles were analysed, which all resulted in a slightly higher Brix, but also considerable decrease in fruit size. Effects of the *CIF1* knockout mutations on invertase activity and the expression of other sugar-related genes varied in an unexpected manner.

INTRODUCTION

Tomato is more attractive for consumers and has a higher value when it tastes sweeter (Tieman et al., 2017). Thus, as in many other fruits, a breeding objective for tomato is to increase the amount of sugar in ripe fruit. In ripe cultivated tomato fruit, sugars are present in equimolar amounts of glucose and fructose and a trace amount of sucrose (Winsor, 1966; Petró-Turza, 1986). These sugars are originally formed as sucrose by photosynthesis in source-tissues, e.g. leaves and green fruit. The sucrose is then translocated via bulk-transport through the phloem to sink-tissues such as the developing fruit. To facilitate import into sink-cells, sucrose is hydrolysed by invertases (β-Fructosidase) to form the hexoses, glucose and fructose (Nguyen-Quoc and Foyer, 2001; Matsukura, 2016). Through this process, invertases play an important role in the sugar accumulation in fruits (Klee and Giovannoni, 2011; Matsukura, 2016). To maintain a concentration gradient from source to sink, which allows bulk-transport via the phloem, the sugars in the sink-tissues are stored in vacuoles or converted to starch (N' tchobo et al., 1999).

As the invertases play a crucial step in the formation of the sugar's glucose and fructose, they have been studied intensively over the past years. Several types of invertases can be distinguished: acidic Cell Wall bound Invertase (CWIN), acidic Vacuolar Invertase and alkaline Cytoplasmic or neutral Invertase (Klann et al., 1992; Klee and Giovannoni, 2011). All invertases contain the conserved catalytic "WEC(V/P)DF" domain and the sucrose-recognition sites "NDPN" and "RDP" (Godt and Roitsch, 1997; Alberto et al., 2004; Lammens et al., 2009; Slugina et al., 2019). In tomato, there are four functional *CWINs*:

lycopersicum Invertases (LIN) 5 to *8* (Fridman and Zamir, 2003). *LIN8* and *LIN6* are located on chromosome 10, and have a role in vegetative tissues, while the *CWINs* on chromosome 9, *LIN5* and *LIN7*, are specific to flowers and fruits (Godt and Roitsch, 1997; Fridman et al., 2002; Fridman and Zamir, 2003; Proels et al., 2003). In addition, five *CWIN* pseudogenes and five non-functional CWINs that lack the "NDPN" motif have been identified recently (Wei et al., 2020). Of the *CWINs*, *LIN5* has received abundant attention in the last two decades. The main reason for this is the identification of an introgression line derived from a cross between cv. M82 and *S. pennellii*, Brix9-2-5 (Eshed and Zamir, 1996). This line had an increased sugar content (+28% glucose, +18% fructose) and a 20-35% increased Brix (the amount of sugar in tomato is highly correlated with the soluble solids, which is measured in degrees Brix), without an apparent negative effect on yield (Fridman et al., 2002). Fine mapping of the QTL revealed linked non-synonymous nucleotide substitutions in the coding sequence of *LIN5*. These lead to three amino acid substitutions as the underlying cause of higher sugar content through increased invertase activity in the *S. pennellii* variant (Fridman et al., 2000, 2004). *LIN5* is expressed in the reproductive organs and in early fruit development. Four hours after pollination, *LIN5* transcript levels increased in the style. At two days after pollination, *LIN5* expression was highest in the ovaries (Shen et al., 2019).

Like other invertases, LIN5 activity is post-translationally regulated by a proteinaceous inhibitor, Cell-wall Inhibitor of β-fructosidase (CIF1), also known as Invertase Inhibitor 1 (INVINH1). *CIF1* is expressed both in vegetative and reproductive tissues. The highest *CIF1* expression levels are in roots and in maturing fruit: the expression increases from time of flowering to 20 days after anthesis (Jin et al., 2009). *CIF1* is co-expressed and the protein is physically interacting with LIN5 during the ovary-to-fruit transition (Le Roy et al., 2013; Palmer et al., 2015). RNA interference (RNAi) lines directed at knocking down *CIF1* expression resulted in increased CWIN activity, measured in leaves, roots and developing fruits. This resulted in an approximately 25% increase in sugar content in ripe fruits, a prolonged leaf life span due to inhibition of abscisic acid-induced senescence, and an increased seed weight (Jin et al., 2009).

Many studies that report increase in sugar content in tomato through genetic modification are often using a single, often poorly described cultivar. Thus, in this study we aimed at verifying the effect of *CIF1* knockdown on sugar content in our model cv. Moneyberg, an indeterminate variety grown in greenhouses. If successful, this would demonstrate the feasibility of a transgene-free approach to higher sugar in tomato by targeting *CIF1* with CRISPR/Cas9-mutagenesis. *CIF1* knockout mutants in cv. Moneyberg showed a modest increase in sugar content, measured by °Brix. However, at the same time the resulting fruits were significantly smaller in size. These results question the universal applicability of this approach without a yield penalty.

RESULTS

CIF1 PROTEIN SEQUENCE IS CONSERVED THROUGHOUT CULTIVATED TOMATO AND ITS NEAREST RELATIVES

To assess the evolutionary history and importance of Cell-wall Inhibitor of β-fructosidase (CIF1), we constructed a phylogenetic tree (**Figure 1a**) using protein sequences of CIF homologs from tomato, potato, pepper, and *Arabidopsis* (Saitou and Nei, 1987; Kumar et al., 2018; Stecher et al., 2020). Tomato CIF1 and its homologs in the three *Solanaceous* species are most homologous to *Arabidopsis* C/VIF1, which was characterized as Vacuolar Invertase Inhibitor (VIF) (Link et al., 2004). In tomato, this group comprises both the Vacuolar invertase inhibitor VIF as well as CIF1, and a closely related homolog, here named CIF1-like. Their tandem arrangement as shown by their consecutive numbering, which is conserved in all three species, suggests a common origin involving gene duplication events (**Figure 1b**). By combining the phylogenetic tree with the synteny information, and with the exon/intron configuration, it seems most likely that a gene duplication in a common ancestor of tomato and potato, but not of pepper resulted in the formation of the *CIF1* and *CIF1-like* genes. A duplication event in a common ancestor of all three *Solanaceous* species may have resulted in the common *VIF* and *CIF* lineages, where the *VIF* lineage lost the intron.

Solanum CIF1 and CIF1-like were compared in a protein alignment (**Figure 2**). Four cysteine residues form a disulphide bridge between seven α-helixes and stabilize the protein (Datir and Ghosh, 2020a). The C-terminal domain also contributes to interface stabilization (Hothorn et al., 2004a, 2004b; Rausch and Greiner, 2004). Three conserved amino acids, Pro117-Lys119-Phe120 (PKF) interact with the active site of invertase (Hothorn et al., 2010). The adjacent conserved Gly-115 and Ala-121 may also play a role in targeting invertase (Datir and Ghosh, 2020a). The alignment of CIF1 and CIF1-like proteins from five different *Solanum* species reveals overall homology between the two groups, with a number of differences that are conserved among the members of the two proteins. A notable exception is a proline residue (P148 or P145, in CIF1 and CIF1-like, respectively). While conserved in both proteins in all other species, it is substituted with glutamine (Q) in CIF1-Like proteins in all 66 cultivated *S. lycopersicum* varieties, as well as in the *S. pimpinellifolium* and *S. cheesmanii* accessions of the re-sequenced collection (Roohanitaziani, 2019). Further exploration of Blast results throughout the *Angiosperms* reveals the conservation of this proline residue. Hence, the prediction of the effect on protein function by Provean resulted in a score of -6.9 for the substitution (where a score below -2.5 is considered deleterious, Choi et al., 2012), suggesting that the substitution in tomato CIF1-Like if expressed, may have a negative effect on protein function. In summary, both *CIF1* as well as *CIF1-Like* are extremely conserved, suggesting that *CIF1* has been under strong selection pressure well before tomato domestication occurred.

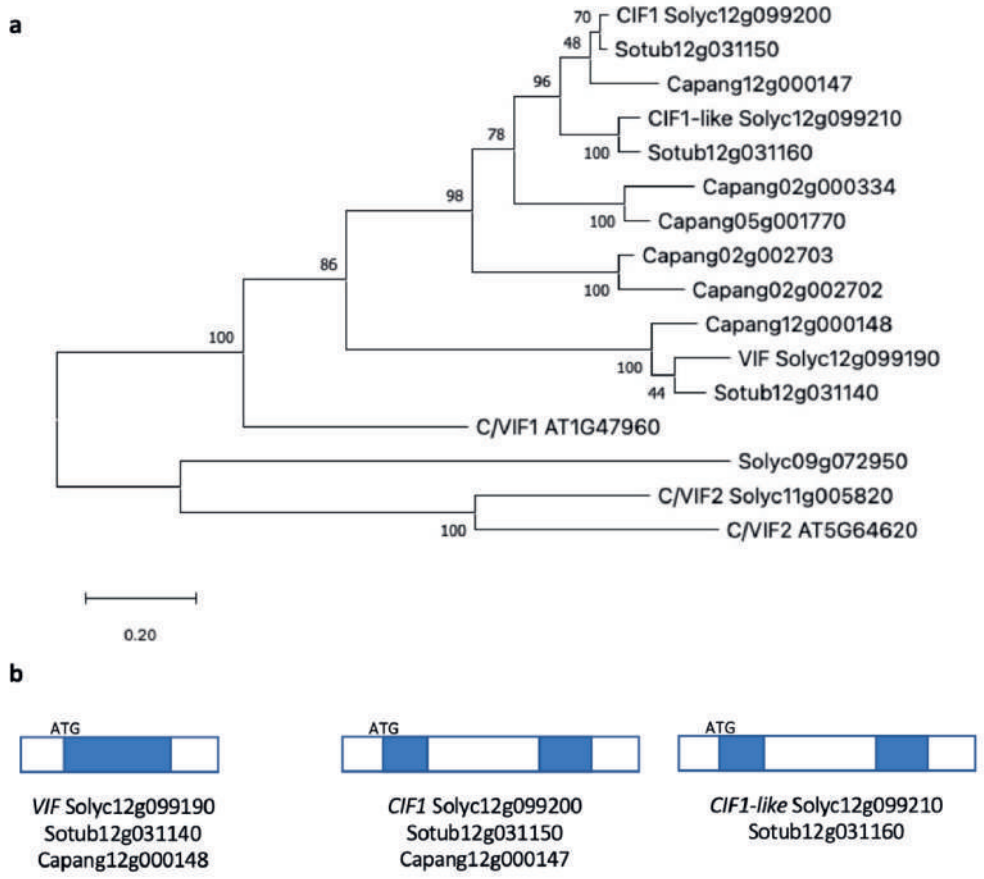


Figure 1: **(a)** Phylogenetic tree of CIFs and Vacuolar Invertase Inhibitors (VIF). The phylogenetic tree based on proteins sequences was constructed in MEGAX program (Kumar et al., 2018; Stecher et al., 2020), by the Maximum Likelihood method and JTT matrix-based model with a 100-fold bootstrap-test (Saitou and Nei, 1987). Bootstrap values (%) are indicated at each branch point. The Scale bar marks 0.2 amino acid substitution per site, which was computed using the Poisson correction method (Zuckerlandl and Pauling, 1965). The C/VIF2 proteins were used as the outgroup. The protein sequences listed in the tree were obtained from the SolGenomics Network website (Solgenomics.net) for tomato (Solyc numbers), potato (Sotub) and pepper (Capang) of the GenBank database for Arabidopsis (https ://www.ncbi.nlm.nih.gov/). **(b)** Synteny of the invertase inhibitor genes on chromosome 12 in three species. The coding sequence is depicted as blue boxes.

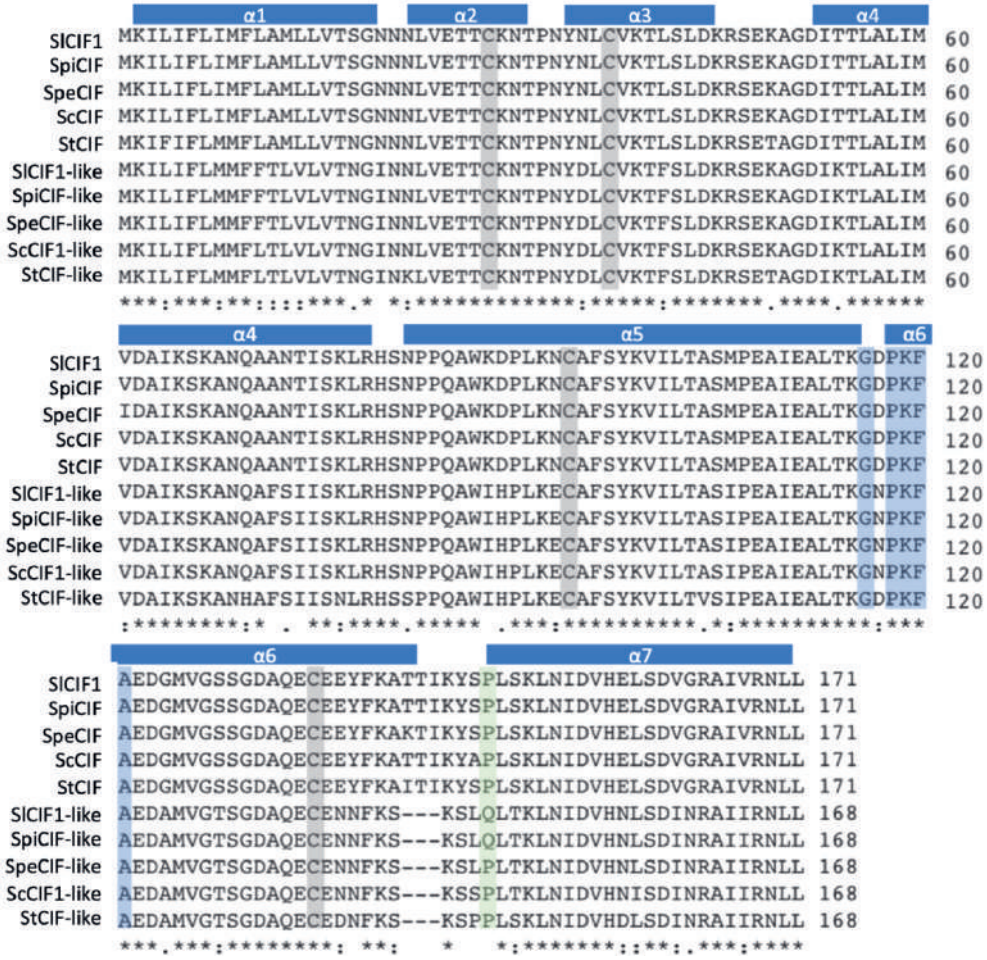


Figure 2: Protein alignment of *Solanum* CIF1 and CIF1-like amino acid sequences. Amino acids sequences were aligned using multiple sequence alignment in the EMBL-EBI search and sequence analysis tool (https://www.ebi.ac.uk/Tools/msa/clustalo/) using ClustalW with default parameters (Madeira et al., 2019). The amino acids are represented by a single letter code. The protein sequences listed in the diagram are SlCIF1 (*S. lycopersicum,* Solyc12g099200), SpiCIF (*S. pimpinellifolium,* Sopim12g099200), SpeCIF (*S. pennellii,* XP_015059317), ScCIF (*S. chilense,* TMW97421), StCIF1 (*S. tuberosum,* Sotub12g031150), SlCIF1-like (*S. lycopersicum,* Solyc12g099210), SpiCIF-like (*S. pimpinellifolium,* Sopim12g099210), SpeCIF-like (*S. pennellii,* XP_015060934), ScCIF-like (*S. chilense,* TMW97422), StCIF-like (*S. tuberosum,* Sotub12g031160). Conserved functional amino acids are indicated with blue boxes. Grey boxes indicate disulphide bridges between the α-helixes. The green box indicates the position of the P to Q substitution in CIF1-like. Secondary structure indications are based on the predictions of (Datir, 2020; Datir and Ghosh, 2020a). Asterisks at the bottom of the alignment indicate identical residues. Amino acid numbers are listed on the right of the alignment.

CIF1-CR MUTANTS HAVE A HIGHER BRIX IN SMALLER FRUITS

CIF1 is conserved in cultivated tomato, which suggests an important function in normal fruit set or development. A study that used RNA interference to target *CIF1* transcripts increased the invertase activity and the sugar content of ripe fruit (Jin et al., 2009). The study reported no effect on fruit size. We set out to test the reproducibility of these results and the effect of a *CIF1* specific knockout in our greenhouse model-variety: cv. Moneyberg. To be able to create transgene-free offspring and true knockouts, we used CRISPR/Cas9 mutagenesis.

We obtained knockout mutations by using binary vectors containing *SpCas9* and four single-guide RNAs (sgRNAs) targeting the first exon of *CIF1* (**Figure 3a, b**). sgRNAs were chosen based on predicted efficiency and the absence of off-targets in the tomato genome with less than 2 mismatches and/or 1 bulge (Bae et al., 2014; Chari et al., 2015; Doench et al., 2016; Concordet and Haeussler, 2018). Stable transformation of tomato cotyledon explants was performed and 15 primary transformants (T_0) were obtained of which 10 had a mutation. In the obtained mutants, different CRISPR alleles occurred with deletions of one to eight base pairs in one or more targets being the most common, as in *cif1-cr1* and *cif1-cr4* (**Figure 3b**). Furthermore, we found an inversion of the sequence between two targets (*cif1-cr2*) and a 114 bp insertion in one target (*cif1-cr4*) (**Figure 3b**). A Blastn search revealed that the insertion in *cif1-cr4* originated from a non-coding region on chromosome 5. The diversity in the positions of the observed mutations showed that all four sgRNAs were active in the experiment. Moreover, the simultaneous occurrence of indels in two overlapping target sequences shows that having overlapping target sequences does not preclude Cas9 activity or Double Stranded Break repair at both sites, as shown in *cif1-cr2* and *cif1-cr3*. Five different alleles were selected for further study (**Figure 3b**). *cif1-cr1* and *cif1-cr2* were found in one bi-allelic T_0 plant and *cif1-cr3* and *cif1-cr4* in another. The fifth line had a large deletion between sgRNA target 1 and target 4 (not shown in Figure 3b). However, this plant produced only small seedless fruits. This was possibly due to a tissue culture-effect, or the severity of the mutation may have caused reduced fertility. The 3-bp deletion in *cif1-cr1* caused a Q85 deletion in the protein, which was predicted to have a neutral effect (Provean score=0.92). The other three mutations resulted in premature stop-codons and predicted truncated proteins lacking the conserved residues that interact with LIN5 (**Figure 3c**). Both the inversion and the insertion of alleles *cif1-cr2* and *cif1-cr4,* respectively, occurred downstream of a frameshift mutation in target 2 causing a premature stop-codon and thus are not expected to result in entirely new protein sequences. All guides were selected to have no additional targets with zero or one mismatches in the protospacer sgRNA target sequence. Nevertheless, to rule out a potential off target mutation in the next best target, *CIF1-Like* with two mismatches (**Table 1**), all lines were genotyped in the *CIF1-Like* region. No mutations were found. The next closest off-targets had three or more mismatches and were not considered likely to have mutations (**Table S1**).

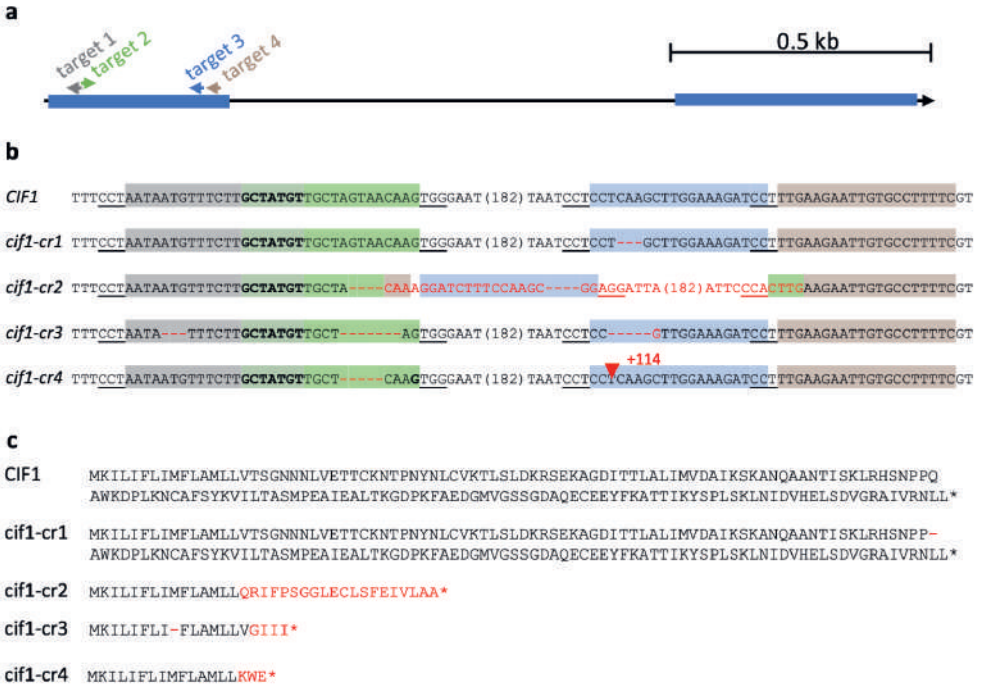


Figure 3: (a) The *CIF1* gene and CRISPR/Cas9 targets. The coding sequence is depicted as blue boxes. **(b)** CRISPR/Cas9 induced mutations. sgRNA target sequences are highlighted in different colours. The text is bold where target 1 and target 2 overlap. Protospacer-adjacent motif (PAM) sequences are underlined. Red dashes and letters indicate mutations. Numbers in parentheses indicate the number of bases not depicted. (**c**) Resulting protein sequences. Red dashes and letters indicate deletions or additional amino acids due to a frame shift caused by an indel.

The selected T_0 plants were self-fertilized, and the progeny segregated in the expected ratios for a single transgene and the homozygous or bi-allelic mutations. Phenotyping was done in the Cas9-free homozygous T_2 plants for the most relevant traits: Brix and size. Three plants per genotype (mutants and WT control) were placed in random positions in a greenhouse. Of these plants, three trusses were labelled, vibrated for pollination, and pruned down to 6 fruits per truss. At breaker+7 days red-ripe fruit were used to determine Brix, width, and weight. All tested mutant plants produced fruit with a significantly higher (3.6-11%) Brix compared to wild type (**Figure 4a, Table 2, S2**). However, the size and weight of fruits from mutant plants was on average approximately 25% lower than that of wild-type fruits (**Figure 4b, c, Table S2**). This result confirms the potential of a *CIF1* knockout for increasing sugar content in tomato fruit, but in contrast to earlier reports, with a major trade-off for fruit size.

Table 1: Genome matches for sgRNA Target 3

sgRNA target sequence	Mismatches	Gene	Region
GGATCTTTCCAAGCTTGAGGAGG	0	Solyc12g099200 (CIF1)	cds
GGAT**G**T**A**TCCAAGCTTGAGGAGG	2	Solyc12g099210 (CIF1-Like)	cds

Underlined sequences represent the PAM site. Bold sequences represent mismatches

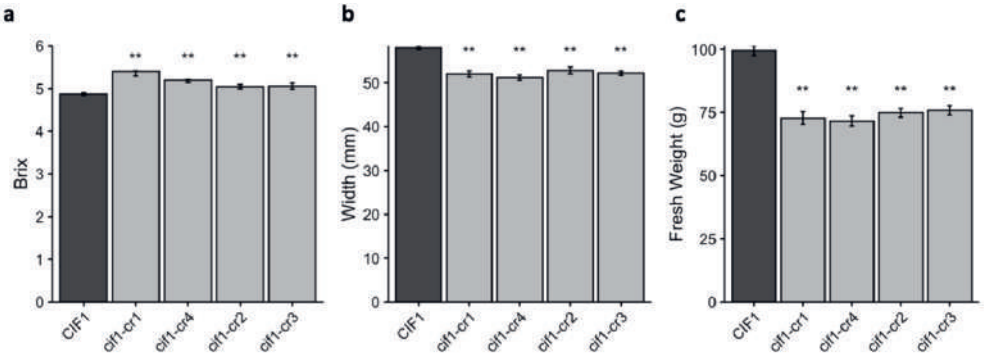


Figure 4: (a) Brix, **(b)** width and **(c)** fresh weight of T_2 fruits from wild-type (CIF1) and homozygous mutants. Six fruits from three trusses were measured for each plant, with three plants per genotype. Mean values (± SE) were compared to CIF1 using one-way ANOVA after a check for normal distribution of the data. Significant differences are represented by asterisks: (∗) P-value < 0.05, (∗∗) P-value < 0.01.

INVERTASE ACTIVITY IS NOT INCREASED IN ALL *CIF1-CR* MUTANTS

The premise for our research was that by generating a knockout of *CIF1,* the sugar content would increase as lifting the repression by CIF1 would increase LIN5 activity. To test this hypothesis, we performed an invertase activity assay on whole fruits at 10 days after anthesis (DAA) (**Figure 5**), 10 DAA being the same timepoint used in the RNA interference (RNAi) study of *CIF1* (Jin et al., 2009). In addition, it is a stage when *LIN5* expression starts to decrease and *CIF1* expression starts to increase (Fridman et al., 2002; Fridman and Zamir, 2003; Proels et al., 2003; Jin et al., 2009). Initially, we assayed fruits of the T_1 generation, one plant per genotype with homozygous or biallelic mutations (**Figure 5a**). This invertase assay demonstrated that the mutation in *cif1-cr2* resulted in an increased activity of CWIN. The *cif1-cr1* mutant with the predicted neutral mutation had no increased activity (**Figure 5a**). As might be expected from these results, the bi-allelic plant containing both the *cif1-cr1* and *cif1-cr2* alleles had an intermediate phenotype. However, the bi-allelic combination of *cif1-cr3* and *cif1-cr4,* surprisingly, did not lead to a change in activity when compared to the wild-type (*CIF1*), while it is expected to be a full null mutant. To confirm these phenotypes, another round of invertase assays was done in the T_3 generation, with three homozygous plants per mutation (**Figure 5b**). The same results as in the T_1 generation were obtained. Although now the *cif1-cr1* allele shows a slight, albeit non-significant, increase in CWIN activity compared to wild-type. Thus, it appears that in our mutants Brix increase and invertase activity are not correlated.

To explore the cause of the discrepancies between the mutant lines, RT-PCR and qPCR expression analyses were performed at 10 DAA (**Figure 5c**). The RT-PCR amplification of the *CIF1* transcript was done in the T_1 generation. All mutations showed the expected transcript size, including that of the *cr4* allele with its 114 nt insertion, giving no evidence of alternative mRNA-splicing (**Figure 5d**).

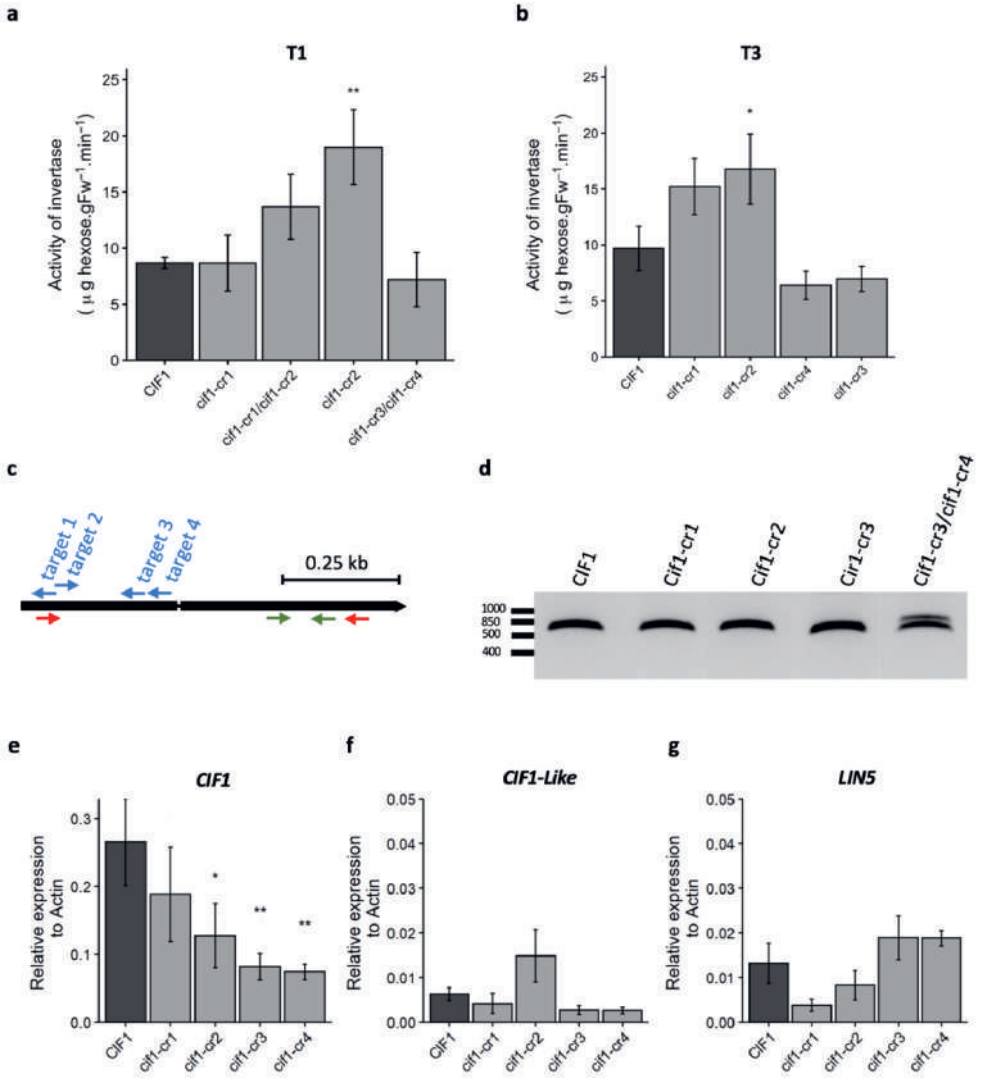


Figure 5: Invertase activity measured in the columella **(a)** of fruit in the T1 generation and **(b)** whole 10 DAA old fruit in the T3 generation. Per genotype, three samples of one plant (T1) or two samples of three plants (T3) were taken as replicates. Each sample was a pool of three fruits from the same truss. **(c)** Location of RT-PCR (red arrows) and qPCR (green arrows) primers on the *CIF1* mRNA. **(d)** RT-PCR of *CIF1*. Expected Wild-type band (CIF1) of 660 bp. For *Cif-cr4* a 114 bp increase is expected. **(e-g)** Expression in 10 DAA T3 fruit of **(e)** *CIF1*, **(f)** *CIF2* and **(g)** *LIN5* in wild-type (CIF1) and *cif1-cr* mutants. Per genotype, three plants were sampled twice. Each sample was a pool of three fruits from the same truss. Expression was measured by qPCR with $2^{-\Delta CT}$, relative to the reference gene *Actin*. Mean values (± SE) were compared to *CIF1*, using one-way ANOVA after a check for normal distribution of the data. Significant differences are represented by asterisks; (*) P-value < 0.05, (**) P-value < 0.01, (***) P-value <0.001.

