Breeding for higher yield in tomato by yield dissection and quantitative trait loci for yield components

Yutaka Tsutsumi-Morita

Propositions

1. A large effect of a trait on yield in an experiment with one genotype does not make this trait a candidate for high-yielding breeding.
   (this thesis)

2. Several genes reducing the greenness of leaves improve biomass production at crop level.
   (this thesis)

3. Genomic prediction has more power than CRISPR-Cas to improve yield.

4. Strong emphasis on journal impact factors when evaluating researchers threatens diversity of research.

5. The academic system discriminates against financially vulnerable people.

6. An international public fund is needed to compensate farmers for yield reduction caused by unpredictable natural disasters.

Propositions belonging to the thesis, entitled:
Breeding for higher yield in tomato by yield dissection and quantitative trait loci for yield components

Yutaka Tsutsumi-Morita
Wageningen, 7 February 2023

Breeding for higher yield in tomato by yield dissection and quantitative trait loci for yield components

Yutaka Tsutsumi-Morita

Thesis committee

Promotors

Prof. Dr F.A. Van Eeuwijk
Professor of Applied Statistics
Wageningen University & Research

Prof. Dr L.F.M. Marcelis
Professor of Horticulture and Product Physiology
Wageningen University & Research

Co-promotor

Dr E. Heuvelink
Associate professor of Horticulture and Product Physiology
Wageningen University & Research

Other members

Prof. Dr R.G.F. Visser, Wageningen University & Research
Dr M. Causse, INRAE, Montfavet, France
Prof. Dr B. Zwaan, Wageningen University & Research
Dr X. Yin, Wageningen University & Research

This research was conducted under the auspices of the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC).

Breeding for higher yield in tomato by yield dissection and quantitative trait loci for yield components

Yutaka Tsutsumi-Morita

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 Tuesday 7 February 2023
at 1:30 p.m. in the Omnia Auditorium.

Yutaka Tsutsumi-Morita

Breeding for higher yield in tomato by yield dissection and quantitative trait loci for yield components,
315 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2023)
With references, with summary in English.

ISBN 978-94-6447-498-5
DOI https://doi.org/10.18174/581306

To Tomiko, my grandmother

Abstract

Yield improvement is an important target for plant breeding. To improve crop yield, it is proposed to introduce genetic variation from wild genetic resources into elite cultivars. However, identifying the genetic loci that increase yield, and the traits that drive this improvement remains challenging. The objective of this thesis is to determine in which direction — via which component traits by using which quantitative trait loci (QTLs) — yield of an elite parent can be improved, and identify the phenotypic and genotypic ideotypes for a high yield in indeterminate tomato by combining yield dissection models and QTLs.

Two tomato-breeding trials were conducted: one in the Netherlands and one in Spain. Each trial tested a particular tomato breeding population developed for each production system — hybrids obtained from crossing around 350 recombinant inbred lines (RILs) derived from multiple parents with two testers. QTLs for yield and yield component traits were detected by using a mixed model approach, where identity-by-descent (IBD) probabilities served as genetic information.

Yield was first dissected from two perspectives — (1) a harvest dissection model that captures yield from fruit-sink capacity and (2) a biomass dissection model capturing yield from biomass production (total plant biomass), allocation (harvest index), and dilution of fruit dry matter by water (fruit fresh-dry weight ratio). The harvest dissection model revealed that the number of fruits (FN) and individual fruit fresh weight (FFW) showed a trade-off (negative correlation between FN and FFW; mostly different QTL parental allele effects between FN QTLs and FFW QTLs at the same position) in both trials. Thus, to develop high-yielding lines, both traits need to be investigated simultaneously. In the biomass dissection model, fruit fresh-dry weight ratio correlated with yield in the Dutch trial but not in the Spanish trial. Nevertheless, the same strategy to breed for high-yielding lines results from the Dutch and Spanish trials; yield improvements can be more attributed to the number of fruits than individual fruit fresh weight, more attributed to total fruit dry weight than fruit fresh-dry weight ratio, and more attributed to total biomass than harvest index.

Total biomass was further dissected into adapted light use efficiency (LUE_{adp}) and the fraction of intercepted light (F_{int}) in the Dutch trial. F_{int} was calculated from the leaf area index, and LUE_{adp} was calculated as total biomass divided by F_{int}. LUE_{adp} and F_{int} showed no correlation with each other, and most of their QTLs were unlinked. Therefore, high-total-biomass lines can be produced by improving both LUE_{adp} and F_{int} or each one of them separately. Wild parents have a large potential to improve the total biomass of the elite parent. Some of the high-total-biomass lines achieved a higher yield than the elite parent.

LUE_{adp} is thought to be primarily determined by photosynthesis. QTLs for photosynthesis-related traits were detected in the Dutch trial. None of the photosynthesis-related traits

correlated with yield, total biomass, or LUE_{adp}. On the other hand, photosynthesis-related QTLs were often at the same position as QTLs for yield, total biomass and/or LUE_{adp}, although the QTL allele effects were not always consistent between photosynthesis-related traits and yield. Nevertheless, combining QTLs for all photosynthesis-related traits resulted in a good yield prediction accuracy.

This thesis demonstrates that wild genetic resources possess beneficial alleles to improve the yield of the elite parent. Yield dissection models with QTLs for these yield components provide a framework for evaluating the effect of introgressed wild germplasm on component traits and subsequently on yield. Plant breeders can design high-yielding lines by the preferable composition of component traits (phenotypic ideotype) by assembling positive alleles for component traits (genotypic ideotype).

Contents

Chapter 1

General introduction

1.1. Importance of improving yield by incorporating genetic variation from wild germplasm

1.1.1. Yield needs to be improved

Yield improvement is one of the most important targets for plant breeding (Duvick *et al.*, 2004; van Eeuwijk *et al.*, 2019). Yield per unit area has increased during the past decades due to several factors, including improved genotypes (cultivars) and management (Duvick *et al.*, 2004; Ortiz *et al.*, 2008; de Gelder *et al.*, 2012). Studies using both modern and historic cultivars have shown that there has been long-term genetic improvement in maize (Duvick *et al.*, 2004) and tomato (Higashide and Heuvelink, 2009). To feed the increasing human population by 2050, the rate of annual yield increase per unit area for major crops must be more than double compared to the current level (Tester and Langridge, 2010; Ray *et al.*, 2012, 2013; Hickey *et al.*, 2017, 2019).

1.1.2. Importance of increasing genetic diversity

To further improve crop yield, it is important to make use of genetic diversity present in wild relatives (Tanksley and McCouch, 1997). Genetic variation had decreased during crop domestication, followed by the initial plant breeding until the last century. Thus, plant breeding has unintentionally reduced the genetic diversity in cultivated germplasm, on which plant breeding based (Tanksley and McCouch, 1997) until the middle of the last century (Schouten *et al.*, 2019). Now, the natural genetic diversity is commonly recognized as the main source of adaptation and crop improvement (Huang and Han, 2014; McAusland *et al.*, 2020). Although wild relatives show low yield at the phenotypic level, recent studies showed that re-introducing genetic variation from wild relatives improved several traits including yield (Gur and Zamir, 2004; Ortiz *et al.*, 2008; Trethowan and Mujeeb-Kazi, 2008; Reynolds *et al.*, 2012; Rothan *et al.*, 2019; Schouten *et al.*, 2019; Schulthess *et al.*, 2022).

1.2. Challenges in evaluation of high-yielding lines

While yield can be improved by introducing genetic variation from wild germplasm into the elite line, the evaluation of such high-yielding lines is still challenging (van Eeuwijk *et al.*, 2019). Yield is a complex trait, which is an emergent outcome of the interaction between many underlying physiological processes and morphological characteristics. Accordingly, yield is controlled by a large number of genes — each with generally only small effects (Dekkers and Hospital, 2002; Eeuwijk *et al.*, 2010). If a gene has a large effect (i.e., for a qualitative trait, which is controlled by the small number of genes), this gene would be easily found by mutant analysis. However, for a quantitative trait (controlled by many genes and therefore the phenotypes of individuals form a continuous distribution), the mutant analysis

does not work because other genes can mask the lack of effect of the mutated gene. Therefore, for a quantitative trait, we try to find an important genomic region (quantitative trait locus; QTL) by statistically comparing phenotype and genotype. Molecular marker technologies and the construction of genetic maps enable the dissection of the variation in quantitative traits into the effects of quantitative trait loci (QTLs) (Paterson *et al.*, 1988; Tanksley *et al.*, 1992). Furthermore, recent advancements in high-throughput and cost-effective genotyping technologies enable genomic selection, which can potentially accelerate the rate of genetic gain in crops (Meuwissen *et al.*, 2001; Heffner *et al.*, 2009; Voss-Fels *et al.*, 2019). However, breeding for high yield remains challenging due to the interaction between genotype and environment (G × E). That is, the high-yield ranking of genotypes is dependent on the environment, for example, year or site (Cooper and Hammer, 1996; Chapman *et al.*, 2003; Tardieu, 2003, 2012; Chenu *et al.*, 2009, 2018; Messina *et al.*, 2011; Cooper *et al.*, 2014; van Eeuwijk *et al.*, 2019). Therefore, the stable selection for high-yielding lines remains difficult.

1.3. Combining physiological knowledge with genetics

1.3.1. Selecting for yield itself

Plant breeders have increased crop yield potential largely through empirical selection for yield *per se* (Figure 1A) (Yin *et al.*, 2000*b*). However, yield needs to be increased further and more rapidly (Ray *et al.*, 2013), which requires new strategies.

1.3.2. Selecting for yield via its component traits

Donald (1968) defined the *"(phenotypic) ideotype"* concept based on multiple traits: "a theoretical biological model which is expected to perform or behave predictably within a defined environment", and more specifically, "a plant model which is expected to yield a greater quantity or quality of grain, oil or other useful product when developed as a cultivar". Plant breeders have tried to identify the ideotypes for higher yields for specific target environments. For example, the ideotype of rice (*Oryza sativa*) in the 1960s was a short semi-dwarf rice cultivar, which did not lodge under increased nitrogen supply (Chandler *et al.*, 1969). With the ideotype concept, breeders try to select for yield by selecting for a set of high-priority traits.

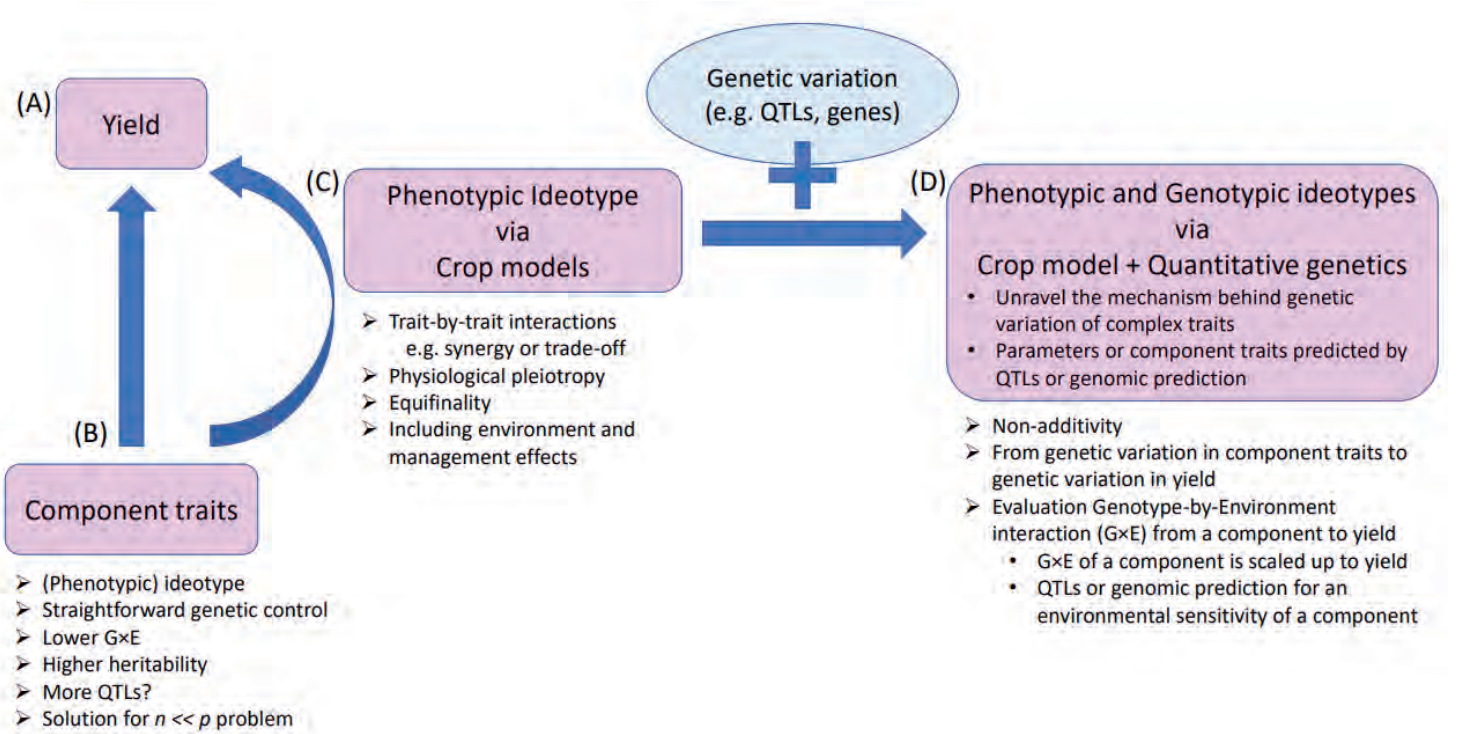


Figure 1 Selection for high yielding genotypes: (A) directly on yield itself, (B) by component traits, (C) using a crop model, and (D) by combining crop model with quantitative genetics.

Selecting for component traits — yield components, which compose yield — rather than for yield itself (Figure 1B) has certain advantages. First, simpler component traits might be under *partially or wholly independent genetic control* (Yin *et al.*, 2000*b*) and may be more directly linked to quantitative trait loci (QTLs) and genes (Fukai *et al.*, 1999; Price and Courtois, 1999; Tardieu, 2003; Hammer *et al.*, 2005, 2006; Chenu *et al.*, 2017). Second, component traits are relatively less context dependent, i.e., they have *lower G×E* than yield (Hammer *et al.*, 2006; Technow *et al.*, 2015; Messina *et al.*, 2018; Cooper *et al.*, 2021). Component traits have narrower genotype-phenotype gaps, which results in *higher heritability* (Lamsal *et al.*, 2018) (heritability is defined as the proportion of phenotypic variance explained by genotypic variance in a breeding population). Consequently, *more QTLs* may be detected for component traits compared to yield (Yin *et al.*, 2002). An additional merit of using component traits is to alleviate the *n << p problem* in genomic prediction. For modern plant breeding, the number of markers (p) is far larger than the number of genotypes (n), which causes fitting problems for models of phenotypes as functions of markers. Thus, penalization of marker effects is required. Additional phenotypes enhance the chances of identifying biologically relevant markers (Crossa *et al.*, 2017; Ramstein *et al.*, 2019; Washburn *et al.*, 2020).

1.3.3. Exploring phenotypic ideotypes by using crop models

Although selecting for a component trait is useful, a component trait is often not independent from other traits. To consider *trait-by-trait relationships* (e.g. *synergy* or *trade-off* among component traits), crop models are useful (Figure 1C). A crop model assembles mathematical algorithms to represent crop growth and development processes as a whole system. It translates input information (environment and management) to yield as an emergent outcome via underlying physiological processes (Yin *et al.*, 2004). Some trait-by-trait relationships are caused by "physiological pleiotropy", which can be understood from a carbon or nitrogen balance (Boote *et al.*, 2021). For example, increased leaf-level photosynthesis rate is often associated with increased specific leaf weight (kg m^{-2}) or specific leaf nitrogen (Boote *et al.*, 2021; Yin *et al.*, 2022). Thus, the photosynthetic rate per unit area may change but when expressed by unit leaf mass or nitrogen mass, it might not be necessarily increased (Boote *et al.*, 2021). *Environmental effects* can be incorporated into crop models by sensitivities of physiological and morphological processes to each environmental factor (Boote *et al.*, 2021). With the aid of crop models, some studies attempted to design ideotypes adapted to specific environments (Kropff *et al.*, 1995; Martre *et al.*, 2015; Génard *et al.*, 2016; Quilot-Turion *et al.*, 2016); other studies analyzed past genetic improvement from experimental data to propose plant ideotypes (Suriharn *et al.*, 2007, 2011; Peng *et al.*, 2011; Boote *et al.*, 2021). Additionally, crop models help to demonstrate *equifinality*; that is, the same yield can be achieved by different compositions of component traits (Lamsal *et al.*, 2018).

1.3.4. Combining crop models with genetics: Unravel the mechanism behind the genetic variation of complex traits

While component analysis and ideotype exploration are useful for breeding high-yielding varieties, they often do not consider the existing *genetic variation* (Hammer *et al.*, 2002; Chenu *et al.*, 2017). Knowing the possible genetic variation will prevent a crop model from simulating unrealistic results, which are outside the range of genetic feasibility (Boote *et al.*, 2021).

Initially, crop models were developed in the 'School of de Wit' as generic models with no reference to species (Bouman *et al.*, 1996; Boote *et al.*, 2021). Incorporating genetic variation in some parameters in a crop model was first attempted by agronomists (White and Hoogenboom, 1996). Currently, some models are species-specific models with no reference to genotypes, but many current crop models include *cultivar-specific parameters* (also called 'genetic coefficients') to represent genetic differences (Boote *et al.*, 2021). These cultivar-specific parameters can be further modelled as *functions of genes or QTLs*. For example, cultivar-specific parameters of a crop model were estimated by linear functions of the known six genetic loci related to photoperiod sensitivity and determinacy (White and Hoogenboom, 1996, 2003; Messina *et al.*, 2006). This attempt to incorporate genetic effects into a parameter causes a shift in these parameters to reflect genetic variation (Boote *et al.*, 2001; Hammer *et al.*, 2002).

Incorporating genetic variation into crop models made it possible for crop models to be used for plant breeding purposes (Figure 1D) (Yin *et al.*, 1999*b*; Chapman *et al.*, 2003; Hammer *et al.*, 2004; Chenu *et al.*, 2009). For example, genetic variation for five-component traits was included in a crop model to predict yield in maize breeding (Cooper *et al.*, 2016).

Non-additivity. Common methodologies of QTL analysis and genomic prediction handle additive genetic effects. Although additive genetic effects can accelerate genetic gain in plant breeding, further improvement is possible by including non-additive effects, namely dominance (intra-locus non-additivity) and epistasis (inter-locus non-additivity) (Voss-Fels *et al.*, 2019). Epistasis is defined by interactions between alleles at different genes or QTLs that affect trait values (Cooper *et al.*, 2021). Some aspects of epistasis have been investigated as trait-by-trait interactions, therefore via crop models (Chapman *et al.*, 2003; Yin *et al.*, 2004; Messina *et al.*, 2011, 2018; Technow *et al.*, 2015; Voss-Fels *et al.*, 2019; Cooper *et al.*, 2021). For example, two traits show a trade-off due to carbon limitation, which might cause epistasis between QTLs for these two traits.

From variation in component traits to variation in yield. Combining crop models with QTLs or genomic prediction is often used to quantify the influence of genetic variation in a trait or parameter on yield variation (Yin *et al.*, 2000*a*, 2002). That is, genetic variation in a trait can

be scaled up to yield variation via crop models, which makes it possible to evaluate how much genetic gain in a component trait transfers to genetic gain in yield (Tardieu and Tuberosa, 2010). In other words, this methodology can quantify the effects of genes or alleles of traits on yield (Cooper *et al.*, 2021).

Evaluation of genotype-by-environment interaction (G×E) from a component to yield. Crop models combined with genetics can also help evaluation of G×E. Crop models include environmental responses of traits or parameters, which enables the evaluation of G×E for each trait, and this can subsequently be scaled up to G×E for yield (Chapman *et al.*, 2003; Technow *et al.*, 2015; Messina *et al.*, 2018).

If the environmental response of a trait is genetic, this environmental sensitivity can be a target for QTL analysis or genomic prediction. For example, in maize, QTLs were detected for three parameters of leaf elongation rate and for one parameter of anthesis-to-silking interval (Reymond *et al.*, 2003). Three parameters of leaf elongation rate were the sensitivities to temperature, vapour pressure deficit, and soil water potential. These parameters were incorporated in a crop model to predict leaf area expansion and grain set, which were subsequently scaled up to maize yield under drought stress (Chenu *et al.*, 2008, 2009). Using such QTLs for environmental sensitivities to model leaf elongation rate avoids complex QTL-environment interactions, which are commonly observed for directly measured traits such as leaf length (Reymond *et al.*, 2004; Chenu *et al.*, 2009).

1.4. Physiological breeding

Once breeders know component traits that can be beneficial for yield, these beneficial component traits can be used for making a cross for a future cultivar and/or in choosing parent lines (Cooper *et al.*, 2014; van Eeuwijk *et al.*, 2019). If a positive genetic effect on a trait can be straightforwardly scaled up to a positive genetic effect on yield, then the selection of high-yielding lines would be relatively simple. In reality, however, there is a trade-off between traits in some cases. For example, the trade-off between seed size and seed number was reported in grain crops (Sadras, 2007). Genetic dissection of the trade-off between these components revealed some progenies, for which the trade-off was missing. This lack of trade-off between components led to extremely high yields in these lines, making these loci good targets for marker assist selection (Bustos *et al.*, 2013; Griffiths *et al.*, 2015; Reynolds and Langridge, 2016). Thus, in so-called ‘Physiological breeding’ parents with different complex but complementary traits are crossed to achieve cumulative gene action for yield (Reynolds *et al.*, 2012; Reynolds and Langridge, 2016).

1.5. High-throughput phenotyping

Genetic dissection of yield into simpler traits with often higher heritability provides the possibility of measuring these underlying traits in high-throughput phenotyping facilities (Tardieu and Tuberosa, 2010; Araus and Cairns, 2014; Cabrera-Bosquet *et al.*, 2016; Rutkoski *et al.*, 2016; Araus *et al.*, 2018). Although high-throughput field phenotyping remains a bottleneck for future plant breeding advances (Araus and Cairns, 2014; Araus *et al.*, 2018), high-throughput phenotyping has a great potential to increase the efficiency of plant breeding programs. It allows for the measurement of a large number of phenotypes at low cost, with less labour demand and at high accuracy (Reynolds and Langridge, 2016; Rutkoski *et al.*, 2016). In addition, if a trait shows a high correlation with yield, this trait can be used as an early selection trait (van Eeuwijk *et al.*, 2019).

1.6. Multi-parent population

Although natural genetic diversity is the main source of adaptation and crop improvement (Huang and Han, 2014), genetic linkage mapping studies traditionally have been conducted especially on bi-parental populations (Arrones *et al.*, 2020). In bi-parental populations (consisting of a set of individuals derived from crosses between two inbred parents (Tripodi, 2021)), genetic diversity is limited. In addition, genetic recombination events are limited, which limits the resolution for QTL detection. Furthermore, the QTLs detected in a bi-parental population might not be expressed in other genetic backgrounds (Arrones *et al.*, 2020). In practice, these genetic resources developed for bi-parental populations are not directly relevant to the breeding community, which limits effective implementation (Cavanagh *et al.*, 2008).

Recently, multi-parent populations have received more attention (Scott *et al.*, 2020; Arrones *et al.*, 2020; Li *et al.*, 2021). Multi-parent populations are produced by crossing more than two inbred founder lines. For example, diallel (Giraud *et al.*, 2017; Turner *et al.*, 2018), Multi-parent Advanced Generation Inter-Cross (MAGIC) (Cavanagh *et al.*, 2008; Huang *et al.*, 2015; Pascual *et al.*, 2015; Arrones *et al.*, 2020), and Nested Association Mapping (NAM) (Yu *et al.*, 2008; McMullen *et al.*, 2009) populations. These populations may provide solutions for the drawbacks of linkage mapping (mainly traditional bi-parental crosses that include limited genetic variability) and association mapping (genome-wide association study: GWAS) approaches (Arrones *et al.*, 2020), for which population structure can lead to spurious associations if not dealt with in an appropriate manner (Yu *et al.*, 2006). The merits of multi-parent populations are the inclusion of wider genetic and phenotypic diversity and high power and resolution for mapping QTLs (Scott *et al.*, 2020; Arrones *et al.*, 2020). Multi-parent populations are closer to practical breeding populations, where many parents have been involved.

1.7. Tomato

1.7.1. Importance of tomato for industry and research

Tomato (*Solanum lycopersicum*) is the most consumed vegetable worldwide after potato (*Solanum tuberosum L.*). It has become an important food in many countries (Causse *et al.*, 2020). Tomato has several advantages as a model plant. It is self-pollinated, diploid (2n = 2x = 24) with a medium-sized genome (approximately 900 Mb) (Rothan *et al.*, 2019). A high-quality reference genome sequence was published in 2012 (Tomato Genome Consortium *et al.*, 2012). It can be transformed with a high success rate. A lot of genetic and genomic resources are available and access to several databases for genetic and phenotypic data exists (Rothan *et al.*, 2019). Thus, the tomato is a model plant for plant genetics, pathology, physiology, fruit development and ripening (Causse et al., 2020; Heuvelink et al., 2020a). Tomato is also a model for the *Solanaceae* family, which includes other important crops, such as potato, pepper (*Capsicum annuum L.*), and eggplant (*Solanum melongena*).

In tomatoes there are both indeterminate and determinate types. Indeterminate tomatoes are cultivated in greenhouses. Plants produce new leaves, flower trusses and fruits continually resulting in (almost) year-round production. Indeterminate tomatoes are mainly consumed as fresh vegetables. Determinate tomato is usually grown in the open field. Initiation of new trusses stops after a few trusses and all fruits ripen roughly at the same time making single machine harvest possible. Determinate tomato is normally grown for processing, such as use in paste, juice, and ketchup (Costa and Heuvelink, 2018). “Indeterminate” or “determinate” is regulated by the *SELF PRUNING (SP)* gene (ortholog of Arabidopsis *TERMINAL FLOWER1*). *SP* reverts “determinate” processing tomato varieties back to the “indeterminate” vine-like sympodial growth that defines wild tomato species (Pnueli *et al.*, 1998; Lippman *et al.*, 2007).

1.7.2. Tomato history of genetic diversity

Tomato originates from the Andes region and has undergone two domestication steps during its history (Blanca *et al.*, 2012, 2015; Lin *et al.*, 2014). Although wild relatives have a large genetic diversity (reviewed by (Bai and Lindhout, 2007), genetic variation in cultivated tomatoes strongly reduced during their domestication and early breeding practices (Miller and Tanksley, 1990; Sim *et al.*, 2009; Blanca *et al.*, 2012). In tomato breeding after about 1940, wild relatives have been deliberately introgressed into cultivated tomatoes for desired traits (Rick and Chetelat, 1995; Bai and Lindhout, 2007). The patterns of single nucleotide polymorphisms (SNP) variation also indicates that modern breeding practices have reintroduced genetic variation into the tomato crop from wild species as a result of introgression to obtain disease resistances, different fruit sizes, and flavour components (Sim *et al.*, 2009; Schouten *et al.*, 2019).

1.7.3. Tomato breeding population and quantitative genetics analysis

At the end of the 1980s, the first tomato genetic maps of molecular markers were constructed to dissect quantitative traits into QTLs (Paterson *et al.*, 1988; Tanksley *et al.*, 1992). In early breeding practices, it was difficult to identify polymorphisms between elite germplasms because of the lack of genetic variation in the cultivated species (Miller and Tanksley, 1990; Sim *et al.*, 2009). Therefore, most of the mapping populations had been based on interspecific crosses between a cultivar and a related wild species, e.g. from the *lycopersicon* group (as reviewed by (Foolad, 2007; Labate *et al.*, 2007; Grandillo *et al.*, 2011), *lycopersicoides* (Pertuzé *et al.*, 2002), and *juglandifolia* group (Albrecht *et al.*, 2010). In modern plant breeding, wild germplasms were re-introduced into the elite materials, which has made it possible to identify polymorphisms between elite materials (Sim *et al.*, 2009; Schouten *et al.*, 2019). Therefore, quantitative genetics studies (QTL analysis, genomic prediction) of tomatoes currently can be conducted on populations developed from wild germplasms, landraces, and/or cultivated genotypes (Pascual *et al.*, 2016; Rothan *et al.*, 2019).

Historically, many tomato studies have been conducted on bi-parental populations (for example, (Muir *et al.*, 2014; Albert *et al.*, 2016). In the last decade, genome-wide association studies (GWAS) also have been conducted in tomatoes (Sauvage *et al.*, 2014; Ruggieri *et al.*, 2014; Sacco *et al.*, 2015; Yamamoto *et al.*, 2016; Phan *et al.*, 2019). However, the number of studies on multi-parent populations is limited. A multi-parent advanced generation inter-cross (MAGIC) population from eight parents was developed (Pascual *et al.*, 2015) and QTLs were reported (Pascual *et al.*, 2015, 2016; Diouf *et al.*, 2018; Bineau *et al.*, 2021). Another MAGIC population was developed, on which QTL studies were planned (Campanelli *et al.*, 2019). QTL mapping on a cross of two commercial F1 hybrids (therefore four founders) was also reported, for which recombinant inbred lines (RILs) were developed (Ohyama *et al.*, 2017).

Until now, numerous genes and QTLs underlying traits of interest have been reported (Labate *et al.*, 2007; Foolad and Panthee, 2012; Causse and Grandillo, 2016; Grandillo and Cammareri, 2016). Major breeding goals of tomatoes are productivity, resistance to biotic and abiotic stresses, and fruit sensorial and nutritional quality (Rothan *et al.*, 2019).

1.8. Knowledge gap

Combining crop models with genetics is a promising approach to increasing the genetic gain of yield. This methodology has been used in several studies to predict yield from QTLs or genomic prediction for some important underlying traits, e.g. time to flower, pod-set and maturity of soybean (Messina *et al.*, 2006), leaf expansion rate under drought stress (Chenu *et al.*, 2008, 2009) or leaf photosynthesis (Gu *et al.*, 2014*a*). It is important to recognize that the models need to improve the explanatory capabilities to scale up QTL or genetic effect of

the component to the effect on yield, before QTL or genetic effects can be incorporated into crop models (Yin *et al.*, 1999*b*,*a*; Hammer *et al.*, 2002; Reymond *et al.*, 2003; Chenu *et al.*, 2008, 2009; Boote *et al.*, 2021). Here, an important consideration for successful application is the decision of which traits should be used within the crop model to represent the genetic variation for yield (Cooper *et al.*, 2016). These target component traits or parameters are usually predetermined based on prior knowledge, and there is no clear methodology on how to determine which traits' variations should be included in the crop models to represent the genetic variation in yield.

1.9. This thesis

Introgression of wild germplasm into the elite parent produces genetic variation in component traits. The genetic variation in a component trait does not necessarily cause genetic variation in yield. Thus, the first important question is whether genetic variation in component traits brings genetic variation in yield and which components are important. If we identify influential component traits, we can incorporate the genetic variation for these component traits into a crop model to predict yield.

The wide genetic variation in component traits and yield is useful for QTL detection and yield prediction. However, if positive allele effects always arise from elite genotypes, this QTL detection and prediction are less useful to breed a high-yielding line. Fortunately, several studies showed that positive alleles also come from wild genotypes in some QTLs or genes (Gur and Zamir, 2004; Ortiz *et al.*, 2008; Trethowan and Mujeeb-Kazi, 2008; Reynolds *et al.*, 2012; Rothan *et al.*, 2019; Schouten *et al.*, 2019; Schulthess *et al.*, 2022). This means that a higher phenotypic value of a component trait compared to the elite parent leads to a higher yield compared to the elite parent. Thus, we can define the phenotypic ideotype for a higher yield. This phenotypic ideotype must be achieved by combining positive alleles, which come from not only elite parents but also wild parents. If we look at component traits, we can identify a higher number of positive alleles than looking at only yield. Therefore, more QTLs will be available to combine positive alleles to breed a higher-yielding line. Furthermore, if we consider trait-by-trait relationships via yield dissection models, we understand how we can effectively breed a high-yielding line. For example, if a component trait has no trade-off with other traits and is positively correlated with yield, the positive alleles for the component trait can be simply used to breed a high-yielding line. On the other hand, if two component traits show a trade-off, positive alleles of these component traits need to be carefully selected to breed a high-yielding line.

1.9.1. Objective

The objective of this thesis is to find in which direction yield of the elite parent can be improved, and identify the phenotypic and genotypic ideotypes for a high yield in

indeterminate tomato. The research questions of this thesis are categorized in three groups (phenotype, QTL analysis and methodology).

A. Phenotype
   (1) To assess which traits underlie the phenotypic variation in yield, and therefore should be incorporated into a crop model to predict yield
   (2) To identify what is the phenotypic ideotype for high yield
B. QTL analysis
   (1) To detect QTLs for yield and yield component traits
   (2) To assess to what extent do QTLs for component traits appear as QTLs for the target trait (i.e., yield or total biomass), and *vice versa*
   (3) To provide information to formulate the genotypic ideotype for a high yield by combining positive alleles of QTLs for component traits
   (4) To assess whether wild parents can bring some positive alleles for component traits and whether these positive alleles of components can lead to a higher yield
C. Methodology
   (1) To demonstrate the usefulness of combining yield dissection models with QTLs for high-yielding breeding
      - To systematically find an influential component trait via yield dissection models
      - To recognize how to combine positive alleles from different component traits
   (2) To investigate whether component traits are more stable and show lower G×E than yield

I used tomato as a model species. Experiments were conducted in the Netherlands and Spain. For each country, a recombinant inbred line (RIL) population of indeterminate tomato was developed from multiple parents − four parents (two elite parents, two wild parents) for the Dutch breeding population, three parents (one elite parent, two wild parents) for the Spanish breeding population. These were the industrial breeding populations developed by a breeding company, a project partner. The wild germplasm introgressed into the elite genetic background causes a large variation in component traits and yield (= fruit yield in tomato case), which allowed me to investigate the contribution of wild alleles to component traits and yield. Experiments were conducted on hybrids between RILs and test parents similar to commercial varieties.

1.9.2. Thesis outline

This thesis consists of six chapters (Figure 2). This **first chapter** serves as a general introduction and provides the thesis aim and research questions.

In **Chapter 2**, objectives A, B, and C(1) were investigated in an experiment in the Netherlands by using two yield dissection models combined with QTLs. An experiment was

conducted in the Netherlands by using a breeding population developed for Dutch tomato production. Yield dissection models were designed from two perspectives; one from the fruit-sink capacity (harvest dissection; Figure 2A) and the other from a combination of biomass production (total biomass), its allocation to fruits (harvest index), and dilution of fruit dry weight by water (fruit fresh-dry weight ratio) (biomass dissection; Figure 2B).

In **Chapter 3**, objectives A, B, C(1) and C(2) were investigated in two experiments in Spain by using two yield dissection models combined with QTLs, similar to the Dutch experiment. Two experiments were conducted in different locations in Spain. Both experiments tested the same breeding population developed for Spanish tomato production. The usefulness of methodology combining yield dissection models with QTLs in Chapter 2 was validated. Phenotypic characteristics and QTLs were compared between Spanish and Dutch experiments. Two experiments in different greenhouses in Spain enabled to examine the objective C(2) — to investigate whether component traits show a lower G×E than yield.

In **Chapter 4**, objectives A, B, and C(1) were investigated related to total plant biomass production in the same experiment as Chapter 2, in the Netherlands. Dissection strategy continues from Chapter 2; total biomass was further dissected into the light interception and light use efficiency (Figure 2C). From phenotypic and QTL analyses, I investigated how to breed a high-biomass line and how it is related to a high-yielding line.

In **Chapter 5**, objectives A and B were investigated related to photosynthesis-related traits in the same experiment as Chapter 2, in the Netherlands. I identified QTLs for photosynthesis-related traits and examined the effect of photosynthesis-related traits on light use efficiency, total biomass, and yield. I discussed the possible role of photosynthesis-related traits for plant breeding to improve yield (Figure 2D).

Chapter 6 is a general discussion, using the obtained knowledge in Chapter 2 till 5 to answer the research questions of Chapter 1, discussing my contribution to scientific progress as well as the implications of my work for plant breeding and providing suggestions for further research.

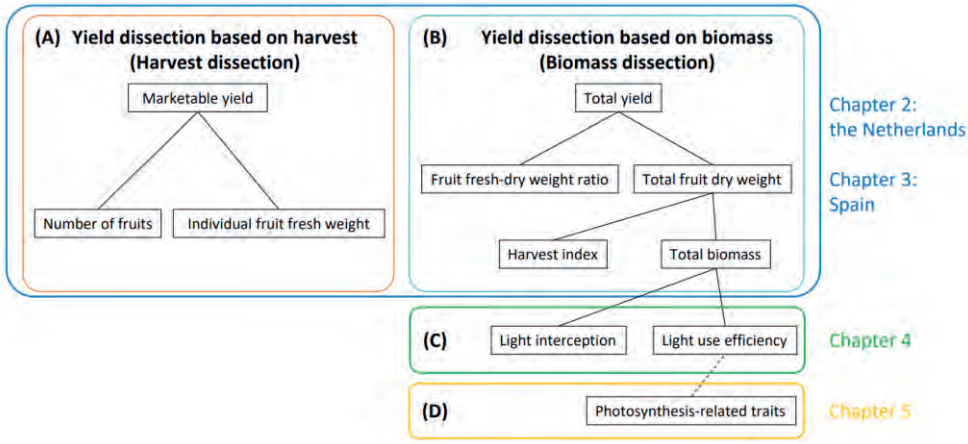


Figure 2 Thesis composition. Yield dissection models based on (A) harvest and (B) biomass, which were used in Chapter 2 (the Dutch trial) and Chapter 3 (the Spanish trial). (C) Chapter 4 dissected total biomass into the light interception and light use efficiency. (D) Chapter 5 examined the effects of photosynthesis-related traits on light use efficiency, total biomass and yield.

Chapter 2

Yield dissection models to improve yield: a case study in tomato

Yutaka Tsutsumi-Morita[1,2], Ep Heuvelink[1], Sedighehsadat Khaleghi[1], Daniela Bustos-Korts[2], Leo F.M. Marcelis[1], Kim M.C.A. Vermeer[3], Hannelore van Dijk[3], Frank F. Millenaar[3], George A.K. Van Voorn[2], Fred A. Van Eeuwijk[2]

[1] Wageningen University & Research, Horticulture and Product Physiology, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[2] Wageningen University & Research, Biometris, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[3] BASF Vegetable seeds – Nunhems, Napoleonsweg 152, 6083 AB Nunhem, The Netherlands

Published in In Silico Plants (2021), *3*(1), 1–18.
https://doi.org/10.1093/insilicoplants/diab012

Abstract

Yield as a complex trait may either be genetically improved directly, by identifying QTLs contributing to yield, or indirectly via improvement of underlying components, where parents contribute complementary alleles to different components. We investigated the utility of two yield dissection models in tomato for identifying promising yield components and corresponding QTLs. In a harvest dissection, marketable yield was the product of number of fruits and individual fruit fresh weight. In a biomass dissection, total yield was the product of fruit fresh-dry weight ratio and total fruit dry weight. Data came from a greenhouse experiment with a population of hybrids formed from four-way RILs. Trade-offs were observed between the component traits in both dissections. Genetic improvements were possible by increasing the number of fruits and the total fruit dry weight to offset losses in fruit fresh weight and fruit fresh-dry weight ratio. Most yield QTLs colocalized with component QTLs, offering options for the construction of high-yielding genotypes. An analysis of QTL allelic effects in relation to parental origin emphasized the complementary role of the parents in the construction of desired genotypes. Multi-QTL models were used for the comparison of yield predictions from yield QTLs and predictions from the products of components following multi-QTL models for those components. Component QTLs underlying dissection models were able to predict yield with the same accuracy as yield QTLs in direct predictions. Harvest and biomass yield dissection models may serve as useful tools for yield improvement in tomato by either or both of combining individual component QTLs and multi-QTL component predictions.

Keywords: Complementation; four-way RIL; indeterminate tomato; multi-parent population; QTLs; yield components; yield dissection; yield prediction.

2.1. Introduction

Continued crop yield improvement is needed to keep pace with the food demand by the increasing world population (Tester and Langridge, 2010; Ray *et al.*, 2013; Hickey *et al.*, 2019). Yield is the complex, integrated outcome of interactions between multiple genetic factors, environmental inputs and management conditions and decisions over time. Yield is a strongly context dependent trait subject to many types of genotype by management by environment (G×M×E) interactions (Cooper and Hammer, 1996; Chapman *et al.*, 2003; Tardieu, 2003, 2012; Chenu *et al.*, 2009, 2018; Messina *et al.*, 2011; Cooper *et al.*, 2014). To reduce these context dependencies, it is suggested to dissect yield in physiologically simpler traits, yield components, that are relatively less context dependent, i.e. that show less G×M×E, following physiological models for yield that create G×M×E as an emergent property from inputs defined by basic, upstream physiological traits and management and environmental factors (Tardieu, 2003; Hammer *et al.*, 2006; Chenu *et al.*, 2009; Cooper *et al.*, 2009; Messina *et al.*, 2011; Bustos-Korts *et al.*, 2019). The success of alternative yield dissections as a starting point for yield improvement strategies by breeding depends strongly on the feasibility of identifying a set of genetic factors, or quantitative trait loci (QTLs), driving phenotypic variation in the yield components. When a limited (QTL mapping) or exhaustive (genomic prediction) set of genetic markers allows the successful prediction of yield components, yield for any genotype and environmental condition can be predicted from marker profiles and environmental inputs. For recent examples of this approach, see Technow et al. (2015), Bustos-Korts et al. (2019) and Millet et al. (2019).

Another reason to try to dissect complex traits in component traits is to address the question of equifinality. Equifinality at a physiological level occurs when different combinations of component traits give rise to the same yield in particular M×E conditions. At a genetic level, equifinality occurs when different multi-locus QTL genotypes produce similar yields. It is obvious that writing yield as a function of simpler component traits, provides extra opportunities to improve yield by identifying yield component configurations that produce equal yields in some M×E conditions, but different yields in other M×E conditions. Moving to the QTL level, colocalizations of yield and component QTLs, open the way to more effective marker-assisted selection strategies. Yield QTLs can contribute positively to yield under one type of M×E conditions, while contributing negatively to yield under other M×E conditions (Boer *et al.*, 2007; Chenu *et al.*, 2009; Tardieu, 2012; Bustos-Korts *et al.*, 2019). Dissection of yield in components will shed light on why a particular yield QTL shows contrasting behaviour across changing M×E conditions. Following up on yield dissection, it will be important to identify and quantify trade-offs between yield components in relation to the M×E conditions and formulate ideotypes for higher yielding genotypes (Yin and Struik, 2008; Chenu *et al.*, 2009, 2017). Reynolds and Langridge (2016) present a good overview of theory and practice of physiological dissection approaches to breeding.

Ramstein et al. (2019) add an interesting aspect to the discussion about the relevance of dissection approaches by observing that the creation of additional phenotypes for a set of genotypes alleviates part of the n<<p problem in the fitting of genotype-to-phenotype models, where *n* is the number of genotypes and *p* is the number of markers. For many modern breeding populations, the number of markers is far larger than the number of genotypes, so that for fitting models of phenotypes as functions of markers, penalization of marker effects is required. Additional phenotypes enhance the chances of identifying biologically relevant markers.

In predictive breeding, the ultimate aim is the prediction of yield within acceptable bounds of uncertainty for new genotypes in new environments (Cooper *et al.*, 2014; Yin *et al.*, 2016; Washburn *et al.*, 2020). A more modest objective looks at a static genetic dissection of yield into upstream components that are less integrated (less complex) and practically measurable or that can be approximated by use of a phenotyping tool, and for which a strong genetic basis can be identified (Alimi et al., 2013a). In this way, trade-offs and negatively correlated components may become visible that hamper breeding efforts to improve yield based on selection for yield itself. At the same time, the use of QTL analysis on the component trait data may lead to the identification of alleles that make positive (or negative) contributions to yield components and hence to yield itself. The construction of a genotype containing several of such positive QTL alleles for different yield components may result in directed yield improvement. It can be visualized by a response surface for yield in which we search for the global maximum (Cooper *et al.*, 2014). When component traits are assessed under controlled conditions in a well-defined management regime, repeatable and predictable G×M interactions may be exploited while unpredictable G×E and G×M×E interactions may be kept small. This constitutes an ideal scenario for improving yield by selecting for promising combinations of yield components.

The main objective of the current paper is to investigate two yield dissections and ways in which these dissections can increase the efficiency of breeding efforts. We use a static yield dissection into simpler components and then identify QTLs for these components and compare the component QTLs to the QTLs for yield itself in terms of location and effects. The immediate questions we like to answer or start answering are:

- What is the phenotypic and genetic variation for yield and its components?
- To which extent do we observe transgression (offspring exceeds performance of parents) in yield and component traits?
- To which extent is yield driven by its components in different dissections?
- Do we observe correlations and trade-offs between component traits that may impair the use of a particular dissection model?
- Can we identify the genetic basis (QTLs) of yield and its component traits and predict traits from their QTLs?

- To what extent do yield and component QTLs colocalize?
- Can we understand trade-offs between yield components in terms of linked and pleiotropic QTLs?
- Can efficient yield improvement strategies be formulated based on yield components?

In this paper, we consider two static yield dissections: one based on harvest, and one based on biomass production. Dissection based on harvest considers yield as the product of number of harvested organs and average organ size and is a standard dissection proposed and discussed in many papers (e.g. barley (Yin *et al.*, 2002), maize (Peng *et al.*, 2011), pea (Timmerman-Vaughan *et al.*, 2005), melon (Zalapa *et al.*, 2007)). In the dissection based on biomass production, yield is the product of total biomass production, harvest index and fruit fresh-dry weight ratio, all yield components in the component hierarchy presented by Higashide and Heuvelink (2009); Gur et al. (2010); Ronga et al. (2017, 2019).

We demonstrate our dissection approach on a population of hybrids of indeterminate tomato (*Solanum lycopersicum*) bred for greenhouse production in the Netherlands, grown and sampled during 2016. Tomato is one of the most important vegetable crops worldwide, and it is the model plant for *Solanaceae* (Heuvelink et al., 2020b), for which genetic and genomic resources are available (Tomato Genome Consortium *et al.*, 2012). The tomato hybrids came from a cross between 342 four-way recombinant inbred lines (RILs) and two testers. The parent lines for the four-way RILs contained two elite lines and two wild-type lines. The four-way cross population was not specifically created for yield improvement, but instead presented a spectrum of genetic material with contributions from all parents. The wild-type material introduced genetic and physiological interactions that allowed us to investigate the contribution of wild yield component alleles to yield in elite genetic backgrounds.

2.2. Materials and methods

2.2.1. Yield dissection models

We considered two static yield dissections: one based on harvest, and another one based on biomass production. They are given in Figure 1.

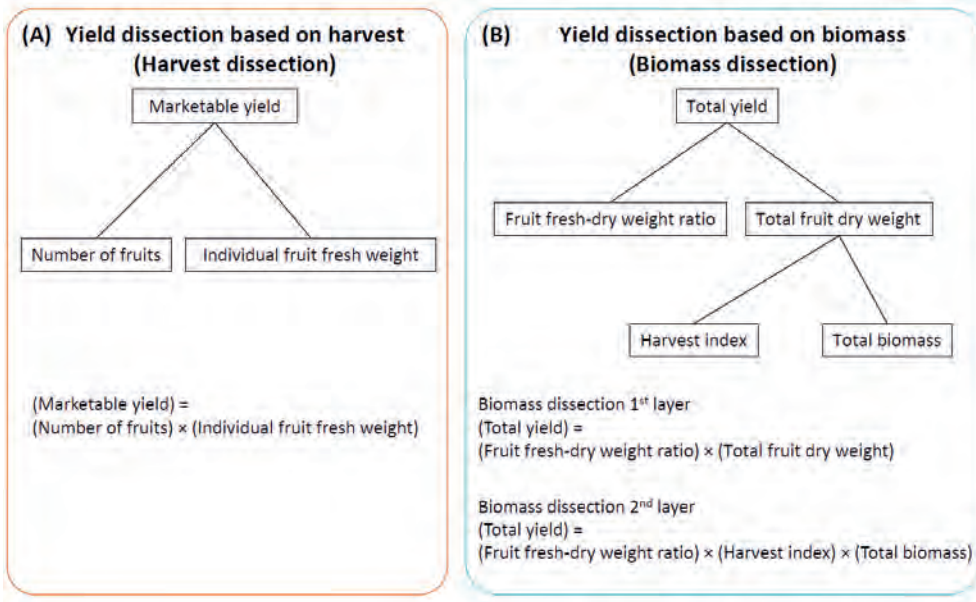


Figure 1 Dissection models for yield. (A) Harvest dissection and (B) Biomass dissection.

The dissection based on harvest (Figure 1A) is straightforward, yield is the number of harvested organs times their individual weight:

$$Marketable\ yield = (Number\ of\ fruits) \times (Individual\ fruit\ fresh\ weight) \quad Eq.\ 1$$

This is a standard dissection in which the most elementary limitation to yield formation is included, where the increase in one component may not necessarily result in an increased total yield because of underlying trade-offs and negative correlations (Yin *et al.*, 2002; Sadras, 2007; Yin and Struik, 2008; Gambín and Borrás, 2010; Sadras and Lawson, 2011; Bustos *et al.*, 2013; Slafer *et al.*, 2014). In contrast, the dissection of biomass (Figure 1B) is based on the component hierarchy of Higashide and Heuvelink (2009). It considers yield as the product of fruit fresh-dry weight ratio and total fruit dry weight

$$Total\ yield = (Fruit\ fresh - dry\ weight\ ratio) \times (Total\ fruit\ dry\ weight) \quad Eq.\ 2$$

The total fruit dry weight can be further dissected into harvest index (HI) and total biomass

$$Total\ fruit\ dry\ weight = (Harvest\ index) \times (Total\ biomass) \quad Eq.\ 3$$

Root weight is not considered in the calculation of harvest index and total biomass, because the measurement of roots is impractical for obvious reasons. Combining equations (2) and (3) gives

$$Total\ yield = (Fruit\ fresh - dry\ weight\ ratio) \times (Harvest\ index) \times (Total\ biomass) \qquad Eq.\ 4$$

By having a static dissection it is (implicitly) assumed that the fractions fruit fresh-dry weight ratio and harvest index are fixed during the growth season and constant within a genotype.

2.2.2. Breeding population

The tomato population used in this experiment consisted of indeterminate tomato hybrids. Two wild-types (W1 and W2) were each crossed to an elite parent line (E1). The F1s of these two wild-type × elite crosses were then crossed with each other. The F1×F1 offspring population was crossed with another elite line (E2), and the resulting offspring population was selfed to produce a four-way RIL population. These four-way RILs were combined with two testers (TP1, TP2), producing two populations of 342 genotypes (RILs × TP1 and RILs × TP2) (Figure 2). Although the phenotypic observations were done on hybrids, the QTL analyses relate the phenotypes to markers in the RILs. In the rest of the paper, we will for convenience discuss results in relation to both RILs and hybrids, in an exchangeable way.

The four-way RILs were genotyped with a set of 279 SNP SOLCAP markers with minor allele frequency > 0.0125 [(Sim *et al.*, 2012), http://solcap.msu.edu/tomato_genotype_data.shtml]. We used a genetic map that was based on the publicly available map, TraitGenetics EXPIMP2012: https://solgenomics.net/cview/map.pl?map_version_id=149. Together with the map, the markers formed the basis for identity by descent (IBD) calculations between parents and offspring that served to create design matrices for genome-wide testing of QTL additive effects. The calculation of IBD probabilities was done via a Hidden Markov Model by the Mathematica package RABBIT as described by (Zheng *et al.*, 2015).

Following these IBD calculations, we found that the DNA in the four-way RIL population contained a 28 % contribution from E1, 52 % from E2, 12 % from W1 and 8 % from W2. The crossing scheme in Figure 2 suggests that without selection, we should expect contributions of 25 % from E1, 50 % from E2 and 12.5 % from W1 and W2. Therefore, some selection against the wild-type alleles appeared in the construction of the RIL population.

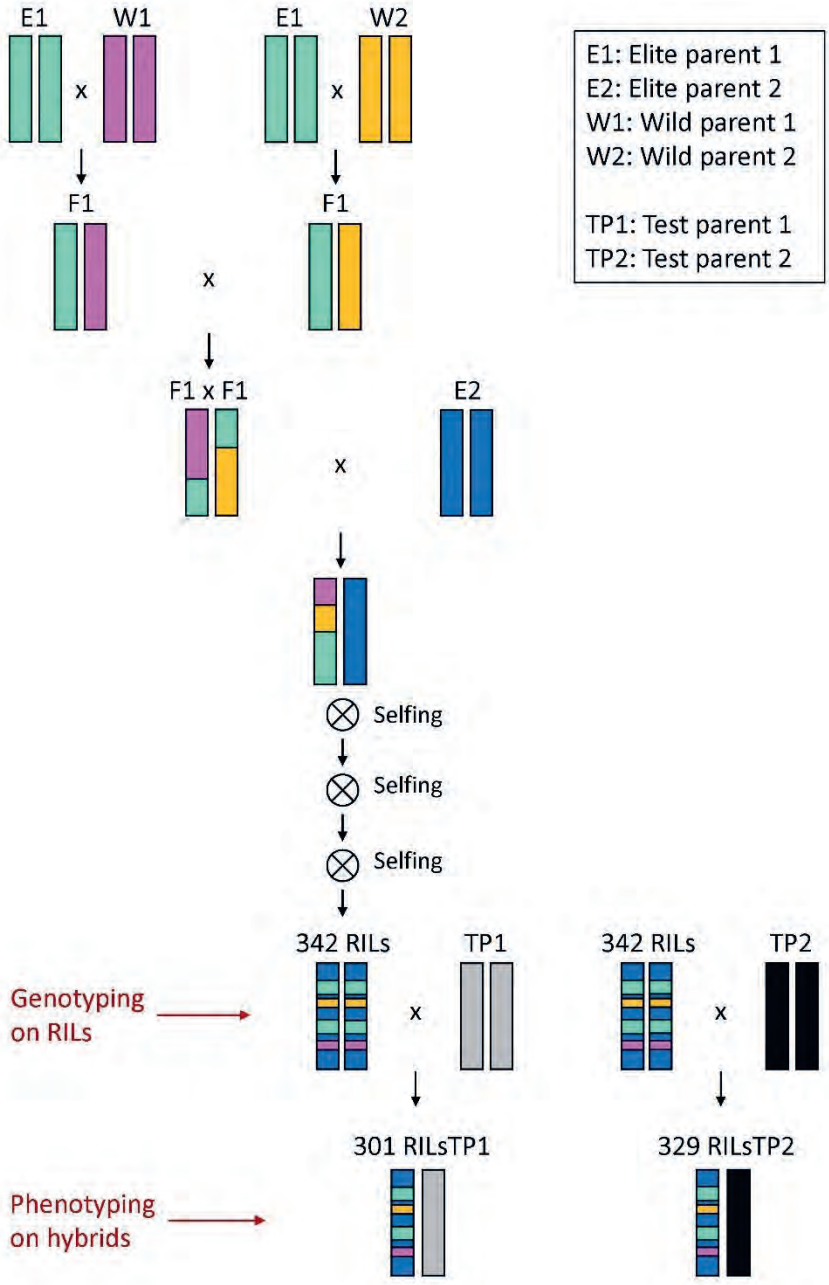


Figure 2 Construction of four-way RILs and corresponding hybrids. Two elite parents (E1, E2) and two wild-type parents (W1, W2) were founders for offspring population of 342 four-way RILs. The RILs were crossed with two testers (TP1, TP2). The resulting hybrids (RILsTP1, RILsTP2) were phenotyped.

2.2.3. Growing conditions and experimental design

The phenotyping experiment was conducted in the Westland area (the main tomato-producing region in the Netherlands) with commercial crop management. The experiment started in the first week of February 2016, when tomato seedlings were transferred to an area of around 3000 m^2 in a commercial tomato greenhouse (6 ha). Fruit harvest started in the middle of April. In September, the shoot tops were removed (decapitation), and the harvest finished in the middle of November. Tomatoes were grown on stone wool, and stems were supported by ropes (high-wire system). A shoot top produced about one truss and three leaves every week. Trusses were pruned to maintain six fruits per truss according to what is customary for this kind of truss tomato in the Netherlands. Plants were grown without supplementary light. Average daily temperature was 21.7 °C during daytime and 17.1 °C for night-time, while average CO_2 concentration during daytime was 629 ppm. Supplementary Figure 1 shows tomato production in the greenhouse.

As experimental design, a randomized complete block design was employed (**see** Supplementary Figure 2). The area was partitioned into two adjacent areas: RILs crossed with tester 1 (TP1) produced hybrids, RILsTP1, which were planted in area 1 (block 1), and hybrids created from the second tester (TP2), RILsTP2, were grown in area 2 (block 2). For quality control and spatial adjustment, several repetitions of crosses of the elite parents with tester 1 and 2 were planted (E1TP1 and E2TP1 in area 1, E1TP2 and E2TP2 in area 2). Also, several duplicates of five commercial cultivars were placed in both areas to be used as reference for correcting phenotypic values for spatial trends in later analysis. An experimental unit, or plot, included nine plants of a genotype. Because three out of nine plants were allowed to keep an additional stem, the resulting stem density was 3.37 m^{-2}, in agreement with commercial practice.

2.2.4. Phenotyping

Per plot, we phenotyped each of the traits as identified in the harvest and biomass dissection models (Table 1). Total yield, marketable yield, number of fruits, and fruit fresh-dry weight ratio were measured directly. All fruits per plot were measured for total yield, and all marketable fruits per plot were measured for marketable yield and the number of fruits. Individual fruit fresh weight was calculated at every harvest from marketable yield divided by the number of fruits. To determine the fruit fresh-dry weight ratio, three fruits were measured together from one side of a truss per plot. After fruit fresh weight was measured, fruits were dried in a ventilated oven at 45 °C for 24 h, 70 °C for the next 24 h and 105 °C for the last 72 h to determine fruit dry weight. Fruit dry weight measurements were conducted three times (in June, August and October), and the mean values per plot of these three measurements were used in the analysis. Also, the individual leaf dry weight was measured three times (in June, August and October). Three fully grown bottom leaves were measured

together from a stem per plot. The leaves were dried in a ventilated oven at 45 °C for 24 h followed by 105 °C for 48 h.

The harvest index was calculated as total fruit dry weight divided by total biomass. The total biomass in turn was taken as the sum of total fruit dry weight, total leaf dry weight and total stem dry weight. Total fruit dry weight was calculated from total yield multiplied by the fruit fresh-dry weight ratio. Total leaf dry weight was calculated from the mean of the three individual leaf dry weight measurements per plot multiplied by the total number of leaves. The total number of leaves in turn was calculated from the number of leaves under the first truss and the number of trusses (three leaves between two adjacent trusses). Total stem dry weight was estimated from stem length and diameter based on a regression model. We used this estimation because it was not feasible to measure the whole stem dry weight for more than 700 plots. For all plots, stem length and diameter were measured on one stem per plot. The mean stem diameter was calculated from six points uniformly distributed over a stem. First, stem volume (cm^3) was calculated by stem length (cm) and mean stem diameter (cm), assuming a perfect cylinder shape:

$$\textit{Stem volume} = \textit{(Stem length)} \times [\textit{(1/2} \times \textit{Mean stem diameter)}]^2 \times \pi \qquad \textit{Eq. 5}$$

In total, 102 plots (45 RILsTP1, 45 RILsTP2, 3 E1TP1, 3 E2TP1, 3 E1TP2, 3 E2TP2, of which 10 RILs were taken equal between RILsTP1 and RILsTP2) were randomly selected based on a method to uniformly cover the genetic space (Bustos-Korts *et al.*, 2016). From each plot, one stem was sampled after tomato production had finished in November. Total stem dry weight (48 h at 105 °C) for these 102 plots was measured and a regression equation (R^2 = 0.99) for total stem dry weight as function of stem volume was determined

$$\textit{Total stem dry weight} = 0.146352 \times \textit{(Stem volume)} \qquad \textit{Eq. 6}$$

Using equation (6), total stem dry weights were calculated for all plots.

Table 1 Trait definitions, units, collection and sampling. Units are expressed at plot level, including 12 stems from 1 genotype at 3.56 m^2.

Abbreviation	Trait	Unit in analysis	Observed / Calculated / Estimated	Sampling
Ytotal	Total yield	kg m^{-2}	Observed	All fruits per plot.
Ymark	Marketable yield	kg m^{-2}	Observed	All marketable fruits per plot.
<Component traits in harvest dissection>				
FN	Number of fruits	m^{-2}	Observed	All marketable fruits per plot.
FFW	Individual fruit fresh weight	g $fruit^{-1}$	Calculated at every harvest (Ymark / FN/1000)	
<Component traits in biomass dissection>				
Fresh_Dry_ratio	Fruit fresh-dry weight ratio	-	Observed	3 fruits measured together from one side of a truss per plot per season. Mean from three seasons.
FDWtotal	Total fruit dry weight	kg m^{-2}	Calculated (Ytotal/Fresh_Dry_ratio)	
HI	Harvest index	-	Calculated (FDWtotal / Biomass)	
Biomass	Total biomass (above ground)	kg m^{-2}	Calculated (FDWtotal + LDWtotal + StemDWtotal)	
<intermediate traits>				
LDWtotal	Total leaf dry weight	kg m^{-2}	Calculated (LN * LDWindv*12/3.56/1000)	
LDWindv	Individual leaf dry weight	g $leaf^{-1}$	Observed	3 full-grown leaves measured together from 1 stem per plot per season. Mean from three seasons.
LN	Total number of leaves	-	Calculated (LNunder1 + (TN-1)*3)	
LNunder1	Number of leaves under the 1st truss		Observed	1 stem per plot.
TN	Number of trusses	-	Observed	1 stem per plot.
StemDWtotal	Total stem dry weight	kg m^{-2}	Estimated per stem by regression (slope = 0.15; 12 stems per plot; 3.56m^2 per plot): (0.15*StemL* (StemD/2)^2)*π *12/3.56/1000	
StemD	Stem diameter	cm	Observed	Mean from 6 positions for 1 stem per plot.
StemL	Stem length	cm	Observed	1 stem per plot.

2.2.5. Phenotypic and QTL analysis

A first step in the phenotypic analysis was the calculation of spatially adjusted genotypic means. We utilized the R-package SpATS (Spatial Analysis of Field Trials with Splines) fitting a two-dimensional Penalized spline (P-spline) mixed model (Velazco *et al.*, 2017; Rodríguez-Álvarez *et al.*, 2018) to calculate the BLUEs, the best linear unbiased estimates corrected for local spatial trends created by variations in environmental and management conditions. The same R-package was used to calculate a generalized heritability (Cullis *et al.*, 2006; Rodríguez-Álvarez *et al.*, 2018). The duplicates of the hybrid genotypes E1TP1, E2TP1, E1TP2, E2TP2 and five commercial cultivars were included in the experiment functioned as reference and facilitated the spatial corrections.

Histograms of BLUEs for yield and components were generated to show the proportion and magnitude of transgression that occurred for single and pairs of traits, i.e. the proportion of offspring RILs that were exceeding the performance of their parents and the amount by which this was the case. Response surface plots for yield as a function of component traits (BLUEs) helped to investigate trade-offs between traits.

Because our mapping population was the result of a non-standard four-way cross, a special purpose QTL mapping procedure was used within a mixed model framework and based on IBD probabilities between parents and RILs as described in detail by (Li *et al.*, 2021). To test for QTLs, at each genomic evaluation position the following models for the observation y were compared in a deviance test, where we use a scalar notation for simplicity: H_0, $y = \mu + \varepsilon$ versus H_a, $y = \mu + \mathbf{z}^t\mathbf{u} + \varepsilon$, with μ a fixed intercept, ε a normally distributed error with mean 0 and variance σ^2, $\mathbf{z}$ a design vector containing scaled IBD probabilities between the four parents and the RILs, $\mathbf{z}^t$ the transpose of $\mathbf{z}$ and $\mathbf{u}$ a coefficients vector of random QTL allele substitution effects with mean zero and variance σ^2_{QTL}. Scaling of $\mathbf{z}$ was such that IBD probabilities were converted to allele substitution effects.

The deviance test to compare models under null (H_0) and alternative hypothesis (H_a) consisted of a loglikelihood ratio test for the variance component, σ^2_{QTL}, within the package ASReml (Butler *et al.*, 2017). A threshold for significance was put at $-\log_{10}P \geq 2.9$ as a compromise between the desire to impose an approximate correction for multiple testing and find some consistency in the occurrence pattern of yield and yield component QTLs across the genome.

The significant QTLs from a genome scan were subsequently inserted in a multi-QTL mixed model: $y = \mu + \sum(\mathbf{z}^t\mathbf{u}) + \varepsilon$, where the summation, $\sum$, indicates that multiple QTLs were included into the model. Each QTL was fitted with a specific variance. Final estimates for QTL effects were obtained from the multi-QTL model fit for QTLs with non-zero variances. Therefore, QTLs for being reported needed to pass a single and multi-locus test.

To investigate colocalization of yield and component QTLs, a window of 20 cM on either side of the yield QTL was defined within which the component QTL should be located. We chose 20 cM because this seemed to be a reasonable choice of confidence interval for QTL location given our type of population, population size and QTL effect magnitude (Darvasi and Soller, 1997; Wu *et al.*, 2007). A further condition to be fulfilled for colocalization of yield and component QTLs was that the parental origin of the yield QTL allele with the largest magnitude coincided with the parental origin of the component QTL allele, or haplotype, with the largest magnitude.

We investigated the predictive accuracy of multi-QTL models for yield and yield components as well as the accuracy of dissection models for yield prediction. For the latter, we first produced the multi-QTL predictions for the yield components in a dissection, and then calculated the product between the multi-QTL predictions of two (or more) yield components. For the evaluation of prediction accuracies, we implemented a cross-validation procedure. The data were partitioned into 5-folds, with 4-folds being included in a training set and 1-fold serving as validation set. This 5-fold cross-validation was repeated 20 times for different random partitionings of the data. An overview of all phenotypic and QTL analyses is given in Figure 3.