For the expression analysis, T_3 plants (three plants per genotype and two samples per plant) were used. *CIF1* mRNA levels were significantly decreased in the three lines with mutations causing premature stop codons (*cif1-cr2-4*), which is consistent with

Nonsense Mediated mRNA Decay caused by nonsense codons preceding an intron (**Figure 5e**). No significant expression changes were found for *CIF1-Like* or *LIN5* (**Figure 5f, g**). Although these results confirm that our mutants behave as true knockouts, they do not explain the diverging effect on CWIN activity.

THE *CIF1* MUTATIONS AFFECT EXPRESSION OF SUGAR RELATED GENES IN DIFFERENT WAYS

The sugar pathway and source-sink dynamics involve many genes that form a tightly regulated pathway. Short distance transport from cell-to-cell is achieved by an array of sugar transporters: SUGARS WILL EVENTUALLY be EXPORTED TRANSPORTERS (SWEET), Sucrose Transporters (SUTs) and Hexose Transporters (Yadav et al., 2015; Zhang and Turgeon, 2018; Fernie et al., 2020). When CIF1 activity is changed, we hypothesized that changes in the expression of other sugar-related genes might occur (Ru et al., 2020). In addition, ectopic expression of CIF1 resulted in a large increase of a Small Heat Shock Protein (*SlHSP17.7*), suggesting regulatory feedback between *CIF1* and *SlHSP17.7* (Zhang et al., 2018b). We hypothesized that in the absence of CIF1, the *SlHSP17.7* expression could be decreased in our mutants. By other pathways than via *LIN5*, the sugar pathway could thus be influenced. Indeed, the two *CIF1* mutants that did not have increased CWIN activity, *cif1-cr3* and *cif1-cr4,* had a severely decreased *SlHSP17.7* expression at 10 DAA, while the *cif1-cr1* and *cif1-cr2* showed a slight but not significant increase in expression **(Figure 6**).

To assess what other sugar related genes were influenced we also performed expression analysis for several other sugar related genes in the T_3 10 DAA fruits (**Figure 6**). Of the Neutral Cytosolic Invertases (CINs), *CIN2* is unaffected in all mutant lines (and very lowly expressed at 10 DAA), while *CIN3* expression is decreased in all lines. Differentiation between our mutants occurred for *CIN4*, where just *cif1-cr1* has a higher expression. *CIN5* had a decreased expression in *cif1-cr1* and *cif1-cr2*. *Defective CWIN* (*deCWIN*), *Hexose Transporter 2* (*HT2*) and *LIN9* were hardly expressed at 10 DAA and there was no difference between the wild type and the mutants. Of the *SWEETs* measured, *SWEET14* was unaffected, *SWEET12c* was downregulated in all mutants and *SWEET7a* was upregulated in *cif1-cr2* and *cif1-cr4.* From these data it appears that that knocking out *CIF1* does affect expression of other genes that are involved in sugar transport or accumulation, but that different knockout mutations have different levels of effect. This difference is most striking for *HSP17.7*, where two mutants show more than 10-fold reduction in expression, while two others show no significant difference in expression.

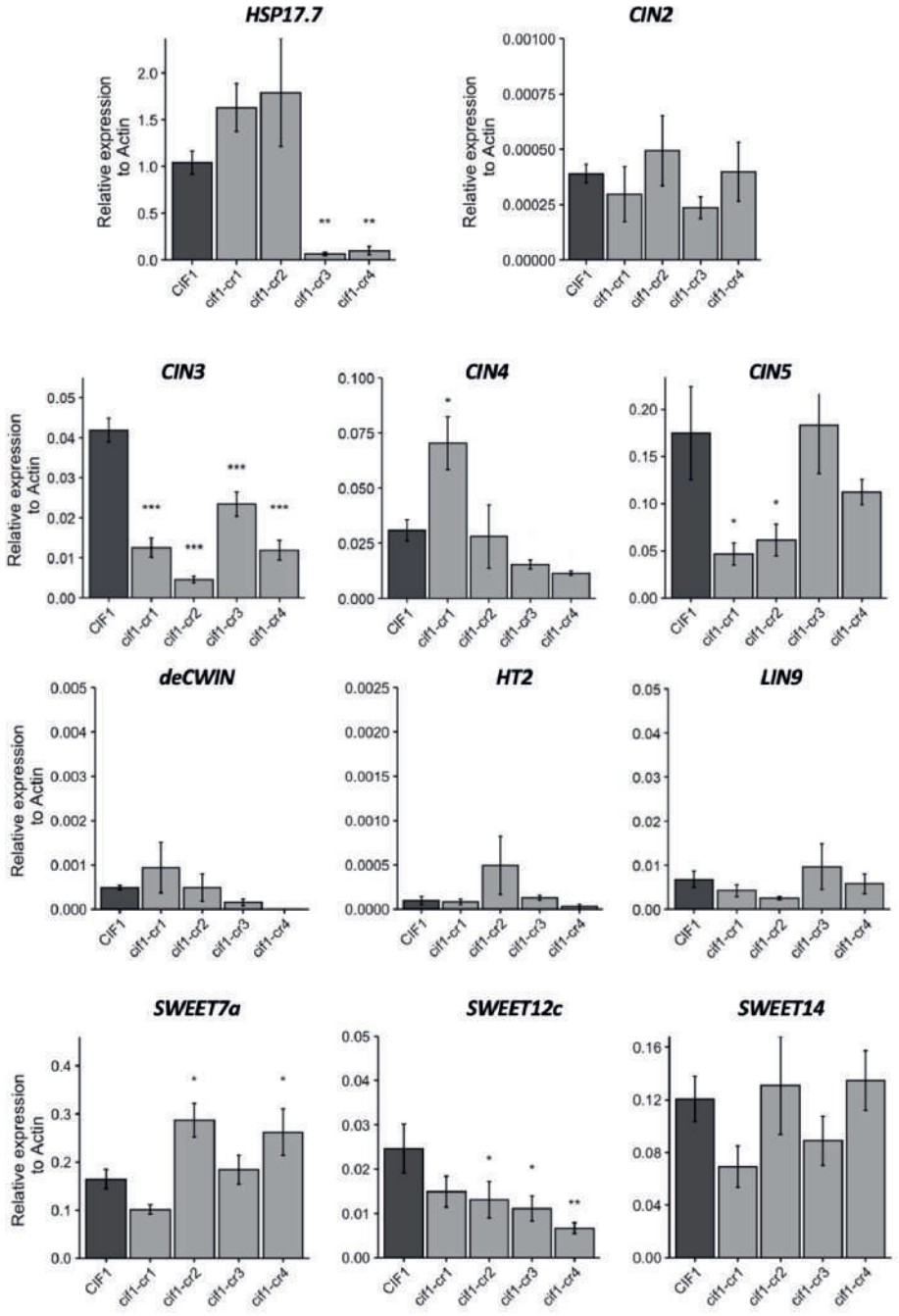


Figure 6: Gene expression in T3 fruit at 10 DAA of in wild-type (CIF1) and cif1-cr mutants. Per genotype, three plants were used. From each plant, two samples were taken. Each sample was a pool of three fruits from the same truss. Expression was measured by qPCR and expressed as 2-ΔCT, relative to the reference gene Actin. Mean values (± SE) were compared to CIF1, using one-way ANOVA after a check for normal distribution of the data. Significant differences are indicated by asterisks; (*) P-value < 0.05, (**) P-value < 0.01, (***) P-value <0.001.

DISCUSSION

It is important to understand the process of sugar accumulation in fruit and the mechanisms controlling it, given the importance in the final quality perception of tomato (Baldwin et al., 2000; Klee and Giovannoni, 2011; Tieman et al., 2012, 2017). Tomato mostly contains glucose and fructose and Cell Wall Invertase (CWIN) hydrolyses sucrose to form these hexoses (Fridman and Zamir, 2003). Increased activity of a fruit specific CWIN, LIN5, has been linked to increased sugar content (Fridman et al., 2002; Baxter et al., 2005; Jin et al., 2009; Zanor et al., 2009). Silencing Cell-wall Inhibitor of β-fructosidase (CIF1 or INVINH1), increasing LIN5 activity and sugar content in the ripe fruit, gave promising results from a breeding perspective (Jin et al., 2009). In this study we have attempted to further support the proposed positive effect of decreasing tomato CIF1 activity on fruit sugar content, by producing true knockout mutants of *CIF1* using CRISPR/Cas-mutagenesis. While the previous study demonstrated that knockdown by RNA interference (RNAi) led to increased Brix and sugar content, we found a modest increase in Brix in all mutants, yet without a clear correlation with increased CWIN activity (Table 2). Moreover, all mutants had lower fruit weight, which was not reported earlier. Effects of the *CIF1* knockout mutations on the expression of other sugar-related genes also varied widely.

Table 2: Overview of mutant phenotypes

Mutant	Mutation	Brix (%)	Weight (%)	CWIN activity T3 (%)	Increased expression	Decreased expression
Cif1-cr1	-3	+11.0**	-26.8**	+56.9	*CIN4*	*CIN3, CIN5*
Cif1-cr2	-8, inversion	+ 3.6**	-24.7**	+89.7*	*SWEET7a*	*CIN3, CIN5, SWEET12c*
Cif1-cr3	-14	+3.8**	-23.8**	-28.8		*HSP17.7, CIN3, SWEET12c*
Cif1-cr4	+109	+6.6**	-28.0**	-28.1	*SWEET7a*	*HSP17.7, CIN3, SWEET12c*

Significant differences are indicated by asterisks; (*) P-value < 0.05, (**) P-value < 0.01

DIFFERENT MUTATIONS IN *CIF1*, INCLUDING SEVERAL KNOCKOUTS, AFFECT CWIN ACTIVITY DIFFERENTLY

We found a clear negative correlation between size and brix in our knockout lines, but the underlying mechanism was not as straightforward as expected. The RNAi study of Jin et al. showed an increase in CWIN activity, attributed to a posttranslational inhibitory effect of CIF1 on LIN5 (Jin et al., 2009). However, when we measured invertase activity at 10 days after anthesis (DAA), the same timepoint as in the RNAi study, we only found increased CWIN activity in one of our lines (*cif1-cr2*). For the other alleles, only a slight, but not significant increase (*cif1-cr1*) or no increase at all (*cif1-cr3, cif1-cr4*), was found. This may be caused by missing the peak of LIN5 activity in our genetic background, cv.

Moneyberg. LIN5 expression and CWIN activity in the fruitlet increases right after pollination, in a window between 4 hours and 2 days (Shen et al., 2019). CIF1 expression is unchanged during this period. This early activity of LIN5, in the cell division stage of development, is key to normal development (Palmer et al., 2015; Ru et al., 2017). Silencing LIN5 in this period led to smaller seeds and abortion rates in both tomato and *Litchi chinensis* (Zanor et al., 2009; Zhang et al., 2018a). *LIN5* expression at 10 DAA did not change in our mutants, which is in line with *CIF1* acting as a post-translational inhibitor.

The differences between the RNAi study and our CRISPR/Cas9 mutagenesis approach, could be explained by the different techniques. RNAi has been a widely used method to evaluate gene function (Saurabh et al., 2014). However, the approach may lead to an incomplete suppression of expression or lack of specificity for the intended target gene. In the *CIF1*-RNAi lines, the RNAi-construct could have affected *CIF1-like* mRNA, as only 25 SNPs compared to *CIF1-like* mRNA were present in the 300 bp fragment used in the inverted repeat leading to small interfering RNAs. The RNAi study does not mention expression differences of *CIF1-like*. However, we do not expect an effect of downregulation of *CIF1-like* if that occurred, since its expression is low in fruit and it contains a possibly deleterious P145Q substitution. A CRISPR/Cas9 mutagenesis approach might not necessarily reveal gene function either (Wang et al., 2020). Expect for *cif1-cr1*, all our mutants were predicted to form severely truncated proteins (**Figure 3c**). However, an alternative start codon might be present and become functional if a mutation occurs in the first exon, as in our case in *cif1-cr2, -cr3* and *-cr4* substantial mutations occurred in the first exon (Merchante et al., 2017). There are two in-frame ATGs in the second exon of CIF1, M106 and M125 (**Figure 2**). Use of either would create a severely truncated protein. Start of translation at M106 would result in a protein containing the PKF domain needed for interaction with the active site of invertase (Hothorn et al., 2010). However, whether this occur and whether the protein is stable without the first five α-helixes, has not been validated in this study. In addition, it would not explain the difference between the mutations. In other studies, exon skipping occurred if the exon contained a Premature Stop Codon after mutagenesis. This exons skipping creates an alternative spliced mRNA that may partially restore protein function (Syed et al., 2012; Mou et al., 2017; Rodriguez-Rodriguez et al., 2018). *CIF1* only consists of two exons, with the Premature Stop Codons occurring in the first exon. No alternative splice variants were detected via RT-PCR. Expression of *CIF1* decreased in our mutants, which is in line with Nonsense Mediated mRNA Decay caused by nonsense codons preceding an intron (Lykke-Andersen and Jensen, 2015). *CIF1-like* could function redundantly with *CIF1*, although the effect is expected to be marginal as *CIF1-like* expression is very low and did not increase in the mutants. These findings combined, indicate that the alleles generated (except *cif1-cr1*) are true knockouts. Thus, the difference in phenotypes remains unexplained. *Cif1-cr1* had an in-frame deletion of one amino acid (Q85), which was predicted to be a neutral deletion (Provean score=0.92). The sequence is conserved in all *Solanum CIF1s* and *CIF1-likes* (**Figure 2**),

but Q85 is not conserved in other CIF and VIF homologs (Datir and Ghosh, 2020b). From our analysis, Q85 does seem to play a role in the sugar accumulation, as *cif1-cr1* had the highest increased Brix and decreased fruit size. In addition, the Q85 deletion led to a decreased *CIN3* and *CIN5* expression, while *CIN4* expression increased (**Table 2**). Thus, our mutants confirm that even a slight disruption can lead to a phenotype difference.

CIF1 KNOCKOUT AFFECTS EXPRESSION OF SEVERAL GENES IN THE SUGAR-PATHWAY

CIF1 is known to increase the expression of the a Small Heat Shock Protein (*HSP17.7*) when ectopically expressed (Zhang et al., 2018b). In two of our knockout mutants, *HSP17.7* expression decreased severely at 10 DAA. These were the two lines (*cif1-cr3, -cr4*) that did not have the increased CWIN activity. The decreased *HSP17.7* expression in these mutants might be the underlying cause of the similar Brix-phenotype, but divergent CWIN activity. In earlier work, downregulation of *HSP17.7* in the sucrose accumulator *S. chmielewskii* decreased sucrose content (Zhang et al., 2018b). Another study in *S. lycopersicum* cv. Moneymaker, overexpression of *HSP17.7* increased sucrose content, but decreased fructose and glucose (Zhang et al., 2020). From this it can be concluded that *HSP17.7* influences sucrose content. As *S. lycopersicum* is a hexose accumulator, the decreased *HSP17.7* expression in *cif1-cr3* and *-cr4* could lead to a decrease in sucrose and concomitant increase in fructose and glucose. This would lead to an increased Brix, as we have measured.

Other sugar-related genes were affected as well in the *CIF1* CRISPR mutants (**Table 2**). *Neutral Cytosolic Invertases 5* (*CIN5*) had the opposite expression pattern as *HSP17.7* in the mutants: in *cif1-cr1* and *-cr2* expression severely decreased, while in the *cif1-cr3* and *-cr4* expression was not different compared to wild type. *CIN3* decreased in all lines and *CIN4* increased just in *cif1-cr1*. Evidently, *CIF1* influences the expression of some *CINs*. The stark increase of *CIN4* in *cif-cr1* could result in the increased hexose content needed to achieve a higher brix in a line that was predicted to have a neutral mutation. On the other hand, *CIN3* and *CIN5* are downregulated in this line. *SUGARS WILL EVENTUALLY be EXPORTED TRANSPORTERS* 12c (*SWEET12c*) was downregulated in all lines, while *SWEET7a* was one of the few tested genes that was upregulated. On average, most differentially regulated genes were downregulated in the mutants. From this data it becomes clear that CIF1 mutations not only affect LIN5 (post-translationally), but also affect other sugar-related genes in an unexpected manner. This may be due to sugar-feedback loops directed on different genes.

THE RELATIONSHIP BETWEEN SIZE AND BRIX

The increase in sugar content that we found in our generated knockout lines are in line with the earlier RNAi results (Jin et al., 2009). Their study found an increase of about 25% sugar content, which is markedly more than the Brix increase we found (about 5%).

Brix measures the total soluble solids content, which is mainly sugars in tomato. Thus, the different measurements should be comparable. In addition, the RNAi study did not report a size decrease. As CIF1 is a post-translational inhibitor of LIN5, it is worthwhile to address known LIN5 sequence variation effects. The Brix9-2-5 QTL derived from a *S. pennellii* introgression, later confirmed to have a more active *LIN5* protein as the underlying mechanism, caused an increase in fruit sugar content but did not result in smaller fruit (Eshed and Zamir, 1996; Fridman et al., 2000, 2004). The experiments were done with cv. M82 as a parent, a determinate processing tomato. The *S. pennellii* allele was also effective in increasing Brix in indeterminate lines, but no mention of size or weight was made (Fridman et al., 2000). In a recent study where one of three amino acid changes of the *S. pennellii* allele was used in a ectopic *LIN5* expression comparison, a negative correlation between yield and sugar content was found (Tieman et al., 2017). Based on our results, it seems that impairing *CIF1* would be unattractive for breeders and growers as the decreased weight is not compensated by an increased Brix.

In conclusion, this study has revealed unexpected outcomes of what were expected to be relatively straightforward knockout mutation experiments. These outcomes may be either due to complex feedback regulatory loops between invertase inhibitor activity and invertase expression or activity, or to unexpected effects on gene activity from Crispr/Cas-mutations in the open reading frame, through yet-to-be-discovered mechanisms.

MATERIALS AND METHODS

PHYLOGENETIC TREE CONSTRUCTION

The phylogenetic tree based on proteins sequences was constructed using MEGAX software (Kumar et al., 2018; Stecher et al., 2020), with the Neighbour-Joining method using a 1000-fold bootstrap-test (Saitou and Nei, 1987). The amino acid substitution per site was computed using the Poisson correction method (Zuckerlandl and Pauling, 1965). The protein sequences listed in the tree were obtained from the SolGenomics Network website (Solgenomics.net) or the GenBank database (https ://www.ncbi.nlm.nih.gov/).

CRISPR/CAS9-MUTAGENESIS DESIGN AND VECTOR ASSEMBLY

Four spacers targeting the first exon of *CIF1* (**Figure 1A, 1B, Table S3**) were selected. Online programs, sgRNA scorer 1.0 (https://crispr.med.harvard.edu/sgRNAScorer/) and the GPP sgRNA designer tool (http://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design), were used for designing and predicting effectiveness of the sgRNA spacers (Chari et al., 2015; Doench et al., 2016). Spacers with an effectiveness score of at least 50% in both tools were selected and sgRNA's with predicted likely off-target

action were excluded by using the online Cas-OFFinder tool (http://www.rgenome.net/cas-offinder/, mismatch = 2, bulge size =1) and CRISPOR (http://crispor.tefor.net/crispor.py) (Bae et al., 2014; Concordet and Haeussler, 2018).

The MoCLo toolkit and Golden Gate cloning were used to assemble the vector (Engler et al., 2009; Weber et al., 2011). Briefly, each sgRNA was fused to a *Arabidopsis* U6 promoter as AtU6p:sgRNA:TTTT and ligated to a level 1 vector. Level 1 constructs plCH47732-NOSp::NPTII::OCST, plCH47742-35Sp::hCas9::NOST, plCH47751-U6p::sgRNA1, plCH47761-U6p::sgRNA2, plCH47772-U6p::sgRNA3, plCH47781-U6p::sgRNA4 and plCH41822-pLE6E were cut and ligated into the level 2 vector *plCSL4723* (Werner et al., 2012). The construct was transformed to *Agrobacterium tumefaciens* C58C1 and grown on 20 mg/L rifampicin, 20 mg/L gentamycin and 100 mg/L kanamycin selection plates.

TRANSFORMATION AND PLANT GROWTH

Transformation of tomato cv. Moneyberg was performed with *Agrobacterium tumefaciens* C59C1 as previously described but with media B supplemented with 1 mg/L 2,4D and with 200 explants as starting material (Van Roekel et al., 1993). Tissue culture was done in a growth chamber with 16 h light and 8 h dark at 25 °C. Rooted transformants were transferred to rockwool and acclimatized in a growth chamber. Alternatively, plants were sown on filter paper and transferred to cubes of rockwool after a week. Five to Eight weeks later seedlings were transplanted into a greenhouse (Unifarm, Wageningen 51.57° N, 5.31° E, The Netherlands) on rockwool slabs at a density of 2.5 plants*m^{-2}. Nutrients were provided by fertigation (EC 4.5, pH 5.6). Climatic conditions in the greenhouse were at ambient temperature (> 20 °C) under a 16h light/8h dark cycle (0.6–28.4 MJ m–2 day–1 natural light supplemented with artificial light using high pressure sodium lamps (SON-T Agro 600 Watt, Philips, Eindhoven, The Netherlands). Side shoots were removed once a week. Flowers were pollinated by vibrating each flower/truss three times a week with an electric toothbrush.

GENOTYPING

Genomic DNA from young leaves was isolated using the hexadecyl trimethylammonium bromide (CTAB) method (Porebski et al., 1997). Detection of transgenes in each generation was done by detection of *Cas9* and/or *NPTII* and for transformed plants the *CIF1* target region was Sanger sequenced using one of the PCR primers **(Table S3)**. Plants with heterozygous or bi-allelic mutations (T_0) in the region of interest were selfed. Segregation in T_1 plants for the presence of a homozygous or bi-allelic mutation in the region of interest, while lacking the T-DNA insertion, was selected in seedlings.

PHENOTYPING

T_0 plants were only used for genotyping and seed collection. In the T_1 generation, one or two plants per genotype were placed randomly on a row in the greenhouse. Seven flowers per truss were vibrated at anthesis, and other flowers were removed. The first truss was used for seed collection and the subsequent four trusses were used for phenotyping. The first four fruits of a truss were grown to the ripe stage for phenotyping. Individual fruits were harvested at breaker+7 (±1 day) for phenotyping. The three remaining fruits were harvested at 10 days (+1) after anthesis (DAA), and the columella was dissected and snap-frozen in liquid nitrogen for invertase activity analysis. In the T_2 generation, three plants per genotype were randomly placed on a row in the greenhouse. Six flowers per truss were vibrated with an electric toothbrush-holder at anthesis. Excess flowers were removed. The first truss was used for seed collection and the three subsequent trusses were used for phenotyping. Individual fruits were harvested at breaker+7 (±1 day) for phenotyping. Fruit width, weight and brix were measured. Brix measurements were done in duplicate per fruit with an Atago PR-32α digital refractometer.

INVERTASE ASSAY

0.3 g of homogenized cryogenic-ground 10 DAA columella T_1 samples was mixed with 0.45 mL extraction buffer (500 mM sodium-acetate (NaAc) buffer at pH 4.7, 100 mM sodium-bisulfite, 100 mM phenylmethylsulfonyl fluoride (PMSF), 340 µl/ml β-mercapto-ethanol) followed by centrifuging for 15 minutes at 4°C, 14,000 rpm. The pellet fraction was used for cell wall invertase activity. It was dissolved in 300 µl extraction buffer, washed using 100% ammoniumsulfate and incubated on ice for 45 minutes. After incubation on ice, samples were centrifuged for 10 minutes at 4°C, 14,000 rpm. Remaining pellet was washed again with 80% ammoniumsulfate and incubated on ice for 45 minutes. Centrifuging at 10 minutes at 4°C, 14,000 rpm yielded a pellet that was dissolved in 150 µl 50mM NaAc buffer. 40 µl of the resulting enzyme mix was then mixed with 40 µl of invertase substrate (500 mM NaAc buffer, 2% sodiumazide, 250mM sucrose). Incubation was done at 30°C. 20 µl samples were taken after 0, 30 and 60 minutes and inactivated at 96°C for 2-5 minutes. 80 µl of Milli-Q water was added and samples were stored at -20°C. Prior to measuring using high performance anion exchange chromatography (Dionex ICS-5000+ Detector/Chromatography Compartment) samples were cleaned using anion exchange buffer. The final dilution was 48.5x. A standard containing glucose, fructose and sucrose was used for sugar measurements. Two to three replicates per genotype were performed, with each replicate consisting of 1-3 fruits.

GENE EXPRESSION ANALYSIS

RNA was isolated from homogenized cryogenic-ground T_3 10 DAA whole fruit ground samples by using the MaqMax™-96 total RNA isolation kit with Plant RNA isolation aid (Thermo Fisher Scientific) with a KingFisher 96 Magnetic Particle Processor. cDNA was synthesized with the iScript cDNA synthesis kit (Bio-Rad). qRT-PCR was performed with iQ SYBR Green Supermix (Bio-Rad) on a Bio-Rad CFX6 qRT-PCR instrument. The following condition was applied for PCR amplification: 3 min 95˚C, 40*[15 s 95˚C, 60 s 60˚C], followed by a melt-curve. *Actin* was used as a refence (**Table S3**). Relative expression changes were calculated according to the $2^{-\Delta CT}$ method as described (Livak and Schmittgen, 2001). For each genotype, three plants were sampled. Each sample consisted of a pool of tree fruits.

RT-PCR was performed with a Q5 polymerase PCR and specific primers for *CIF1* (**Table S3**). The following condition was applied: 1 min 98°C, 37*[10 s 98°C, 20 s 61°C, 1 min 72°C], 5 min 72°C.

STATISTICAL ANALYSIS

Statistical analysis was performed with R, package version 3.5. For all data, normal distribution was confirmed, and ANOVAs were used to test for significant differences between mutants and wild type.

ACKNOWLEDGMENTS

We thank Jeroen van Arkel and Hanny Hakkert for their valuable help with the invertase assay and sugar measurements. We thank Geurt Versteeg, Teus van den Brink and Sean Geurts for their expertise and care for the tomato growth.