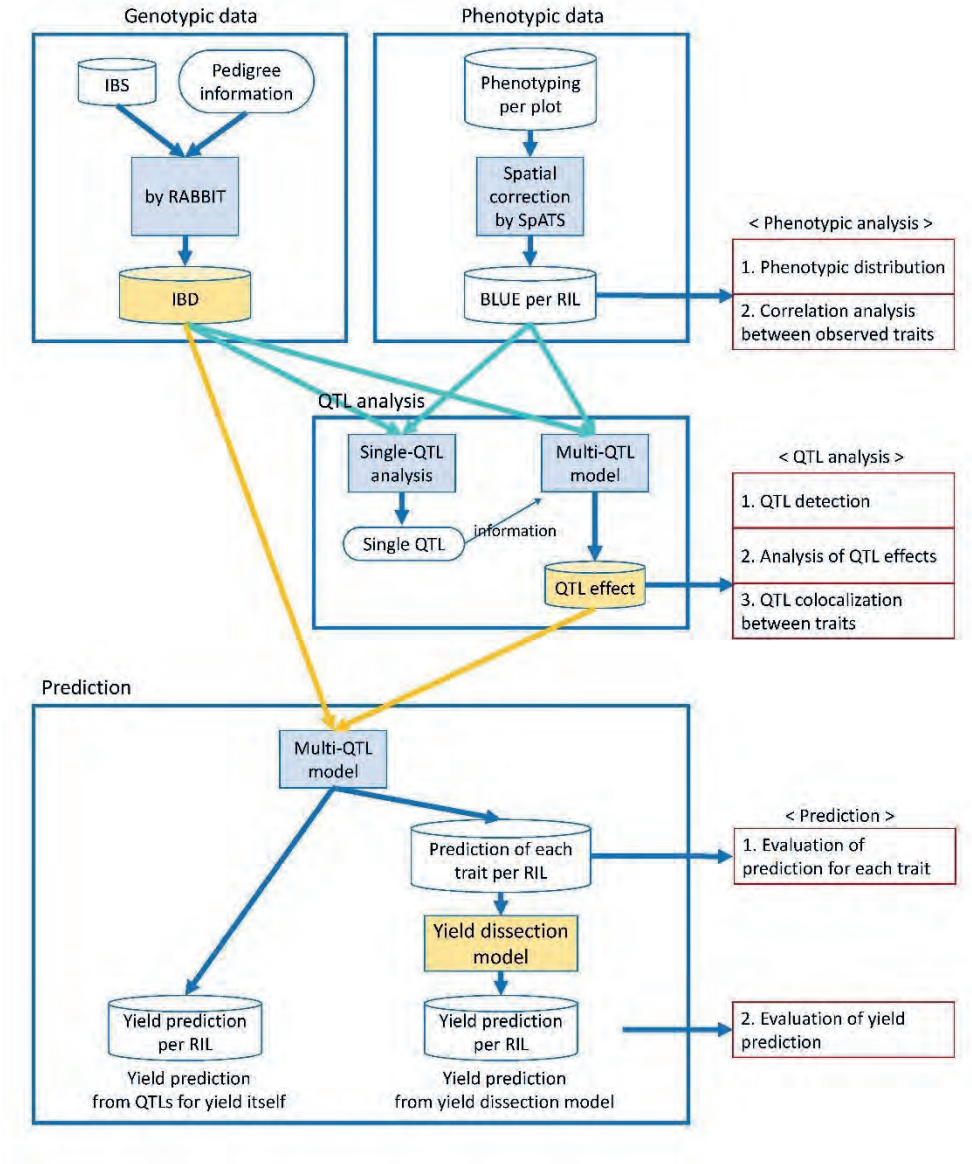


Figure 3 Overview of statistical analyses as presented in the Materials and Methods section. BLUE, best linear unbiased estimate; IBD, identity by descent; IBS, identity by state; RIL, recombinant inbred line. Software: RABBIT, a Mathematica software to calculate IBD; SpATS, an R-package to estimate spatially adjusted genotypic means.

2.3. Results

2.3.1. Heritabilities

The experiment was analysed by various mixed models with as most important outputs, heritabilities and spatially corrected genotypic means. The heritabilities will give an impression of the genetic determination of the traits and the possibilities for identifying QTLs. Table 2 shows that heritabilities for total yield and marketable yield were moderately high, 0.62 and 0.64, respectively. Therefore, direct selection for yield on the basis of yield QTLs looks feasible within this RIL population. When we look at the heritabilities of the yield components in the yield dissections, we note that the heritabilities of the components in the harvest dissection are higher than those for yield: 0.77 for number of fruits and 0.80 for individual fruit fresh weight. In contrast, for the components in the biomass dissection, the heritabilities were lower than for yield, 0.57 for fruit fresh-dry weight ratio, 0.47 for total fruit dry weight, 0.52 for harvest index and 0.48 for total biomass.

Table 2 Trait heritabilities and numbers of QTLs.

Trait	Unit	Heritability	Number of QTLs	Number of QTLs colocalized with yield QTLs with the same haplotype showing the highest effect
Total yield	kg m^{-2}	0.62	16	
Marketable yield	kg m^{-2}	0.64	14	
<Component traits in harvest dissection>				
Number of fruits (marketable fruits)	m^{-2}	0.77	19	2
Individual fruit fresh weight (marketable fruits)	g $fruit^{-1}$	0.80	21	8
<Component traits in biomass dissection>				
Fruit fresh-dry weight ratio	-	0.57	15	7
Total fruit dry weight	kg m^{-2}	0.47	12	7
Harvest index	-	0.52	8	1
Total biomass (above ground)	kg m^{-2}	0.48	12	5

2.3.2. Genotypic means, transgressions and correlations

The distributions of the genotypic means for total yield and marketable yield reveal that relatively little transgression occurred for RILs beyond the elite parents (Figure 4AB), i.e., few RILs had higher yields than the elite parents.

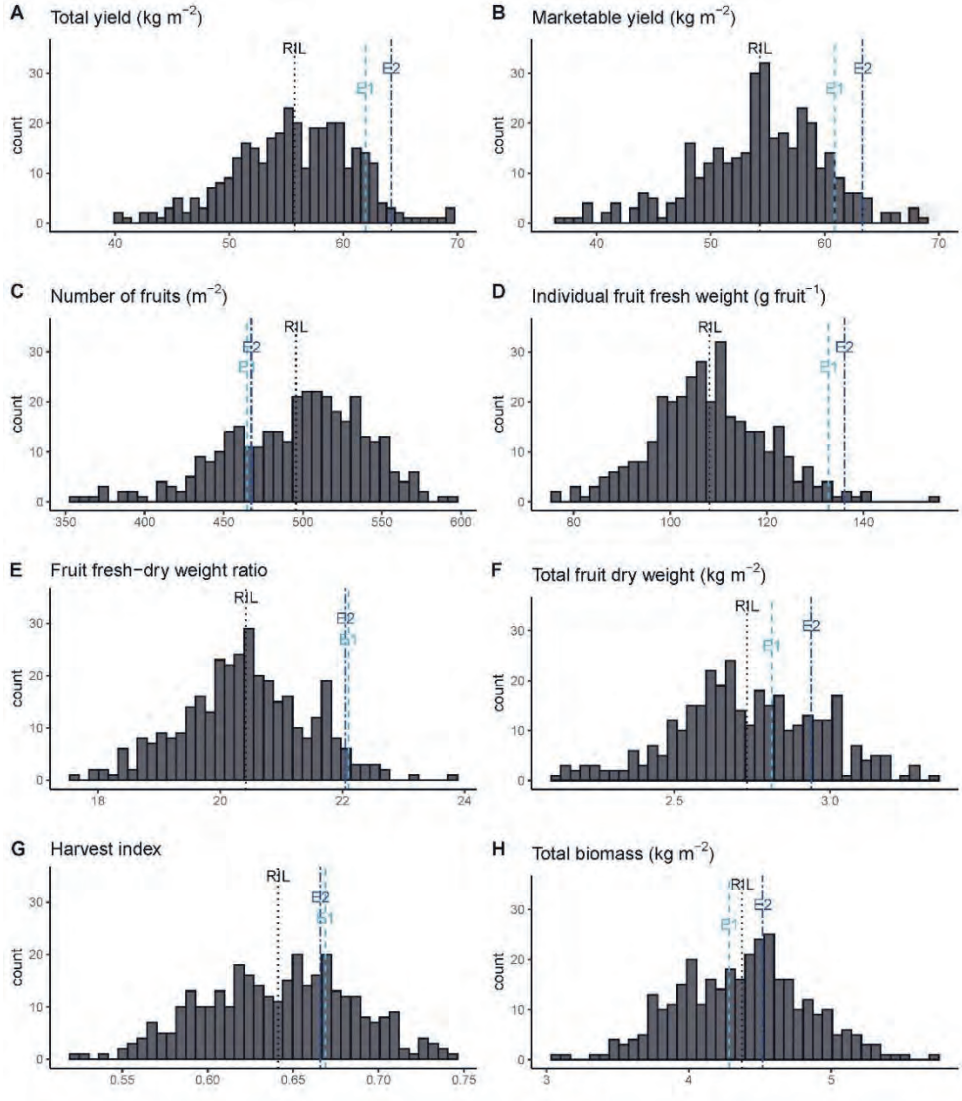


Figure 4 Distribution of RIL means for yield and yield components. Mean of all RILs indicated by vertical black dashed line, performance of the two elite parents by vertical blue dashed lines. Traits with standard errors: A, Total yield (3.3); B, Marketable yield (3.5); C, Number of fruits (22.4); D, Individual fruit fresh weight (5.4); E, Fruit fresh-dry weight ratio (0.7); F, Total fruit dry weight (0.2); G, Harvest index (0.031); H, Total biomass (0.33).

The distributions of the components in the harvest dissection (Figure 4C-H) point to an interesting phenomenon. For individual fruit fresh weight, again few RILs were better than the parents, but for the number of fruits, many RILs produced more fruits than the elite parents. This suggests that yield may be improved by increasing the number of fruits for genotypes with sufficiently high fruit fresh weight. This would be straightforward when number of fruits and fruit fresh weight are independent traits.

Figure 5 presents predicted marketable yields from the product of the BLUEs for number of fruits and individual fruit fresh weight. The marginal histograms for number of fruits and fruit fresh weight in Figure 5 correspond to the histograms in Figure 4. In Figure 5, we observe a trade-off between number of fruits and fruit fresh weight (Pearson correlation was -0.44). The scatter plot of predicted yields with isoclines illustrates that despite the moderately negative correlation between the components, a small set of RILs achieved superior yields because of a higher number of fruits. This higher number of fruits compensated for a slightly lower fruit fresh weight in a limited number of RILs. This observation is a first success in the application of yield dissection. The harvest dissection provides ideas for possible improvement strategies of yield that would be missed when yield was studied only by itself.

In the biomass dissection, total yield is created from fruit fresh-dry weight ratio and total fruit dry weight. The latter is then the outcome of harvest index combined with total biomass. Figure 4 shows little transgression for the fruit fresh-dry weight ratio, but a reasonable amount of transgression for the total fruit dry weight. Therefore, yield improvement is suggested via increasing total fruit dry weight. The response surface for yield as a function of total fruit dry weight and fruit fresh-dry weight ratio in Figure 5 demonstrates more opportunities for improving yield via total fruit dry weight than via the fruit fresh-dry weight ratio. Again, the trade-off between the components makes that it is almost impossible to achieve higher yields by simultaneous improvements in both components. At best one component, total fruit dry weight, can be increased in such a way that decreases in the other component, fruit fresh-dry weight ratio, are more than compensated for. This insight could only be obtained via a dissection.

Further insight in alternative strategies for improving total yield via biomass dissection follow from the decomposition of total fruit dry weight in harvest index and total biomass, Figure 4 and 5. Figure 4 leads to the conclusion that for both harvest index and total biomass substantial transgression occurred, suggesting that higher total fruit dry weight and higher total yield is possible via either or both of harvest index and total biomass. However, the yield response surface of Figure 5 demonstrates that more yield improvement can be obtained via higher total biomass than via higher harvest index.

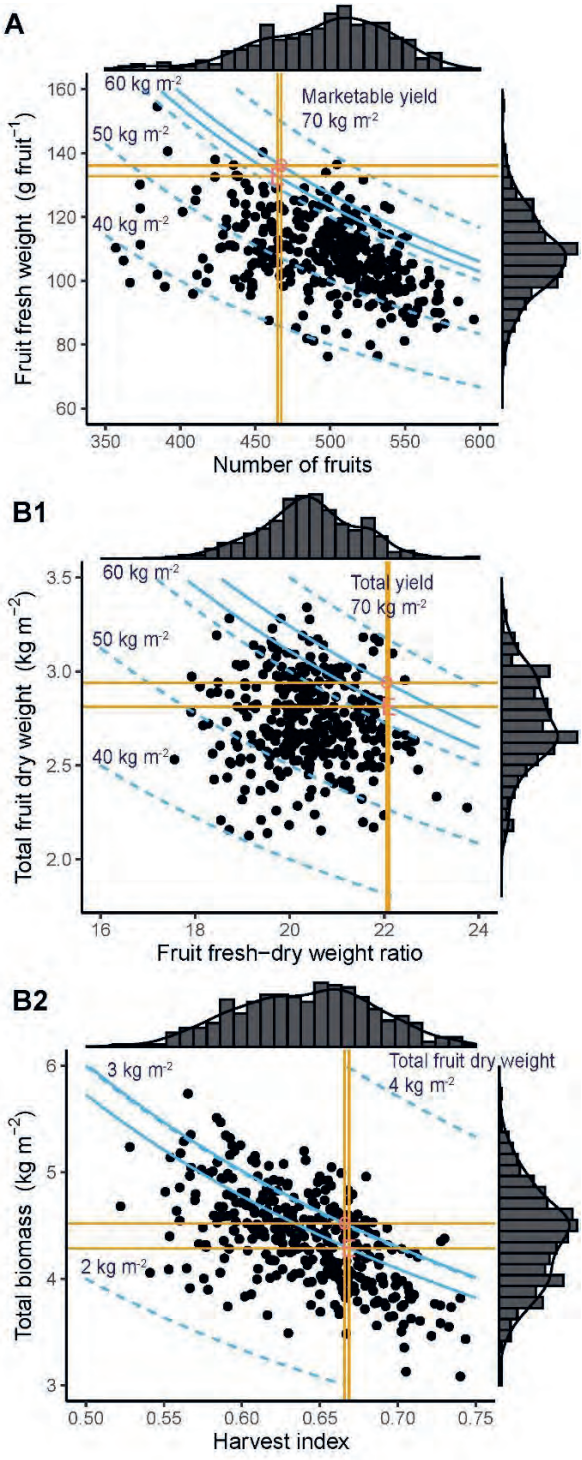


Figure 5 Response surfaces. (A) Marketable yield in harvest dissection. (B1) Total yield in biomass dissection. (B2) Total fruit dry weight in second level of biomass dissection. Horizontal and vertical red lines define quadrants determined by positions of the elite lines in the response surfaces. Curved blue lines define yield isoclines, with solid lines referring to the elite parents and dashed lines referring to 40, 50, 60 and 70 kg m^{-2} isoclines.

2.3.3. QTL analysis

Results of QTL analysis are presented in especially Table 2 and Figure 6 and 7, and see Supplementary Table 1.

Table 2 gives the numbers of QTLs for all traits and the numbers of colocalizing QTLs between yield and its components. We decided on colocalization between yield and component QTL when the component QTL appeared within a 20 cM window to the left or right from the yield QTL and the parental origin of the alleles for maximum yield and component coincided. Note that this is a rather strict criterion. For yield and its components, 8–21 QTLs were detected. We will refer to yield QTLs as follows: y1.1 is the first yield QTL on chromosome 1, y1.2 is the second yield QTL on chromosome 1, y2.3 is the third yield QTL on chromosome 2, etc. Of the 16 QTLs detected for total yield and 14 for marketable yield, 10 QTLs colocalized, resulting in 20 yield QTLs in total. All yield QTLs, except for one (y3.2; see Figure 6), colocalized with component trait QTLs. More yield QTLs colocalized with QTLs for component traits in the biomass dissection than in the harvest dissection. This result is important for plant breeders, as it allows them to get a better understanding of the nature of yield QTLs and the utility of individual QTLs for improving yield, where this utility is determined by QTL alleles increasing certain yield components without compromising other yield components. Figure 6 and especially Supplementary Table 1, which contains physical positions of QTLs, can be used in the construction of a strategy to arrive at desirable genotypes (ideotypes), where the choice of which QTLs to target for yield improvement needs to take into account linkage and pleiotropic effects between yield and yield components.

Yield QTLs were identified on most of the chromosomes, with the exception of chromosome 11, as this chromosome did not contain a QTL for yield although it did hold a QTL for fruit fresh-dry weight ratio. For 17 out of 19 QTLs (y1.1 and y2.1 were the exceptions), the parental allele or haplotype that showed the highest positive effect for yield also presented the highest positive effect for at least one component trait, an extremely useful result for breeders. Of these 17 QTLs, eight yield QTLs (y1.2, y3.1, y4.1, y5.1, y6.1, y7.1, y10.1, y12.2) corresponded to QTLs for component traits in both the harvest and the biomass dissection (black boxes in Figure 6), two yield QTLs (y2.2, y10.2) corresponded to QTLs for component traits in the harvest dissection but not in the biomass dissection (orange box), and seven yield QTLs (y1.3, y1.4, y6.2, y7.2, y8.1, y9.1, y12.1) corresponded to QTLs for component traits in the biomass dissection but not in the harvest dissection (green box).

Figure 7 shows parental chromosomal effects as well as multi-QTL predictions for the parent genotypes, two wild-type and two elite lines. In this case, parental haplotypes were constructed by combining the QTL genotypes across the QTL loci on single chromosomes to create a kind of chromosomal haplotypes for the four parents. Chromosomal effects per parent were calculated by summing allelic effects. The representation in Figure 7 makes it

possible to appreciate the individual chromosomal contributions to yield and component traits, expressed as a deviation from the mean (intercept in multi-QTL model), as well as the predicted phenotype for the parent.

For total yield (Figure 7A), marketable yield (Figure 7B), individual fruit fresh weight (Figure 7D), total fruit dry weight (Figure 7F) and total biomass (Figure 7H), the two elite parents had an overall positive predicted trait value, while the two wild-type parents had a negative overall prediction, with elite parent 1, being best, followed by elite parent 2, wild-type parent 1 and wild-type parent 2 being the least productive. Still, in the elite parents some small negative effects were present on some chromosomes, while some sizeable positive effects occurred in wild-type parent 1. The distribution of these effects indicates that yield increases seem possible by replacing some elite alleles by wild-type alleles. This phenomenon is more outspoken in the components than yield itself. Furthermore, when considering number of fruits (Figure 7C), fruit fresh-dry weight ratio (Figure 7E) and harvest index (Figure 7G), the usefulness of wild-type parent 1 as bearer of positive yield component alleles becomes even more evident.

2.3.4. Predictions from multi-QTL models

Although the merit of dissecting yield into components was already demonstrated above by the extra options that were offered to compile superior genotypes for yield from QTL alleles for components with different parental origin, we were curious to see whether yield can be predicted from the components in a dissection model. We built multi-QTL models for the components, took the multi-QTL predictions for the components and applied the function to the component predictions that according to the dissection model should result in yield. In this paper, product formulations were defined for yield in the harvest and biomass dissection models.

Figure 8 contains the prediction accuracies for yield (black box in Figure 8) and component traits from their corresponding QTLs (white boxes in Figure 8) based on 20 realizations of a 5-fold cross-validation. For marketable yield, the (median) accuracy from predictions by multi-locus yield QTL models was 0.63, while the accuracy for the components number of fruits and individual fruit fresh weight was 0.73 and 0.69, respectively. When the phenotypic predictions for the components were multiplied to approximate yield (grey box in Figure 8), the accuracy for the harvest dissection model was slightly lower than the accuracy for predicting yield from its own QTLs, 0.56. The correlation between the predictions of yield from its own QTLs and from the harvest dissection model with component QTLs was 0.84 (see Supplementary Figure 3A).

In the biomass dissection, we observed an accuracy of 0.61 for total yield, and 0.65 for fruit fresh-dry weight ratio, while for total fruit dry weight it was 0.50. Accuracy for total yield prediction using the product of predicted components in the biomass dissection model was a

little bit better than direct prediction of total yield from its own QTLs, 0.63. The correlation between total yield prediction from its own QTLs and total yield from the biomass dissection model was 0.89 (see Supplementary Figure 3B). Extending the biomass dissection model with a decomposition of total fruit dry weight into harvest index and total biomass led to a somewhat lower accuracy than the other two prediction models for total yield, 0.58 (Figure 8; see Supplementary Figure 3C).

The conclusion of these multi-locus prediction analyses is that the genetic basis of the component traits can be estimated well enough to try to predict yield from the components using a dissection model. When the components can be measured or approximated earlier or cheaper than yield, whether by genomic prediction, secondary phenotyping or a combination of those two, prediction of yield via dissection models appears attractive.

Figure 6 QTL locations and effects for yield and components. On vertical axis, chromosome 1 to 12, on horizontal axis, position within chromosome. QTL positions are indicated by dots. Dashed vertical lines represent markers. Colour of dot shows haplotype with highest positive effect. Size of dot represents proportion (%) of largest haplotype effect to mean for the trait. Boxes indicate colocalized yield and component QTLs, with black boxes for components belonging to both harvest and biomass dissection, orange for components belonging to harvest dissection and green for components belonging to biomass dissection.

Figure 6 Continued.

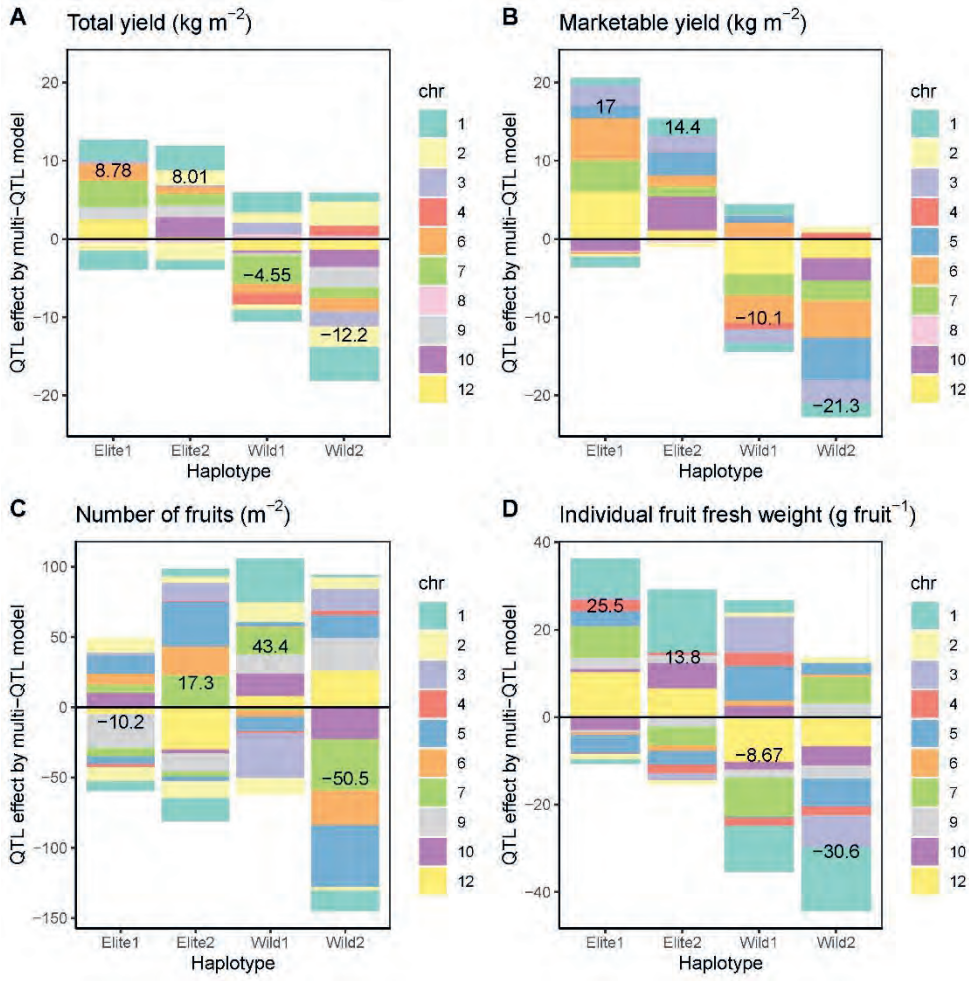


Figure 7 Haplotype effects per chromosome and multi-locus predictions for the parents. Haplotype effects have colour of chromosome. Multi-locus prediction presented as number.

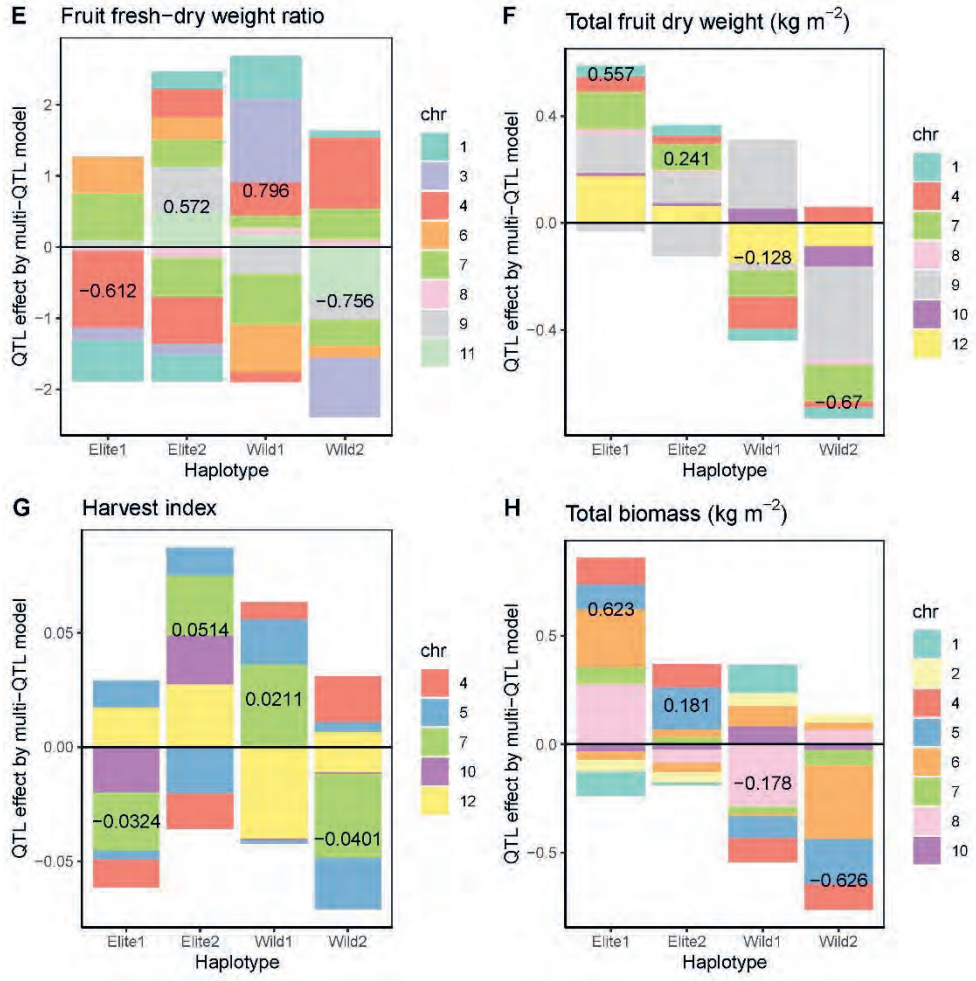


Figure 7 Continued.

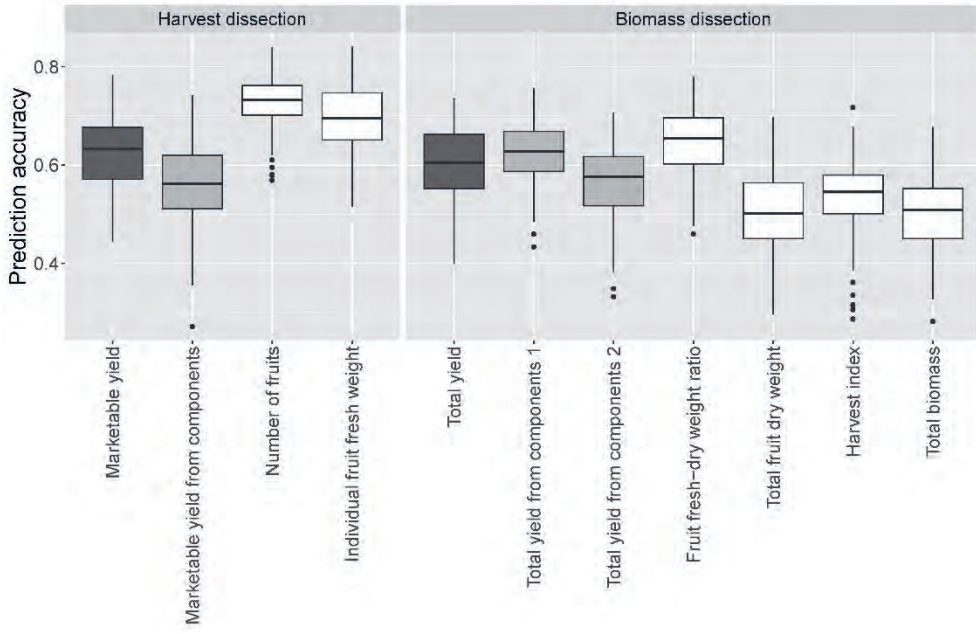


Figure 8 Prediction accuracies for yield and yield components as well as for yield as predicted from a dissection model. On the left, marketable yield from components indicates the accuracy of prediction of marketable yield by the product of the predictions for number of fruits and individual fruit fresh weight. On the right, total yield from components 1 shows the accuracy for prediction of total yield from the product of the predictions for fruit fresh-dry weight ratio and total fruit dry weight, while total yield from components 2 gives the prediction of total yield from the product of the predictions for fruit fresh-dry weight ratio, harvest index and total biomass. See also Supplementary Figure 3.

2.4. Discussion

2.4.1. Limitations of yield on the underlying levels

By following a yield dissection, we obtain insight in the limitations working on yield from genetic contributions at upstream levels. As a follow up, we should target QTLs for component traits such that observed trade-offs and negative correlations can be circumvented or 'broken' to further improve yield.

One apparent yield limitation is the trade-off between the number of fruits and individual fruit fresh weight (the two underlying components of marketable yield in the harvest-based dissection). The wild-type parents mainly had negative effects on individual fruit fresh weight, but these negative effects seemed to be mitigated by positive effects on the number of fruits. Hybrids with a higher number of fruits tended to have a lower individual fruit fresh

weight, and vice versa, resulting in a negative correlation between the two traits. Some chromosomes contained QTLs that had opposite effects on number of fruits and individual fruit fresh weight (e.g. QTLs on chromosome 7 and 12). This suggests a necessity to consider both component traits simultaneously to increase yield. Several hybrids had a combination of individual fruit fresh weight and fruit number that led to higher yield than that of the elite parents, presenting favourable combinations of QTLs for the underlying component traits that should be targetted by breeders.

The underlying component traits of yield in the biomass-based dissection, fruit fresh-dry weight ratio and total fruit dry weight, were only weakly correlated. This might be explained by the observation, that the QTLs for the two component traits were detected on different chromosomes (chromosome 3, 6 and 11 for fruit fresh-dry weight ratio, and 10 and 12 for total fruit dry weight, respectively). In addition, QTLs on 'shared' chromosomes had differentiating effects on the two component traits (QTLs on chromosome 1 and 4 affected fruit fresh-dry weight ratio much more than total fruit dry weight, whereas the opposite was true for QTLs on chromosome 9). This suggests the two traits are largely independent from each other and can as such be increased independently and/or simultaneously to increase total yield, which offers opportunities for breeding.

Total fruit dry weight was further dissected into harvest index and total biomass. In our population, total biomass correlated strongly with total fruit dry weight. By far, the largest QTL effects on total biomass were contributed by elite parent 1, although at least one QTL haplotype on chromosome 1 from wild-type parent 1 also looked promising. On the other hand, both wild-type parents as well as elite parent 2 had considerable positive and negative effects on harvest index from QTLs on different chromosomes, most importantly chromosome 7 (the highest positive effect from elite parent 2 and wild-type parent 1, and the most negative effect from wild-type parent 2, respectively). Five out of 12 QTLs for total biomass colocalized with yield QTLs with the same haplotype showing the highest positive effect, whereas only one out of eight QTLs for harvest index colocalized with yield QTLs. The question that arises is why harvest index QTLs seem to contribute so little to yield in this population? Does the contribution of harvest index QTLs to yield depend on genetic background or management? Or, is it a consequence of our dissection model that may be too simple and violates certain causal relations?

A previous study comparing eight Dutch indeterminate tomato cultivars released between 1950 and 2003 also showed yield was positively correlated with total biomass but not with harvest index (Higashide and Heuvelink, 2009). On the contrary, another study comparing six Japanese indeterminate tomato cultivars released in the past 80 years showed that yield highly positively correlated with harvest index but not with total biomass (Higashide *et al.*, 2012). These conflicting observations may be the result of the use of different genotypes or of genotype by environment effects (genotypic differences being dependent on the environmental and/or management conditions, van Eeuwijk et al., 2016). A higher harvest

index results in a lower biomass allocation to leaves and leaf area index, which could mean a lower total biomass unless the specific leaf area (leaf area per leaf dry mass) and/or light use efficiency (biomass production per unit intercepted light) would improve. In major staple crops, improvements of harvest index have historically increased yield potentials (Hay, 1995; Smith *et al.*, 2018). It is reported that modern crop cultivars have a higher harvest index, although total dry matter production is most often very similar (Evans, 1993).

2.4.2. Contributions from wild-types

Many genes contribute to yield, making yield improvement a considerable challenge for breeders. In tomato, domestication and improvement have increased the productivity by narrowing its genetic basis (Bai and Lindhout, 2007). Increasing the genetic diversity may be beneficial for improving yield along different component traits. Wild germplasm has been used as a source of new alleles for tomato breeding in recent decades (Rick, 1978; Rick and Chetelat, 1995; Tanksley and McCouch, 1997; Lin *et al.*, 2014). In our experiment, two wild-type parents provided new alleles. Many hybrids performed worse in terms of marketable yield (for the harvest dissection) or total yield (for the biomass dissection). Negative effects of wild-type parents are to be expected, as individual fruit fresh weight has increased 100-fold during the domestication and improvement of tomato (Lin *et al.*, 2014). There was, however, a subset of hybrids in our population that performed better than the elite lines. This is because wild-type parent 1 also contributed positive QTL effects to different component traits underlying yield, such as number of fruits and fruit fresh-dry weight ratio. This emphasizes that by only considering yield and not looking at underlying components, positive effects to yield via component traits remain obscured. Moreover, genetic effects present an actionable target for breeders, who can cross QTLs for component traits into elite lines. In addition, it suggests that wild parents are useful as genetic resources, not only for incorporating multiple disease-resistant loci in elite lines, but also to improve yield. Contributions of wild parents to tomato yield have been reported for biparental populations (Zamir, 2001; Gur and Zamir, 2004; Lippman *et al.*, 2007), whereas our study reports it for a multi-parent population.

2.4.3. Future methodological possibilities

Beneficial component traits can be used for making a cross for future cultivars and/or in choosing parent lines (Cooper *et al.*, 2014; van Eeuwijk *et al.*, 2019), and in this paper, we present a proof of principle for using yield dissection models for identifying such beneficial component traits. The usefulness of dissecting traits along component traits has received more attention with the emergence of phenotyping technology, such as high-throughput phenotyping. Ideally, a phenotype is highly correlated with QTLs. We can then perform a selection of high-yielding RILs/hybrids in an early stage by extracting DNA from seedlings, eliminating the need for following full breeding cycles, which would reduce breeding efforts considerably.

Yield is the integrated result of many genetic and gene–environment interactions, and a strong correlation between QTLs and phenotypic traits is more likely when the traits are on a lower organizational level, i.e. are upstream, which excludes several genetic interactions. Because of the simpler genetic basis, it is also expected that prediction of component traits is more accurate than prediction of yield itself. In our analysis, the yield prediction models based on QTLs for component traits performed about as good as the ones based on QTLs for yield itself, suggesting that genetic dissection is viable. It should be noted, that the prediction accuracy may be highly influenced by the quality and quantity of genotypic and environmental sampling. In our experiment, number of fruits and individual fruit fresh weight were phenotyped for all fruits in our study.

Additional advantages of using a genetic dissection with component traits may be:

- 1) The heritability may be higher than for yield due to the simpler genetic mechanism of component traits. In our results, the number of fruits and the individual fruit fresh weight showed higher heritability than yield (Table 2);
- 2) Some component traits do not need a complete harvest season (8 months) to be determined, possibly shortening the breeding cycle, e.g. individual fruit fresh weight and the fruit fresh-dry weight ratio can be measured for early yield.

In particular, this latter point may become more relevant as the measurement of several component traits is becoming easier and cheaper, and feasible for larger numbers of genotypes by new phenotyping technologies (e.g. by using drones, or high throughput phenotyping platforms).

The use of a four-parent cross involving also two wild-type parents generates a larger ‘genetic space’ for possible genotypes to expose possible trade-offs and negative correlations, which might not have been found when only a biparental population would have been used. Depending on the cross, some traits may not be sufficiently contrasting, e.g. by crossing two parents with high yield, one will most likely not find promising QTLs for improving yield. In this four-parent cross involving very different genotypes, both the aggregated trait ‘yield’ and underlying component traits show strong contrasts.

Another relevant item is the set-up of the experiment. By growing the population in a greenhouse, management and environmental conditions are under relative control. The advantage of this is that we have replicable conditions. It, however, also means that the application of the QTLs we found may be limited when considering tomato populations grown in other environments. Other experiments could focus on alternative designs to evaluate which component-trait QTLs are consistently expressed across environmental and management gradients. This does not affect the methodological approach of combining static yield dissection modelling with QTL analysis for component traits, which we assume to be generally applicable. To include genotype by environment interactions into yield prediction, both static yield dissection models with environmental covariates (Millet *et al.*, 2016, 2019)

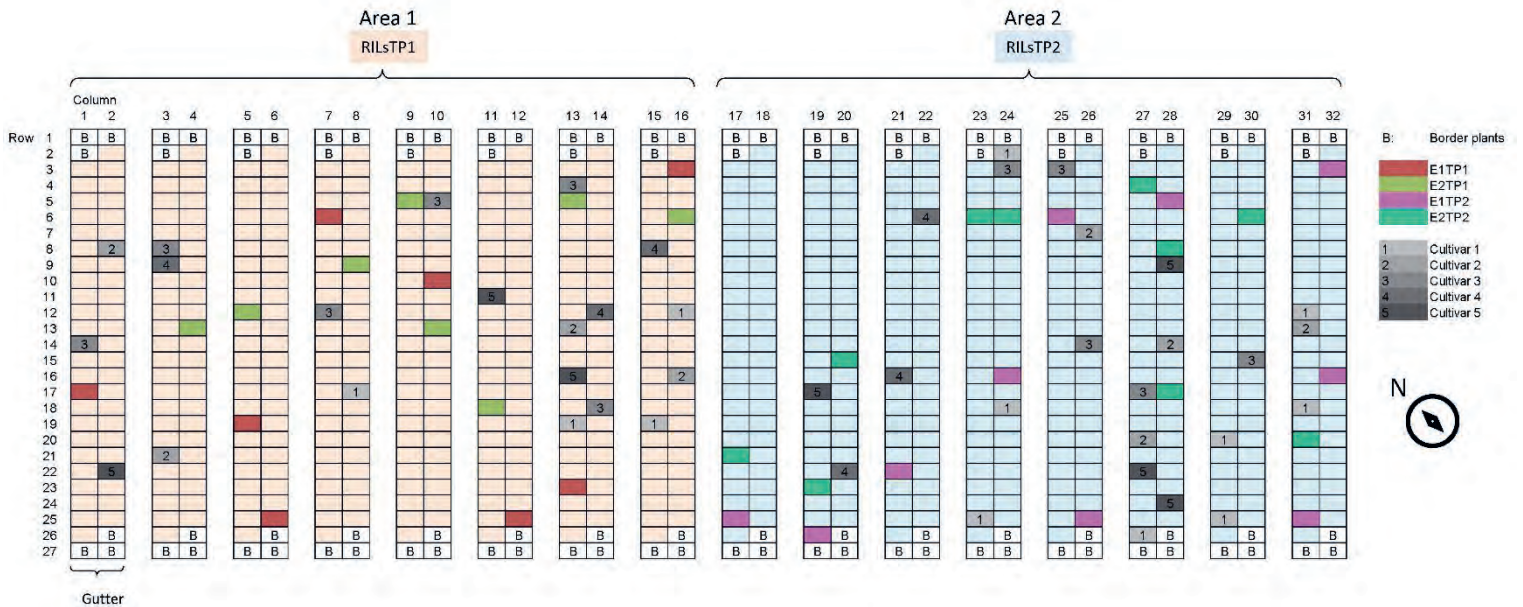
and dynamic crop growth models that include environmental factors may be useful (Cooper *et al.*, 2016). In a follow-up paper to the current paper, we will report on results obtained in two experiments in Spanish greenhouses and come back to the question of the consistency of yield and component QTLs and the utility of yield dissection models.

2.5. Concluding remarks

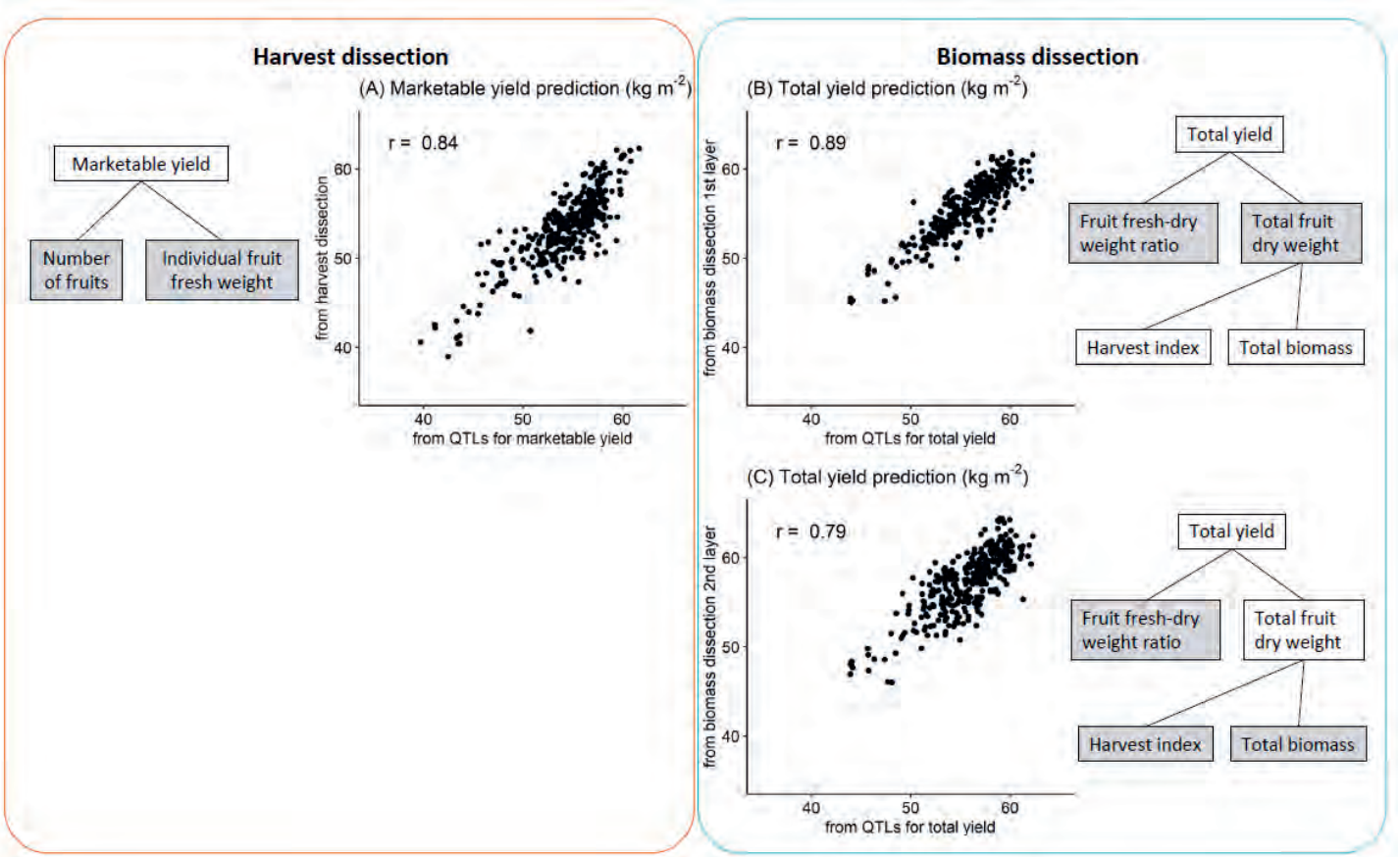
In this paper, we investigated the combination of static genetic yield dissection and QTL analyses for the possible detection of QTLs for component traits underlying yield. A primary outcome of such an approach for breeders is insights into yield response surfaces as functions of component traits in which alternative scenarios for yield improvement can be investigated under genetic and management constraints as well as cost considerations. The high accuracies of yield prediction from yield dissection models, where each component trait was predicted by its own QTLs, suggests that breeding for high yields can be done by selecting for component traits. A multi-parent population with sufficiently diverse parent genotypes enforces strong contrasts in component traits. Of interest to pre-breeding is that wild germplasm has more to offer than disease resistance genes and can contribute to higher yields.

Supporting information

Supplementary Figure 1 Indeterminate tomato grown in greenhouse for phenotyping. A shoot top produced a truss and three leaves constantly at around 1-week interval. Some flowers at top were pruned to keep six fruits per truss.



Supplementary Figure 2 Experimental design in the greenhouse.



Supplementary Figure 3 Scatter plots and correlations for yield predictions via dissection models versus direct yield predictions. (A) Harvest dissection, (B) Biomass dissection highest level, (C) Biomass dissection two levels.

Supplementary Table 1 Yield and component QTLs. QTL positions, SolCAP marker names, $-log_{10}(p)$ values for test on variance component of haplotype effects, and haplotype effects.

						QTL effect from multi-QTL model			
Trait	Chromosome	cM	Position (bp)	SolCAP.marker	- log_{10}P	Elite parent 1	Elite parent 2	Wild parent 1	Wild parent 2
Total yield (kg m^{-2})									
	1	25.6	2,879,156	solcap_snp_sl_60089	6.9	-1.31	-0.10	0.32	1.09
	1	38.8	4,940,680	solcap_snp_sl_33676	8.2	2.72	3.11	-1.46	-4.37
	1	53.2	79,245,417	solcap_snp_sl_25931	6.2	-1.07	-0.38	1.50	-0.06
	1	77.5	84,839,564	solcap_snp_sl_63558	4.3	-0.05	-0.73	0.75	0.03
	2	11	36,167,998	solcap_snp_sl_15706	4.2	-0.74	1.93	1.35	-2.54
	2	29	40,136,280	solcap_snp_sl_10552	5.9	-0.25	-2.17	-0.72	3.13
	3	32.2	3,540,929	solcap_snp_sl_19514	3.1	0.36	0.17	1.45	-1.99
	4	39.4	7,963,140	solcap_snp_sl_45951	6.2	0.02	0.09	-1.38	1.28
	6	42.6	42,434,661	CL017618-0174	5.7	2.15	0.83	-1.28	-1.70
	7	23.2	2,881,365	solcap_snp_sl_15786	7.7	1.29	1.13	-1.78	-0.64
	7	47.6	61,442,206	solcap_snp_sl_53538	7.9	2.12	0.39	-1.89	-0.63
	8	18	2,587,919	solcap_snp_sl_14531	2.9	-0.29	-0.50	0.41	0.38
	9	108.2	70,828,802	solcap_snp_sl_22327	3.0	1.44	1.42	-0.23	-2.64
	10	6.8	1,395,637	solcap_snp_sl_46339	6.9	0.00	0.16	0.16	-0.32
	10	21	3,092,679	SGN-U603133_snp167	6.2	-0.17	2.45	-0.34	-1.94
	12	55.7	63,393,430	solcap_snp_sl_55530	7.5	2.56	0.19	-1.44	-1.31
Marketable yield (kg m^{-2})									
	1	38.8	4,940,680	solcap_snp_sl_33676	7.7	0.92	1.86	-1.12	-1.66
	1	53.2	79,245,417	solcap_snp_sl_25931	5.5	-1.41	0.25	1.27	-0.10

	2	23.4	38,960,460	Le001778_68_solcap_snp_sl_33474	8.2	-0.23	-0.59	0.15	0.68
	3	54.5	57,784,509	solcap_snp_sl_35369	4.0	2.62	2.16	-1.81	-2.96
	4	39.4	7,963,140	solcap_snp_sl_45951	3.7	0.09	0.07	-0.82	0.67
	5	48	8,354,784	solcap_snp_sl_51094	5.3	1.55	2.94	0.93	-5.42
	6	13.4	36,268,579	solcap_snp_sl_68871	3.4	-0.32	0.95	1.74	-2.36
	6	42.6	42,434,661	CL017618-0174	7.5	5.40	0.50	-3.55	-2.35
	7	23.2	2,881,365	solcap_snp_sl_15786	7.9	2.11	0.71	-1.35	-1.47
	7	47.6	61,442,206	solcap_snp_sl_53538	10.6	1.83	0.53	-1.27	-1.09
	8	18	2,587,919	solcap_snp_sl_14531	3.3	-0.08	-0.23	0.19	0.12
	10	21	3,092,679	SGN-U603133_snp167	9.0	-1.55	4.37	0.10	-2.92
	12	15	1,971,483	solcap_snp_sl_12656	5.2	3.40	1.07	-3.25	-1.22
	12	55.7	63,393,430	solcap_snp_sl_55530	5.1	2.67	-0.21	-1.26	-1.20
Number of fruits (m^{-2})									
	1	25.6	2,879,156	solcap_snp_sl_60089	26.3	-5.16	-16.64	19.86	1.94
	1	117.1	95,296,848	solcap_snp_sl_54758	3.2	-1.94	5.23	11.21	-14.50
	2	11	36,167,998	solcap_snp_sl_15706	4.5	-10.05	-11.93	13.99	7.99
	2	25	39,514,373	solcap_snp_sl_8510	3.3	10.57	4.61	-12.26	-2.91
	3	96.4	-	SL10678_608	3.2	2.81	13.23	-31.96	15.93
	4	35	5,295,726	solcap_snp_sl_41609	3.9	-2.36	0.56	-1.21	3.00
	5	2	712,311	solcap_snp_sl_69075	6.9	5.98	1.25	1.65	-8.88
	5	15	2,153,070	solcap_snp_sl_49210	4.5	-2.75	-3.28	-10.04	16.07
	5	43.6	6,571,837	solcap_snp_sl_50925	24.7	6.10	21.16	1.65	-28.92
	5	64	62,533,914	solcap_snp_sl_12213	3.3	-2.54	9.03	-0.50	-5.99
	6	17.3	37,583,093	solcap_snp_sl_55858	5.5	1.19	2.68	0.69	-4.56
	6	42	42,258,236	solcap_snp_sl_39073	5.8	7.06	17.75	-4.66	-20.16

	7	10.8	1,557,551	solcap_snp_sl_68398	11.3	1.40	1.74	-1.70	-1.43
	7	42.8	60,093,644	solcap_snp_sl_14023	13.3	4.04	21.09	7.34	-32.48
	7	68.6	65,232,683	solcap_snp_sl_71085	5.3	-6.04	-3.71	11.93	-2.18
	9	64	14,485,136	solcap_snp_sl_41476	4.8	-23.87	-12.77	13.43	23.20
	10	21	3,092,679	SGN-U603133_snp167	10.9	9.62	-2.92	16.19	-22.89
	12	2	715,204	CL015154-0161_solcap_snp_sl_45733	5.8	0.66	-24.67	0.06	23.95
	12	40.8	38,735,374	solcap_snp_sl_38512	14.3	-4.95	-5.08	7.75	2.27
Individual fruit fresh weight (g fruit^{-1})									
	1	25.6	2,879,156	solcap_snp_sl_60089	14.8	-1.07	1.89	2.75	-3.57
	1	40.8	18,523,311	solcap_snp_sl_41282	12.9	8.82	12.54	-10.46	-10.89
	2	29	40,136,280	solcap_snp_sl_10552	3.6	-1.21	-1.06	0.91	1.36
	3	37	4,623,969	solcap_snp_sl_23196	6.7	0.65	-1.56	8.37	-7.45
	4	35	5,295,726	solcap_snp_sl_41609	3.4	2.54	0.60	-1.95	-1.19
	4	62	60,271,155	solcap_snp_sl_11520	5.5	-0.20	-2.00	3.00	-0.80
	5	21	2,662,318	solcap_snp_sl_49084	3.3	2.87	-1.11	3.72	-5.48
	5	30	3,980,854	solcap_snp_sl_23734	8.1	0.55	-0.64	0.92	-0.83
	5	39	5,309,739	solcap_snp_sl_50722	6.4	-0.85	-0.83	-0.25	1.93
	5	72.4	63,351,010	solcap_snp_sl_12274	3.7	-3.44	-0.53	3.28	0.69
	6	13.4	36,268,579	solcap_snp_sl_68871	3.5	-0.47	-1.18	1.11	0.54
	6	28	39,619,494	solcap_snp_sl_1346	5.1	-0.01	-0.13	0.04	0.10
	7	35	56,593,537	solcap_snp_sl_38868	6.4	7.20	-4.24	-8.94	5.97
	9	2	651,775	solcap_snp_sl_28404	3.4	-0.46	-2.13	-0.50	3.09
	9	25	2,673,326	solcap_snp_sl_39806	7.4	2.48	1.79	-1.23	-3.04
	10	6.8	1,395,637	solcap_snp_sl_46339	13.3	-2.96	3.97	2.56	-3.57
	10	21	3,092,679	SGN-U603133_snp167	7.0	0.84	1.79	-1.88	-0.75
	10	69	64,037,481	solcap_snp_sl_60967	6.8	0.00	0.02	0.03	-0.05

12	15	1,971,483	solcap_snp_sl_12656	12.0	0.61	3.95	-3.44	-1.12
12	40.8	38,735,374	solcap_snp_sl_38512	15.0	1.76	2.66	-1.93	-2.49
12	68	64,806,488	solcap_snp_sl_25007	13.4	7.85	0.00	-4.77	-3.07
Fruit fresh-dry weight ratio								
1	57	80,017,664	solcap_snp_sl_38290	3.3	-0.58	0.25	0.29	0.04
1	91	88,524,928	ZDS_Promoter_SNP1	3.3	0.00	-0.38	0.32	0.06
3	37	4,623,969	solcap_snp_sl_23196	4.7	-0.18	-0.15	1.18	-0.85
4	44	53,605,655	CL015899-0460_solcap_snp_sl_42811	4.8	-0.61	0.40	-0.13	0.35
4	62	60,271,155	solcap_snp_sl_11520	4.5	-0.24	-0.61	0.33	0.52
4	74	62,550,134	CL016270-0427_solcap_snp_sl_47056	4.7	-0.22	-0.05	0.13	0.13
6	42.6	42,434,661	CL017618-0174	6.2	0.53	0.31	-0.68	-0.16
7	23.2	2,881,365	solcap_snp_sl_15786	4.5	0.29	0.24	-0.25	-0.28
7	37	57,927,159	solcap_snp_sl_51838	5.6	0.06	-0.14	-0.18	0.27
7	47.6	61,442,206	solcap_snp_sl_53538	4.1	0.22	0.13	-0.28	-0.08
7	68.6	65,232,683	solcap_snp_sl_71085	4.6	0.08	-0.41	0.17	0.16
8	29	54,444,193	solcap_snp_sl_13447	3.6	0.00	-0.10	0.09	0.01
8	46	58,745,796	solcap_snp_sl_15381	4.4	-0.05	-0.05	0.01	0.09
9	69	63,812,814	solcap_snp_sl_43102	4.4	0.04	0.65	-0.38	-0.31
11	11.8	1,988,809	solcap_snp_sl_10611	3.3	0.06	0.48	0.17	-0.72
Total fruit dry weight (kg m^{-2})								
1	38.8	4,940,680	solcap_snp_sl_33676	4.1	0.041	0.041	-0.043	-0.040
4	35	5,295,726	solcap_snp_sl_41609	5.0	0.045	0.005	-0.109	0.059
4	74	62,550,134	CL016270-0427_solcap_snp_sl_47056	3.1	0.011	0.022	-0.011	-0.022
7	23.2	2,881,365	solcap_snp_sl_15786	3.1	0.052	0.042	-0.075	-0.019
7	42	59,724,392	solcap_snp_sl_5862	4.5	0.087	0.058	-0.025	-0.120
8	46	58,745,796	solcap_snp_sl_15381	3.2	0.019	0.008	-0.007	-0.021

	9	15	1,514,198	solcap_snp_sl_7731	2.9	0.009	0.007	-0.015	-0.001
	9	62	7,156,689	solcap_snp_sl_45141	3.3	-0.031	-0.123	0.165	-0.011
	9	103	70,334,215	solcap_snp_sl_69669	3.0	0.136	0.105	0.090	-0.332
	10	6.8	1,395,637	solcap_snp_sl_46339	5.9	0.011	0.012	0.054	-0.077
	12	15	1,971,483	solcap_snp_sl_12656	4.9	0.130	0.020	-0.109	-0.041
	12	55.7	63,393,430	solcap_snp_sl_55530	5.1	0.046	0.043	-0.043	-0.046
Harvest index									
	4	71	61,874,267	solcap_snp_sl_3687	8.3	-0.0010	0.0000	0.0007	0.0003
	4	80.4	63,919,364	solcap_snp_sl_47385	7.5	-0.0113	-0.0157	0.0071	0.0198
	5	42.8	6,247,585	CL016061-0368	8.3	-0.0040	-0.0202	0.0198	0.0043
	5	76	63,803,770	solcap_snp_sl_185	10.7	0.0119	0.0121	-0.0015	-0.0225
	7	80.2	67,162,714	solcap_snp_sl_70636	3.2	-0.0252	0.0263	0.0357	-0.0368
	10	6.8	1,395,637	solcap_snp_sl_46339	3.6	-0.0199	0.0215	-0.0009	-0.0007
	12	15	1,971,483	solcap_snp_sl_12656	3.8	0.0159	0.0250	-0.0298	-0.0111
	12	52.4	62,821,376	solcap_snp_sl_42775	4.4	0.0014	0.0023	-0.0102	0.0065
Total biomass (kg m^{-2})									
	1	53.2	79,245,417	solcap_snp_sl_25931	3.2	-0.111	-0.015	0.128	-0.002
	2	23.4	38,960,460	Le001778_68_solcap_snp_sl_33474	6.3	-0.055	-0.045	0.062	0.038
	4	71	61,874,267	solcap_snp_sl_3687	8.8	0.123	0.109	-0.114	-0.118
	5	43.6	6,571,837	solcap_snp_sl_50925	9.0	0.114	0.194	-0.098	-0.210
	6	6.3	2,600,146	solcap_snp_sl_35072	4.0	-0.038	-0.044	0.048	0.034
	6	17.3	37,583,093	solcap_snp_sl_55858	4.5	0.116	0.000	-0.011	-0.105
	6	46	43,804,153	solcap_snp_sl_57810	3.6	0.154	0.025	0.011	-0.191
	6	59.6	47,291,843	solcap_snp_sl_57155	3.3	0.000	0.010	0.032	-0.042
	7	23.2	2,881,365	solcap_snp_sl_15786	5.6	0.009	0.017	-0.014	-0.012
	7	47.6	61,442,206	solcap_snp_sl_53538	5.2	0.066	0.015	-0.023	-0.058

8	46	58,745,796	solcap_snp_sl_15381	3.3	0.277	-0.059	-0.284	0.066
10	52.9	62,277,373	solcap_snp_sl_59337	4.2	-0.032	-0.026	0.085	-0.026

Chapter 3

Yield dissection models with quantitative genetics to improve yield: a case study in greenhouse-grown tomato in the Mediterranean climate

Yutaka Tsutsumi-Morita[1,2], Ep Heuvelink[1], Daniela Bustos-Korts[2], Martin P. Boer[2], Carmen Saez Cuesta[1], Kim M.C.A. Vermeer[3], Hannelore van Stappen-van Dijk[3], Frank F. Millenaar[3], George A.K. Van Voorn[2], Leo F.M. Marcelis[1], Fred A. Van Eeuwijk[2]

[1] Wageningen University & Research, Horticulture and Product Physiology, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[2] Wageningen University & Research, Biometris, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[3] BASF's vegetable seeds business (Nunhems), Napoleonsweg 152, 6083 AB Nunhem, The Netherlands

Abstract

Introgression of alleles from wild genetic resources has been proposed as a strategy to increase yield via the improvement of its underlying traits. However, identifying the genetic loci that increase yield and the traits that drive this improvement remains challenging. To address this challenge, in our previous study we dissected yield into component traits via yield dissection models (harvest or biomass dissections) for greenhouse-grown indeterminate tomato in the Netherlands (NL). Harvest dissection models capture fruit-sink capacity, and biomass dissection models represent biomass production (total plant biomass), allocation (harvest index) and dilution of fruit dry matter (fruit fresh-dry weight ratio). Yield dissection models combined with quantitative genetics revealed how wild germplasm improved component traits, which subsequently improved yield. In this study, we validated the usefulness of this methodology by using another tomato population in another environment, Spain. The hybrids between two test parents and 346 recombinant inbred lines (RILs) originated from three parents (one elite and two wild parents) were grown in two greenhouses in Spain. The aim of this study was (1) to identify component traits and quantitative trait loci (QTLs) that contribute to improving the yield of the elite parent by introducing wild germplasm in the Spanish breeding population; (2) to compare the resulting breeding strategy for high yield in Spain with our previous study in NL; (3) to evaluate whether yield component traits show less G×E than yield itself. Phenotypic and QTL analyses showed that wild parents brought much more yield improvement in Spain than in NL. Introgressed wild germplasms created a trade-off between the number of fruits and individual fruit fresh weight in the harvest dissection, similarly to NL. However, the fruit fresh-dry weight ratio did not correlate with yield in the biomass dissection in Spain, despite a positive correlation observed in NL. The strategy to breed a high-yielding line appeared to be similar between Spain and NL. That is, yield improvements can be more attributed to the number of fruits than individual fruit fresh weight, more attributed to total fruit dry weight than fruit fresh-dry weight ratio, and more attributed to total biomass than harvest index. Some QTLs were at the same positions between Spain and NL, which are more likely to be real QTLs. Some component traits (e.g., number of fruits per truss, individual fruit fresh weight, fruit fresh-dry weight ratio) showed less G×E than yield — higher heritability, higher correlation and higher QTL colocalization between both greenhouses in Spain, and higher prediction accuracy. Less G×E in some component traits indicates that these component traits might be more stable in selections over years. Yield prediction via the harvest dissection model using QTLs for component traits showed a similar accuracy as yield prediction by QTLs for yield itself. Thus, the allele effects of component traits can be transformed into the effects on yield via yield dissection models. These results can be used by breeders to design high-yielding lines with a more advantageous composition of component traits with positive allele combinations.

3.1. Introduction

Yield improvement is one of the most important targets for plant breeding (Duvick *et al.*, 2004; van Eeuwijk *et al.*, 2019). To further improve crop yield, it is frequently suggested to introduce genetic variation from wild germplasm, including new yield alleles, into elite cultivars (Tanksley and McCouch, 1997; Schulthess *et al.*, 2022). However, the evaluation of such high-yielding lines remains challenging (van Eeuwijk *et al.*, 2019). Yield is a complex trait, which is an integrated outcome of many underlying physiological processes, each of which is controlled by many genes under the influence of the environment and crop management (Dekkers and Hospital, 2002; Eeuwijk *et al.*, 2010). One way to disentangle these complex relationships is by dissecting yield into underlying component traits (Jackson *et al.*, 1996; Yin *et al.*, 2002). Component traits are often simpler, controlled by fewer genes, and influenced by the environment in a less complex way (less interaction between genotype and environment (G×E)) than yield (Fukai *et al.*, 1999; Hammer *et al.*, 2005; Cooper *et al.*, 2009; Parent and Tardieu, 2014). If a component trait has a large effect on yield, it might be used to indirectly select for improved yield (Reymond *et al.*, 2003; Chenu *et al.*, 2008; Maccaferri *et al.*, 2008; Velazco *et al.*, 2019). The effect of a component trait on the target trait can be reduced if there are strong trade-offs between component traits (Yin and Struik, 2008). Trade-offs can have physiological or genetic reasons. Examples of physiological trade-offs are common among traits that compete for resources (Acreche and Slafer, 2009; Li *et al.*, 2015). In contrast, genetic trade-offs occur when several pleiotropic QTLs have opposite effect signs. Yield dissection models with quantitative trait loci (QTLs) are useful to reveal the relationships between traits and to identify QTLs for complementary traits — traits forming an upper trait (e.g., yield) in a complementary way, where different combinations of component traits give rise to the same yield (equifinality). By using the information about QTLs for complementary component traits, breeders can inform their selection process, reducing trade-offs and leading to improved yield (Yin *et al.*, 2002; Bustos *et al.*, 2013; Reynolds and Langridge, 2016; Tsutsumi-Morita *et al.*, 2021).

This idea was evaluated in our previous paper on indeterminate greenhouse-grown tomato (*Solanum lycopersicum*) in one environment, Westland in the Netherlands (NL) (Tsutsumi-Morita *et al.*, 2021). We used two yield dissection models; harvest dissection and biomass dissection. The harvest dissection model dissects yield into the number of harvested fruits and average fruit weight. Thus, the harvest dissection model captures fruit-sink capacity. The fruit-sink capacity is not identical to fruit-sink strength (Marcelis, 1996), which is difficult to measure for many genotypes in plant breeding. The strategy of harvest dissection has also been used in cereal crops (Yin *et al.*, 2002; Millet *et al.*, 2016, 2019). On the other hand, the biomass dissection model captures source strength: biomass production (total plant biomass), allocation (harvest index) and dilution of fruit dry matter with water (fruit fresh-dry weight ratio). The biomass dissection model is a simplified version of crop growth models, which have been broadly used in agronomy and horticulture (Jones *et al.*, 1991; Heuvelink and

Bertin, 1994; Scholberg *et al.*, 1997). The combination of harvest and biomass dissection models with QTLs, can guide breeders how to improve the yield of the elite parent by introducing wild germplasm. It answers the question which component traits to focus on and which QTLs.

In this paper, we validated the methodology of using yield dissection models with QTLs for another tomato population in another environment, Almeria in Spain. We investigated the strategy of plant breeding for high yield in Spain and assessed the consistency of the breeding strategy between Spanish and Dutch trials (previous study (Tsutsumi-Morita *et al.*, 2021)). Almeria and Westland are two of the largest production regions of indeterminate greenhouse-grown tomato for fresh consumption in Europe. The warm Mediterranean climate of Almeria is suitable for greenhouse tomato production during winter without heating, but too hot to produce tomato during summer. In contrast, the cool climate of Westland has the advantage to produce tomato during summer, but in winter the amount of radiation is too little to produce tomato without supplementary lighting. Not only climate but also production systems and management are different between Almeria and Westland, making both production systems an interesting case study to evaluate the potential of the yield dissection approach. The most relevant differences between these production regions are; tomatoes are planted in sand and soil in Almeria, where accumulated salt stress is a major concern (Luis Caparrós-Martínez *et al.*, 2020); whereas in NL tomatoes are grown on substrates, mostly stonewool, where nutrient solutions are controlled automatically. The environment inside the greenhouse is controlled manually by opening top and side windows in Almeria; whereas the environment is computer-controlled by opening top windows, heating, and carbon dioxide fertilization in NL (Heuvelink, 2018). Because of these differences in production systems and environment, a breeding company developed different breeding populations for each region and used in this study.

The aim of this study was:

1. To validate the usefulness of the methodology of combining yield dissection models with quantitative genetics for high-yielding breeding.
    1-1. To reveal which component traits explain more phenotypic variation in yield and have more colocalized QTLs with yield.
    1-2. To reveal how to improve the yield of the elite parent by introducing wild germplasm, in which direction (via which component trait), by using which QTLs.
2. To assess whether the breeding strategies for high yield in Spain are similar or different compared to the Dutch trial (different populations in different environments).