SUPPLEMENTARY TABLES

Supplementary Table 1: Off-targets of the four used sgRNAs up to 3 mismatches

Target	Off target Sequence	Mismatch Position	Mismatch count	Chromosome
1	ACGAAGAAAGAAACATTATTGGG	..**..*.....	3	SL2.40ch01
1	TCATAAAAAGAAACATTATTTAG	*....**.....	3	SL2.40ch01
1	ACATAGGAAGAAAAAGTATTGGG	......*.....*.*....	3	SL2.40ch01
1	ATAAAGCAAGAAATATTATTTGG	.*.*.....*.....	3	SL2.40ch04
1	GCATACCAAGAAATATTATTGGG	*....*.....*.....	3	SL2.40ch05
1	ATATCGCAAGAAACATAATTCGA	.*..*.....*...	3	SL2.40ch05
1	TCATAGCAAAAAAGATTATTCGG	*.....*...*.....	3	SL2.40ch05
1	ACATGGCAAGAAACATTATGGGG	....*.....*	2	SL2.40ch05
1	ACTGAGCAAAAAACATTATTTGG	..**....*.....	3	SL2.40ch07
1	TCATAGCAAGACACATTAATAGG	*.....*.....*.	3	SL2.40ch07
1	TCAAAGTAAGAAACATTATTAGA	*..*..*.....	3	SL2.40ch08
1	ACATAGAAAGAAACATTATAGAG	......*.....*	2	SL2.40ch09
2	AATATGTTGCTAGTAACAAATAG	**.....*	3	SL2.40ch09
3	GGATCCTTCCAAGATTGAAGCGG	.....*.....*....*.	3	SL2.40ch00
3	GGATCTTTTTAAGTTTGAGGTGG	.....**...*.....	3	SL2.40ch01
3	GGAACTTTCTAAACTTGAGGGGG	...*.....*..*.....	3	SL2.40ch04
3	GGATCTTATCAAGCTTTAGGAGG	.....**.....*...	3	SL2.40ch04
3	GGGTGATTCCAAGCTTGAGGAGG	..*.**.....	3	SL2.40ch05
3	GGATCCTTCCAAGATTGAGACGG	.....*.....*.....*	3	SL2.40ch06
3	GGAACTTTCTAAACTTGAGGAGG	...*.....*..*.....	3	SL2.40ch07
3	GGATCTTTCTCAACTTGAGGAGG	.....**.*.....	3	SL2.40ch07
3	GGAACTTGCTAAGCTTGAGGGGG	...*...*.*.....	3	SL2.40ch07
3	GGATCTTTCTAAGATTGAGACGG	.....*...*.....*	3	SL2.40ch07
3	GGATCCTTCCAAGATTGAAGCGG	.....*.....*....*.	3	SL2.40ch07
3	GGATGTCTCAAAGCTTGAGGGAG	....*.*..*.....	3	SL2.40ch08
3	GGATGTCTCAAAGCTTGAGGGAG	....*.*..*.....	3	SL2.40ch10
3	AGATGTTTCTAAGCTTGAGGGAG	*...*....*.....	3	SL2.40ch11
3	GGATCTTGCTAAACTTGAGGAGG	.....*.*..*.....	3	SL2.40ch12
3	GGATCTTATCAAGCTTTAGGAGG	.....**.....*...	3	SL2.40ch12
3	AGATGTTTCTAAGCTTGAGGGAG	*...*....*.....	3	SL2.40ch12
3	**GGATGTATCCAAGCTTGAGGAGG**	....*.*.....	2	SL2.40ch12
4	GAAAAGGCACAACCTTTCAAAGG	.....***.....	3	SL2.40ch01
4	GAAAACGCACAATCTTTCAAGGG	.....*.....**.....	3	SL2.40ch03
4	GAAAAGGCACAACTTTTAAACGG	.....*.*..*..	3	SL2.40ch05
4	CAAAACTCACAATTCTTCAAGGA	*....**.....	3	SL2.40ch05
4	GCATAGTCACAATTCTTCAAAAG	.*.*..*.....	3	SL2.40ch08
4	ACAAAGGCACAATGCTTCAATGG	**.....*.....	3	SL2.40ch09
4	GCAAAGTCACAATTCTTCAGGAG	.*...*.....*	3	SL2.40ch10
4	GAAAAATCACAATTCTTCAAAAG	.....**.....	2	SL2.40ch10
4	GAATAGGAACAACTCTTCAAAAG	...*...*....*.....	3	SL2.40ch12
4	GCAAAGGCAGAATTATTCAAAGG	.*.....*....*.....	3	SL2.40ch12
4	GGAAAGGCACAATTATTCCACGG	.*.....*...*.	3	SL2.40ch12

Target 1: ACATAGCAAGAAACATTATT; Target 2: GCTATGTTGCTAGTAACAAG: Target 3: GGATCTTTCCAAGCTTGAGG; Target 4: GAAAAGGCACAATTCTTCAA. The bold off-target is in *CIF1-like*.

Supplementary Table 2: T_2 phenotyping of knock out lines

Line	Average °Brix	°Brix increase	% °Brix difference	Average weight (g)	Weight (g) Increase	% weight difference
CIF1	4.87			99		
cif1-cr1	5.41	0.54	11.0	73	-26.7	-26.8
cif1-cr2	5.05	0.18	3.6	75	-24.6	-24.7
cif1-cr3	5.06	0.19	3.8	76	-23.7	-23.8
cif1-cr4	5.20	0.32	6.6	72	-27.9	-28.0

Supplementary Table 3: Primers used in this study

Description	Name	Sequence (5'-3')
Vector construction		
Fw primer used to amplify the sgRNA for pL1 assembly	sgRNA1_F	TGTGGTCTCAATTGACATAGCAAGAAACATTATT GTTTTAGAGCTAGAAATAGCAAG
Fw primer used to amplify the sgRNA for pL1 assembly	sgRNA2_F	TGTGGTCTCAATTGGCTATGTTGCTAGTAACAA GGTTTTAGAGCTAGAAATAGCAAG
Fw primer used to amplify the sgRNA for pL1 assembly	sgRNA3_F	TGTGGTCTCAATTGGGATCTTTCCAAGCTTGAG GGTTTTAGAGCTAGAAATAGCAAG
Fw primer used to amplify the sgRNA for pL1 assembly	sgRNA4_F	TGTGGTCTCAATTGGAAAAGGCACAATTCTTCAA GTTTTAGAGCTAGAAATAGCAAG
Universal Rv primer used to amplify sgRNAs for pL1 assembly	sgRNA_R	TGTGGTCTCAAGCGTAATGCCAACTTTGTAC
Validating pL2 construction	pL2_F	TCATCAGTCAATTACGGGGCT
	pL2_R	CGCACGGCTGGCACATACAA
Genotyping		
Genotyping *NPTII*	NPTII_F	AGACAATCGGCTGCTCTGAT
	NPTII_R	AGCCAACGCTATGTCCTGAT
Genotyping *Cas9*	Cas9_F	CTGGCCAATGGAGAGATTCG
	Cas9_R	GCTCCCTGATGGGCTTATCC
Genotyping *CIF1* (Solyc12g099200)	CIF1_F	CCCACCGAAAACACAAAGCA
	CIF1_R	GCAACCAGTAAGATAGGGTCG
Genotyping *CIF1-Like* (Solyc12g099210)	CIF1-Like_F	ATCCAAATTCATTGTTATCC
	CIF1-Like_R	TATTTATTAAGAGAGGCTTATTG
qRT-PCR		
qRT-PCR *Actin* (Solyc11g005330)	Actin_F	TGTCCCTATTTACGAGGGTTATGC
	Actin_R	AGTTAAATCACGACCAGCAAGAT
qRT-PCR *CIF1* (Solyc12g099200)	CIF1_F	GTTGGTAGAGCCATTGTAAGAAATTT
	CIF1_R	TGATCATAATGTGACGAATCGAAT
qRT-PCR *CIF1-Like* (Solyc12g099210)	CIF1-Like_F	CTGGAGATGCACAAGAATGTGA
	CIF1-Like_R	TATTAAGAGAGGCTTATTGATTCAC
qRT-PCR *CIN2* (Solyc01g058020)	CIN2_F	GGCAGTGACCCTAAGAATACG
	CIN2_R	AGCCAAATCCACCGCTTT
qRT-PCR *CIN3* (Solyc01g100810)	CIN3_F	TGGTCCATTGTGTCGTCTT
	CIN3_R	GGTGTGTTCTTAGGGTCGCT
qRT-PCR *CIN4* (Solyc01g111100)	CIN4_F	CGTAGAATGGGTGTGTATGG
	CIN4_R	GAACTTGTTGACTGCTGTGTG
qRT-PCR *CIN5* (Solyc04g081440)	CIN5_F	AGATGTGCCTTGTTTCTCCT
	CIN5_R	CCGATGAAGTAACCACCAC

Supplementary Table 3 continued

Description	Name	Sequence (5'-3')
qRT-PCR *deCWIN* (Solyc03g121680)	deCWIN_F	AGTATGAGAGGATGGGCTGG
	deCWIN_R	TCTGCCTGTGTAGCATTGAC
qRT-PCR *HSP17.7* (Solyc06g076540)	HSP17.7_F	ACTTGGCATCGTGTGGAAC
	HSP17.7_R	ACTTGACATCAGGCTTCTTCAC
qRT-PCR *HT2* (Solyc09g075820)	HT2_Fw	TCAACTACGGAACAGCCAAG
	HT2_Rv	TCAGGTTCAATGTTGTCGGT
qRT-PCR *LIN5* (Solyc09g010080)	LIN5_Fw	TGGGGTTGGTCAAATGAATCCG
	LIN5_Rv	GAATACCTTGAATTCCAGCCCATCC
qRT-PCR *LIN9* (Solyc08g079080)	LIN9_Fw	ACTGGGTCAACCAACGAATC
	LIN9_Rv	TGCCCTCATACTTGATCCAT
qRT-PCR *SWEET7a* (Solyc08g082770)	SWEET7_F	TGATGCCTACATTCTCGCACC
	SWEET7_R	TCCTTTAGCCTCTCTTGCTGCC
qRT-PCR SWEET12c (Solyc05g024260)	SWEET12c_F	GCATCGTGTTTCAAGTGGTTCG
	SWEET12c_R	TCTATCGCTGGCTTTGCGTT
qRT-PCR SWEET14 (Solyc03g097560)	SWEET14_F	GCCAAAGGCAATCATAGAGG
	SWEET14_R	AGGCACACACAATCAGACCT
RT-PCR		
RT-PCR *CIF1*	CIF1_Fw	CTAATAATGTTTCTTGCTATGTTGC
RT-PCR cif1-cr3	$CIF1^{cif1\text{-}cr3}$_Fw	TCCTAATATTTCTTGCTATGTTGCT
RT-PCR *CIF1*	CIF1_Rv	AGAGGGGATACACACATAACATT

F: Forward, R: Reverse, pL1: level 1 plasmid Golden Gate, pL2: level 2 plasmid Golden Gate. Genes in this study can be found in the Sol Genomics Network website (https://solgenomics.net/) with the listed accession numbers

REFERENCES

Alberto, F., Bignon, C., Sulzenbacher, G., Henrissat, B., and Czjzek, M. (2004). The Three-dimensional Structure of Invertase (β-Fructosidase) from *Thermotoga maritima* Reveals a Bimodular Arrangement and an Evolutionary Relationship between Retaining and Inverting Glycosidases. J. Biol. Chem. **279**: 18903–18910.

Bae, S., Park, J., and Kim, J.-S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics **30**: 1473–1475.

Baldwin, E.A., Scott, J.W., Shewmaker, C.K., and Schuch, W. (2000). Flavour trivia and tomato aroma: Biochemistry and possible mechanisms for control of important aroma components. HortScience **35**: 1013–1022.

Baxter, C.J., Carrari, F., Bauke, A., Overy, S., Hill, S.A., Quick, P.W., Fernie, A.R., and Sweetlove, L.J. (2005). Fruit carbohydrate metabolism in an introgression line of tomato with increased fruit soluble solids. Plant Cell Physiol. **46**: 425–437.

Chari, R., Mali, P., Moosburner, M., and Church, G.M. (2015). Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. Nat. Methods **12**.

Choi, Y. (2012). A fast computation of pairwise sequence alignment scores between a protein and a set of single-locus variants of another protein. In 2012 ACM Conference on Bioinformatics, Computational Biology and Biomedicine, BCB 2012 (ACM Press: New York, New York, USA), pp. 414–417.

Concordet, J.P. and Haeussler, M. (2018). CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res. **46**: W242–W245.

Datir, S. and Ghosh, P. (2020a). In silico analysis of the structural diversity and interactions between invertases and invertase inhibitors from potato (*Solanum tuberosum* L.). 3 Biotech **10**: 178.

Datir, S. and Ghosh, P. (2020b). In silico analysis of the structural diversity and interactions between invertases and invertase inhibitors from potato (*Solanum tuberosum* L.). 3 Biotech **10**: 178.

Datir, S.S. (2020). Invertase inhibitors in potato: towards a biochemical and molecular understanding of cold-induced sweetening. Crit. Rev. Food Sci. Nutr.

Doench, J.G. et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. **34**: 184–191.

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes. PLoS One **4**: e5553.

Eshed, Y. and Zamir, D. (1996). Less-than-additive epistatic interactions of quantitative trait loci in tomato. Genetics **143**: 1807–1817.

Fernie, A.R., Bachem, C.W.B., Helariutta, Y., Neuhaus, H.E., Prat, S., Ruan, Y.L., Stitt, M., Sweetlove, L.J., Tegeder, M., Wahl, V., Sonnewald, S., and Sonnewald, U. (2020). Synchronization of developmental, molecular and metabolic aspects of source–sink interactions. Nat. Plants **6**: 55–66.

Fridman, E., Carrari, F., Liu, Y.-S., Fernie, A.R., and Zamir, D. (2004). Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science **305**: 1786–1789.

Fridman, E., Liu, Y.S., Carmel-Goren, L., Gur, A., Shoresh, M., Pleban, T., Eshed, Y., and Zamir, D. (2002). Two tightly linked QTLs modify tomato sugar content via different physiological pathways. Mol. Genet. Genomics **266**: 821–826.

Fridman, E., Pleban, T., and Zamir, D. (2000). A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc. Natl. Acad. Sci. U. S. A. **97**: 4718–4723.

Fridman, E. and Zamir, D. (2003). Functional divergence of a syntenic invertase gene family in tomato, potato, and *Arabidopsis*. Plant Physiol. **131**: 603–9.

Godt, D.E. and Roitsch, T. (1997). Regulation and tissue-specific distribution of mRNAs for three extracellular invertase isoenzymes of tomato suggests an important function in establishing and maintaining sink metabolism. Plant Physiol. **115**: 273–282.

Hothorn, M., D'Angelo, I., Márquez, J.A., Greiner, S., and Scheffzek, K. (2004a). The Invertase Inhibitor Nt-CIF from Tobacco: A Highly Thermostable Four-helix Bundle with an Unusual N-terminal Extension. J. Mol. Biol. **335**: 987–995.

Hothorn, M., Van Den Endec, W., Lammens, W., Rybin, V., and Scheffzek, K. (2010). Structural insights into the pH-controlled targeting of plant cell-wall invertase by a specific inhibitor protein. Proc. Natl. Acad. Sci. U. S. A. **107**: 17427–17432.

Hothorn, M., Wolf, S., Aloy, P., Greiner, S., and Scheffzek, K. (2004b). Structural insights into the target specificity of plant invertase and pectin methylesterase inhibitory proteins. Plant Cell **16**: 3437–3447.

Jin, Y., Ni, D.-A., and Ruan, Y.-L.Y.-L. (2009). Posttranslational Elevation of Cell Wall Invertase Activity by Silencing Its Inhibitor in Tomato Delays Leaf Senescence and Increases Seed Weight and Fruit Hexose Level. Plant Cell **21**: 2072–89.

Klann, E., Yelle, S., and Bennett, A.B. (1992). Tomato Fruit Acid Invertase Complementary DNA : Nucleotide

and Deduced Amino Acid Sequences. PLANT Physiol. **99**: 351–353.

Klee, H.J. and Giovannoni, J.J. (2011). Genetics and Control of Tomato Fruit Ripening and Quality Attributes. Annu. Rev. Genet. **45**: 41–59.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. **35**: 1547–1549.

Lammens, W., Le Roy, K., Schroeven, L., Van Laere, A., Rabijns, A., and Van Den Ende, W. (2009). Structural insights into glycoside hydrolase family 32 and 68 enzymes: Functional implications. J. Exp. Bot. **60**: 727–740.

Link, M., Rausch, T., and Greiner, S. (2004). In Arabidopsis thaliana, the invertase inhibitors AtC/VIF1 and 2 exhibit distinct target enzyme specificities and expression profiles.

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.

Lykke-Andersen, S. and Jensen, T.H. (2015). Nonsense-mediated mRNA decay: An intricate machinery that shapes transcriptomes. Nat. Rev. Mol. Cell Biol. **16**: 665–677.

Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., and Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. **47**: W636–W641.

Matsukura, C. (2016). Sugar accumulation in tomato fruit and its modification using molecular breeding techniques. In Biotechnology in Agriculture and Forestry (Springer International Publishing), pp. 141–154.

Merchante, C., Stepanova, A.N., and Alonso, J.M. (2017). Translation regulation in plants: an interesting past, an exciting present and a promising future. Plant J. **90**: 628–653.

Mou, H. et al. (2017). CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. Genome Biol. **18**: 108.

N' tchobo, H., Dali, N., Nguyen Quoc, B., Foyer, C.H., and Yelle, S. (1999). Starch synthesis in tomato remains constant throughout fruit development and is dependent on sucrose supply and sucrose synthase activity. J. Exp. Bot. **50**: 1457–1463.

Nguyen-Quoc, B. and Foyer, C.H. (2001). A role for "futile cycles" involving invertase and sucrose synthase in sucrose metabolism of tomato fruit. J. Exp. Bot. **52**: 881–889.

Palmer, W.M.M., Ru, L., Jin, Y., Patrick, J.W.W., and Ruan, Y.-L. (2015). Tomato Ovary-to-Fruit Transition is Characterized by a Spatial Shift of mRNAs for Cell Wall Invertase and its Inhibitor with the Encoded Proteins Localized to Sieve Elements. Mol. Plant **8**: 315–328.

Petró-Turza, M. (1986). Flavour of tomato and tomato products. Food Rev. Int. **2**: 309–351.

Porebski, S., Bailey, L.G., and Baum, B.R. (1997). Modification of a CTAB DNA Extraction Protocol for Plants Containing High Polysaccharide and Polyphenol Components. Plant Mol. Biol. Report. **15**: 8–15.

Proels, R.K., Hause, B., Berger, S., and Roitsch, T. (2003). Novel mode of hormone induction of tandem tomato invertase genes in floral tissues. Plant Mol. Biol. **52**: 191–201.

Rausch, T. and Greiner, S. (2004). Plant protein inhibitors of invertases. Biochim. Biophys. Acta - Proteins Proteomics **1696**: 253–261.

Rodriguez-Rodriguez, J.A., Lewis, C., McKinley, K.L., Sikirzhytski, V., Corona, J., Maciejowski, J., Khodjakov, A., Cheeseman, I.M., and Jallepalli, P. V. (2018). Distinct Roles of RZZ and Bub1-KNL1 in Mitotic Checkpoint Signaling and Kinetochore Expansion. Curr. Biol. **28**: 3422-3429.e5.

Van Roekel, J.S.C., Damm, B., Melchers, L.S., Hoekema, A., and Davey, M.R. (1993). Factors influencing transformation frequency of tomato (*Lycopersicon esculentum*). Plant Cell Rep. **12**: 644–647.

Roohanitaziani, R. (2019). Genetic analysis of fruit quality in tomato (Wageningen University, Doctoral dissertation).

Le Roy, K., Vergauwen, R., Struyf, T., Yuan, S., Lammens, W., Mátrai, J., De Maeyer, M., and Van den Ende, W. (2013). Understanding the role of defective invertases in plants: Tobacco nin88 fails to degrade sucrose. Plant Physiol. **161**: 1670–1681.

Ru, L., He, Y., Zhu, Z., Patrick, J.W., and Ruan, Y.-L. (2020). Integrating Sugar Metabolism With Transport: Elevation of Endogenous Cell Wall Invertase Activity Up-Regulates *SlHT2* and *SlSWEET12c* Expression for Early Fruit Development in Tomato. Front. Genet. **11**.

Ru, L., Osorio, S., Wang, L., Fernie, A.R., Patrick, J.W., and Ruan, Y.-L. (2017). Transcriptomic and metabolomics responses to elevated cell wall invertase activity during tomato fruit set. J. Exp. Bot. **68**: 4263–4279.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**: 406–425.

Saurabh, S., Vidyarthi, A.S., and Prasad, D. (2014). RNA interference: Concept to reality in crop improvement. Planta **239**: 543–564.

Shen, S., Ma, S., Liu, Y., Liao, S., Li, J., Wu, L., Kartika, D., Mock, H.P., and Ruan, Y.L. (2019). Cell wall invertase and sugar transporters are differentially activated in tomato styles and ovaries during pollination and fertilization. Front. Plant Sci. **10**: 506.

Slugina, M.A., Dzhos, E.A., Shchennikova, A. V, and Kochieva, E.Z. (2019). Up-regulation of the vacuolar invertase *TA1* gene may contribute to the accumulation of carotenoids in tomato fruits. Curr. Challenges Plant Genet. Genomics, Bioinformatics, Biotechnol. **24**: 146–149.

Stecher, G., Tamura, K., and Kumar, S. (2020). Molecular evolutionary genetics analysis (MEGA) for macOS. Mol. Biol. Evol. **37**: 1237–1239.

Syed, N.H., Kalyna, M., Marquez, Y., Barta, A., and Brown, J.W.S. (2012). Alternative splicing in plants - coming of age. Trends Plant Sci. **17**: 616–623.

Tieman, D. et al. (2012). The chemical interactions underlying tomato flavour preferences. Curr. Biol. **22**: 1035–1039.

Tieman, D. et al. (2017). A chemical genetic roadmap to improved tomato flavour. Science (80-.). **355**: 391–394.

Wang, R., Angenent, G.C., Seymour, G., and de Maagd, R.A. (2020). Revisiting the Role of Master Regulators in Tomato Ripening. Trends Plant Sci. **25**: 291–301.

Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. PLoS One **6**: e16765.

Wei, H., Chai, S., Ru, L., Pan, L., Cheng, Y., Ruan, M., Ye, Q., Wang, R., Yao, Z., Zhou, G., Chen, Y., and Wan, H. (2020). New insights into the evolution and expression dynamics of invertase gene family in *Solanum lycopersicum*. Plant Growth Regul.: 1–13.

Werner, S., Engler, C., Weber, E., Gruetzner, R., and Marillonnet, S. (2012). Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. Bioengineered **3**: 38–43.

Winsor, G.W. (1966). Some factors affecting the composition, flavour and firmness of tomatoes. Sci. Hortic. **18**: 27–35.

Yadav, U.P., Ayre, B.G., and Bush, D.R. (2015). Transgenic approaches to altering carbon and nitrogen partitioning in whole plants: Assessing the potential to improve crop yields and nutritional quality. Front. Plant Sci. **6**: 275.

Zanor, M.I. et al. (2009). RNA Interference of LIN5 in Tomato Confirms Its Role in Controlling Brix Content, Uncovers the Influence of Sugars on the Levels of Fruit Hormones, and Demonstrates the Importance of Sucrose Cleavage for Normal Fruit Development and Fertility. Plant Physiol. **150**: 1204–1218.

Zhang, C. and Turgeon, R. (2018). Mechanisms of phloem loading. Curr. Opin. Plant Biol. **43**: 71–75.

Zhang, J., Wu, Z., Hu, F., Liu, L., Huang, X., Zhao, J., and Wang, H. (2018a). Aberrant seed development in Litchi chinensis is associated with the impaired expression of cell wall invertase genes. Hortic. Res. **5**: 39.

Zhang, N., Shi, J., Zhao, H., and Jiang, J. (2018b). Activation of small heat shock protein (SlHSP17.7) gene by cell wall invertase inhibitor (SlCIF1) gene involved in sugar metabolism in tomato. Gene **679**: 90–99.

Zhang, N., Zhao, H., Shi, J., Wu, Y., and Jiang, J. (2020). Functional characterization of class I SlHSP17.7 gene responsible for tomato cold-stress tolerance. Plant Sci. **298**: 110568.

Zuckerlandl, E. and Pauling, L. (1965). Evolutionary Divergence and Convergence in Proteins. In Evolving Genes and Proteins (Elsevier), pp. 97–166.

CHAPTER 6

Expanding the CRISPR/Cas9-toolbox with Gene Targeting in tomato

Vera Veltkamp[1,2], Giorgio Gullotta[1,2], Jeroen Busscher[3], Michiel Lammers[1,2], William de Martines[4], Ellen Slaman[1,2], Xiaobing Jiang[1,2], Marian Bemer[2], Gerco C. Angenent[1,2], and Ruud A. de Maagd[2]

[1]Laboratory of Molecular Biology, Wageningen University, The Netherlands

[2]Bioscience, Wageningen Plant Research, The Netherlands

[3]Laboratory of Plant Physiology, Wageningen University, The Netherlands

[4]Plant Breeding, Wageningen University, The Netherlands

ABSTRACT

In this chapter we investigated a gene targeting approach for modification of LIN5, the tomato gene encoding a cell wall invertase involved in sugar accumulation. The goal was to substitute the three amino acids that are most likely underlying the higher LIN5 activity in the *Brix9-2-5* quantitative trait locus (QTL). We created a construct for expressing *SpCas9* and multiple sgRNAs each targeting one amino acid and creating a double stranded break at each site. A geminiviral replicon system was used for the simultaneous production of the donor template for homologous repair. In protoplast assays we obtained gene targeting rates between 0.2 and 1.5%, while a stable transformation led to one plant with gene targeting events in two targets and an indel in the third target. The use of three sgRNAs led to high rates of indels in combination with the intended amino acid substitutions. Because of the low rates of HR and high rates indel formation, no stable transformant was obtained with all three intended amino acids that would lead to a higher Brix. Our conclusion is that successful Gene Targeting with the geminiviral replicon system, Cas9, and multiple sgRNAs in transgenic tomato is feasible, but needs to be optimized to achieve higher frequencies.

INTRODUCTION

In the previous chapters we have used CRISPR/Cas9 mutagenesis to apply variation in the *LIN5* and *AGPL1* promoters, respectively. Our goal was to modulate *Cis*-regulatory Elements (CREs), to create lines that had increased expression of our target genes. We hypothesized that by removing CREs, expression could be modified locally without severe pleiotropic effects. Instead of increasing expression of the gene, an approach targeting the protein for higher enzymatic activity could also give the desired phenotype. For the *Brix9-2-5* QTL, increased Brix was attributed to increased enzymatic activity, not increased expression of LIN5. Three amino acid substitutions, Glu348Asp, Asn366Asp and Leu373Val, in the *S. pennellii* LIN5 protein compared to *S. lycopersicum* were linked to the increased Brix phenotype (Fridman et al., 2000, 2004). It was hypothesized that the Glu to Asp substitution at position 348 would be the most important, as it is closest to the active site of the invertase. The conservation of this residue in other invertase proteins further supports an essential role. The ASN to ASP substitution at position 366 was also correlated with an increase in sugar when constitutively expressed (Tieman et al., 2017).

In planta modification of the involved amino acids would be an ideal method to study the relevance of each or all amino acids. The rapid development of CRISPR/Cas tools inspired us to re-create the three amino acid substitutions by gene targeting. Effects of all three or single amino acid substitutions could then be validated and studied in any desired cultivar. Additionally, the use of gene targeting would allow us to create a desired genotype in two generations without the need of back-crossing and

introgression lines. The first generation would contain the CRISPR/Cas gene targeting construct and the desired substitution. In the second generation, the CRISPR/Cas-construct could be segregated out, leaving a modified heterozygous or homozygous *LIN5* allele in the absence of any foreign DNA.

Gene targeting is defined as the modification of a gene using Homologous Recombination (HR) (Paszkowski et al., 1988), and efficient HR has been a long-coveted goal in plant genetics. Past attempts with *Agrobacterium*-mediated transformation have been successful, but with low success rates of approximately one targeting event per 10^4 to 10^5 transformation events using different transformation methods and plant species (Paszkowski et al., 1988; Offringa et al., 1992). HR can occur after a double stranded break (DSB) in the DNA. The preferred repair mechanisms of a DSB in plants is Non-homologous end joining (NHEJ), which can be imprecise and result in small indels or larger deletions (Shukla et al., 2009; Symington and Gautier, 2011; Deriano and Roth, 2013; Knoll et al., 2014; Schiml et al., 2014; Čermák et al., 2015; van Overbeek et al., 2016; Tan et al., 2020). With HR, available homologous DNA is used as a repair template. This can be in the form of a sister chromatid, a tandem duplication or a piece of exogenously supplied donor template (Puchta, 2004; Symington and Gautier, 2011). The donor template can contain anything from small modifications in an otherwise identical whole gene or the insertion of entirely new sequence between two arms with homology to the target (D'Halluin et al., 2013; Čermák et al., 2015).

Sequence-specific nucleases such as meganucleases, Zinc-finger nucleases, and most recently Cas nucleases have given us the possibility to generate DSBs at a predefined target site (Shukla et al., 2009; Townsend et al., 2009; Wang et al., 2012; Jinek et al., 2012; D'Halluin et al., 2013). This allows the control of the exact location of intended HR (Puchta and Fauser, 2014). The amount of donor template able to reach the nucleus and the DSB site has always been a bottleneck for achieving high rates of HR. Recently, a promising approach using geminivirus-based DNA replicons achieved a high frequency HR without the use of a selection marker for insertion in tomato (Dahan-Meir et al., 2018). In this system, replication initiator protein (Rep) generates a high amount of a circular dsDNA fragment, the replicon. Rep uses a Long Intergenic Region (LIR) as the origin of replication and Short Intergenic Regions (SIR) as the origin of replication for the complementary and virion-sense transcripts, as well as transcription termination and polyadenylation (Fondong, 2013). Between the SIR and LIR, a fragment of interest, e.g., a donor sequence for HR, can be inserted. Upon transformation to a plant, Rep produces a large amount of the donor template in the replicon. This replicon can then serve as the donor template for HR at a CRISPR/Cas generated DSB site.

We used this replicon system to produce the template for modification of the three amino acids of interest in LIN5. For this we used a multiplexed setup, where each amino acid substitution was targeted by one single guide RNA (sgRNA) and a single donor

template that contained the substitutions for all three amino acids. The system was tested in a stable transformation and in a high-throughput protoplast assay.