3. To investigate whether yield component traits show less G×E than yield itself.

A tomato breeding population, developed for Spanish production, consisted of hybrids between two test parents and 346 recombinant inbred lines (RILs) originating from three parents (one elite parent and two wild parents). This population was tested in two greenhouses (two locations) in Almeria (Spain) enabling us to evaluate whether yield component traits show less G×E than yield itself. We determined the genetic basis of yield and component traits by quantitative trait locus (QTL) analysis in a mixed model approach (Li *et al.*, 2021) in which we constructed design matrices to test for QTL effects from identity by descent (IBD) probabilities that were calculated using observed genetic marker information (Zheng *et al.*, 2015). IBD probabilities present probabilities of parental origin, which are more informative for QTL mapping purposes than the original observed marker data (identity by state: IBS) in a multi-parent population. We assessed the consistencies and differences of high-yield composition by component traits and underpinning QTLs for these relationships, which leads to consistent or different strategies of high-yielding breeding between Spanish and Dutch trials. We validated the usefulness of combining yield dissection models with quantitative genetics for plant breeding for high yield.

3.2. Materials and methods

3.2.1. Yield dissection models

We used two types of yield dissection models; one based on harvest (harvest dissection) and another based on biomass production (biomass dissection), largely the same as in our previous study (Tsutsumi-Morita et al., 2021) (Figure 1).

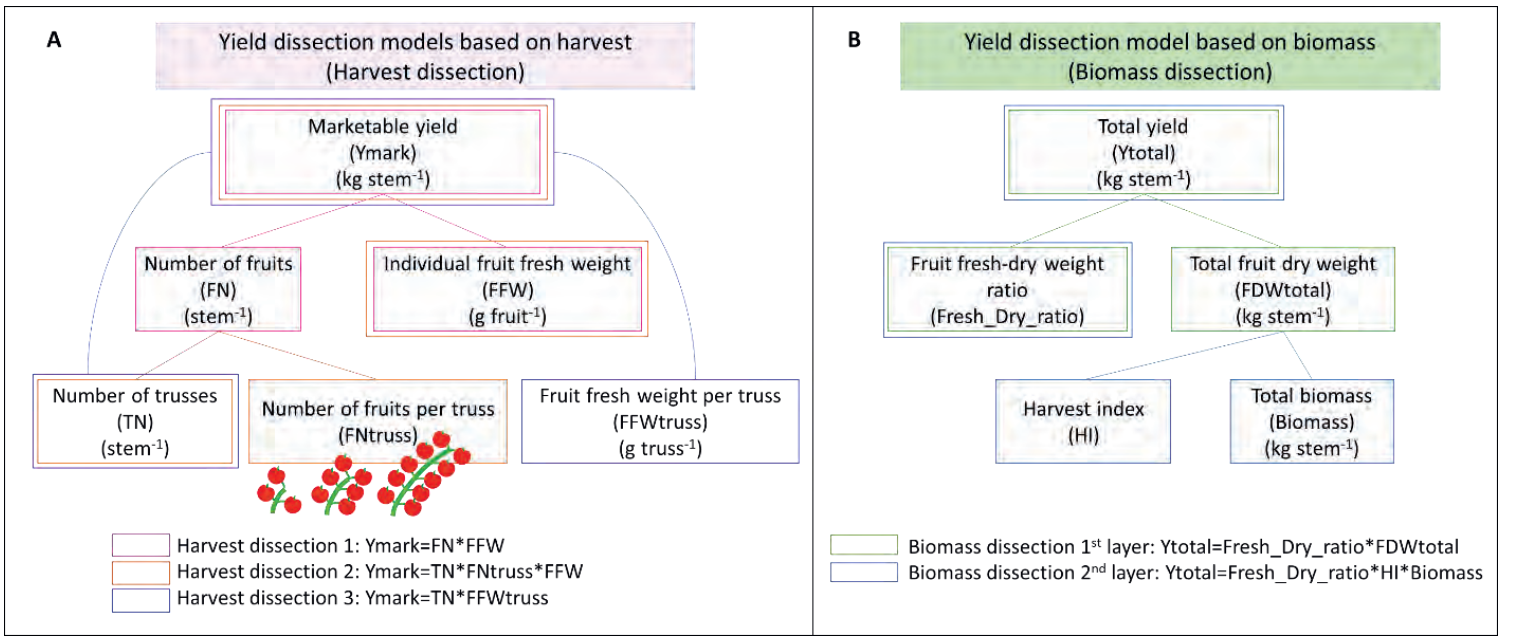


Figure 1 Dissection models for yield. (A) Harvest dissection and (B) Biomass dissection.

In the Spanish experiments, flowers were not pruned. Accordingly, there are three harvest dissection models. The first one is the same as in our previous study (Tsutsumi-Morita *et al.*, 2021); marketable yield (*Ymark*) is the product of the number of fruits (*FN*) and the individual fruit fresh weight (*FFW*) (Eq. 1).

Harvest dissection 1:

$$Ymark = FN \times FFW \qquad Eq.\ 1$$

FN is the product of the number of trusses (*TN*) and the number of fruits per truss (*FNtruss*).

Harvest dissection 2:

$$Ymark = TN \times FNtruss \times FFW \qquad Eq.\ 2$$

There is another possibility to dissect *Ymark* into the harvested fruits; *Ymark* is the product of *TN* and the fruit fresh weight per truss (*FFWtruss*).

Harvest dissection 3:

$$Ymark = TN \times FFWtruss \qquad Eq.\ 3$$

In the biomass dissection, total yield (*Ytotal*) is considered from another perspective; how much total biomass was produced, which fraction of total biomass has been allocated to fruits, and how much water went into fruits to dilute the fruit dry mass. *Ytotal* is first the product of fruit fresh-dry weight ratio (*Fresh_Dry_ratio*) and the total fruit dry weight (*FDWtotal*).

Biomass dissection 1st layer:

$$Ytotal = Fresh\_Dry\_ratio \times FDWtotal \qquad Eq.\ 4$$

FDWtotal is further the product of harvest index (*HI*) and total biomass above ground (*Biomass*). *HI* is the ratio between *FDWtotal* and *Biomass*. This relationship is incorporated into Eq. 4.

Biomass dissection 2nd layer:

$$Ytotal = Fresh\_Dry\_ratio \times HI \times Biomass \qquad Eq.\ 5$$

The yield dissection models in this study are comparable to the yield dissection models in the previous study (Tsutsumi-Morita *et al.*, 2021) except for the inclusion of *TN*, *FNtruss*, and *FFWtruss* in the harvest dissection in this study. This difference occurred because flowers

were not pruned in this study in Spain, but flowers were pruned and *FNtruss* was fixed to six in the previous study in the Netherlands (Tsutsumi-Morita *et al.*, 2021).

3.2.2. Breeding population

The population used in this study consisted of indeterminate tomato hybrids, specifically developed for production in greenhouse in Spain. Two wild parents (W1, W2) were introgressed into an elite parent (EP), and 346 recombinant inbred lines (RILs) were developed. The RILs were crossed with two testers (TP1, TP2), producing two populations of 346 genotypes (RILs × TP1 and RILs × TP2) (Figure 2). Although phenotyping was conducted on hybrids, the QTL analyses related the phenotypes to markers in the RILs.

The RILs were genotyped with a set of 192 SNP SOLCAP markers (Sim *et al.*, 2012). We used a genetic map based on a publicly available map, TraitGenetics EXPIMP2012 (https://solgenomics.net/cview/map.pl?map_version_id=149). Together with the map, the markers formed the basis for identity by descent (IBD) calculations between parents and offspring that served to create design matrices for genome-wide testing of QTL additive effects. The calculation of IBD probabilities was done via a Hidden Markov Model by the Mathematica package RABBIT (Zheng *et al.*, 2015). Monomorphic markers and segregation-distorted markers (97% of RILs originated from the elite parent with more than 90% probability) were removed, and the resulting 171 polymorphic markers were used for QTL analyses. At chromosome 9, marker positions were reordered by RABBIT (Zheng et al., 2019).

3.2.3. Phenotypic data collection

We phenotyped the traits of the yield dissection models in two greenhouses. Each trait was measured at plot level. Then, the observed phenotypic values were corrected for spatial trends. Finally, the spatially-corrected mean per RIL per greenhouse was used for further analyses.

3.2.3.1. Growing conditions

Plants were transplanted on soil in two commercial tomato greenhouses, Fernando (Fer) and Juan Manuel (JM), in Almeria in Spain in September (week 37) in 2016. JM greenhouse was located closer to the sea than Fer greenhouse. Plants were cultivated according to the commercial practice (Supplementary Figure 1) — hanging system, where plants were supported by ropes and the positions of ropes were adjusted according to stem growth (high-wire system); no flower pruning; drip irrigation system; no automatic climate control, where top and side windows were opened and closed manually; shading by white wash, which was sprayed on top of the roof in the middle of February; shoot top removal, which was conducted at the beginning of March. Harvest was conducted every two weeks. The first harvest was in week 48 (28 November 2016) and the last harvest was in week 19 (8 May 2017) in the Fer

greenhouse. The first harvest was in week 49 (5 December 2016) and the last harvest was in week 18 (1 May 2017) in JM greenhouse.

3.2.3.2. Experimental design

A randomized complete block design was employed (Supplementary Figure 2). Our population (RILsTP1, RILsTP2) was repeated in two commercial greenhouses (Fer, JM). Each greenhouse had two blocks; RILsTP1 were planted in block 1 and RILsTP2 were in block2 for each greenhouse respectively, and RILs were randomized over the planting positions. The Fer greenhouse consisted of 754 plots and JM greenhouse consisted of 740 plots, which included 346 RILs crossed by test parent 1 (RILsTP1), 346 RILs crossed by test parent 2 (RILsTP2), elite parent crossed by test parent 1 (EPTP1: 10 times in Fer, 11 times in JM), elite parent crossed by test parent 2 (EPTP2: 7 times in Fer, 11 times in JM) and 5 cultivars (each cultivar was repeated 6 times in block 1 in Fer, 3 times in block 2 in Fer, 2-3 times respectively in block 1 and block 2 in JM) (Table in Supplementary Figure 2). Each plot included four plants (one stem per plant).

3.2.3.3. Phenotyping: Measurements of yield and component traits per plot

Yield and component traits in yield dissection models were phenotyped at plot level. Overview of phenotyping was shown in Supplementary Table 1. Every two weeks, tomatoes were harvested. When they were harvested, Total yield (Ytotal), marketable yield (Ymark), and component traits in harvest dissection (the number of harvested marketable fruits (FN), the number of trusses (TN), and the number of fruits per truss (FNtruss), individual fruit fresh weight (FFW), fruit fresh weight per truss (FFWtruss)) were measured (phenotyped) for all plots of 4 plants in each greenhouse through the whole growing season. Component traits in the biomass dissection were phenotyped once for 1 plant per plot in a subset of plots. The subset was selected by uniform coverage of genetic space (Bustos-Korts *et al.*, 2016), which included 110 genotypes in 160 plots (Supplementary Figure 3). In this subset, 10 RILs were observed in both test-parent crosses (RILsTP1, RILsTP2) in both greenhouses, 10 RILs were in crosses of test parent 1 (RILsTP1) in both greenhouses, 10 RILs were in crosses of test parent 2 (RILsTP2) in both greenhouses, and 20 RILs respectively in each test-parent cross in each greenhouse (Supplementary Figure 3).

3.2.4. Phenotypic and QTL analyses

An overview of the analyses and corresponding results are presented in Figure 3. Phenotypic and QTL analyses were conducted per greenhouse (Fer, JM). G×E was evaluated by comparing both greenhouses, in Spain.

3.2.4.1. Phenotypic analysis

First, spatially adjusted genotypic means (BLUEs: best linear unbiased estimates) and the generalized heritability were calculated by the R-package SpATS (Spatial Analysis of Field Trials with Splines) using a two-dimensional Penalized spline (P-spline) mixed model (Velazco *et al.*, 2017; Rodríguez-Álvarez *et al.*, 2018) similar to our previous study (Tsutsumi-Morita *et al.*, 2021). Additionally-replicated genotypes (EPTP1, EPTP2, and five cultivars) (Table in Supplementary Figure 2) facilitated the spatial corrections.

Histograms of the BLUEs for yield and component traits were generated to present the proportion and magnitude of transgression. Pearson correlations between both greenhouses per trait were calculated to investigate G×E for each trait at the phenotypic level. Pearson correlations between traits were adopted to evaluate the general relationships between traits, whereas response surface plots for yield as a function of component traits (BLUEs) enabled us to investigate each RIL. A RIL may be superior to the elite parent for both or only one component at the phenotypic level.

3.2.4.2. QTL analysis

The IBD probabilities defined haplotype IBD matrices of dimensions equal to the number of RILs by the number of parents that formed the basis for a mixed model QTL mapping procedure as described by Li et al. (2021, 2022) in the same way as our previous study (Tsutsumi-Morita *et al.*, 2021).

A single locus QTL scan was performed to test for the significance of the variance of the haplotype, or QTL allele, effects corresponding to the three parent alleles at each marker position. The test consisted of a loglikelihood ratio test for the variance component within the package ASReml (Butler et al., 2017). A threshold for significance was put at $-\log_{10}P \geq 2.9$ for yield and component traits in the harvest dissection, whereas lower thresholds ($-\log_{10}P \geq 0.9$ *or* 0.8) were employed for component traits in the biomass dissection because phenotyping was conducted on a subset of RILs and it was therefore difficult to detect QTLs (Table 1). Two adjacent QTLs were identified as different QTLs if their distance exceeded 20 cM, or when the $-\log_{10}P$ value dropped by more than 0.5 between them. Following a single-QTL model scan, a multi-QTL model was developed for each trait by a backward elimination from a model including all markers that were identified as significant in the initial single-QTL scan. QTLs were removed from the final multi-QTL model when their variance component was zero.

To investigate the colocalization of QTLs for each trait between both greenhouses, and between the Spanish and Dutch experiments, the positions of single-locus QTLs were used. For comparison of two experiments, a window of 20 cM was applied. For comparison among three experiments (two Spanish and one Dutch experiment), three QTLs were considered as

colocalized if a QTL located within 20 cM from at least one QTL of the other two experiments, and the distance between two distant QTLs were less than 25 cM. Haplotype effects were not inspected to determine co-localizations. The percentage of colocalized QTLs per trait was used to assess G×E at the genotypic level.

To investigate the colocalization of QTLs between yield and component traits, and also between component traits, within the same greenhouse, multi-QTL models and their haplotype effects were used. Two QTLs were considered to be colocalized, if these two QTLs located within 20 cM from each other. Additionally, the parental origins of the QTL allele with the largest magnitude (haplotype) were investigated to decide whether they coincided between traits. These QTL-allele effect comparisons were not possible for single-QTL comparison between the Spanish and Dutch experiments because of the different populations developed from different parents.

3.2.4.3. Prediction

Predictive qualities of QTLs for yield and QTLs for component traits were assessed by predictions from the multi-QTL models. First, we predicted yield and component traits by multi-QTL models for the trait itself. Subsequently, we also predicted yield from component traits that were predicted by corresponding component-trait multi-QTL models and were inserted into dissection models. Prediction accuracies (correlation coefficient between observation and prediction) were evaluated in two ways: (1) by cross-validation within each greenhouse and (2) by validation over two greenhouses.

A five-fold cross-validation scheme was implemented. Data were randomly partitioned into five folds, with four folds being included in a training set to develop a multi-QTL model and one fold serving as a validation set. For traits phenotyped on 346 RILs (total yield, marketable yield, and component traits in the harvest dissection), this five-fold cross-validation was repeated 20 times for different random partitionings of the data. For traits phenotyped on 70 RILs per greenhouse (component traits in biomass dissection), the cross-validation was repeated 130 times, where around 50 RILs were included in a training set and 20 RILs serving as a validation set (Table 2; Supplementary Figure 8). This cross-validation (50 RILs for training, 20 RILs for validation) was also applied to traits phenotyped on 346 RILs for a fair comparison between the harvest dissection and the biomass dissection (Supplementary Figure 8).

To further investigate to what extent QTL effects were consistent between both greenhouses (Fer and JM greenhouses), we predicted traits in a greenhouse by the multi-QTL models determined for the other greenhouse. First, as a reference, the prediction accuracy of the fitted model was evaluated, in which the same RILs were used for a training set and a validation set within each greenhouse. Second, the same RILs were used for a training set and a validation set, but the training set and the validation set employed phenotypic values from

the other greenhouse. That is, a multi-QTL model was developed on data from a greenhouse, which was used to predict the same RILs in the other greenhouse (Table 2). Third, 70 RILs in a greenhouse were used for a training set; and all RILs, for which observations were available, in the other greenhouse were used for a validation set. That is, 70 RILs were used for developing a multi-QTL model in a greenhouse, which was used for predicting RILs in the other greenhouse including RILs that were not used in the training set (Supplementary Table 4; Supplementary Table 5).

The data from the experiment in the Netherlands were already presented in our previous study (Tsutsumi-Morita *et al.*, 2021). The data from the Netherlands were included in this study to compare the results between the Spanish and Dutch experiments (different breeding populations in different environments).

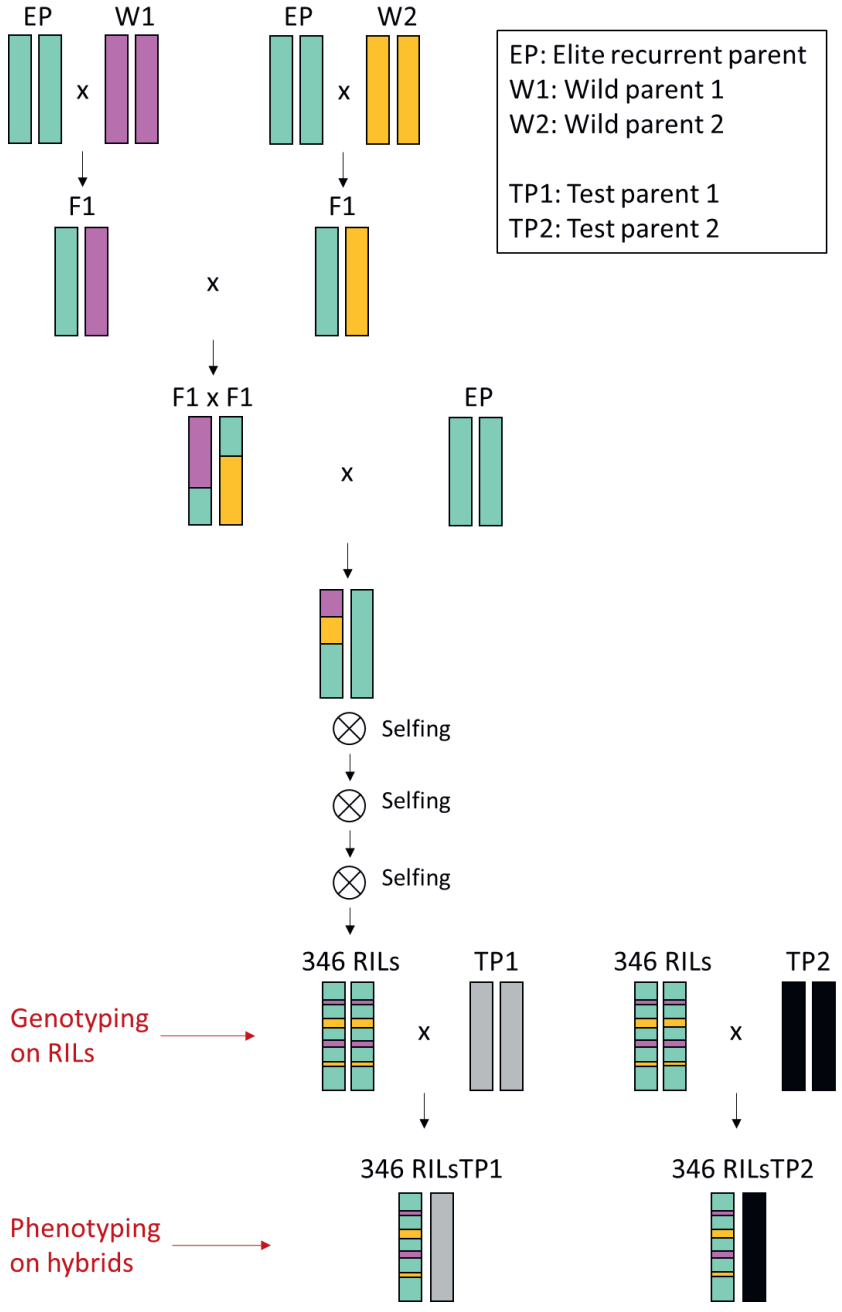


Figure 2 Construction of RILs and corresponding hybrids. An elite parent (EP) and two wild-type parents (W1, W2) were founders for offspring population of 346 RILs. The RILs were crossed with two testers (TP1, TP2). The resulting hybrids (RILsTP1, RILsTP2) were phenotyped.

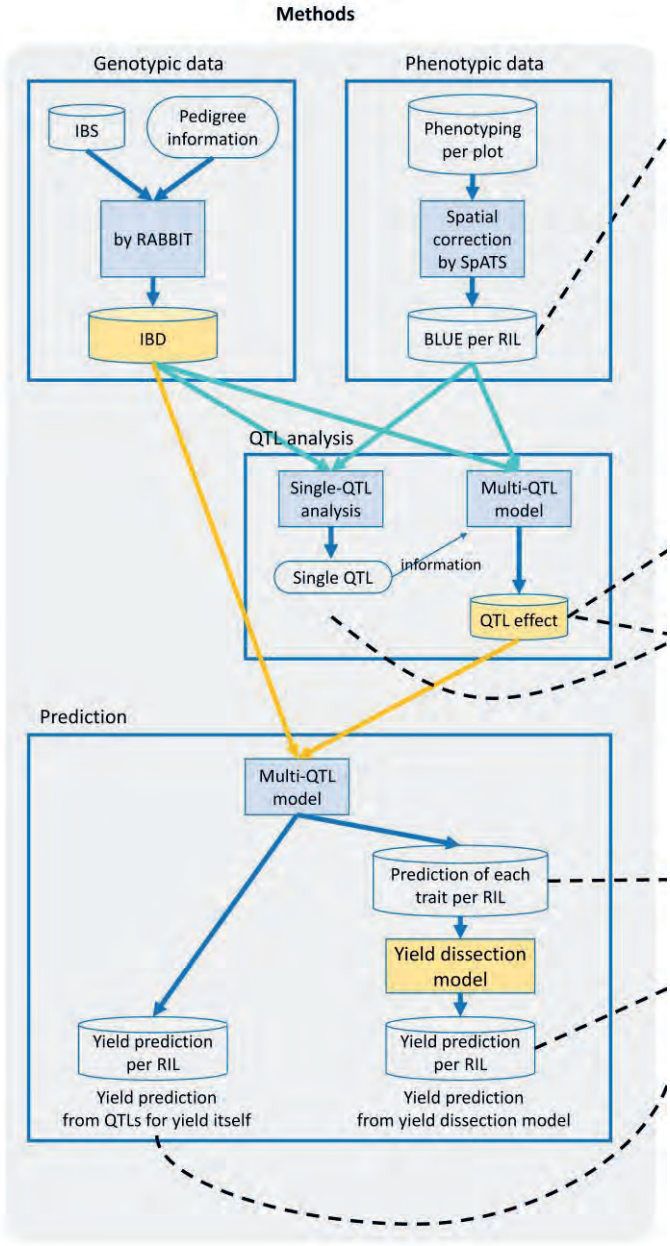


Figure 3 Overview of statistical analyses as presented in the Material and Methods section and corresponding Results section. Abbreviations: BLUE, best linear unbiased estimate; IBD, identity by descent; IBS, identity by state; RIL, recombinant inbred line. Software: RABBIT, a Mathematica software to calculate IBD; SpATS, an R-package to estimate spatially adjusted genotypic means.

Results

< Phenotypic analysis >

Results Section	Spanish trial		Comparison with a different population in the Netherlands
		G×E evaluation	
3.1.1.	Heritabilities (Table 1)		
3.1.2.	Genotypic means, transgression (Fig. 4)	Correlation of each trait between two Spanish greenhouses (Fig. 4)	
3.1.3.	Correlations between traits (Fig 6. AB)		Comparison with NL (Fig. 6 AB)
3.1.4.	Transgression on yield dissection models (Fig. 7)		Comparison with NL (Fig. 7)

< QTL analysis >

Results Section	Spanish trial		Comparison with a different population in the Netherlands
		G×E evaluation	
3.2.1.	QTL detection (Table 1; Fig. 8)		
3.2.2.		Single-QTLs at the same position between two greenhouses in Spain (Fig. 9)	Comparison with NL (Fig. 9)
3.2.3.	Multi-QTLs colocalization between yield and components (Fig. 8; Fig. 10)		Comparison with NL (Fig. 8; Fig. 10)
3.2.4.	Co-localized QTLs alignment with phenotypic correlations between traits (Fig. 6)		Comparison with NL (Fig. 6)

< Prediction >

1. Prediction of each trait by own QTLs

2. Yield prediction from yield dissection models by using QTLs for yield component

Results Section	Spanish trial		Comparison with a different population in the Netherlands
		G×E evaluation	
3.3.1.	Cross-validation (Table 2)		
3.3.2.		Prediction in a greenhouse by QTL effects from the other greenhouse (swap greenhouses) (Table 2)	

Figure 3 Continued.

Table 1 Trait heritabilities and numbers of QTLs. Haplotype indicates the number of multi-QTLs, for which each parent showing the highest positive allele effect.

Trait	Abbreviation	Number of RILs	Heritability	Threshold (- $\log_{10}$ P)	Number of peaks	Number of multi-QTLs	Haplotype			Prediction accuracy of fitted model (r)
							Elite parent	Wild parent 1	Wild parent 2	
Fer greenhouse										
Total yield	Ytotal	346	0.12	2.9	7	6	3	1	2	0.40
Marketable yield	Ymark	346	0.12	2.9	7	7	4	1	2	0.40
Number of fruits	FN	346	0.20	2.9	20	10	2	4	4	0.59
Number of trusses	TN	346	0.11	2.9	17	9	5	3	1	0.50
Number of fruits per truss	FNtruss	346	0.62	2.9	27	17	5	5	7	0.76
Individual fruit fresh weight	FFW	346	0.65	2.9	26	17	12	2	3	0.81
Fruit fresh weight per truss	FFWtruss	346	0.38	2.9	26	15	9	2	4	0.66
Fruit fresh-dry weight ratio	Fresh_Dry_ratio	70	0.62	0.9	22	11	5	3	3	0.88
Total fruit dry weight	FDWtotal	70	0.55	0.9	17	10	3	3	4	0.75
Harvest index	HI	69	0.27	0.9	16	7	3	4	0	0.72
Total biomass	Biomass	69	0.02	0.9	10	6	4	2	0	0.71

JM greenhouse										
Total yield	Ytotal	346	0.26	2.9	15	11	5	4	2	0.58
Marketable yield	Ymark	346	0.30	2.9	16	9	4	4	1	0.59
Number of fruits	FN	346	0.42	2.9	18	10	4	3	3	0.68
Number of trusses	TN	346	0.53	2.9	24	14	7	2	5	0.70
Number of fruits per truss	FNtruss	346	0.52	2.9	24	15	6	6	3	0.76
Individual fruit fresh weight	FFW	346	0.60	2.9	22	10	7	1	2	0.69
Fruit fresh weight per truss	FFWtruss	346	0.39	2.9	13	9	5	3	1	0.54
Fruit fresh-dry weight ratio	Fresh_Dry_ratio	70	0.87	0.9	16	11	6	3	2	0.75
Total fruit dry weight	FDWtotal	70	0.68	0.9	9	6	4	0	2	0.62
Harvest index	HI	70	0.83	0.9	16	9	5	1	3	0.77
Total biomass	Biomass	70	0.26	0.8	8	6	4	1	1	0.47

3.3. Results

3.3.1. Phenotypic analyses

3.3.1.1. Heritabilities for the Spanish greenhouses

Heritabilities and spatially corrected genotypic means were first computed by mixed models using SpATS R-package. The heritabilities give an impression of traits genetic determination and the possibilities for identifying QTLs from the phenotype. The heritabilities for total yield and marketable yield were low (both 0.12) in the Fernando (Fer) greenhouse and relatively low in the Juan Manuel (JM) greenhouse (0.26 and 0.30 respectively) (Table 1). Therefore, for this RIL population, direct selection for yield based on yield QTLs would be difficult in the Fer greenhouse; whereas it might be feasible in the JM greenhouse. The heritabilities of most component traits were higher than those for yield, except for the number of trusses in Fer (0.11) and total biomass in both greenhouses (0.02 in Fer, 0.26 in JM). Especially, the heritabilities of some components were much higher than those for yield: number of fruits per truss (0.62 in Fer, 0.52 in JM), individual fruit fresh weight (0.65 in Fer, 0.60 in JM), fruit fresh-dry weight ratio (0.62 in Fer, 0.87 in JM), and total fruit dry weight (0.55 in Fer, 0.68 in JM). Heritability of harvest index was very high in JM (0.83), however rather low in Fer (0.27).

3.3.1.2. Genotypic means, transgression and correlation of each trait between two Spanish greenhouses

We examined the distribution of the traits to quantify the amount of transgression (progeny showing larger phenotypic values than the elite parent). Around half of RILs showed transgression for total yield and marketable yield (Figure 4A, B), fruit fresh-dry weight ratio, total fruit dry weight in both greenhouses, and for harvest index in the JM greenhouse (Figure 4H, I, J). More than half of the RILs showed transgression for the number of fruits, number of trusses, number of fruits per truss, and total biomass in both greenhouses (Figure 4C, D, E, K). In contrast, less transgression was observed for individual fruit fresh weight, fruit fresh weight per truss in both greenhouses, and harvest index in Fer greenhouse (Figure 4F, G, J), for which more than half of the RILs showed lower phenotypic values than the elite parent. The large amount of transgression for many traits reflects the large complementarity of parents in these crosses, making it a good data for studying the dissection and QTL approach.

Plasticity was observed for component traits in the harvest dissection. The number of fruits and number of trusses were higher in Fer greenhouse than in JM greenhouse (Figure 4c, d), whereas it was the other way around for the number of fruits per truss, individual fruit fresh weight, and fruit fresh weight per truss (Figure 4e, f, g).

Next, we examined the phenotypic correlations of traits between both Spanish greenhouses, to get an impression of potential genotype × environment interaction (G×E). For total yield and marketable yield, the correlations between Fer and JM were relatively low, 0.36 (Figure 4a, b), indicating the presence of G×E with potential rank changes across greenhouses, which challenges the selection of stable and high-yielding varieties. In contrast, correlations between Fer and JM greenhouses for the number of fruits per truss (0.60), individual fruit fresh weight (0.62) and fresh-dry weight ratio (0.72) (Figure 4e, f, h) were higher than for yield suggesting a more similar RILs ranking between both greenhouses, supporting the idea that a dissection approach might be valuable to inform the selection of candidate varieties.

3.3.1.3. Correlations between traits for yield dissection models

– *Spanish greenhouses*

The same yield can be achieved via different component traits, so-called equifinality (Figure 5A, B). Hence, the genetic gain can be accelerated when explicitly selecting for complementary component traits. This strategy works best when yield and a component trait are correlated, which indicates that traits are informative about each other and can potentially be co-selected. It also informs whether there is a trade-off between traits, expressed as a negative correlation at the population level.

In the harvest dissection, marketable yield (Ymark) positively correlated ($r > 0.3$) with all component traits except for the number of fruits per truss (FNtruss), which did not correlate (Figure 6A; Supplementary Figure 4A, B). However, a negative correlation between the number of fruits (FN) and individual fruit fresh weight (FFW) was observed, which mitigated the positive correlations of marketable yield with these two components. This negative correlation constrains the range of yield response to selection when selecting for yield only and indicates that an explicit consideration of FN and FFW as selection target shows promise to achieve higher yield by a selection of complementary component traits. The number of fruits was more correlated with the number of trusses (TN) than with the number of fruits per truss (FNtruss), and these two component traits were weakly negatively correlated with each other. A weak negative correlation between the number of trusses and fruit fresh weight per truss (FFWtruss) was observed, which mitigated the positive correlations of marketable yield with these two component traits.

In the biomass dissection, total yield (Ytotal) was positively correlated with total fruit dry weight (FDWtotal) and total biomass (Biomass) (Figure 6B; Supplementary Figure 4C, D). Fruit fresh-dry weight ratio (Fresh_Dry_ratio) and total fruit dry weight (FDWtotal) negatively correlated with each other, suggesting a trade-off between both traits. Total fruit dry weight correlated with harvest index (HI) and also total biomass (Biomass). However, in the Fer greenhouse, both traits were negatively correlated, indicating that co-selecting for harvest index and biomass might be difficult, unless an explicit selection pressure is applied

on both traits. The relationship between harvest index and total biomass was not consistent between Fer and JM greenhouses.

– *Comparison with Dutch trial (different population in different environment)*

In the harvest dissection (Figure 6A), Pearson correlations were similar for the Spanish and Dutch (NL) trials (different populations in different environments). Both trials show that yield can benefit from explicitly selecting for increased number of fruits and individual fruit fresh weight simultaneously because of the negative correlation between these component traits.

The results for the biomass dissection model (Figure 6B) were less consistent between Spain and NL than those for the harvest dissection. The following relationships were similar between these two trials: yield correlated with total fruit dry weight (in JM) and total biomass; yield was more correlated with total fruit dry weight than with fruit fresh-dry weight ratio; yield did not correlate with harvest index; total fruit dry weight correlated with total biomass. The following relationships were different between both trials: yield did not correlate with fruit fresh-dry weight ratio in Spain, while it did correlate in the Netherlands (NL); fruit fresh-dry weight ratio and total fruit dry weight negatively correlated with each other in Spain, whereas there was no correlation in NL; total fruit dry weight and harvest index positively correlated with each other in Spain, while they did not correlate in NL; harvest index and total biomass negatively correlated with each other in Fer and NL, while they did not correlate in JM.

3.3.1.4. Trade-off or synergy between component traits by introgression of wild germplasm, and resulting transgression

– *Spanish greenhouses*

Yield dissection models revealed (1) how the introgressed wild chromosome segments changed yield-component phenotypes of the elite parent in the segregating RIL population in the relationship with the other component trait, (2) how these changes affected yield, and (3) which component traits show promise to select for RILs showing a higher yield than the elite parent (Figure 5C; Figure 7). As mentioned, around half of the RILs showed transgressive segregation, achieving a higher yield than the elite parent. To investigate the reason for the higher yield of the superior RILs, the phenotypes of two underlying component traits were plotted on the x-axis and y-axis. Isoclines illustrating the same yield values, but achieved via a different combination of component traits were added (Figure 5; Figure 7). RILs above the isocline of the elite parent showed transgression of yield (yield is higher than for the elite parent) (located in area I (yellow colored), II (purple colored), and IV (green colored) in Figure 5C), RILs on the right side of the vertical line drawn at the x-value of the elite parent

showed transgression of the component trait on the x-axis (area I and IV in Figure 5C), and RILs above the horizontal line drawn at the y-value of the elite parent showed transgression of the component trait on the y-axis (area I and II in Figure 5C). These lines can be used by breeders to be informed about potential yield outcomes, depending on the selection pressure that is applied on each of the component traits.

In the harvest dissection, introgression of wild-parents chromosome segments mainly increased number of fruits at the same time with decreasing individual fruit fresh weight, resulting in a negative correlation between these two component traits (Figure 7A, B). In spite of this overall negative correlation, around half of the RILs showed a higher yield than the elite parent (RILs located at the upper side of the solid contour line; Figure 7A, B). Yield had more gain from a large number of fruits than from a large individual fruit fresh weight (the number of RILs in area IV was higher than the number of RILs in area II; Figure 5C; Figure 7A, B).

The number of fruits was further dissected into the number of trusses and the number of fruits per truss (Figure 7D, E). The number of fruits per truss and the number of trusses were only slightly correlated, which implies that there is a large set of possible configurations of both traits to achieve a similar number of fruits.

Yield also can be dissected into the number of trusses and fruit fresh weight per truss (Figure 7F, G). Yield had more gain from a large number of trusses than from high fruit fresh weight per truss (the number of RILs in area IV was higher than the number of RILs in area II; Figure 5C; Figure 7F, G).

In the biomass dissection, introgression of wild-parents chromosome segments tended to have opposite effects between fruit fresh-dry weight ratio and total fruit dry weight, which resulted in a negative correlation between these two component traits (Figure 7H, I). Yield had more gain from total fruit dry weight than from fruit fresh-dry weight ratio (the number of RILs in area II was higher than the number of RILs in area IV; Figure 5C; Figure 7H, I). The relationship between harvest index and total biomass showed different patterns between Fer and JM greenhouses (Figure 7K, L). Nevertheless, in both greenhouses, total fruit dry weight (isocline in Figure 7K, L) had more gain from total biomass than from harvest index (the number of RILs in area II was higher than the number of RILs in area IV; Figure 5C; Figure 7K, L).

- *Comparison with Dutch trial (different population in different environment)*

In the trial in the Netherlands (NL), much smaller transgression of yield was observed (3.2% for total yield, 3.8% for marketable yield; RILs located at upper side of the solid isocline in

Figure 7C) than in Spain. Nevertheless, the attributes of high-yielding lines were similar between Spain and NL.

In the harvest dissection in NL, most of exceeding RILs showed higher yield than the elite parents due to higher number of fruits to offset loss in individual fruit fresh weight, similar to Spain (Figure 7A, B, C).

In the biomass dissection, a large difference between trials was observed for the fruit fresh-dry weight ratio. In NL, most of the RILs showed a lower fruit fresh-dry weight ratio than the elite parent (Figure 7J), whereas the elite parent was located in the middle of the phenotypic distribution of RILs in Spain (Figure 7H, I). In other words, there was room for increasing fruit fresh-dry weight ratio in Spain via introgression of wild-type alleles into the elite parents, but this was not feasible for the Dutch population. Interestingly, it seems at first glance possible in Spain to improve yield by increasing fruit fresh-dry weight ratio, however, the negative correlation between fruit fresh-dry weight ratio and total fruit dry weight mitigated the positive effect of fruit fresh-dry weight ratio on yield. In spite of only a few RILs showing higher yield than elite parents in NL, the main attribute of high-yielding lines was similar as for Spain: the high total fruit dry weight with the same or slightly lower fruit fresh-dry weight ratio compared to the elite parent (area II in Figure 7J). Regarding harvest index and total biomass, Fer greenhouse showed a similar pattern as for NL, i.e. the negative correlation between harvest index and total biomass, less than half of the RILs showed transgression in harvest index (Figure 7K, M). However, a higher transgression in total biomass in Fer than that in NL, resulted in a higher transgression of total fruit dry weight (RILs above the solid isocline; Figure 7K) in Fer than that in NL (Figure 7M). RILs showing a higher total fruit dry weight than the elite parent (RILs located at upper side of the solid isocline) was attributed to the high total biomass with high or mildly lower harvest index compared to the elite parent (area I and II in Figure 7K, L, M), which was similar between Spain and NL.

3.3.2. QTL analyses

3.3.2.1. QTL detection in Spanish greenhouses

For an effective selection strategy based on complementary component traits, it is useful to link these component traits to specific QTL alleles that can be used for a more directed selection. Multi-QTLs were detected for yield and component traits, and positive allele effects derived from both elite and wild parents (Table 1; Figure 8). This supports the idea that transgressive segregation can be achieved by combining yield-increasing alleles from all parents.

In the harvest dissection, the number of QTLs for component traits was higher than the number of QTLs for yield itself in the Fer greenhouse, whereas the number of QTLs was

similar between yield and component traits in the JM greenhouse (Table 1). Because a smaller number of RILs were considered in the biomass dissection, few QTLs were detected for the component traits, despite using a low threshold of $-log_{10}P$. Nevertheless, high prediction accuracies of fitted models (multi-QTL models were developed by using all available observations of RILs) indicated that this threshold choice was reasonable, not leading to too many false positives. Prediction accuracies of fitted models for component traits were higher than that for yield, except for fruit fresh weight per truss and total biomass in JM greenhouse.

Spanish and Dutch breeding populations originated from different parents. All parents had some positive as well as negative QTL allele effects for all traits in both Spanish and Dutch (NL) trials (different populations in different environments), except for wild parent 1 for total fruit dry weight in JM (only negative effect) (Supplementary Figure 5). Therefore, wild parents bring genetic gain to improve the elite parent. High yielding can be achieved by combining yield-increasing alleles in all traits from all parents. In contrast to NL, where elite parents had more positive and wild parents had more negative QTL allele effects on yield, the difference between the elite parent and wild parents was small in Spain (Supplementary Figure 5A, G). This corresponded to the finding that half of the RILs showed transgression in yield in Spain, whereas little transgression was observed in NL (Figure 4A, B). Although parental chromosomal effect was generally not very consistent between both greenhouses in Spain, there were some consistencies. For example, chromosome 7 for yield, 1 for the number of fruits, 1 and 3 for individual fruit fresh weight, 8 for total fruit dry weight. Especially for fruit fresh-dry weight ratio, several chromosomes had the same effect between both greenhouses, i.e. chromosome 1, 3, 5, and 8 (Supplementary Figure 5H1, H2). This might reflect the high phenotypic correlation between both greenhouses for this trait (0.72) (Figure 4h).

3.3.2.2. Single-QTLs at the same position between two greenhouses in Spain, and between Spanish and Dutch trials

- *Spanish greenhouses*

In the Spanish trial, single-QTL comparison between the Fer and JM greenhouses enabled us to evaluate whether QTL detection was more stable for component traits than for yield itself over different environments (Figure 9; Supplementary Figure 6). More single-QTLs were at the same position in individual fruit fresh weight, number of fruits per truss, fruit fresh weight per truss, fruit fresh-dry weight ratio, and harvest index compared to yield.

– *Comparison with Dutch trial (different population in different environment)*

Single-QTL comparison between the Spanish and Dutch trials enabled us to evaluate whether the same chromosome positions played important roles between different populations and environments. This comparison was done for yield and for component traits to assess whether traits differed in the colocalization frequency between Spain and NL (Figure 9; Supplementary Figure 6). Around 29% of the QTLs were at the same position for the Spanish and Dutch trials, which did not show the large differences between yield and component traits.

3.3.2.3. Multi-QTLs colocalization between yield and component traits

We evaluated which component-trait QTLs underlie a yield QTL (including total yield and marketable yield) within each experiment. We decided on colocalization between yield and component-trait QTLs when yield and component-trait QTLs were at the same position (within 20 cM), and haplotype, which showed the highest positive allele effect, was the same between yield and component traits. Some yield QTLs colocalized with both QTLs for component traits in the harvest and biomass dissections (black boxes in Figure 8; Figure 10). More yield QTLs colocalized with QTLs for component traits in the harvest dissection (orange boxes in Figure 8) than QTLs for component traits in the biomass dissection (light-blue boxes in Figure 8) in both greenhouses in Spain (Figure 8A, B; Figure 10). The results in Spain were opposite to those of NL; more yield QTLs were supported by QTLs for component traits in the biomass dissection than by QTLs for component traits in the harvest dissection in NL (Figure 8D; Figure 10).

– *Spanish greenhouses*

In Spain, two yield QTLs were underlain by the same component-trait QTLs between both greenhouses (Fy7.1 and Jy7.2 by fruit fresh weight per truss; Fy12.1 and Jy12.1 by individual fruit fresh weight and fruit fresh weight per truss) (Figure 8). Among them, only for Fy7.1 and Jy7.2, haplotype was the same between Fer and JM greenhouses. Interestingly, for another yield QTL in chromosome 7 (Fy7.2 and Jy7.3), underlying two component traits showed the same haplotype between Fer and JM (wild parent 1 for number of trusses and elite parent for number of fruits per truss). However, the resultant haplotype for yield was different between both greenhouses; the same haplotype as for the number of fruits per truss for Fy7.2, whereas the same haplotype as for the number of trusses for Jy7.3.

– *Comparison with Dutch trial (different population in different environment)*

Although Spanish and Dutch populations were developed separately from different parents, there were still some similarities. Three yield QTLs were underlain by the same component-trait QTLs between JM greenhouse and Dutch trial (Jy3.1 and Ny3.1 by individual fruit fresh

weight; Jy4.1 and Ny4.1 by number of fruits; Jy7.2 and Ny7.2 by total biomass) (Figure 8). Additionally, Jy5.1 (25.2 cM) and Ny5.1 (48 cM) were both underlain by number of fruits, although these two yield QTLs were just outside of 20 cM window. Interestingly, this yield QTL was also underlain by total fruit dry weight and harvest index in JM, but by total biomass in NL. Regarding Jy3.1 and Ny3.1, in addition to individual fruit fresh weight, fruit fresh-dry weight ratio also might underlie the yield QTL (QTL for fruit fresh-dry weight ratio in JM (54.5 cM) was just outside of 20 cM window of Jy3.1 (32.2 cM)).

3.3.2.4. Some co-localized QTLs aligned with the phenotypic correlations between traits

Some multi-QTLs were at the same position (within 20 cM) between traits in the yield dissection models, except for between total yield (Ytotal) and harvest index (HI) in Fer greenhouse (denominator in Figure 6C, D; Supplementary Figure 7). In some of multi-QTLs at the same position, the same haplotype showed the highest positive allele effect between two traits (numerator in Figure 6C, D). Some co-localization of multi-QTLs with same haplotype corresponded to Pearson correlations between two traits (Figure 6A, B).

- *Spanish greenhouses*

Negative correlations between traits (Figure 6A, B) corresponded to that only 0 or 1 colocalized multi-QTL with the same haplotype between FN and FFW, between TN and FNtruss, and between TN and FFWtruss in the harvest dissection (Figure 6C), and between Fresh_Dry_ratio and FDWtotal in the biomass dissection (Figure 6D).

- *Comparison with Dutch trial (different population in different environment)*

The results were similar between the Spanish and Dutch trials in the harvest dissection (Figure 6C), whereas different in the biomass dissection (Figure 6D). For example for Ytotal and Fresh_Dry_ratio more co-localized multi-QTLs showed the same haplotype in NL than in Spain (Figure 6D). This corresponded to the correlation between these two traits, which was high in NL but low or zero in Spain (Figure 6B). Another difference was between FDWtotal and HI, where zero colocalized multi-QTL showed the same haplotype in NL however 3 out of 4 in Spain (Figure 6D). This corresponded to the correlation between these two traits which was low in NL but high in Spain (Figure 6B).

3.3.3. Prediction

We investigated whether predictions of component traits show higher accuracies than prediction of yield itself, and whether yield prediction by yield dissection models using multi-QTLs for component traits can show higher prediction accuracies than yield prediction by multi-QTLs for yield itself. For this purpose, first, we conducted cross-validations within

each greenhouse in the Spanish trial and second we conducted predictions in a greenhouse by using the multi-QTL model developed in the other greenhouse (swap greenhouses).

3.3.3.1. Cross-validation

All component traits in Fer and many component traits (except for fruit fresh weight per truss, total fruit dry weight, and total biomass) in JM showed higher prediction accuracies than yield (Table 2; Supplementary Figure 8). This supports the idea that traits that are lower in the trait hierarchy (i.e. depend on less underlying traits), might be more predictable than those upper in the trait hierarchy.

In the harvest dissection, despite of the higher prediction accuracies of component traits than yield, yield predictions via yield dissection models by using QTLs for component traits, showed similar accuracies as yield prediction by QTLs for yield itself. In the biomass dissection in JM, yield prediction via yield dissection models showed lower accuracies than yield prediction by QTLs for yield itself. In the biomass dissection in Fer, prediction accuracies were too low to be compared between yield prediction by QTLs for yield itself and yield prediction via yield dissection models (Table 2; Supplementary Figure 8).

3.3.3.2. Prediction in a greenhouse by QTL effects from the other greenhouse

To investigate to which extent QTL effects were consistent between both greenhouses, we predicted traits in one greenhouse by using the multi-QTL model developed in the other greenhouse (Table 2; Supplementary Table 4). Prediction accuracies were reasonable for most traits. Many component traits (except for the number of trusses and total biomass in both greenhouses, total fruit dry weight in Fer) showed higher prediction accuracies than yield.

Although many component traits showed higher prediction accuracies than yield, in the harvest dissection in both greenhouses and the biomass dissection 1st layer in JM, yield predictions via yield dissection models showed similar accuracies as yield prediction by QTLs for yield itself. In the biomass dissection in Fer and the biomass dissection 2nd layer in JM, yield predictions via yield dissection models showed lower accuracies than yield prediction by QTLs for yield itself (Table 2; Supplementary Table 5).

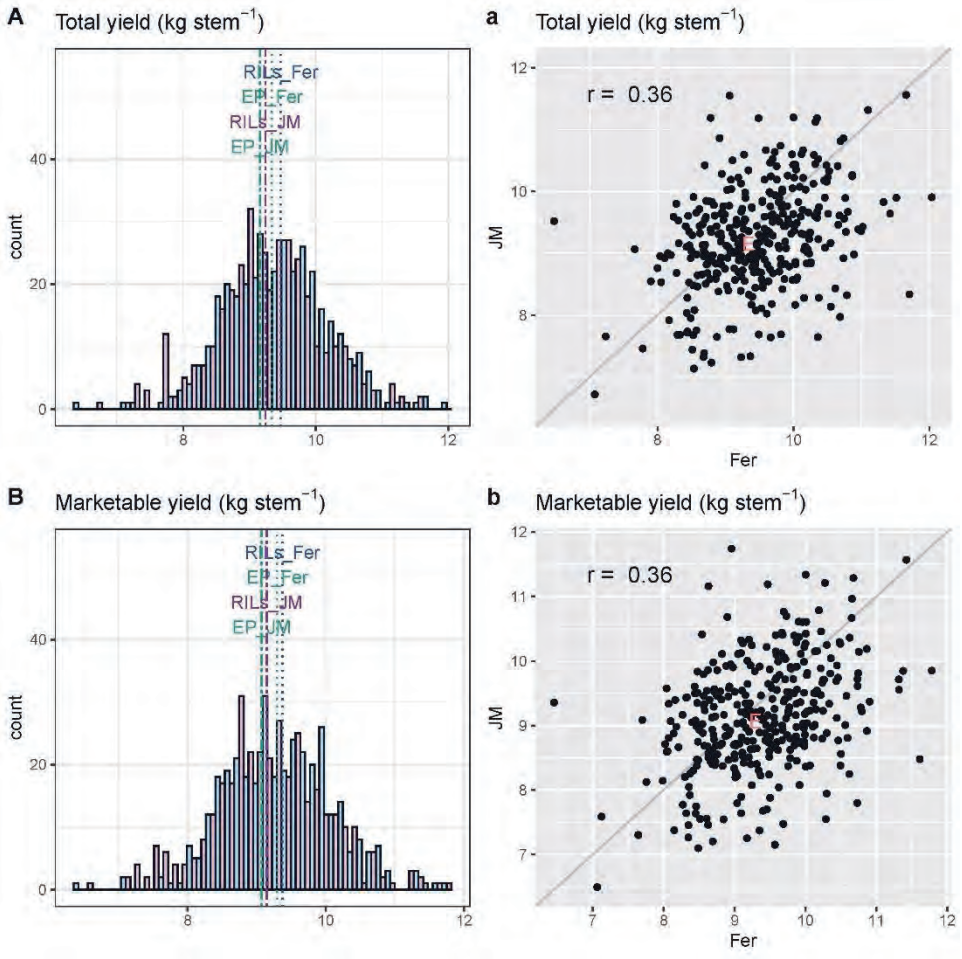


Figure 4 Distributions of RIL means for yield and component traits (A-K) and correlation between Fer and JM greenhouses for each trait (a-k). Vertical lines on histogram show the mean for the RIL population and the performance of the elite parent (EP) in each greenhouse (A-K). E denotes the elite parent (a-k). In the biomass dissection, 70 RILs were phenotyped in each greenhouse, of which 30 RILs were overlapping between both greenhouses (h, i, j, k).

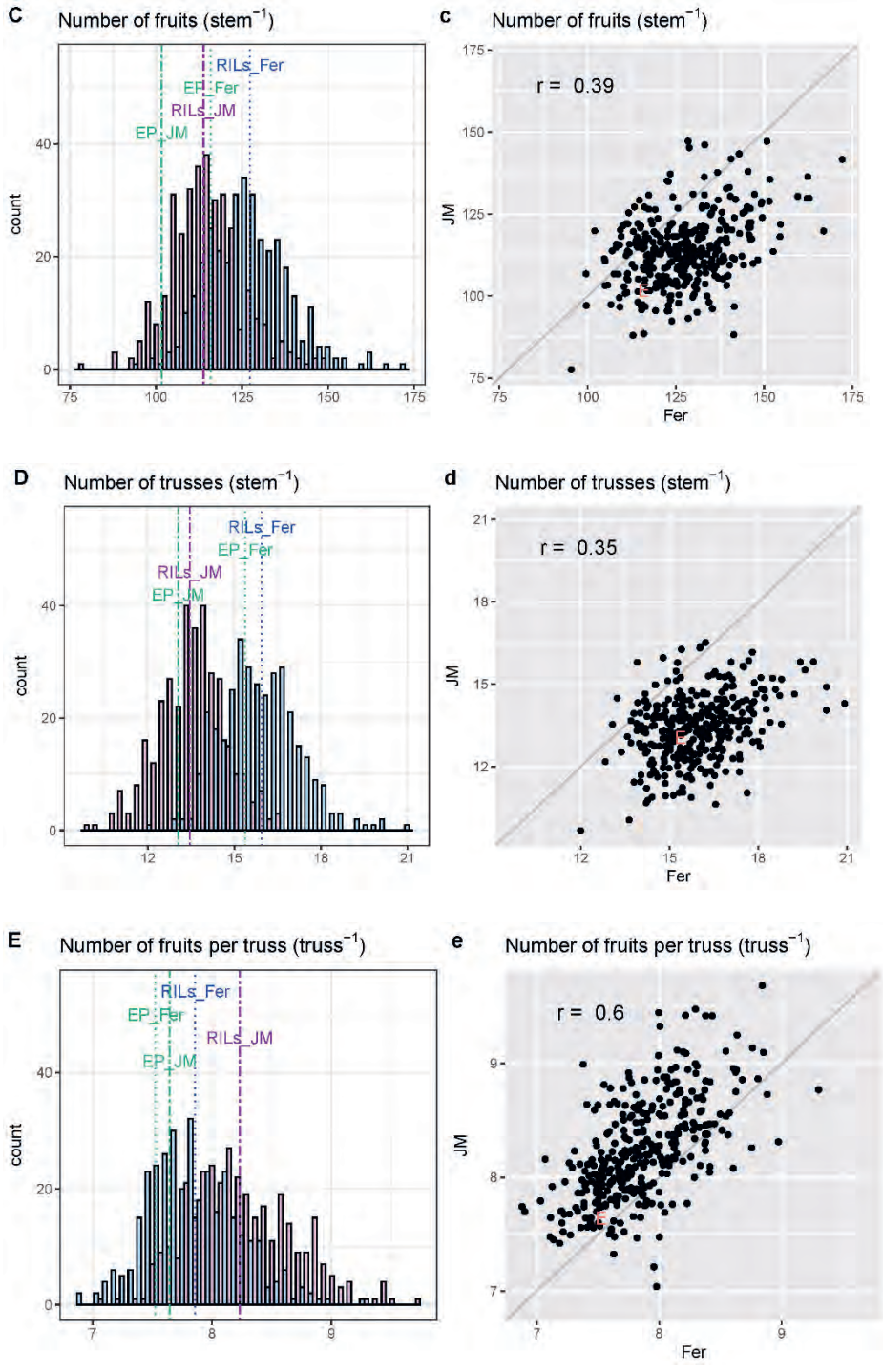


Figure 4 Continued.

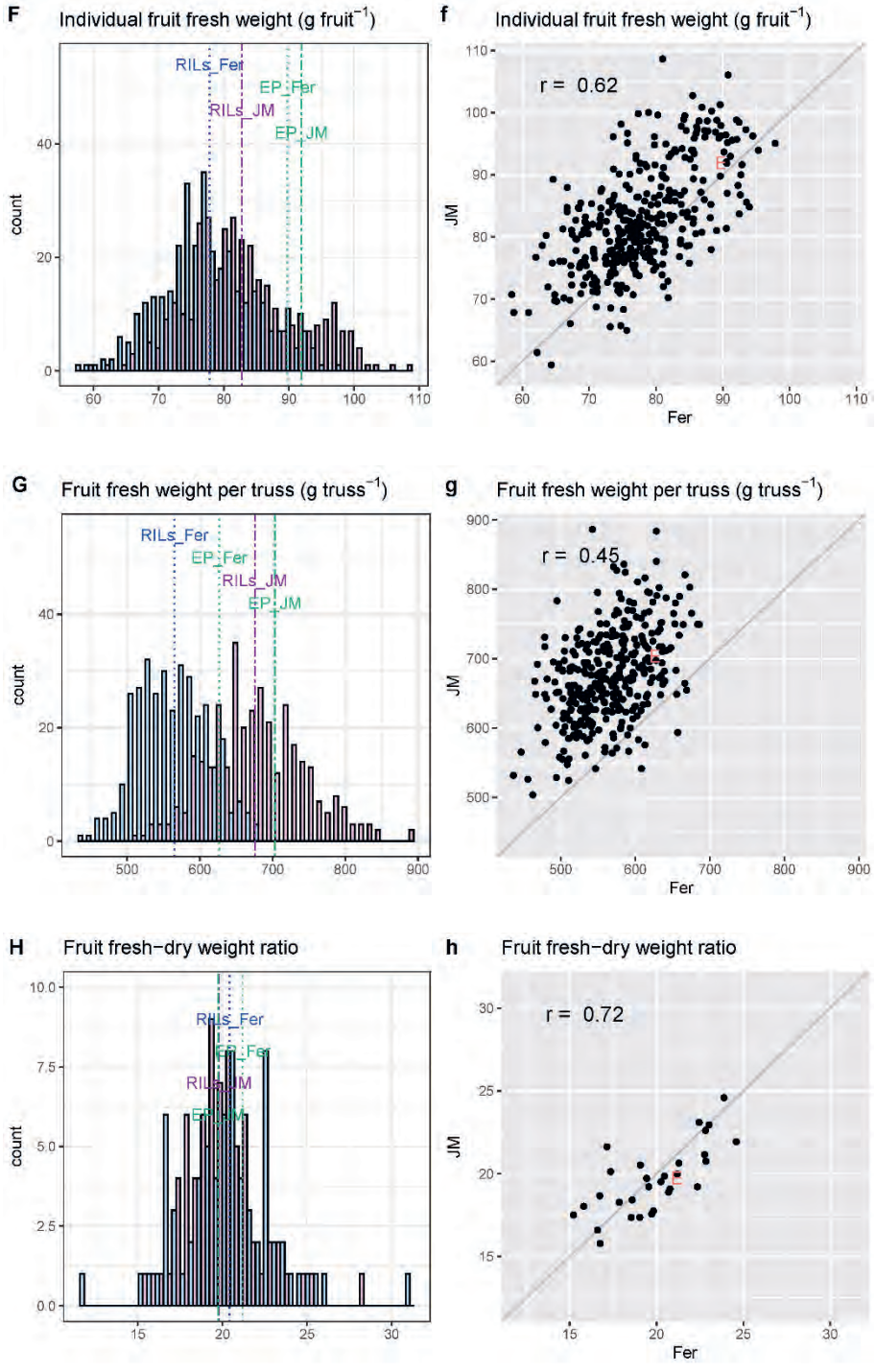


Figure 4 Continued.

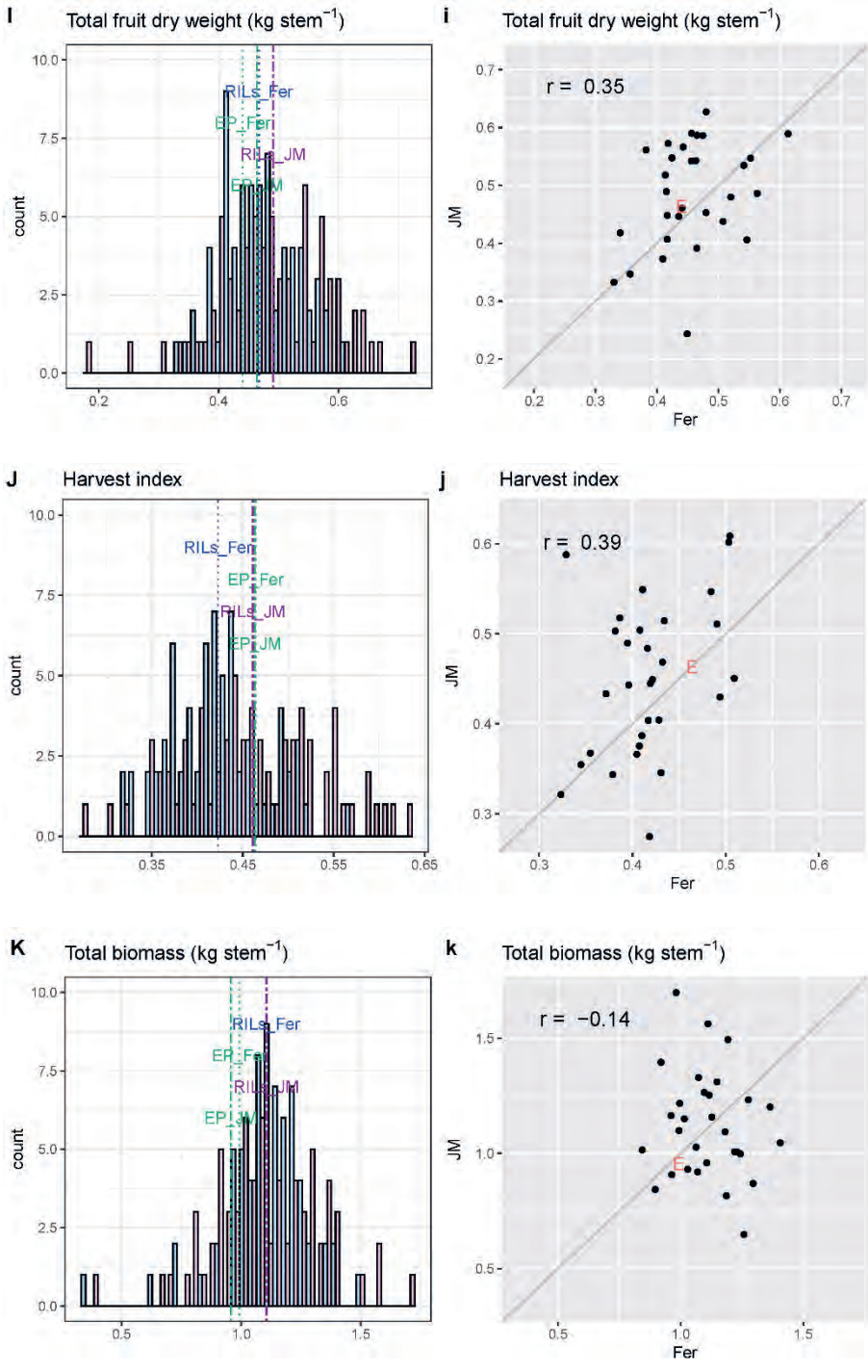


Figure 4 Continued.

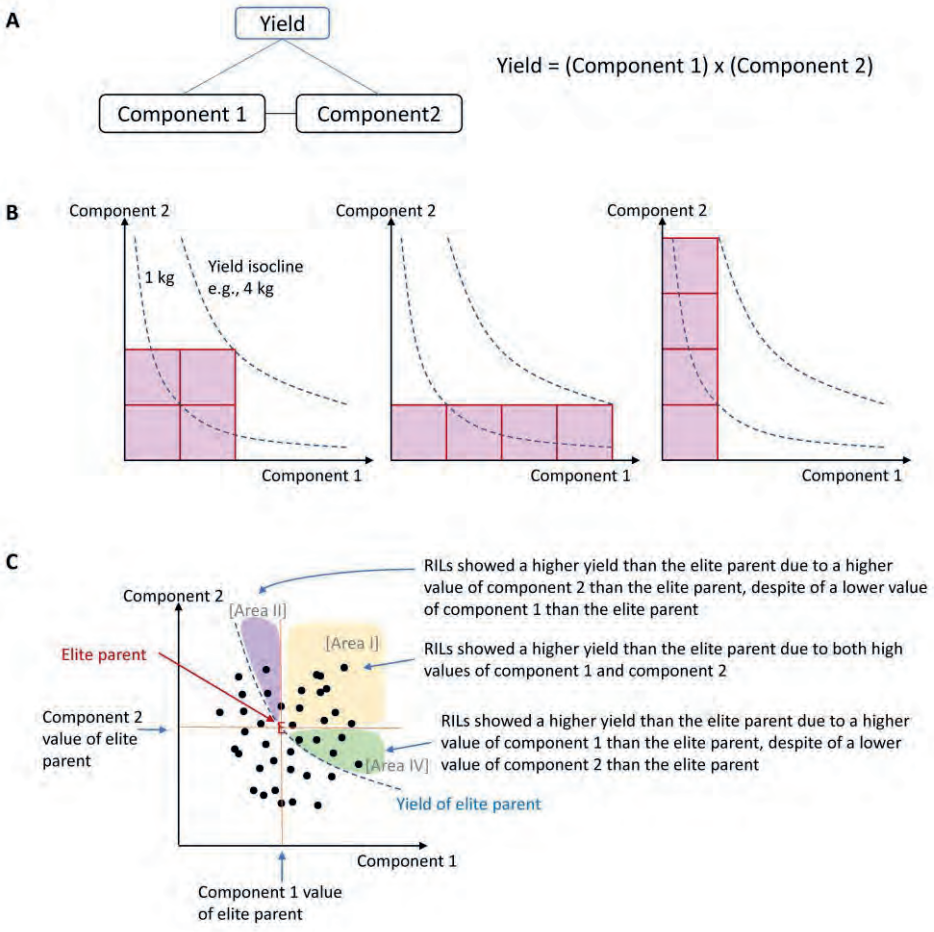


Figure 5 Schematic presentation of equifinality presented at each layer of the yield dissection model. (A) Yield as the product of two underlying component traits (components). (B) Equifinality. The same yield (4 kg in this example) can be achieved via different compositions of component traits. (C) Response surface plot showing the positions of RILs and the elite parent in the context of the target trait (presented by isoclines) and two underlying component traits (x-axis and y-axis).

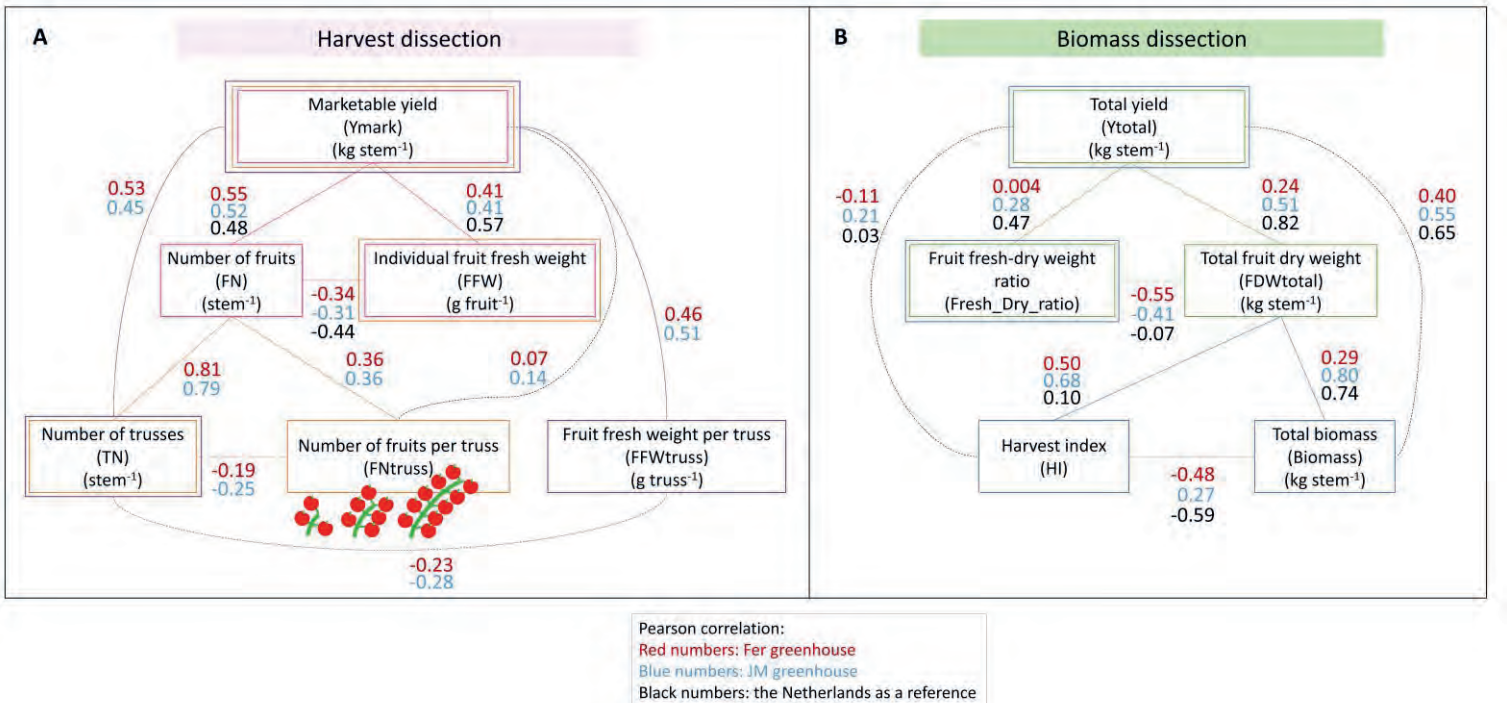


Figure 6 Relationships between traits. (A, B) Pearson correlation. (C, D) Multi-QTL colocalization. Colours of numbers show greenhouse: red for Fer, blue for JM greenhouses in Spain; black for Dutch trial (population for Dutch production in Dutch environment) as a reference. In the multi-QTL colocalization (C, D), numbers in squares show the number of multi-QTLs for each trait. The denominator shows the number of multi-QTLs at the same position between traits, the numerator shows the number of colocalized multi-QTLs with the same allele showing the highest positive effect between traits.

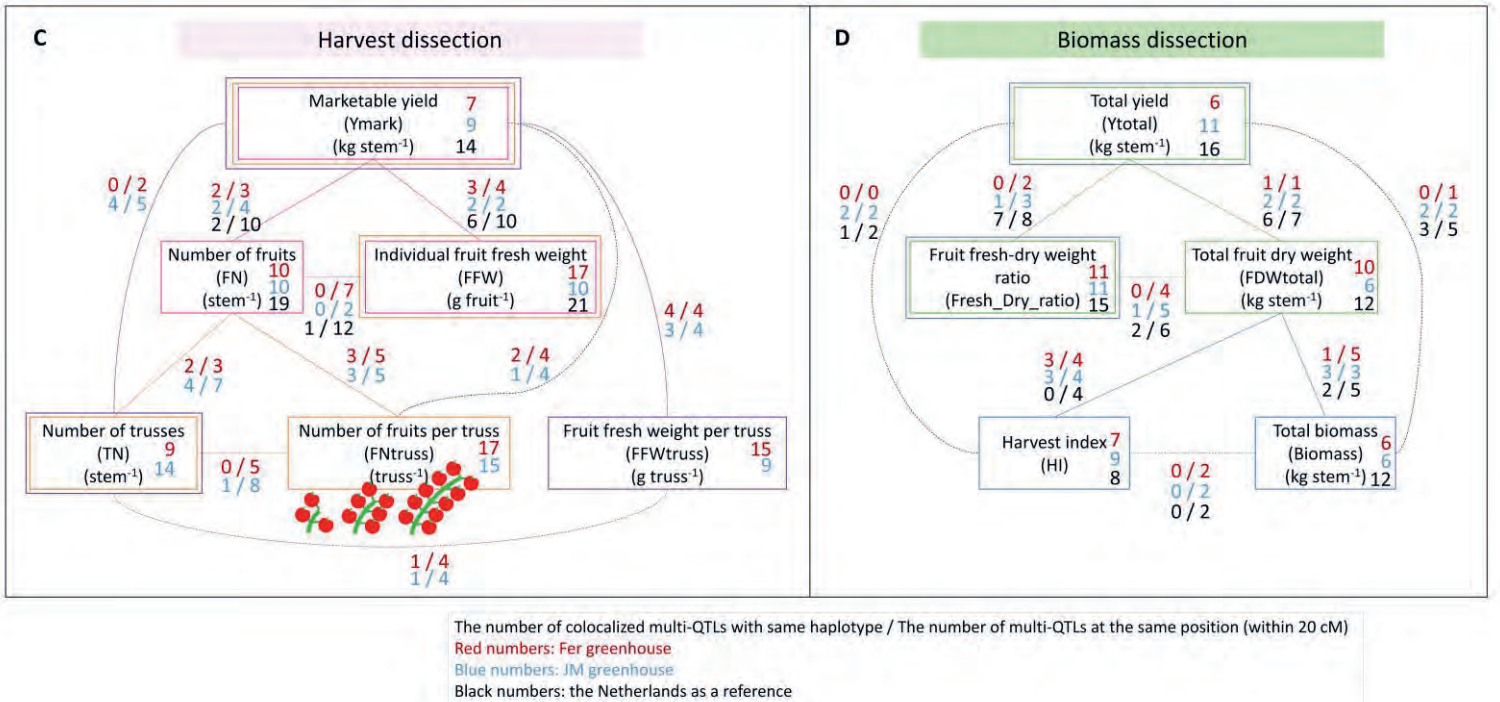


Figure 6 Continued.

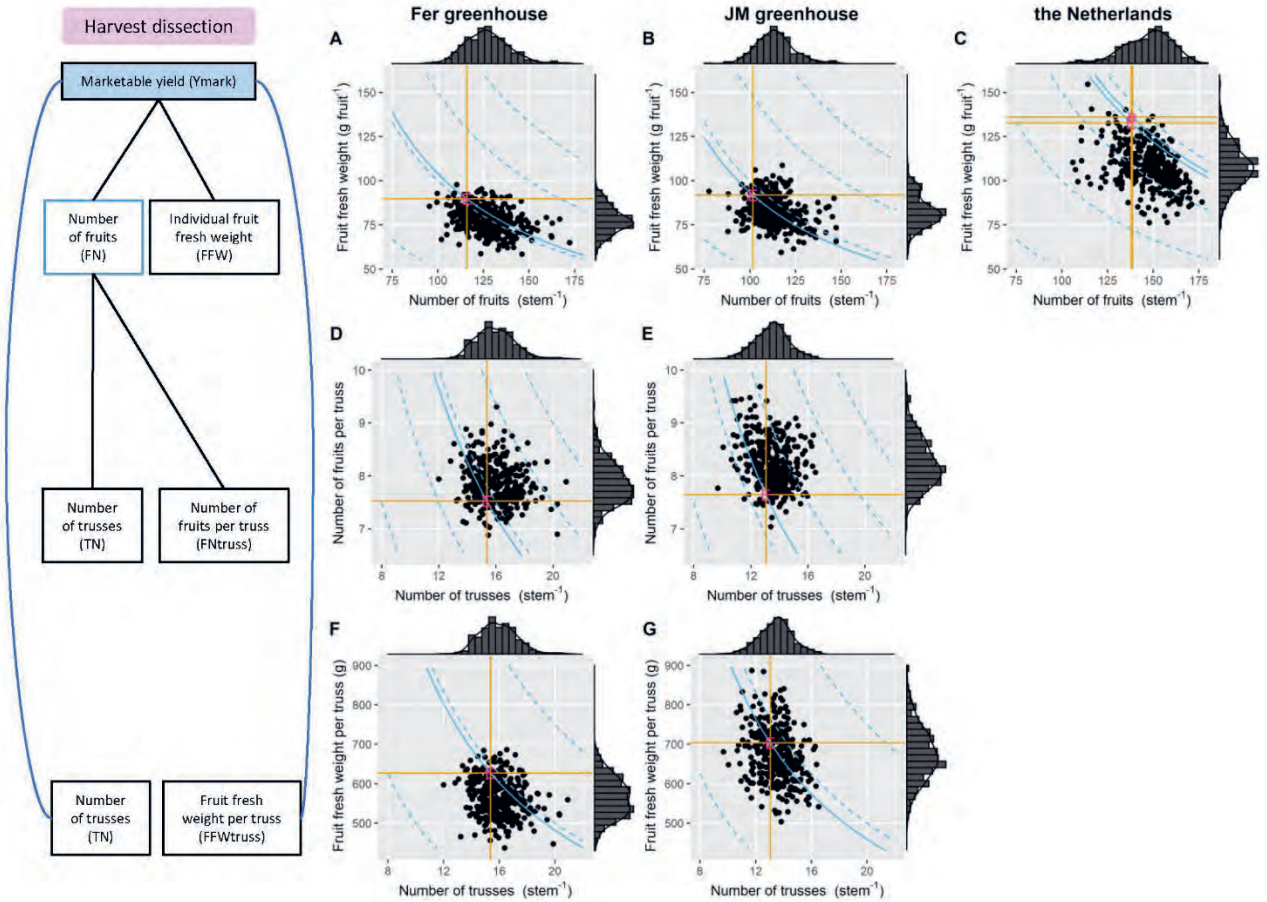


Figure 7 Caption next page.

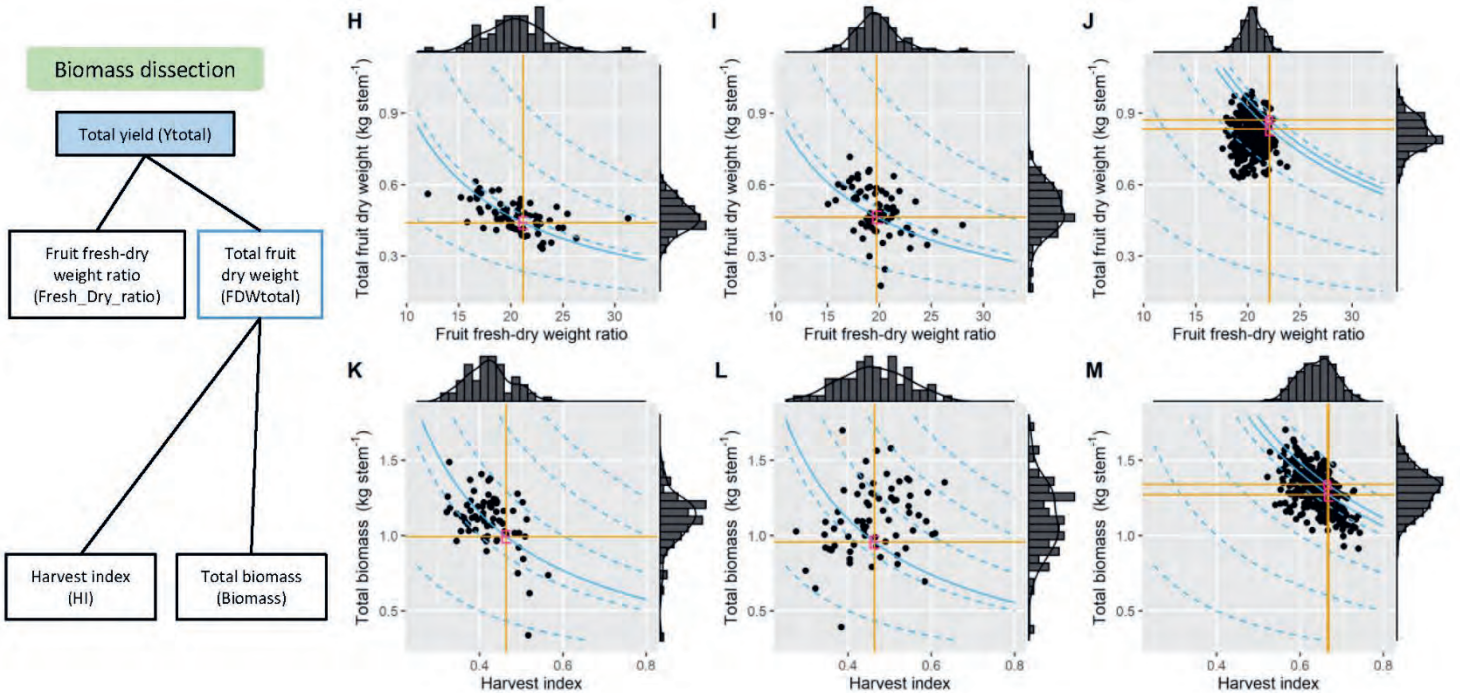


Figure 7 Response surfaces. (A-G) Traits in harvest dissection. (H-M) Traits in biomass dissection. Three greenhouses are compared: Fer greenhouse (A, D, F, H, K), JM greenhouse (B, E, G, I, L) in Spain, and Dutch trial (population for Dutch production in Dutch environment) as a reference (C, J, M). Horizontal and vertical red lines define quadrants determined by the position of the elite line (E) in the response surfaces. Curved blue lines define isoclines of the upper trait, with solid lines referring to the elite parent. For more explanation how to read the plot, please see Figure 5 (C).

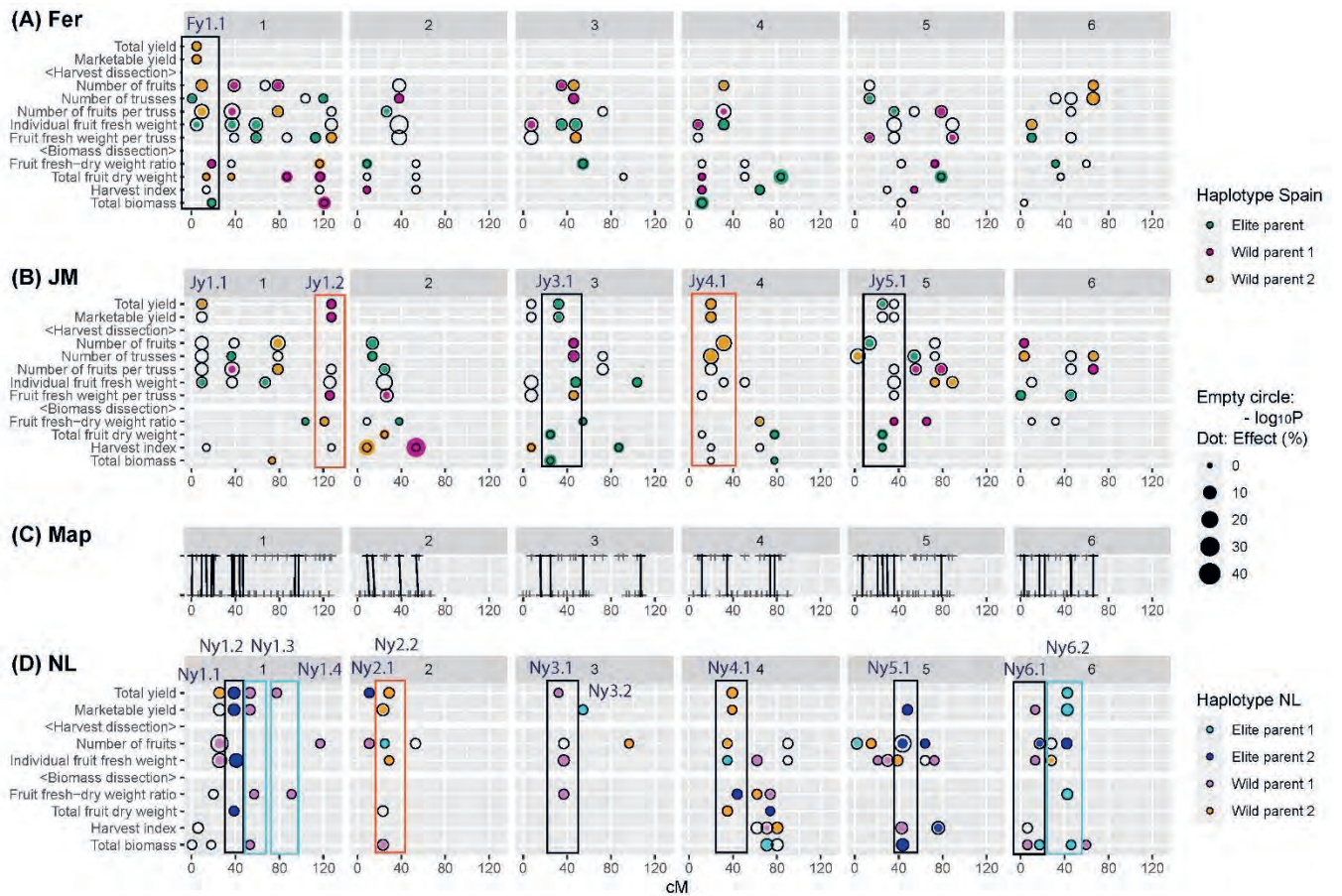


Figure 8 QTL locations and effects. Caption continued to the below Figure 10.

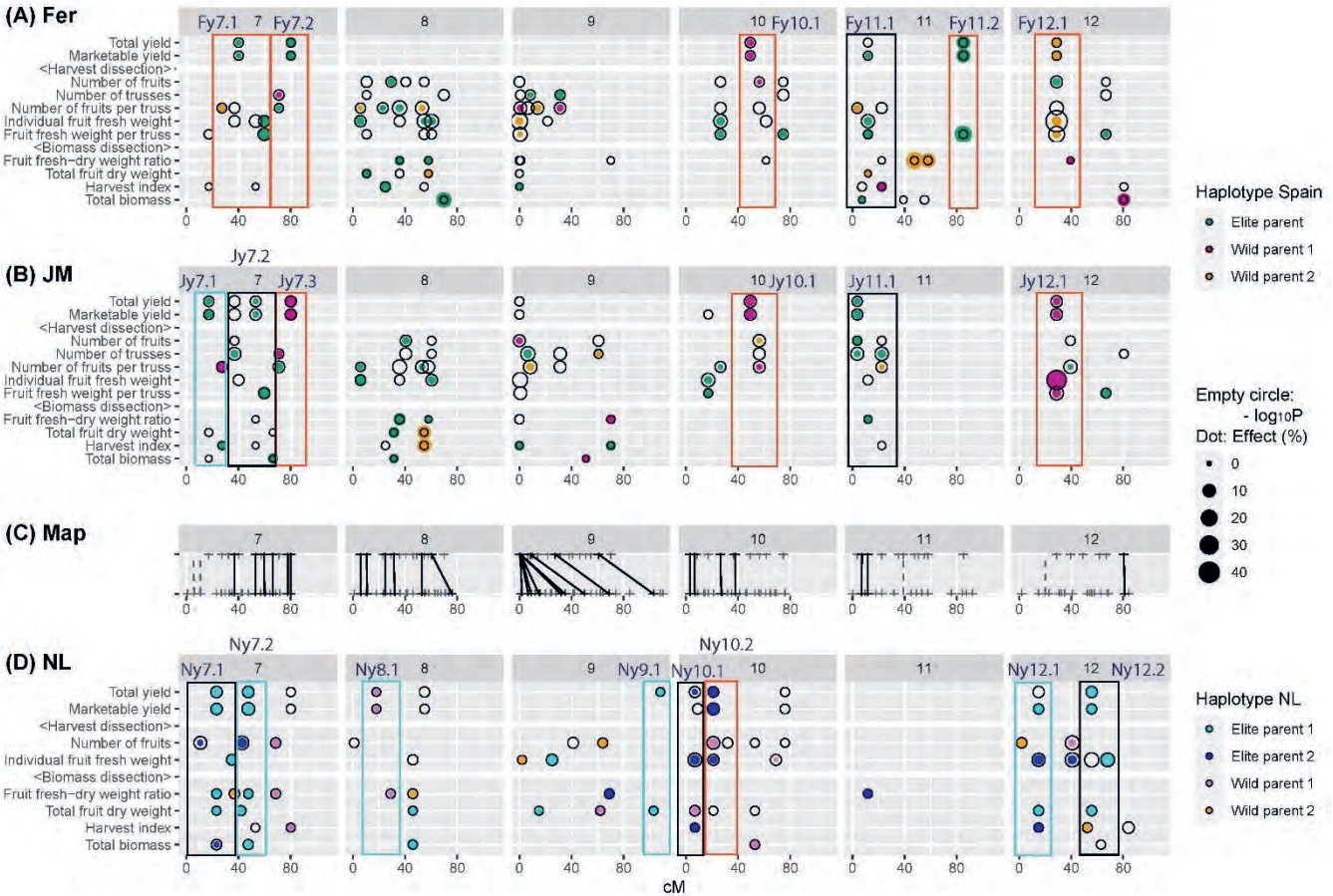


Figure 8 QTL locations and effects. Caption continued below Figure 10.

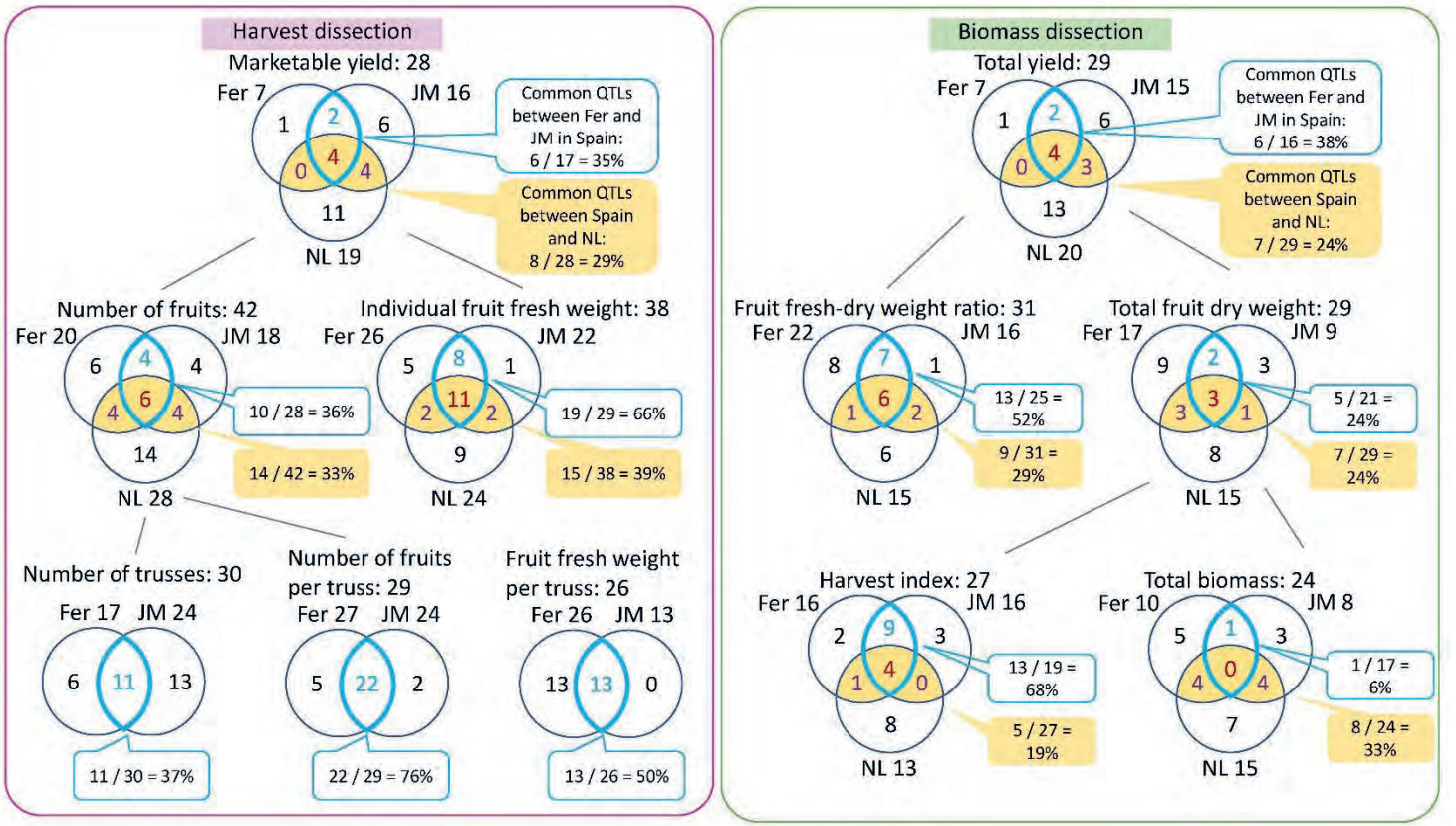


Figure 9 Single-QTL colocalization for each trait between Fer and JM greenhouses in Spain (coloured by blue), and between Spanish and Dutch trials (different populations and environments) (filled by yellow). The numbers next to trait names indicate the total number of single QTLs for each trait. The numbers next to Fer, JM, and NL indicate the total number of single QTLs for each experiment respectively. Please see Supplementary Figure 6 for QTL plots.

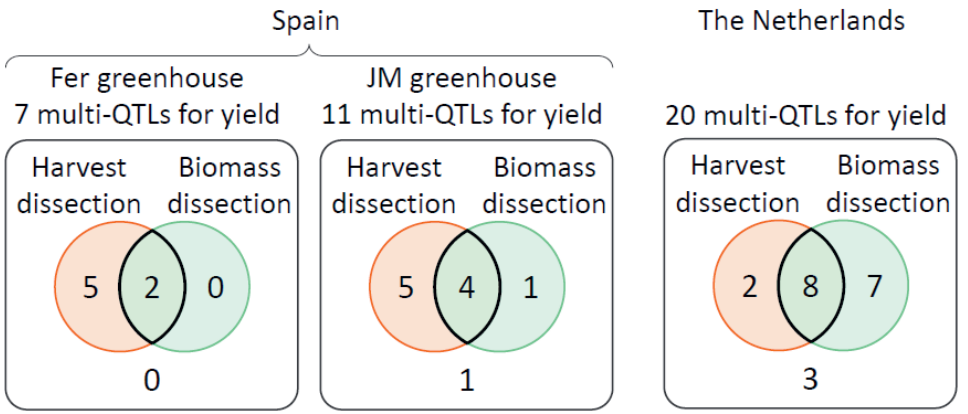


Figure 10 Venn diagram showing the number of colocalized multi-QTLs with the same haplotype showing the highest positive allele effect between yield and component traits in the harvest dissection and/or biomass dissection, in two Spanish greenhouses and in Dutch trial. Yield QTLs include both QTLs for total yield and marketable yield.

Figure 8 Caption continued. QTL locations and effects for yield and component traits. (A) Fer greenhouse. (B) JM greenhouse. (C) Genetic map comparison between Spanish and Dutch populations. (D) Dutch trials (population for Dutch production in Dutch environment). On horizontal axis, chromosome 1 to 12, position within chromosome. Single-QTL positions are indicated by empty circles and multi-QTL positions are indicated by dots. Size of empty circle shows -log$_{10}$P value at single-QTL analysis. Colour of dot shows haplotype with highest positive effect, and the size of dot represents proportion (%) of the largest haplotype effect to mean for the trait in the multi-QTL model. Boxes indicate colocalized yield and component-trait QTLs with the same haplotype showing the highest positive allele effect between yield and component traits: black boxes for component traits belonging to both harvest and biomass dissections, orange for component traits belonging to harvest dissection, and light blue for component traits belonging to biomass dissection. Ticks upper and bottom in (C) represent markers in Spanish and Dutch populations respectively. Common markers between two populations were connected by solid lines in (C). Dotted lines in (C) indicate that these common markers were not used for Spanish analyses (due to distorted markers) but were illustrated here for comparison of genetic maps between Spanish and Dutch populations.

Table 2 Prediction accuracies in fitted model, cross-validation, and swap greenhouses. Heritability, correlation and QTLs colocalization were re-presented as reference for considering G×E.

	(Heritability)		(Correlation)	(QTLs colocalization)		Prediction accuracy					
						Fitted model		Cross-validation*		Swap greenhouses**	
	Fer	JM	between Fer and JM	between Fer and JM	between Spain and NL	Fer	JM	Fer	JM	Fer (pred) - JM (obs)	JM (pred) - Fer (obs)
Marketable yield	0.12	0.30	0.36	35%	29%	0.40	0.59	0.31	0.52	0.46	0.36
Number of fruits	**0.20**	**0.42**	**0.39**	**36%**	**33%**	**0.59**	**0.68**	**0.53**	**0.63**	**0.47**	**0.40**
Number of trusses	0.11	**0.53**	0.35	**37%**		**0.50**	**0.70**	**0.42**	**0.62**	0.41	0.33
Number of fruits per truss	**0.62**	**0.52**	**0.60**	**76%**		**0.76**	**0.76**	**0.70**	**0.70**	**0.68**	**0.67**
Individual fruit fresh weight	**0.65**	**0.60**	**0.62**	**66%**	**39%**	**0.81**	**0.69**	**0.76**	**0.64**	**0.66**	**0.74**
Fruit fresh weight per truss	**0.38**	**0.39**	**0.45**	**50%**		**0.66**	0.54	**0.59**	0.46	**0.51**	**0.61**
Harvest dissection 1						**0.43**	0.55	**0.32**	0.50	**0.47**	0.36
Harvest dissection 2						0.40	0.52	0.31	0.46	0.42	0.34
Harvest dissection 3						0.37	0.50	0.30	0.43	0.39	0.32

Total yield	0.12	0.26	0.36	38%	24%	0.40	0.58	0.07	0.47	0.46	0.36
Fruit fresh-dry weight ratio	**0.62**	**0.87**	**0.72**	**52%**	**29%**	**0.88**	**0.75**	**0.67**	**0.47**	**0.73**	**0.78**
Total fruit dry weight	**0.55**	**0.68**	0.35	24%	24%	**0.75**	**0.62**	**0.48**	0.33	0.35	**0.43**
Harvest index	**0.27**	**0.83**	**0.39**	**68%**	19%	**0.72**	**0.77**	**0.57**	**0.50**	**0.49**	**0.60**
Total biomass	0.02	0.26	-0.14	6%	**33%**	**0.71**	0.47	**0.35**	0.21	0.18	0.20
Biomass dissection 1st layer						0.04	0.55	-0.05	0.34	-0.06	**0.38**
Biomass dissection 2nd layer						0.19	0.48	**0.10**	0.31	0.19	0.22

**Cross-validation: Median of prediction accuracies (correlation coefficient between observation and prediction) was shown. Five-fold cross-validation by 346RILs for marketable yield and components in Harvest dissection. 50:20 cross-validation by 70RILs for total yield and components in Biomass dissection.*

***Swap greenhouses: For total and marketable yield, and for component traits in the harvest dissection, multi-QTL model by 346RILs in a greenhouse was used for predictions of 346RILs in the other greenhouse. For component traits in the biomass dissection, multi-QTL model by 70RILs in a greenhouse was used for predictions of 30RILs in the other greenhouse, which were overlapped between both greenhouses. For yield prediction by the biomass dissection, multi-QTL model by 70RILs in a greenhouse was used for predictions of 70RILs in the other greenhouse.*

3.4. Discussion

3.4.1. High transgression in Spanish trial (population for Spanish production in Spanish environment)

Half of the RILs showed a higher yield than the elite parent (transgression; Figure 4) in the Spanish trial (population for Spanish production in Spanish environment), which was a big difference from the Dutch trial (population for Dutch production in Dutch environment) where only a small portion of RILs showed a higher yield than the elite parent (Tsutsumi-Morita *et al.*, 2021). The high transgression indicated that there is plenty of room for yield improvement in Spanish breeding population. This might be partly because the elite parent in the Dutch trial already achieved a high yield (0.56 and 0.58 kg stem^{-1} week^{-1} of total yield for elite parent 1 and elite parent 2 respectively) whereas the elite parent in the Spanish trial did not (0.41 and 0.44 kg stem^{-1} week^{-1} of total yield in the Fer and JM greenhouses respectively). However, high-yielding breeding seems still challenging in Spain. The correlation for yield between the Fer and JM greenhouses was not high (0.36 (Figure 4a, b)), which corresponded to the low percentage of common single-QTLs for yield between both greenhouses (35% for marketable yield, 38% for total yield (Figure 9)).

3.4.2. Yield dissection models revealed characteristics of high yielding from phenotype and QTL

In this study, we tried to decompose yield complexity by using yield dissection models and explored possible pathways to improve yield. The population of our study was developed by introgressing two wild germplasms into an elite parent background. Two important aspects need to be considered. The first one is the position of the elite parent relative to the RILs distribution for each trait (transgression), that is, how was the phenotype of the elite parent changed by introgressed wild germplasm. The second is how was the change in a component connected to changes in yield and other components. In our yield dissection models, increment in a component leads to increment in the upper trait (e.g., yield). However, if there is a trade-off (negative correlation) between both component traits, an increase in a component trait does not necessarily improve the upper trait (e.g., yield). Thus, first we looked at the correlation between both component traits at the population level. Then, we investigated whether the transgression of the upper trait (e.g., yield) could be attributed to either one of the component traits.

3.4.2.1. Harvest dissection — trade-off between the number of fruits and individual fruit fresh weight

In the harvest dissection, in both Spanish and Dutch trials, wild germplasms tended to increase the number of fruits (FN) while decreasing individual fruit fresh weight (FFW) (Figure 6A; Figure 7A, B, C). This negative correlation was explained by QTL allele effects. None of the colocalized QTLs between FN and FFW showed the same haplotype having the largest allele effect in Spain, and only 1 out of 12 colocalized QTLs showed the same haplotype having the largest allele effect in the Netherlands (NL) (Figure 6C). Some haplotype effects per chromosome were opposite between FN and FFW (e.g. effect of elite parent and wild parent 2 on chromosome 1; effect of elite parent on chromosome 3) (Supplementary Figure 5B, C). This indicates that yield cannot be simply increased by increasing either FN or FFW, but breeders need to consider both component traits simultaneously. This negative correlation between FN and FFW was also observed in other tomato studies (Chakrabarti *et al.*, 2013), and a negative correlation between seed number and seed size was observed in other crops (Sadras, 2007; Yin and Struik, 2008). Thus, this relationship always needs to be considered by breeders.

3.4.2.2. Biomass dissection 1st layer — fruit fresh-dry weight ratio did not correlate with yield in Spain

Tomato consumers often think that high-yielding lines are less tasty. This happened in Dutch production cultivars until early 1990's, when the media in Germany complained about the taste of Dutch tomato as "Wasserbomben (water bombs)" (Hendriks, 2016; Schouten *et al.*, 2019). However, the link between high yield and watery taste was not observed in the Spanish trial because fruit fresh-dry weight ratio did not correlate with yield in Spain, regardless of a much wider variation of fruit fresh-dry weigh ratio than for NL (Figure 7H, I, J). The lack of correlation between yield and fruit fresh-dry weight ratio in Spain was partly explained by zero or only one colocalized QTL showing the same haplotype between fruit fresh-dry weight ratio and yield (Figure 6D).

Salt stress is one of the major problems in greenhouse production in Almeria in Spain (Luis Caparrós-Martínez *et al.*, 2020). High salinity often results in low fruit fresh-dry weight ratio (strong taste) (Navarro *et al.*, 2005; Cuartero *et al.*, 2006; Diouf *et al.*, 2018). In line with this, the fruit fresh-dry weight ratio of the elite parent in Spain was lower than that in NL. Surprisingly, some RILs in Spain showed a higher fruit fresh-dry weight ratio beyond RILs in the Dutch trial (Figure 7H, I, J). Nevertheless, fruit fresh-dry weight ratio did not correlate with yield at the population level in Spain, which was different from NL (Figure 6B). However, regarding RILs' selection of high yielding, Spain and NL had the same conclusion: more high-yielding RILs showed higher yield than the elite parent due to higher total fruit dry weight rather than due to higher fruit fresh-dry weight ratio.

3.4.2.3. Biomass dissection 2nd layer — total biomass is more important than harvest index

Harvest index (HI) showed a high positive correlation with total fruit dry weight (FDWtotal) in Spain, which was the major difference compared to NL, where there was no correlation (Figure 6B). This phenotypic relationship was supported by allele effects of colocalizing QTLs (Figure 6D) and similar pattern of haplotype effect in chromosome 8 between FDWtotal and HI (Supplementary Figure 5I, J). One possible reason why HI had a positive influence on FDWtotal in Spain but not in NL may be that the HI of the elite parent in Spain was much lower (0.46 in Fer; 0.46 in JM) than that in NL (0.67 for elite parent 1 and elite parent 2). Thus, the elite parent in Spain had much more room to improve HI. RILs distribution in HI was lower and wider in Spain (0.32-0.56 in Fer; 0.27-0.63 in JM) than that in NL (0.52-0.74) (Figure 7K, L, M).

The positive correlation of HI with FDWtotal in Spain, however, did not lead to a positive correlation between HI and yield (Figure 6B; Supplementary Figure 4). This might be due to the negative correlation between FDWtotal and fruit fresh-dry weight ratio (Figure 6B). Indeed, fruit fresh-dry weight ratio and HI were weakly negatively correlated with each other in JM (-0.23 (not significant) in Fer; -0.36 in JM (Supplementary Figure 4)). For major staple crops, improvements in HI have historically increased yield potential (Hay, 1995; Smith *et al.*, 2018), and it is widely recognized that harvest index is an appropriate target for increasing yield in breeding (Reynolds *et al.*, 2011; Sadras and Richards, 2014). HI was a less important target in our populations of indeterminate tomato. This is consistent with a previous study comparing eight Dutch indeterminate-tomato cultivars, where HI did not correlate with yield (Higashide and Heuvelink, 2009). However, another study comparing six Japanese indeterminate-tomato cultivars showed an opposite result; HI highly correlated with yield (Higashide *et al.*, 2012). Therefore, HI might become a target for yield improvement in some populations in some environment, which was not the case for our trials.

RILs showing higher FDWtotal than the elite parent(s) had this attribute thanks to more total biomass rather than more HI in both Spain and NL (Figure 7K, L, M). In addition, total biomass was also positively correlated with yield in Spain and NL (Figure 6B). Therefore, total biomass seems to be an important factor for high-yielding breeding in both countries.

There is a caution about the reliability of total biomass measurement in the Spanish trial. Compared to HI, total biomass showed low heritability, low correlation and low QTL colocalization between the Fer and JM greenhouses, and low prediction accuracy. These differences might come from environmental variance of the number of trusses, which was used to calculate total leaf dry weight and total stem dry weight. The number of trusses showed a low heritability in Fer, relatively low correlation and low QTL colocalization between the Fer and JM greenhouses, and low prediction accuracy in Fer. The high

environmental variance of number of trusses seemed to have almost no influence on HI because HI is almost the same between whole plant and a part of plant in indeterminate tomato.

3.4.3. Plasticity to the environment and genotype × environment interaction (G×E) in Spanish trial

Component traits in the harvest dissection showed plasticity to the environment, that is, the phenotypic value is higher in one environment than in the other environment (Figure 4c-g). This might be due to the different temperatures between Fer and JM greenhouses. Although we did not have environmental data, high temperature is known to accelerate plant development, which results in a high number of trusses with small fruits (De Koning, 1994; Savvides *et al.*, 2013). Thus, the temperature in Fer greenhouse might have been higher than that in JM, which was located closer to the sea. Although the higher number of trusses in Fer than in JM was mitigated by a lower number of fruits per truss, the resulting number of fruits was still higher in Fer than in JM (Figure 4c-g). Nevertheless, yield did not show much plasticity to the environment (Figure 4a, b). Thus, plasticity to the environment for each component trait in the harvest dissection was negated with each other.

Regardless of the high plasticity to the environment, the correlation between Fer and JM greenhouses was high (less G×E) for number of fruits per truss and individual fruit fresh weight (Figure 4e, f). That is, the ranking of RILs was consistent between both greenhouses for these traits. The high correlation between both greenhouses corresponded to the high heritabilities (Table 1), high colocalizations of QTLs between both greenhouses (Figure 9), and high prediction accuracies for these traits (Table 2).

On the other hand, component traits in the biomass dissection did not show plasticity to the environment (Figure 4h-k). High salinity is known to stimulate soluble solid content (less fruit fresh-dry weight ratio) (Navarro *et al.*, 2005; Cuartero *et al.*, 2006; Diouf *et al.*, 2018). No plasticity for fruit fresh-dry weight ratio might mean that salinity level was similar between Fer and JM greenhouses, although no soil data was available. Correlation between both greenhouses was high for fruit fresh-dry weight ratio (Figure 4h), which corresponded to the high heritability, high colocalization of QTLs between both greenhouses, and high prediction accuracies even when different greenhouses were used for the training set and the validation set (Table 2).

3.4.4. Accuracy of yield prediction

Accuracy of yield prediction by yield dissection models using QTLs for component traits did not exceed the accuracy of yield prediction by QTLs for yield, although more QTLs were used for yield dissection models and several component traits showed higher prediction accuracies than yield (Table 2). Uncertainty in yield prediction by crop models with

quantitative genetics may arise from crop model structure (Yin *et al.*, 2000*a*) and prediction accuracy of the component traits. Because no uncertainty arises from model structure in the case of yield dissection models (Yin *et al.*, 2002), uncertainty might come from the prediction accuracy of component traits. If a component trait shows low prediction accuracy, the low prediction accuracy of the component trait reduces the accuracy of yield prediction by yield dissection models, even if other component traits show high prediction accuracies. Additionally, uncertainty also arises from the multiplication of the uncertainties of component traits. Nevertheless, yield prediction by the harvest dissection 1 showed the same accuracy as yield prediction by QTLs for yield itself. Therefore, it is possible to predict yield by yield dissection models. Here, QTL allele effects for component traits can be converted to effects on yield.

3.4.5. Why do we use yield dissection models?

3.4.5.1. Design of high-yielding lines

Although the direct prediction of yield by yield QTLs showed better or equivalent performance than the use of yield dissection models, yield dissection models have several advantages. Studying yield alone does not reveal the underlying reasons for high yield. Yield dissection models give additional insights and options for high-yielding breeding. The same yield can be achieved by different compositions of component traits (equifinality). Yield dissection models make the equifinality explicitly visible and expose trade-offs between component traits. These relationships were underpinned by QTLs. Therefore, breeders can design high-yielding lines with preferable composition of component traits by understanding the allele effects of component traits on yield.

3.4.5.2. Harvest dissection (sink capacity) vs biomass dissection (source strength)

In our study, we used the harvest dissection to capture the sink capacity, and the biomass dissection to capture the source limitation. On the one hand, improving source activity has been recognized as a major option to increase crop yield potential further, but on the other hand, yield is limited by sink capacity rather than source strength during seed filling in major crops (Borrás *et al.*, 2004; Chang *et al.*, 2017; Smith *et al.*, 2018). In tomato genetic research, it was shown that fruit size increased ~100 times compared to its ancestor during domestication and improvement (Lin *et al.*, 2014). However, it is not clear whether the increment of fruit size (individual fruit fresh weight) was due to the increase in capacity of fruit size and/or more biomass becoming available for fruits. Fruiting tomato plants in Dutch greenhouses when grown according to commercial practice are known to be source-limited, as is shown in fruit pruning experiments and carbohydrate contents measurement (Cockshull and Ho, 1995*a*; Iglesias *et al.*, 2002; Li *et al.*, 2015). Considering the high plasticity of tomato

fruit size (i.e. pruning increases fruit size), it is possible that QTLs involved in fruit size also have a function in biomass production or biomass allocation to fruits.

In our study, a negative correlation was observed between the number of fruits and individual fruit fresh weight (Figure 6A, Figure 7A, B, C). This might be explained by source limitation, although this might also be explained by the conflicting allele effects of some QTLs at the same position between both traits (Figure 6C). Some yield QTLs colocalized with both component-trait QTLs in the harvest and biomass dissections (Figure 8; Figure 10). There were also many other QTLs showing pleiotropic effects between component traits in the harvest and biomass dissections (Figure 8). These pleiotropic QTLs were associated with both sink capacity and source strength.

Considering the source limitation of fruiting tomato plants (Li *et al.*, 2015), the biomass dissection seems to be suitable for high-yielding breeding. However, the harvest dissection is also useful. The harvest dissection model seemed to capture the influence of source strength. In spite of the plasticity to the environment (Figure 4), harvest dissection model 1 showed the same prediction accuracy as yield prediction by QTLs for yield itself (Table 2). In addition, from the perspective of phenotyping, component traits in the harvest dissection are easier to be measured than component traits in the biomass dissection. In major staple crops, yield is usually more sink limited than source limited during seed filling, but this changes to source limitation if resource availability is strongly reduced (Borrás *et al.*, 2004). Thus, sink and source limitations are changeable depending on species, population, management and environment. We recommend to use both harvest and biomass dissection models for high-yielding breeding.

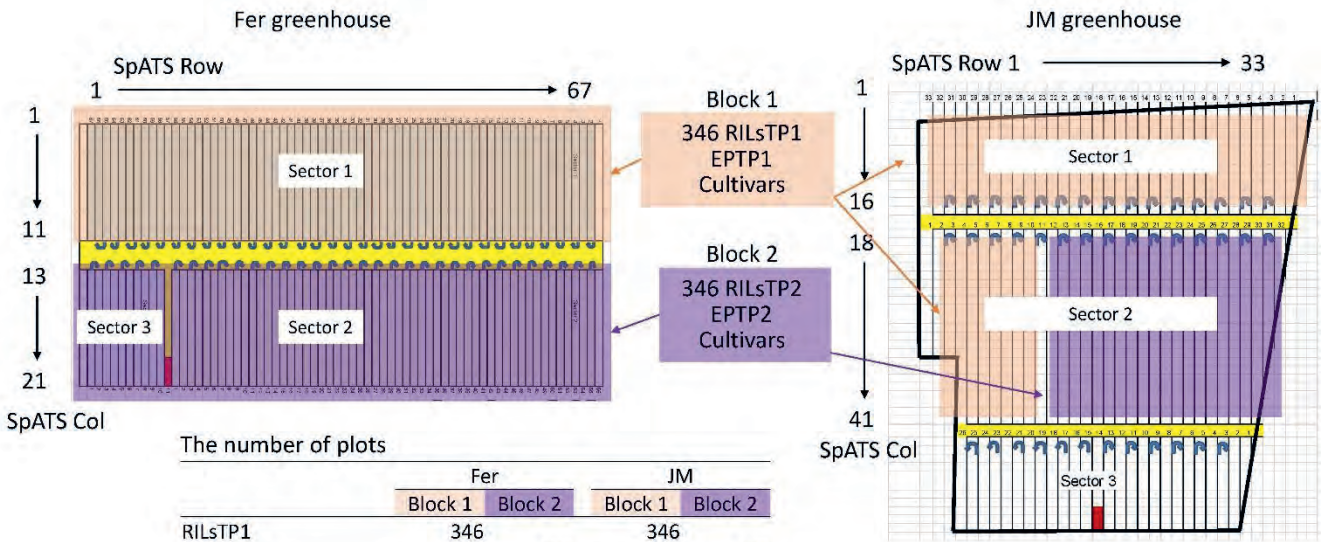
3.5. Conclusion

This paper presents the usefulness of combining yield dissection models with quantitative genetics for making a strategy for high-yielding breeding. We demonstrated this for a Spanish tomato trial (population suitable for Spanish production and tested in Spanish environment) and compared it with our earlier work on a Dutch trial (population suitable for Dutch production and tested in Dutch environment). In Spain, half of RILs showed a higher yield than the elite parent (transgression), whereas in the Netherlands (NL) this was only 3%. The harvest dissection model revealed the trade-off between number of fruits and individual fruit fresh weight in both Spain and NL. Thus, explicit selection based on positive alleles for both traits combined is necessary. The biomass dissection model showed that fruit fresh-dry weight ratio was not correlated with yield in Spain, whereas a positive correlation was shown in NL. More high-yielding lines were attributed to the number of fruits than individual fruit fresh weight, more attributed to total fruit dry weight than fruit fresh-dry weight ratio, and more attributed to total biomass than harvest index, in both Spain and NL. Most relationships between traits were underpinned by QTL colocalizations and QTL allele effects. Most yield

QTLs were underlain by component QTLs and some of them showed pleiotropic effects between components in the harvest and biomass dissections, which illustrated that the role of a QTL might appear as both sink capacity and biomass availability for yield. Although some components (e.g., number of fruits per truss, individual fruit fresh weight, fruit fresh-dry weight ratio) showed less G×E (higher heritability, higher correlation and higher QTLs colocalization between both greenhouses, and higher prediction accuracy) than yield, accuracy of yield prediction via yield dissection models did not exceed the accuracy of yield prediction by QTLs for yield itself. Yet, the similar accuracy of yield prediction via the harvest dissection to the accuracy of yield prediction by yield QTLs indicated that QTL allele effects of component traits can be transformed into the effects on yield. Accuracy of yield prediction was higher by the harvest dissection model than by the biomass dissection model in most cases in Spain, whereas in NL it was the other way around although the difference was small. Because sink limitation and source limitation are changeable depending on species, population, management and environment, we recommend to use both harvest and biomass dissection models together to find the place to improve yield in terms of component traits and QTLs. Yield dissection models together with quantitative genetics revealed how to improve the yield of the elite parent and answered the question on which component traits to focus and on which QTLs. Plant breeders can design high-yielding lines with preferable composition of component traits by combining positive alleles for component traits.

Supporting information

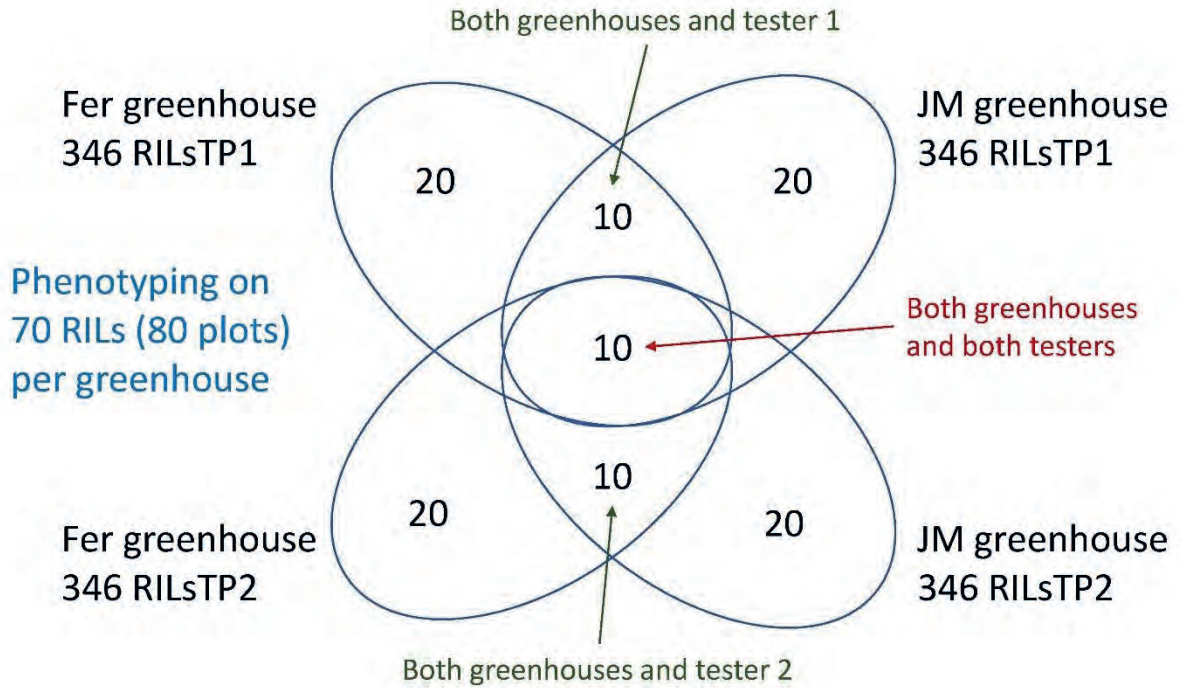
Supplementary Figure 1 Indeterminate tomato grown in a greenhouse in Spain for phenotyping.



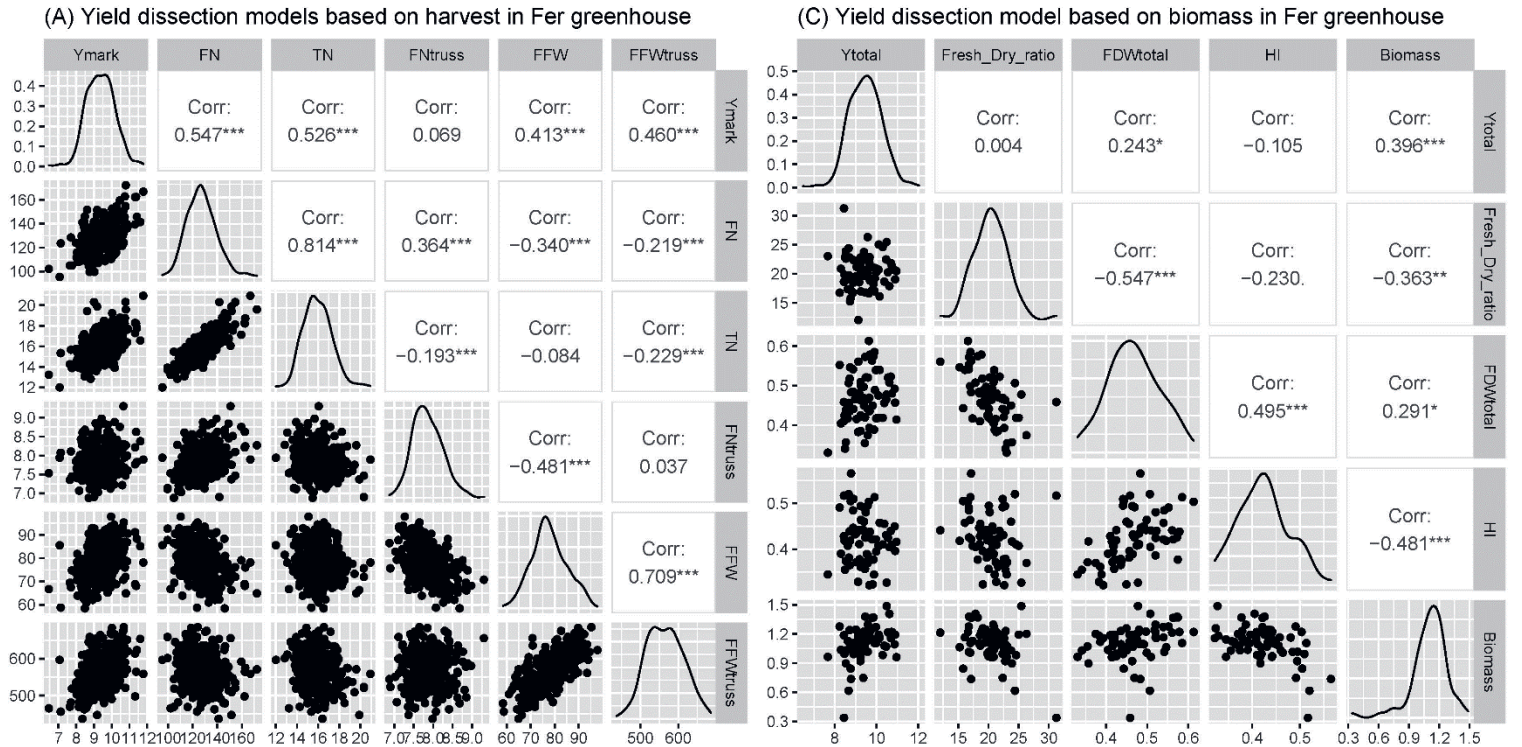
The number of plots

	Fer		JM	
	Block 1	Block 2	Block 1	Block 2
RILsTP1	346		346	
RILsTP2		346		346
EPTP1	10		11	
EPTP2		7		11
Cultivar A	6	3	3	3
Cultivar B	6	3	2	3
Cultivar C	6	3	3	2
Cultivar D	6	3	2	3
Cultivar E	6	3	3	2
Total per block	386	368	370	370
Total per greenhouse		754		740

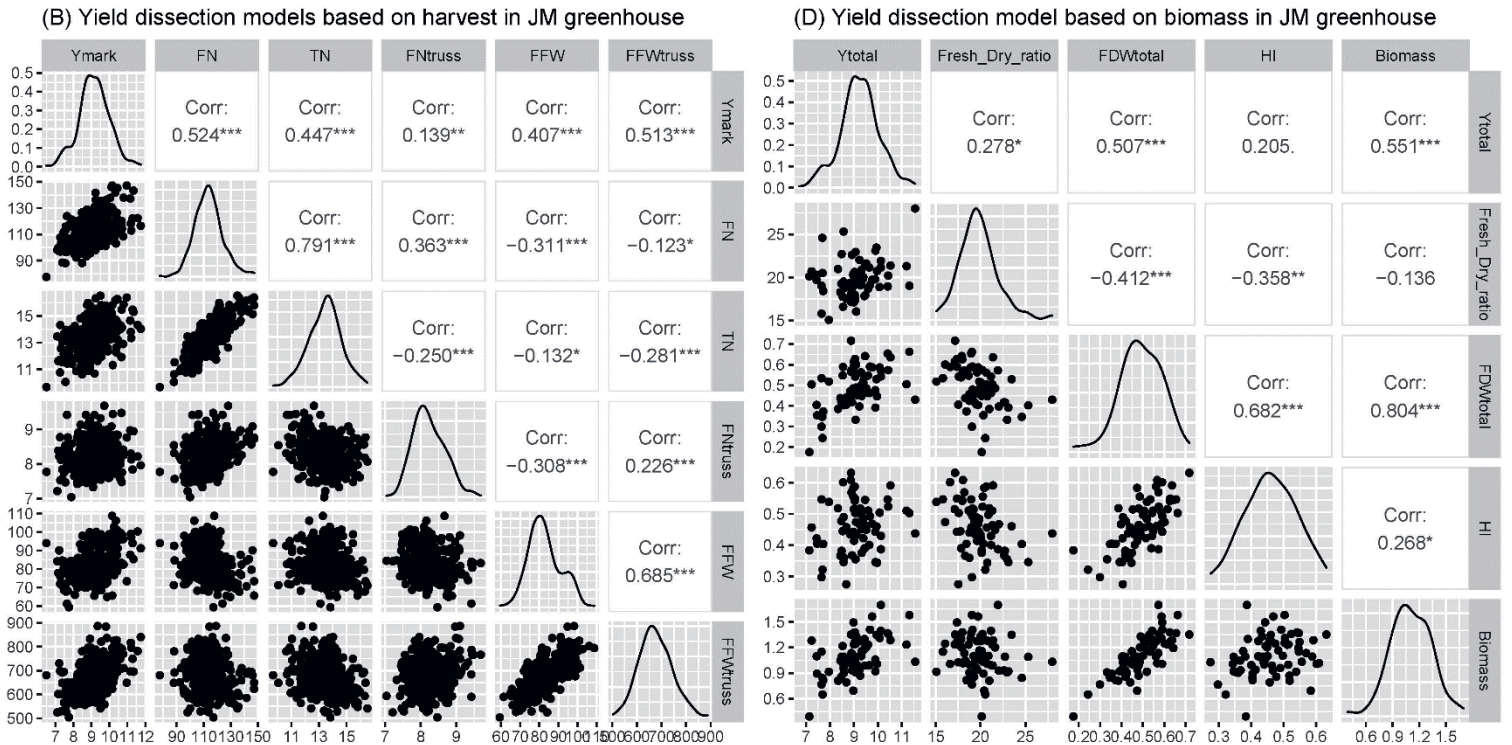
Supplementary Figure 2 Experimental design in greenhouse. SpATS Row and SpATS Col indicate the row and column numbers, which were used for correction of spatial trend by R-package SpATS.



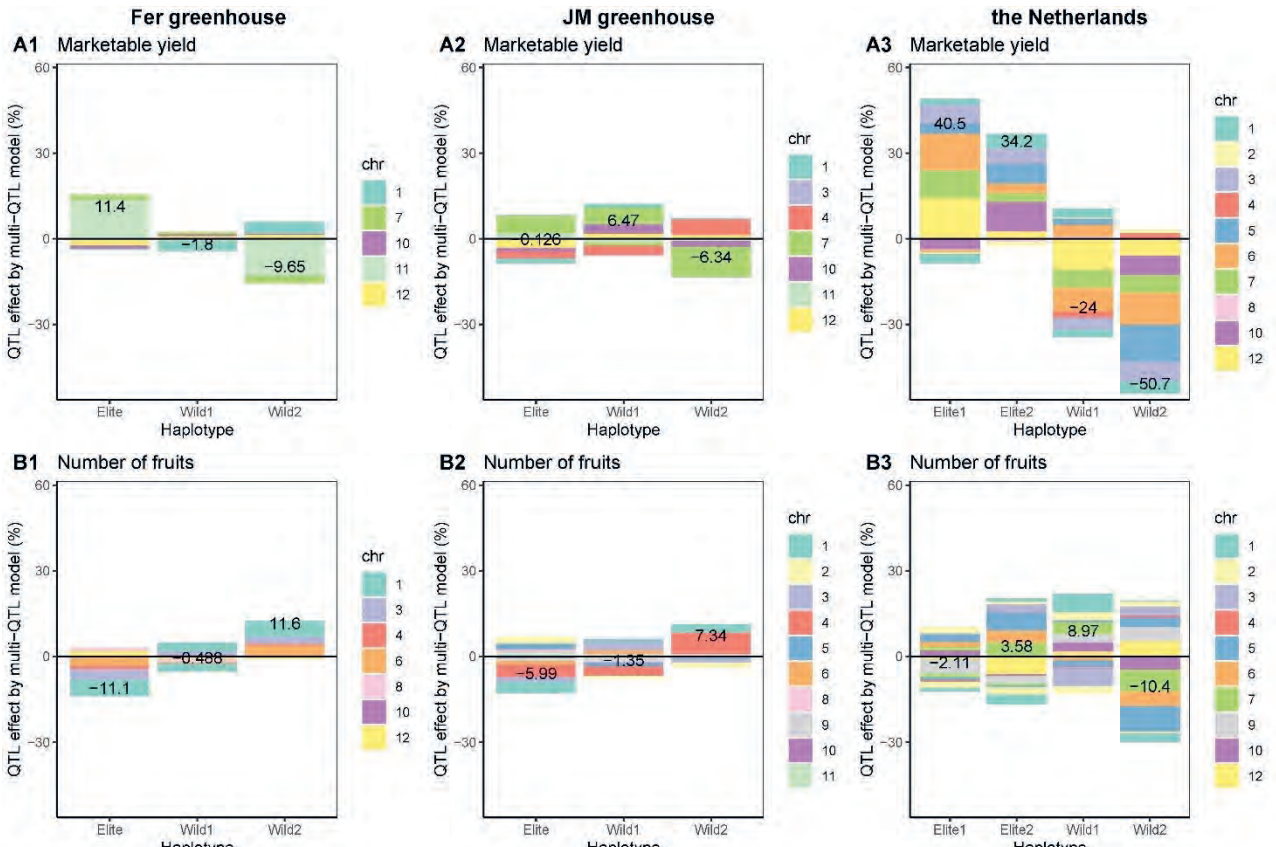
Supplementary Figure 3 Subset comprised of 110 genotypes 160 plots for phenotyping of component traits in the biomass dissection.



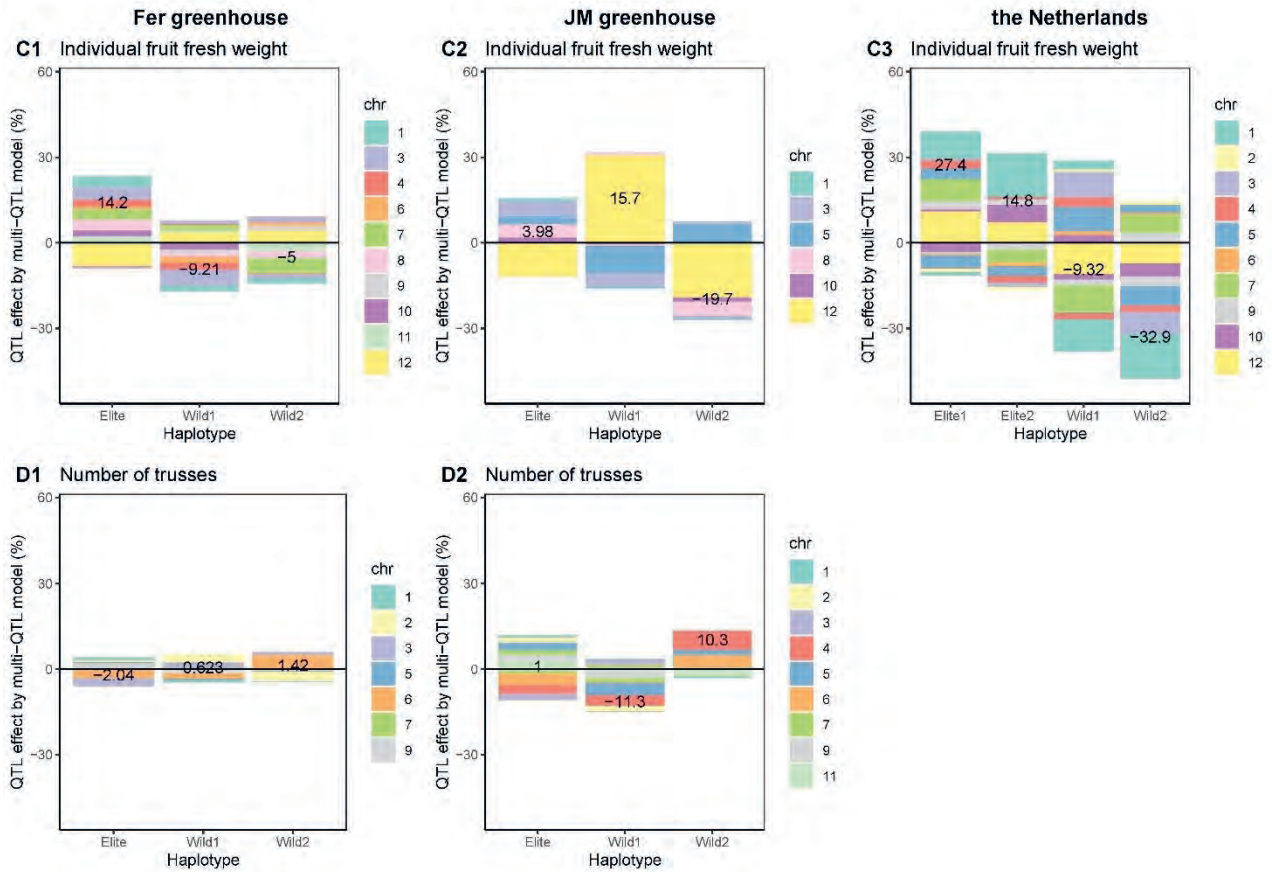
*Supplementary Figure 4 Correlation between traits. (A) Harvest dissection in Fer greenhouse. (B) Harvest dissection in JM greenhouse. (C) Biomass dissection in Fer greenhouse. (D) Biomass dissection in JM greenhouse. *significant correlation at $P < 0.05$.*



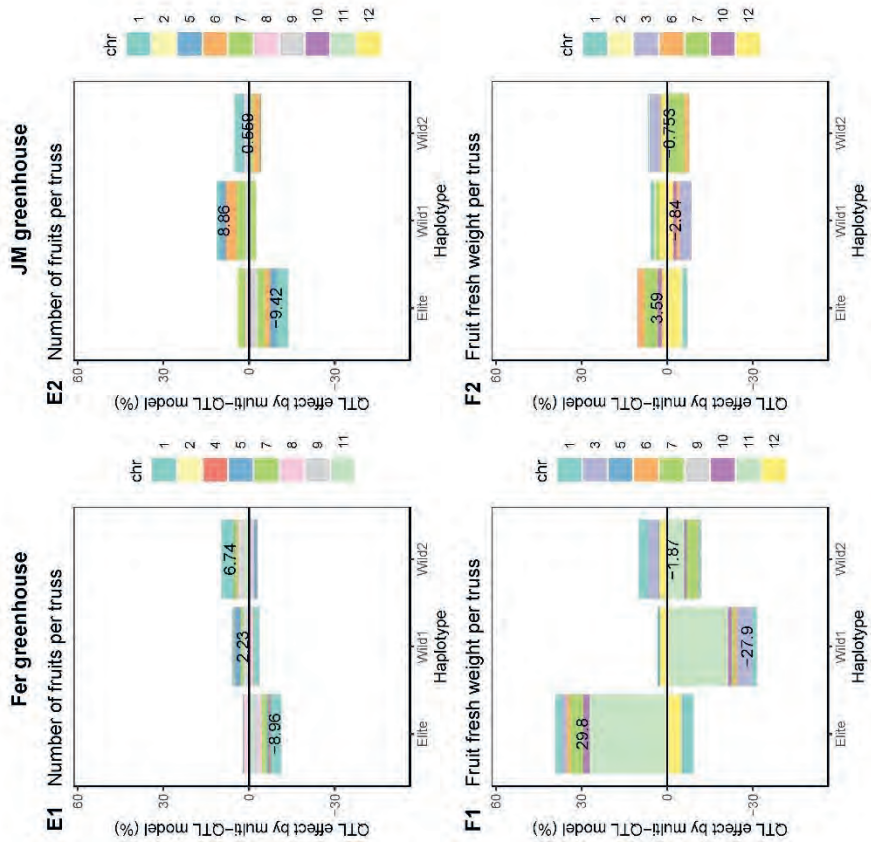
Supplementary Figure 4 Continued.