RESULTS AND DISCUSSION

GENE TARGETING WITH THE GEMINIVIRAL-REPLICON DONOR DELIVERY SYSTEM IS NOT EFFICIENT IN STABLE TRANSFORMATION

Our goal was to induce three targeted nucleotide substitutions in *LIN5* leading to three amino acid substitutions in the protein (**Figure 1a, b**). These amino acid substitutions were chosen as they are thought to be the cause of a higher Brix phenotype in the *S. pennellii* -introgression *Brix-9-2-5* (Fridman et al., 2004; Zanor et al., 2009). We used the effective CRISPR/Cas9 mutagenesis system to create a Double Stranded Break (DSB) to initiate Homologous Recombination (HR) with a provided template. Four single guide RNA (sgRNAs) were designed to target the three amino acids of interest (**Figure 1a**). SgRNA1a or sgRNA1b targeted the Glu348Asp substitution, sgRNA2 targeted the Asn366Asp substitution and sgRNA3 targets the Leu373Val substitution. The donor template was designed to have the three necessary nucleotide substitutions that would result in the desired amino acids (**Figure 1b**). In all cases, the predicted Cas9 DSB site (-3 from the Protospacer-Adjacent Motif; PAM) were chosen to be as close as possible to the intended substitution, at least within the "seed" region (10-12 nucleotides from the PAM). The proximity of the substituted nucleotide to the PAM, should abolish sgRNA binding activity on the target site in case of a successful HR event (Semenova et al., 2011; Wiedenheft et al., 2011; Pattanayak et al., 2013; Jiang and Doudna, 2017). The donor template contained one additional nucleotide substitution that removed a *BpiI*-site, which interfered with cloning (**Figure 1a**).

A left (969 bp) and right (792 bp) homology arm was added to make the final donor template (**Figure 2a)**. Large amounts of donor template were generated by using the bean yellow dwarf virus geminiviral rolling circle replicon (**Figure 2b**) (Dahan-Meir et al., 2018). The system was cloned in both a GreenGate vector as well as in a Golden Gate vector (**Figure 2c**) (Weber et al., 2011; Lampropoulos et al., 2013; Engler et al., 2014). The T-DNA constructs contained the CRISPR/Cas9 system with three different *sgRNAs*, *aCas9* expressed under transcriptional control of the *Petroselinum crispum* (parsley) *PcUbi4* promoter, and the replicon machinery consisting of *replication initiator protein (Rep)* and two *Short Intergenic Repeats (SIR)* harbouring the donor template and a *Long Intergenic Repeat (LIR).* For target 1, two sgRNAs were possible and one was tested in the GreenGate system (1a) and one in the Golden Gate system (1b). A Golden Gate control vector was made without the *pcUbi4:aCas9* ("No Cas9"). NPTII and GFP were used as selection and screening markers for transformants.

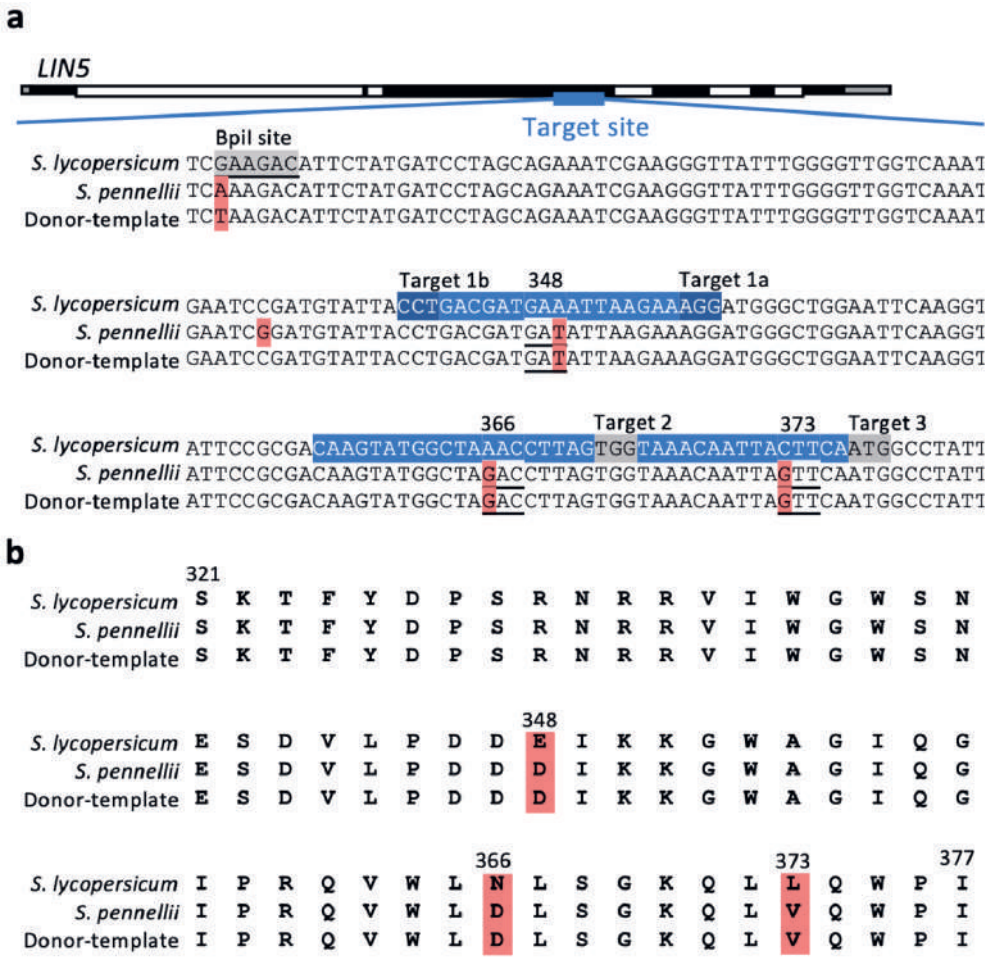


Figure 1: **(a)** A close-up of the third *LIN5* exon, aligned with *S. pennellii* and the eventual donor template. Differential nucleotides are highlighted red and underlined. Three different guides were identified in the regions of interest, highlighted in blue. The PAM regions are highlighted grey, as well as a *BpI* site in *S. lycopersicum*. Target 1a: GreenGate. Target 1b: Golden Gate **(b)** the resulting amino acid alignment of the same region. Mismatched amino acids are highlighted.

First, we performed a stable transformation with the GreenGate vector. There, the transformation efficiency was extremely low. Out of 375 co-cultured explants, only 35 showed a GFP signal (9% transformation efficiency) three weeks after transformation. From these explants, we obtained seven rooted, diploid primary transformants (T_0). Five of these (*lin5-cr-01, -03, -04, -06* and -07) produced the viral replicon DNA as was shown by PCR and a qPCR, both with primers specific for the circular replicon (**Figure 3a, 3b**). The qPCR indicated a ~150-fold increase in copy number of the replicon. Thus, the supply of abundant donor template was successful. Six plants had indels at target 2 and target 3, but none had the intended Gene Targeting substitutions (**Figure 3c**).

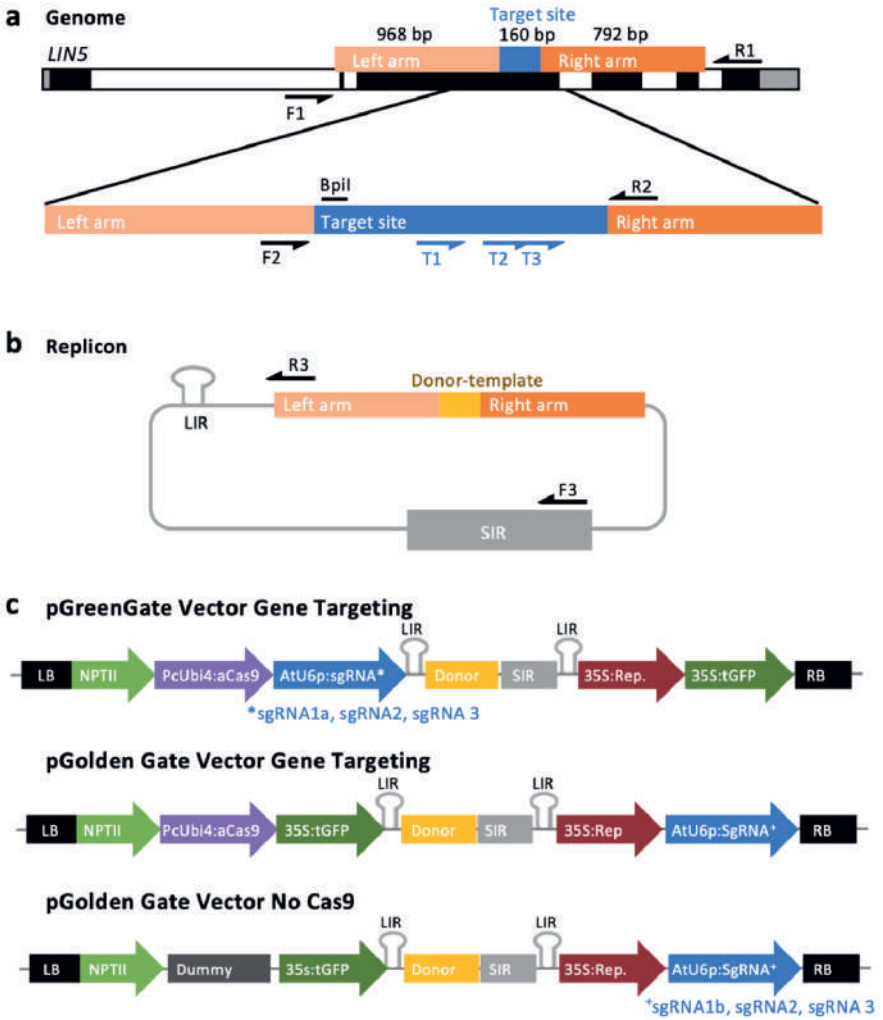


Figure 2: (a) Diagrams showing the genomic region of the *LIN5* gene. A DSB is induced three times in the target-region by the CRISPR/Cas9 system. Primers for genotyping are represented by half arrows, as well as the region of the sgRNA targets. **(b)** Recombination in the regions of homology between the donor and the broken target results in the replacement of the deletion allele by the modified donor template. The donor template is generated by in the geminiviral replicon **(c)** Plants were transformed with a single construct containing Cas9 under the expression of the *PcUbiquitin4* (*Ubi4*) promoter, three sgRNA (*AtU6p:t1a/b, :t2, :t3*), Rep, the modified donor template and short intergenic sequence (SIR) in the geminiviral replicon defined by the large intergenic region (LIR) sequences. Kanamycin resistance for the *NPTII* gene and turboGFP (*tGFP*) were used as transformant selection markers. The control construct has the geminiviral replicon and sgRNAs but no Cas9 endonuclease.

An identical 6-bp deletion in target 2 in all mutants is possibly a result of preferred microhomology-mediated end-joining at that position (Deriano and Roth, 2013; van Overbeek et al., 2016; Tan et al., 2020). Interestingly, all mutants appeared to be homozygous, which if not coincidental, could indicate repair of a DSB by HR with an already mutated sister chromatid as template.

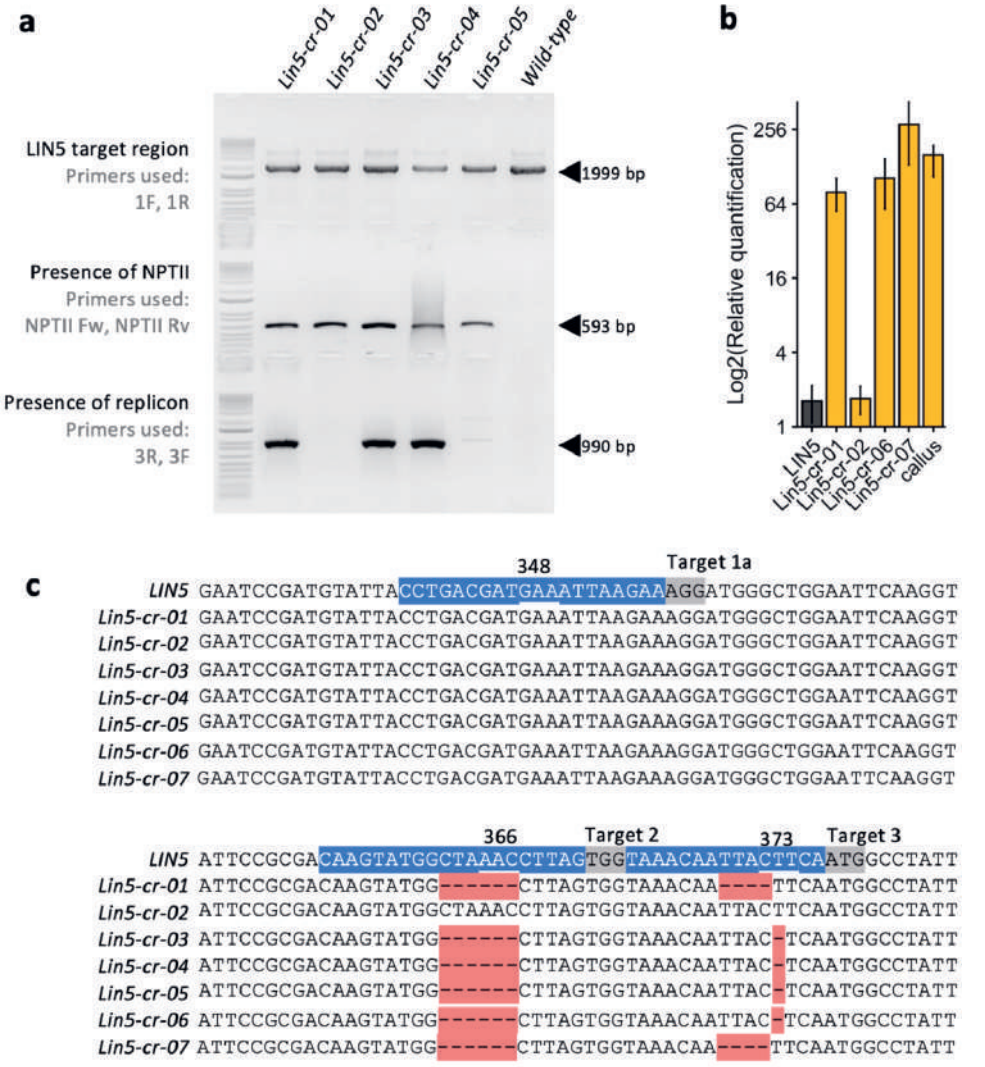


Figure 3: GreenGate vector transformation **(a)** Electrophoresis gel showing the PCR products from genotyping. Genomic DNA was used, obtained from leaf tissue **(b)** Relative quantification of the circular replicon from genomic DNA. Primers 4R and 4F were used for the replicon and compared to a primer pair on the *LIN5* promoter (**Table S2**). Two replicates of a non-transformed (LIN5), T_0 plants and pieces of callus from tissue culture. Bars represent the standard error of the mean of the replicates. Genomic DNA was used, obtained from leaf tissue. **(c)** Alignments of the wild-type (*LIN5*) and mutant coding sequences. sgRNA target regions are highlighted with a grey PAM. Mutations are highlighted.

We used 1017 explants for a stable transformation with the Golden Gate vector. This vector backbone proved much more efficient in transformation, with 99% of the explants developing GFP positive callus (**Figure 4a**). 212 transformed shoots were genotyped, of which 125 produced the replicon (**Figure 4b, c**). Out of these 125, the majority (105) had indels in one or more of the targets. Only one shoot had a heterozygous Gene Targeting event at both target 1 and target 2.

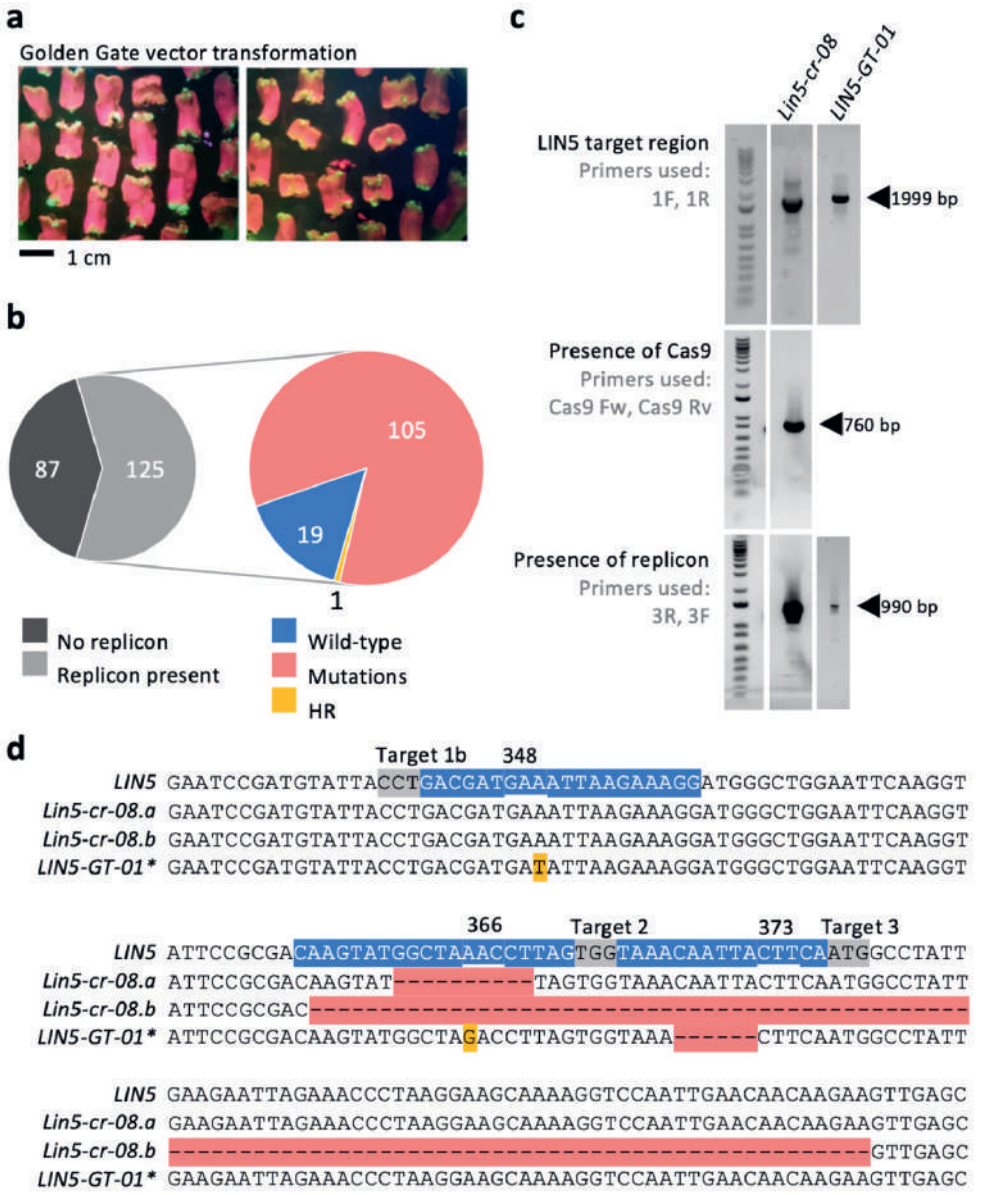


Figure 4: Golden Gate vector Transformation **(a)** Two representative pictures of explants with early callus development during tissue culture, three weeks after cocultivation with the Golden Gate construct. Picture is taken with an UV-torch. The transformation is highly efficient, with 99% of explants showing GFP-positive callus development. **(b)** Representation of the number of shoots genotyped and either or not containing the replicon. Of the plants that contained the replicon, the number of shoots that had a wild-type genotype, displayed indels or evidence of homologous recombination (HR), are shown. **(c)** Electrophoresis gel showing the PCR products of genotyping. **(d)** Alignments of the wild-type (*LIN5*) and mutant coding sequences. sgRNA target regions are highlighted with a grey PAM. Mutations are highlighted. The two alleles (a, b) are shown for *lin5-cr-08*.
*LIN5-GT-01 was bi-allelic or heterozygous mutant/wild-type. However, we did not resolve the individual alleles. For each target, one strand had a wild-type sequence, but the linkage between the alleles was not determined. See **Figure S1** for the trace-alignment.

However, an indel was also present in target 3, resulting in a disruption of the *LIN5* open reading frame (**Figure 4d, Figure S1**). In addition, this plant failed to root and could not be propagated. From the stable transformation experiment, we concluded that the sgRNAs were highly effective in creating DSBs and that the replicon was produced. However, as expected, NHEJ was the dominant repair mode, and a much higher number of shoots would have to be screened to detect true Gene Targeting events.

Several mutant plants from the GreenGate vector transformation (*lin5-cr-01, -03, -04*), and one (*-08*) from the Golden Gate vector transformation were grown for propagation. These all produced the replicon and had homozygous or bi-allelic indels disrupting the *LIN5* open reading frame (**Figure 4c, d**). All mutants had fertility issues. The GreenGate plants (*lin5-cr-01, -03, -04*) showed aberrant stigma development and no or little pollen (**Figure S2**). These plants either gave small, seedless fruit or no fruit-set at all. *Lin5-cr-08* did produce small fruit, but with only a few seeds. As all phenotypes were observed in the primary transformants, there is a high chance of a tissue culture effect. However, the observed *lin5* knockout phenotype is in line with previous studies. *LIN5* is known to have a role in early fruit set and development and is expressed in the reproductive organs and in early fruit development (4 hour till 5 days after pollination) (Shen et al., 2019). *LIN5* silencing by RNA interference (RNAi) led to a lower Brix, smaller fruit, fewer seeds, and increased fruit abortion (Zanor et al., 2009). In other plant species a critical role for cell wall invertases has been found as well (Wan et al., 2018; Liao et al., 2020). We obtained a few seeds from a cross between the GreenGate vector mutants and wild-type for phenotyping in the Back-Crossed generation. Combined with the few seeds from *lin5-cr-08*, the true phenotype of a *LIN5* knockout remains to be determined in a follow-up study.

HIGH-THROUGHPUT GENE TARGETING IN PROTOPLASTS RESULTS IN LOW, BUT MEASURABLE RATES OF HR

To get a better understanding of feasibility and HR rates, we used protoplast transfection with the *LIN5* GT vectors, combined with Next Generation Sequencing (NGS) for a high-throughput analysis of DSB repair events. The GreenGate vector was transfected to three different batches of protoplasts. Two mock transfections were taken as a negative control. The experiment was performed twice, once with 20 hours incubation and once with 50 hours incubation after transfection. A first PCR was performed with primers F1 and R1 (**Figure 2a, Figure S4a**) that were designed to anneal outside the donor template. In the second nested PCR, the NGS-amplicon was amplified with barcoded primers F2 and R2 (**Figure S4b**). Between 16,000 and 29,000 aligned reads were obtained per sample with Illumina NGS sequencing. Analysis of the amplicon sequences revealed that, as with stable transformation, the NHEJ-events leading to indels arose frequently, at approximately 25% (**Figure 5a**). Indel frequencies did not differ between 20 or 50h post-transfection incubation. HR-rates, comprising the sum of all substitution events in the amplicon, were 0.76% after 20 hours and 0.83%

after 50 hours, which were significantly higher than the background rate of HR-like substitutions (0.13%) events found in the mock treatments (**Figure 5b**). HR events were often combined with indels in one or more of the targets, similar as to what was seen in the one stable transformant with evidence of HR. A combination with indels occurred in 34% of the HR events at 20 hours and 52% at 50 hours.

The HR-like events in the mock treatments could be due to a combination of random occurring mutations, sequencing errors, PCR template-switching, and cross-contamination of samples during and/or after transfection. Cross-contamination during transfection is not likely, since then the NHEJ-levels would have been higher in the mock treatments. In the Gene Targeting samples, only 0.07% or reads resulted in all three amino acid substitutions (Sub1,2,3), while none of the mock treatments resulted in the perfect HR repair or the repair of the complete replicon (Sub0,1,2,3). From this, we think it is highly likely that the HR-like events in the mock treatment are most likely naturally occurring mutations or sequencing errors. Different combinations of Sub1, 2 and 3 were present at 0.27% (20 hours) and 0.43% (50 hours) (**Figure S3**). The substituted *BpiI* site in the replicon (sub0) occurred in nearly half the total HR events. This could be the result of PCR contamination of the replicon or of HR events using the complete donor template. To estimate the level of replicon contamination, we used a "no cas9" control vector (**Figure 2c**) that did contain the replicon in the second protoplast assay.

The second assay was performed with the Golden Gate vectors (**Figure 2b**). Only 50 hours were taken as post-transfection incubation. 41% indels were detected (**Figure 5c**). This time, between 37,000 and 67,000 reads were aligned per sample. The rate of HR was 3.17 % and 0.46% contained all intended substitutions (Sub1,2,3, **Figure 5d**). The rates of the HR-like events in the "no Cas9" and mock treatment were 1.61 and 1.12% respectively, with perfect HR rates of 0.24% and 0.13% respectively. Since we can observe both the intended HR substitutions (Sub1,2,3) and the complete donor template (Sub0,1,2,3) in the mock and "no Cas9" treatment, this gives an indication that some form of contamination is occurring. As the mock treatment should not contain any replicon, the contamination probably occurred from low-level (approximately 1%) cross-contamination between samples. This would have happened after the transfection, since the NHEJ levels of the control samples were low. In the "No Cas9" treatment, the presence of the replicon seems to give some additional background noise over the mock treatment. Thus, subtraction of the HR-like events of the "No Cas9" treatment from the HR-events in the Gene Targeting treatment was done to get an estimate of the net rate of HR efficiency. This was 1.55% total HR, 0.22% with all three amino acids substituted (sub1,2,3) and 0.57% with the complete donor template (sub0,1,2,3). Even though the probable contamination between samples and contamination from the replicon during sample processing was obstructive, the rates of HR-like events in the controls were low enough to discern the HR rates in the Gene Targeting treatment.

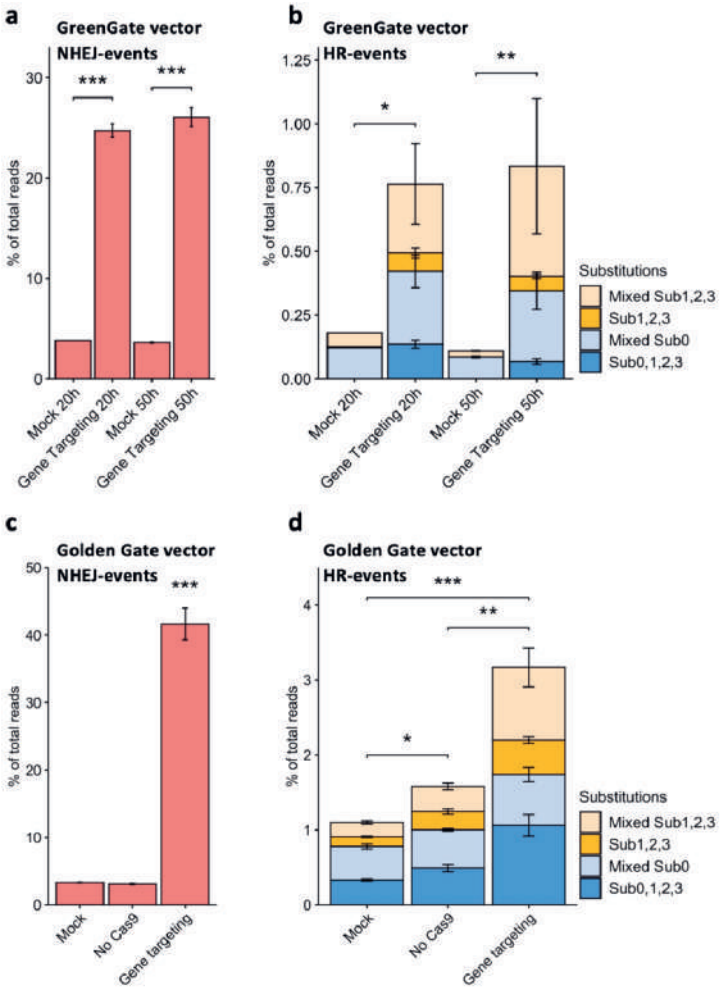


Figure 5: Analysis of GT experiments for GreenGate **(a,b)** and Golden Gate **(c,d)** vectors, respectively. **(a, c)** NHEJ-events in the GreenGate and Golden Gate vector protoplast transfections. **(b, d)** Substitution events in the GreenGate and Golden Gate vector protoplast transfections. Substitution 0 (Sub0) stands for an altered *Bpi*I site, as in the Replicon. Sub1, 2 and 3 are the intended HR substitutions in the open reading frame of *LIN5*. The mean of 3 independent transfections is shown. Error bars represent the standard error of the mean. Statistically significant differences are represented by asterisks (* p<0.05, ** p<0.01, *** p<0.001), determined by pairwise t-tests.

CONCLUSION AND FUTURE PERSPECTIVE

We intended to modify three specific amino acids in the LIN5 protein to replicate *the S. pennelli* allele of the *Brix-9-2-5* QTL. In this QTL, one or more of these amino acid substitutions were shown to be the underlying cause of the higher Brix phenotype. A promising strategy for achieving this goal was Gene Targeting using CRISPR/Cas9-induced homologous recombination (HR) with a geminiviral replicon donor template

delivery system (Baltes et al., 2014; Dahan-Meir et al., 2018). However, we did not achieve the 25% Gene Targeting events in transformed plants that were described in the work of Dahan-Meir et al., nor was this achievement equalled elsewhere since then. The estimated frequency of Gene Targeting of the *LIN5* open reading frame at our chosen target region ranged from 0.2 to 1.5% in protoplast transfection assays. In theory, the rates were high enough to obtain a stable transformant, if the same frequency as in protoplasts applies. However, in a stable transformation, we obtained just one transformant out of 212 genotyped that had a partial HR-event.