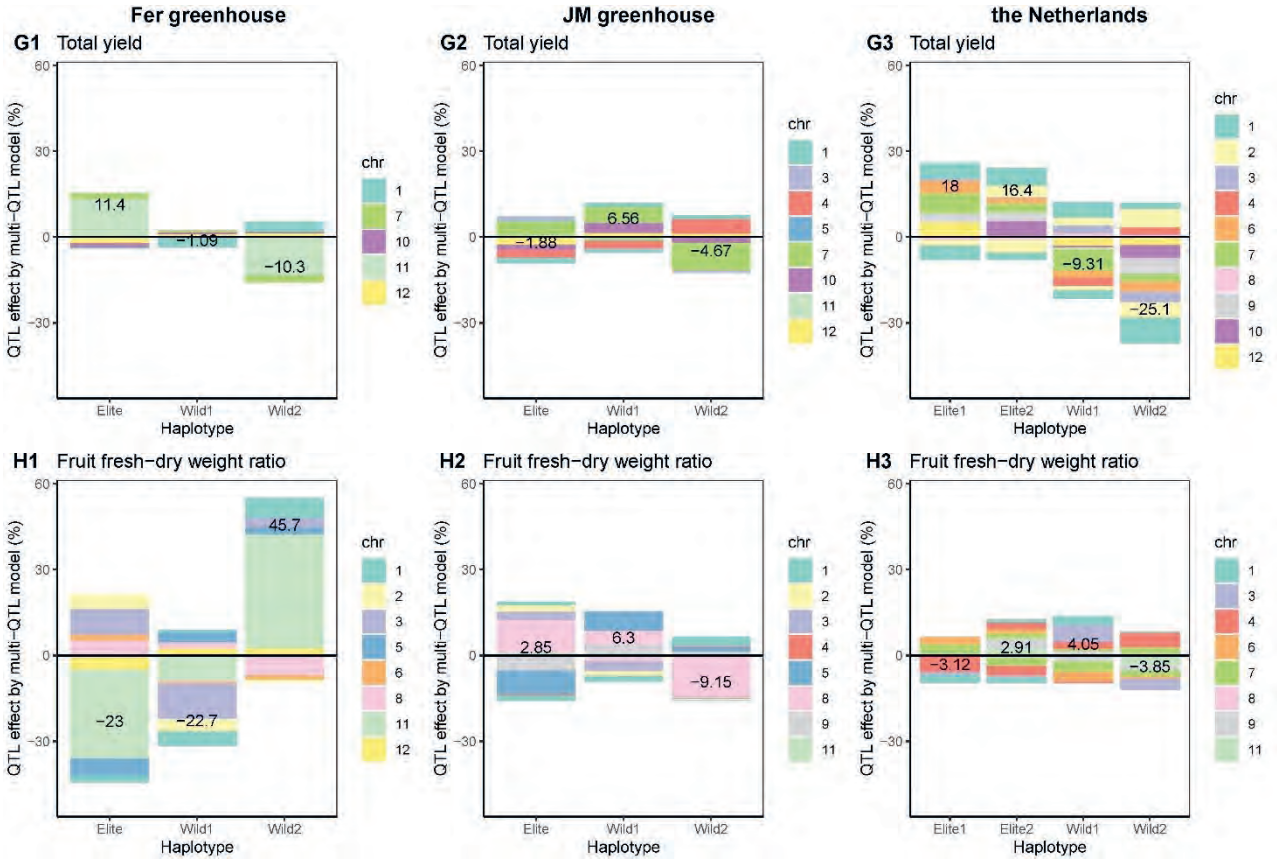
Supplementary Figure 5 Haplotype effects per chromosome and multi-locus predictions for the parents. (1) Fer greenhouse. (2) JM greenhouse. (3) Dutch trials (population for Dutch production in Dutch environment) as a reference. Haplotype effects are shown by proportion (%) to the general mean in multi-QTL model for each trait, and have colour of chromosome. Multi-locus prediction presented as number.



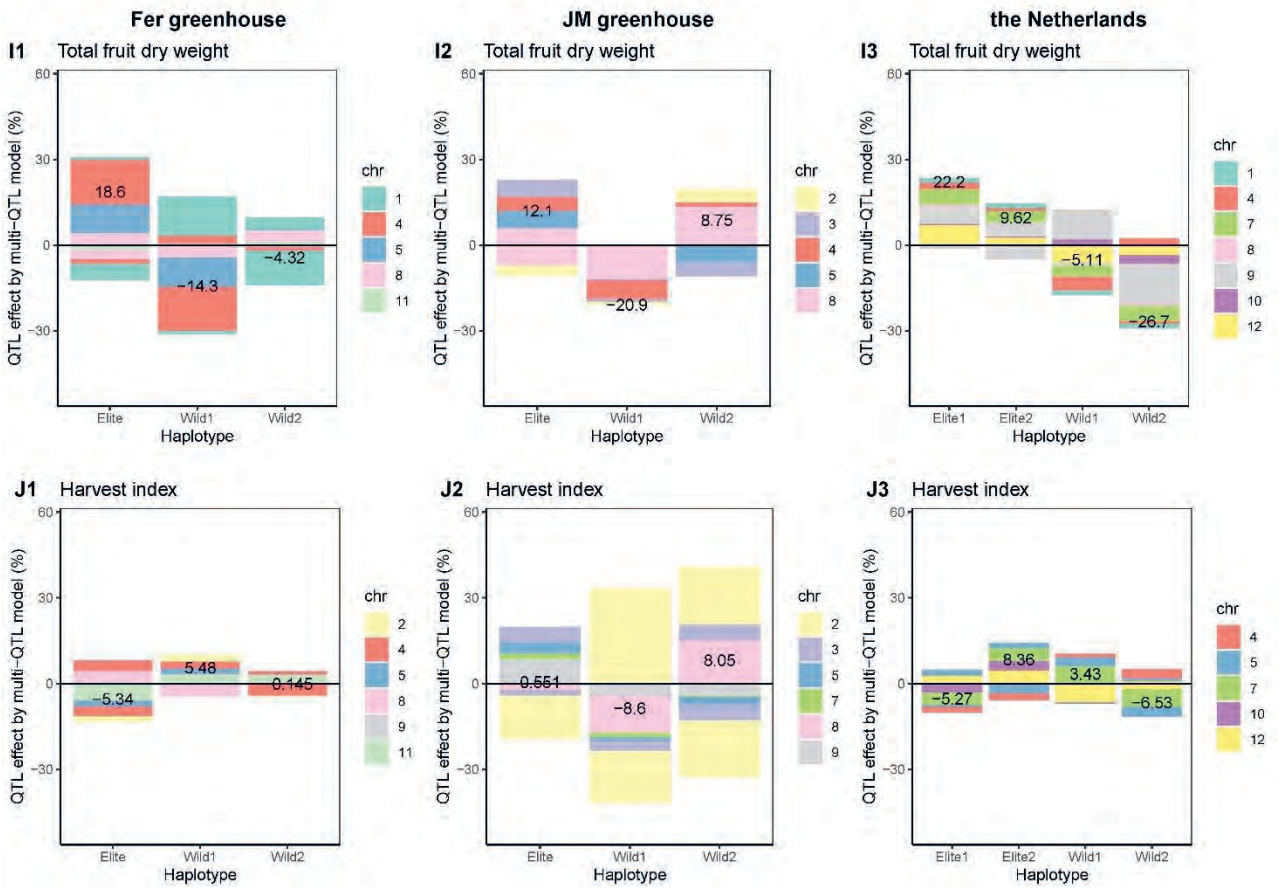
Supplementary Figure 5 Continued.



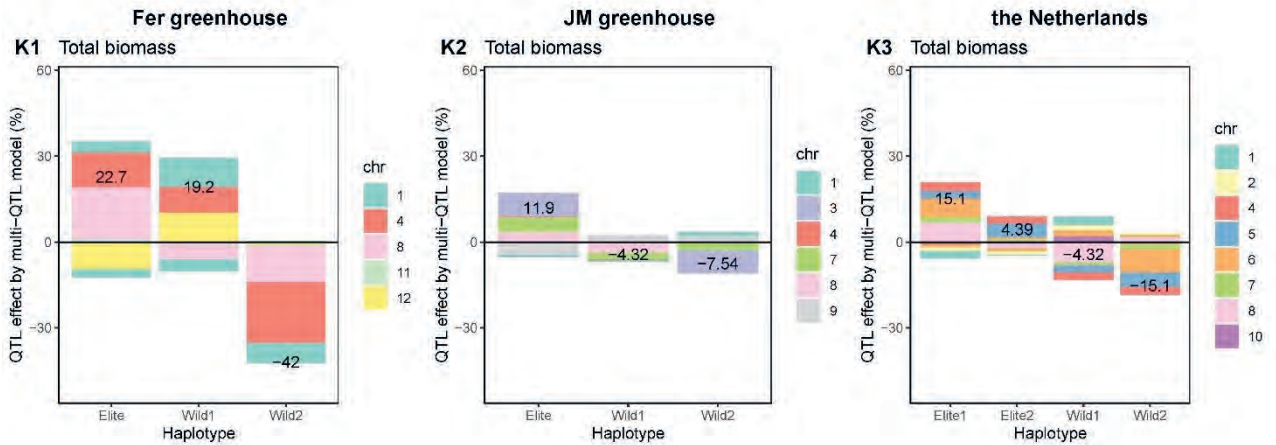
Supplementary Figure 5 Continued.



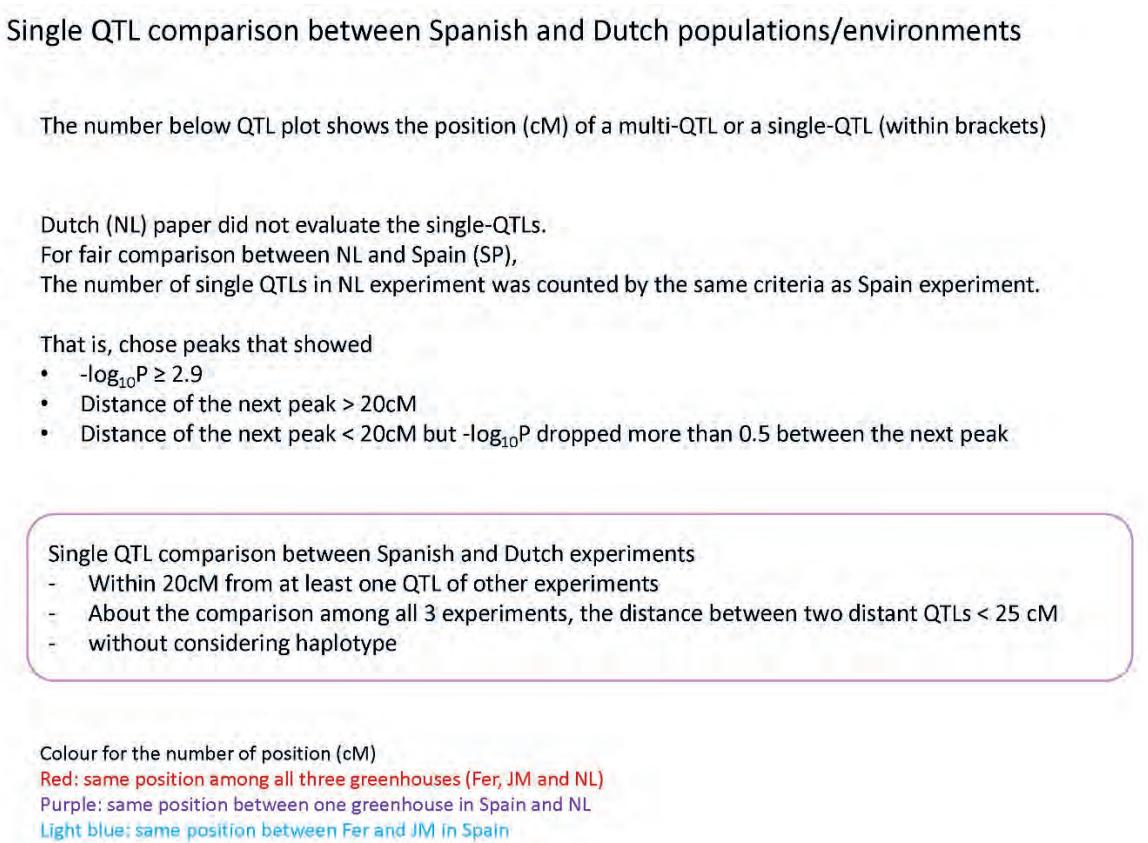
Supplementary Figure 5 Continued.



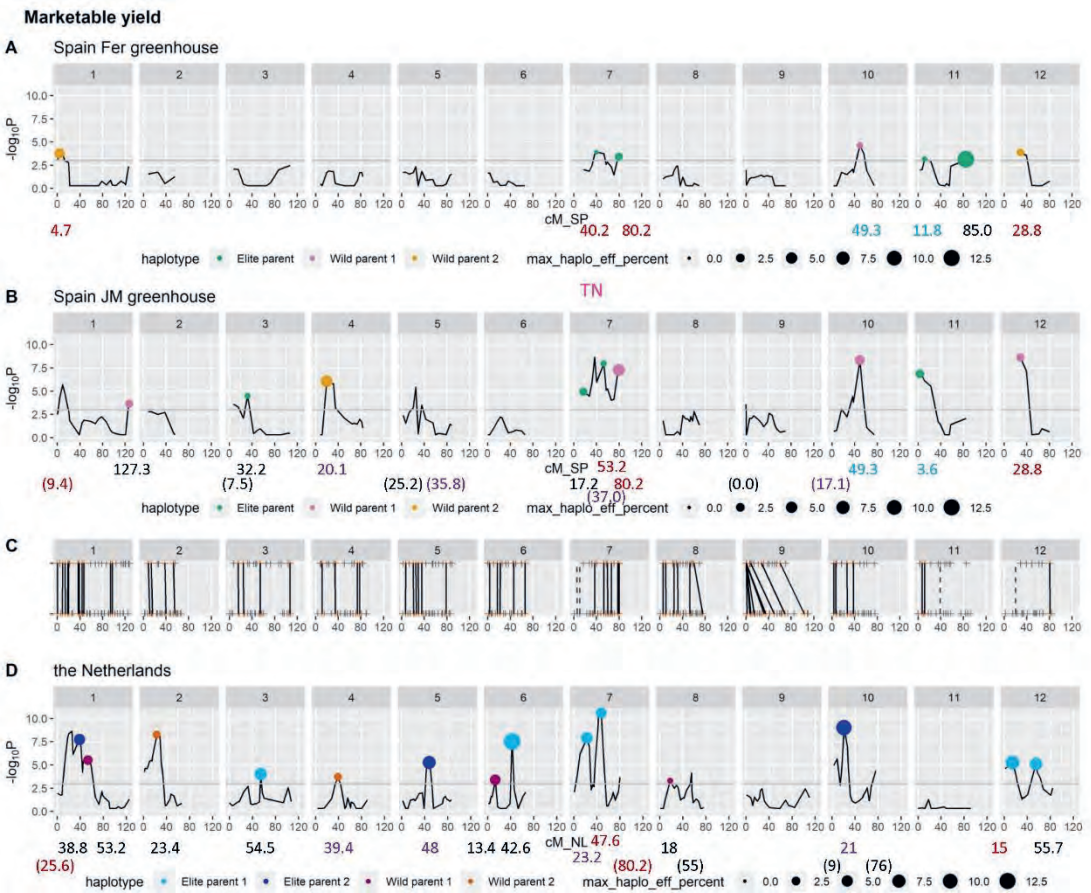
Supplementary Figure 5 Continued.



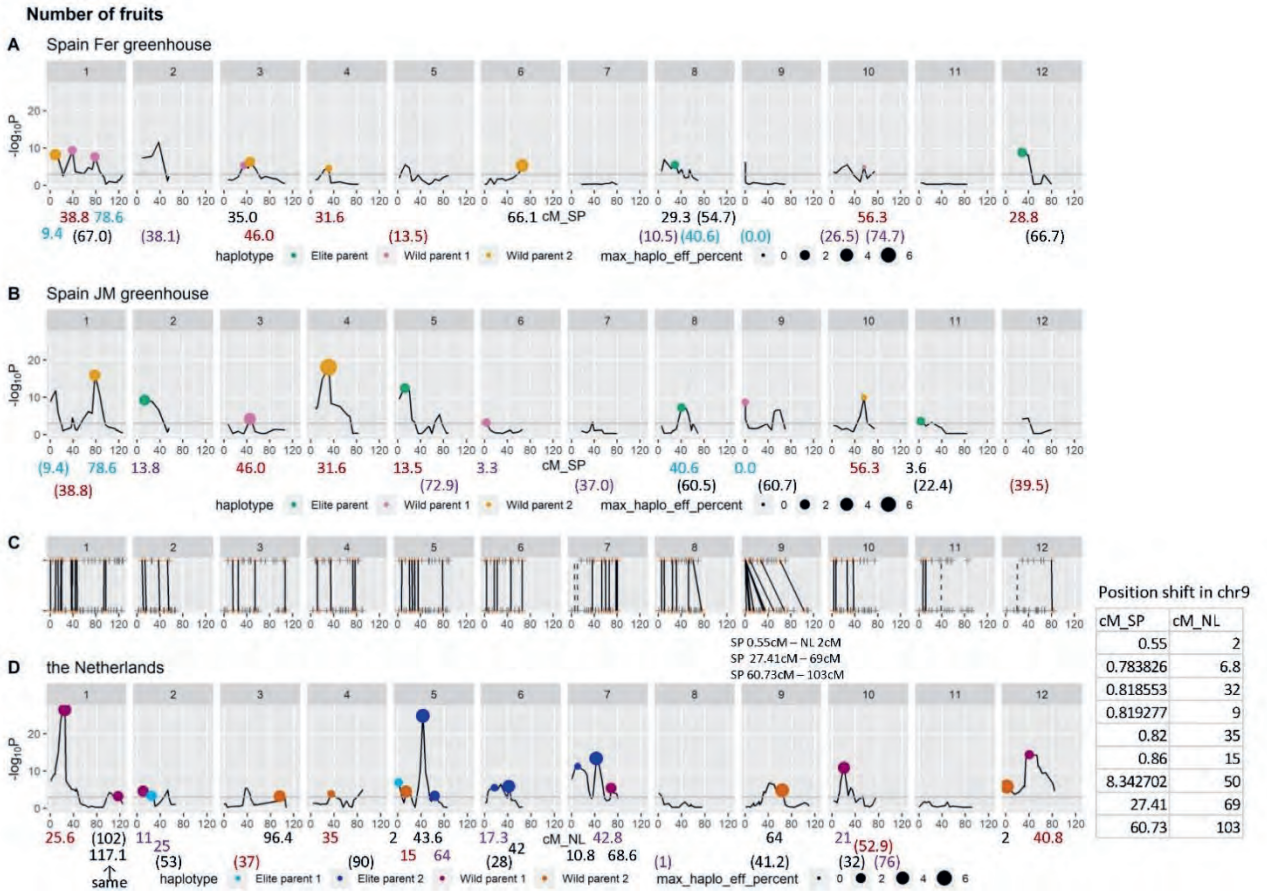
Supplementary Figure 5 Continued.



Supplementary Figure 6 Single-QTL comparison for each trait between Fer and JM greenhouses in Spain, and between Spanish and Dutch trials (different populations and environments). On vertical axis, -log10P value in single-QTL analysis; on horizontal axis, chromosome 1 to 12, with position (cM) within chromosome.



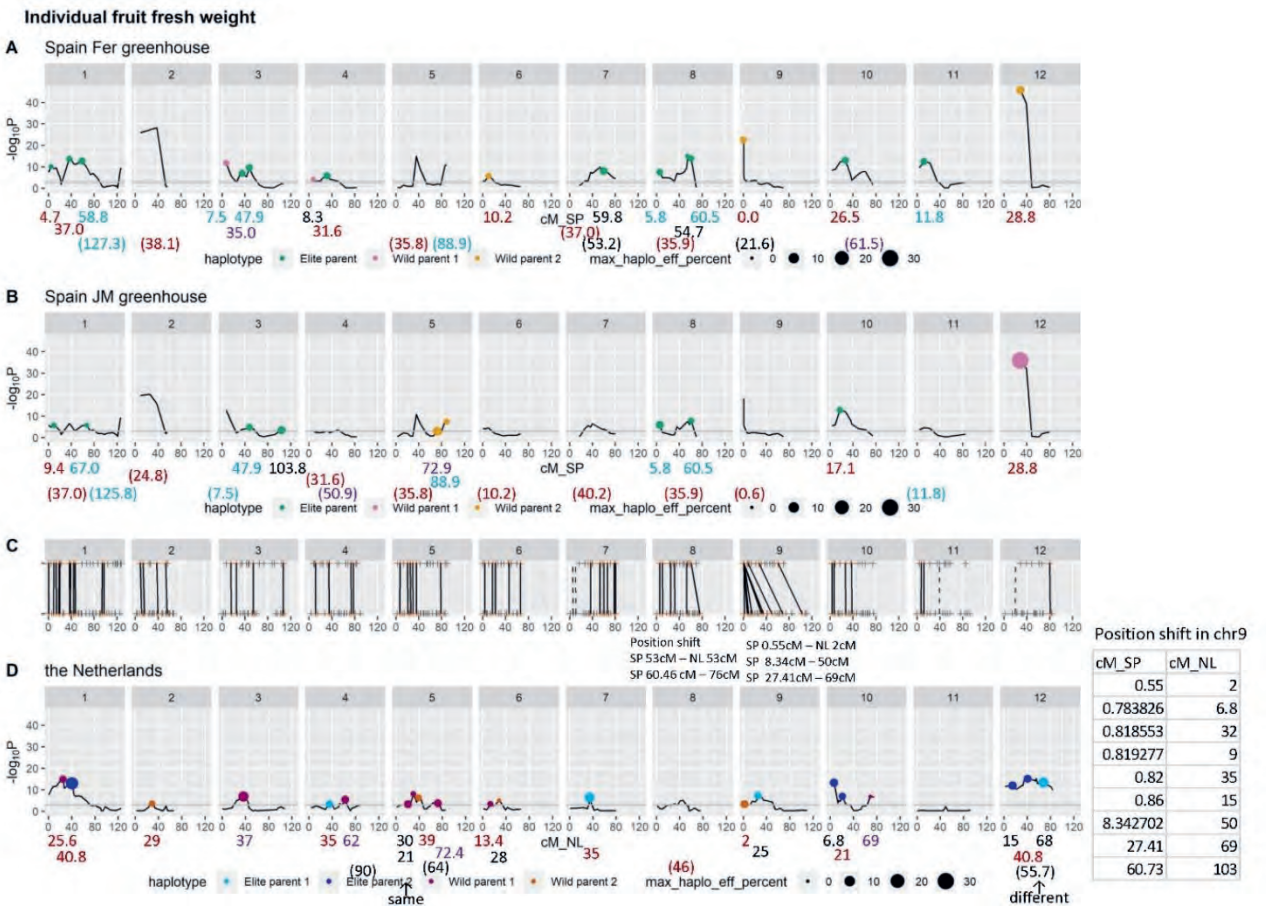
Supplementary Figure 6 Continued.



Position shift in chr9

cM_SP	cM_NL
0.55	2
0.783826	6.8
0.818553	32
0.819277	9
0.82	35
0.86	15
8.342702	50
27.41	69
60.73	103

Supplementary Figure 6 Continued.

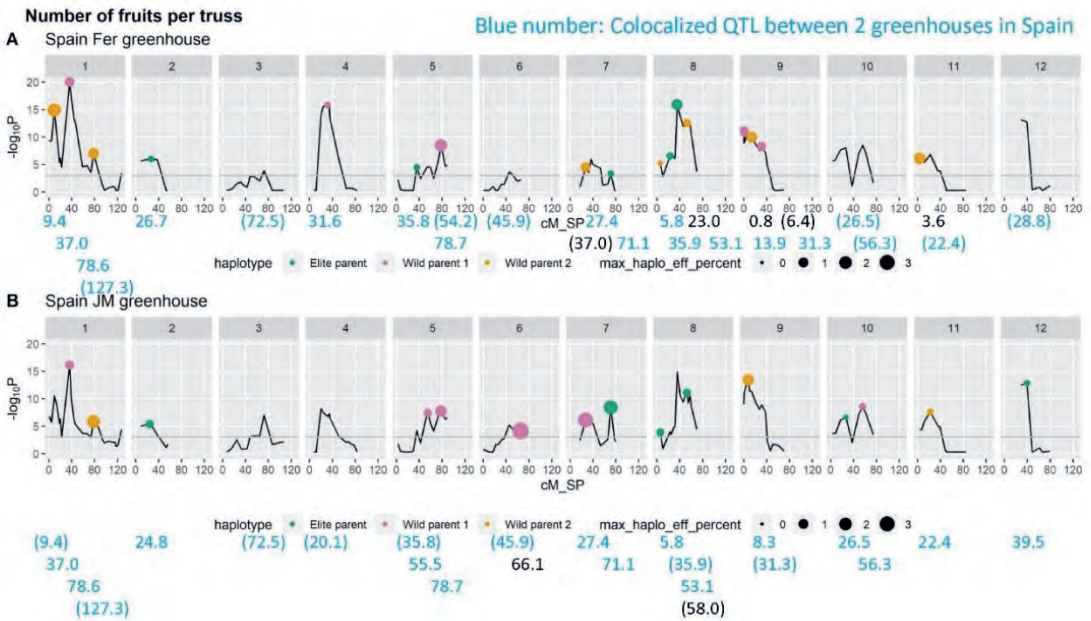


Position shift in chr9

cM_SP	cM_NL
0.55	2
0.783826	6.8
0.818553	32
0.819277	9
0.82	35
0.86	15
8.342702	50
27.41	69
60.73	103

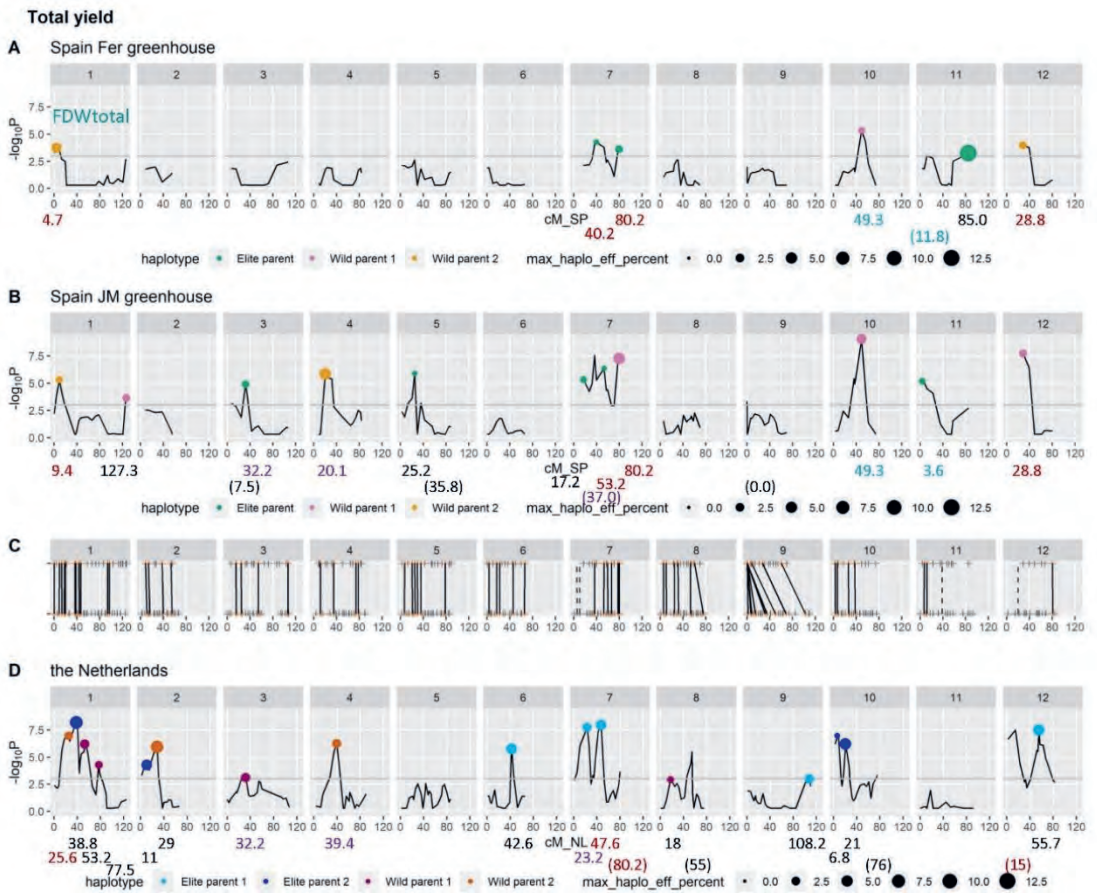
Supplementary Figure 6 Continued.

Supplementary Figure 6 Continued.

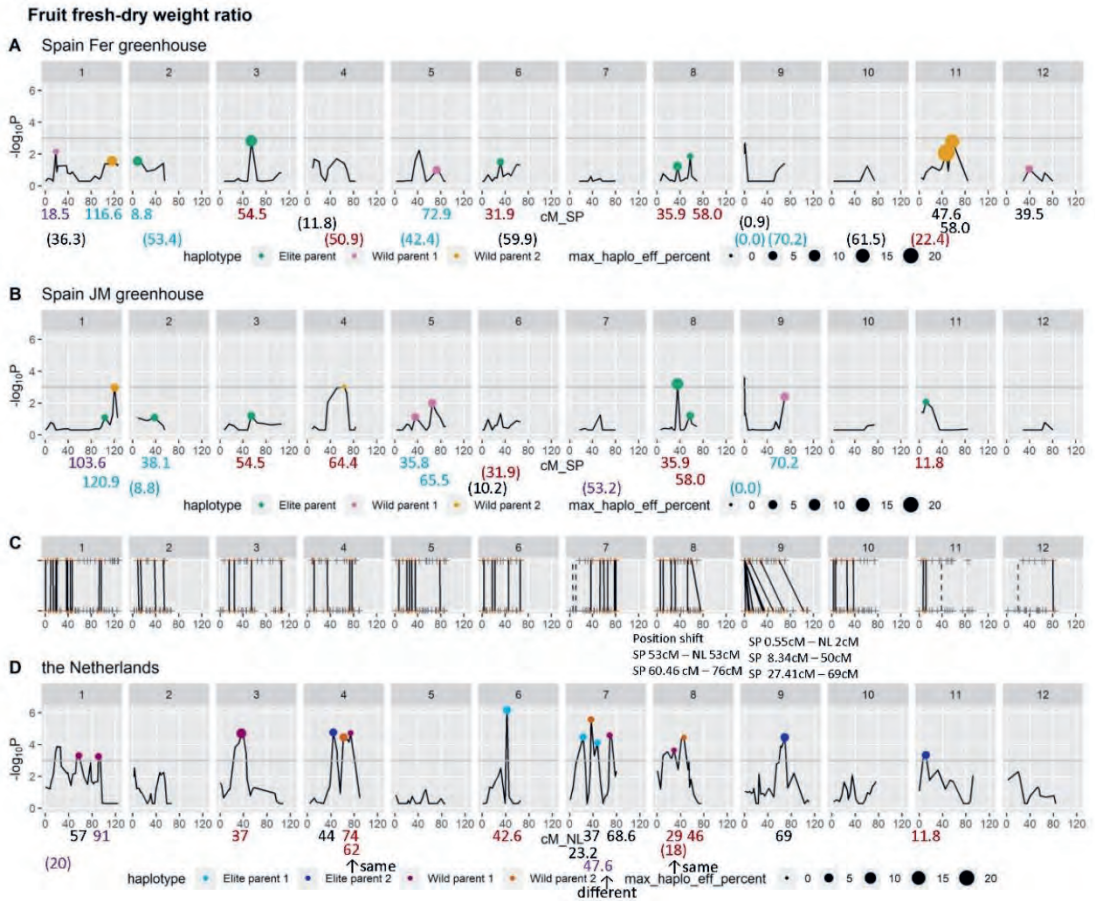


Supplementary Figure 6 Continued.

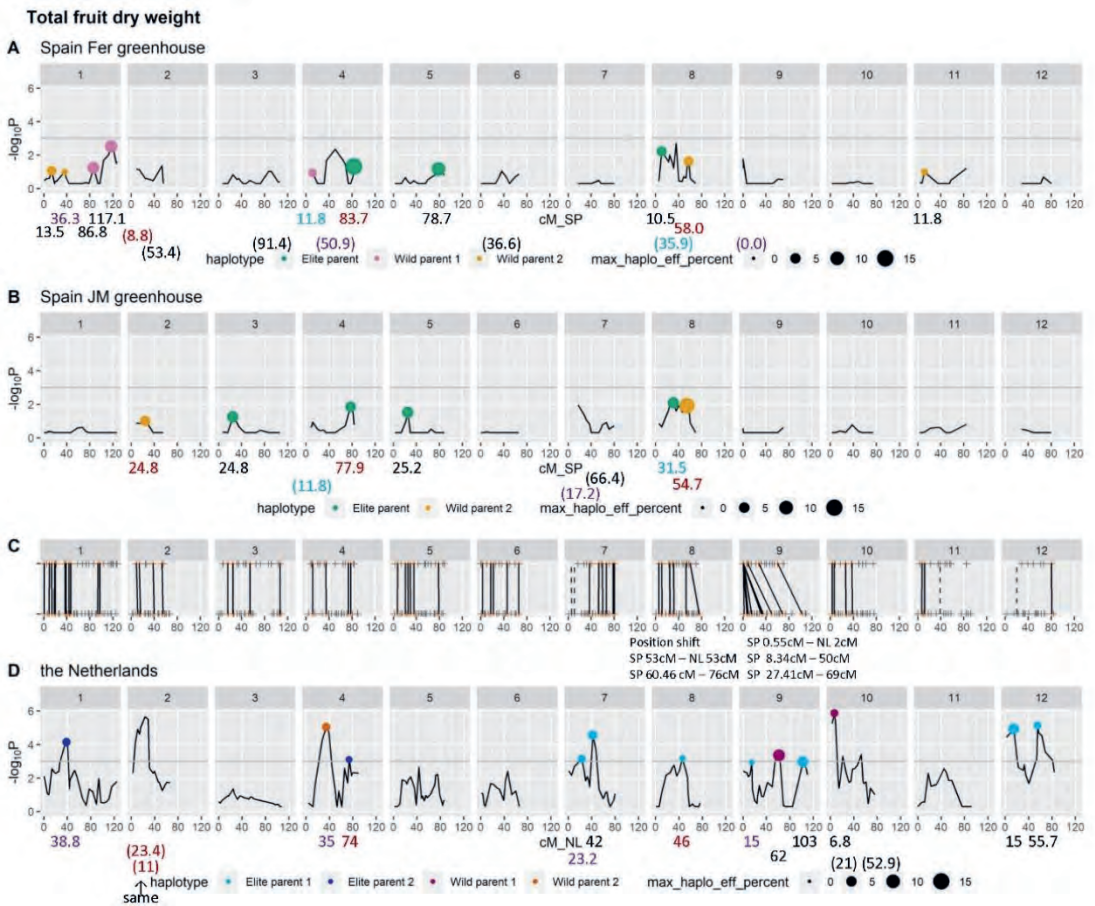
Supplementary Figure 6 Continued.



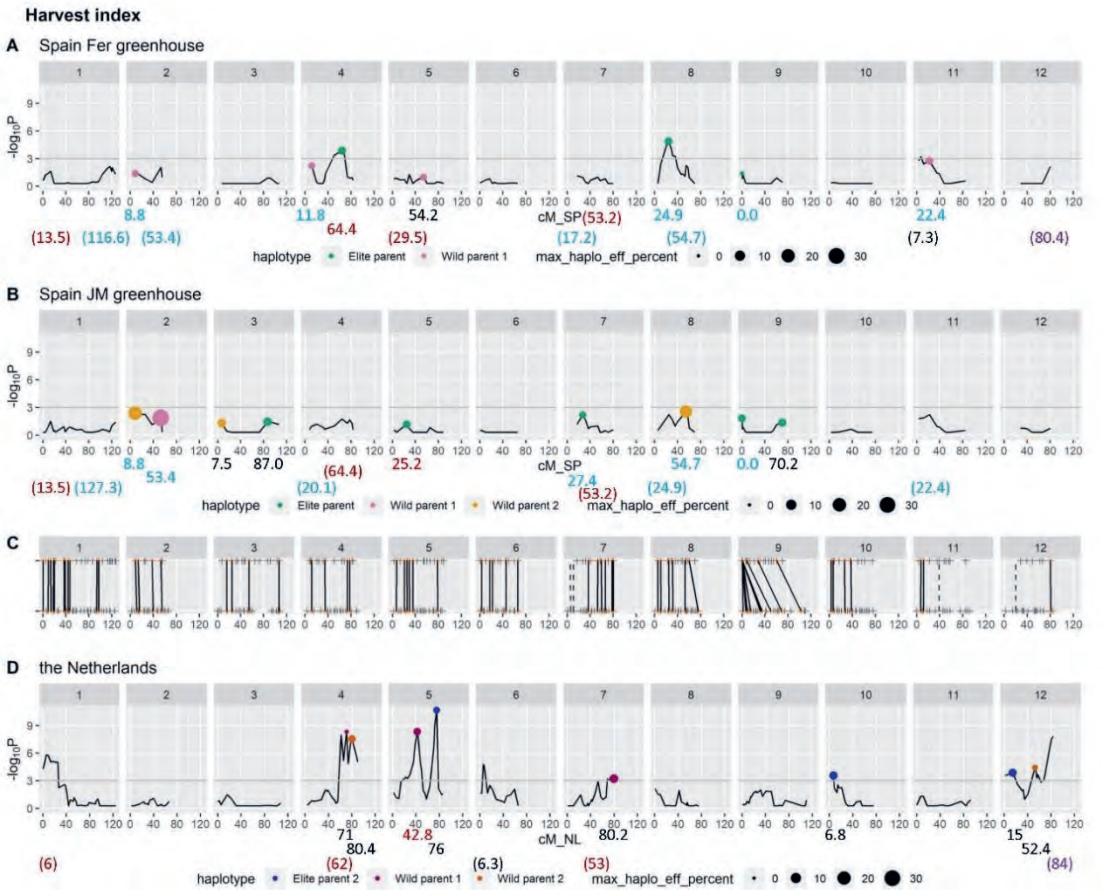
Supplementary Figure 6 Continued.



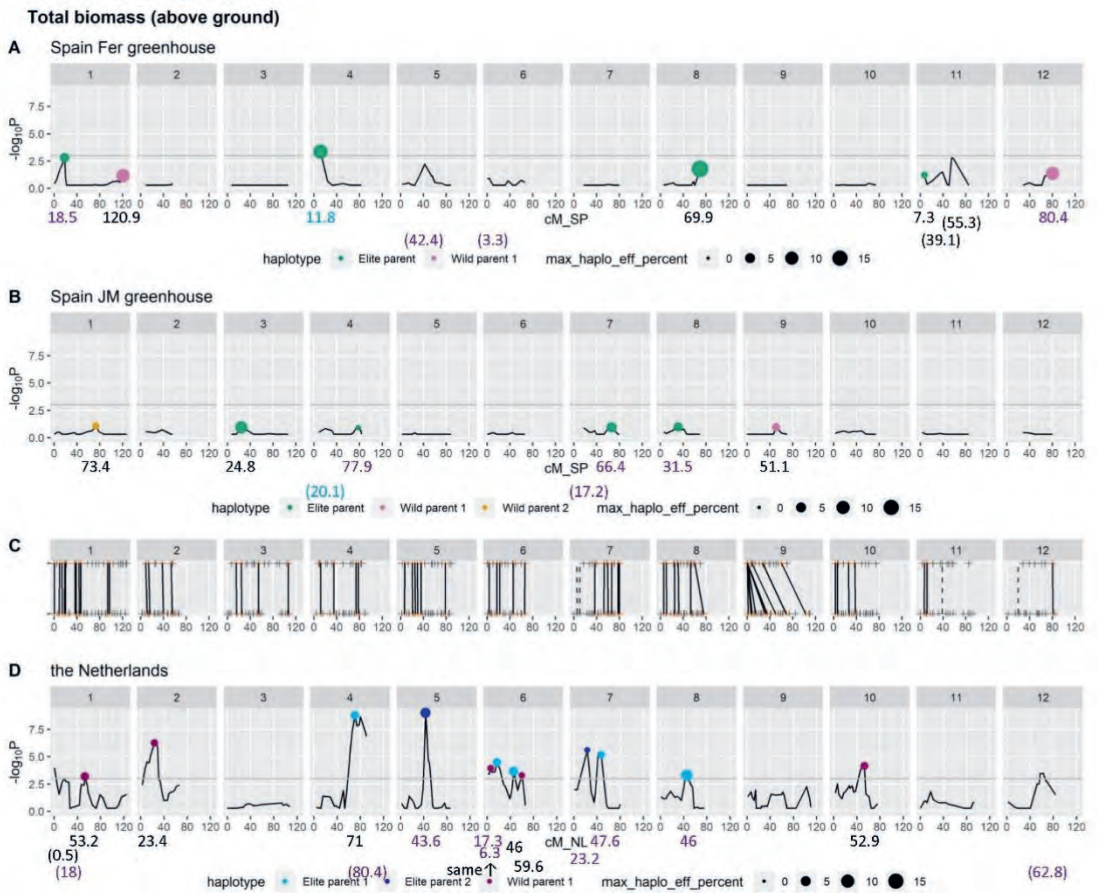
Supplementary Figure 6 Continued.



Supplementary Figure 6 Continued.



Supplementary Figure 6 Continued.

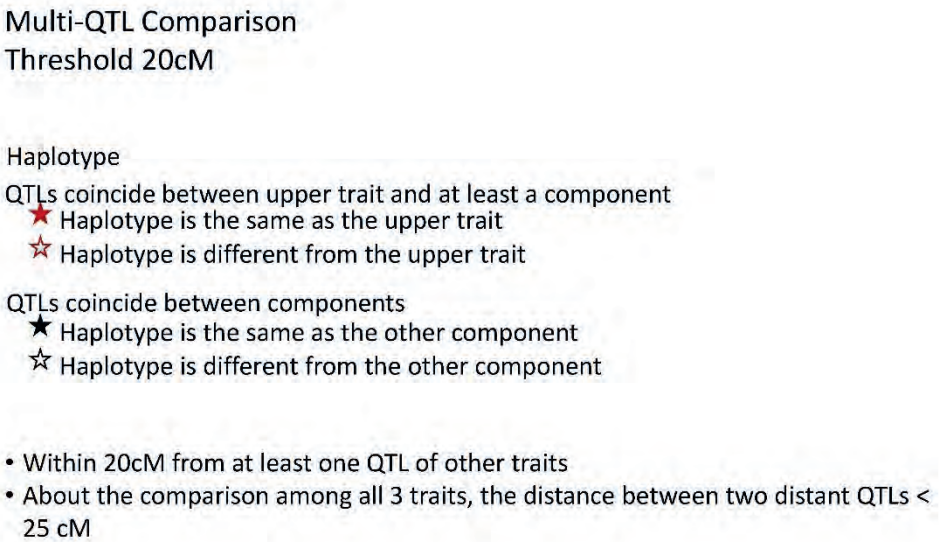


Supplementary Figure 6 Continued.

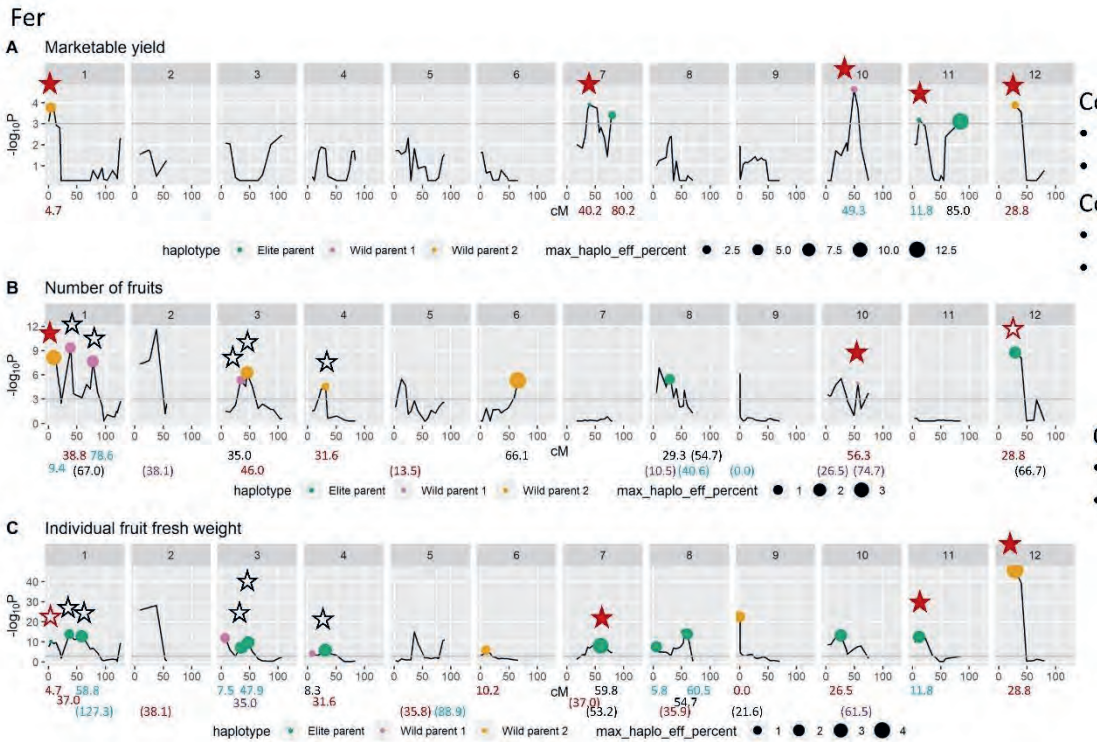
Single QTL comparison without considering haplotype

	red	purple	purple	blue	Fer only	JM only	NL only			
Spain (Fer)	√	√		√	√					
Spain (JM)	√		√	√		√				
NL	√	√	√				√	Fer total	JM total	NL total
Total yield	4	0	3	2	1	6	13	7	15	20
Marketable yield	4	0	4	2	1	6	11	7	16	19
Number of fruits	6	4	4	4	6	4	14	20	18	28
Number of trusses	-	-	-	11	6	13	-	17	24	-
Number of fruits per truss	-	-	-	22	5	2	-	27	24	-
Individual fruit fresh weight	11	2	2	8	5	1	9	26	22	24
Fruit fresh weight per truss	-	-	-	13	13	0	-	26	13	-
Fruit fresh-dry weight ratio	6	1	2	7	8	1	6	22	16	15
Total fruit dry weight	3	3	1	2	9	3	8	17	9	15
Harvest index	4	1	0	9	2	3	8	16	16	13
Total biomass	0	4	4	1	5	3	7	10	8	15

Supplementary Figure 6 Continued.



Supplementary Figure 7 Multi-QTL comparison between traits within each experiment. On vertical axis, -log$_{10}$P value in single-QTL analysis; on horizontal axis, chromosome 1 to 12, with position (cM) within chromosome. Colour of dot shows haplotype with highest positive effect, and the size of dot represents proportion (%) of the largest haplotype effect to mean for the trait in the multi-QTL model.

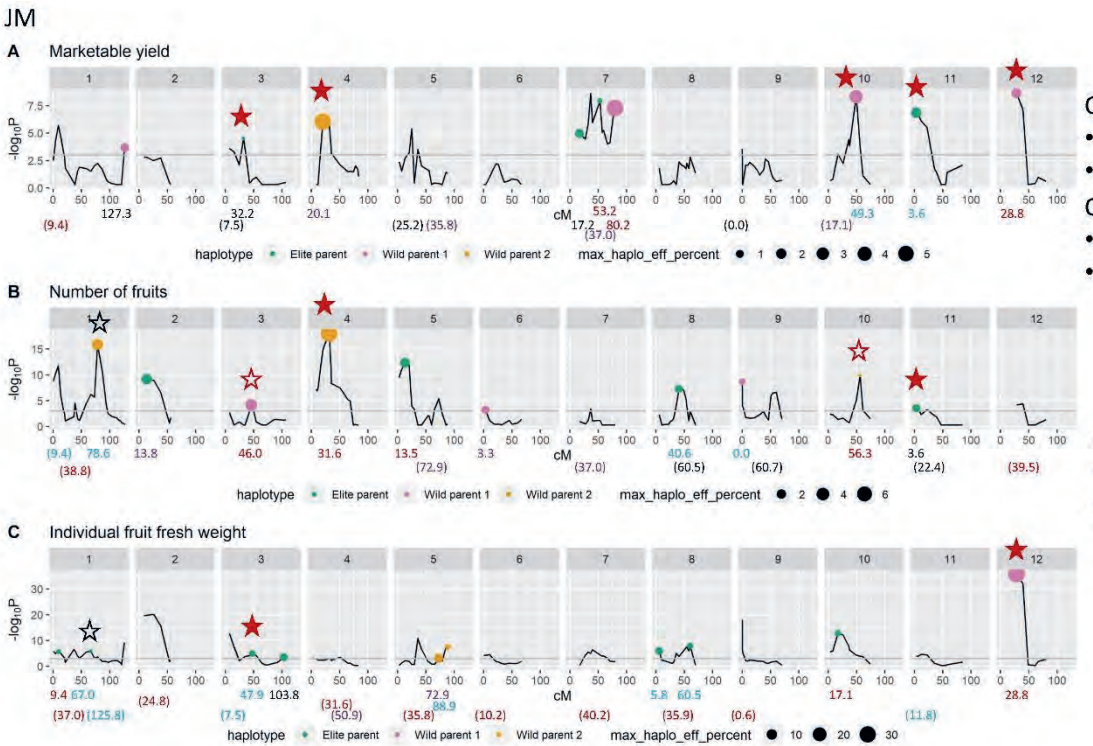


Common multi-QTL between A and B
- With the same haplotype: 2
- With different haplotype: 1

Common multi-QTL between A and C
- With the same haplotype: 3
- With different haplotype: 1

Common multi-QTL between B and C
- With the same haplotype: 0
- With different haplotype: 7

Supplementary Figure 7 Continued.

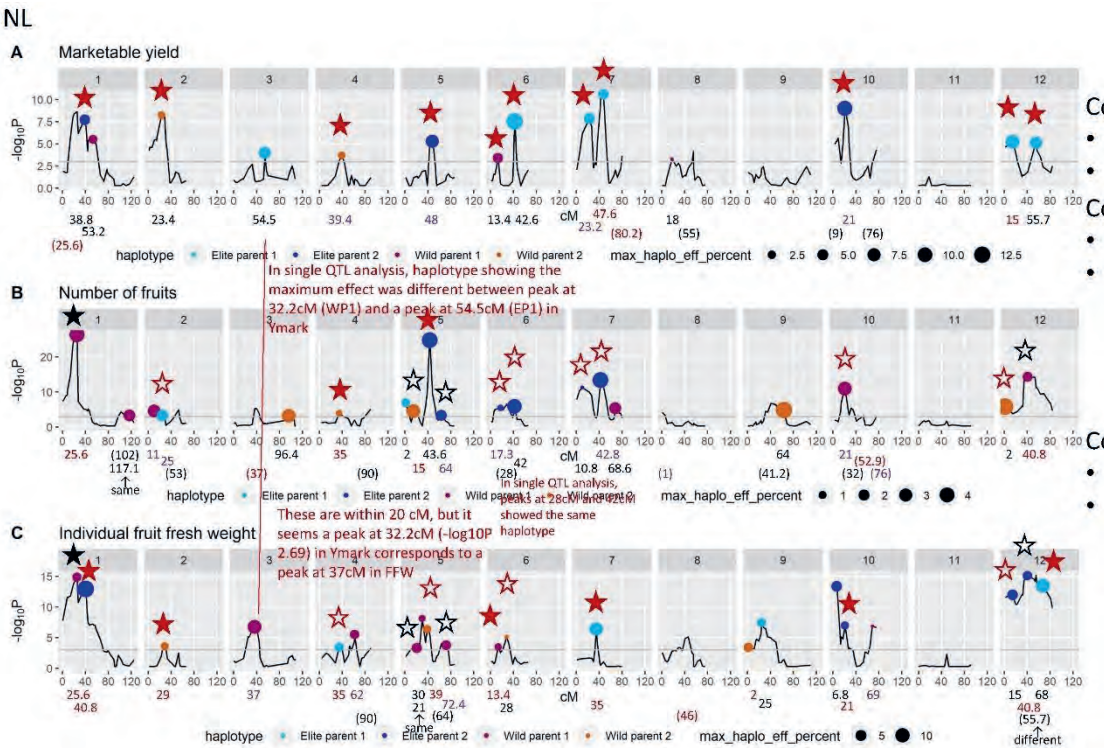


Common multi-QTL between A and B
- With the same haplotype: 2
- With different haplotype: 2

Common multi-QTL between A and C
- With the same haplotype: 2
- With different haplotype: 0

Common multi-QTL between B and C
- With the same haplotype: 0
- With different haplotype: 2

Supplementary Figure 7 Continued.

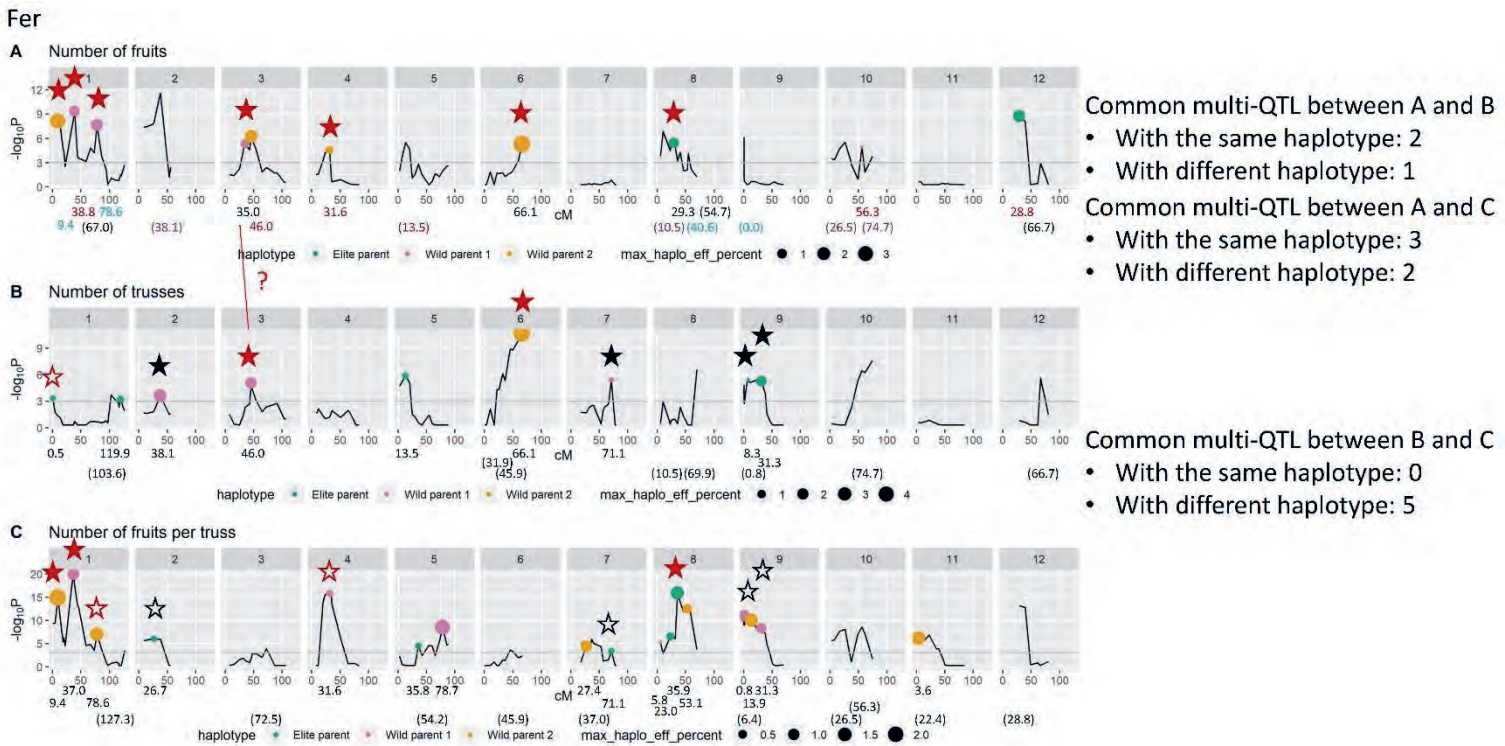


Common multi-QTL between A and B
- With the same haplotype: 2
- With different haplotype: 8

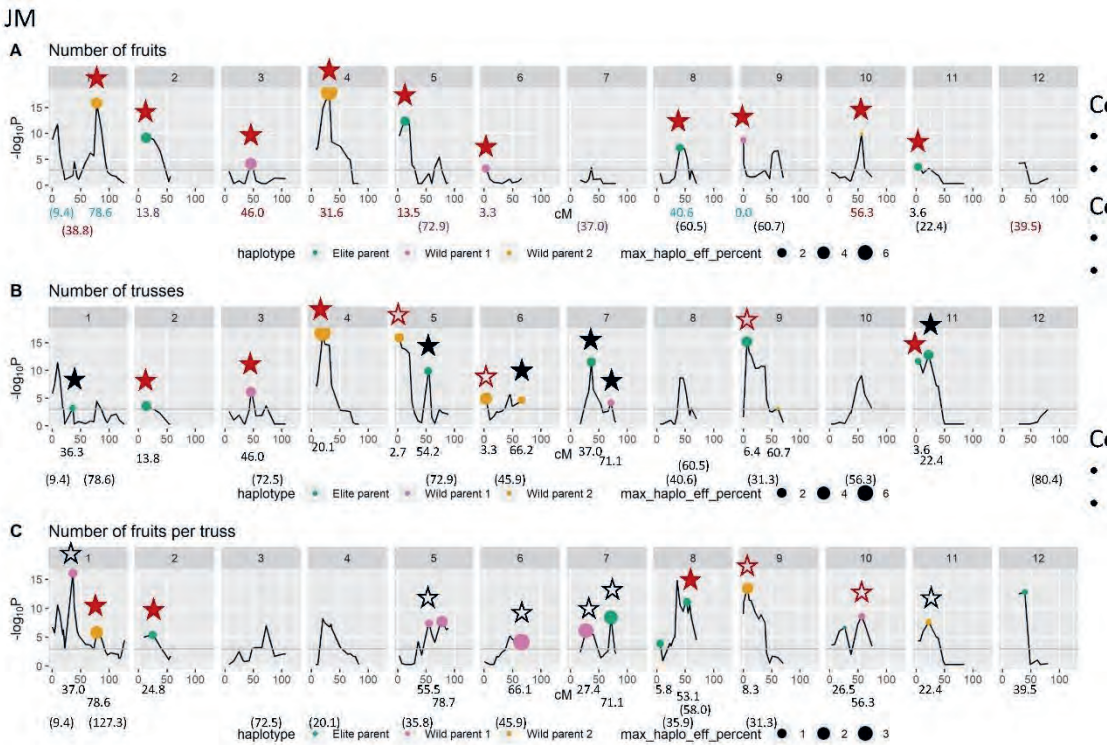
Common multi-QTL between A and C
- With the same haplotype: 6
- With different haplotype: 4

Common multi-QTL between B and C
- With the same haplotype: 1
- With different haplotype: 11

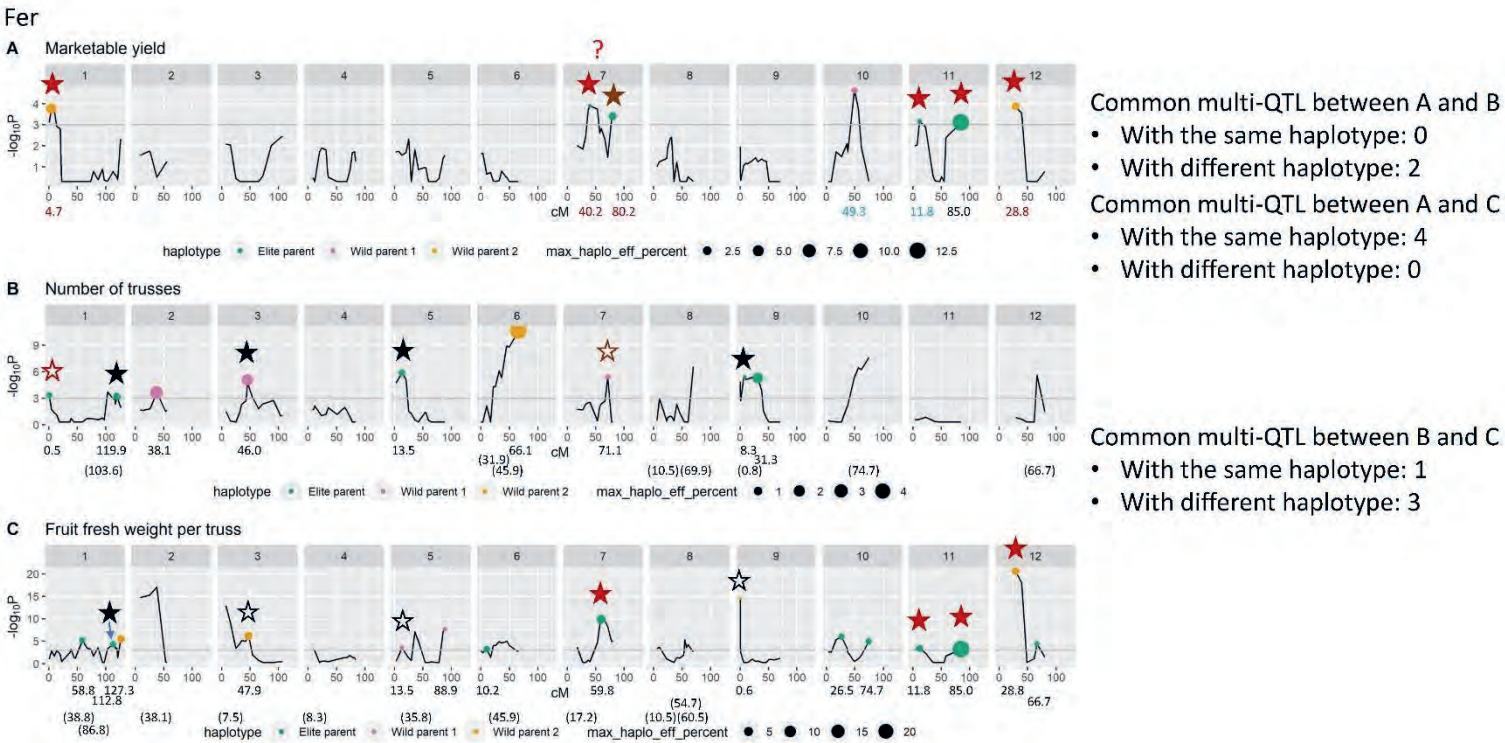
Supplementary Figure 7 Continued.



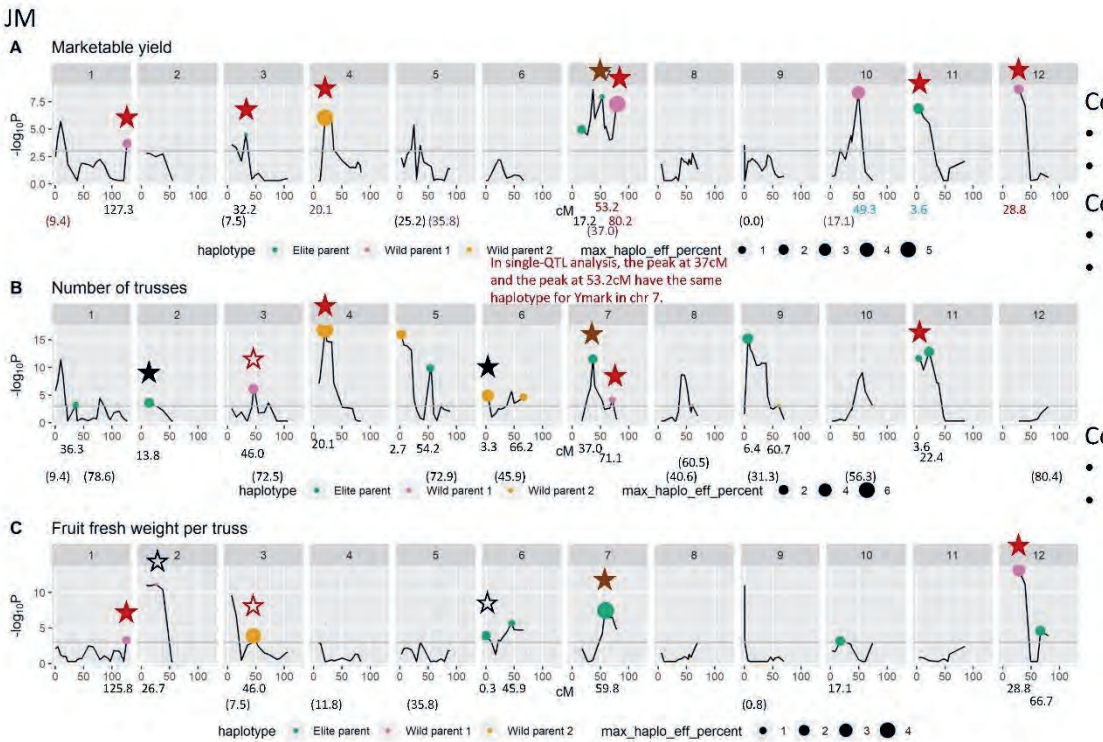
Supplementary Figure 7 Continued.



Supplementary Figure 7 Continued.



Supplementary Figure 7 Continued.

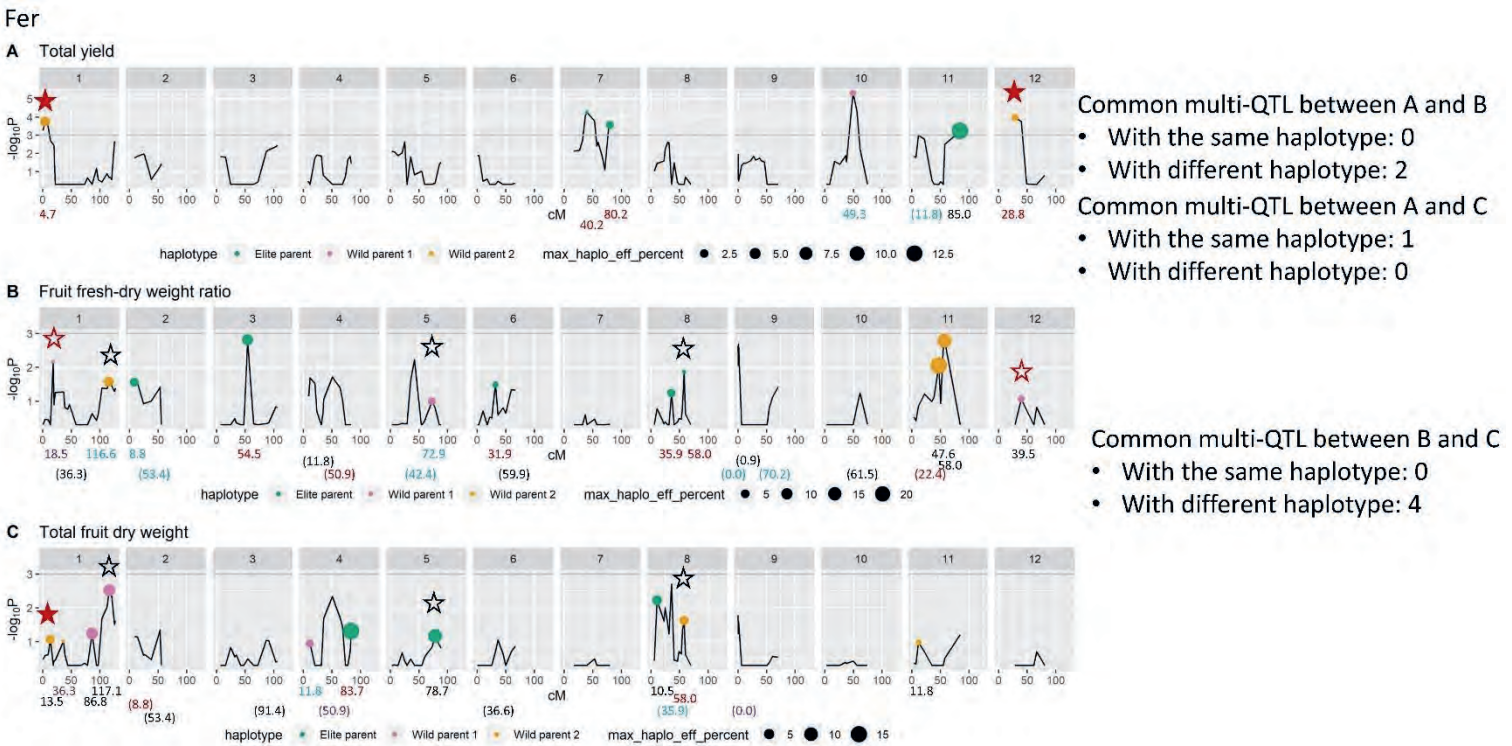


Common multi-QTL between A and B
- With the same haplotype: 4
- With different haplotype: 1

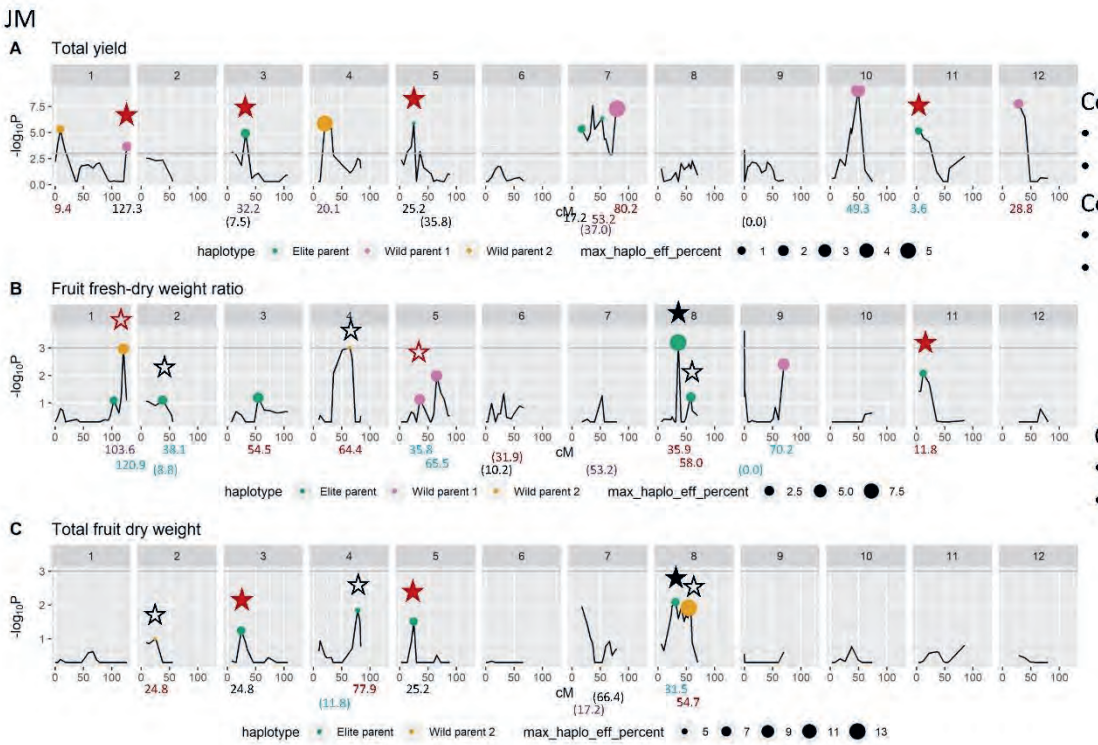
Common multi-QTL between A and C
- With the same haplotype: 3
- With different haplotype: 1

Common multi-QTL between B and C
- With the same haplotype: 1
- With different haplotype: 3

Supplementary Figure 7 Continued.



Supplementary Figure 7 Continued.

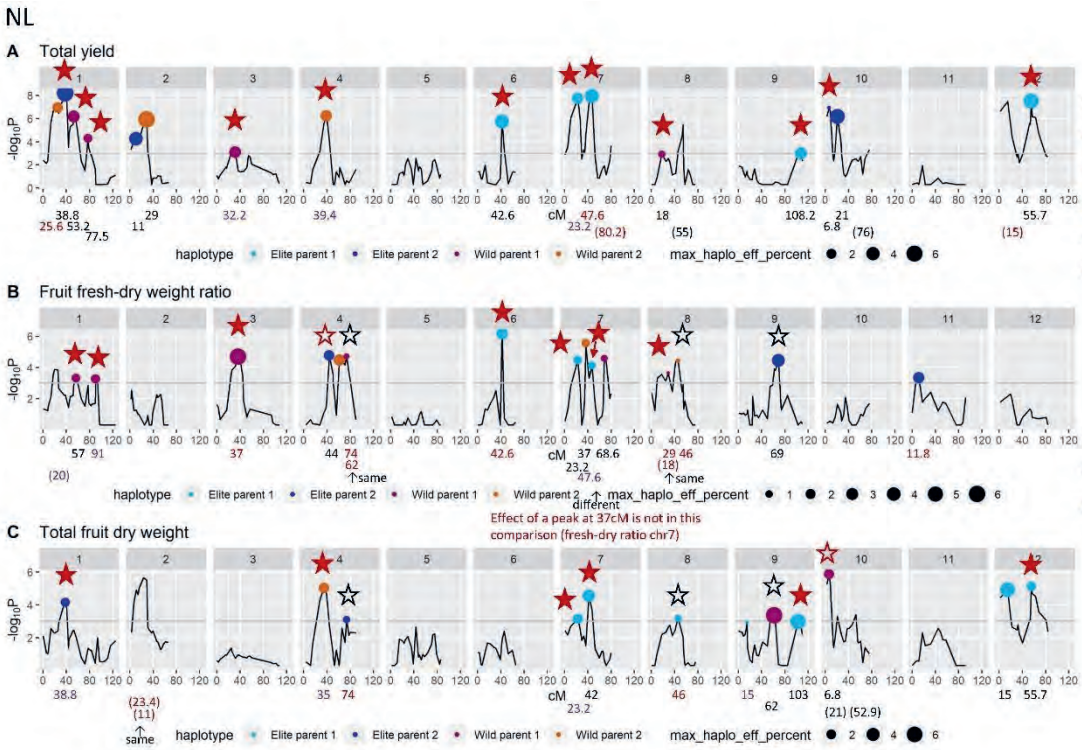


Common multi-QTL between A and B
- With the same haplotype: 1
- With different haplotype: 2

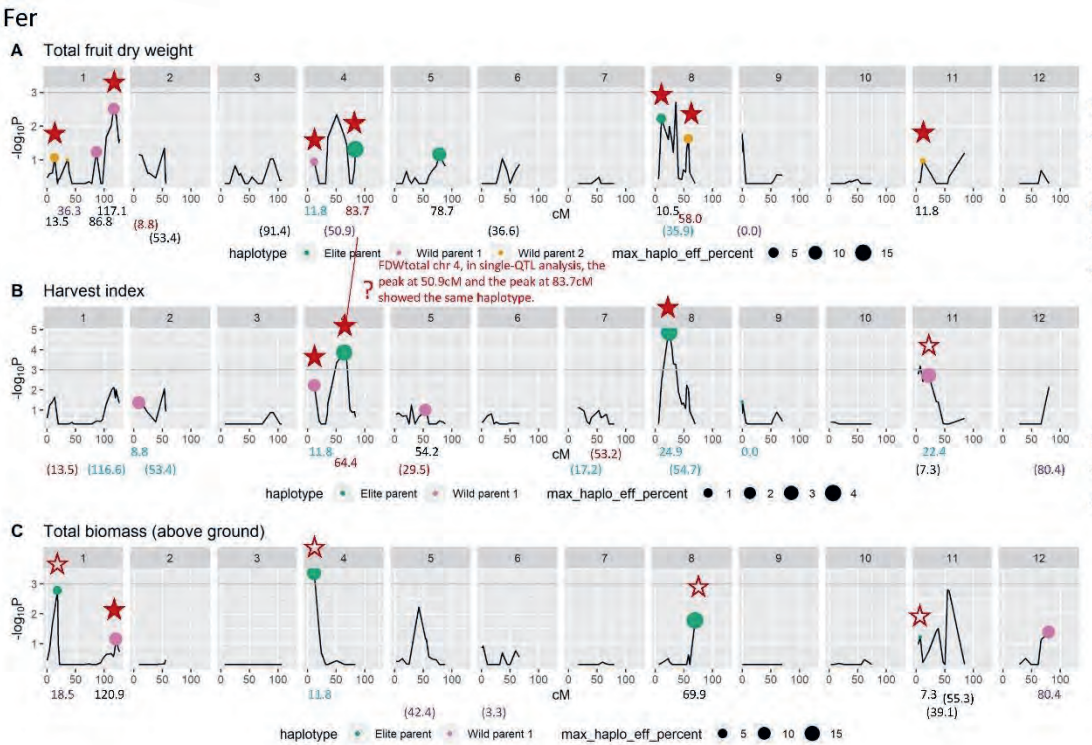
Common multi-QTL between A and C
- With the same haplotype: 2
- With different haplotype: 0

Common multi-QTL between B and C
- With the same haplotype: 1
- With different haplotype: 4

Supplementary Figure 7 Continued.



Supplementary Figure 7 Continued.

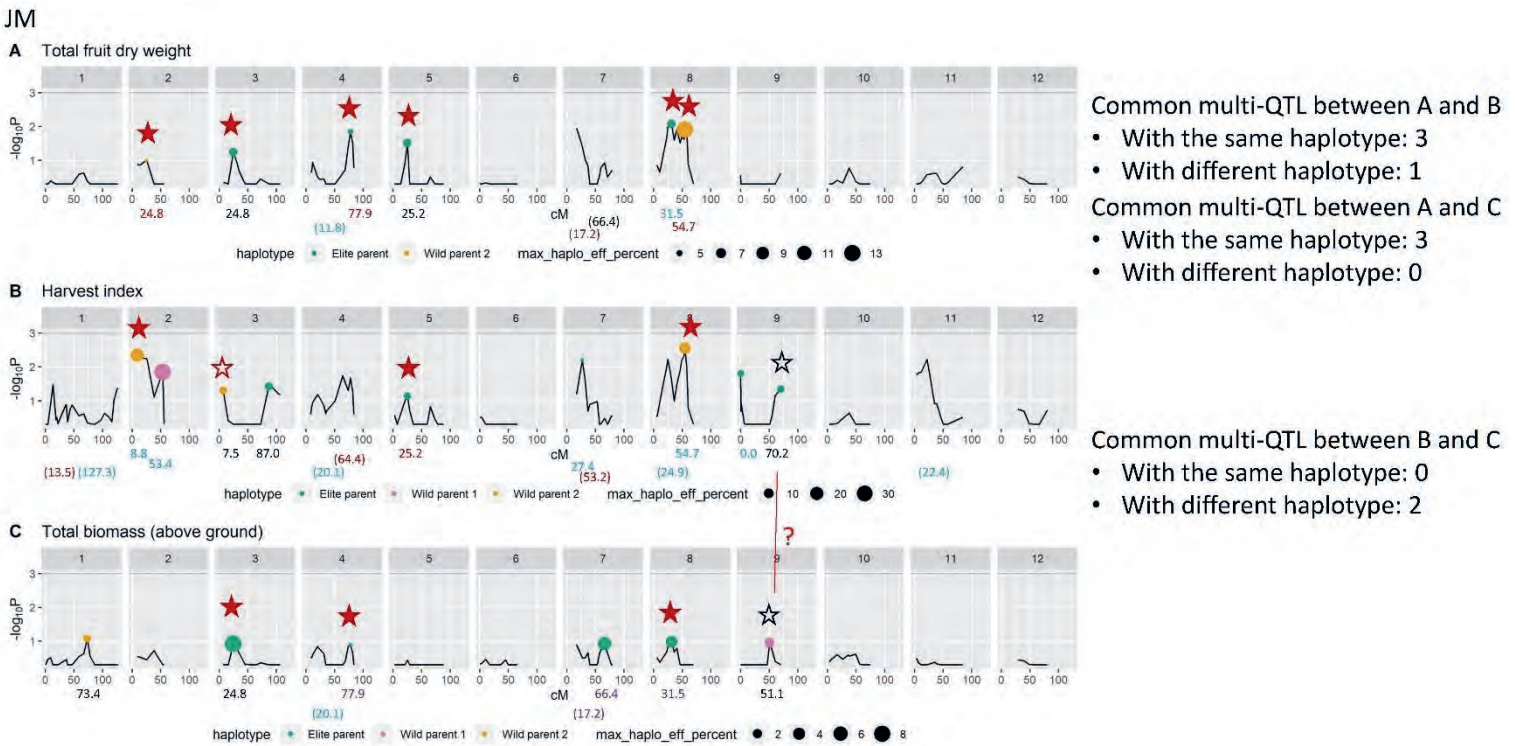


Common multi-QTL between A and B
- With the same haplotype: 3
- With different haplotype: 1

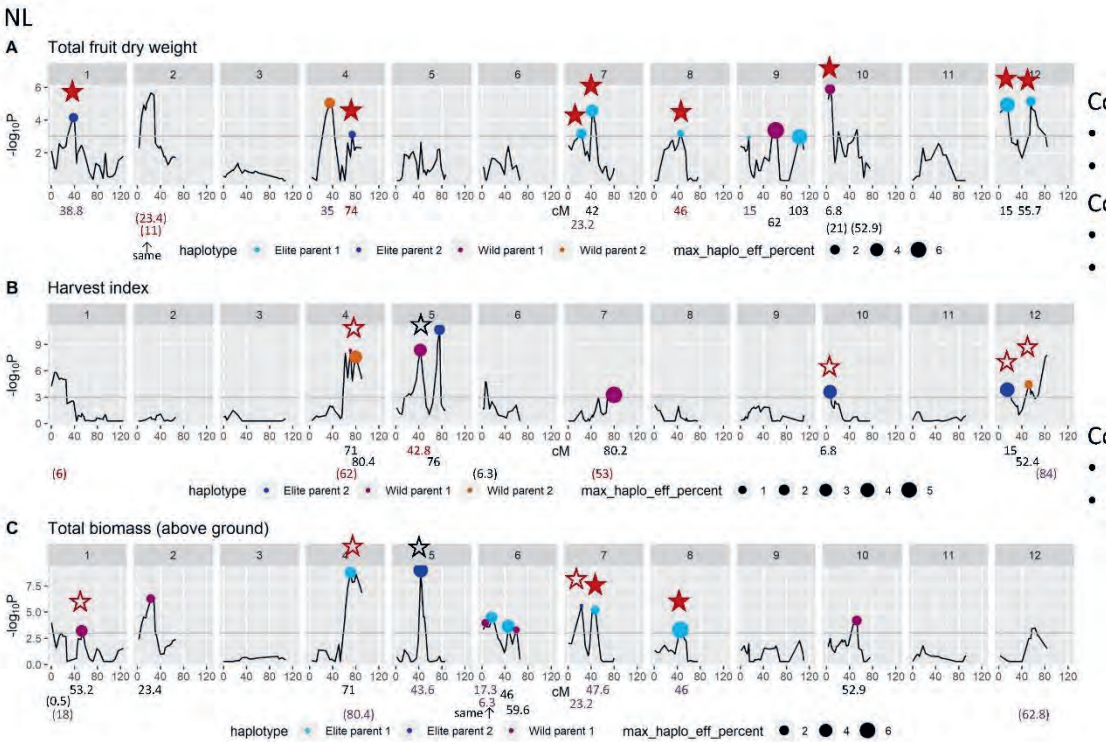
Common multi-QTL between A and C
- With the same haplotype: 1
- With different haplotype: 4

Common multi-QTL between B and C
- With the same haplotype: 0
- With different haplotype: 2

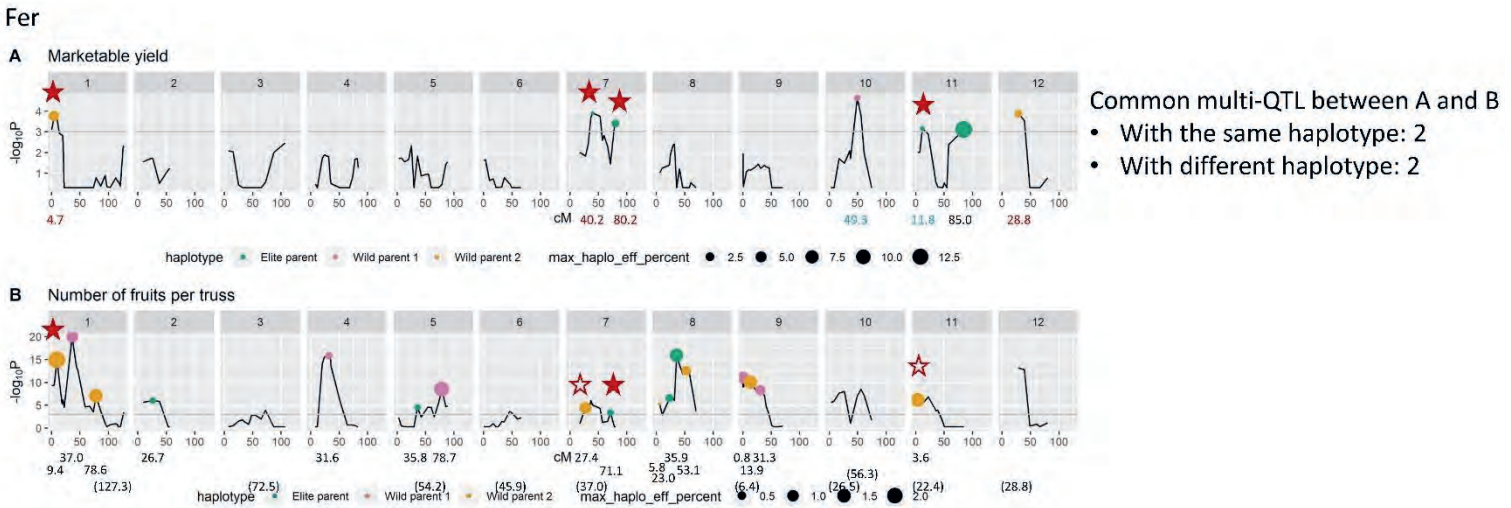
Supplementary Figure 7 Continued.



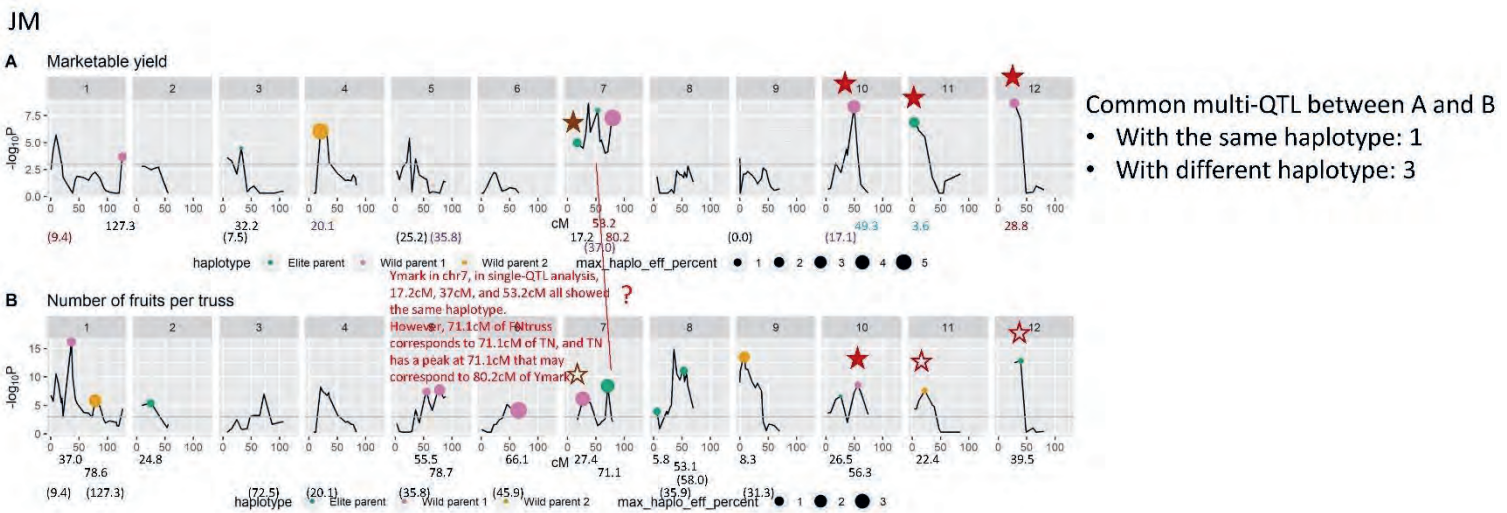
Supplementary Figure 7 Continued.



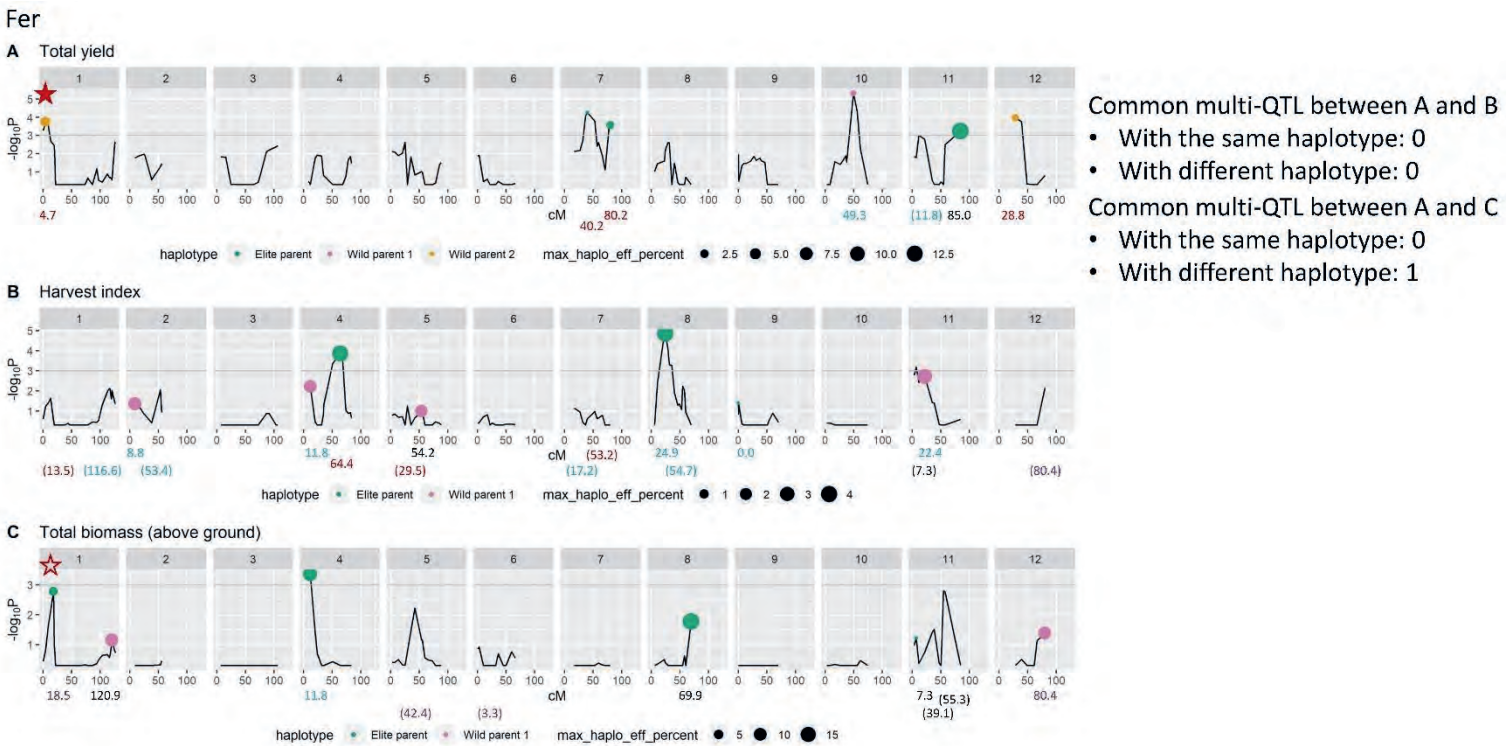
Supplementary Figure 7 Continued.



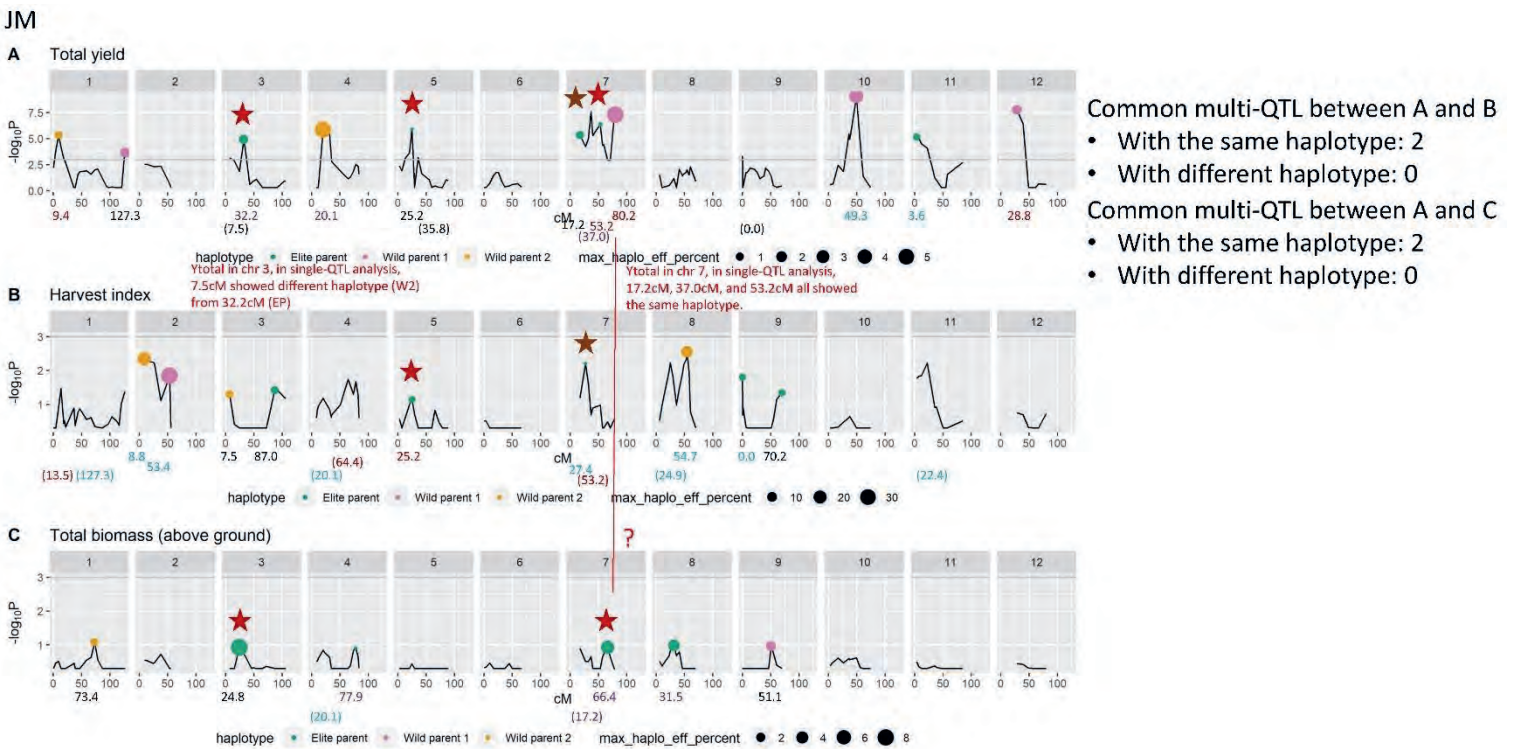
Supplementary Figure 7 Continued.



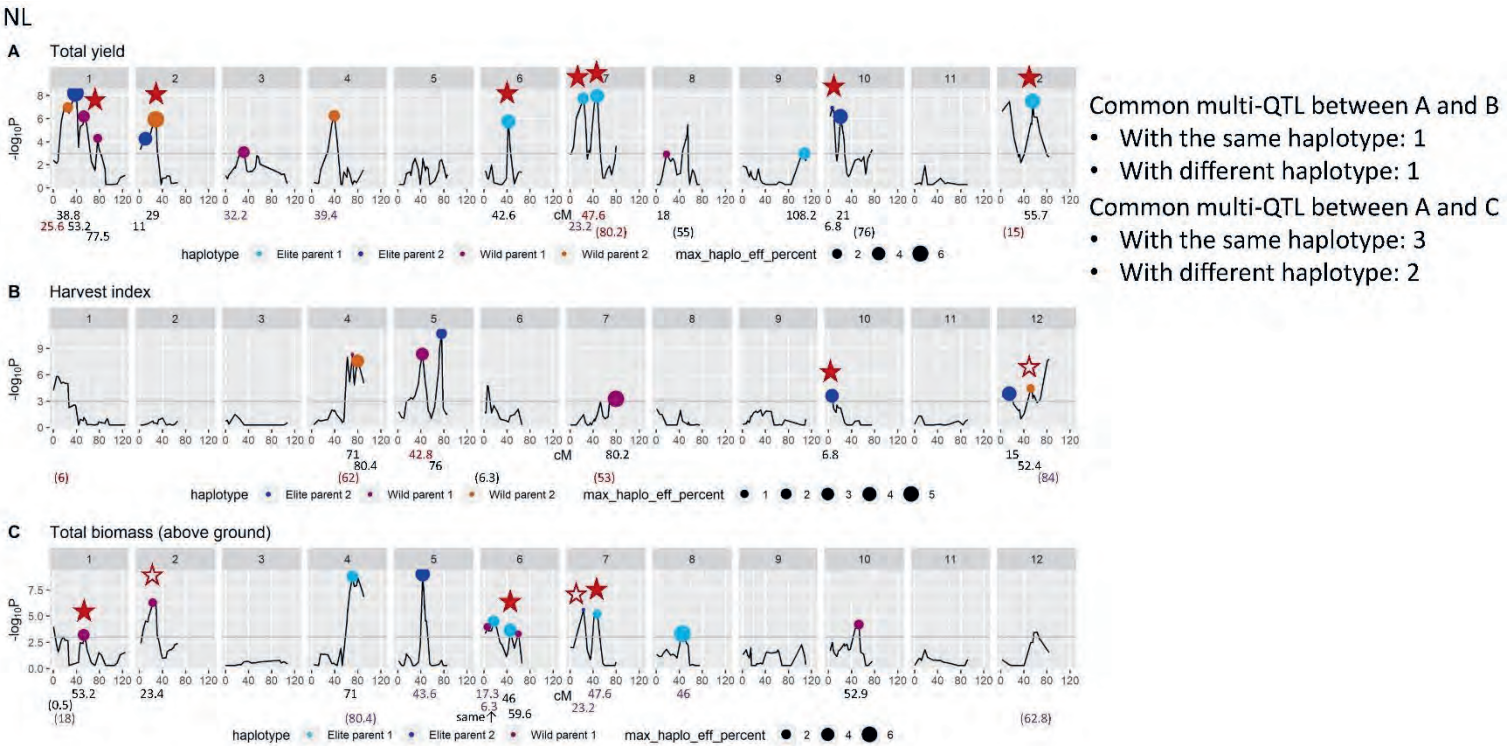
Supplementary Figure 7 Continued.



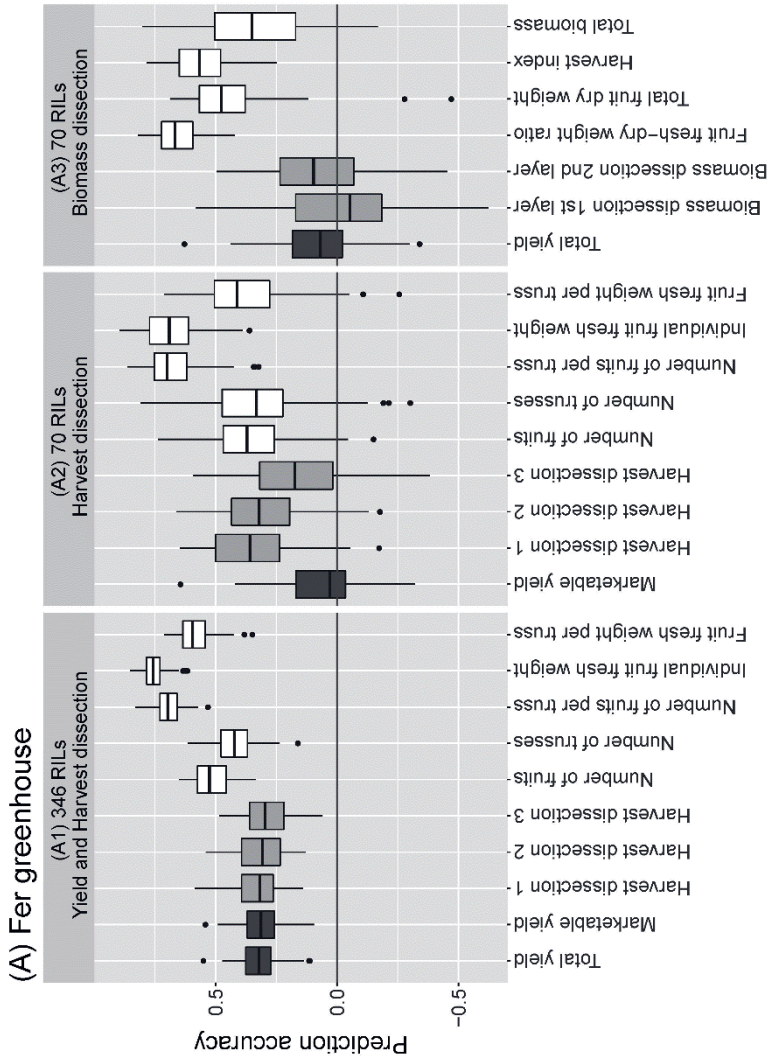
Supplementary Figure 7 Continued.



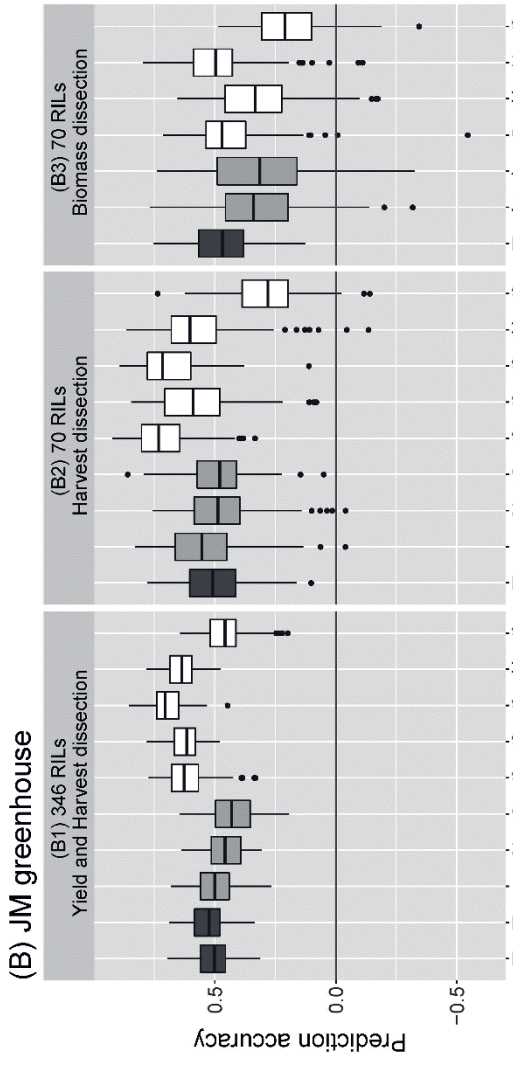
Supplementary Figure 7 Continued.



Supplementary Figure 7 Continued.



Supplementary Figure 8 Prediction accuracies in cross validation for yield and component traits, as well as for yield as predicted from yield dissection models. (A) Fer greenhouse. (B) JM greenhouse. (1) five-fold cross-validation by using 346 RILs for yield and component traits in the harvest dissection. (2) Cross validation by 50 RILs for training set and 20 RILs for validation set for component traits in the harvest dissection. (3) Cross validation by 50 RILs for training set and 20 RILs for validation set for component traits in the biomass dissection. Colours of boxes indicate the type of prediction: yield prediction by QTLs for yield itself (black), prediction of component traits by own QTLs (white), and yield prediction by yield dissection models using QTLs for component traits (grey).



Supplementary Figure 8 Continued.

Supplementary Table 1 Overview of phenotyping.

Abbreviation	Trait	Unit in analysis		Phenotyping	The number of RILs					
					Fer Block 1	Fer Block 2	Fer Total	JM Block 1	JM Block 2	JM Total
Ytotal	Total yield	kg stem^{-1}	Observed	All fruits per plot were measured biweekly.	346	346	346	346	346	346
Ymark	Marketable yield	kg stem^{-1}	Observed	All marketable fruits per plot were measured biweekly.	346	346	346	346	346	346
<Component traits in the yield dissection model based on harvest>										
FN	Number of fruits	stem^{-1}	Observed	All marketable fruits per plot at every harvest.	346	346	346	346	346	346
TN	Number of trusses	stem^{-1}	Observed	All trusses with marketable fruits per plot at every harvest.	346	346	346	346	346	346
FNtruss	Number of fruits per truss	truss^{-1}	Calculated	Calculated at every harvest (FN/TN). Weekly data of individual plot including the outlier of FFW or missing data of TN were removed. Mean of calculated values from all harvests (every two weeks) was calculated for each plot.	346	346	346	346	346	346
FFW	Individual fruit fresh weight	g fruit^{-1}	Calculated	Calculated at every harvest (FFW = Ymark / FN * 1000). After calculation, outliers (3x IQR) per plot were removed. Then, two plots that showed FFW>500g were also removed. Median of calculated values at every harvest (every two weeks) was calculated for each plot.	346	346	346	346	346	346
FFWtruss	Fruit fresh weight per truss	g truss^{-1}	Calculated	Calculated at every harvest (FFWtruss = Ymark / TN * 1000). After outliers of FFW and missing data of TN were removed, median of calculated values at every harvest (every two weeks) was calculated for each plot.	346	346	346	346	346	346
<Component traits in the yield dissection model based on biomass>										
Fresh_Dry_ratio	Fruit fresh-dry weight ratio	-	Observed	3 fruits (1 per plant) from 3 plants were sampled per plot. 3 fruits measured together.	40	39	70	40	40	70

FDWtotal	Total fruit dry weight	kg stem^{-1}	Calculated	FDWtotal = Ytotal/FDMCrcp	40	39	70	40	40	70
HI	Harvest index	-	Calculated	HI = FDWtotal / Biomass	39	39	69	40	40	70
Biomass	Total biomass (above ground)	kg stem^{-1}	Calculated	Biomass = FDWtotal + LDWtotal + StemDWtotal	39	39	69	40	40	70
<intermediate traits>										
LDWtotal	Total leaf dry weight	kg stem^{-1}	Calculated	LDWtotal = LDW_3leaves*TN Tomato has 3 leaves between adjacent trusses. Therefore, the leaves under the lowest truss were ignored.	39	39	69	40	40	70
LDW_3leaves	Leaf dry weight of 3 leaves	kg for 3 leaves	Observed	3 full-grown leaves measured together from 1 stem per plot.	39	39	69	40	40	70
StemDWtotal	Total stem dry weight	kg stem^{-1}	Calculated	StemDWtotal = StemDWseg* TN	39	39	69	40	40	70
StemDWseg	Stem dry weight per segment	kg segment^{-1}	Estimated	Estimated from stem volmume for a main stem per plot. Fer greenhouse: StemDWseg = 1.06858 * StemVolseg * 10^{-4} (Adjusted R-squared: 0.989) JM greenhouse: StemDWseg= 1.39058 * StemVolseg * 10^{-4} (Adjusted R-squared: 0.981) Regression model for each greenhouse was developted by using 20 axillary stems for Fer greenhouse and 18 axillary stems for JM greenhouse. StemD and StemL_seg were measured and StemVol_seg was calculated in the same way as explained below. StemDWseg was measured after drying samples at 105 °C for 2 days.	39	39	69	40	40	70
StemVolseg	Stem volume per segment	cm^3 segment^{-1}	Calculated	StemVolseg = (StemD/2)^2 * pi * StemLseg	39	39	69	40	40	70
StemD	Stem diameter	cm	Observed	Mean from 2 positions for 1 main stem per plot.	39	39	69	40	40	70
StemLseg	Stem length per segment	cm	Observed	Stem length for two segments (between truss n and truss n+2) was measured on 1 main stem per plot, and the value was devided by 2 to get the value per segment.	39	39	69	40	40	70

Supplementary Table 2 Yield and component-trait QTLs in Fer greenhouse in Spain. QTL positions, SolCAP marker names, -log$_{10}$(p) values in single-QTL analysis, and haplotype effects in multi-QTL model.

			Position (bp)				QTL effect from multi-QTL model		
Trait	Chromosome	cM	ver. 2.5	ver. 4.0	SolCAP.marker	- log$_{10}$P	Elite parent	Wild parent 1	Wild parent 2
Total yield (kg stem^{-1})									
	1	4.7	986,514	994,295	solcap_snp_sl_60417	3.74	-0.034	-0.271	0.305
	7	40.2	59,231,709	59,118,881	solcap_snp_sl_10686	4.25	0.026	-0.020	-0.006
	7	80.2	67,162,714	66,990,224	solcap_snp_sl_70636	3.58	0.143	0.062	-0.205
	10	49.3	61,842,466	60,952,055	solcap_snp_sl_33166	5.31	-0.121	0.069	0.051
	11	11.8	1,988,809	2,024,684	solcap_snp_sl_10611	3.01	NA	NA	NA
	11	85.0	55,196,665	53,263,567	SL1_00sc6004_2094360_solcap_snp_sl_44897	3.23	1.126	-0.003	-1.123
	12	28.8	3,768,296	3,806,963	CL009187-0469	3.97	-0.174	0.071	0.103
Marketable yield (kg stem^{-1})									
	1	4.7	986,514	994,295	solcap_snp_sl_60417	3.75	-0.045	-0.295	0.340
	7	40.2	59,231,709	59,118,881	solcap_snp_sl_10686	3.89	0.009	-0.007	-0.002
	7	80.2	67,162,714	66,990,224	solcap_snp_sl_70636	3.39	0.138	0.070	-0.207
	10	49.3	61,842,466	60,952,055	solcap_snp_sl_33166	4.63	-0.089	0.057	0.033
	11	11.8	1,988,809	2,024,684	solcap_snp_sl_10611	3.17	0.015	-0.004	-0.011
	11	85.0	55,196,665	53,263,567	SL1_00sc6004_2094360_solcap_snp_sl_44897	3.10	1.118	-0.044	-1.074
	12	28.8	3,768,296	3,806,963	CL009187-0469	3.87	-0.188	0.074	0.114
Number of fruits (stem^{-1})									
	1	9.4	1,511,380	1,518,201	solcap_snp_sl_60330	8.11	-0.34	-3.79	4.13

1	38.8	4,940,680	4,963,691	solcap_snp_sl_33676	9.33	-3.31	1.83	1.47
1	67.0	82,553,933	74,799,309	solcap_snp_sl_17063	4.79	NA	NA	NA
1	78.6	85,141,066	77,384,601	solcap_snp_sl_21283	7.62	-4.46	2.29	2.18
2	38.1	43,252,729	41,236,467	CL016172-0398_solcap_snp_sl_29710	11.56	NA	NA	NA
3	35.0	4,327,152	4,235,663	Psy1_250_b	5.29	-1.83	1.01	0.82
3	46.0	35,965,114	32,004,652	solcap_snp_sl_30590	6.24	-3.66	0.88	2.78
4	31.6	4,635,914	4,657,293	solcap_snp_sl_27161	4.55	-1.32	0.58	0.74
5	13.5	2,069,395	2,117,900	solcap_snp_sl_49234	5.47	NA	NA	NA
6	66.1	48,838,772	46,411,582	solcap_snp_sl_31676	5.27	-4.62	-0.65	5.27
8	10.5	1,703,791	1,711,177	solcap_snp_sl_56679	6.86	NA	NA	NA
8	29.3	54,296,108	52,392,793	solcap_snp_sl_13456	5.43	1.55	-1.55	0.00
8	40.6	57,576,758	55,676,173	solcap_snp_sl_29413	4.27	NA	NA	NA
8	54.7	60,405,484	58,517,969	solcap_snp_sl_64706	4.10	NA	NA	NA
9	0.0	411,890	413,568	solcap_snp_sl_17510	6.15	NA	NA	NA
10	26.5	3,991,802	3,783,084	solcap_snp_sl_34373	5.52	NA	NA	NA
10	56.3	62,638,578	61,745,624	solcap_snp_sl_59444	4.91	-0.05	0.04	0.01
10	74.7	64,978,978	64,130,104	solcap_snp_sl_8808	3.78	NA	NA	NA
12	28.8	3,768,296	3,806,963	CL009187-0469	8.75	2.46	-1.33	-1.13
12	66.7	64,671,315	64,156,637	solcap_snp_sl_62972	2.93	NA	NA	NA
Number of trusses (stem^{-1})								
1	0.5	310,894	324,884	CL004303-0524	3.33	0.064	-0.079	0.016
1	103.6	92,043,676	84,331,398	solcap_snp_sl_40199	3.70	NA	NA	NA
1	119.9	96,037,443	88,382,128	C2_At2g15890_430_b	3.20	0.104	-0.041	-0.063

2	38.1	43,252,729	41,236,467	CL016172-0398_solcap_snp_sl_29710	3.62	0.109	0.458	-0.567
3	46.0	35,965,114	32,004,652	solcap_snp_sl_30590	5.04	-0.498	0.337	0.161
5	13.5	2,069,395	2,117,900	solcap_snp_sl_49234	5.86	0.077	-0.069	-0.008
6	31.9	40,659,989	38,300,758	solcap_snp_sl_12771	6.09	NA	NA	NA
6	45.9	43,804,153	41,468,982	solcap_snp_sl_57810	8.90	NA	NA	NA
6	66.1	48,838,772	46,411,582	solcap_snp_sl_31676	10.63	-0.466	-0.334	0.800
7	71.1	65,597,754	65,439,722	solcap_snp_sl_12101	5.41	-0.044	0.053	-0.009
8	10.5	1,703,791	1,711,177	solcap_snp_sl_56679	2.95	NA	NA	NA
8	69.9	64,137,467	62,259,249	solcap_snp_sl_10181	6.54	NA	NA	NA
9	0.8	748,255	767,180	solcap_snp_sl_14676	4.81	NA	NA	NA
9	8.3	4,706,233	4,738,076	CL016894-0326_solcap_snp_sl_39506	5.39	0.038	-0.028	-0.010
9	31.3	64,905,399	61,002,505	solcap_snp_sl_42973	5.26	0.283	-0.195	-0.088
10	74.7	64,978,978	64,130,104	solcap_snp_sl_8808	7.60	NA	NA	NA
12	66.7	64,671,315	64,156,637	solcap_snp_sl_62972	5.60	NA	NA	NA
Number of fruits per truss								
1	9.4	1,511,380	1,518,201	solcap_snp_sl_60330	14.88	-0.041	-0.153	0.194
1	37.0	4,597,945	4,614,026	solcap_snp_sl_59890	19.95	-0.154	0.082	0.072
1	78.6	85,141,066	77,384,601	solcap_snp_sl_21283	6.97	-0.123	0.002	0.121
1	127.3	98,090,338	90,442,389	CL015806-0417_solcap_snp_sl_53692	3.47	NA	NA	NA

2	26.7	40,022,938	38,037,260	solcap_snp_sl_36037	6.00	0.014	-0.008	-0.006
3	72.5	63,163,824	57,590,779	CL015346-0143_solcap_snp_sl_62495	3.89	NA	NA	NA
4	31.6	4,635,914	4,657,293	solcap_snp_sl_27161	15.83	-0.043	0.022	0.021
5	35.8	4,891,796	4,941,398	solcap_snp_sl_48814	4.52	0.016	-0.009	-0.007
5	54.2	61,442,463	60,839,876	solcap_snp_sl_37183	4.57	NA	NA	NA
5	78.7	64,086,547	63,470,159	solcap_snp_sl_224	8.47	-0.057	0.155	-0.099
6	45.9	43,804,153	41,468,982	solcap_snp_sl_57810	3.61	NA	NA	NA
7	27.4	3,747,320	3,720,215	solcap_snp_sl_11067	4.43	-0.153	0.068	0.085
7	37.0	57,927,159	57,809,287	solcap_snp_sl_51838	5.96	NA	NA	NA
7	71.1	65,597,754	65,439,722	solcap_snp_sl_12101	3.37	0.015	-0.017	0.002
8	5.8	985,676	998,051	solcap_snp_sl_7326	5.23	-0.011	0.004	0.007
8	23.0	3,350,101	3,272,192	solcap_snp_sl_5426	6.56	0.028	-0.038	0.010
8	35.9	56,243,052	54,345,666	solcap_snp_sl_18196	15.88	0.112	-0.049	-0.062
8	53.1	60,063,143	58,175,394	solcap_snp_sl_64642	12.51	-0.067	0.016	0.051
9	0.8	748,255	767,180	solcap_snp_sl_14676	11.02	-0.008	0.081	-0.073
9	6.4	4,561,319	4,595,285	SGN-U329408_snp291_solcap_snp_sl_39516	10.68	NA	NA	NA
9	13.9	5,191,470	5,233,823	solcap_snp_sl_39407	9.98	-0.105	-0.014	0.118
9	31.3	64,905,399	61,002,505	solcap_snp_sl_42973	8.26	-0.065	0.062	0.003
10	26.5	3,991,802	3,783,084	solcap_snp_sl_34373	8.03	NA	NA	NA
10	56.3	62,638,578	61,745,624	solcap_snp_sl_59444	8.53	NA	NA	NA
11	3.6	941,262	982,796	solcap_snp_sl_66528	6.08	-0.113	-0.017	0.130

11	22.4	3,158,623	3,188,167	solcap_snp_sl_21767	6.82	NA	NA	NA
12	28.8	3,768,296	3,806,963	CL009187-0469	13.13	NA	NA	NA
Individual fruit fresh weight (g fruit^{-1})								
1	4.7	986,514	994,295	solcap_snp_sl_60417	10.14	0.19	-0.11	-0.08
1	37.0	4,597,945	4,614,026	solcap_snp_sl_59890	13.71	0.90	-0.67	-0.23
1	58.8	80,542,671	73,127,585	solcap_snp_sl_38356	12.72	1.53	-0.62	-0.92
1	127.3	98,090,338	90,442,389	CL015806-0417_solcap_snp_sl_53692	9.53	NA	NA	NA
2	38.1	43,252,729	41,236,467	CL016172-0398_solcap_snp_sl_29710	28.13	NA	NA	NA
3	7.5	1,390,288	1,427,470	solcap_snp_sl_63290	11.74	-0.44	0.89	-0.45
3	35.0	4,327,152	4,235,663	Psy1_250_b	7.08	1.49	-0.74	-0.75
3	47.9	52,628,897	47,197,611	solcap_snp_sl_30678	9.60	1.64	-3.27	1.63
4	8.3	1,242,445	1,277,483	solcap_snp_sl_63899	4.20	-0.23	0.32	-0.10
4	31.6	4,635,914	4,657,293	solcap_snp_sl_27161	5.82	1.69	-1.64	-0.05
5	35.8	4,891,796	4,941,398	solcap_snp_sl_48814	14.82	NA	NA	NA
5	88.9	65,747,765	65,139,785	solcap_snp_sl_37971	11.16	NA	NA	NA
6	10.2	35,008,382	32,780,157	solcap_snp_sl_44686	5.74	0.68	-1.48	0.79
7	37.0	57,927,159	57,809,287	solcap_snp_sl_51838	7.45	NA	NA	NA
7	53.2	62,093,144	62,018,573	solcap_snp_sl_53425	9.86	NA	NA	NA
7	59.8	63,644,286	63,561,908	solcap_snp_sl_55514	8.01	2.59	1.15	-3.74
8	5.8	985,676	998,051	solcap_snp_sl_7326	7.53	1.18	-0.89	-0.29
8	35.9	56,243,052	54,345,666	solcap_snp_sl_18196	6.90	NA	NA	NA

8	54.7	60,405,484	58,517,969	solcap_snp_sl_64706	14.85	0.34	-0.01	-0.33
8	60.5	62,345,855	63,310,880	CL016130-0295	13.88	1.16	-0.02	-1.15
9	0.0	411,890	413,568	solcap_snp_sl_17510	22.45	-0.18	-1.00	1.17
9	21.6	7,377,236	7,399,225	solcap_snp_sl_45148	3.78	NA	NA	NA
10	26.5	3,991,802	3,783,084	solcap_snp_sl_34373	13.11	1.51	-1.52	0.01
10	61.5	63,163,119	62,325,614	solcap_snp_sl_61192	7.90	NA	NA	NA
11	11.8	1,988,809	2,024,684	solcap_snp_sl_10611	12.46	1.47	0.53	-2.01
12	28.8	3,768,296	3,806,963	CL009187-0469	45.74	-5.55	2.59	2.96
Fruit fresh weight per truss (g truss^{-1})								
1	38.8	4,940,680	4,963,691	solcap_snp_sl_33676	3.22	NA	NA	NA
1	58.8	80,542,671	73,127,585	solcap_snp_sl_38356	5.19	4.2	-4.8	0.6
1	86.8	87,334,152	79,595,123	solcap_snp_sl_28173	3.53	NA	NA	NA
1	112.8	94,361,361	86,668,843	3284_1_60_solcap_snp_sl_54547	4.35	5.7	-1.8	-3.9
1	127.3	98,090,338	90,442,389	CL015806-0417_solcap_snp_sl_53692	5.40	-15.5	2.0	13.5
2	38.1	43,252,729	41,236,467	CL016172-0398_solcap_snp_sl_29710	17.02	NA	NA	NA
3	7.5	1,390,288	1,427,470	solcap_snp_sl_63290	12.88	NA	NA	NA
3	47.9	52,628,897	47,197,611	solcap_snp_sl_30678	6.12	7.5	-23.8	16.3
4	8.3	1,242,445	1,277,483	solcap_snp_sl_63899	3.15	NA	NA	NA
5	13.5	2,069,395	2,117,900	solcap_snp_sl_49234	3.56	-1.7	1.0	0.8
5	35.8	4,891,796	4,941,398	solcap_snp_sl_48814	7.11	NA	NA	NA

5	88.9	65,747,765	65,139,785	solcap_snp_sl_37971	7.67	-1.1	1.2	-0.1
6	10.2	35,008,382	32,780,157	solcap_snp_sl_44686	3.23	5.2	-3.5	-1.7
6	45.9	43,804,153	41,468,982	solcap_snp_sl_57810	5.00	NA	NA	NA
7	17.2	2,131,288	2,103,765	solcap_snp_sl_22102	3.72	NA	NA	NA
7	59.8	63,644,286	63,561,908	solcap_snp_sl_55514	9.84	20.0	-4.3	-15.7
8	10.5	1,703,791	1,711,177	solcap_snp_sl_56679	3.64	NA	NA	NA
8	54.7	60,405,484	58,517,969	solcap_snp_sl_64706	5.34	NA	NA	NA
8	60.5	62,345,855	63,310,880	CL016130-0295	4.20	NA	NA	NA
9	0.6	651,775	650,804	solcap_snp_sl_28404	14.51	-0.6	0.2	0.4
10	26.5	3,991,802	3,783,084	solcap_snp_sl_34373	5.98	4.9	-2.3	-2.6
10	74.7	64,978,978	64,130,104	solcap_snp_sl_8808	4.95	4.4	-3.3	-1.1
11	11.8	1,988,809	2,024,684	solcap_snp_sl_10611	3.44	7.0	-1.0	-6.1
11	85.0	55,196,665	53,263,567	SL1_00sc6004_2094360_solcap_snp_sl_44897	3.20	108.8	-88.8	-20.0
12	28.8	3,768,296	3,806,963	CL009187-0469	20.57	-21.3	10.4	10.9
12	66.7	64,671,315	64,156,637	solcap_snp_sl_62972	4.39	2.7	-3.3	0.6
Fruit fresh-dry weight ratio								
1	18.5	2,063,522	2,014,003	solcap_snp_sl_33701	2.16	-0.27	0.22	0.06
1	36.3	4,404,784	4,418,255	solcap_snp_sl_59897	1.27	NA	NA	NA
1	116.6	95,122,270	87,430,927	CL009293-0681	1.57	-0.38	-1.44	1.82
2	8.8	36,167,998	34,191,377	solcap_snp_sl_15706	1.56	1.41	-1.16	-0.26
2	53.4	47,123,587	45,124,478	solcap_snp_sl_49965	1.41	NA	NA	NA
3	54.5	57,784,509	52,357,459	solcap_snp_sl_35369	2.81	2.50	-3.44	0.94
4	11.8	1,562,490	1,594,008	solcap_snp_sl_63976	1.68	NA	NA	NA
4	50.9	56,086,497	55,434,838	solcap_snp_sl_3111	1.71	NA	NA	NA

5	42.4	6,127,298	6,207,099	solcap_snp_sl_50851	2.23	NA	NA	NA
5	72.9	63,379,877	62,769,283	solcap_snp_sl_12279	0.99	-1.68	1.00	0.68
6	31.9	40,659,989	38,300,758	solcap_snp_sl_12771	1.50	0.52	-0.14	-0.38
6	59.9	47,400,898	45,081,474	solcap_snp_sl_19851	1.34	NA	NA	NA
8	35.9	56,243,052	54,345,666	solcap_snp_sl_18196	1.23	1.19	0.58	-1.77
8	58.0	61,713,636	59,822,513	solcap_snp_sl_34763	1.86	0.23	-0.10	-0.13
9	0.0	411,890	413,568	solcap_snp_sl_17510	2.64	NA	NA	NA
9	0.9	1,514,198	1,533,785	solcap_snp_sl_7731	2.69	NA	NA	NA
9	70.2	71,185,917	67,281,966	solcap_snp_sl_25745	1.42	NA	NA	NA
10	61.5	63,163,119	62,325,614	solcap_snp_sl_61192	1.24	NA	NA	NA
11	22.4	3,158,623	3,188,167	solcap_snp_sl_21767	1.21	NA	NA	NA
11	47.6	7,926,326	8,024,502	solcap_snp_sl_6901	2.05	-5.95	-0.63	6.58
11	58.0	52,092,467	50,122,139	solcap_snp_sl_5992	2.78	-2.57	-1.77	4.33
12	39.5	9,972,015	10,060,803	solcap_snp_sl_40606	1.06	-1.31	0.66	0.66
Total fruit dry weight (kg stem^{-1})								
1	13.5	1,809,372	1,814,874	solcap_snp_sl_60276	1.06	-0.0119	-0.0027	0.0147
1	36.3	4,404,784	4,418,255	solcap_snp_sl_59897	1.00	-0.0005	-0.0023	0.0028
1	86.8	87,334,152	79,595,123	solcap_snp_sl_28173	1.23	-0.0104	0.0265	-0.0161
1	117.1	95,218,539	87,527,104	solcap_snp_sl_54732	2.51	0.0041	0.0276	-0.0317
2	8.8	36,167,998	34,191,377	solcap_snp_sl_15706	1.15	NA	NA	NA
2	53.4	47,123,587	45,124,478	solcap_snp_sl_49965	1.35	NA	NA	NA
3	91.4	67,103,991	61,553,883	solcap_snp_sl_20721	1.03	NA	NA	NA
4	11.8	1,562,490	1,594,008	solcap_snp_sl_63976	0.94	-0.0064	0.0107	-0.0043
4	50.9	56,086,497	55,434,838	solcap_snp_sl_3111	2.34	NA	NA	NA
4	83.7	64,360,786	62,336,912	solcap_snp_sl_47487	1.30	0.0614	-0.0610	-0.0004

	5	78.7	64,086,547	63,470,159	solcap_snp_sl_224	1.16	0.0396	-0.0401	0.0005
	6	36.6	41,710,575	39,348,124	SGN-U316705_snp691	1.04	NA	NA	NA
	8	10.5	1,703,791	1,711,177	solcap_snp_sl_56679	2.22	0.0163	-0.0135	-0.0028
	8	35.9	56,243,052	54,345,666	solcap_snp_sl_18196	2.70	NA	NA	NA
	8	58.0	61,713,636	59,822,513	solcap_snp_sl_34763	1.62	-0.0124	-0.0034	0.0158
	9	0.0	411,890	413,568	solcap_snp_sl_17510	1.78	NA	NA	NA
	11	11.8	1,988,809	2,024,684	solcap_snp_sl_10611	0.97	-0.0064	0.0019	0.0045
Harvest index									
	1	13.5	1,809,372	1,814,874	solcap_snp_sl_60276	1.62	NA	NA	NA
	1	116.6	95,122,270	87,430,927	CL009293-0681	2.12	NA	NA	NA
	2	8.8	36,167,998	34,191,377	solcap_snp_sl_15706	1.36	-0.009	0.010	-0.001
	2	53.4	47,123,587	45,124,478	solcap_snp_sl_49965	2.04	NA	NA	NA
	4	11.8	1,562,490	1,594,008	solcap_snp_sl_63976	2.22	-0.015	0.010	0.006
	4	64.4	60,772,015	58,731,439	solcap_snp_sl_22225	3.85	0.017	0.001	-0.019
	5	29.5	3,980,854	4,029,867	solcap_snp_sl_23734	1.25	NA	NA	NA
	5	54.2	61,442,463	60,839,876	solcap_snp_sl_37183	0.98	-0.009	0.009	0.000
	7	17.2	2,131,288	2,103,765	solcap_snp_sl_22102	1.14	NA	NA	NA
	7	53.2	62,093,144	62,018,573	solcap_snp_sl_53425	0.97	NA	NA	NA
	8	24.9	3,781,365	3,700,890	solcap_snp_sl_51919	4.84	0.018	-0.019	0.001
	8	54.7	60,405,484	58,517,969	solcap_snp_sl_64706	2.23	NA	NA	NA
	9	0.0	411,890	413,568	solcap_snp_sl_17510	1.39	0.001	-0.001	0.000
	11	7.3	1,347,455	1,386,107	solcap_snp_sl_66431	3.18	NA	NA	NA
	11	22.4	3,158,623	3,188,167	solcap_snp_sl_21767	2.72	-0.026	0.014	0.012
	12	80.4	66,256,781	65,786,764	SGN-U603632_snp167	2.15	NA	NA	NA
Total biomass (above ground) (kg stem^{-1})									

1	18.5	2,063,522	2,014,003	solcap_snp_sl_33701	2.77	0.035	-0.039	0.004
1	120.9	96,218,696	88,563,470	solcap_snp_sl_13402	1.14	-0.026	0.092	-0.065
4	11.8	1,562,490	1,594,008	solcap_snp_sl_63976	3.35	0.108	0.080	-0.188
5	42.4	6,127,298	6,207,099	solcap_snp_sl_50851	2.21	NA	NA	NA
6	3.3	1,579,723	1,595,537	solcap_snp_sl_15516	0.91	NA	NA	NA
8	69.9	64,137,467	62,259,249	solcap_snp_sl_10181	1.77	0.159	-0.046	-0.113
11	7.3	1,347,455	1,386,107	solcap_snp_sl_66431	1.21	0.009	-0.006	-0.003
11	39.1	5,354,992	5,411,068	solcap_snp_sl_9507	1.50	NA	NA	NA
11	55.3	51,523,371	49,571,263	solcap_snp_sl_6044	2.80	NA	NA	NA
12	80.4	66,256,781	65,786,764	SGN-U603632_snp167	1.37	-0.083	0.090	-0.007

Supplementary Table 3 Yield and component-trait QTLs in JM greenhouse in Spain. QTL positions, SolCAP marker names, -log$_{10}$(p) values in single-QTL analysis, and haplotype effects in multi-QTL model.