We used a multiplexed system with three sgRNAs, each targeting one of the amino acids of interest, as we hypothesized that combining targets and thus subsequent DSBs could increase the chance of HR occurring. Whether this assumption is correct remains to be tested in a larger experiment, comparing different numbers and combinations of sgRNAs. However, the multiplex approach may be compromised by the observation that even when successful HR occurred at one of three target sites, one or two of the other targets contained an indel. This may also the case for our one successful HR-event in the stable transformation, although we did not determine the two individual alleles. A benefit of the method is that the protoplast assay showed that also single amino acid substitutions were present, albeit at low frequencies. This means that all combination of substitutions could be obtained from a single vector and transformation, allowing the production of a range of alleles and assessment of individual amino acid functions.

A possible way to increase HR rates is to optimize the replicon-based system further, for instance by using different, more specific, or more active, promoters. However, we showed that the number of NHEJ-induced mutations was high, indicative of effective sgRNAs and successful inductions of double strand breaks. The replicon was also produced in large amounts, although the optimal concentration is not known. Thus, it appears that the system up to actual HR, was working. Enhancement of the HR-pathway or, alternatively blocking the NHEJ pathway could increase the rates (Even-Faitelson et al., 2011; Kwon et al., 2012), as would the use of a selection marker or reporter for insertion, which ideally could also be removed later on (Shaked et al., 2005; Čermák et al., 2015; Butler et al., 2016; Gil-Humanes et al., 2017; Wang et al., 2017). The DSBs created by Cas12/Cpf1, which upon repair may allow another round of break and repair could potentially increase HR frequency, as was apparently the case in a recent study in tomato (Van Vu et al., 2020). Completely different approaches, such as the promising Prime Editing technique (Anzalone et al., 2019; Xu et al., 2020; Veillet et al., 2020; Hassan et al., 2020; Lu et al., 2020), or base editing (Shimatani et al., 2017; Veillet et al., 2019) may also work for cases such as described here with relatively few substitutions but are not likely to be useful for substitution of longer stretches.

METHODS

CRISPR/CAS9 DESIGN AND ASSEMBLY

Four single guide RNAs (sgRNAs) targeting the three amino acids of interest were selected. The CRISPOR tool (http://crispor.tefor.net/crispor.py, spCas9, target: *Solanum lycopersicum*) was used to predict efficacy and potential off-targets (Concordet and Haeussler, 2018). None of the selected sgRNAs had off-target sites with 2 or fewer mismatches in the tomato genome. For the first amino acid substitution, two sgRNAs were used, sgRNA1a and sgRNA1b. sgRNA2 and sgRNA3 targeted the other two substitutions.

The donor template was designed around the three amino acids substitutions. The target region with the desired modification in the amino acids and with a *Bpi*I site removed was *de novo* synthesized (160 bp, GenScript (Leiden), **Figure 1a**). As flanks or homology arms 969 bp left arm and 792 bp right arm were amplified from genomic DNA with specific primers (**Table S2**).

The GreenGate toolkit was used to assemble the GreenGate Gene Targeting vector consisting of *pNOS:NPTII:tOCS* (pGGA), *PcUbi4:aCas9:tNOS* (pGGB), *AtU6:sgRNA1a, AtU6:sgRNA2*, *AtU6:sgRNA3* (pGGC), *LIR-Donor template-LIR-SIR* (pGGD), *2x:p35S:Rep:tNOS* (pGGE) and *2xp35S:turboGFP:t35S* (pGGF) cut-ligated in the pGGZ001 backbone. All components were made GreenGate compatible by removing *Bsa*I sites. The different components were assembled in level 1 or level 2 Golden Gate vectors (Engler et al., 2009; Weber et al., 2011). For instance, each sgRNA was fused to an *Arabidopsis* U6 promoter as *AtU6:sgRNA:TTTT* and cut-ligated to a level 1 vector. Correct parts were amplified from Golden Gate plasmids with a GreenGate compatible overhang (**Table S2**) and the PCR product was cut-ligated in the correct GreenGate Vector. The final construct was transformed to *E. coli* DH5α. All MoClo kit plasmids were a gift from Sylvestre Marillonnet. pICSL01009:AtU6p was a gift from Sophien Kamoun. PcUbi4:aCas9 (pICH47742) was a gift from Renze Heidstra. pRep, pSIR and pLIR were a gift from Daniel Voytas (Baltes et al., 2014).

The MoCLo toolkit and Golden Gate recombination were used to assemble the vectors (Engler et al., 2009; Weber et al., 2011). Level 1 constructs *pNOS:NPTII:tOCS* (pICH47732-pL1), *p35S:aCas9:tNOS* (pICH47742-pL2), *p2xp35S:turboGFP:t35S* (pICH47751-pL3), *LIR* (pICH47761-pL1-F4), *Donor template* (pICH47772-pL1-F5), *SIR-LIR* (pICH47781-pL1-F6), *p35S-Rep-tNOS* (pICH47791-pL1-F7), *AtU6:sgRNA1b* (pICH47732-pL1-F1), *AtU6:sgRNA2* (pICH447742-pL1-F2), *AtU6:sgRNA3* (pICH47751-pL1-F3) and *pLE6E* (pICH41766) were cut-ligated into the level 2 vector pICSL4723 (pL2). To obtain the whole construct, a two-step Golden Gate recombination was performed as described in (Weber et al., 2011; Werner et al., 2012). Additionally, the same vector was made with a dummy plasmid instead of *Cas9* in the pL1-F2 position (pICH54022).

All vectors were confirmed by sequencing, transformed to *Agrobacterium tumefaciens* C58c1, and grown with rifampicin, gentamycin, and kanamycin selection. In *Agrobacterium*, the presence of the correct construct was confirmed by restriction analysis of re-isolated plasmid.

PLANT TRANSFORMATION

Transformation was done with *Agrobacterium tumefaciens* C58C1 as previously described (van Roekel et al., 1993; Gupta and Van Eck, 2016). 375 and 1017 *Solanum lycopersicum* L. 'Moneyberg' cotyledon explants were used in the GreenGate transformation, and in the Golden Gate transformation, respectively. Tissue culture was done in a growth chamber with 16 h light and 8 h dark at 25 °C. Once shoots were formed, GFP positive shoots were selected and rooted on Rooting Inducing Medium (as described in van Roekel et al., 1993). Rooted shoots were placed on rockwool and moved to a growth chamber (16 h light and 8 h dark at 25 °C). A ploidy test was done by Iribov Analytical Services BV. Diploid shoots were genotyped for presence of the transgene and mutations in the target region.

GENOTYPING

Genomic DNA from young leaves was isolated using the CTAB-method as described (Porebski et al., 1997). Alternatively, a PCR was done directly on sampled leaf tissue by using the Phire Plant Direct PCR kit (Thermo Scientific, Catalog number F130WH). Detection of transgenes in each generation was done by a PCR on NPTII. The target region of transformed plants was amplified, sequenced, and aligned in Benchling (https://benchling.com, for primers used see **Table S2**). SYNTHEGO ICE deconvolution analysis (https://ice.synthego.com/#/) was used to resolve bi-allelic or heterozygous mutants, using a wild-type sequence as control file.

QPCR OF REPLICON PRODUCTION

Relative quantification of the circular replicon was performed on genomic DNA. Primers 4R and 4F were used for the replicon and compared to a primer pair on the *LIN5* promoter (**Table S2**). Two replicates of a non-transformed (LIN5), T_0 plants and pieces of callus from tissue culture. Bars represent the standard error the replicates. qRT-PCR was performed with iQ SYBR Green Supermix (Bio-Rad) on a Bio-Rad CFX6 qRT-PCR instrument. The following condition was applied for PCR amplification: 3 min 95°C, 40*[15 s 95°C, 60 s 60°C], followed by a melt-curve). Relative expression changes were calculated according to the $2^{-\Delta CT}$ method as described (Livak and Schmittgen, 2001).

GROWTH CONDITIONS

Shoots grown from callus were transferred to rockwool once they had roots. They were grown in a growth chamber (16 h light and 8 h dark at 25 °C). Five to Eight weeks later seedlings were transplanted into a greenhouse (Unifarm, Wageningen 51.57° N, 5.31° E, The Netherlands) on rockwool slabs at a density of 2.5 plants/m^2. Nutrients were provided by fertigation (EC 4.5, pH 5.6). Climatic conditions in the greenhouse were at ambient temperature (> 20 °C) under a 16h light/8h dark cycle (0.6–28.4 MJ m–2 day–1 natural light supplemented with artificial light using high pressure sodium lamps (SON-T Agro 600 Watt, Philips, Eindhoven, The Netherlands). Side shoots were removed once a week. Flowers were pollinated by vibrating each flower/truss three times a week with an electric toothbrush.

PROTOPLAST TRANSFECTION

E.coli containing the Gene Targeting vectors were grown in in 50 ml tubes under kanamycin selection. DNA was then isolated and purified using PureYield™ Plasmid Miniprep System from Promega to obtain around 200 ng/µl plasmid. The plasmids were concentrated by centrifuging open tubes at 6000 RPM for 30 min to obtain a concentration of 400-450 ng/µl. Protoplasts were isolated from plants grown under sterile conditions for two to three months. A couple of leaves were placed in a petri dish containing 10 ml digestion buffer (72.86 g/L mannitol, 3.90 g/L MES, 1.49 g/L KCl, 1.47 g/L $CaCl_2$ at a of pH 5.7, autoclaved) supplemented with 10 g/L cellulase R10 and 2.5 g/L of macerozyme R10. The leaves were cut in a feather-like pattern and the digestion buffer was refreshed. Petri dishes were covered in aluminium foil and incubated overnight at 25°C. The incubated leaves were shaken 30 times in a circular motion to release the protoplasts from the leaf. The protoplasts suspension was then gently transferred to a 50 ml tube through a 100 µm cell sieve. Remaining protoplasts were obtained by adding 10 Washing buffer 5 (W5, 154 mM NaCl, 125 mM $CaCl_2$.2H2O, 5 mM KCl, 2 mM MES) to the petri dish and adding them to the 50 ml tube through the cell sieve. The protoplasts were washed multiple times using. protoplasts were centrifuged at low speed (100xg for 3 minutes). Supernatant was poured off. 10 ml of W5 buffer was added gently and the protoplasts were carefully resuspended. The cells were centrifuged again. Then the pellet was resuspended in Magnesium-Mannitol solution (MMG, 0.4 M mannitol, 15 mM $MgCl_2$, 4 mM MES), centrifuged again and resuspended in MMG again. Protoplasts were counted in a haemocytometer and the volume was adjusted to correspond to 1 million protoplasts/ml.

For transfection, three biological replicates were used for each construct as three separate transfections. 8 µg of purified plasmid in 20 µl MilliQ was used to transfect 200 µl protoplast suspension in a 2 ml tube. Protoplasts were gently pipetted with wide orifice tips. 200 µl fresh PEG solution (0.4 g/mL PEG-4000 (Fluka), 0.15 g/mL 0.8M mannitol solution and 0.15 g/mL 1M $CaCl_2$ in MilliQ) was added to each transfection

tube and carefully mixed by inversion until the solution was homogeneous. The mixture was incubated for 10-20 minutes at room temperature. Subsequently, 500 µl Washing buffer I (WI, 0.5 M mannitol, 20 mM KCl and 4 mM MES) was added twice droplet wise and mixed by careful inversion. The mix was then centrifuged at low speed (200xg for 3 minutes) and the supernatant carefully removed. 1 ml WI was added again followed by resuspension and centrifuging. Finally, the pellet was resuspended in 150 µl WI. The transfections were incubated at 25°C for 20 or 50 hours in the dark. Transfected protoplasts were checked under the confocal microscope for the percentage of GFP positive protoplasts (**Table S1**). Transfection efficiency was determined using ImageJ software.

NEXT GENERATION SEQUENCING (NGS) AND ANALYSIS

DNA from the protoplasts was extracted using NucleoMag Plant kit for DNA purification from plants from Macherey-Nagel. The target region was amplified using PCR with primers F1 and R1 (**Table S2, Figure S4a** and **S5a,** 0.2 µl Q5 polymerase, 4 µl 5x Buffer, 1 µl dNTPs, 1 µl F1, 1 µl R1, 0.75 µl DNA, 10.8 µl MilliQ. PCR program: 98°C 30sec, [98°C 10sec, 63°C 20sec, 72°C 1.10 min]x36, 72°C 2 min). These primers were designed outside the donor-region to avoid contamination of the replicon. Purified product of this outer PCR was used in a second, inner PCR with barcoded primers to create a 250 bp amplicon for NGS (**Table S2, Figure S4b** and **S5b,** 0.2 µl Q5 polymerase, 4 µl 5x Buffer, 1 µl dNTPs, 1 µl F2a/F2b, 1 µl R2, 0.75 µl PCR product, 10.8 µl MilliQ. PCR program: 98°C 30sec, [98°C 10sec, 65°C 20sec, 72°C 20sec]x20, 72°C 2 min). For each sample, a unique combination of barcoded F2 and R2 primers was used. The barcode consisted of eight random nucleotides at the 5' end of the primer. 4 ul of each PCR product was combined in one tube and the mix was purified using Machery Nagel PCR + Gel purification kit. 100 ng of purified mix was shipped to Eurifins™ to perform NGS (NGSelect Amplicons on Illumina with HiSeq Adapter ligation, 5 M read pairs (150 bp)).

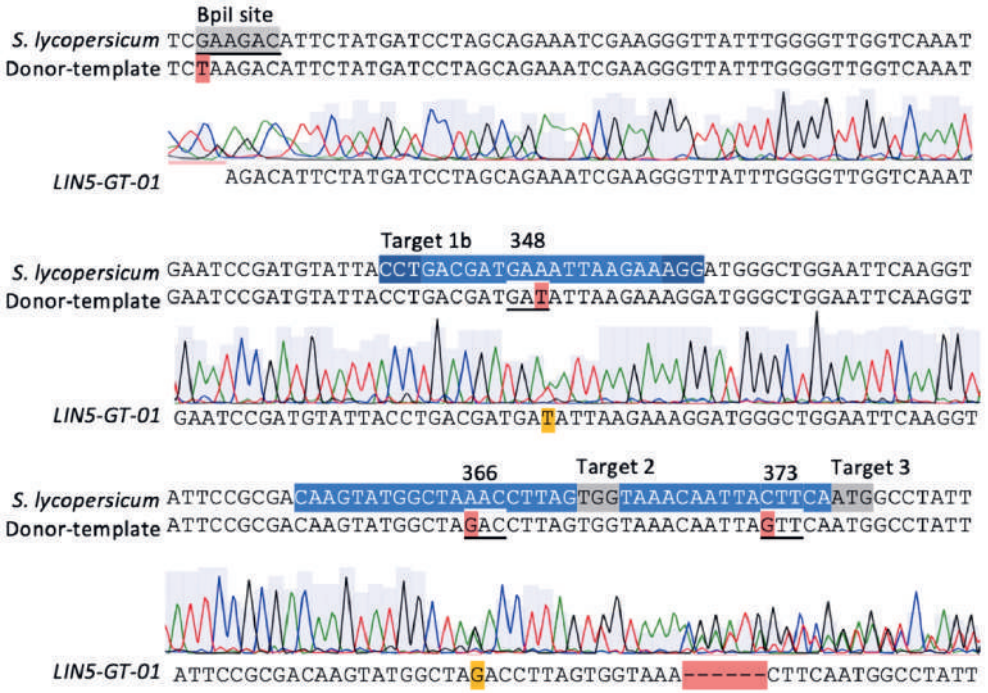
Reads were processed in CIAGEN CLC Genomics Workbench 12. Imported reads were aligned (mismatch cost = 2, gap cost = 3, maximum unaligned end mismatches = 0, minimum score = 8), de-multiplexed based on the barcodes and the barcodes were removed. The resulting fastq amplicon files were aligned to the reference sequence using CRISPREsso2 (Single end reads, 60% homology, -3 bp quantification window, 9 bp window size, minimum average read quality > 10 and minimum single bp quality > 10) run in HDR mode (Pinello et al., 2016; Clement et al., 2019). All three sgRNA targets were supplied. The NHEJ CRISPResso2 output was used to quantify the percentage of NHEJ events. However, CRISPresso2 could not handle the multiplexing we applied and the addition of the *Bbp*I site substitution in the amplicon (substitution 0, sub0). As such, not all (partial) HR events were recognized. CRISPresso2 did provide all unique alignments of the amplicon with the number of reads for each alignment. Thus, we

used these unique aligned amplicon sequences and manually identified the combination of targeted substitutions (sub1,2,3) and sub0.

ACKNOWLEDGMENTS

We thank Jin Muraoka for assistance with the protoplast assays.

SUPPLEMENTARY FIGURES



Supplemental Figure 1: Alignments of the wild-type (LIN5) and mutant coding sequences. sgRNA target regions are highlighted with a grey PAM. Mutations are highlighted red (mutations) and yellow (HR). LIN5-GT-01 was bi-allelic or heterozygous. However, we did not resolve the individual strands. An estimate of the Target 3 mutation was determined by deconvolution of the T_0 mutant with SYNTHEGO ICE Analysis (https://ice.synthego.com/#/), using a wild-type sequence as control file. For each target, the other strand most likely had a wild-type sequence, but the linkage between them is not known.

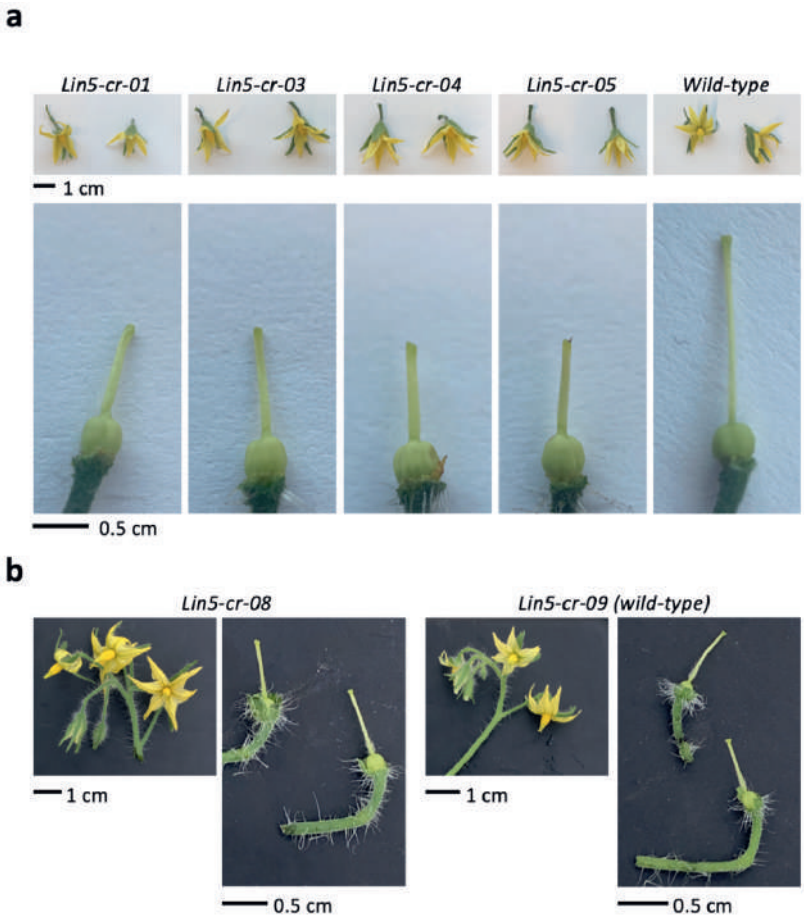
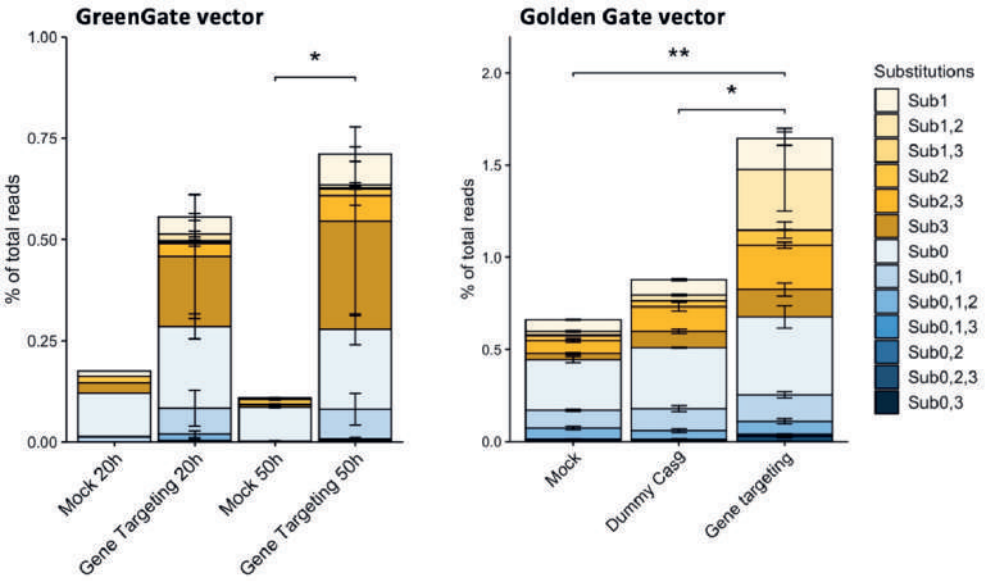
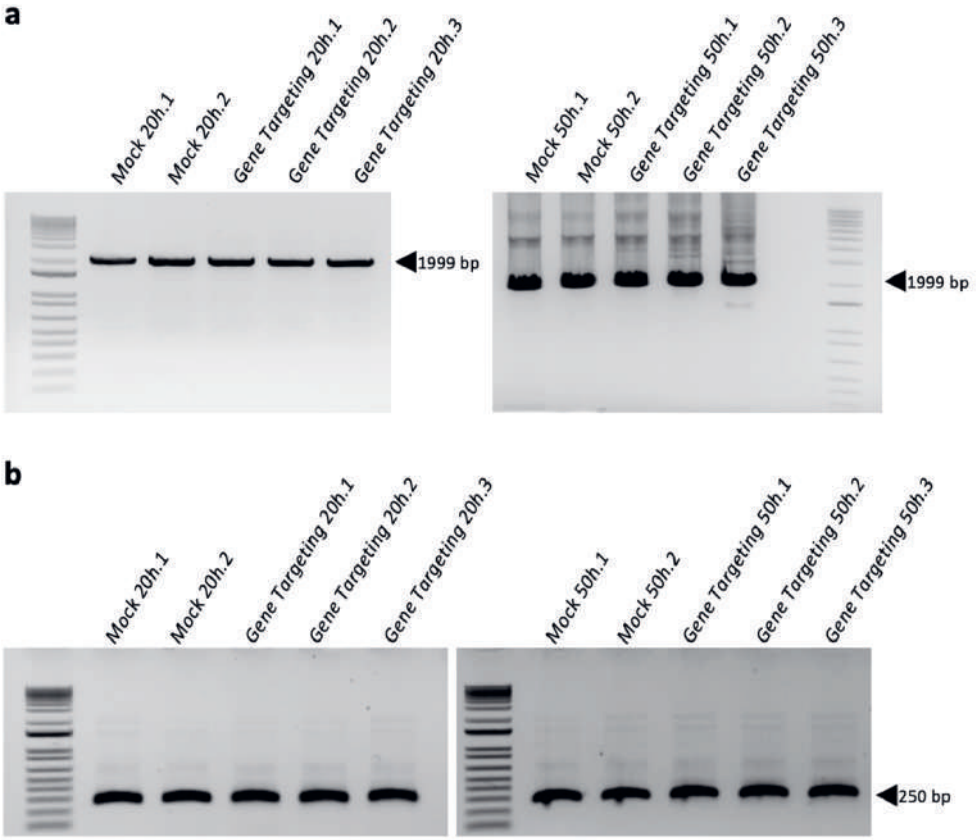


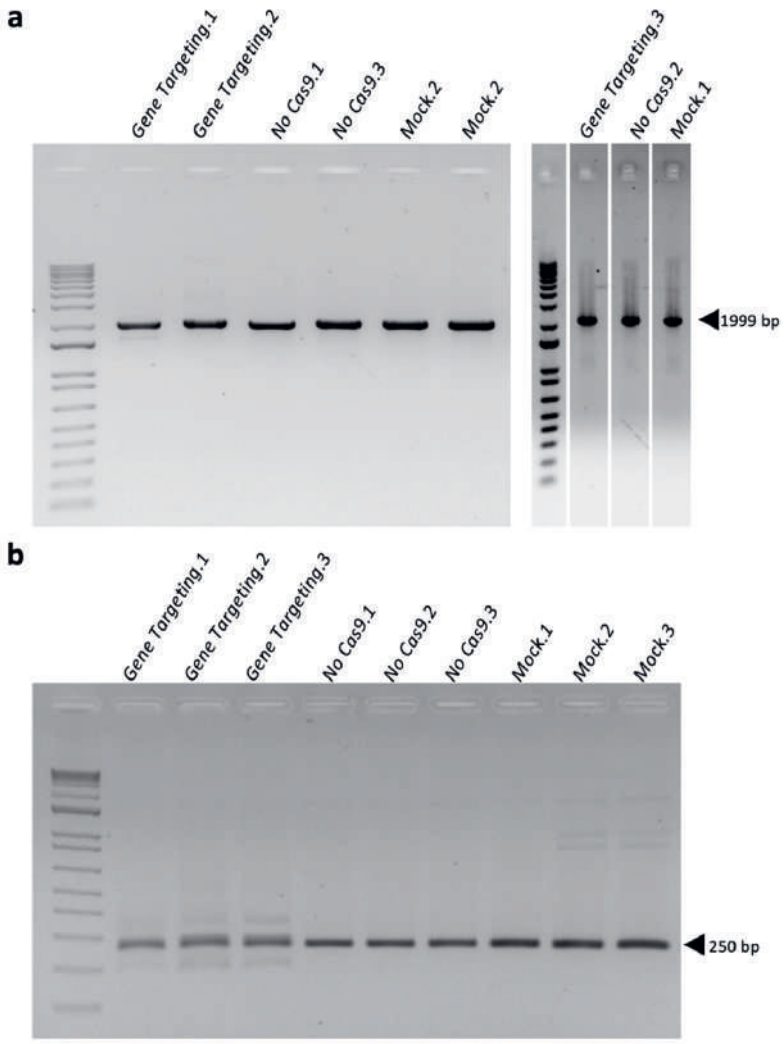
Figure S2: Representative photographs of flowers and emasculated embryos with the stigma of T_0 plants from **(a)** the GreenGate transformed plants and **(b)** the Golden Gate transformed plants.



Supplemental Figure 3: Composition of the mixed substitution events in the GreenGate and Golden Gate protoplast transfections. Substitution 0 (Sub0) stands for an altered *BpiI* site, as in the Replicon. Sub1, 2 and 3 are the intended HR substitutions in the ORF *of LIN5*. The average of 3 independent transfections is shown. Error bars represent the standard error of the mean. Statistically significant differences are represented by asterisks (* p<0.05, ** p<0.01, *** p<0.001), determined by pairwise t-tests.