			Position (bp)				QTL effect from multi-QTL model		
Trait	Chromosome	cM	ver. 2.5	ver. 4.0	SolCAP.marker	- log$_{10}$P	Elite parent	Wild parent 1	Wild parent 2
Total yield (kg stem^{-1})									
	1	9.4	1,511,380	1,518,201	solcap_snp_sl_60330	5.31	0.024	-0.114	0.090
	1	127.3	98,090,338	90,442,389	CL015806-0417_solcap_snp_sl_53692	3.63	-0.193	0.145	0.048
	3	7.5	1,390,288	1,427,470	solcap_snp_sl_63290	3.13	NA	NA	NA
	3	32.2	3,607,939	3,592,814	solcap_snp_sl_19508	4.89	0.124	-0.037	-0.087
	4	20.1	2,923,218	2,959,057	solcap_snp_sl_21374	5.83	-0.241	-0.231	0.472
	5	25.2	3,416,161	3,463,230	CL017464-0100	5.88	0.023	-0.020	-0.003
	5	35.8	4,891,796	4,941,398	solcap_snp_sl_48814	3.16	NA	NA	NA
	7	17.2	2,131,288	2,103,765	solcap_snp_sl_22102	5.33	0.078	-0.049	-0.029
	7	37.0	57,927,159	57,809,287	solcap_snp_sl_51838	7.56	NA	NA	NA
	7	53.2	62,093,144	62,018,573	solcap_snp_sl_53425	6.35	0.025	-0.014	-0.011
	7	80.2	67,162,714	66,990,224	solcap_snp_sl_70636	7.28	0.334	0.502	-0.836
	9	0.0	411,890	413,568	solcap_snp_sl_17510	3.33	NA	NA	NA
	10	49.3	61,842,466	60,952,055	solcap_snp_sl_33166	9.05	-0.169	0.332	-0.163
	11	3.6	941,262	982,796	solcap_snp_sl_66528	5.18	0.064	-0.042	-0.022
	12	28.8	3,768,296	3,806,963	CL009187-0469	7.75	-0.242	0.132	0.110
Marketable yield (kg stem^{-1})									
	1	9.4	1,511,380	1,518,201	solcap_snp_sl_60330	5.69	NA	NA	NA
	1	127.3	98,090,338	90,442,389	CL015806-0417_solcap_snp_sl_53692	3.66	-0.173	0.131	0.042
	3	7.5	1,390,288	1,427,470	solcap_snp_sl_63290	3.60	NA	NA	NA

3	32.2	3,607,939	3,592,814	solcap_snp_sl_19508	4.48	0.039	-0.019	-0.020
4	20.1	2,923,218	2,959,057	solcap_snp_sl_21374	6.05	-0.207	-0.288	0.494
5	25.2	3,416,161	3,463,230	CL017464-0100	5.39	NA	NA	NA
5	35.8	4,891,796	4,941,398	solcap_snp_sl_48814	3.49	NA	NA	NA
7	17.2	2,131,288	2,103,765	solcap_snp_sl_22102	4.94	0.128	-0.079	-0.049
7	37.0	57,927,159	57,809,287	solcap_snp_sl_51838	8.64	NA	NA	NA
7	53.2	62,093,144	62,018,573	solcap_snp_sl_53425	7.96	0.046	-0.026	-0.020
7	80.2	67,162,714	66,990,224	solcap_snp_sl_70636	7.26	0.377	0.512	-0.889
9	0.0	411,890	413,568	solcap_snp_sl_17510	3.56	NA	NA	NA
10	17.1	2,686,835	2,563,006	CL015957-0328	3.07	NA	NA	NA
10	49.3	61,842,466	60,952,055	solcap_snp_sl_33166	8.32	-0.120	0.306	-0.185
11	3.6	941,262	982,796	solcap_snp_sl_66528	6.84	0.176	-0.112	-0.065
12	28.8	3,768,296	3,806,963	CL009187-0469	8.63	-0.279	0.162	0.117
Number of fruits (stem^{-1})								
1	9.4	1,511,380	1,518,201	solcap_snp_sl_60330	11.72	NA	NA	NA
1	38.8	4,940,680	4,963,691	solcap_snp_sl_33676	4.47	NA	NA	NA
1	78.6	85,141,066	77,384,601	solcap_snp_sl_21283	15.83	-4.67	1.18	3.50
2	13.8	37,687,512	35,703,901	solcap_snp_sl_14946	9.13	2.97	-1.04	-1.94
3	46.0	35,965,114	32,004,652	solcap_snp_sl_30590	4.14	-2.28	3.82	-1.54
4	31.6	4,635,914	4,657,293	solcap_snp_sl_27161	17.94	-5.11	-3.91	9.01
5	13.5	2,069,395	2,117,900	solcap_snp_sl_49234	12.32	2.37	-1.78	-0.59
5	72.9	63,379,877	62,769,283	solcap_snp_sl_12279	5.37	NA	NA	NA
6	3.3	1,579,723	1,595,537	solcap_snp_sl_15516	3.18	-1.61	1.51	0.11
7	37.0	57,927,159	57,809,287	solcap_snp_sl_51838	3.35	NA	NA	NA
8	40.6	57,576,758	55,676,173	solcap_snp_sl_29413	7.18	1.48	-0.99	-0.49

	8	60.5	62,345,855	63,310,880	CL016130-0295	2.96	NA	NA	NA
	9	0.0	411,890	413,568	solcap_snp_sl_17510	8.70	-1.31	0.76	0.55
	9	60.7	70,334,215	66,429,240	solcap_snp_sl_69669	6.62	NA	NA	NA
	10	56.3	62,638,578	61,745,624	solcap_snp_sl_59444	9.88	-0.55	0.18	0.37
	11	3.6	941,262	982,796	solcap_snp_sl_66528	3.50	1.47	-1.37	-0.10
	11	22.4	3,158,623	3,188,167	solcap_snp_sl_21767	3.24	NA	NA	NA
	12	39.5	9,972,015	10,060,803	solcap_snp_sl_40606	4.34	NA	NA	NA
Number of trusses (stem^{-1})									
	1	9.4	1,511,380	1,518,201	solcap_snp_sl_60330	11.39	NA	NA	NA
	1	36.3	4,404,784	4,418,255	solcap_snp_sl_59897	3.14	0.095	-0.040	-0.055
	1	78.6	85,141,066	77,384,601	solcap_snp_sl_21283	4.35	NA	NA	NA
	2	13.8	37,687,512	35,703,901	solcap_snp_sl_14946	3.53	0.253	-0.248	-0.005
	3	46.0	35,965,114	32,004,652	solcap_snp_sl_30590	6.10	-0.300	0.286	0.013
	3	72.5	63,163,824	57,590,779	CL015346-0143_solcap_snp_sl_62495	3.60	NA	NA	NA
	4	20.1	2,923,218	2,959,057	solcap_snp_sl_21374	16.82	-0.400	-0.516	0.916
	5	2.7	718,884	764,076	solcap_snp_sl_15889	15.91	0.201	-0.454	0.253
	5	54.2	61,442,463	60,839,876	solcap_snp_sl_37183	9.84	0.153	-0.143	-0.010
	5	72.9	63,379,877	62,769,283	solcap_snp_sl_12279	2.92	NA	NA	NA
	6	3.3	1,579,723	1,595,537	solcap_snp_sl_15516	4.86	-0.469	0.022	0.447
	6	45.9	43,804,153	41,468,982	solcap_snp_sl_57810	5.56	NA	NA	NA
	6	66.2	48,691,982	46,264,764	solcap_snp_sl_24255	4.62	-0.093	-0.037	0.129
	7	37.0	57,927,159	57,809,287	solcap_snp_sl_51838	11.52	0.224	-0.193	-0.031
	7	71.1	65,597,754	65,439,722	solcap_snp_sl_12101	4.16	-0.134	0.098	0.035
	8	40.6	57,576,758	55,676,173	solcap_snp_sl_29413	8.65	NA	NA	NA
	8	60.5	62,345,855	63,310,880	CL016130-0295	3.11	NA	NA	NA

9	6.4	4,561,319	4,595,285	SGN-U329408_snp291_solcap_snp_sl_39516	15.21	0.315	-0.297	-0.018
9	31.3	64,905,399	61,002,505	solcap_snp_sl_42973	10.82	NA	NA	NA
9	60.7	70,334,215	66,429,240	solcap_snp_sl_69669	3.08	-0.075	0.029	0.046
10	56.3	62,638,578	61,745,624	solcap_snp_sl_59444	9.04	NA	NA	NA
11	3.6	941,262	982,796	solcap_snp_sl_66528	11.61	0.090	0.056	-0.146
11	22.4	3,158,623	3,188,167	solcap_snp_sl_21767	12.78	0.275	-0.111	-0.164
12	80.4	66,256,781	65,786,764	SGN-U603632_snp167	2.92	NA	NA	NA
Number of fruits per truss								
1	9.4	1,511,380	1,518,201	solcap_snp_sl_60330	10.51	NA	NA	NA
1	37.0	4,597,945	4,614,026	solcap_snp_sl_59890	16.05	-0.141	0.074	0.067
1	78.6	85,141,066	77,384,601	solcap_snp_sl_21283	5.79	-0.218	0.036	0.182
1	127.3	98,090,338	90,442,389	CL015806-0417_solcap_snp_sl_53692	4.44	NA	NA	NA
2	24.8	39,810,804	37,825,149	solcap_snp_sl_33636	5.35	0.051	-0.008	-0.043
3	72.5	63,163,824	57,590,779	CL015346-0143_solcap_snp_sl_62495	6.96	NA	NA	NA
4	20.1	2,923,218	2,959,057	solcap_snp_sl_21374	8.22	NA	NA	NA
5	35.8	4,891,796	4,941,398	solcap_snp_sl_48814	4.19	NA	NA	NA
5	55.5	61,581,473	60,979,304	solcap_snp_sl_22605	7.37	-0.095	0.053	0.042
5	78.7	64,086,547	63,470,159	solcap_snp_sl_224	7.68	-0.101	0.134	-0.033
6	45.9	43,804,153	41,468,982	solcap_snp_sl_57810	5.13	NA	NA	NA
6	66.1	48,838,772	46,411,582	solcap_snp_sl_31676	4.10	-0.161	0.342	-0.181
7	27.4	3,747,320	3,720,215	solcap_snp_sl_11067	6.13	-0.235	0.256	-0.021
7	71.1	65,597,754	65,439,722	solcap_snp_sl_12101	8.37	0.215	-0.158	-0.058
8	5.8	985,676	998,051	solcap_snp_sl_7326	3.89	0.038	-0.034	-0.004
8	35.9	56,243,052	54,345,666	solcap_snp_sl_18196	14.81	NA	NA	NA
8	53.1	60,063,143	58,175,394	solcap_snp_sl_64642	11.10	0.043	0.009	-0.052

8	58.0	61,713,636	59,822,513	solcap_snp_sl_34763	10.52	NA	NA	NA
9	8.3	4,706,233	4,738,076	CL016894-0326_solcap_snp_sl_39506	13.40	-0.207	0.073	0.134
9	31.3	64,905,399	61,002,505	solcap_snp_sl_42973	8.87	NA	NA	NA
10	26.5	3,991,802	3,783,084	solcap_snp_sl_34373	6.64	0.007	-0.006	-0.001
10	56.3	62,638,578	61,745,624	solcap_snp_sl_59444	8.59	-0.033	0.027	0.006
11	22.4	3,158,623	3,188,167	solcap_snp_sl_21767	7.58	-0.018	-0.001	0.018
12	39.5	9,972,015	10,060,803	solcap_snp_sl_40606	12.81	0.017	-0.010	-0.007
Individual fruit fresh weight (g fruit^{-1})								
1	9.4	1,511,380	1,518,201	solcap_snp_sl_60330	5.72	0.56	-0.13	-0.42
1	37.0	4,597,945	4,614,026	solcap_snp_sl_59890	6.43	NA	NA	NA
1	67.0	82,553,933	74,799,309	solcap_snp_sl_17063	5.82	0.27	-0.31	0.03
1	125.8	97,461,827	89,815,320	SGN-U573826_snp1014_solcap_snp_sl_53877	8.99	NA	NA	NA
2	24.8	39,810,804	37,825,149	solcap_snp_sl_33636	20.08	NA	NA	NA
3	7.5	1,390,288	1,427,470	solcap_snp_sl_63290	12.64	NA	NA	NA
3	47.9	52,628,897	47,197,611	solcap_snp_sl_30678	4.83	1.64	-2.26	0.62
3	103.8	69,555,207	63,993,538	solcap_snp_sl_33838	3.47	2.52	-1.70	-0.82
4	31.6	4,635,914	4,657,293	solcap_snp_sl_27161	2.99	NA	NA	NA
4	50.9	56,086,497	55,434,838	solcap_snp_sl_3111	3.44	NA	NA	NA
5	35.8	4,891,796	4,941,398	solcap_snp_sl_48814	10.74	NA	NA	NA
5	72.9	63,379,877	62,769,283	solcap_snp_sl_12279	2.94	1.64	-6.13	4.49
5	88.9	65,747,765	65,139,785	solcap_snp_sl_37971	7.50	0.57	-1.36	0.79
6	10.2	35,008,382	32,780,157	solcap_snp_sl_44686	4.44	NA	NA	NA
7	40.2	59,231,709	59,118,881	solcap_snp_sl_10686	6.38	NA	NA	NA
8	5.8	985,676	998,051	solcap_snp_sl_7326	5.86	2.73	0.91	-3.63
8	35.9	56,243,052	54,345,666	solcap_snp_sl_18196	5.00	NA	NA	NA

8	60.5	62,345,855	63,310,880	CL016130-0295	7.88	1.07	-0.64	-0.43
9	0.6	651,775	650,804	solcap_snp_sl_28404	17.93	NA	NA	NA
10	17.1	2,686,835	2,563,006	CL015957-0328	12.78	1.39	-0.25	-1.14
11	11.8	1,988,809	2,024,684	solcap_snp_sl_10611	4.54	NA	NA	NA
12	28.8	3,768,296	3,806,963	CL009187-0469	35.78	-9.23	24.33	-15.11
Fruit fresh weight per truss (g truss^{-1})								
1	125.8	97,461,827	89,815,320	SGN-U573826_snp1014_solcap_snp_sl_53877	3.27	-10.8	7.8	3.0
2	26.7	40,022,938	38,037,260	solcap_snp_sl_36037	11.09	-6.2	4.2	1.9
3	7.5	1,390,288	1,427,470	solcap_snp_sl_63290	9.59	NA	NA	NA
3	46.0	35,965,114	32,004,652	solcap_snp_sl_30590	3.88	2.4	-27.5	25.1
4	11.8	1,562,490	1,594,008	solcap_snp_sl_63976	3.01	NA	NA	NA
5	35.8	4,891,796	4,941,398	solcap_snp_sl_48814	3.02	NA	NA	NA
6	0.3	394,870	413,654	solcap_snp_sl_39959	3.86	9.7	-3.7	-6.0
6	45.9	43,804,153	41,468,982	solcap_snp_sl_57810	5.67	5.6	-2.6	-3.0
7	59.8	63,644,286	63,561,908	solcap_snp_sl_55514	7.42	29.7	7.4	-37.1
9	0.8	748,255	767,180	solcap_snp_sl_14676	10.94	NA	NA	NA
10	17.1	2,686,835	2,563,006	CL015957-0328	3.14	10.4	-7.6	-2.8
12	28.8	3,768,296	3,806,963	CL009187-0469	13.15	-28.2	17.2	11.0
12	66.7	64,671,315	64,156,637	solcap_snp_sl_62972	4.58	10.8	-13.7	2.9
Fruit fresh-dry weight ratio								
1	103.6	92,043,676	84,331,398	solcap_snp_sl_40199	1.09	0.28	-0.23	-0.06
1	120.9	96,218,696	88,563,470	solcap_snp_sl_13402	2.96	-0.45	-0.20	0.65
2	8.8	36,167,998	34,191,377	solcap_snp_sl_15706	1.07	NA	NA	NA
2	38.1	43,252,729	41,236,467	CL016172-0398_solcap_snp_sl_29710	1.09	0.41	-0.31	-0.10
3	54.5	57,784,509	52,357,459	solcap_snp_sl_35369	1.18	0.60	-0.68	0.08

	4	64.4	60,772,015	58,731,439	solcap_snp_sl_22225	3.01	-0.07	0.03	0.04
	5	35.8	4,891,796	4,941,398	solcap_snp_sl_48814	1.11	-0.95	0.63	0.32
	5	65.5	62,819,450	62,212,955	solcap_snp_sl_123	1.98	-0.70	0.71	-0.01
	6	10.2	35,008,382	32,780,157	solcap_snp_sl_44686	0.97	NA	NA	NA
	6	31.9	40,659,989	38,300,758	solcap_snp_sl_12771	1.33	NA	NA	NA
	7	53.2	62,093,144	62,018,573	solcap_snp_sl_53425	1.25	NA	NA	NA
	8	35.9	56,243,052	54,345,666	solcap_snp_sl_18196	3.19	1.84	0.95	-2.79
	8	58.0	61,713,636	59,822,513	solcap_snp_sl_34763	1.20	0.46	-0.26	-0.20
	9	0.0	411,890	413,568	solcap_snp_sl_17510	3.63	NA	NA	NA
	9	70.2	71,185,917	67,281,966	solcap_snp_sl_25745	2.39	-1.06	0.81	0.26
	11	11.8	1,988,809	2,024,684	solcap_snp_sl_10611	2.07	0.21	-0.17	-0.04
Total fruit dry weight (kg stem^{-1})									
	2	24.8	39,810,804	37,825,149	solcap_snp_sl_33636	0.99	-0.0166	-0.0046	0.0211
	3	24.8	2,704,993	2,681,265	solcap_snp_sl_14355	1.24	0.0274	-0.0052	-0.0222
	4	11.8	1,562,490	1,594,008	solcap_snp_sl_63976	0.93	NA	NA	NA
	4	77.9	63,407,578	61,390,203	solcap_snp_sl_47259	1.84	0.0220	-0.0296	0.0076
	5	25.2	3,416,161	3,463,230	CL017464-0100	1.50	0.0270	-0.0012	-0.0258
	7	17.2	2,131,288	2,103,765	solcap_snp_sl_22102	1.96	NA	NA	NA
	7	66.4	64,739,981	64,587,131	solcap_snp_sl_71171	0.92	NA	NA	NA
	8	31.5	55,321,421	53,417,147	solcap_snp_sl_48643	2.07	0.0267	-0.0257	-0.0011
	8	54.7	60,405,484	58,517,969	solcap_snp_sl_64706	1.91	-0.0319	-0.0280	0.0599
Harvest index									
	1	13.5	1,809,372	1,814,874	solcap_snp_sl_60276	1.47	NA	NA	NA
	1	127.3	98,090,338	90,442,389	CL015806-0417_solcap_snp_sl_53692	1.37	NA	NA	NA
	2	8.8	36,167,998	34,191,377	solcap_snp_sl_15706	2.35	-0.009	-0.084	0.093

2	53.4	47,123,587	45,124,478	solcap_snp_sl_49965	1.84	-0.061	0.152	-0.091
3	7.5	1,390,288	1,427,470	solcap_snp_sl_63290	1.30	-0.009	-0.017	0.026
3	87.0	66,140,542	60,594,445	solcap_snp_sl_61881	1.43	0.025	0.000	-0.025
4	20.1	2,923,218	2,959,057	solcap_snp_sl_21374	1.18	NA	NA	NA
4	64.4	60,772,015	58,731,439	solcap_snp_sl_22225	1.73	NA	NA	NA
5	25.2	3,416,161	3,463,230	CL017464-0100	1.14	0.018	-0.005	-0.012
7	27.4	3,747,320	3,720,215	solcap_snp_sl_11067	2.20	0.009	-0.007	-0.002
7	53.2	62,093,144	62,018,573	solcap_snp_sl_53425	0.97	NA	NA	NA
8	24.9	3,781,365	3,700,890	solcap_snp_sl_51919	2.23	NA	NA	NA
8	54.7	60,405,484	58,517,969	solcap_snp_sl_64706	2.55	-0.010	-0.058	0.068
9	0.0	411,890	413,568	solcap_snp_sl_17510	1.81	0.019	-0.010	-0.008
9	70.2	71,185,917	67,281,966	solcap_snp_sl_25745	1.34	0.020	-0.010	-0.010
11	22.4	3,158,623	3,188,167	solcap_snp_sl_21767	2.21	NA	NA	NA
Total biomass (above ground) (kg stem^{-1})								
1	73.4	84,060,941	76,298,637	solcap_snp_sl_602	1.07	-0.008	-0.006	0.013
3	24.8	2,704,993	2,681,265	solcap_snp_sl_14355	0.92	0.087	-0.001	-0.087
4	20.1	2,923,218	2,959,057	solcap_snp_sl_21374	0.84	NA	NA	NA
4	77.9	63,407,578	61,390,203	solcap_snp_sl_47259	0.89	0.003	-0.003	0.000
7	17.2	2,131,288	2,103,765	solcap_snp_sl_22102	0.89	NA	NA	NA
7	66.4	64,739,981	64,587,131	solcap_snp_sl_71171	0.92	0.051	-0.023	-0.028
8	31.5	55,321,421	53,417,147	solcap_snp_sl_48643	0.98	0.038	-0.038	0.000
9	51.1	68,604,010	64,704,353	solcap_snp_sl_58253	0.95	-0.047	0.025	0.022

Supplementary Table 4 Prediction accuracies for each trait in a greenhouse by using multi-QTLs detected in the other greenhouse.

	(A) Comparison of fitted				(B) Multi-QTL model by 70 RILs → Predict RILs of which observations were available			
	Fer (pred) - Fer (obs)	JM (pred) - JM (obs)	Fer (pred) - JM (obs)	JM (pred) - Fer (obs)	Fer (pred) - Fer (obs)	JM (pred) - JM (obs)	Fer (pred) - JM (obs)	JM (pred) - Fer (obs)
	Multi-QTL model by 346 RILs → Predicted 346 RILs				Multi-QTL model by 70 RILs → Predicted 346 RILs			
Total yield	0.40	0.58	0.46	0.36	0.33	0.43	0.41	0.25
Marketable yield	0.40	0.59	0.46	0.36	0.32	0.46	0.41	0.27
Number of fruits	0.59	0.68	0.47	0.40	0.48	0.57	0.45	0.33
Number of trusses	0.50	0.70	0.41	0.33	0.40	0.60	0.40	0.21
Number of fruits per truss	0.76	0.76	0.68	0.67	0.64	0.65	0.60	0.62
Individual fruit fresh weight	0.81	0.69	0.66	0.74	0.66	0.61	0.60	0.72
Fruit fresh weight per truss	0.66	0.54	0.51	0.61	0.49	0.43	0.47	0.57
	Multi-QTL model by 70 RILs → Predicted 70 RILs		Multi-QTL model by 70 RILs → Predicted 30 RILs				Multi-QTL model by 70 RILs → Predicted 70 RILs	
Fruit fresh-dry weight ratio	0.88	0.75	0.73	0.78			0.60	0.59
Total fruit dry weight	0.75	0.62	0.35	0.43			0.27	0.37
Harvest index	0.72	0.77	0.49	0.60			0.48	0.45
Total biomass	0.71	0.47	0.18	0.20			-0.01	-0.06

Supplementary Table 5 Accuracies for yield predictions in a greenhouse by multi-QTLs detected in the other greenhouse.

	(A) Comparison of fitted				(B) Multi-QTL model by 70 RILs → Predict 346 RILs			
	Fer (pred) - Fer (obs)	JM (pred) - JM (obs)	Fer (pred) - JM (obs)	JM (pred) - Fer (obs)	Fer (pred) - Fer (obs)	JM (pred) - JM (obs)	Fer (pred) - JM (obs)	JM (pred) - Fer (obs)
	Multi-QTL model by 346 RILs → Predicted 346 RILs							
By QTLs for total yield	0.40	0.58	0.46	0.36	0.33	0.43	0.41	0.25
By QTLs for marketable yield	0.40	0.59	0.46	0.36	0.32	0.46	0.41	0.27
Harvest dissection 1	0.43	0.55	0.47	0.36	0.16	0.43	0.37	0.30
Harvest dissection 2	0.40	0.52	0.42	0.34	0.14	0.35	0.26	0.20
Harvest dissection 3	0.37	0.50	0.39	0.32	0.27	0.33	0.41	0.20
	Multi-QTL model by 70 RILs → Predicted 70 RILs							
Biomass dissection 1st layer	0.04	0.55	-0.06	0.38	0.08	0.15	0.09	0.11
Biomass dissection 2nd layer	0.19	0.48	0.19	0.22	0.02	0.11	0.13	0.13

Chapter 4

Introgressions of wild germplasm boosted total plant biomass to improve yield by mostly unlinked light-use-efficiency QTLs and light-interception QTLs in tomato

Yutaka Tsutsumi-Morita[1,2], Sedighehsadat Khaleghi[1], Martin P. Boer[2], Frank F. Millenaar[3], Daniela Bustos-Korts[2], George A.K. Van Voorn[2], Leo F.M. Marcelis[1], Fred A. Van Eeuwijk[2], Ep Heuvelink[1]

[1] Wageningen University & Research, Horticulture and Product Physiology, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[2] Wageningen University & Research, Biometris, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[3] BASF's vegetable seeds business (Nunhems), Napoleonsweg 152, 6083 AB Nunhem, The Netherlands

Abstract

Identifying the genetic loci that improve biomass production and yield remains a challenge in plant breeding. Total (plant) biomass can be dissected into the light use efficiency (LUE) and the amount of intercepted light, which depends on the leaf area index (LAI). This study aimed to (1) identify quantitative trait loci (QTLs) for LUE, light interception, LAI, and leaf-level traits, (2) assess which traits share their genetic basis with fruit yield and total biomass, and (3) discuss how the detected QTLs can contribute to the selection of higher-yielding genotypes in indeterminate tomato (*Solanum lycopersicum*). Phenotyping was conducted on hybrids obtained from crossing two testers with 342 recombinant inbred lines (RILs) derived from four parents (two elite parents and two wild parents). Fraction of intercepted light (F_{int}) was calculated from LAI, and adapted light use efficiency (LUE_{adp}) was calculated as total biomass divided by F_{int}. We detected 9 QTLs for LUE_{adp}, 10 QTLs for F_{int} or LAI, and 3 to 12 QTLs for each leaf-related trait. Phenotypic analysis by using best linear unbiased estimate (BLUE) per RIL showed that both LUE_{adp} and F_{int} positively correlated with total biomass ($r = 0.72$ and $r = 0.69$ respectively). Four QTLs for LUE_{adp} and 7 QTLs for F_{int} or LAI colocalized with QTLs for total biomass. LUE_{adp} and F_{int} did not show a trade-off at the BLUE level, and most of their QTLs were unlinked. Therefore, both LUE_{adp} QTLs and F_{int} QTLs are useful to improve total biomass by combining positive alleles. On the other hand, fruit yield showed a higher correlation with LUE_{adp} ($r = 0.64$) than with F_{int} ($r = 0.26$). Four QTLs for LUE_{adp} and three QTLs for F_{int} or LAI colocalized with QTLs for yield. The reduced influence of F_{int} on yield compared to the influence of F_{int} on total biomass arose from the negative correlation between F_{int} and harvest index ($r = -0.71$), because leaf area was mainly determined by leaf dry weight ($r = 0.93$) and not by specific leaf area ($r = -0.07$, not significant). For RILs with a higher yield than the elite parents, this was attributed to both LUE_{adp} and F_{int}, or to LUE_{adp} alone. Wild parents showed positive allele effects on LUE_{adp} and F_{int}, some of which also showed positive allele effects on yield. Our results indicate that wild parents possess beneficial alleles to improve total biomass via LUE_{adp} and F_{int}, which provided the way for some RILs to achieve a higher yield than the elite parent.

4.1. Introduction

Introducing wild germplasm into elite cultivars has been proposed as a promising strategy to improve yield (Tanksley and McCouch, 1997; Schulthess *et al.*, 2022). However, identifying genetic bases to improve yield and biomass production remains a challenge because yield and total (plant) biomass are complex traits, which integrate many underlying physiological processes (Reynolds *et al.*, 2012). Fruit yield (dry mass) can be considered as the product of total biomass and the fraction of biomass partitioned into fruits (harvest index) (Eq. 1) (Hay, 1995). Increased harvest index has contributed substantially to yield improvement in the 20th century in major agricultural crops, hence there might be little room left for further improvement (Unkovich *et al.*, 2010; Foulkes *et al.*, 2011). Therefore, several studies suggest that further yield improvement must come from increased biomass production (Zhu *et al.*, 2010; Reynolds *et al.*, 2012). In our previous study, we used yield dissection models in combination with quantitative genetic modelling to consider the strategy of high-yield breeding in tomato (*Solanum lycopersicum*) (Tsutsumi-Morita *et al.*, 2021). We found that yield was much more correlated to total biomass than to harvest index, indicating that selecting for higher total biomass may be a promising strategy to indirectly improve yield.

Dissection of a complex trait into underlying component traits reveals the relationship between component traits (i.e., trade-off, synergy, or independence) and how to combine quantitative trait loci (QTLs) for these component traits to improve the upper trait (Tsutsumi-Morita *et al.*, 2021). Total biomass is determined as the product of light interception and the efficiency of the conversion of intercepted light into biomass (light use efficiency; LUE) (Eq. 2) (Monteith, 1977). Light interception is further considered as the product of cumulative incident light during cultivation and the fraction of intercepted light by the crop. The fraction of intercepted light is often modelled using Beer–Lambert's law. Light attenuates exponentially through the canopy, which is modelled via a leaf area index (LAI) and a light extinction coefficient k (Eq. 3) (Monsi and Saeki, 1953; Hirose, 2005).

$$Yield = Total\ biomass \times Harvest\ index \qquad Eq.\ 1$$

$$Total\ biomass = Light\ interception\ \times LUE \qquad Eq.\ 2$$

$$\begin{aligned} Light\ interception & \\ &= (Incident\ light) \times (Fraction\ of\ intercepted\ light) \\ &= (Incident\ light) \times (1 - e^{-k \times LAI}) \end{aligned} \qquad Eq.\ 3$$

LUE and light interception might be connected either positively or negatively. Increasing leaf thickness, and thus more tissue per leaf area, may result in a higher content of nitrogen, ribulose-1,5-bisphosphate carboxylase-oxygenase (Rubisco), chlorophyll content, and more electron transport machinery per unit leaf area (Mathan *et al.*, 2021; Yin *et al.*, 2022). Thus, increasing leaf thickness (reducing specific leaf area) might increase LUE (per unit leaf area).

However, the increased leaf thickness may reduce leaf area and light interception (Austin, 1989), which might cause a reduction in whole-plant photosynthesis (Richards, 2000). On the other hand, a higher LUE without reducing specific leaf area might increase early leaf expansion (Yin *et al.*, 2022) and light interception. Improved leaf photosynthesis (thus improved LUE) by genetic modification also resulted in larger leaf areas or larger plants (Kromdijk *et al.*, 2016; Driever *et al.*, 2017). Therefore, increased biomass might be caused directly by higher leaf photosynthetic rates but also indirectly by increased light interception (Yin *et al.*, 2022). Thus, it is worthwhile investigating the relationship between LUE and light interception in a breeding population to find a strategy to improve total biomass.

Quantitative trait loci (QTLs) for LUE were reported for wheat (Molero *et al.*, 2019). Genomic predictions of radiation use efficiency (RUE, which is similar to LUE, but based on global radiation rather than visible light) and three other traits have been used in yield prediction for maize using a crop growth model (Technow *et al.*, 2015). So far, to our knowledge, no study reported QTLs for LUE in tomato. Light interception is another important determinant of total biomass. The main determinant of light interception is LAI, for which QTLs have been reported in wheat (Wang *et al.*, 2022), pepper (Alimi et al., 2013b), and chickpea (Sivasakthi *et al.*, 2018). QTLs for canopy cover dynamics, which is related to LAI, were reported in potato (Khan *et al.*, 2019). In maize, QTLs for parameters of leaf models, which predict the vertical distribution of leaf area, were reported (Lacube *et al.*, 2020). In tomato, QTLs for leaf shape (Holtan and Hake, 2003; Frary *et al.*, 2004; Chitwood *et al.*, 2013; Fulop *et al.*, 2016; Nakano *et al.*, 2016) (e.g., leaf length, leaf width, and the number of leaflets), whole-genome predictions for leaf length and leaf width (Yamamoto *et al.*, 2016), QTLs for individual leaf area (Muir *et al.*, 2014), and QTLs for the total number of leaves (Shirasawa *et al.*, 2013; Yamamoto *et al.*, 2016) were reported. However, to our knowledge, no QTLs for LAI have been reported in tomato.

This study aimed to: (i) determine QTLs for the crop-level traits LUE, light interception, LAI, and several leaf-level traits such as the number of leaves, leaf width and leaf dry weight; (ii) assess whether there are shared QTLs between crop-level traits, leaf-level traits, biomass and fruit yield; (iii) discuss how the detected QTLs could contribute to the selection of higher-total-biomass genotypes and/or higher-yielding genotypes that combine positive alleles for crop-level and leaf-level traits in tomato.

4.2. Materials and methods

4.2.1. Breeding Population

Four parents (two elite parents (E1, E2) and two wild parents (W1, W2)) of indeterminate tomato (*Solanum lycopersicum*) were crossed to develop a 342 RILs population. These RILs were crossed with two testers (TP1, TP2), producing two populations of 342 genotypes (RILs

× TP1 and RILs × TP2), which are the same populations as in our previous study[6]. The phenotypes of the hybrids of RILs and testers were used to identify QTLs and estimate QTL effects on the RILs.

Identity by descent (IBD) probabilities between parents and offspring (Zheng *et al.*, 2015) at 279 SNP SOLCAP marker positions were used as genetic information as explained in a previous study (Tsutsumi-Morita *et al.*, 2021).

4.2.2. Growing conditions and experimental design

This was the same experiment as reported by Tsutsumi-Morita et al., 2021, but here we used different data related to leaf. The phenotyping experiment was conducted in a high-tech greenhouse with a climate control system (e.g., heating, ventilation, and CO_2 injection) in the Netherlands from February to November 2016, with commercial crop management. Growing conditions and experimental design were described in a previous paper (Tsutsumi-Morita *et al.*, 2021).

4.2.3. Phenotyping

Phenotyping was conducted for leaf-related traits (Table 1). Some traits were directly measured: the number of leaves on the plant (LNplant); leaf width (Lwidth), leaf area (LA), and leaf dry weight (LDW) of three bottom leaves. Specific leaf area (SLA) was calculated by dividing LA by LDW. Leaf area index (LAI), fraction of intercepted light based on LAI (F_{int}), and adapted light use efficiency (LUE_{adp}) were estimated as explained below.

4.2.3.1. Leaf area index (LAI) estimation

LAI was estimated in three steps based on non-destructive measurements. First, leaf width was measured on a selected number of leaves of each plant. From this, we estimated the width of all leaves on the plant. Second, based on the estimated width we estimated area per leaf, which finally allowed us to calculate leaf area per plant and per floor area.

We estimated leaf width (defined as the biggest width of the composite leaf) from leaf position (1 is the top leaf, which is the youngest leaf longer than 5 cm) because it was not feasible to measure all leaf widths for more than 700 plots. All leaves on a stem were measured in spring for a subset of 102 plots. This subset was aimed at representing well the genetic diversity in the population (Bustos-Korts *et al.*, 2016). A Holling type III equation was used for the relation between leaf width and leaf position because the shape of the curve fitted well with our data (median of R-squared = 0.91; median of RMSE = 4.0) (Supplementary Figure 1a). We used the nls function in R (version 4.0.2) (R Core Team, 2020) to fit Eq. 4 to the observations.

$$Leaf\ width\ (cm) = a \times (Leaf\ position)^2 / (b^2 + (Leaf\ position)^2) \qquad Eq.\ 4$$

where *a* is the asymptote, and *b* indicates the leaf position at which the leaf attains half of the maximum width (Bolker, 2008). For all other plots in spring and for all plots in summer, measurements were conducted at a limited number of leaf positions (3, 5, 7, 8, 9, 10) and the three leaves from the bottom.

Leaf area was estimated from the square of leaf width (Schwarz and Kläring, 2001) (Supplementary Figure 1b).

$$Leaf\ area\ (cm^2) = c \times (Leaf\ width)^2 \qquad Eq.\ 5$$

where *c* is a proportionality factor. To confirm whether this relationship was appropriate for our experiment, we randomly sampled 78 leaves varying in size in spring, in addition to three bottom leaves for all plots. Leaf area and leaf width were measured using a leaf area meter (LI-3100; LI-COR Inc., Lincoln, NE, USA) and tape measure (a flexible ruler), respectively. Leaf area and square of leaf width showed a linear relationship (R-squared = 0.84) (Supplementary Figure 2). We assumed that *c* may be genotype dependent, and it was estimated for each plot based on the three bottom leaves on a stem of a plant, separately for spring and summer. Leaf width was measured for each bottom leaf. Leaf area was measured for each bottom leaf separately in some plots (177 plots for spring and 20 plots for summer), while leaf areas of three leaves were measured together for other plots (532 plots for spring and 690 plots for summer), where it was assumed that the proportionality factor between leaf area and the square of leaf width was the same for all three leaves in the same plot.

Finally, LAI was computed by summing the leaf area of all leaves on a stem multiplied by the number of stems per m^2.

$$Leaf\ area\ index\ (m^2\ m^{-2}) = \sum_{i=1}^{n} Leaf\ area_i \times 12/(3.56\ \times 10000) \qquad Eq.\ 6$$

where *i* is the leaf position and *n* is the number of leaves on a stem. One plot (3.56 m^2) included 12 stems of one genotype. 10000 is the unit change from cm^2 to m^2. For further analyses, for each trait, we used the mean values (the mean of three seasons for LNplant, Lwidth, LA, LDW, SLA; and for two seasons for LAI, *a*, *b*, *c*).

4.2.3.2. Fraction of intercepted light and adapted light use efficiency

LAI was used to estimate the fraction of intercepted light (F_{int}) by Beer–Lambert's law (Monsi and Saeki, 1953; Monteith, 1981).

$$Fraction\ of\ intercepted\ light\ (F_{int}) = 1 - e^{-k \times LAI} \qquad Eq.\ 7$$

where *k* is the light extinction coefficient. In this study, we could not determine a genotype-specific *k* because each plot area was relatively small, and the light condition of a plot was influenced by neighbouring plots. Thus, we assumed *k* = 0.75 for all plots as a reasonable value for greenhouse tomato in the Netherlands (Heuvelink and Buiskool, 1995; Heuvelink, 1996).

Total plant biomass (total biomass) (kg m^{-2}) was the sum of fruit dry mass, leaf dry mass, and stem dry mass as described in our previous study (Tsutsumi-Morita *et al.*, 2021). Because total biomass (kg m^{-2}) can be calculated as the product of cumulative incident light above the canopy (MJ m^{-2}), the fraction of light intercepted by the crop (F_{int}), and the efficiency of the conversion of light into biomass (light use efficiency; LUE (kg m^{-2} / MJ m^{-2})) (Monteith, 1977) (Eq. 2; Eq. 3), LUE can be derived as follows:

$$Light\ use\ efficiency\ (LUE) = \frac{Total\ biomass}{F_{int} \times Incident\ light} \qquad Eq.\ 8$$

Because the incident light was the same for all plots, evaluation of genetic effect of LUE is equivalent for the evaluation of *(Light use efficiency)* × *(Incident light)*, that is, total biomass divided by F_{int} (kg m^{-2}). In this paper, we evaluated *(Light use efficiency)* × *(Incident light)* as adapted light use efficiency (LUE_{adp}) (kg m^{-2}).

$$LUE_{adp} = \qquad Eq.\ 9$$

$$(LUE) \times (Incident\ light) = \frac{Total\ biomass}{F_{int}}$$

4.2.4. Phenotypic and QTL analyses

Phenotypic and QTL analyses followed a similar procedure as in our previous study (Tsutsumi-Morita *et al.*, 2021). In brief, phenotypic data was corrected for local spatial trends created by variations in environmental and management conditions. Spatially adjusted genotypic means per RIL (BLUE: the best linear unbiased estimate) were calculated by R-package SpATS (Spatial Analysis of Field Trials with Splines) using a two-dimensional Penalized spline (P-spline) mixed model (Velazco *et al.*, 2017; Rodríguez-Álvarez *et al.*, 2018). The same R-package was used to calculate the generalized heritability (Cullis *et al.*, 2006; Rodríguez-Álvarez *et al.*, 2018). The BLUEs per RIL were used for phenotypic analyses and QTL detection. The fraction of the trait phenotypic variance that was explained by its component traits was calculated as the square of the correlation coefficient.

BLUEs per RIL were used as responses in models for single and multiple QTLs. A mixed model framework was applied by using IBD probabilities between parents and RILs described in our previous study (Tsutsumi-Morita *et al.*, 2021; Li *et al.*, 2021) within the package ASReml version 4 (Butler *et al.*, 2017) in R version 3.6.1 (R Core Team, 2020). A threshold for significance in the single-QTL analysis was defined at $-log_{10}P \geq 2.9$. Two adjacent QTLs were recognized as different QTLs, if the distance between two QTLs was larger than 20 cM or if $-log_{10}P$ value dropped by more than 0.5 between two adjacent QTLs. All single QTLs were then fitted together in a multi-QTL model. We excluded some QTLs from the multi-QTL model when the variance component for the QTL was estimated as zero in the multi-QTL model. QTLs for different traits were considered to colocalize, if they were located within 20 cM and the same haplotype showed the highest positive allele effect at these QTLs. Statistical analyses and data visualization were performed using R version 4.0.2 (R Core Team, 2020).

Table 1 Trait definitions, units, and phenotyping (measurements).

Abbreviation	Trait	Unit	Phenotyping	
LUE_{adp}	Adapted light use efficiency	kg m^{-2}	Estimated	(Total biomass)/F_{int}
F_{int}	Fraction of intercepted light (based on LAI)	-	Estimated	1−exp(−k × LAI) under assumption of k = 0.75
LAI	Leaf area index	cm^2_{leaf} cm^{-2}_{soil}	Estimated	Described in main text. Mean of two seasons (May, Jul).
LN_{plant}	The number of leaves on plant	$stem^{-1}$	Measured	Measured on a stem per plot. Mean of three seasons (Jun, Jul, Aug).
L_{width}	Leaf width	cm	Measured	Mean of 3 bottom leaves on a stem per plot. Mean of three seasons (May, Jul, Sep).
LA	Leaf area	cm^2 $leaf^{-1}$	Measured	Mean of 3 bottom leaves on a stem per plot. Mean of three seasons (May, Jul, Sep).
LDW	Leaf dry weight	g $leaf^{-1}$	Measured	Mean of 3 bottom leaves on a stem per plot. Mean of three seasons (May, Jul, Sep).
SLA	Specific leaf area	cm^2 g^{-1}	Calculated	(LA of 3 bottom leaves) / (LDW of 3 leaves). Mean of three seasons (May, Jul, Sep).
a_asymp	Parameter a (asymptote)	-	Parameter	Leaf width model described in text. Mean of two seasons (May, Jul).
b	Parameter b (leaf position at which leaf width = a/2)	-	Parameter	Leaf width model described in text. Mean of two seasons (May, Jul).
c_prop	Proportionality factor c	-	Parameter	Leaf area model described in text (Leaf area = c × (Leaf width)^2). Mean of two seasons (May, Jul).

4.3. Results

4.3.1. All high-yielding lines showed higher total biomass than the elite parents

The best linear unbiased estimates (BLUEs) per recombinant inbred line (RIL) showed that all lines with higher fruit yield (hereafter denoted as yield) than the elite parents showed higher total plant biomass (hereafter denoted as total biomass) than the elite parents (Figure 1). This indicated that total biomass is one of the important traits to breed high-yielding lines in this population.

4.3.2. Large transgression

Elite parents were located close to the mean of the RIL distribution for most traits, except for leaf width and parameter "*a (asymptote)*" (less transgression), and proportionality factor "*c*" (larger transgression) (Figure 2). This indicates that wild parents brought both positive and negative genetic effects into the genetic background of the elite parents. Less transgression in leaf width and parameter "*a*" (the asymptote of leaf width growth) indicated that introgression of wild genomes had higher effects on reducing leaf width than increasing it (Figure 2e,i). Although more than half of the RILs showed shorter leaf width than the elite parents, the elite parents were located close to the mean of RIL distribution for leaf area. This corresponded to a large transgression in the proportionality factor "*c*" (ratio of leaf area/(leaf width)^2), indicating that more than half of the RILs showed a narrower leaf shape than the elite parents (Figure 2 k). The introgression of wild germplasms brought a wide range of phenotypic diversity to this population.

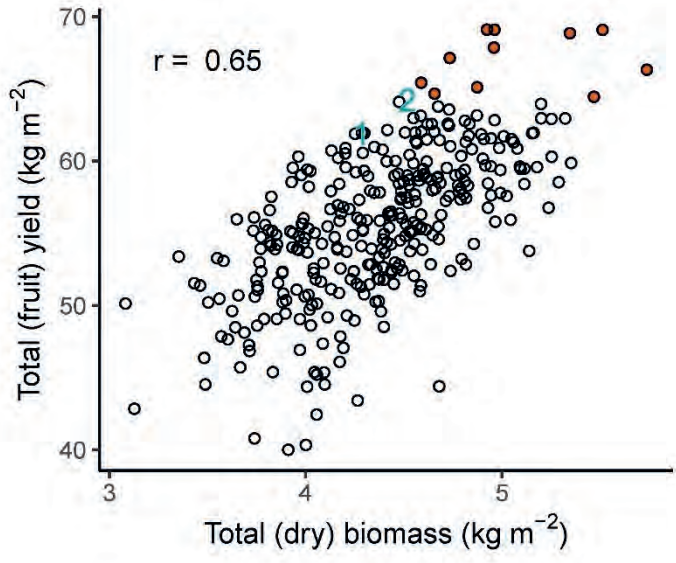


Figure 1 The correlation between total (fruit) yield and total (dry) biomass among tomato recombinant inbred lines (RILs). Green numbers indicate the elite parents (1, elite parent 1; 2, elite parent 2). Red-filled circles indicate higher-yielding lines than the elite parent 2. (This figure was drawn based on the data of Tsutsumi-Morita et al., 2021(Tsutsumi-Morita et al., 2021).)

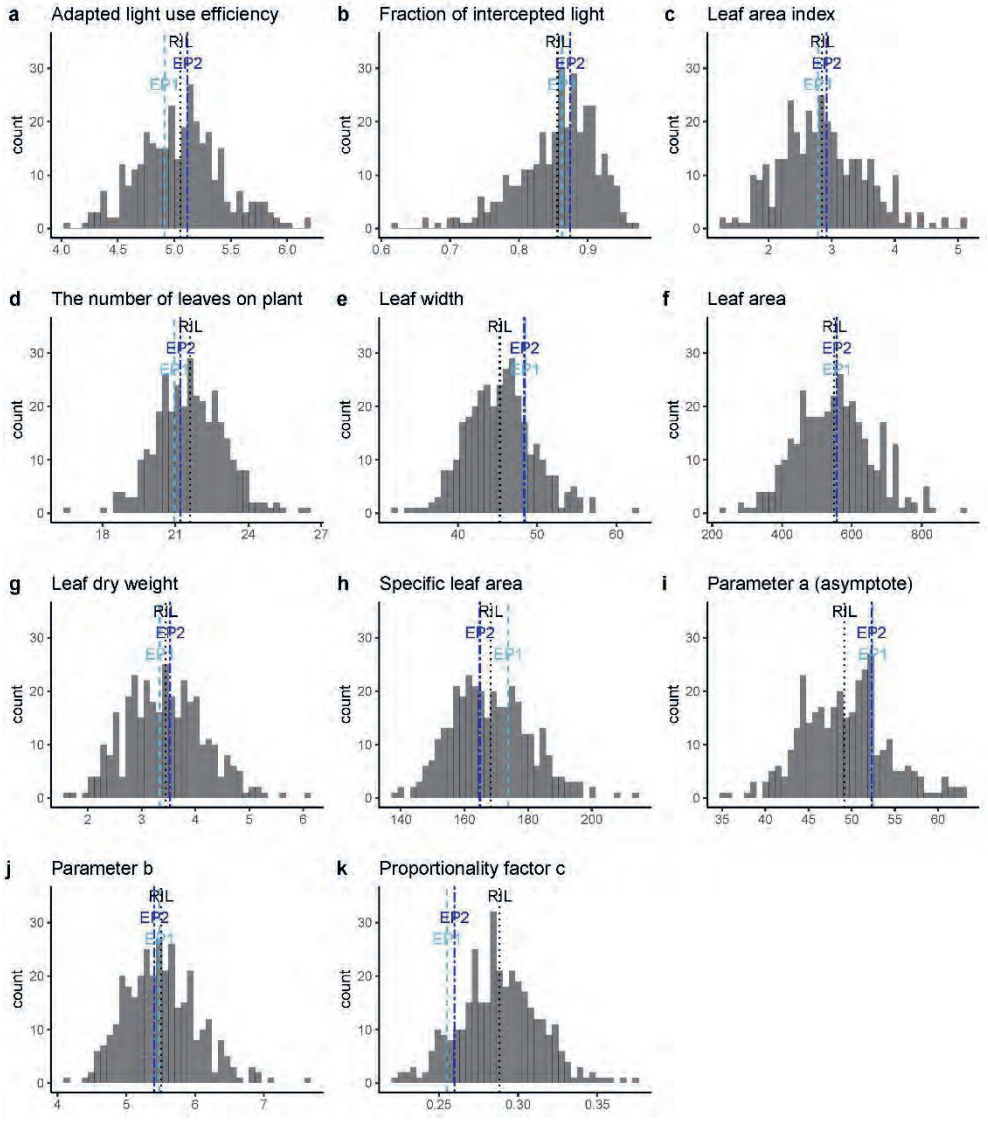


Figure 2 Distributions of the best linear unbiased estimates (BLUEs) per recombinant inbred line (RIL). The mean of RILs is indicated by the vertical black dashed line, the performance of the two elite parents by vertical blue dashed lines (the light blue colour for the elite parent 1 (EP1) and the blue colour for the elite parent 2 (EP2)). Traits were explained in Table 1.

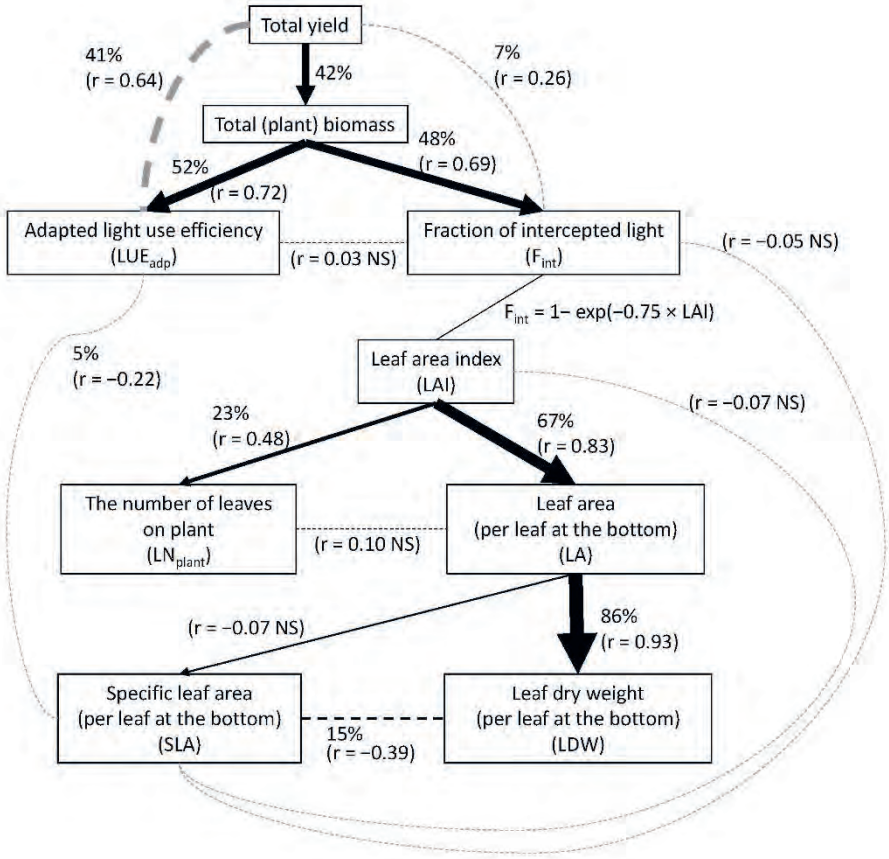


Figure 3 Fraction of phenotypic variance of an upper trait explained by each component trait (= square of correlation) (%), and correlation between traits (r). Total plant biomass (total biomass) is the multiplication of the adapted light use efficiency (LUE_{adp}) and the fraction of intercepted light (F_{int}). Leaf area index (LAI) is the summation of leaf area (LA) of all leaves per m^2, for which the number of leaves (LN_{plant}) and leaf area (LA) are two elements (Eq. 6). Specific leaf area (SLA) is leaf area (LA) divided by leaf dry weight (LDW), thus leaf area (LA) is the multiplication of specific leaf area (SLA) and leaf dry weight (LDW). Significance of correlation coefficient $P < 0.05$. Traits were explained in Table 1. The line thickness represents the fraction of phenotypic variance explained by the other trait. Abbreviation: NS, not significant.

4.3.3. Phenotypic analyses

4.3.3.1. No correlation between adapted light use efficiency (LUE_{adp}) and the fraction of intercepted light (F_{int})

No correlation was found between adapted light use efficiency (LUE_{adp}) and the fraction of intercepted light (F_{int}) (Figure 3; Figure 4b,c; Supplementary Figure 3). This may be due to no effect of specific leaf area (SLA) on F_{int} and only a weak effect on LUE_{adp} (correlation (r) = −0.22). SLA is often considered as an important determinant for leaf area. However, leaf

area (per leaf) (LA) was not correlated with SLA but was mainly determined by dry weight (per leaf) (LDW) (r = 0.93). The phenotypic variance of leaf area index (LAI) was explained by LA more (r = 0.83) than by the number of leaves (LN_{plant}) (r = 0.48). Accordingly, SLA did not show a correlation with LAI or F_{int}. On the other hand, SLA weakly negatively correlated with LUE_{adp} (r = −0.22) and LDW (r = −0.39). Thus, thicker leaves tended to have larger biomass per leaf and higher LUE_{adp}, although these correlations were weak. No correlation between LUE_{adp} and F_{int} suggests that both traits are likely to be regulated by different sets of QTLs.

4.3.3.2. Phenotypic variance of total biomass was explained by LUE_{adp} and F_{int} but the phenotypic variance of yield mainly was explained by LUE_{adp}

Phenotypic variance of total biomass was equally explained by LUE_{adp} (52%; r = 0.72) and F_{int} (48%; r = 0.69) (Figure 3; Supplementary Figure 3). Because of no correlation between LUE_{adp} and F_{int}, both traits are effective to improve total biomass without a trade-off. However, the phenotypic variance of yield was mainly explained by LUE_{adp} (41%; r = 0.64) and hardly by F_{int} (7%; r = 0.26). This is probably due to the negative correlation between F_{int} and harvest index (r = −0.71) (Figure 4c; Supplementary Figure 3), which came from the negative correlation between LAI and harvest index (r = −0.72) (Supplementary Figure 3). The negative correlation between LAI and harvest index might be caused by LAI being mainly determined by total leaf dry weight rather than SLA.

4.3.3.3. Characteristics of high-yielding lines

For breeding high-yielding lines, it is important that some RILs show higher yields than the elite parents (Figure 4). Although total biomass increased with increasing LAI (r = 0.69) (Figure 4a; Supplementary Figure 3), not all high-LAI lines showed a high yield. Accordingly, not all high-F_{int} lines showed a high yield (Figure 4b), partially due to a low harvest index (Figure 4c). Among the 10 lines with higher yield than the elite parents (red diamonds in Figure 4bc), 7 high-yielding lines showed higher F_{int} than the elite parents, and the other 3 high-yielding lines showed slightly lower F_{int} than the elite parents. On the other hand, all 10 high-yielding lines showed higher LUE_{adp} than the elite parents. Hence, high-yielding lines can be established by both LUE_{adp} and F_{int}. However, because of its stronger correlation with yield, selecting for high LUE_{adp} is a more certain strategy to obtain a high yielding line than selecting for high F_{int}.

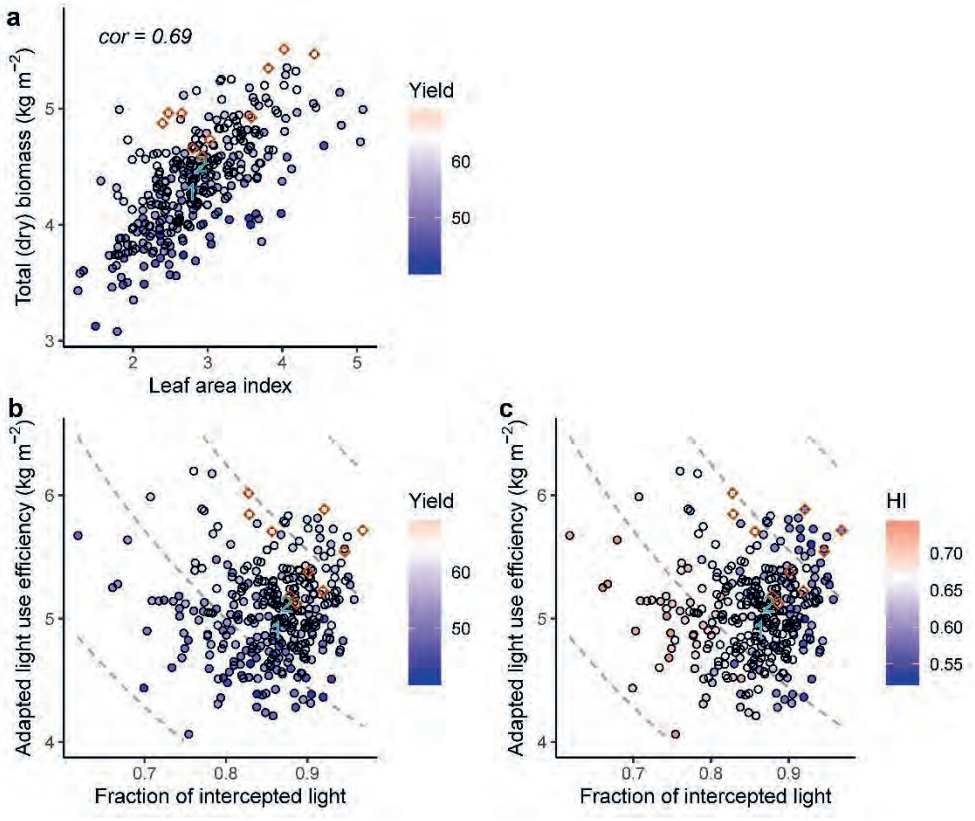


Figure 4 Characteristics of high-yielding lines, which show higher yield than the elite parents. (a) Relationship between total biomass and leaf area index. (b)(c) Response surface of total biomass (dashed isoclines) from the multiplication of adapted light use efficiency and the fraction of intercepted light. Green numbers denote the elite parents (1, elite parent 1; 2, elite parent 2). The colours of dots show yield (b) or harvest index (HI) (c), where the white colour corresponds to the elite parent 2 (recurrent parent). For ease of recognition, red diamonds indicate higher-yielding lines than the elite parents.

4.3.4. QTL detection

We detected three to 12 QTLs for each trait (Table 2). Low numbers of QTLs were detected for specific leaf area (3 QTLs) and parameter *b* (5 QTLs), both of which showed low heritabilities (Table 2). High numbers of QTLs were detected for parameter *a* (the asymptote of leaf width growth: 12 QTLs) and leaf width (11 QTLs), both of which showed high heritabilities. Prediction accuracies were relatively high for leaf-area-related traits (median in cross-validation 0.50 – 0.55 for F_{int}, LAI, leaf width, leaf area, leaf dry weight, and parameter *a*), which corresponded to high heritabilities (0.47 – 0.59).

Table 2 Heritabilities, numbers of QTLs, and prediction accuracies (correlations between observations and predictions). Traits were explained in Table 1.

		Number of QTLs		Prediction accuracy	
Trait	Heritability	Peaks in single-QTL analysis	Multi-QTLs	Fitted	Median in cross-validation
Adapted light use efficiency	0.41	11	9	0.51	0.40
Fraction of intercepted light	0.51	19	10	0.62	0.54
Leaf area index	0.47	19	10	0.62	0.55
The number of leaves on plant	0.30	14	9	0.56	0.47
Leaf width	0.59	16	11	0.65	0.54
Leaf area	0.54	12	7	0.59	0.53
Leaf dry weight	0.50	12	9	0.61	0.53
Specific leaf area	0.37	3	3	0.37	0.32
Parameter a (asymptote)	0.57	17	12	0.63	0.50
Parameter b	0.17	5	5	0.39	0.27
Proportionality factor c	0.39	10	8	0.51	0.38

4.3.4.1. Hardly any colocalization between adapted light use efficiency (LUE_{adp}) QTLs and fraction-of-intercepted-light (F_{int}) QTLs

Among the 9 QTLs for adapted light use efficiency (LUE_{adp}) and the 10 QTLs for the fraction of intercepted light (F_{int}), only one locus colocalized (within 20 cM with the same parental haplotype showing the highest positive effect). This colocalized locus corresponded to $LUE_{adp}7.2$ (the 2nd QTL for LUE_{adp} on chromosome 7) and $F_{int}7.1$, with elite parent 1 showing the highest positive effect (Figure 5). On the other hand, only $LUE_{adp}12.1$ and $F_{int}12.1$ were at the same position (within 20 cM) but wild parent 1 showed the most positive effect for F_{int} but the most negative effect for LUE_{adp}. In addition, $LUE_{adp}8.1$ and LAI8.2 were at the same position, but the trait-increasing haplotype allele came from a different parent ($LUE_{adp}8.1$, elite parent 2 (E2) > E1 > 0 > wild parent 2 (W2) > W1; LAI8.2, E1 > E2 > 0 > W1 > W2). The other LUE_{adp} QTLs were not located closely to F_{int} or LAI QTLs. Therefore, combining positive alleles for LUE_{adp} and F_{int} is largely possible via recombination to produce high-biomass lines.

4.3.4.2. More F_{int} or LAI QTLs than LUE_{adp} QTLs colocalized with biomass QTLs whereas this was the opposite for yield QTLs

Only one QTL on chromosome 7 detected as a QTL for both LUE_{adp} and F_{int}, was also a yield QTL and total biomass QTL (black box with a star in Figure 5). Among the 12 QTLs for total biomass, which we reported in our previous study (Tsutsumi-Morita *et al.*, 2021), seven were colocalized with QTLs for F_{int} or LAI (6 green and 2 black boxes (1 was overlapped) in Figure 5), whereas only four were colocalized with QTLs for LUE_{adp} (2 red and 2 black boxes in Figure 5). On the other hand, among the 16 QTLs for total yield, three were colocalized with QTLs for F_{int} or LAI (2 green and 1 black stars in Figure 5), and four were colocalized with QTLs for LUE_{adp} (3 red and 1 black stars in Figure 5). Thus, more F_{int} or LAI QTLs than LUE_{adp} QTLs colocalized with total-biomass QTLs, whereas more LUE_{adp} QTLs than F_{int} or LAI QTLs colocalized with yield QTLs. The high colocalization of F_{int} QTLs with total-biomass QTLs was not taken over to the colocalization of F_{int} QTLs with yield QTLs (Figure 5), which reflected that the high correlation of F_{int} with total biomass (0.69) was not taken over to the correlation of F_{int} with yield (0.26), although total biomass and yield showed a high correlation (0.65) (Figure 3; Supplementary Figure 3).

4.3.4.3. A specific leaf area (SLA) QTL colocalized with a LUE_{adp} QTL but not with F_{int} QTLs

None of the 3 QTLs for specific leaf area (SLA) were colocalized with QTLs for leaf area, LAI, or F_{int} (Figure 5). On the other hand, SLA1.2 was at the same position as LUE_{adp}1.2 with wild parent 1 showing the most negative effect for SLA but the most positive effect for LUE_{adp}. This is in line with SLA having no correlation with leaf area, LAI or F_{int} but a weak negative correlation with LUE_{adp} (Figure 3). SLA1.1 was at the same position as QTL for leaf dry weight (LDW1.1); however, the trait-increasing allele came from a different parent.

4.3.4.4. LUE_{adp} QTLs colocalization with leaf-shape related traits

LUE_{adp} might include some effects related to the light extinction coefficient *k* because we determined F_{int} from LAI and *k*, but *k* was assumed to be the same for all genotypes (Eq. 7). LUE_{adp} was derived from total biomass divided by F_{int} (Eq. 9), which means the genetic differences from *k* were included in values of LUE_{adp}. *k* is related to light penetration through the canopy, thus, leaf shape (leaf width, a (asymptote of leaf width model), proportionality factor c) might be related to LUE_{adp}. According to this concept, LUE_{adp}2.1 and c2.1 were colocalized, and wild parent 1 showed the most positive effect on both traits (narrower-shape leaves have higher LUE_{adp}). LUE_{adp}10.1, Leaf-width10.1 and parameter-a10.1 were also at the same position, with wild parent 1 showing the most positive effect for LUE_{adp} and the most negative effect for leaf width and parameter a (wild parent 1 genotype produces smaller leaf width and higher LUE_{adp}). LUE_{adp}7.2 and Leaf-width7.1 also colocalized but the

haplotype allelic effects did not fit this concept (wild parent 2 had the most negative effect on both traits).

4.3.4.5. Leaf area index (LAI) QTLs

We detected 10 QTLs for LAI, of which one was underlain by both a QTL for the number of leaves (LN_{plant}) and a QTL for leaf area (LA) (purple "AN" in Figure 5), three by QTLs for LN_{plant} (purple "N" in Figure 5), and another three by QTLs for LA (purple "A" in Figure 5). Note, for LAI2.1, haplotype effects were similar between wild parent 1 (0.057) and wild parent 2 (0.058). Therefore, the QTL for LAI was judged to be underlain by both LN_{plant} QTL (haplotype of wild parent 1) and LA QTL (haplotype of wild parent 2) (Supplementary Table 1).

4.3.5. Similar haplotype effects per chromosome between adapted light use efficiency (LUE_{adp}) and yield, and between the fraction of intercepted light (F_{int}) and total biomass

Wild parents brought both positive and negative effects on both LUE_{adp} and F_{int} (Figure 6). The total effects of each wild parent were negative for F_{int}, whereas the total effect of wild parent 1 was positive for LUE_{adp}. Positive effects of wild parents indicated that the elite parents can be improved by introducing wild germplasm for these two traits. Only chromosomes 2, 7, and 12 were involved in both traits; chromosomes 1, 8, and 10 were involved in only LUE_{adp}; chromosomes 4, 5, and 6 were involved in only F_{int}. The small overlapping of QTL-harboring chromosomes between LUE_{adp} and F_{int} makes it easier to combine favourable alleles for these two traits in selective breeding.

Haplotype effects per chromosome showed similar tendencies to QTL-colocalization analyses; F_{int} showed similar patterns to total biomass, and LUE_{adp} showed similar patterns to yield (Figure 6; Supplementary Figure 4). Haplotype effects per chromosome of F_{int} showed similar patterns to those of total biomass in chromosomes 2, 4, 5, and 6. Among these chromosomes, the pattern was also similar to yield only in chromosome 2. On the other hand, haplotype effects per chromosome of LUE_{adp} showed similar patterns to those of total biomass only in chromosome 7 and showed similar patterns to those of yield in chromosomes 7 and 12. Therefore, haplotype effects per chromosome of F_{int} were more similar to those of total biomass than those of yield; haplotype effects per chromosome of LUE_{adp} were more similar to those of yield than those of total biomass.