Supplemental Figure 4: GreenGate vector transfection **(a)** Outer PCR with primers 1F and 1R on 0.75 ul genomic DNA isolated from protoplasts. **(b)** Inner PCR with primers 2F and 2R on 0.75 ul of cleaned outer PCR product



Supplemental Figure 5: Golden Gate vector transfection **(a)** Outer PCR with primers 1F and 1R on 0.75 ul genomic DNA isolated from protoplasts. **(b)** Inner PCR with primers 2F and 2R on 0.75 ul of cleaned outer PCR product

SUPPLEMENTARY TABLES

Supplementary table 1: Overview of the protoplast transfection efficiency, average of 3 biological replicates with the standard error (SE)

Sample	Transfection efficiency ± SE
GreenGate vector (20h incubation)	45% ± 9.1
GreenGate (50h incubation)	84% ± 10.7
Golden Gate vector (50h incubation)	37% ± 2.5
Golden Gate "No Cas" (50h incubation)	62% ± 6.1
Mock	0

Supplementary table 2: Used primers

Description target	name	Sequence (5'-3')
Cloning		
LIN5-GT1a GreenGate	sgRNA.1a	TGTGGTCTCAATTGCCTGACGATGAAATTAAGAA GTTTTAGAGCTAGAAATAGCAAG
LIN5-GT1b Golden Gate	sgRNA.1b	TGTGGTCTCAATTGCCTTTCTTAATTTCATCGTC GTTTTAGAGCTAGAAATAGCAAG
LIN5-GT2	sgRNA.2	TGTGGTCTCAATTGCAAGTATGGCTAAACCTTAG GTTTTAGAGCTAGAAATAGCAAG
LIN5-GT3	sgRNA.3	TGTGGTCTCAATTGGTGGTAAACAATTACTTCAA GTTTTAGAGCTAGAAATAGCAAG
Universal sgRNA	sgRNA Rv	TGTGGTCTCAAGCGTAATGCCAACTTTGTAC
5' Donor	5'flank Fw	CTGGAAGACCTGGAGTCTCCATTGGATGCCTTTC
	5'flank Rv	CTGGAAGACTTGCATAGAAATTACCATAGTCG
3' Donor inner (nested)	3'flank Fw	GGGAAGACTTCAATGGCCTATTGAAGAATTAG
	3'flank Rv	GCGGAAGACAGAGCGTAAATCTACATCTACATATCC
NOSp:NPTII:OCST pL1 for pGGA	NPTIIg Fw	TGTGGTCTCAACCTGAACCGCAACGTTGAAGGAGC
	NPTIIg Rv	TGTGGTCTCATGTTCGGCTGAGCCTCGACATGTTG
PcUbi4:Cas9:tNOS pL1 for pGGB	Cas9g Fw	TGTGGTCTCAAACATTCAAAAATTACGGATATGAATATAG
	Cas9g Rv	TGTGGTCTCAAGCCTCGATCTAGTAACATAGATGAC
sgRNAs PCR pL2 for pGGC	sgRNA Fw	CGCGGTCTCGGGGCTGAATTCGGATCCGGAGTGA
	sgRNA Rv	CCGGGTCTCTCTGATCCCTCGGTCACATGTGC
Donor PCR pL2 for pGGD	Donor Fw	TGTGGTCTCAGCAGCACATGTGCATCCTCTCTG
	Donor Rv	CGAGGTCTCATCAGGGTTGAACACTCTGTGCC
2x35sp:Rep PCR pL1 for pGGE – part 1	Rep1 Fw	TGTGGTCTCACTGCCAAGGCGATTAAGTTGGGTA
	Rep1 Rv	TGTGGTCTCTATGATCTCGGAAATCTCCTCTGGTTTTAA
2x35sp:Rep PCR pL1 for pGGE – part 2	Rep2 Fw	TGTGGTCTCATCATAAGGTTTCTCCTCGCAAATC
	Rep2 Rv	TGTGGTCTCATAGTAATTCCCGATCTAGTAACATAGATG
2x35Sp:turboGFP:35St pL1 fpr pGGF	GFP Fw	TGTGGTCTCAACTACAACATGGTGGAGCACGACACTC
	GFP Rv	TGTGGTCTCAATACATCTGGATTTTAGTACTGGATTTTGG
Genotyping primers		
NPTII	NPTII Fw	AGACAATCGGCTGCTCTGAT
	NPTII Rv	AGCCAACGCTATGTCCTGAT
Cas9	Cas9 Fw	CTGGCCAATGGAGAGATTCG
	Cas9 Rv	GCTCCCTGATGGGCTTATCC

Supplementary table 2 continued

Description target	name	Sequence (5'-3')
GT outer PCR	1F	AGTTTCTGTGTTTTTCACCTTTCAT 3916
	1R	AAAGCAAAACTTACCAAACTCCT 3917
GT inner PCR (without barcode)	2F	GGTTGGAAGGGATTGAGAATCG P3918
	2R	TCAATTGGACCTTTTGCTTCCT
Replicon Fw	3F	TCCAGTCTTCGTCAGGATTGC
Replicon Rv	3R	GGCCCATCCAAGCTGTAGTT
qPCR primers		
LIN5	qLIN5 Fw	CCAAAATTCTTCTCAAAGGCGGAAT
	qLIN5 Rv	CCGACACGACCTTTCAAGACAT
Replicon	4F	GCATCACAGGTTTCAACTTTTCC
	4R	TAGGATCAAATTGTTCGGCCTCG

The underlined sequence represents the spacers in the sgRNA primers. Fw: Forward, Rv: Reverse, pL1: level 1 plasmid Golden Gate, pL2: level 2 plasmid Golden Gate; GT: Gene Targeting

REFERENCES

Anzalone, A. V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A., and Liu, D.R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature **576**: 149–157.

Baltes, N.J., Gil-Humanes, J., Cermak, T., Atkins, P. a, and Voytas, D.F. (2014). DNA Replicons for Plant Genome Engineering. Plant Cell **26**: 151–163.

Butler, N.M., Baltes, N.J., Voytas, D.F., and Douches, D.S. (2016). Geminivirus-mediated genome editing in potato (*Solanum tuberosum L.*) using sequence-specific nucleases. Front. Plant Sci. **7**: 1–13.

Čermák, T., Baltes, N.J., Čegan, R., Zhang, Y., and Voytas, D.F. (2015). High-frequency, precise modification of the tomato genome. Genome Biol. **16**: 232.

Clement, K., Rees, H., Canver, M.C., Gehrke, J.M., Farouni, R., Hsu, J.Y., Cole, M.A., Liu, D.R., Joung, J.K., Bauer, D.E., and Pinello, L. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat. Biotechnol. **37**: 224–226.

Concordet, J.P. and Haeussler, M. (2018). CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res. **46**: W242–W245.

D'Halluin, K., Vanderstraeten, C., Van Hulle, J., Rosolowska, J., Van Den Brande, I., Pennewaert, A., D'Hont, K., Bossut, M., Jantz, D., Ruiter, R., and Broadhvest, J. (2013). Targeted molecular trait stacking in cotton through targeted double-strand break induction. Plant Biotechnol. J. **11**: 933–941.

Dahan-Meir, T., Filler-Hayut, S., Melamed-Bessudo, C., Bocobza, S., Czosnek, H., Aharoni, A., and Levy, A.A. (2018). Efficient in planta gene targeting in tomato using geminiviral replicons and the CRISPR/Cas9 system. Plant J. **95**: 165–171.

Deriano, L. and Roth, D.B. (2013). Modernizing the nonhomologous end-joining repertoire: Alternative and classical NHEJ share the stage. Annu. Rev. Genet. **47**: 433–455.

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes. PLoS One **4**: e5553.

Engler, C., Youles, M., Gruetzner, R., Ehnert, T.M., Werner, S., Jones, J.D.G., Patron, N.J., and Marillonnet, S. (2014). A Golden Gate modular cloning toolbox for plants. ACS Synth. Biol. **3**: 839–843.

Even-Faitelson, L., Samach, A., Melamed-Bessudo, C., Avivi-Ragolsky, N., and Levy, A. a. (2011). Localized egg-cell expression of effector proteins for targeted modification of the Arabidopsis genome. Plant J. **68**: 929–937.

Fondong, V.N. (2013). Geminivirus protein structure and function. Mol. Plant Pathol. **14**: 635–649.

Fridman, E., Carrari, F., Liu, Y.-S., Fernie, A.R., and Zamir, D. (2004). Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science **305**: 1786–1789.

Fridman, E., Pleban, T., and Zamir, D. (2000). A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc. Natl. Acad. Sci. U. S. A. **97**: 4718–4723.

Gil-Humanes, J., Wang, Y., Liang, Z., Shan, Q., Ozuna, C. V, Sánchez-León, S., Baltes, N.J., Starker, C., Barro, F., Gao, C., and Voytas, D.F. (2017). High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. Plant J. **89**: 1251–1262.

Gupta, S. and Van Eck, J. (2016). Modification of plant regeneration medium decreases the time for recovery of *Solanum lycopersicum* cultivar M82 stable transgenic lines. Plant Cell. Tissue Organ Cult. **127**: 1–7.

Hassan, M.M., Yuan, G., Chen, J.-G., Tuskan, G.A., and Yang, X. (2020). Prime Editing Technology and Its Prospects for Future Applications in Plant Biology Research. BioDesign Res. **2020**: 1–14.

Jiang, F. and Doudna, J.A. (2017). CRISPR-Cas9 Structures and Mechanisms. Annu. Rev. Biophys. **46**: 505–529.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science (80-.). **337**: 816–821.

Knoll, A., Fauser, F., and Puchta, H. (2014). DNA recombination in somatic plant cells: Mechanisms and evolutionary consequences. Chromosom. Res. **22**: 191–201.

Kwon, Y.I., Abe, K., Osakabe, K., Endo, M., Nishizawa-Yokoi, A., Saika, H., Shimada, H., and Toki, S. (2012). Overexpression of *OsRecQl4* and/or *OsExo1* enhances DSB-induced homologous recombination in rice. Plant Cell Physiol. **53**: 2142–2152.

Lampropoulos, A., Sutikovic, Z., Wenzl, C., Maegele, I., Lohmann, J.U., and Forner, J. (2013). GreenGate - A novel, versatile, and efficient cloning system for plant transgenesis. PLoS One **8**.

Liao, S., Wang, L., Li, J., and Ruan, Y.L. (2020). Cell wall invertase is essential for ovule development through sugar signaling rather than provision of carbon nutrients. Plant Physiol. **183**: 1126–1144.

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.

Lu, Y., Tian, Y., Shen, R., Yao, Q., Zhong, D., Zhang, X., and Zhu, J. (2020). Precise Genome Modification in Tomato Using an Improved Prime Editing System. Plant Biotechnol. J.: pbi.13497.

Offringa, R., van den Elzen, P.J.M., and Hooykaas, P.J.J. (1992). Gene targeting in plants using the *Agrobacterium* vector system. Transgenic Res. **1**: 114–123.

van Overbeek, M. et al. (2016). DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. Mol. Cell **63**: 633–646.

Paszkowski, J., Baur, M., Bogucki, A., and Potrykus, I. (1988). Gene targeting in plants. EMBO J. **7**: 4021–4026.

Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., and Liu, D.R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat. Biotechnol. **31**: 839–843.

Pinello, L., Canver, M.C., Hoban, M.D., Orkin, S.H., Kohn, D.B., Bauer, D.E., and Yuan, G.C. (2016). Analyzing CRISPR genome-editing experiments with CRISPResso. Nat. Biotechnol. **34**: 695–697.

Porebski, S., Bailey, L.G., and Baum, B.R. (1997). Modification of a CTAB DNA Extraction Protocol for Plants Containing High Polysaccharide and Polyphenol Components. Plant Mol. Biol. Report. **15**: 8–15.

Puchta, H. (2004). The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. J. Exp. Bot. **56**: 1–14.

Puchta, H. and Fauser, F. (2014). Synthetic nucleases for genome engineering in plants: Prospects for a bright future. Plant J. **78.5**: 727–741.

van Roekel, J.S.C., Damm, B., Melchers, L.S., Hoekema, A., and Davey, M.R. (1993). Factors influencing transformation frequency of tomato (*Lycopersicon esculentum*). Plant Cell Rep. **12**: 644–647.

Schiml, S., Fauser, F., and Puchta, H. (2014). The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in Arabidopsis resulting in heritable progeny. Plant J. **80**: 1139–1150.

Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., Van Der Oost, J., Brouns, S.J.J., and Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc. Natl. Acad. Sci. U. S. A. **108**: 10098–10103.

Shaked, H., Melamed-Bessudo, C., and Levy, A.A. (2005). High-frequency gene targeting in *Arabidopsis* plants expressing the yeast *RAD54* gene. Proc. Natl. Acad. Sci. U. S. A. **102**: 12265–12269.

Shen, S., Ma, S., Liu, Y., Liao, S., Li, J., Wu, L., Kartika, D., Mock, H.P., and Ruan, Y.L. (2019). Cell wall invertase and sugar transporters are differentially activated in tomato styles and ovaries during pollination and fertilization. Front. Plant Sci. **10**: 506.

Shimatani, Z. et al. (2017). Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. Nat. Biotechnol. **35**: 441–443.

Shukla, V.K. et al. (2009). Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. Nature **459**: 437–441.

Symington, L.S. and Gautier, J. (2011). Double-strand break end resection and repair pathway choice. Annu. Rev. Genet. **45**: 247–271.

Tan, J. et al. (2020). Efficient CRISPR/Cas9-based plant genomic fragment deletions by microhomology-mediated end joining. Plant Biotechnol. J.: 1–3.

Tieman, D. et al. (2017). A chemical genetic roadmap to improved tomato flavour. Science (80-.). **355**: 391–394.

Townsend, J.A., Wright, D.A., Winfrey, R.J., Fu, F., Maeder, M.L., Joung, J.K., and Voytas, D.F. (2009). High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature **459**: 442–445.

Veillet, F., Kermarrec, M.-P., Chauvin, L., Guyon-Debast, A., Chauvin, J.-E., Gallois, J.-L., and Nogué, F. (2020). Prime editing is achievable in the tetraploid potato, but needs improvement. bioRxiv: 2020.06.18.159111.

Veillet, F., Perrot, L., Chauvin, L., Kermarrec, M.P., Guyon-Debast, A., Chauvin, J.E., Nogué, F., and Mazier, M. (2019). Transgene-free genome editing in tomato and potato plants using Agrobacterium-mediated delivery of a CRISPR/Cas9 cytidine base editor. Int. J. Mol. Sci. **20**.

Van Vu, T., Sivankalyani, V., Kim, E., Doan, D.T.H., Tran, M.T., Kim, J., Sung, Y.W., Park, M., Kang, Y.J., and Kim, J. (2020). Highly efficient homology-directed repair using CRISPR/Cpf1-geminiviral replicon in tomato. Plant Biotechnol. J. **18**: 2133–2143.

Wan, H., Wu, L., Yang, Y., Zhou, G., and Ruan, Y.L. (2018). Evolution of Sucrose Metabolism: The Dichotomy of Invertases and Beyond. Trends Plant Sci. **23**: 163–177.

Wang, M., Lu, Y., Botella, J.R., Mao, Y., Hua, K., and Zhu, J. (2017). Gene Targeting by Homology-Directed Repair in Rice Using a Geminivirus-Based CRISPR/Cas9 System. Mol. Plant **10**: 1007–1010.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2012). Genetic screens in human cells using the CRISPR-Cas9 system. BMJ Support. Palliat. Care **2**: 256–263.

Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. PLoS One **6**: 1–11.

Werner, S., Engler, C., Weber, E., Gruetzner, R., and Marillonnet, S. (2012). Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. Bioengineered **3**: 38–43.

Wiedenheft, B., Van Duijn, E., Bultema, J., Waghmare, S., Zhou, K., Barendregt, A., Westphal, W., Heck, A., Boekema, E., Dickman, M., and Doudna, J.A. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. Proc. Natl. Acad. Sci. U. S. A. **108**: 10092–10097.

Xu, W., Zhang, C., Yang, Y., Zhao, S., Kang, G., He, X., Song, J., and Yang, J. (2020). Versatile Nucleotides Substitution in Plant Using an Improved Prime Editing System. Mol. Plant **13**: 675–678.

Zanor, M.I. et al. (2009). RNA interference of *LIN5* in tomato confirms its role in controlling brix content, uncovers the influence of sugars on the levels of fruit hormones, and demonstrates the importance of sucrose cleavage for normal fruit development and fertility. Plant Physiol. **150**: 1204–1218.

CHAPTER 7

General discussion and future perspective

Soluble solids (Brix), a measure of sugar content is an important quality trait for the tomato industry. In **Chapter 2,** I discussed several Quantitative Trait Loci (QTLs) and genes that influence Brix. From this chapter, it became clear that wild relatives have been a valuable source of new alleles for breeding populations. However, to introduce a desired trait, a breeder needs to create an introgression line containing the genomic fragment with the allele from the wild relative, which involves crossing and multiple backcrosses to lose the unwanted parts of the wild parent's genome. New breeding methods can speed up the process or introduce entirely new alleles (Klee and Tieman, 2018). With the rapid development of tools such as Clustered Regularly Interspaced Short Palindromic Repeats /Crispr-ASsociated protein (CRISPR/Cas) targeted mutagenesis, I could now revisit the QTLs and other targets to improve Brix in an indeterminate tomato cultivar, cv. Moneyberg, by creating new variation.

One of the genes I chose to target was *Cell-wall Inhibitor of β-Fructosidase* (*CIF1* or *INVINH1*). In a previous study, silencing *CIF1* led to higher invertase activity and sugar levels in tomato fruit (Jin et al., 2009). In **Chapter 5** of this thesis, I used CRISPR/Cas9 to knockout *CIF1* in an indeterminate tomato variety, cv. Moneyberg. Four mutant alleles were analysed, which all resulted in a slightly higher Brix but also in a considerable decrease of fruit size, which was not reported in the RNAi study. Effects of the *cif1* knockout mutations on invertase activity varied, which was also unexpected based on the RNAi study. Mutating the open reading frame of a gene can have unexpected results, as has been reviewed by (Wang et al., 2020). In this case, these outcomes may be either due to complex feedback regulatory loops between invertase inhibitor activity and invertase expression or activity or to unexpected effects on gene activity from Crispr/Cas mutations. Several mechanisms can alter gene function when a mutation occurs, such as an alternative translation start or genetic compensation (Rossi et al., 2015). Using CRISPR/Cas9 to remove the entire gene from start-to-stop codon could circumvent unexpected outcomes. However, removing introns could influence the expression of nearby genes as well. Deleting a large part of the proximal promoter, perhaps the first 1000 bp upstream of the start codon, could result in a more effective form of downregulation.

Apart from creating new alleles in the open reading frame, this thesis mainly explored CRISPR/Cas9 mutagenesis in regulatory regions, aiming to increase expression of genes involved in in the final soluble solids content. The two target genes for this work were *ADP-Glucose Pyrophosphorylase Large Subunit 1 (AGPL1,* **Chapter 3**) and *Lycopersicum INvertase 5 (LIN5,* **Chapter 4**) (Fridman et al., 2002, 2000; Schaffer et al., 2000). These genes were of interest because they underlie QTLs for higher Brix (Eshed and Zamir, 1996; Baxter et al., 2005b, 2005a; Schaffer et al., 2000). A superior *LIN5* allele with increased enzymatic activity due to three amino acid substitutions was identified in an introgression from *S. pennellii* (Fridman et al., 2000, 2004), and a superior *AGPL1* allele with extended temporal expression was identified in an introgression from *S. habrochaites* (Petreikov et al., 2006, 2009). My goal was to

increase the expression of both these genes in a non-determinate cultivar, cv. Moneyberg. Showing that this method is feasible opens the doors for fast implementation of the desired trait in any cultivar. For *LIN5*, even though the QTL is determined by the protein sequence, I wanted to explore the possibility of increasing the expression by deleting suppressive elements in the promoter. The *AGPL1* QTL is determined by expression difference, but unlike *LIN5,* the exact region that defined the superior allele was not reported. This made both *LIN5* and *AGPL1* as good candidates to study their regulation and to try to modify their expression.

PROMOTER ACTIVITY MODULATION IS A PROMISING STRATEGY FOR CREATING NEW PHENOTYPES

Many crop QTLs affecting quality and yield result from changes in the expression levels of the underlying genes rather than of changes in the gene products themselves (Doebley et al., 2006; Hufford et al., 2012; Ye et al., 2017). Variation in gene expression due to changes in *Cis*-regulatory elements (CREs) has been a particular driving force in evolution and breeding (Hufford et al., 2012; Meyer and Purugganan, 2013; Swinnen et al., 2016; Wittkopp and Kalay, 2012). Also, in tomato, large-scale structural variation in promoters and concurrent expression changes have occurred during domestication and improvement by breeding in tomato (Alonge et al., 2020). Examples are naturally-occurring promoter mutations leading to increased expression found in *Al-Activated Malate Transporter 9* (*SlALMT9*) (Ye et al., 2017), *CLAVATA 3* (*SlCLV3*) and *WUSCHEL* (*SlWUS*) (Somssich et al., 2016; Huang and van der Knaap, 2011; Xu et al., 2015, **Chapter 1**). The *S. habrochaites AGPL1* allele is also an example of differential expression being key to the desired phenotype although it did not play a role in domestication and was only more recently introduced in cultivated tomato by breeding (Lin et al., 2014). These examples show that small, naturally occurring variations in the promoter of a gene can lead to changed expression and to different phenotypes.

HIGHER BRIX AND ALTERED EXPRESSION WERE ACHIEVED WITH PROMOTER DELETIONS

Instead of relying on existing natural variation, methods such as targeted mutagenesis with CRISPR/Cas enable us to create variation ourselves. Changing the promoter and the expression pattern of the target gene is less likely to incur detrimental pleiotropic effects than targeting the coding sequence. Additionally, mutating the coding region usually disrupts protein function, while increased or prolonged activity might be the desired effect. In this work (**Chapter 3** and **Chapter 4**), I chose to investigate if mutagenesis of the promoter could yield increased expression. Random systematic allelic variation was created in both promoters by multiplexed CRISPR/Cas9-mutagenesis. I intended to disrupt repressing CREs by mutation, which was predicted to result in increased expression. 26 *AGPL1* and 17 *LIN5* alleles in primary transformant (T_0) lines were selected and analysed in T_1 segregant lines without the transgene. Our promoter deletions led to Brix varying between 10% decrease to a 20% increase. Gene

expression concurrently increased in a selection of promising mutants with increased Brix. Thus, a random systematic mutagenesis approach can indeed result in the desired phenotype.

However, the effects on expression were marginal, and it was difficult to establish reliable reproducibility for lines with a higher Brix. Additionally, when the *AGPL1* promoter was targeted, a Brix increase always led to decreased fruit weight. This contrasts with the phenotype of lines with the *S. habrochaites AGPL1* allele, which led to a higher sugar content and increased size (Petreikov et al., 2006, 2009). Either the use of a different variety or growing system could be the cause of this difference. Another reason could be that not only the expression of *AGPL1* was affected in the introgression line with *S. habrochaites*. In the *LIN5* promoter mutants, no correlation was found between an increased Brix and weight change. The original *Brix9-2-5* allele also had no negative effect on total yield however, fruit weight did decrease in the introgression line (Eshed and Zamir, 1996; Fridman et al., 2002, 2000). In general, higher sugar content is often associated with yield drag. As such, the sugar content of tomato has been one of the major traits that have suffered during domestication, as was demonstrated by Genome-Wide association studies (GWAS) (Zhao et al., 2019; Tieman et al., 2017). In this GWAS, *LIN5* specifically had significant associations with sugar content but was also identified as a region influenced by domestication and improvement sweeps during the selection of larger fruits (Tieman et al., 2017). Our study has shown that increasing the Brix through *LIN5* promoter mutations can result in lines that do not have a weight penalty. However, modulating *AGPL1* expression did result in weight loss.

A similar approach of promoter mutagenesis for improving another tomato trait was more successful (Rodríguez-Leal et al., 2017). However, the latter study had a few significant advantages compared to our work: open field cultivation allowing large plant numbers and a clearly visible and easily scorable trait phenotype (e.g., locule number) being less labour-intensive. In addition, the observed phenotypes were on a gradient scale between wild-type and knockout of the targeted gene. This was the case with CLV3, where locule number increased with decreasing *CLV3* expression, *COMPOUND INFLORESCENCE (S)*, where mutants cause excessive branching, *SELF PRUNING (SP)*, a flowering repressor and *WUSCHEL HOMEOBOX9* (*WOX9*), involved in Shoot Apical Meristem (SAM) termination (Rodríguez-Leal et al., 2017; Hendelman et al., 2021). Even though we achieved moderate success with a lower number of transformants and a non-visible, gain-of-function phenotype, promoter modulation probably has a higher chance of success when a quantitative loss-of-function is desired. Additionally, when that study focused on a smaller region of the *CLV3* promoter with three single guide RNAs (sgRNAs), a strategy similar to ours, it proved more challenging to find a substantial effect on locule number (Wang et al., 2021).

THE CHOICE OF CRISPR-STRATEGY DEPENDS ON THE OBJECTIVE

Our strategy was to deploy either three, six, or eight CRISPR/Cas9 sgRNAs targeting small regions or larger segments of a promoter in one transformation. What became clear is that the more targets are used, the larger were the overall deletions. The four promoter regions of *AGPL1* targeted by a set of 3 sgRNAs all resulted in mutations within the targets. Only the region with eight targets produced target-to-target deletions. In the *LIN5* promoter, I saw a similar result, with only a single event of a target-to-target deletion achieved in a transformant with three sgRNAs. A target-to-target deletion results from non-homologous end joining-repair of two Double Stranded Breaks (DSBs) occurring simultaneously or existing together long enough for the repair to occur between distant breaks. With more targets, the rate of any two simultaneous cuts increases, although other factors such as speed of repair of individual double stranded breaks may also play a role. If multiple sgRNAs are effective, target-to-target deletions can occur. Using three sgRNAs allowed targeting narrowly defined promoter regions but at the same time decreased the likelihood of two breaks coinciding. Thus, only mutations in the most effective sgRNA were common for each set of sgRNAs used. This also meant that we could not obtain many variations in alleles, as mostly the same sgRNA kept producing (small) mutations.

Using more targets could have given a better picture of which regions are relevant. Ideally, one would want small deletions scattered throughout the promoter, or at least in the suspected important (conserved) regions, combined with larger deletions. A drawback of targeting a large promoter region with multiple guides would be that the activity of sgRNAs directed at the outermost targets would yield large deletions, resulting in the loss of any smaller deletions or mutations occurring at targets in between the extreme positions. It would require a probably impossible precise control of the DSB formation and repair processes to achieve the ideal mix of small and large promoter deletions with a single mutagenic construct. Some of the drawbacks of using too many guides at once might be mitigated by targeting multiple regions in random combinations by co-transformation using two or more Agrobacterium strains each targeting one region (Jacobs et al., 2017). The best strategy also heavily depends on the available time, the space for obtaining and growing transformed plants, and how easy the phenotype can be scored. Alternatively, an efficient screening and selection strategy at an early stage of tissue culture might reduce this dependency. Instead of combining many sgRNAs in one vector or making a lot of different vectors, the effort invested in cloning and transformations might also be reduced by using "allele factories". For allele factories, a mutant containing the mutagenesis construct and wild-type plants are crossed to generate new mutations in the fresh wild-type allele in the next generation (Rodríguez-Leal et al., 2017; Wang et al., 2021). Alternatively, two plants, each containing a different construct targeting the same promoter could be crossed to widen the targeted promoter area for mutagenesis in the offspring. These approaches require the mutagenic genes to be active in the germline of the progeny

for the new mutations to be heritable, something which has not yet been tested extensively.

The work on the promoter of both *LIN5* and *AGPL1* provided leads for fruit quality improvement and gained us insight into the potential for altering gene expression (up- and down-regulation) in crops by targeted mutagenesis of candidate CREs. My conclusion is that the described untargeted mutagenesis method is feasible, yet growing mutants and phenotyping them is space- and labour-intensive. Thus, in my case not all transformants could be phenotyped, and it would be better to screen a larger number of plants in the T_1 generation. Even though a Brix measurement is rapid, it is not as fast as determining a visual developmental phenotype. To readily find important CREs in genes with difficult to measure phenotypes, transformation with a promoter fusion with a visual reporter, e.g. GFP, could be an intermediate step. These transformants could then be used in a systematic mutagenesis screen as described in thesis. However, this approach does not have the advantage of assessing the effect on the promoter in its native context and on the targeted trait that the approach of this thesis potentially has. Overall, my work shows that while promoter activity modification is feasible, the approach needs fine-tuning. A more narrow or precise definition of the target area by predicting promising promoter regions, would greatly reduce the number of mutants required and increase the feasibility and success of the approach. Combining CRISPR/Cas mutagenesis with a predictive strategy could achieve this narrowing down of the target area.

PREDICTING THE IMPORTANT PROMOTER REGIONS WOULD INCREASE FEASIBILITY OF PROMOTER ACTIVITY MODULATION

My approach was to try to generate a sufficient variety of promoter alleles in a semi-random fashion that would allow identification and disruption of promoter elements without prior prediction of putative functional elements. In parallel with producing and analysing the promoter mutants, I also performed an analysis of both promoters. Conservation analysis and DNAse I hypersensitive site sequencing (DNAseI-seq), data predicting accessible chromatin allowed us to identify a conserved, open region in the *AGPL1* promoter. Transformants with mutations in this region performed well on Brix, demonstrating the value of this analysis. Thus, I recommend that for future promoter activity modulation studies lacking the option of high-throughput mutant generation or visual phenotyping, it is worthwhile to first narrow down the area(s) of interest. This should be done with available information on conservation, accessibility, and presence of potential Transcription Factor (TF) binding sites.

Ideally, it would be possible to accurately predict the critical regulatory regions and gene regulatory network of interacting TFs by bioinformatics analysis. The following detailed information would be needed: (1) multiple sequenced genomes for comparison of promoter sequence conservation in orthologs ("phylogenetic

footprinting"); (2) Experimentally obtained knowledge of binding sequence motifs for transcription factors or groups of related TFs. Such knowledge can be derived from analysing *in situ* binding, for example, by Chromatin immune Precipitation sequencing (ChIP-seq), or by *in vitro* target enrichment and sequencing such as systematic evolution of ligands by exponential enrichment (SELEX)-seq, Chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) (Meyer et al., 2012; Bartlett et al., 2017; O'Malley et al., 2016). Applications as DNA affinity purification sequencing (DAP-seq), Yeast-one-Hybrid assays (Y1H), ChIP-seq and Electrophoretic Mobility Shift Assays (EMSAs) indicate physical interactions between regions of DNA and a TF; (3) Expression of TFs combined with the expression of genes of interest for biological relevance (co-expression analysis); (4) the epigenetic state of the DNA (DNA methylation, histone modification). Conversely, methods and databases for unbiased identification of transcription factors bound to a promoter of interest are being developed and improved (Lu et al., 2017; Chow et al., 2019; Van Bel et al., 2018; Khan et al., 2018; Aghamirzaie et al., 2017). With good prediction tools, CRISPR/Cas could be used to make a targeted mutation in the region of interest and create the desired phenotype. The subsequent effort of transforming and especially phenotyping would be greatly reduced compared to going in blind, as fewer lines would need to be screened. Thus, I envision that promoter modification to increase expression will become more feasible if the promoter-prediction tools further improve and become more accurate.

In this thesis, I have used Y1H and promoter reporter assays (Kaufmann and Mueller-Roeber, 2018; Bargmann et al., 2013), to study the regulatory network of *LIN5* and *AGPL1*. Overall, these methods gave abundant hits, of which many are likely to be false-positive hits. One reason for this is that the methods are *in vitro* based and do not reflect the *in situ* chromatin state of the two genes during fruit development (Zeitlinger, 2020; Gordân et al., 2013; Mathelier et al., 2016). This artificial situation will most likely increase the binding capabilities of the promoter fragments used in the different assays, reflected especially by the number of hits obtained in the Y1Hseq. Nevertheless, the tools were still valuable as a starting point, and interactors were confirmed in the promoter reporter assay after selection for co-expression. In addition, there was an overlap between the results of the different methods. Most of the Y1H hits based on the *Arabidopsis* library were also identified in the tomato Y1Hseq experiment. The TFs selected based on the promoter mutations, were present in the Y1Hseq results as well, indicating that a Y1H assay can give leads for upstream TFs.

GENE TARGETING WITH CRISPR/CAS9 IS NOT YET EFFICIENT ENOUGH

Besides promoter modification, I also investigated the applicability of the latest developments in CRISPR/Cas-mediated gene targeting to re-create the *S. pennellii LIN5* allele in cv. Moneyberg (**Chapter 6**). The underlying sequence variation that is thought to be responsible for the QTL effect comprises three amino acid substitutions, Glu348Asp, Asn366Asp and Leu373Val, in the *S. pennellii* LIN5 protein compared to *S.*

lycopersicum (Fridman et al., 2000, 2004). *In planta* modification of the sequence encoding the involved amino acids would be an ideal for studying the relevance of each amino acid. Three base pair-substitutions were required to induce the three intended substitutions in the cv. Moneyberg LIN5 protein. All three substitutions are different, which would make an approach by base editing challenging (Veillet et al., 2019; Shimatani et al., 2017). A recently described, promising technique is Homologous Recombination (HR)-based Gene Targeting using donor template-delivery on a geminiviral replicon (Dahan-Meir et al., 2018; Baltes et al., 2014). With this approach, CRISPR/Cas9 creates a DSB at a target site of interest, and the HR-pathway uses a donor template supplied in abundance by a geminiviral-based rolling circle replicon. Dahan-Meir et al. described rates of 25% without using a selection marker. However, several other reports achieve much lower frequencies and in the range of what we have achieved in tomato protoplasts (between 0.2 and 1.5% of sequenced amplicons). We used three sgRNA simultaneously and in proximity, with the intent to increase DSB frequency as well as HR-mediated repair. The major effect of that appeared to be that the few successful gene targeting events were often combined with Non-Homologous End Joining (NHEJ)-induced Indels (between 34 and 52% of the gene targeting events). Thus, for future experiments, I would recommend also using a vector with just one target. This should decrease the combinations of Indels and target substitutions. However, it might reduce overall HR-efficiency as well, as the number of DSB are reduced.

The high rates achieved by Dahan-Meir have not been reported elsewhere. Van Vu et al. reported about 4% Gene Targeting efficiency with Cas9 in the *ANT1* gene, a key transcription factor controlling the anthocyanin pathway, showed when cultured at 31 °C. At 25 °C the rates were similar to our observations, at about 2.5 % (Van Vu et al., 2020). Cas12/Cpf1 outperformed Cas9 with approximately 10% HR rates. Thus, exploring different Cas alternatives and using higher incubation temperatures could by an effective strategy to improve HR rates (Wolter and Puchta, 2019). Alternatively, other strategies remain to be explored as well. These include base editing (Veillet et al., 2019; Shimatani et al., 2017), and Prime Editing (Anzalone et al., 2019), which was already achieved for tomato and other crops (Lu et al., 2020; Xu et al., 2020; Veillet et al., 2020; Hassan et al., 2020).

THE GENE-EDITING REGULATION STANDS IN THE WAY OF THE INTRODUCTION OF EDITED CROPS IN EUROPE

The use of gene editing and, more specifically, CRISPR/Cas technology clearly has enormous potential for plant breeding. However, its use and regulation are a topic of intense debate worldwide and especially in Europe. The issue is if varieties obtained by targeted mutagenesis, such as gene editing (SDN1 or SDN2 in the definition of EFSA, Naegeli et al., 2020) should fall under the EU GMO directive. In Europe, the Directive 2001/18/EC states that Genetically Modified Organisms (GMOs) are subject to an

environmental risk assessment prior to market release, followed by traceability, labelling and post-release monitoring (European Parliament, 2001). In the directive, a GMO was defined as "an organism, with the exception of human beings, in which the genetic material has been altered in a way that does not occur naturally by mating and/or natural recombination". In an annex (1B) organisms obtained through certain techniques of genetic modification were exempted. These techniques are *in vitro* fertilisation, natural processes such as conjugation, transduction and transformation, polyploidy induction and techniques/methods that modify the genetics of an organisms on the condition that they do not involve the use of recombinant nucleic acid molecules. These latter techniques are listed as mutagenesis and cell fusion (including protoplast fusion).

When the GMO directive was written, the use of targeted mutagenesis, such as with CRISPR/Cas, had not yet been discovered. Mutagenesis induces random mutations using of ionizing radiation or mutagenic chemicals, hereafter referred to as "untargeted" mutagenesis. From a pool of plants created by untargeted mutagenesis, a selection for interesting genotypes and phenotypes could be made before or after back-crossing to get rid of all the unwanted mutations. CRISPR/Cas has the enormous advantage of only inducing mutations at the site(s) of your choice: "targeted" mutagenesis. With CRISPR/Cas, breeders and scientists can make the mutation of choice and either segregate out the mutagenic CRISPR/Cas insert or use sgRNA-loaded Cas proteins in protoplasts followed by regeneration. Both methods result in transgene free-offspring (Svitashev et al., 2015; Zhang et al., 2016; Andersson et al., 2018). This makes the resulting CRISPR/Cas products indistinguishable from conventional equivalents on a genetic and phenotypic level.

Many stakeholders in Europe hoped that targeted mutagenesis would fall under this annex and be exempted from the Directive 2001/18/EC. However, in July 2018, the European Court of Justice ruled that all organisms obtained through mutagenesis, including untargeted mutagenesis and targeted mutagenesis mediated by, for instance, CRISPR/Cas, are to be classified as GMOs and would fall under the GMO directive (Case C-528/16, Court of Justice of the European Union., 2018). An exception was made for the organisms obtained through techniques with a substantial historical safety record, e.g. untargeted mutagenesis. In addition, EU member states may lay down additional rules on exempted products, for instance, obtained through untargeted mutagenesis.

Very recently, in April 2021, the European Commission published a study on the use of New Genomics Techniques (NGT), such as CRISPR/Cas-mutagenesis. This study concluded that NGT products have the potential to contribute to sustainable agri-food systems. NGT-products are subject to GMO legislation, but the current legislation is outdated and needs clarification (European Commission, 2021). In conclusion, only time will tell if CRISPR/Cas mutagenesis in Europe will be regarded as safe enough for consumption and the environment.

CONCLUSIONS

The availability of CRISPR/Cas-mediated mutagenesis has renewed interest in modulating the expression of genes by targeting the promoter. This thesis has explored the possibilities of using systematic promoter deletions to increase the expression of genes involved in the sugar content of tomato. Our findings are that it is possible to find mutants with increased expression and a corresponding phenotype. However, to find desired mutants, many different mutant alleles need to be screened. If the target phenotype cannot easily be screened by eye, such as with soluble solids content, phenotyping requires a significant effort. To decrease the number of mutants required, narrowing the target area would increase efficiency tremendously. I suggest that improved predictions on which regions are likely to harbour CREs, would narrow the target area. Combining promoter deletions with TF-promoter interaction assays such as Y1Hs and promoter reporter assays can complete the identification of a gene's regulatory network.

REFERENCES

Aghamirzaie, D., Raja Velmurugan, K., Wu, S., Altarawy, D., Heath, L.S., and Grene, R. (2017). Expresso: A database and web server for exploring the interaction of transcription factors and their target genes in Arabidopsis thaliana using ChIP-Seq peak data. F1000Research **6**: 372.

Alonge, M. et al. (2020). Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. Cell **182**: 1–17.

Andersson, M., Turesson, H., Olsson, N., Fält, A.-S., Ohlsson, P., Gonzalez, M.N., Samuelsson, M., and Hofvander, P. (2018). Genome editing in potato via CRISPR-Cas9 ribonucleoprotein delivery. Physiol. Plant. **164**: 378–384.

Anzalone, A. V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A., and Liu, D.R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature **576**: 149–157.

Baltes, N.J., Gil-Humanes, J., Cermak, T., Atkins, P. a, and Voytas, D.F. (2014). DNA Replicons for Plant Genome Engineering. Plant Cell **26**: 151–163.

Bargmann, B.O.R., Marshall-Colon, A., Efroni, I., Ruffel, S., Birnbaum, K.D., Coruzzi, G.M., and Krouk, G. (2013). TARGET: A Transient Transformation System for Genome-Wide Transcription Factor Target Discovery. Mol. Plant **6**: 978–980.

Bartlett, A., O'Malley, R.C., Huang, S.S.C., Galli, M., Nery, J.R., Gallavotti, A., and Ecker, J.R. (2017). Mapping genome-wide transcription-factor binding sites using DAP-seq. Nat. Protoc. **12**: 1659–1672.

Baxter, C.J., Carrari, F., Bauke, A., Overy, S., Hill, S.A., Quick, P.W., Fernie, A.R., and Sweetlove, L.J. (2005a). Fruit carbohydrate metabolism in an introgression line of tomato with increased fruit soluble solids. Plant Cell Physiol. **46**: 425–437.

Baxter, C.J., Sabar, M., Quick, W.P., and Sweetlove, L.J. (2005b). Comparison of changes in fruit gene expression in tomato introgression lines provides evidence of genome-wide transcriptional changes and reveals links to mapped QTLs and described traits. J. Exp. Bot. **56**: 1591–1604.

Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van De Peer, Y., Coppens, F., and Vandepoele, K. (2018). PLAZA 4.0: An integrative resource for functional, evolutionary and comparative plant genomics. Nucleic Acids Res. **46**: D1190–D1196.

Chow, C.-N., Lee, T.-Y., Hung, Y.-C., Li, G.-Z., Tseng, K.-C., Liu, Y.-H., Kuo, P.-L., Zheng, H.-Q., and Chang, W.-C. (2019). PlantPAN3.0: a new and updated resource for reconstructing transcriptional regulatory networks from ChIP-seq experiments in plants. Nucleic Acids Res. **47**: D1155–D1163.

Court of Justice of the European Union. (2018). Judgment in Case C-528/16.

Dahan-Meir, T., Filler-Hayut, S., Melamed-Bessudo, C., Bocobza, S., Czosnek, H., Aharoni, A., and Levy, A.A. (2018). Efficient in planta gene targeting in tomato using geminiviral replicons and the CRISPR/Cas9 system. Plant J. **95**: 165–171.

Doebley, J.F., Gaut, B.S., and Smith, B.D. (2006). The Molecular Genetics of Crop Domestication. Cell **127**: 1309–1321.

Eshed, Y. and Zamir, D. (1996). Less-than-additive epistatic interactions of quantitative trait loci in tomato. Genetics **143**: 1807–1817.

European Commission (2021). Study on the status of new genomic techniques under Union law and in light of the Court of Justice ruling in Case C-528/1.

European Parliament (2001). Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC. 32001L0018 - EN - EUR-Lex.

Fridman, E., Carrari, F., Liu, Y.-S., Fernie, A.R., and Zamir, D. (2004). Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science **305**: 1786–1789.

Fridman, E., Liu, Y.S., Carmel-Goren, L., Gur, A., Shoresh, M., Pleban, T., Eshed, Y., and Zamir, D. (2002). Two tightly linked QTLs modify tomato sugar content via different physiological pathways. Mol. Genet. Genomics **266**: 821–826.

Fridman, E., Pleban, T., and Zamir, D. (2000). A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc. Natl. Acad. Sci. U. S. A. **97**: 4718–4723.

Gordân, R., Shen, N., Dror, I., Zhou, T., Horton, J., Rohs, R., and Bulyk, M.L. (2013). Genomic Regions Flanking E-Box Binding Sites Influence DNA Binding Specificity of bHLH Transcription Factors through DNA Shape. Cell Rep. **3**: 1093–1104.

Hassan, M.M., Yuan, G., Chen, J.-G., Tuskan, G.A., and Yang, X. (2020). Prime Editing Technology and Its Prospects for Future Applications in Plant Biology Research. BioDesign Res. **2020**: 1–14.

Hendelman, A. et al. (2021). Conserved pleiotropy of an ancient plant homeobox gene uncovered by cis-regulatory dissection. Cell.

Huang, Z. and van der Knaap, E. (2011). Tomato fruit weight 11.3 maps close to fasciated on the bottom of

chromosome 11. Theor. Appl. Genet. **123**: 465–474.
Hufford, M.B. et al. (2012). Comparative population genomics of maize domestication and improvement. Nat. Genet. **44**: 808–811.
Jacobs, T.B., Zhang, N., Patel, D., and Martin, G.B. (2017). Generation of a Collection of Mutant Tomato Lines Using Pooled CRISPR Libraries. Plant Physiol. **174**: 2023–2037.
Jin, Y., Ni, D.-A., and Ruan, Y.-L. (2009). Posttranslational elevation of cell wall invertase activity by silencing its inhibitor in tomato delays leaf senescence and increases seed weight and fruit hexose level. Plant Cell **21**: 2072–2089.
Kaufmann, K. and Mueller-Roeber, B. (2018). Plant Gene Regulatory Networks Methods and Protocols Methods in Molecular Biology K. Kaufmann and B. Mueller-Roeber, eds (Humana Press).
Khan, A. et al. (2018). JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. **46**: D260–D266.
Klee, H.J. and Tieman, D.M. (2018). The genetics of fruit flavour preferences. Nat. Rev. Genet. **19**: 347–356.
Lin, T. et al. (2014). Genomic analyses provide insights into the history of tomato breeding. Nat. Genet. **46**: 1220–1226.
Lu, P. et al. (2017). The fruitENCODE project sheds light on the genetic and epigenetic basis of convergent evolution of climacteric fruit ripening. bioRxiv: 231258.
Lu, Y., Tian, Y., Shen, R., Yao, Q., Zhong, D., Zhang, X., and Zhu, J. (2020). Precise Genome Modification in Tomato Using an Improved Prime Editing System. Plant Biotechnol. J.: pbi.13497.
Mathelier, A., Xin, B., Chiu, T.P., Yang, L., Rohs, R., and Wasserman, W.W. (2016). DNA Shape Features Improve Transcription Factor Binding Site Predictions In Vivo. Cell Syst. **3**: 278-286.e4.
Meyer, C.A., Tang, Q., and Liu, X.S. (2012). Minireview: Applications of next-generation sequencing on studies of nuclear receptor regulation and function. Mol. Endocrinol. **26**: 1651–1659.
Meyer, R.S. and Purugganan, M.D. (2013). Evolution of crop species: Genetics of domestication and diversification. Nat. Rev. Genet. **14**: 840–852.
Naegeli, H. et al. (2020). Applicability of the EFSA Opinion on site-directed nucleases type 3 for the safety assessment of plants developed using site-directed nucleases type 1 and 2 and oligonucleotide-directed mutagenesis. EFSA J. **18**: 6299.
O'Malley, R.C., Huang, S.S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell **165**: 1280–1292.
Petreikov, M., Shen, S., Yeselson, Y., Levin, I., Bar, M., and Schaffer, A.A. (2006). Temporally extended gene expression of the *ADP-Glc pyrophosphorylase large subunit (AgpL1)* leads to increased enzyme activity in developing tomato fruit. Planta **224**: 1465–1479.
Petreikov, M., Yeselson, L., Shen, S., Levin, I., Schaffer, A.A., Dagan, B., Efrati, A., Bar, M., and Co, G.S. (2009). Carbohydrate balance and accumulation during development of near-isogenic tomato lines differing in the AGPase-L1 allele. J. Am. Soc. Hortic. Sci. **134**: 134–140.
Rodríguez-Leal, D., Lemmon, Z.H., Man, J., Bartlett, M.E., and Lippman, Z.B. (2017). Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing. Cell **171**: 470-480.e8.
Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Hölper, S., Krüger, M., and Stainier, D.Y.R. (2015). Genetic compensation induced by deleterious mutations but not gene knockdowns. Nature **524**: 230–233.
Schaffer, A.A., Levin, I., Oguz, I., Petreikov, M., Cincarevsky, F., Yeselson, Y., Shen, S., Gilboa, N., and Bar, M. (2000). ADPglucose pyrophosphorylase activity and starch accumulation in immature tomato fruit: The effect of a Lycopersicon hirsutum-derived introgression encoding for the large subunit. Plant Sci. **152**: 135–144.
Shimatani, Z. et al. (2017). Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. Nat. Biotechnol. **35**: 441–443.
Somssich, M., Je, B. Il, Simon, R., and Jackson, D. (2016). CLAVATA-WUSCHEL signaling in the shoot meristem. Dev. **143**: 3238–3248.
Svitashev, S., Young, J., Schwartz, C., Gao, H., Falco, S.C., and Cigan, a M. (2015). Targeted Mutagenesis, Precise Gene Editing and Site-Specific Gene Insertion in Maize Using Cas9 and Guide RNA. Plant Physiol. **169**: 931–945.
Swinnen, G., Goossens, A., and Pauwels, L. (2016). Lessons from Domestication: Targeting Cis-Regulatory Elements for Crop Improvement. Trends Plant Sci. **21**: 506–515.
Tieman, D. et al. (2017). A chemical genetic roadmap to improved tomato flavour. Science (80-.). **355**: 391–394.
Veillet, F., Kermarrec, M.-P., Chauvin, L., Guyon-Debast, A., Chauvin, J.-E., Gallois, J.-L., and Nogué, F. (2020). Prime editing is achievable in the tetraploid potato, but needs improvement. bioRxiv: 2020.06.18.159111.
Veillet, F., Perrot, L., Chauvin, L., Kermarrec, M.P., Guyon-Debast, A., Chauvin, J.E., Nogué, F., and Mazier, M. (2019). Transgene-free genome editing in tomato and potato plants using Agrobacterium-mediated delivery of a CRISPR/Cas9 cytidine base editor. Int. J. Mol. Sci. **20**.

Van Vu, T., Sivankalyani, V., Kim, E., Doan, D.T.H., Tran, M.T., Kim, J., Sung, Y.W., Park, M., Kang, Y.J., and Kim, J. (2020). Highly efficient homology-directed repair using CRISPR/Cpf1-geminiviral replicon in tomato. Plant Biotechnol. J. **18**: 2133–2143.

Wang, R., Angenent, G.C., Seymour, G., and de Maagd, R.A. (2020). Revisiting the Role of Master Regulators in Tomato Ripening. Trends Plant Sci. **25**: 291–301.

Wang, X., Aguirre, L., Rodríguez-Leal, D., Hendelman, A., Benoit, M., and Lippman, Z.B. (2021). Dissecting cis-regulatory control of quantitative trait variation in a plant stem cell circuit. Nat. Plants **7**: 419–427.

Wittkopp, P.J. and Kalay, G. (2012). Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. Nat. Rev. Genet. **13**: 59–69.

Wolter, F. and Puchta, H. (2019). *In planta* gene targeting can be enhanced by the use of <scp>CRISPR</scp> /Cas12a. Plant J.: tpj.14488.

Xu, C. et al. (2015). A cascade of arabinosyltransferases controls shoot meristem size in tomato. Nat. Genet. **47**: 784–792.

Xu, W., Zhang, C., Yang, Y., Zhao, S., Kang, G., He, X., Song, J., and Yang, J. (2020). Versatile Nucleotides Substitution in Plant Using an Improved Prime Editing System. Mol. Plant **13**: 675–678.

Ye, J., Wang, X., Hu, T. xu, Zhang, F. xia, Wang, B., Li, C. xin, Yang, T. xia, Li, H. xia, Lu, Y. en, Giovannoni, J.J., Zhang, Y., and Ye, Z. (2017). An InDel in the Promoter of Al-ACTIVATED MALATE TRANSPORTER9 Selected during Tomato Domestication Determines Fruit Malate Contents and Aluminum Tolerance. Plant Cell **29**: 2249-2268.

Zeitlinger, J. (2020). Seven myths of how transcription factors read the cis-regulatory code. Curr. Opin. Syst. Biol. **23**: 22–31.

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., Qiu, J.-L., and Gao, C. (2016). Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. Nat. Commun. **7**: 12617.

Zhao, J., Sauvage, C., Zhao, J., Bitton, F., Bauchet, G., Liu, D., Huang, S., Tieman, D.M., Klee, H.J., and Causse, M. (2019). Meta-analysis of genome-wide association studies provides insights into genetic control of tomato flavour. Nat. Commun. **10**: 1–12.

SUMMARY

Throughout the world, the tomato is a valuable and loved vegetable. Its appeal to consumers is important to breeders and growers as well as for promoting consumer health. Tomato sweetness is a trait that consumers generally like. The sugar content, often measured as soluble solids content (SSC or Brix), is an essential contributor to the experience of sweetness. The fruit's soluble solids are mainly imported as sucrose, which is converted to equimolar amounts of glucose and fructose in the fruit. Sucrose is transported from the sources, e.g. leaves, to fruits (the sink), in a tightly coordinated fashion. Final sugar content in ripe fruit is determined by the production and export of photosynthates from the source and the ability of the fruit to import and accumulate sugars (such as in starch) during growth and finally, release them during ripening. The aim of this thesis was to explore new ways of increasing sugar content using targeted mutagenesis and editing with CRISPR/Cas9. Expression modulation by promoter mutation, the manipulation of post-translational regulation and, gene targeting were explored.

Chapter 2 reviewed and explored the regulation of sugar metabolism in tomato, focussing on genes that would be amenable for editing using CRISPR/Cas to increase sugar content in the fruit. Identifying and understanding the role of the genes underlying the production and distribution of sugars is crucial for crop breeding. It is even more important because of the rapid development of new breeding technologies that enable precise modification of these genes. The combination of synthesis, transport, metabolism, and storage determines the final sugar content in tomato. I outlined Quantitative Trait Loci (QTLs) for sugar content and their underlying genes.

Moreover, I discussed possible new targets for breeding that have emerged recently. Some need further research and development because not all putative targets have been validated for an overall positive effect (in tomato), especially when negative trade-offs have to be balanced. The combination of QTL discovery and molecular analysis of the sugar pathway(s) has yielded a wide array of either proven or promising targets for increasing the tomato fruits' sugar content.

From this review, it became clear that both *ADP-Glucose Pyrophosphorylase Large subunit 1 (AGPL1)* and *Cell Wall INvertase 5* (*LIN5*) are essential genes for sugar accumulation. *AGPL1* encodes the large subunit of ADP-glucose pyrophosphorylase (AGPase). This tetrameric AGPase protein complex is involved in starch synthesis in plastids during fruit development. LIN5 hydrolyses sucrose into glucose and fructose. The gradient created is a major driving force for importing sugars into the fruit (sink strength). Both genes underlie a QTL for Brix, with both having an allele from a wild tomato relative conferring the positive effect on sugar content. One of this thesis' aims was to study and modify the transcriptional regulation of expression of *AGPL1* and *LIN5*.

I studied the *Cis*-regulatory Elements (CREs) involved in the regulation of these two genes. One approach was to identify CREs and their interacting Transcription Factors (TFs) using various *in silico* and experimental tools. Simultaneously, I applied CRISPR/Cas9 multiplexed mutagenesis to create variation in both promoters. By studying mutants for their effect on the target gene expression, I intended to discover new types of CRE functionalities in an *in vivo* system. I hypothesized that when a mutation occurred in a CRE it would disrupt a Transcription Factor's binding. If the disrupted site is the binding site for a repressor, the target gene expression would be expected to increase. By increasing the expression of either *LIN5* or *AGPL1*, we concomitantly tried improving the tomato flavour by increasing sugar content.

An *in silico* analysis of the *AGPL1* promoter in **Chapter 3** resulted in identifying conserved and accessible promoter regions, and Yeast-one-hybrid (Y1H) assays identified interacting TFs. I studied the TFs effect on *AGPL1* transcription with promoter-reporter assays. Two potential repressors (FRUITFULL 2 and the CCCH-type Zinc Finger C3H13) and several activators (Abscisic Acid Responsive Elements-Binding Factor 1 (AREB1), the homolog of the Arabidopsis B-box zinc-finger TF BBX19, the homolog of the Arabidopsis GATA-motif containing TF GATA9, Jasmonic Acid 1 (JA1), Nuclear Factor-YA10 (NF-YA10), the homolog of the Arabidopsis Telomerase Repeat Binding Factor-Like 3 (TRFL), and the WRKY motif-containing WRKY24, WRKY41 and WRKY81) were found. I used multiplexed mutagenesis with CRISPR/Cas9 on the *AGPL1* promoter in cv. Moneyberg to modulate expression *in situ* and create plants with higher sugar content. Brix was increased in several of the obtained mutant lines, however, always occurring with a decreased fruit weight. For two out of three of the lines with the highest Brix, an increase of *AGPL1* expression was demonstrated.

Chapter 4 focussed on *LIN5*. Y1H and promoter reporter assays revealed several TFs interacting with the *LIN5* promoter. GATA9, the homolog of pepper MYB48, 'MYB48', R2R3MYB58, MYB76, and a NAC TF behaved as repressors, while C3H13, NF-YA10 and a homolog of pepper NPY1, 'NPY1', increased expression in the promoter reporter assay. I made promoter mutants and found increased Brix in several of the generated mutant lines. Several of these were characterized in the T_2 generation for *LIN5* expression, where I could confirm altered expression in the mutants compared to the wild type. In a promoter reporter assay comparing the mutant promoters with a wild-type promoter, HAT1, MYB HYH and WRKY24 were identified as weak activators of the mutant's promoters.

The work on the promoters of *LIN5* and *AGPL1* provided leads for fruit quality improvement and gained us some insight into the potential of genome editing for altering gene expression (up or down-regulation) in crops by targeting candidate gene CREs. My conclusion is that this approach is feasible, although growing and phenotyping mutants is space- and labour-intensive when targeting the promoter in a

non-biased manner. Reducing the target area by predicting promising promoter regions would reduce the number of mutants required.

In Chapter 5, I explored the role of *Cell-wall Inhibitor of β-fructosidase* (*CIF1* or *INVINH1*), a post-translational inhibitor of LIN5, in regulating LIN5 activity. I generated knockout-mutants of *CIF1* and studied the effects on final soluble solids content (Brix°) and invertase activity, as a previous report showed a positive effect of down-regulation by RNAi on sugar content. CRISPR/Cas-generated knockout mutants did have an increased Brix, but their fruit size decreased considerably. In addition, the mutants did not show a consistent relation between predicted decreased CIF1 activity due to knockout mutations and increased LIN5 activity in an enzyme assay.

Chapter 6 focused on improving the CRISPR/Cas9-toolbox by applying multiplexed Gene Targeting on *LIN5*. *LIN5* is the underlying gene of the high Brix QTL, Brix9-2-5, that originates from *Solanum pennellii*. Three amino acid substitutions in the *S. pennellii* allele are thought to be responsible for a more active LIN5 protein. My goal was to create these three amino acid changes in our cultivar, cv. Moneyberg via homologous recombination, with a geminiviral-based replicon donor delivery system to simultaneously target all three amino acids. I found that the approach results in Gene Targeting in experiments in protoplasts and in stable transformation, albeit at a very low efficiency (1-3%).

In conclusion, this thesis explored the various possibilities of CRISPR/Cas mutagenesis to improve tomato flavour. Targeting CREs to increase expression and using geminiviral-based Gene Targeting were both feasible. However, the frequency of success needs improvement for a broader application, especially in a commercial setting.

SAMENVATTING

Tomaat is wereldwijd een geliefde en commercieel belangrijke groente. Aantrekkelijkheid van tomaat voor de consument is zowel belangrijk voor veredelaars en telers als voor de gezondheid van de consument. De zoetheid van tomaten is een eigenschap die consumenten over het algemeen waarderen. Het suikergehalte, vaak gemeten als gehalte aan oplosbare vaste stoffen (gemeten in Brix), levert een essentiële bijdrage aan de beleving van de zoetheid. Suikers worden voornamelijk geïmporteerd in de vrucht als sucrose, dat daar wordt omgezet in equimolaire hoeveelheden glucose en fructose. Sucrose wordt op een strak gecoördineerde manier naar de vrucht getransporteerd vanuit fotosynthetische bronnen, b.v. bladeren. Het uiteindelijke suikergehalte in de rijpe vrucht wordt bepaald door de productie en export van fotosyntheseproducten uit de bron, het vermogen van het fruit om suikers te importeren en op te slaan tijdens de groei (bijvoorbeeld als zetmeel) en deze uiteindelijk weer vrij te geven tijdens het rijpen. Het doel van dit proefschrift was om nieuwe manieren om het suikergehalte te verhogen met behulp van gerichte mutagenese met CRISPR/Cas9 te onderzoeken. Expressiemodulatie door promotormutaties, maar ook de manipulatie van post-translationele regulatie, evenals gentargeting werden hiervoor onderzocht.

Hoofdstuk 2 onderzocht de regulatie van het suikermetabolisme in tomaat, met de nadruk op genen die veelbelovend zijn voor bewerking met CRISPR/Cas om het suikergehalte in het fruit te verhogen. Het identificeren en begrijpen van de rol van de genen die ten grondslag liggen aan de productie en distributie van suikers is cruciaal voor de veredeling van gewassen. Het is des te belangrijker vanwege de snelle ontwikkeling van nieuwe veredelingstechnologieën die nauwkeurige modificatie van deze genen mogelijk maken. De combinatie van synthese, transport, stofwisseling en opslag bepaalt het uiteindelijke suikergehalte in tomaat. In dit hoofdstuk heb ik "Quantitative Trait Loci" (QTL's, Kwantitatieve kenmerken loci) voor het suikergehalte en hun onderliggende genen uitgelicht.

Daarnaast besprak ik nieuwe genen of locaties op het genoom waarvan recentelijk is ontdekt dat ze invloed hebben op het suikergehalte en kunnen worden ingezet in de veredeling. Sommige van deze genen hebben verder onderzoek nodig omdat deze niet allemaal zijn gevalideerd voor een algeheel positief effect (in tomaat), vooral wanneer negatieve neveneffecten moeten worden meegewogen. De combinatie van QTL-ontdekking en moleculaire analyse van de suikerroute(s) heeft een breed scala aan bewezen of veelbelovende doelen opgeleverd voor het verhogen van het suikergehalte van tomaat.

Uit dit review werd duidelijk dat zowel *ADP-Glucose Pyrofosforylase Large subunit 1 (AGPL1)* als *Cell Wall INvertase 5 (LIN5)* essentiële genen zijn voor suikeraccumulatie.

AGPL1 codeert voor de grote subeenheid van ADP-glucosepyrofosforylase (AGPase). Dit tetramere AGPase-eiwitcomplex is betrokken bij de zetmeelsynthese in plastiden tijdens de vruchtontwikkeling. LIN5 hydrolyseert sucrose tot glucose en fructose. De gecreëerde suiker gradiënt die ontstaat door deze omzetting is een belangrijke drijvende kracht voor de import van suikers in de vrucht (de "sink"). Beide genen liggen ten grondslag aan een eigen QTL voor Brix, waarbij beide genen een allel in een wilde tomatensoort hebben dat een positief effect op het suikergehalte heeft. Eén van de doelstellingen van dit proefschrift was het bestuderen en wijzigen van de transcriptionele regulatie van *AGPL1* en *LIN5* om hun expressie te verhogen. Ik heb de *Cis*-regulerende Elementen (CREs) bestudeerd die betrokken zijn bij de regulatie van deze twee genen. Eén benadering was het identificeren van CREs en hun interacterende transcriptiefactoren (TFs) met behulp van verschillende *in silico*- en experimentele proeven. Tegelijkertijd heb ik gemultiplexte CRISPR/Cas9 mutagenese toegepast om variatie in beide promotorens te creëren. Door het effect op de expressie van *AGPL1* en *LIN5* in mutanten te bestuderen, wilde ik nieuwe soorten CRE-functionaliteiten in een *in vivo* systeem ontdekken. De hypothese was dat wanneer een mutatie optrad in een CRE, dit de binding van een TF zou verstoren. Als de verstoorde plaats de bindingsplaats is voor een repressor, zou de expressie van de twee doelgenen naar verwachting toenemen. Door de expressie van *LIN5* en *AGPL1* te verhogen, probeerde ik tegelijkertijd de tomatensmaak te verbeteren omdat het suikergehalte dan hoger zou worden.

Een *in silico* analyse van de *AGPL1* promotor in **Hoofdstuk 3** resulteerde in de identificatie van geconserveerde en toegankelijke promotor regio's en yeast-one-hybrid (Y1H) experimenten identificeerden TFs die interactie vertoonde met deze promoter. Ik heb het effect van deze TFs op *AGPL1*-transcriptie bestudeerd met promotor-reporter-testen. Twee potentiële repressors (FRUITFULL 2 en de CCCH-type Zinc Finger C3H13) en verschillende activators (Abscisic Acid Responsive Elements-Binding Factor 1 (AREB1), de homoloog van de *Arabidopsis* B-box zinkvinger TF BBX19, de homoloog van de *Arabidopsis* GATA-motief met TF GATA9, Jasmonic Acid 1 (JA1), Nuclear Factor-YA10 (NF-YA10), de homoloog van de *Arabidopsis* Telomerase Repeat Binding Factor-Like 3 (TRFL), en het WRKY-motief met WRKY24, WRKY41 en WRKY81 waren gevonden. Daarnaast gebruikte ik gemultiplexte mutagenese met CRISPR/Cas9 op de *AGPL1*-promotor in de tomaat cultivar Moneyberg om de expressie van *AGPL1* in de vrucht te veranderen en planten te creëren met een hoger suikergehalte. De Brix was inderdaad verhoogd in verschillende van de verkregen mutantlijnen, maar trad altijd gelijktijdig op met een verlaagd vruchtgewicht. Voor twee van de drie lijnen met de hoogste Brix werd ook een toename van *AGPL1*-expressie aangetoond.

Hoofdstuk 4 was gericht op *LIN5*. Y1H en promotor-reporter-testen onthulden verschillende TFs die interactie vertoonde met de *LIN5*-promotor. GATA9, de homoloog van paprika MYB48, R2R3MYB58, MYB76 en een NAC TF gedroegen zich als repressors, terwijl C3H13, NF-YA10 en een homoloog van paprika NPY1 de expressie

in de promotor-reporter-testen verhoogden. Net als voor *AGPL1* maakte ik promotormutanten en vond verhoogde Brix in verschillende van de gegenereerde mutantlijnen. Verschillende hiervan werden gekarakteriseerd voor *LIN5*-expressie, waar ik veranderde expressie in de mutanten in vergelijking met het wildtype (een ongemuteerde lijn) kon bevestigen. In een promotor-reporter-test waarbij de mutant promotors werden vergeleken met een wildtype promotor, werden HAT1, MYB HYH en WRKY24 geïdentificeerd als zwakke activatoren van de promotors van de mutant.

Het werk aan de promotors van *LIN5* en *AGPL1* leverde aanknopingspunten op voor de verbetering van de tomatenkwaliteit en gaf ons inzicht in het potentieel van genoombewerking voor het veranderen van genexpressie (hoger of lager) in gewassen door zich te richten op kandidaatgen-CREs. Mijn conclusie is dat deze benadering haalbaar is, hoewel het groeien en fenotyperen van mutanten ruimte- en arbeidsintensief is, zeker wanneer de promotor op willekeurige wijze wordt getarget voor mutagenese. Het verkleinen van het doelgebied door het voorspellen van veelbelovende promotorregio's zou het aantal benodigde mutanten drastisch kunnen verlagen.

In **Hoofdstuk 5** heb ik de rol van *Cell-wall Inhibitor of β-fructosidase (CIF1* of *INVINH1*), een post-translationele remmer van invertases zoals LIN5, onderzocht. Aangezien een eerder rapport een positief effect aantoonde van een verlaagde *CIF1* expressie door RNAi op het suikergehalte, genereerde ik volledige knock-out-mutanten van *CIF1* en bestudeerde de effecten op het uiteindelijke gehalte aan oplosbare vaste stoffen (Brix) en invertase-activiteit. De CRISPR/Cas-gegenereerde knock-outmutanten hadden inderdaad een verhoogde Brix, maar hun vruchtgrootte nam ook aanzienlijk af. Bovendien vertoonden de mutanten geen consistent verband tussen de voorspelde verminderde *CIF1*-activiteit als gevolg van knock-outmutaties en verhoogde *LIN5*-activiteit in een enzymtest.

Hoofdstuk 6 richtte zich op het uitbreiden en verbeteren van de CRISPR/Cas9 toepassingen door gemultiplexte Gene Targeting toe te passen op *LIN5*. *LIN5* is het onderliggende gen van de Brix QTL, Brix9-2-5, waarbij het allel van *LIN5* dat afkomstig is uit *Solanum pennellii* een hoge Brix geeft. Aangenomen wordt dat drie aminozuursubstituties in dit *S. pennellii*-allel verantwoordelijk zijn voor een actiever LIN5-eiwit. Mijn doel was om deze drie aminozuurveranderingen in onze variëteit, cv. Moneyberg, te verkrijgen door gelijktijdig op alle drie de doelaminozuren te richten via homologe recombinatie met een van een geminivirusl replicon afgeleid donorsysteem. Door middel van Gene Targeting experimenten in protoplasten en via stabiele transformaties en regeneratie ontdekte ik dat de aanpak resulteert in succesvolle verandering van de drie doelaminozuren, zij het met een zeer lage efficiëntie (1-3%).

Concluderend onderzocht dit proefschrift de verschillende mogelijkheden van CRISPR/Cas9-mutagenese om de smaak van tomaten te verbeteren. Het richten op

CRE's om de expressie te verhogen en het gebruik van op een geminivirus gebaseerde gentargeting waren beide haalbaar. De frequenties van succes moeten echter worden verbeterd voor een bredere toepassing, vooral bij een commerciële applicatie.

ACKNOWLEDGMENTS

The acknowledgements, the last but certainly not the least chapter to write. So many of you have made my PhD an amazing journey! I have often felt I had one of the best jobs in the world, which was in no small part due to my amazing colleagues. Here, I want to take some time to thank all of you! Although just a few pages are by far not enough to express all the gratitude and joy I am feeling while writing this section.

First, I would like to thank my supervisor, Ruud. Where to begin! After our first meeting, I was very excited to work with you. Together we wrote the proposal, using the hot new topic CRISPR/Cas combined with improved flavour in tomato as the target. I like to think this combination helped us to win the grant. The years that followed never had me regretting the choice to work with you. Your open-door, endless knowledge and a healthy dose of patience have really helped me these last couple of years. Most importantly, you always had positive encouragement for me. While at the same time making comments on every slide and paragraph off course. But these only improved my presentation and writing skills.

Gerco, thank you for welcoming me to the wonderful group of Plant Developmental Systems (PDS). You are an excellent leader of PDS. I am very grateful for all our progress-meetings, which kept me on track. You gave me a lot of opportunities to explore the things I wanted, both scientific and otherwise, which has helped me to develop both as scientist and as a person!

Michiel and Lena, I couldn't do without you as my paranymphs! Michiel, always ready to help. As my neighbour in the lab, we had many a talk over the most trivial matters to dampen the monotony of some of the work, such as measuring Brix. Thank you for all your explanations, remembering the most trivial numbers (you are way faster than Elab), patience and all your jokes. Lena, we were neighbours, horse-riding buddies, council-colleagues, co-editors and many more things. Thank you for every moment of talking about work and other stuff while sipping tea or Aperol Spritz.

Thank you to all my students. Without you, my PhD journey would have been a lot less interesting and fun. Most importantly, I would have never managed to write this thesis without all your work! Christine, always cheerful, looking fabulous and never complaining about all the tissue culture you did. Marrit, you are very bright, and I hope our time together has shown you that you have every reason to be highly confident in yourself! Douwe, I am still cherishing the plant you gave me at the end of your thesis and hope that it will have blooming future, as I am sure your future will be too. Jeroen, as the last of the "Busschers" to arrive in our lab, you lived up to the name! Julia, through tough personal times, I greatly admired your work-ethic and cheerfulness. Moniek and Giorgio, both of you started at worst possible moment. Your first day was also the first

day of the Covid-quarantine! Nevertheless, both of you showed great enthusiasm, flexibility and worked hard, with the results to show for it.

I had the pleasure of being part of the "tomato club' at PDS. Rufang, what a delight it has been to be your colleague and friend! Scientifically, I thank you for teaching me everything about plant transformation, tomato crossings and DNA isolation. Thank you for the countless times you have taken out a culture for me on the weekend. Your ability to work so fast, for instance in grinding samples for RNA extraction, has been a great opportunity for me to take on the challenge of becoming faster myself - often without successfully outcompeting you. I thank you even more for making it a joy to work at PDS. Ellen, our super smart CRISPR expert! Thank you for helping me with the protoplast work, without you, this thesis would have one chapter less. Jin, thank you for your patience with showing me a protoplast transfection and lending me all your buffers. Thank you, Julia, for all the nice discussions on CRISPR-possibilities. Marian thanks for your answers on all my questions, for bringing new energy to the tomato group and for being so open, calm, and patient with all us PhDs. Xiaobing, your fun jokes in the lab and our (endless) negotiations for greenhouse space, which always turned out good, made many days more interesting! Gul, thank you for your wonderful Christmas wishes last year! They really warmed my heart. Iris, Kai, Victor and Chris, the future of the tomato club! Thank you for all the discussions on the transcription factor techniques and off course for the *Frietjes Stoofvlees*.

Mieke, Tjitske, Marco, Jacqueline and Froukje thank you all for keeping the PDS lab going. You are the – not so secret – power behind every of our experiments! Keep up the good work, even when the growth chambers are a mess (again), everything is finished (again), and gel-combs are lying in the sink (again). Jan, thank you being my buddy in advocating the benefits of Elab, as well as for your seed-related advice. For this latter, I thank Steven as well! Marie-Jose, Maria, and especially Hana thank you for all the administrative support! Wilma, Richard, Kim, and Martijn "de beste", thank you for all your advice and comments! You have made me do better.

Those that came before me at PDS, Leonie, Susanne, Anneke, Sam (my desk-neighbour!) and Suraj, you have been great examples, you have given me loads of tips, and perhaps most importantly, you have set an example for me and continued the great social and scientific environment of PDS. Suraj especially, thank you for all our fun game-evenings, amazing Indian dinners, and endless advice.

The remaining PhDs and Post-docs, all together the beating heart of PDS. Charlotte, Baojian, Mengfan, Suze, Francesca, Judit, Amalia, Annemarie, Manjunath and Amit, with the addition of PDS guests Patricia, Sylvia, Han and Tati. I cannot stress enough how all of you, as part of the "open-workspace-crew", have contributed to the joy of doing a PhD in PDS. Our amazing lunch-club (with the best shrimp fried rice from Xiaobing and mussels from Han) has been sorely missed in Covid-times. Thank you all for all the

lunches, coffees, lengthy discussion on any topic, random talks in the lab, dinners, hot-pots, Sinterklaas celebrations and ice-skating sessions. Together we could share joy and lament about failed experiments. Every day in the lab or office was a good one thanks to all of you!

Geurt, Sean, Maarten, Teus and Henk thank you all for the excellent plant care, as well as your flexibility in dealing with all our crazy experiments! Thank you, Arnaud for giving me advice and motivation as my external supervisor, Ep for patiently explaining about tomato phenotyping, Rumyana for all your advice, especially on tobacco infiltrations and Renze for the many glowmax analyses you finished for me! Hanny and Jeroen, a big thank you for the help with sugar measurements! Ingrid, thank you for being a wonderful and involved leader of the Bioscience Business Unit. On many occasions, you made me proud to be a part of Bioscience.

Remco and Cheryl, I want to thank you for encouraging me to sign up for the EPS Graduate Program, which led to me doing this PhD. Your supervisions during my master Thesis and confidence in me have prepared me perfectly.

Damian, Nikita, Juriaan, and William, while you might not exactly have contributed to me finishing this thesis, it was a wonderful experience to lay the foundations of GeneSprout together with you. I am glad we started the initiative and hopefully, with our help, CRISPR/Cas assisted mutagenesis will be allowed soon in plant breeding.

I would also like to thank the EPS PhD Council and WPC members. Together, we have made a little contribution to make the life better of all PhDs! It was great to work beside you. A special thanks to Ingrid, Susan, Lars, Ivo, Tieme (my amazing co-chair!), Valerie (also, my amazing co-chair), Stuart, Daan, Michelle, Hao, Daniel, Jasper, Octavina, Mandy, Davy, Irene and Sietkse.

My family, Marcel, Rosanne, Jorel, Aloïs and most of all my mother, Yvonne, thank you all for all your support and confidence in me! I highly appreciate your interest in my topic and all the discussions we had on the benefits of CRISPR/Cas in plant breeding.

Then there are some people of whom I doubt will read this. That has been the whole point! You have been the perfect escape from my work, helped me keep a clear mind and gave me the energy to keep on going. After a long, sometimes unsuccessful day in the lab, it has been a joy, every single time, to put on my horse-riding pants and boots and go to the stable. Thank you, Carli, Willemijn, Naomi, Jasper, Nick, Biba, Maxime, Natasja, Myriam, Anne-Marie, Susanne, Monique, Mirthe, Lieve, Lara, Anneke, Herrie, Lotte and so many more! Most thankful, I am of Fien, both a little angel and a little devil, with whom I have flown many a time through the forest to our combined delight.

I have saved the best for last: Joost. My pillar and rock! You have shown endless support. Nothing was too much: moving with me to Wageningen and now to Rotterdam, cooking many dinners if my experiments or writing got out of hand and joining a variety of activities with my colleagues. Not to mention helping me with (bio)-informatics related issues and doing proof-reading of my chapters. Thank you, you are the best and I am lucky to have you in my life.

ABOUT THE AUTHOR

Vera Veltkamp was born on the 27th of March in 1991, in Nijmegen, the Netherlands. After few years there, she and her family moved to Crolles, France, where they lived for five years. After a brief time in the Netherlands, the family moved to Singapore for a period of three years. The time abroad was highly important for the foundation of interest in biology. Exploring the mountains in France or swimming between the mangroves in Malaysia, she loved it all. Her teacher in the final grade of elementary school already predicted that she would become a biologist.

After graduating from the *Stedelijk Gymnasium Nijmegen*, she did not heed her teacher's advice though, and started studying Industrial Design at the *TU Delft* in 2010. After successfully completing her first year, she acknowledged though that her true passion lied with living beings. Thus, she switched to study Biology at *Leiden University*. As the bachelor progressed, she had the option to focus on a particular field of Biology, microbiology, and with this possibility her enthusiasm grew. From the very first lecture she had attended on Plant Biology and Breeding, it was clear for her that this was one of the most exciting and relevant topics in Biology! She continued with a Master Plant Biotechnology in Leiden, where she got to do a thesis on the auxin pathway during somatic embryogenesis of *Arabidopsis* with Cheryl Philipsen and Remco Offringa, and an internship in the Seed Research department of *Rijk Zwaan*, de Lier. She finished her Master in 2016 Cum laude.

During her master thesis, Cheryl and Remco encouraged Vera to take part in the EPS Graduate program, a program where talented master students are allowed to explore their own ideas and write a PhD proposal. This was the perfect opportunity to write a proposal on a subject she had grown to be passionate about, flavour! She found the perfect research group to write this proposal with in Plant Developmental Systems at *Wageningen University*. The proposal was written in collaboration with Gerco Angenent and Ruud de Maagd. The NWO funded the proposal and the work performed based on that proposal resulted in the publication of this thesis. After finalizing her work on the thesis, she found the perfect opportunity to continue in the field as a (junior) tomato breeder at *Totam Seeds*.

Education Statement of the Graduate School

Experimental Plant Sciences

The Graduate School EXPERIMENTAL PLANT SCIENCES

Issued to: Vera Veltkamp
Date: 17 September 2021
Group: Plant Developmental Systems (Bioscience)
University: Wageningen University & Research

1) Start-Up Phase	date	cp
► **First presentation of your project**		
Enjoying the fruits of knowledge: cracking the Cis-regulatory code of gene regulation	2 Mar 2017	1,5
► **Writing or rewriting a project proposal**		
► **MSc courses**		
Subtotal Start-Up Phase		*1,5*

2) Scientific Exposure	date	cp
► **EPS PhD student days**		
Get2Gether 2017, Soest, NL	9, 10 Feb 2017	0,6
Get2Gether 2018, Soest, NL	15, 16 Feb 2018	0,6
Get2Gether 2019, Soest, NL	11, 12 Feb 2019	0,6
Get2Gether 2020, Soest, NL	10, 11 Feb 2020	0,6
► **EPS theme symposia**		
EPS theme 1 symposium 'Developmental Biology of Plants', Leiden, NL	28 Feb 2017	0,3
EPS theme 1 symposium 'Developmental Biology of Plants', Wageningen, NL	30 Jan 2018	0,3
EPS theme 1 symposium 'Developmental Biology of Plants', Leiden, NL	31 Jan 2019	0,3
EPS theme 4 symposium 'Genome Biology', online	11 Dec 2020	0,2
► **Lunteren Days and other national platforms**		
Annual Experimental Plant Sciences meeting, Lunteren, NL	10, 11 Apr 2017	0,6
Annual Experimental Plant Sciences meeting, Lunteren, NL	9, 10 Apr 2018	0,6
Annual Experimental Plant Sciences meeting, Lunteren, NL	8, 9 Apr 2019	0,6
Annual Experimental Plant Sciences meeting, GatherTown, online	12, 13 Apr 2021	0,5
► **Seminars (series), workshops and symposia**		
Seminar: Dr. Sotirios Fragkostefanakis, Alternative splicing of a heat stress transcription factor mediates thermotolerance in tomato	2 Nov 2016	0,1
Seminar: Prof Lars Ostergaard, Fruit shape determination in Brassicaceae	26 Dec 2016	0,1
Seminar: Dr. Gerben van Ooijen, Clocks across taxa	29 May 2017	0,1
Seminar: Prof.Dr. U. Wyss, highlights of hidden insect-worlds	2 Oct 2017	0,1
Seminar: B-Wise Seminar Katy Wolstencroft & Dennis van Muijen	3 Oct 2017	0,2
Seminar: Acces to Plant Data	11 Dec 2017	0,1
Seminar: B-wise Seminar Anton Feenstra & Ehsan Motazedi	9 Jan 2018	0,2
Seminar: CRISPR-Cas - from evolution to revolution	8 Mar 2018	0,1
Seminar: Ronald Snijder (Syngenta): Modern domestication of Pelargonium in a commercial environment	9 May 2018	0,1
Seminar: Dr. Victoria Mironova, What we learnt about auxin transcriptional regulation from meta-analysis of whole genome data	27 Jun 2018	0,1
Seminar: Rosanna Petrella, Transcriptional and epigenetic regulation of STK during flower development in Arabidopsis	13 Nov 2018	0,1
Seminar: Prof. Wolf Frommer, Logistics: Allocation of carbon and energy for yield and pathogen resistance	17 Jan 2019	0,1
Seminar: Plantae presents: Zach Lippman and Arjun Khakhar	9 Dec 2020	0,2
Workshop: Masterclass 100th Dies: Maurits de Klepper - Flavor perception, Wageningen, NL	9 Mar 2018	0,1
Workshop: Plants and Patents, Wageningen, NL	21 Oct 2019	0,2
Symposium: 1st Symposium "WURomics: Technology-driven innovation for plant Breeding, Wageningen, NL	15 Dec 2016	0,3
Symposium: KeyGene "Keys Symposium - Boosting Crop Producitivty", Wageningen, NL	14 Jun 2017	0,2
Symposium: KeyGene & GENALICE: Solutions for agrigenomic Big Data Challenges, Wageningen, NL	30 Nov 2017	0,2
Symposium: Science Week 'What is life' and 'Editing the genome', Wageningen, NL	12, 14 Mar 2018	0,3
Symposium: Networking event of TKI Horticulture & Propagation Materials (TKI TU), Nieuwegein, NL	3 Apr 2018	0,3
Symposium: Future of Food (100 years WUR), Wageningen, NL	22, 23 Jun 2018	0,4
Symposium: When Wizards meet Prophets, Wageningen, NL	29 Mar 2019	0,3
Symposium: COGEM 'Gene edited crops; global perspectives and regulation', First Chamber, the Hague, NL	10 Oct 2019	0,3
► **Seminar plus**		
► **International symposia and congresses**		
10th European Plant Sciences Retreat, Utrecht, NL	3-6 Jul 2018	1,0
CRISPRcon 2019 - Conversations on science, society and the future of gene editing, Wageningen, NL	20, 21 Jun 2019	0,6
XVI Solanaceae Conference: Yield and Nutrition, Jerusalem, Israel	15-19 Sep 2019	1,5

	1st PlantEd Conference (COSTaction), Novi Sad, Serbia	5-7 Nov 2019	0,9
	SOL, Solanaceae conference, International Online Meeting 2020	9-11 Nov 2020	0,6
►	**Presentations**		
	Presentation: Enjoying the fruits of knowledgeCracking the Cis-regulatory code of gene regulation, TKI TU 2018	3 Apr 2018	1,0
	Poster: 10th European Plant Sciences Retreat	3-6 Jul 2018	1,0
	Presentation: CRISPR/Cas9: toepassing in plantenonderzoek en -veredeling, Fascination of Plants Day 2018, Utrecht, NL	18 May 2018	1,0
	Presentation: Enjoying the fruits of Knowledge: a sweeter tomato, XVI Solanaceae Conference	15-19 Sep 2019	1,0
	Presentation: 1st PlantEd Conference (COSTaction), Novi Sad, Serbia	5-7 Nov 2019	1,0
►	**3rd year interview**		
►	**Excursions**		
	EPS Company Visit at KeyGene 2017	12 Oct 2017	0,2
	EPS Company Visit at Dümmen Orange 2018	15 Jun 2018	0,3
	EPS Company Visit at Nunhems (BASF) 2019	25 Oct 2019	0,2
	Subtotal Scientific Exposure		*20,3*

3) In-Depth Studies	*date*	*cp*
► **Advanced scientific courses & workshops**		
Transcription Factors and Transcriptional Regulation, Wageningen, NL	12-14 Dec 2016	1,0
Basic R course, Wageningen, NL	8-9 May 2017	0,6
The Power of RNA-seq, Wageningen, NL	11-13 Jun 2018	0,9
Linux basic course, Wageningen, NL	16 Aug 2018	0,1
Gentle hands-on introduction to Python programming, online	2, 3 Jul 2020	0,6
► **Journal club**		
Literature Discussion group at Plant Developmental Systems (Bioscience)	Oct 2016 - Oct 2020	3,0
► **Individual research training**		
Subtotal In-Depth Studies		*6,2*

4) Personal Development	*date*	*cp*
► **General skill training courses**		
PhD Competence Assessment, Wageningen, NL	19 Apr 2017	0,3
Course: EPS Introduction Course, Wageningen, NL	16 Feb 2017	0,3
Course: Project and Time Management, Wageningen, NL	Oct-Dec 2017	1,5
Course Scientific Artwork - Vector graphics and images, Wageningen, NL	15, 16 Apr 2019	0,6
Course: Scientific Writing, Wageningen, NL	Sep-Dec 2019	1,8
Workshop: Negotiation, Wageningen, NL	28 Jun 2018	0,1
Workshop: Masterclass - Stratego for Women, Wageningen, NL	9 Oct 2018	0,2
Workshop: WGS PhD Workshop Carousel, Wageningen, NL	24 May 2019	0,3
Workshop: Young WUR: David McCandless - Data Visualization, online	24 Nov 2020	0,3
Symposium: Publish for Impact, Wageningen, NL	7 Feb 2017	0,2
► **Organisation of meetings, PhD courses or outreach activities**		
Outreach: Support at Fascination of Plants Day Event, The Hague	19 May 2017	0,0
Outreach: Support at Fascination of Plants Day Event, The Hague	17 may 2019	0,0
Organisation: Get2Gether 2019	Mar 2018 - Mar 2019	1,5
Organisation: Get2Gether 2020	Mar 2019 - Mar 2020	1,5
Organisation: side event GeneSprout at CRISPRcon 2019	20 Jun 2019	0,0
Organisation: Plants & Patents Workshop	21 Oct 2019	0,0
► **Membership of EPS PhD Council**		
Council member EPS PhD Council (including Secretary Wageningen PhD Council)	Mar 2018 - Mar 2020	1,4
Chair EPS PhD council	Jul 2018 - Sep 2019	1,4
Subtotal Personal Development		*11,4*

5) Teaching & Supervision Duties	*date*	*cp*
► **Courses**		
► **Supervision of BSc/MSc students**		
Christine de Vries (Internship)	4 Sep 2017- 22 Jan 2018	0,0
Marrit Alderkamp (Internship)	5 Feb 2018 - 22 Jun 2018	0,0
Douwe Zantinge (MSc Thesis)	3 Sep 2018 - 29 Mar 2019	3,0
Jeroen Busscher (Internship)	4 Feb 2019 - 22 Jun 2019	0,0
Julia Ruiz Cappella (MSc Thesis)	2 Sep 2019 - 28 Feb 2020	0,0
Moniek Schippers (BSc Thesis)	16 Mar 2020 - 29 Jun 2020	0,0
Giorgio Gullotta (MSc Thesis)	11 May 2020 - 26 Oct 2020	0,0
Subtotal Teaching & Supervision Duties		*3,0*

TOTAL NUMBER OF CREDIT POINTS*		**42,4**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS with a minimum total of 30 ECTS credits.

**A credit point (cp) represents a normative study load of 28 hours of study.*

This research was financially supported by an NWO-EPS Graduate School Master talent grant with co-financing and support of Plantum, ENZA zaden Research & Development B.V., Hazera Seeds B.V., Rijk Zwaan Breeding B.V., Bejo Zaden B.V., Syngenta Seeds B.V.

Financial support from Wageningen University for designing and printing this thesis is gratefully acknowledged

Cover design by Evelien Jagtman | evelienjagtman.com

Thesis layout by the author

Printed by proefschriftmaken.nl