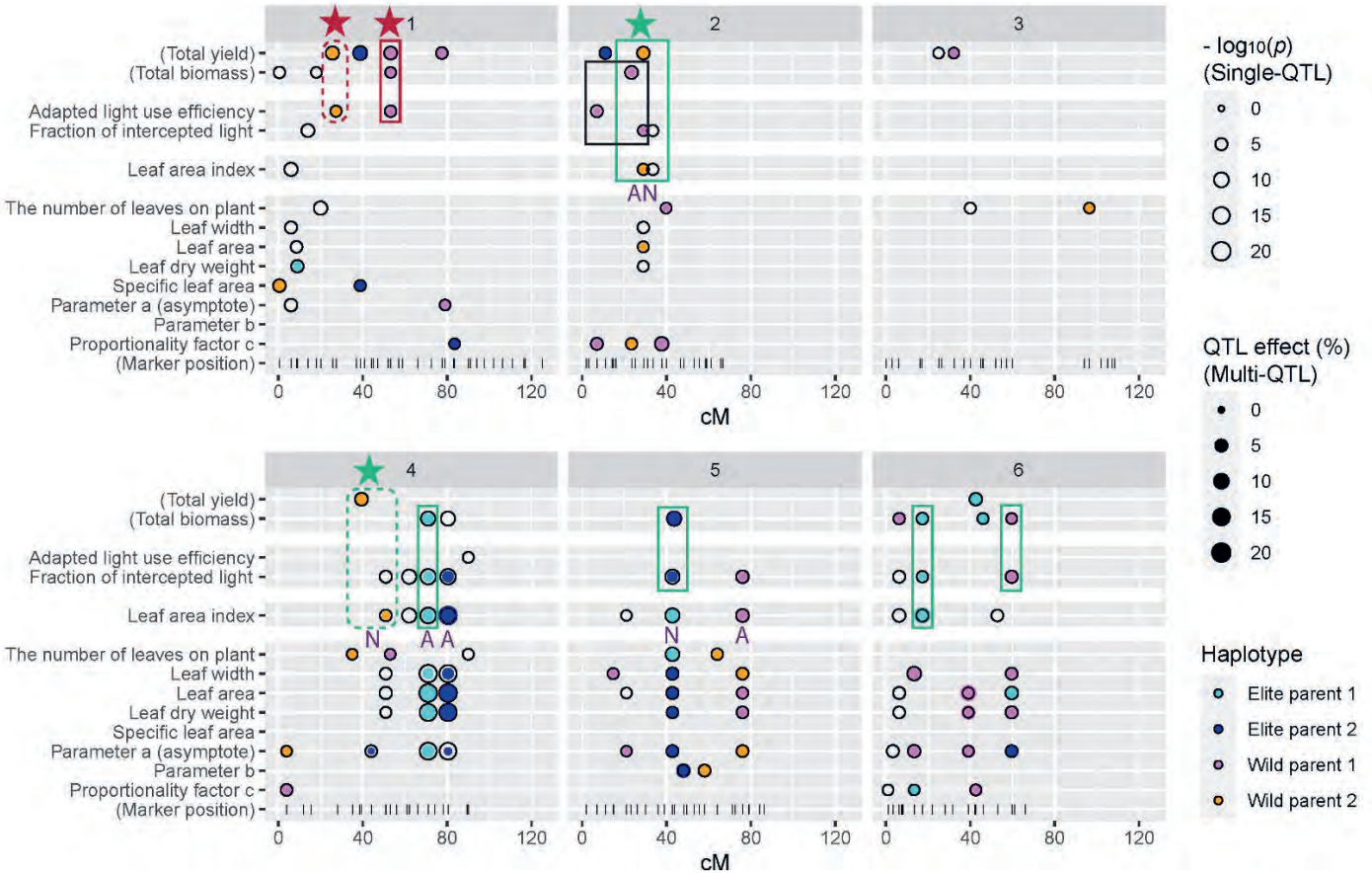


Figure 5 Caption two page later.

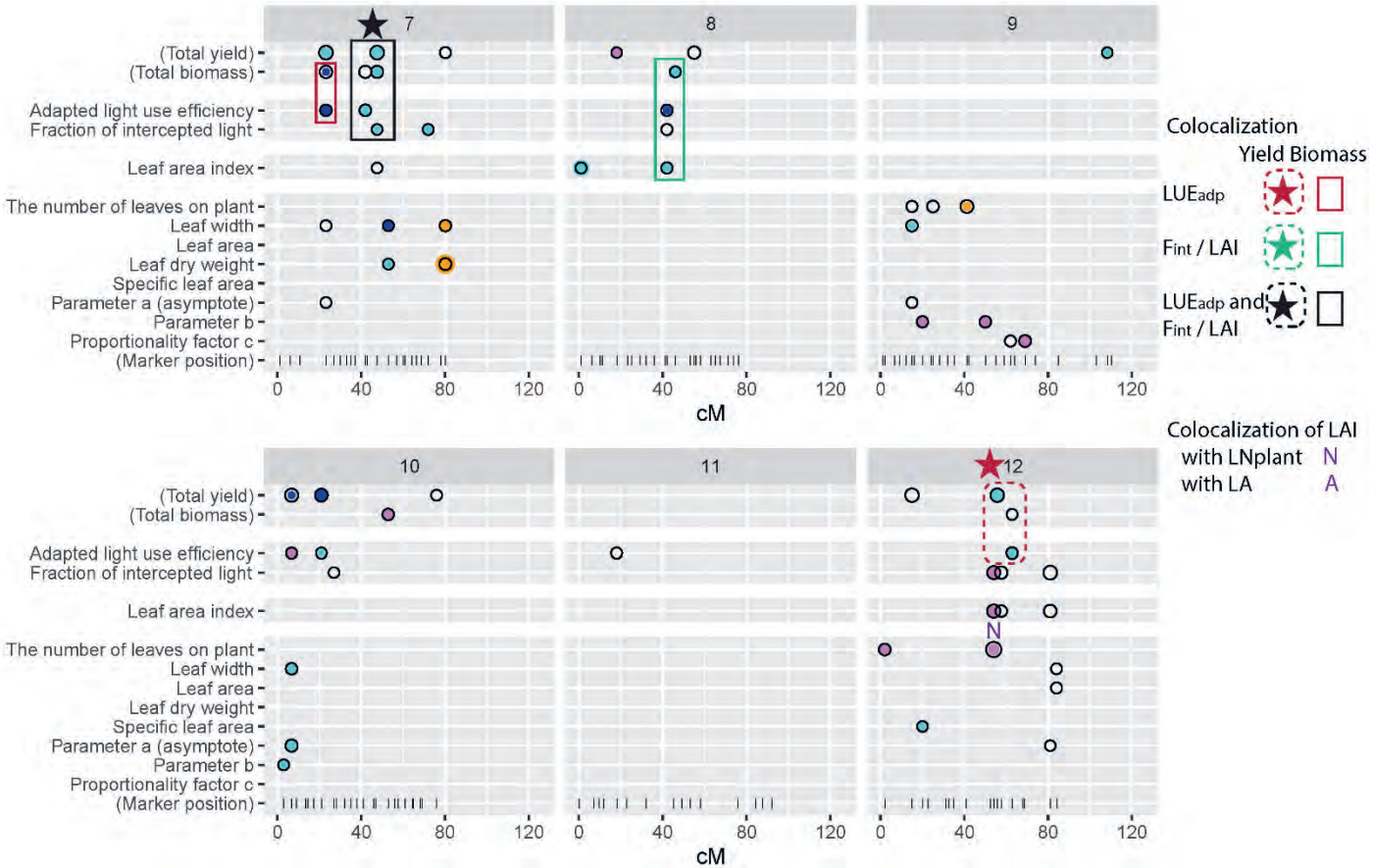


Figure 5 Caption next page

Figure 5 QTL locations and effects. The horizontal axis indicates chromosome number (1 – 12) and genetic position (cM) within the chromosome. Single-QTL positions are indicated by empty circles, where the size of the empty circle indicates -log$_{10}$(p) value. QTL effects in the multi-QTL model for each trait are indicated by dots, where the colour of dot shows the haplotype with the highest positive effect and the size of dot represents proportion (%) of the largest haplotype effect to the mean for the trait. Stars indicate yield-QTL colocalization, and boxes indicate total-biomass-QTL colocalization with components: adapted light use efficiency (LUE_{adp}, red colour); fraction of intercepted light (F_{int}) or leaf area index (LAI) (green colour); or both LUE_{adp} and F_{int} (black colour). Purple letters indicate colocalization of LAI QTLs with QTLs for the number of leaves on plant (N) and/or QTLs for leaf area (A). Traits were explained in Table 1. (The results of total yield and total biomass were based on the data of Tsutsumi-Morita et al., 2021 as references.)

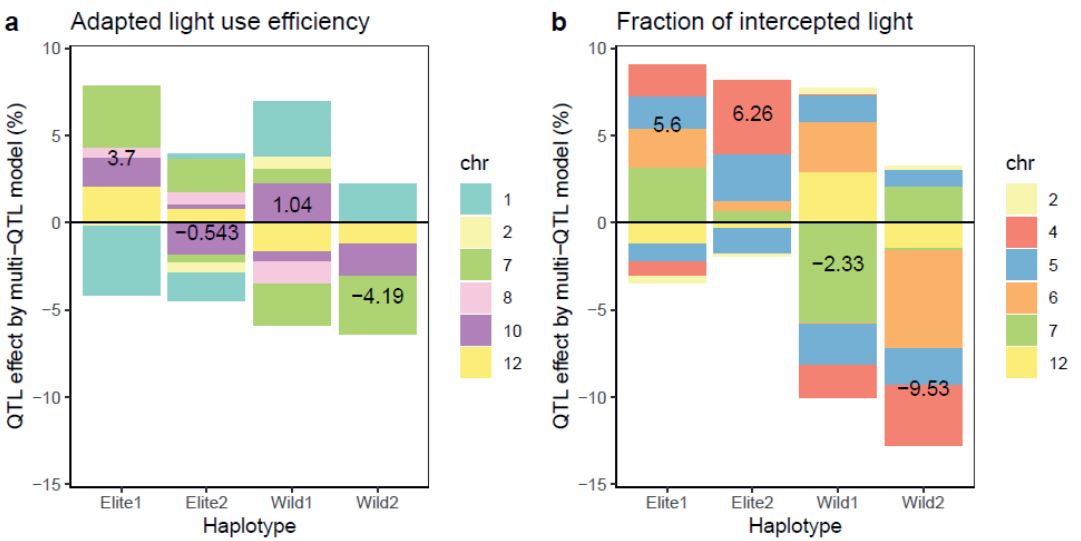


Figure 6 Haplotype effects per chromosome and multi-locus predictions for the parents. Haplotype effects are coloured by chromosome. Numbers on the bars and positions along the y-axis indicate the value for the multi-locus prediction. Traits were explained in Table 1.

4.4. Discussion

To our knowledge, this is the first study to report the genetic basis of leaf area index (LAI), the fraction of intercepted light (F_{int}), and adapted light use efficiency (LUE_{adp}) in tomato. We evaluated the effect of these traits on total biomass and yield by phenotypic and QTL analyses.

4.4.1. No trade-off between adapted light use efficiency (LUE_{adp}) and the fraction of intercepted light (F_{int}) to improve total biomass

Adapted light use efficiency (LUE_{adp}) and the fraction of intercepted light (F_{int}) did not show a correlation (r = 0.03, not significant; Figure 3; Supplementary Figure 3). They shared only one colocalized QTL with the same haplotype effect (LUE_{adp}7.2 and F_{int}7.1) and only one QTL at the same position with different allele effects (LUE_{adp}12.1 and F_{int}12.1). Thus, most QTLs for LUE_{adp} and F_{int} were located separately. LUE_{adp} and F_{int} potentially have positive or negative relationships (Yin *et al.*, 2022). In field crops, the variation of maximum photosynthetic capacity may be related to leaf thickness or nitrogen content (Mathan *et al.*, 2021). An increase in photosynthetic capacity by increasing leaf thickness, however, is not desirable (Austin, 1989) because of a trade-off with leaf expansion, which potentially reduces whole-plant photosynthesis (Richards, 2000; Yin *et al.*, 2022). We did not observe this trade-off between LUE_{adp} and F_{int} because SLA had no correlation and no QTL at the same position with leaf area, LAI or F_{int}. No correlation between SLA and leaf area was also reported in an introgression line population of determinate tomato (Muir *et al.*, 2014).

On the other hand, a high LUE_{adp} could produce more biomass and therefore might increase leaf area, LAI and F_{int}. Then, a positive relationship between LUE_{adp} and F_{int} might be observed (Yin *et al.*, 2022). For example, improving photosynthesis by genetic modification resulted in greater leaf area or larger plants (Kromdijk *et al.*, 2016; Driever *et al.*, 2017). We observed a RIL with higher LUE_{adp} tends to show a slightly higher leaf dry weight (LDW) (r = 0.29 for LUE_{adp} and LDW ; 1 QTL colocalization, LUE_{adp}7.2 and LDW7.1), a slightly lower SLA (r = −0.22 for LUE_{adp} and SLA1.2; LUE_{adp}1.2 and SLA1.2 at the same position with wild parent 1 increasing LUE_{adp} and decreasing SLA), and a slightly higher leaf area (r = 0.24 for LUE_{adp} and leaf area; 0 QTL colocalization). However, no positive correlation was found between LUE_{adp} and F_{int} due to no correlation of SLA with F_{int}. No trade-off between LUE_{adp} and F_{int}, and mostly unlinked QTL locations between LUE_{adp} and F_{int} QTLs, make it possible to combine positive alleles for LUE_{adp} and F_{int} to improve total biomass.

4.4.2. Selecting for adapted light use efficiency (LUE_{adp}) is a more successful strategy to obtain a high yielding line than selecting for fraction of intercepted light (F_{int})

Compared to LUE_{adp}, which showed a high potential to improve yield, F_{int} did not show a positive correlation with yield (r = 0.26). One might think that this is because F_{int} is already high and almost all light is intercepted in all RILs. However, F_{int} showed a wide distribution: ranged from 0.62 to 0.97 for all RILs, and from 0.83 to 0.97 for high-yielding RILs. F_{int} correlated with total biomass (r = 0.69), and many F_{int} QTLs colocalized with total-biomass QTLs. Less effect of F_{int} on yield comes from a trade-off between F_{int} and harvest index (HI) (r = −0.71; Supplementary Figure 3). That is, the biomass allocated to the leaves cannot be used for fruits. No effect of SLA on leaf area (r = −0.07, not significant) produced this clear trade-off between F_{int} and harvest index. Thus, F_{int} increased total biomass but decreased harvest index, resulting in less positive effects of F_{int} on yield (r = 0.26). On the other hand, LUE_{adp} did not show a trade-off with harvest index (r = −0.15). Thus, positive effects of LUE_{adp} on total biomass were taken over to positive effects on yield.

The importance of LUE for high-yielding breeding was suggested by previous studies (Van Der Ploeg *et al.*, 2007; Higashide and Heuvelink, 2009) comparing several Dutch cultivars released in different years. The yield of modern tomato cultivars was on average 40% higher than yield of 'Moneymaker' released in 1950. This increase in production resulted from a higher LUE under an assumption of a fixed light extinction coefficient of k = 0.75 (Van Der Ploeg *et al.*, 2007). Another study compared eight Dutch cultivars, where the light interception was directly measured, showing that yield improvement over the past 50 years was attributed to total biomass production by LUE (Higashide and Heuvelink, 2009). Thus, LUE has been improved to increase yield in tomato in the past. In this study, we showed that LUE_{adp} is still an important target trait to increase yield. LUE_{adp} showed a large variation in our population, and LUE_{adp} of the current elite parents can be improved by introgression of wild genomes. We detected LUE_{adp} QTLs, which will be useful to increase total biomass and yield.

4.4.3. Adapted light use efficiency (LUE_{adp}) includes the effect of light extinction coefficient *k*

One caution is that the light extinction coefficient k was fixed for F_{int} calculation (Eq. 7) in our analysis, whereas k is most likely genotype dependent (Higashide and Heuvelink, 2009). Because LUE_{adp} was derived from total biomass divided by F_{int}, the genetic effect of k was included in the variation in LUE_{adp} in this study. k is determined by morphological traits (e.g., leaf shape, leaf angle distribution, and internode length) as well as leaf optical properties (chlorophyll content) (Sarlikioti *et al.*, 2011; Yin *et al.*, 2022). A higher k increases light interception at the same LAI; however, it reduces light penetration through the canopy, which

reduces total biomass production because lower leaves cannot receive enough light whereas upper leaves are light saturated (Yin *et al.*, 2022). A previous study comparing eight Dutch tomato cultivars reported that modern cultivars had lower *k*, which means light penetrates deeper in the canopy, resulting in higher total biomass (Higashide and Heuvelink, 2009). Simulation of tomato breeding by genomic prediction in 96 big-fruited tomato varieties revealed that selection for yield and flavour would result in morphological changes, which were related to the increase in total biomass (Yamamoto *et al.*, 2016). In our study, most of RILs showed narrower leaves (larger proportionality factor *c*) than elite parents. Narrower leaves allowed deeper light penetration through the canopy, which reduces *k*. Deeper light penetration through the canopy was expected to increase canopy photosynthesis and total biomass (Sarlikioti *et al.*, 2011). In this study, two QTLs for LUE_{adp} might represent the effect of *k* because LUE_{adp}2.1 colocalized with Proportionality-factor-c2.1, and LUE_{adp}10.1 was at the same position with Leaf-width10.1 and Parameter-a10.1 (asymptote of leaf width model). In both cases, wild parent 1 increased LUE_{adp} and at the same time produced narrower-shaped leaves or decreased leaf width. For the other LUE_{adp} QTLs, further research is necessary to elucidate whether these LUE_{adp} QTLs are related to *k*, and/or photosynthesis.

4.4.4. Appropriate leaf area index (LAI) and light extinction coefficient *k* might be environment dependent

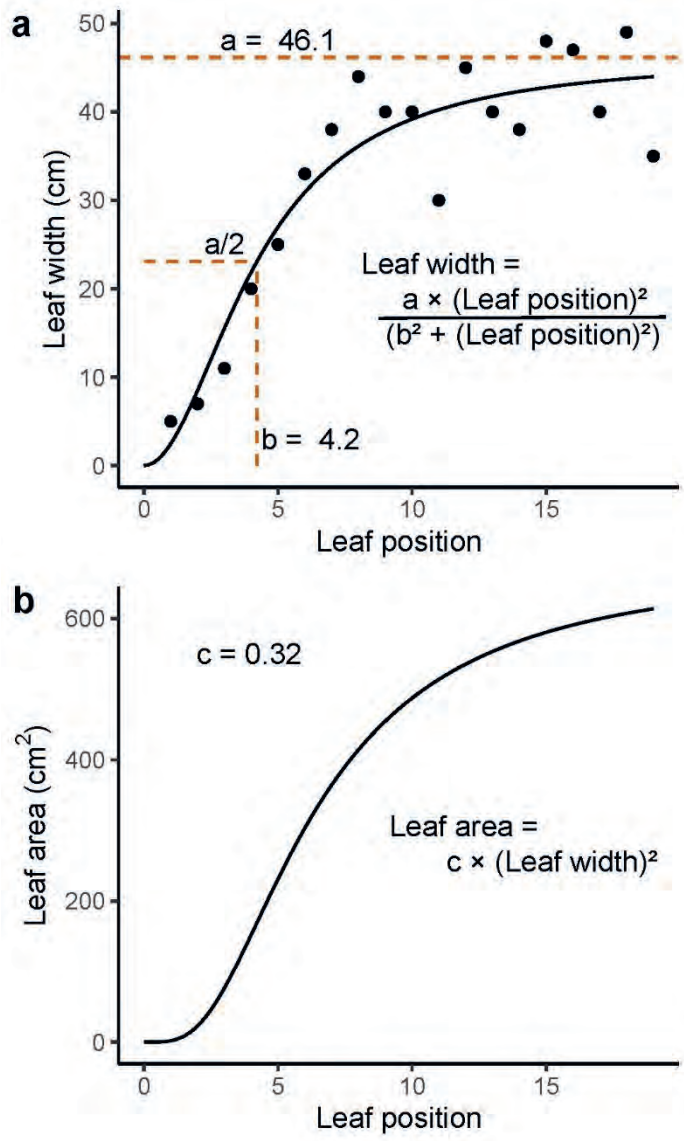
In this study, LAI showed a linear relationship with total biomass, at least until LAI = 4 (Figure 4a). The optimal LAI is light-intensity dependent; under higher light intensity, higher LAI is recommended (Saito *et al.*, 2020). Our experiment was conducted in a typical high-tech greenhouse in the Netherlands with a high transmissivity of light. Thus, we expected the optimal LAI should be higher than other types of greenhouses with a lower transmissivity in the same region. Under a higher LAI, a lower *k* will achieve a more uniform light distribution, which is beneficial for whole-plant photosynthesis because photosynthesis may be light-saturated in the upper canopy whereas not saturated in the lower canopy (Yin *et al.*, 2022). Therefore, in this study, a smaller *k* (a deeper light penetration) seems beneficial. Under a lower light environment, the optimum LAI will become lower (Saito *et al.*, 2020); therefore, optimum *k* will be higher to keep a high fraction of intercepted light (F_{int}). In such a case, a different genotype might be appropriate to achieve high total biomass. This genotype × environment interaction still needs to be investigated.

4.5. Conclusion

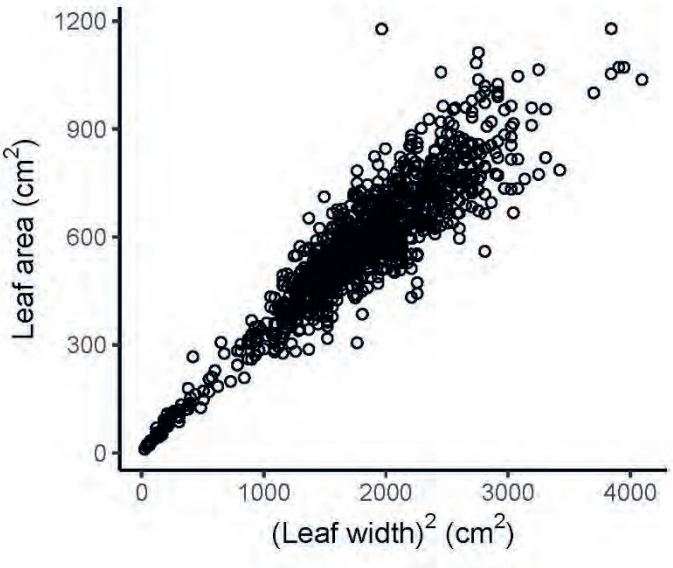
This study aimed to find QTLs for adapted light use efficiency (LUE_{adp}), the fraction of intercepted light (F_{int}), and leaf area index (LAI) to assess which of these QTLs are shared with total (plant) biomass and fruit yield (yield), and to discuss how these detected QTLs could be combined to breed high-yielding lines. We detected several QTLs for these traits,

for which wild germplasm had both positive and negative effects. The phenotypic variation in total biomass was explained equally by variation in F_{int} and variation in LUE_{adp}, and 4 QTLs for LUE_{adp} and 7 QTLs for F_{int} or LAI colocalized with QTLs for total biomass. F_{int} and LUE_{adp} were not correlated with each other and were largely controlled by different QTLs. Therefore, total biomass can be improved by F_{int} and/or LUE_{adp} by combining favourable alleles for these two traits. Although F_{int} contributed to improving total biomass, the positive effect of F_{int} on total biomass resulted in a less positive effect of F_{int} on yield, due to the negative correlation between F_{int} and harvest index because leaf area was mainly determined by leaf dry weight but not by specific leaf area. Consequently, selecting for LUE_{adp} is a more successful strategy than selecting for F_{int} for high-yielding breeding. Wild germplasm brought much improvement in total biomass via QTLs for LUE_{adp} and QTLs for F_{int}, which opened the way to improve the yield of the elite parent.

Supporting information

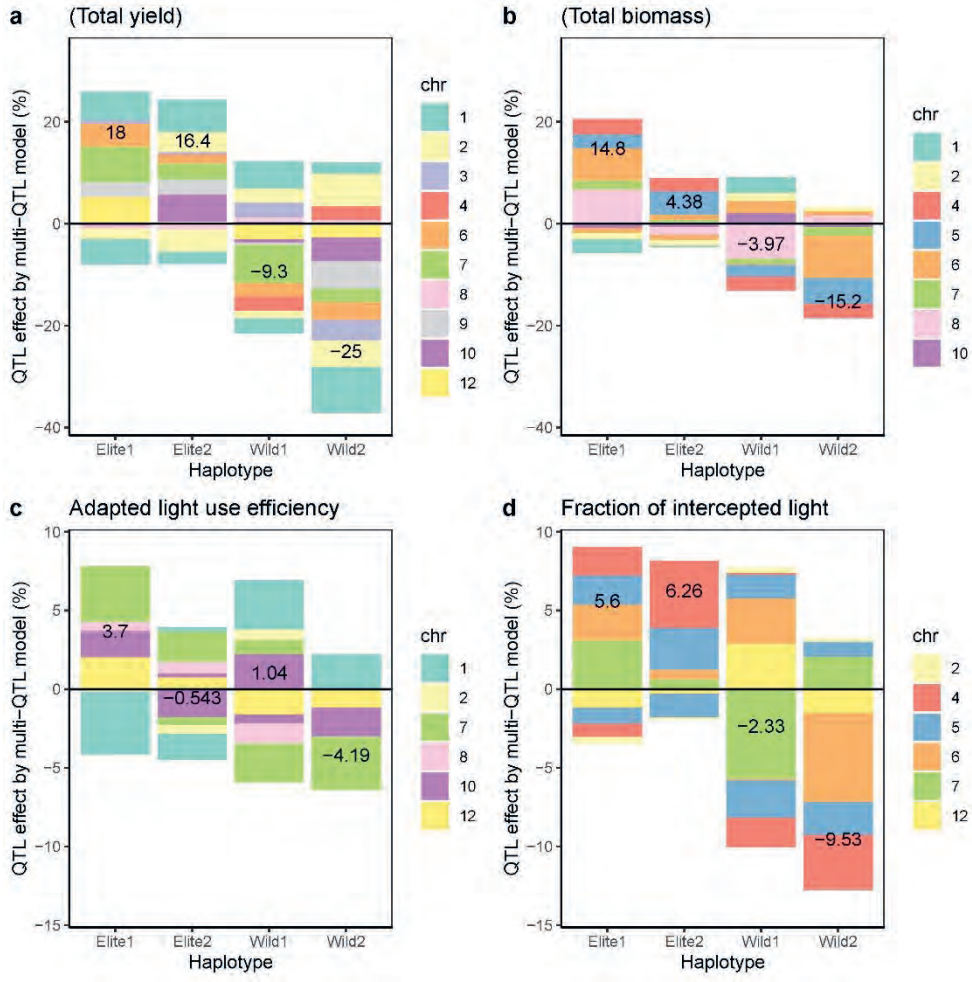


Supplementary Figure 1 Example of leaf width model (a) and leaf area model (b) in a plot in spring.

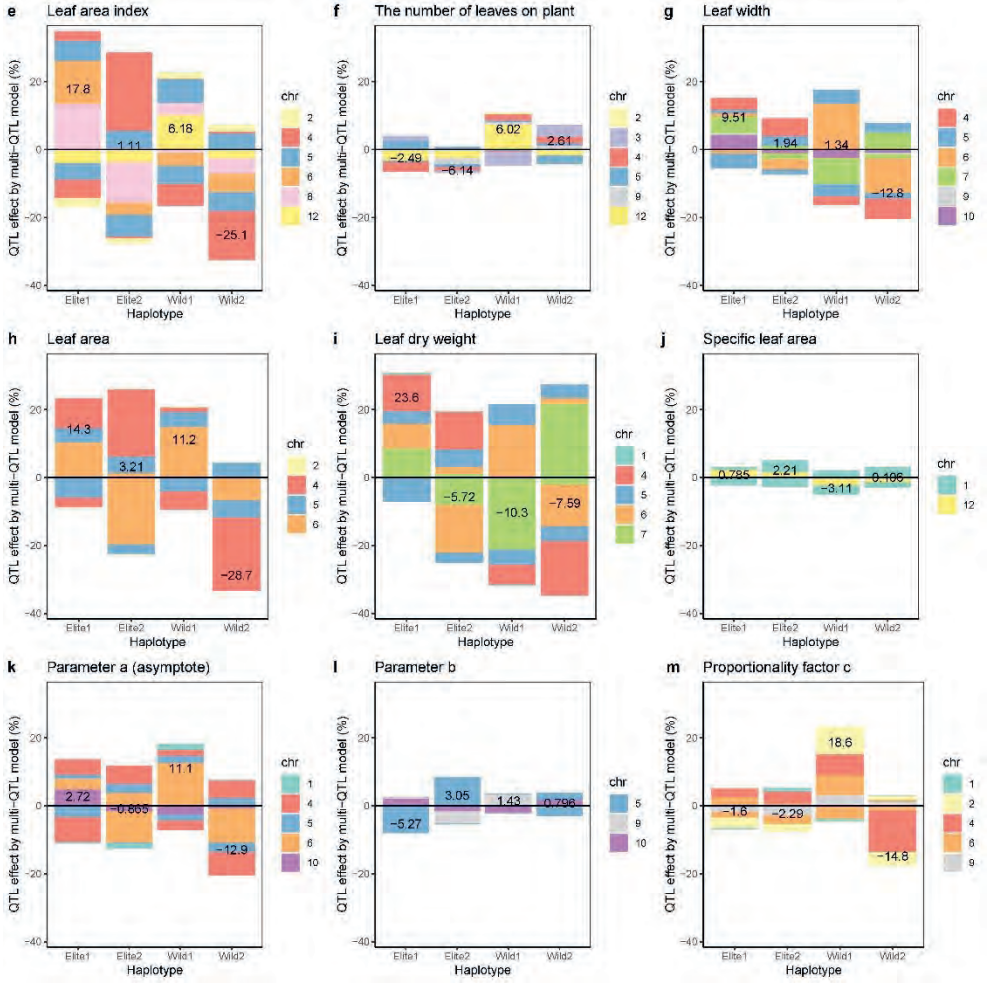


Supplementary Figure 2 Leaf area plotted against the square of leaf width using 78 leaves from random sampling and the bottom three leaves for all plots in spring.

*Supplementary Figure 3 Phenotypic Pearson correlations among traits by the best linear unbiased estimates (BLUEs) per RIL. Ytotal, total yield; Biomass, total biomass; HI, harvest index; LUE_{adp}, adapted light use efficiency; F_{int}, fraction of intercepted light (based on leaf area index); LAI, leaf area index; LNplant, the number of leaves on plant; Lwidth, leaf width; LA, leaf area; LDW, leaf dry weight; SLA, specific leaf area; a_asymp, parameter a (asymptote); b, parameter b; c_prop, proportionality factor c. Traits were explained in Table 1. *significant correlation at $P < 0.05$. (Data of total yield, total biomass and harvest index were based on Tsutsumi-Morita et al., 2021 as references.)*



Supplementary Figure 4 Haplotype effects per chromosome and multi-locus predictions for the parents. Haplotype effects have the colour of the chromosome. Numbers on the bars and positions along the y-axis indicate the value for the multi-locus prediction. Traits were explained in Table 1. Data of total yield (a) and total biomass (b) were based on the previous study (Tsutsumi-Morita et al., 2021) as references. Data of adapted light use efficiency (c) and the fraction of intercepted light (d) were added from Figure 6 for convenience of comparison.



Supplementary Figure 4 Continued.

Supplementary Table 1 QTLs list. QTL positions, SolCAP marker names, -log$_{10}$(p) values for test on variance component of haplotype effects, and haplotype effects. Traits were explained in Table 1.

						QTL effect from multi-QTL model			
Trait	Chromosome	cM	Position (bp)	SolCAP.marker	-log$_{10}$(*p*)	Elite parent 1	Elite parent 2	Wild parent 1	Wild parent 2
(Total yield)									
	1	25.6	2,879,156	solcap_snp_sl_60089	6.9	-1.31	-0.10	0.32	1.09
	1	38.8	4,940,680	solcap_snp_sl_33676	8.2	2.71	3.11	-1.46	-4.36
	1	53.2	79,245,417	solcap_snp_sl_25931	6.2	-1.07	-0.38	1.50	-0.06
	1	77.5	84,839,564	solcap_snp_sl_63558	4.3	-0.05	-0.73	0.75	0.03
	2	11	36,167,998	solcap_snp_sl_15706	4.2	-0.73	1.92	1.34	-2.53
	2	29	40,136,280	solcap_snp_sl_10552	5.9	-0.25	-2.16	-0.71	3.12
	3	25	2,704,993	solcap_snp_sl_14355	3.1	NA	NA	NA	NA
	3	32.2	3,540,929	solcap_snp_sl_19514	3.1	0.36	0.17	1.46	-1.99
	4	39.4	7,963,140	solcap_snp_sl_45951	6.2	0.02	0.09	-1.38	1.28
	6	42.6	42,434,661	CL017618-0174	5.7	2.15	0.83	-1.28	-1.70
	7	23.2	2,881,365	solcap_snp_sl_15786	7.7	1.29	1.13	-1.77	-0.64
	7	47.6	61,442,206	solcap_snp_sl_53538	7.9	2.12	0.39	-1.89	-0.63
	7	80.2	67,162,714	solcap_snp_sl_70636	3.6	NA	NA	NA	NA
	8	18	2,587,919	solcap_snp_sl_14531	2.9	-0.29	-0.50	0.41	0.38
	8	55	60,545,538	solcap_snp_sl_64736	5.5	NA	NA	NA	NA
	9	108.2	70,828,802	solcap_snp_sl_22327	3.0	1.45	1.42	-0.23	-2.64
	10	6.8	1,395,637	solcap_snp_sl_46339	6.9	0.00	0.17	0.17	-0.34
	10	21	3,092,679	SGN-U603133_snp167	6.2	-0.17	2.44	-0.34	-1.93
	10	76	65,126,274	Le013158s_161	3.3	NA	NA	NA	NA

	12	15	1,971,483	solcap_snp_sl_12656	7.5	NA	NA	NA	NA
	12	55.7	63,393,430	solcap_snp_sl_55530	7.5	2.55	0.19	-1.43	-1.31
(Total biomass)									
	1	0.5	310,894	CL004303-0524	4.0	NA	NA	NA	NA
	1	18	2,063,522	solcap_snp_sl_33701	3.0	NA	NA	NA	NA
	1	53.2	79,245,417	solcap_snp_sl_25931	3.2	-0.112	-0.015	0.129	-0.002
	2	23.4	38,960,460	Le001778_68_solcap_snp_sl_33474	6.3	-0.055	-0.046	0.063	0.038
	4	71	61,874,267	solcap_snp_sl_3687	8.8	0.123	0.108	-0.114	-0.118
	4	80.4	63,919,364	solcap_snp_sl_47385	8.6	NA	NA	NA	NA
	5	43.6	6,571,837	solcap_snp_sl_50925	9.0	0.113	0.194	-0.098	-0.210
	6	6.3	2,600,146	solcap_snp_sl_35072	4.0	-0.039	-0.045	0.049	0.035
	6	17.3	37,583,093	solcap_snp_sl_55858	4.5	0.114	0.001	-0.011	-0.103
	6	46	43,804,153	solcap_snp_sl_57810	3.6	0.152	0.025	0.011	-0.188
	6	59.6	47,291,843	solcap_snp_sl_57155	3.3	0.000	0.013	0.039	-0.052
	7	23.2	2,881,365	solcap_snp_sl_15786	5.6	0.009	0.017	-0.014	-0.012
	7	42	59,724,392	solcap_snp_sl_5862	5.4	NA	NA	NA	NA
	7	47.6	61,442,206	solcap_snp_sl_53538	5.2	0.063	0.014	-0.022	-0.056
	8	46	58,745,796	solcap_snp_sl_15381	3.3	0.275	-0.059	-0.282	0.066
	10	52.9	62,277,373	solcap_snp_sl_59337	4.2	-0.033	-0.027	0.086	-0.027
	12	62.8	64,157,980	solcap_snp_sl_19630	3.5	NA	NA	NA	NA
Adapted light use efficiency									
	1	27.3	3,101,943	solcap_snp_sl_60043	3.7	-0.060	-0.084	0.039	0.105
	1	53.2	79,245,417	solcap_snp_sl_25931	4.1	-0.142	0.014	0.122	0.006
	2	7	35,704,585	solcap_snp_sl_22259	4.8	-0.008	-0.029	0.036	0.001
	4	90	65,402,987	solcap_snp_sl_3894	3.4	NA	NA	NA	NA

7	23.2	2,881,365	solcap_snp_sl_15786	4.2	0.069	0.098	-0.125	-0.042
7	42	59,724,392	solcap_snp_sl_5862	4.4	0.112	-0.025	0.043	-0.130
8	42	57,915,097	solcap_snp_sl_48398	4.1	0.030	0.037	-0.066	-0.001
10	6.8	1,395,637	solcap_snp_sl_46339	3.4	0.067	-0.090	0.114	-0.090
10	21	3,092,679	SGN-U603133_snp167	3.3	0.018	0.013	-0.028	-0.002
11	18	2,443,617	solcap_snp_sl_10596	3.5	NA	NA	NA	NA
12	62.8	64,157,980	solcap_snp_sl_19630	4.0	0.103	0.037	-0.081	-0.059
Fraction of intercepted light								
1	14	1,809,372	solcap_snp_sl_60276	6.3	NA	NA	NA	NA
2	29	40,136,280	solcap_snp_sl_10552	3.3	-0.0035	-0.0013	0.0030	0.0019
2	33.5	40,887,642	solcap_snp_sl_21744	3.6	NA	NA	NA	NA
4	50.9	57,883,659	solcap_snp_sl_2179	5.0	NA	NA	NA	NA
4	62	60,271,155	solcap_snp_sl_11520	8.8	NA	NA	NA	NA
4	71	61,874,267	solcap_snp_sl_3687	11.1	0.0151	0.0075	-0.0156	-0.0071
4	80.4	63,919,364	solcap_snp_sl_47385	11.3	-0.0068	0.0279	0.0007	-0.0217
5	42.8	6,247,585	CL016061-0368	9.0	0.0151	0.0216	-0.0195	-0.0172
5	76	63,803,770	solcap_snp_sl_185	5.7	-0.0083	-0.0121	0.0125	0.0079
6	6.3	2,600,146	solcap_snp_sl_35072	5.0	NA	NA	NA	NA
6	17.3	37,583,093	solcap_snp_sl_55858	3.9	0.0022	0.0014	-0.0009	-0.0027
6	59.6	47,291,843	solcap_snp_sl_57155	5.3	0.0163	0.0038	0.0235	-0.0436
7	47.6	61,442,206	solcap_snp_sl_53538	3.1	0.0028	0.0009	-0.0031	-0.0007
7	72	65,631,063	solcap_snp_sl_70992	3.4	0.0226	0.0041	-0.0434	0.0167
8	42	57,915,097	solcap_snp_sl_48398	3.7	NA	NA	NA	NA
10	27	3,991,802	solcap_snp_sl_34373	3.1	NA	NA	NA	NA
12	54	63,006,395	solcap_snp_sl_42740	6.7	-0.0095	-0.0024	0.0236	-0.0117

	12	57.6	63,580,692	solcap_snp_sl_31953	5.6	NA	NA	NA	NA
	12	81	66,256,781	SGN-U603632_snp167	7.9	NA	NA	NA	NA
Leaf area index									
	1	6	1,116,348	solcap_snp_sl_20499	6.8	NA	NA	NA	NA
	2	29	40,136,280	solcap_snp_sl_10552	3.9	-0.0755	-0.0390	0.0566	0.0579
	2	33.5	40,887,642	solcap_snp_sl_21744	3.8	NA	NA	NA	NA
	4	50.9	57,883,659	solcap_snp_sl_2179	4.4	-0.0068	-0.0128	0.0045	0.0151
	4	62	60,271,155	solcap_snp_sl_11520	8.7	NA	NA	NA	NA
	4	71	61,874,267	solcap_snp_sl_3687	11.9	0.0788	0.0372	-0.0889	-0.0272
	4	80.4	63,919,364	solcap_snp_sl_47385	11.3	-0.1401	0.5994	-0.0920	-0.3674
	5	21	2,662,318	solcap_snp_sl_49084	3.0	NA	NA	NA	NA
	5	42.8	6,247,585	CL016061-0368	8.8	0.1609	0.1391	-0.1470	-0.1530
	5	76	63,803,770	solcap_snp_sl_185	6.2	-0.1356	-0.1814	0.1884	0.1286
	6	6.3	2,600,146	solcap_snp_sl_35072	6.1	NA	NA	NA	NA
	6	17.3	37,583,093	solcap_snp_sl_55858	5.1	0.3470	-0.0954	-0.1020	-0.1497
	6	52.8	45,035,130	CL015545-0057_solcap_snp_sl_57521	4.5	NA	NA	NA	NA
	7	47.6	61,442,206	solcap_snp_sl_53538	3.1	NA	NA	NA	NA
	8	1	230,545	CL016245-0267	3.5	0.3245	-0.3265	0.0969	-0.0949
	8	42	57,915,097	solcap_snp_sl_48398	3.9	0.0437	0.0098	-0.0253	-0.0282
	12	54	63,006,395	solcap_snp_sl_42740	7.0	-0.1087	-0.1001	0.2780	-0.0692
	12	57.6	63,580,692	solcap_snp_sl_31953	4.5	NA	NA	NA	NA
	12	81	66,256,781	SGN-U603632_snp167	6.0	NA	NA	NA	NA
The number of leaves on plant									
	1	20	2,441,831	SL20259_223	7.3	NA	NA	NA	NA
	2	39.8	43,252,729	CL016172-0398_solcap_snp_sl_29710	3.2	-0.016	-0.026	0.105	-0.063

	3	40	4,981,481	solcap_snp_sl_16433	3.5	NA	NA	NA	NA
	3	96.4	-	SL10678_608	3.0	0.345	-0.163	-0.936	0.754
	4	35	5,295,726	solcap_snp_sl_41609	3.2	-0.155	-0.088	-0.014	0.257
	4	53	58,290,645	solcap_snp_sl_43569	3.1	-0.477	-0.092	0.415	0.154
	4	90	65,402,987	solcap_snp_sl_3894	3.5	NA	NA	NA	NA
	5	42.8	6,247,585	CL016061-0368	7.5	0.518	0.154	-0.106	-0.566
	5	64	62,533,914	solcap_snp_sl_12213	4.2	-0.080	-0.204	0.111	0.173
	9	15	1,514,198	solcap_snp_sl_7731	3.7	NA	NA	NA	NA
	9	25	2,673,326	solcap_snp_sl_39806	4.5	NA	NA	NA	NA
	9	41.2	4,543,446	solcap_snp_sl_39525	6.5	0.032	-0.404	0.135	0.237
	12	2	715,204	CL015154-0161_solcap_snp_sl_45733	4.4	-0.072	-0.489	0.933	-0.373
	12	54	63,006,395	solcap_snp_sl_42740	11.2	-0.646	-0.046	0.688	0.003
Leaf width									
	1	6	1,116,348	solcap_snp_sl_20499	5.0	NA	NA	NA	NA
	2	29	40,136,280	solcap_snp_sl_10552	4.2	NA	NA	NA	NA
	4	50.9	57,883,659	solcap_snp_sl_2179	4.5	NA	NA	NA	NA
	4	71	61,874,267	solcap_snp_sl_3687	16.7	1.09	0.70	-0.81	-0.98
	4	80.4	63,919,364	solcap_snp_sl_47385	16.2	0.45	1.68	-0.38	-1.74
	5	14.7	2,152,107	solcap_snp_sl_49215	3.6	-0.14	-0.11	0.81	-0.57
	5	42.8	6,247,585	CL016061-0368	5.2	0.47	1.29	-1.56	-0.20
	5	76	63,803,770	solcap_snp_sl_185	4.8	-1.84	-0.51	1.10	1.25
	6	13.4	36,268,579	solcap_snp_sl_68871	7.8	-0.53	-1.25	3.92	-2.14
	6	59.6	47,291,843	solcap_snp_sl_57155	4.7	0.44	-0.12	2.04	-2.37
	7	23.2	2,881,365	solcap_snp_sl_15786	3.6	NA	NA	NA	NA
	7	53	62,093,144	solcap_snp_sl_53425	3.4	0.24	0.41	-0.05	-0.60

	7	80.2	67,162,714	solcap_snp_sl_70636	4.1	2.00	-0.79	-3.40	2.19
	9	15	1,514,198	solcap_snp_sl_7731	4.3	0.15	0.03	0.00	-0.18
	10	6.8	1,395,637	solcap_snp_sl_46339	4.1	1.92	-0.47	-1.07	-0.37
	12	84	66,885,302	solcap_snp_sl_31342	3.4	NA	NA	NA	NA
Leaf area									
	1	8.6	1,470,863	solcap_snp_sl_8704	4.3	NA	NA	NA	NA
	2	29	40,136,280	solcap_snp_sl_10552	3.4	-1.2	-1.2	1.1	1.3
	4	50.9	57,883,659	solcap_snp_sl_2179	5.3	NA	NA	NA	NA
	4	71	61,874,267	solcap_snp_sl_3687	17.4	47.9	11.1	-29.8	-29.2
	4	80.4	63,919,364	solcap_snp_sl_47385	17.4	-15.8	97.6	6.9	-88.8
	5	21	2,662,318	solcap_snp_sl_49084	3.1	NA	NA	NA	NA
	5	42.8	6,247,585	CL016061-0368	5.0	23.3	26.3	-22.0	-27.5
	5	76	63,803,770	solcap_snp_sl_185	4.0	-32.1	-16.2	24.4	23.9
	6	6.3	2,600,146	solcap_snp_sl_35072	4.9	NA	NA	NA	NA
	6	39.1	41,994,425	solcap_snp_sl_27215	4.0	44.3	-107.1	72.3	-9.5
	6	59.6	47,291,843	solcap_snp_sl_57155	5.8	11.6	7.0	8.4	-27.1
	12	84	66,885,302	solcap_snp_sl_31342	3.6	NA	NA	NA	NA
Leaf dry weight									
	1	9	1,511,380	solcap_snp_sl_60330	5.3	0.02	0.00	-0.02	-0.01
	2	29	40,136,280	solcap_snp_sl_10552	3.3	NA	NA	NA	NA
	4	50.9	57,883,659	solcap_snp_sl_2179	3.6	NA	NA	NA	NA
	4	71	61,874,267	solcap_snp_sl_3687	15.4	0.23	0.01	-0.09	-0.15
	4	80.4	63,919,364	solcap_snp_sl_47385	16.3	0.14	0.38	-0.12	-0.41
	5	42.8	6,247,585	CL016061-0368	4.0	0.13	0.18	-0.16	-0.15
	5	76	63,803,770	solcap_snp_sl_185	4.5	-0.25	-0.11	0.22	0.14

	6	6.3	2,600,146	solcap_snp_sl_35072	4.4	NA	NA	NA	NA
	6	39.1	41,994,425	solcap_snp_sl_27215	3.0	0.13	-0.50	0.31	0.06
	6	59.6	47,291,843	solcap_snp_sl_57155	5.2	0.13	0.08	0.23	-0.44
	7	53	62,093,144	solcap_snp_sl_53425	3.3	0.03	0.03	0.00	-0.07
	7	80.2	67,162,714	solcap_snp_sl_70636	4.8	0.26	-0.28	-0.74	0.76
Specific leaf area									
	1	0.5	310,894	CL004303-0524	5.9	-3.55	-4.59	3.20	4.94
	1	38.8	4,940,680	solcap_snp_sl_33676	3.5	1.28	5.73	-4.41	-2.60
	12	20	2,494,987	solcap_snp_sl_41068	3.1	3.55	2.47	-3.84	-2.17
Parameter a (asymptote)									
	1	6	1,116,348	solcap_snp_sl_20499	5.8	NA	NA	NA	NA
	1	79	85,136,490	solcap_snp_sl_9756	3.1	-0.20	-0.78	0.91	0.06
	4	3.9	776,552	solcap_snp_sl_9838	3.2	-3.69	0.21	0.88	2.59
	4	44	53,605,655	CL015899-0460_solcap_snp_sl_42811	4.5	0.02	0.04	-0.01	-0.05
	4	71	61,874,267	solcap_snp_sl_3687	15.4	2.19	2.17	-1.22	-3.13
	4	80.4	63,919,364	solcap_snp_sl_47385	14.9	0.12	0.29	-0.19	-0.23
	5	21	2,662,318	solcap_snp_sl_49084	3.0	0.29	-0.01	0.41	-0.69
	5	42.8	6,247,585	CL016061-0368	4.7	0.26	1.35	-0.90	-0.71
	5	76	63,803,770	solcap_snp_sl_185	4.7	-1.63	-0.14	0.64	1.13
	6	3.3	1,579,723	solcap_snp_sl_15516	5.2	NA	NA	NA	NA
	6	13.4	36,268,579	solcap_snp_sl_68871	5.4	0.14	-0.29	1.72	-1.57
	6	39.1	41,994,425	solcap_snp_sl_27215	3.2	0.90	-4.31	3.67	-0.26
	6	59.6	47,291,843	solcap_snp_sl_57155	4.7	0.50	1.85	0.92	-3.27
	7	23.2	2,881,365	solcap_snp_sl_15786	3.3	NA	NA	NA	NA
	9	15	1,514,198	solcap_snp_sl_7731	3.3	NA	NA	NA	NA

10	6.8	1,395,637	solcap_snp_sl_46339	5.2	2.46	-0.81	-1.30	-0.35
12	81	66,256,781	SGN-U603632_snp167	3.0	NA	NA	NA	NA
Parameter b								
5	48	8,354,784	solcap_snp_sl_51094	4.0	-0.320	0.460	-0.003	-0.137
5	58	61,885,368	solcap_snp_sl_22622	4.8	-0.121	-0.013	0.013	0.121
9	20	2,052,762	solcap_snp_sl_39886	3.8	-0.001	-0.048	0.067	-0.018
9	50	4,706,233	CL016894-0326_solcap_snp_sl_39506	4.2	0.041	-0.146	0.115	-0.011
10	3	1,163,895	solcap_snp_sl_13202	3.3	0.111	-0.086	-0.114	0.088
Proportionality factor c								
1	83.5	86,549,187	solcap_snp_sl_2218	3.0	-0.0009	0.0030	-0.0023	0.0002
2	7	35,704,585	solcap_snp_sl_22259	4.8	-0.0063	-0.0012	0.0102	-0.0027
2	23.4	38,960,460	Le001778_68_solcap_snp_sl_33474	4.0	-0.0028	-0.0013	0.0013	0.0029
2	37.7	42,435,573	solcap_snp_sl_29545	7.5	0.0008	-0.0038	0.0124	-0.0093
4	3.9	776,552	solcap_snp_sl_9838	4.5	0.0072	0.0101	0.0178	-0.0351
6	1	1,111,897	solcap_snp_sl_22860	3.1	NA	NA	NA	NA
6	13.4	36,268,579	solcap_snp_sl_68871	3.0	0.0073	0.0021	-0.0112	0.0018
6	42.6	42,434,661	CL017618-0174	3.4	-0.0050	-0.0074	0.0166	-0.0042
9	62	7,156,689	solcap_snp_sl_45141	3.6	NA	NA	NA	NA
9	69	63,812,814	solcap_snp_sl_43102	4.8	-0.0049	-0.0080	0.0092	0.0037

Chapter 5

Quantitative trait loci (QTLs) for photosynthesis-related traits to understand and predict yield in tomato

Yutaka Tsutsumi-Morita[1,2], Ep Heuvelink[1], Jan F. H. Snel[3], Martin P. Boer[2], Sedighehsadat Khaleghi[1], Frank F. Millenaar[4], Daniela Bustos-Korts[2], George A.K. Van Voorn[2], Leo F.M. Marcelis[1], Fred A. Van Eeuwijk[2],

[1] Wageningen University & Research, Horticulture and Product Physiology, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[2] Wageningen University & Research, Biometris, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[3] Adviesbureau JFH Snel, Wageningen, The Netherlands
[4] BASF's vegetable seeds business (Nunhems), Napoleonsweg 152, 6083 AB Nunhem, The Netherlands

Abstract

Despite the fundamental importance of photosynthesis for biomass production, it has been a long-standing discussion whether photosynthesis is a useful target trait for plant breeding to improve yield. The present study aimed to (1) identify quantitative trait loci (QTLs) for photosynthesis-related traits and (2) reveal the possible role of photosynthesis in improving yield in indeterminate tomato for greenhouse production, by using hybrids from 2 testers × 342 recombinant inbred lines originating from 4 parents (2 elite and 2 wild parents). We detected 10 QTLs for chlorophyll content index and 0 to 7 QTLs for each of the 14 traits measured by chlorophyll fluorescence. Photosynthesis-related QTLs were often at the same position as QTLs for yield, total plant biomass and/or light use efficiency. However, allele effects were not always consistent between yield and photosynthesis-related traits, which may explain why we observed the absence of phenotypic correlation between yield and photosynthesis-related traits. Nevertheless, yield was predicted from QTLs for all photosynthesis-related traits together with a good prediction accuracy (correlation between observation and prediction; 0.74 for fitted model; 0.60 for median in cross-validation). Several candidate genes related to electron transport or photosynthesis were found. This study demonstrates that wild relatives harbour beneficial alleles to improve the photosynthesis-related traits in the genetic background of elite parents. The variation in yield can be predicted with good accuracy from photosynthesis-related QTLs, which is an invaluable tool for high-yielding breeding.

List of abbreviations

Trait	Abbreviation
(Total fruit yield)	(Yield)
(Total dry biomass)	(Biomass)
(Adapted light use efficiency)	(LUEadp)
Photosynthesis-related traits	
Chlorophyll content index (measured by SPAD)	Chlorophyll content, (ChlC)
Traits measured by chlorophyll fluorescence (ChlF)	
F_v/F_m	F_v/F_m
Rapid light response curve (rLRC) related	
Parameter alpha (initial slope) of rLRC	Slope of rLRC
Parameter theta (convexity) of rLRC	Convexity of rLRC
Parameter J_{max} of rLRC	J_{max} of rLRC
ETR at PPFD = 1500 µmol m^{-2} s^{-1} (fitted)	ETR1500 (fit)
ETR at PPFD = 1500 µmol m^{-2} s^{-1} (observed)	ETR1500 (obs)
Y(II) at PPFD = 1500 µmol m^{-2} s^{-1} (observed)	Y(II)1500
Fv'/Fm' at PPFD = 1500 µmol m^{-2} s^{-1} (observed)	Fv'/Fm'1500
qP at PPFD = 1500 µmol m^{-2} s^{-1} (observed)	qP1500
Induction curve (IC) related	
t50 for ETR Induction Curve	t50 ETR induction
Saturated ETR of Induction Curve (ETR from rLRC at inducing irradiance of IC)	ETR_IC_saturated
ETR at the last point of Induction Curve (observed)	ETR_IC_last
Non-photochemical quenching (NPQ) related	
t50 for NPQ Relaxation Curve	t50 NPQ relaxation
NPQ at the last point of Induction Curve	NPQ_IC_last
NPQ at PPFD = 1500 µmol m^{-2} s^{-1}	NPQ1500

5.1. Introduction

By 2050, crop yield needs to be substantially improved to meet the food demand of the increasing human population (Ort *et al.*, 2015). Photosynthesis has recently gained renewed interest as the key to improving crop yield (Foyer *et al.*, 2017) because it is the first step of biomass production. Yield is mainly determined by three crop physiological components: light interception, the conversion efficiency of intercepted light energy into biomass (light use efficiency; LUE), and biomass partitioning into harvested products (harvest index) (Monteith, 1977). Because the harvest index and efficiency of the light capture are close to their theoretical maxima (Long et al., 2006; Zhu et al., 2010), the remaining component to improve yield is LUE, which is thought to be primarily determined by photosynthesis (Long et al., 2006; Zhu et al., 2010; Reynolds et al., 2011, 2012).

The importance of leaf photosynthesis for crop yield has been demonstrated by CO_2 enrichment experiments. Comparison of the growth of a genotype in current and future projected elevated CO_2 concentration environments in the field has supported the idea of increasing leaf photosynthesis to increase yield (Ainsworth and Long, 2005; Sun *et al.*, 2009; Ainsworth and Bush, 2011). For tomato production in high-tech greenhouses, CO_2 injection into the air is a common practice to increase leaf photosynthesis and yield (Heuvelink et al., 2020a).

However, except for a few studies (Watanabe *et al.*, 1994; Fischer *et al.*, 1998; Higashide and Heuvelink, 2009), photosynthesis per unit leaf area is often poorly correlated with yield when quantifying these traits in different cultivars or genotypes in a breeding population (Evans, 1993, 1998; Driever *et al.*, 2014; Gu *et al.*, 2014*b*). The poor correlation has raised questions about the idea that leaf photosynthesis limits crop production in plant breeding (Richards, 2000). The difficulty in finding a correlation between leaf photosynthesis and yield might arise from the difference between yield and total plant biomass production (i.e., respiration and harvest index), and also from the difference between total plant biomass production and leaf photosynthesis. The latter difference might be due to at least three reasons. First, whole plant photosynthesis is different from leaf photosynthesis because leaf area per plant and plant architecture are different among genotypes (Long et al., 2006). Second, leaf photosynthesis shows variation among leaves and within a leaf (Xiong *et al.*, 2015). Therefore, leaf photosynthesis measured on a (part of a) leaf would not be representative of all leaves. Third, photosynthesis shows a non-linear response to each environmental factor (e.g., light intensity, CO_2 concentration, and temperature). Thus, photosynthesis in one environmental condition or the capacity of photosynthesis (derived from a response curve) does not match with the operational photosynthesis during a day (Yin *et al.*, 2022).

The recent advances in the genetic basis of photosynthesis were made in a narrow range of germplasm by synthetic biology using bioengineered plants (Kromdijk *et al.*, 2016; Simkin *et al.*, 2019; López-Calcagno *et al.*, 2020). For example, increment in photosynthesis by

increasing the capacity for ribulose-1,5- bisphosphate (RuBP) regeneration (Driever *et al.*, 2017) or by accelerating recovery from photoprotection (Kromdijk *et al.*, 2016) boosted biomass, yield or both. These transgenic plants showed promise that this strategy can provide a solution for food security in the near future (Parry *et al.*, 2013; Foyer *et al.*, 2017).

Plant breeding to improve crop performance requires a broad genetic variation linked to the traits of interest (Tanksley and McCouch, 1997; Mujeeb-Kazi *et al.*, 2013; Huang and Han, 2014). However, photosynthesis studies have been conducted in a narrow range of germplasm (Long et al., 2006). Natural genetic variation in photosynthesis has been explored, and quantitative trait loci (QTLs) for various photosynthetic processes have been reported in rice, soybean, maize, and *Arabidopsis* (Flood *et al.*, 2011; Lawson *et al.*, 2012; van Rooijen *et al.*, 2017; van Bezouw *et al.*, 2019). Regarding tomato, one study reported QTLs for photosynthesis in *Solanum pennellii* introgression lines (Silva *et al.*, 2018). A study in rice incorporating leaf photosynthesis QTLs into a crop model showed that genetic variation of photosynthesis scaled up from leaf to crop biomass (Gu *et al.*, 2012*a*, 2014*a*). However, to the best of our knowledge, no studies have shown the extent to which these QTL alleles impact crop yield (van Bezouw *et al.*, 2019). More quantitative genetic studies should be conducted to reveal the genetic variation of photosynthesis and its role in breeding for improved yield.

The aim of the present study was to identify the quantitative genetic basis for photosynthesis-related traits and reveal the possible role of photosynthesis in improving yield in indeterminate tomato for greenhouse production. We conducted QTL analyses for photosynthesis-related traits on hybrids of 342 recombinant inbred lines (RILs) of tomato (*Solanum lycopersicum*) (342 RILs × test parent 1, 342 RILs × test parent 2) grown in a commercial tomato greenhouse. This RIL population was developed from four parents (two elite and two wild parents), which allowed us to investigate the effect of introgression of wild germplasm in elite parents' genetic background. We measured chlorophyll content index and chlorophyll fluorescence with the focus on the following aspects:

(1) Chlorophyll content index measured by SPAD meter
(2) Maximum quantum efficiency of photosystem II (PSII) photochemistry (F_v/F_m)
(3) Rapid light response curve (rLRC) of electron transport rate
(4) Induction curve (IC) of electron transport rate
(5) Non-photochemical quenching (NPQ) relaxation

Chlorophylls play a pivotal role in photosynthesis. Some studies reported that high chlorophyll content correlated with high leaf photosynthesis rate (Peng *et al.*, 2008). However, it has also been argued that a reduced leaf chlorophyll content improves light penetration into canopies, which reduces light saturation and thus would improve canopy photosynthesis (Melis, 2009; Zhu *et al.*, 2010; Ort *et al.*, 2011, 2015; Song *et al.*, 2017).

Photosynthesis can be separated into three main biochemical components: light reactions (electron transport), CO_2 fixation, and photorespiration (Sharwood *et al.*, 2022). Although CO_2 fixation and photorespiration can be measured by the gas exchange rate, gas exchange measurement in high throughput is difficult. On the other hand, the estimation of photosynthetic electron transport can be measured in high throughput using chlorophyll fluorescence techniques (Baker, 2008; Murchie and Lawson, 2013; Flood *et al.*, 2016; Furbank *et al.*, 2020). In the absence of photorespiration (under high CO_2 concentration), the electron transport rate shows a linear relationship with CO_2 fixation (Genty *et al.*, 1989). Recently, a study using a crop model demonstrated the importance of electron transport parameters rather than the maximum CO_2 fixation to improve crop yield (Yin *et al.*, 2022). F_v/F_m, a chlorophyll fluorescence parameter is a good proxy for the quantum efficiency of PSII photochemistry in dark-adapted leaves (Yin *et al.*, 2022). Another chlorophyll fluorescence parameter, photosystem II (PSII) operating efficiency (Y(II)) can be used to derive electron transport rate (Baker, 2008). Electron transport rate can be used to measure a rapid light response curve (rLRC) (Figure 1b), which provides information on the saturation characteristics of electron transport and overall photosynthetic performance (Thornley, 1998; Ralph and Gademann, 2005). In addition, a rapid and tuned response of photosynthesis to fluctuations in light has been identified as a key target for improvement (Zhu *et al.*, 2004; Lawson *et al.*, 2012; Kromdijk *et al.*, 2016; Taylor and Long, 2017; Murchie *et al.*, 2018). That is, photosynthesis can be more efficient if it increases quickly in transition from low to high light conditions (Lawson *et al.*, 2012; Taylor and Long, 2017) (lower t50, which is the time required to reach 50% of full induction; Figure 1c). Photosynthesis can be improved also by accelerating the recovery from photoprotection (rapid decrease in nonphotochemical quenching (NPQ)) in transition from high to low light conditions (Zhu *et al.*, 2004; Kromdijk *et al.*, 2016) (lower t50, which is the time required to reach 50% of NPQ observed at the last time point measured in high light; Figure 1d). We detected photosynthesis-related QTLs and compared these QTLs with previously found yield QTLs (Tsutsumi-Morita *et al.*, 2021). Then, we attempted to predict yield from photosynthesis-related QTLs to assess the usefulness of photosynthesis-related QTLs for plant breeding to improve yield.

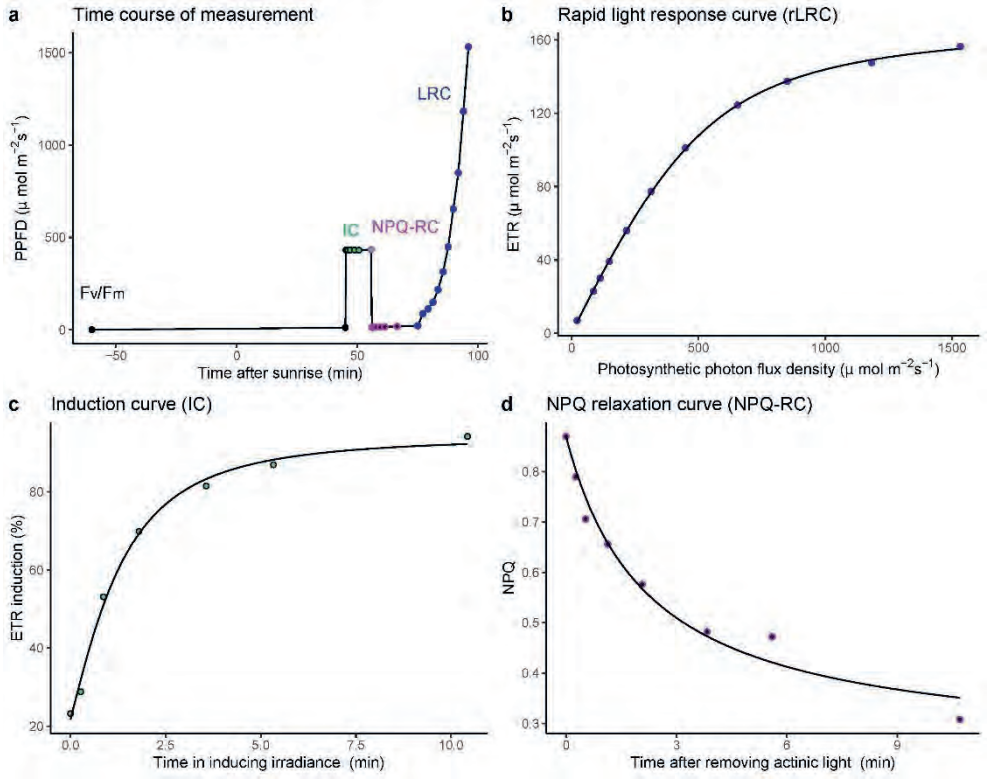


Figure 1 Representation of a chlorophyll fluorescence measurement. (a)Time course of measurement, (b) Rapid light response curve, (c) ETR induction curve from low to high irradiance, (d) NPQ relaxation curve after high to low irradiance. Dots indicate observed values. Curves in b-d shows fitted curves of each model. Abbreviations: PPFD, photosynthetic photon flux density; ETR, electron transport rate; NPQ, non-photochemical quenching.

5.2. Materials and methods

5.2.1. Breeding Population

The indeterminate-tomato population used in this study was the same as in a previous study (Tsutsumi-Morita *et al.*, 2021). Four parents (two elite parents (E1, E2) and two wild parents (W1, W2)) were crossed to develop a 342 RILs population. These RILs were crossed with two testers (TP1, TP2), producing two populations of 342 hybrid genotypes each (RILs × TP1 and RILs × TP2). These 684 genotypes were phenotyped three times, late spring, summer, and autumn, in one production period. Although phenotyping was conducted on these hybrids, QTL analyses related the phenotypes to markers in the RILs. Identity by descent (IBD) probabilities between parents and offspring (Zheng *et al.*, 2015) at 279 SNP SOLCAP marker positions were used as genetic information as mentioned in the previous study (Tsutsumi-Morita *et al.*, 2021).

5.2.2. Growing conditions and experimental design

The phenotyping experiment was the same as our previous study (Tsutsumi-Morita *et al.*, 2021), conducted in a modern glasshouse in the Netherlands from February to November in 2016, with commercial crop management. Growing conditions and experimental design were described in the previous study (Tsutsumi-Morita *et al.*, 2021).

5.2.3. Chlorophyll fluorescence and chlorophyll content measurement

Chlorophyll fluorescence was measured by four customized Monitoring-PAM chlorophyll fluorometers (Hex-PAM, Heinz Walz GmbH, Germany), each of which has six measuring heads (the main modifications of the Hex-PAM comprise the use of a white LED as excitation source and modified leaf-clip to allow for larger leaves and easier handling. Apart from the white LED light source, the Hex-PAM measures chlorophyll fluorescence using the same Pulse-Amplitude-Modulation principle (PAM) used in the PAM fluorometers (Heinz Walz GmbH, Germany) (Porcar-Castell *et al.*, 2008; Murchie and Lawson, 2013). Because we mostly used three of four Hex-PAMs at a time (due to maintenance not all four could be used simultaneously), we measured around 18 leaves (18 genotypes) every day from May to mid-November in 2016.

On the day before the chlorophyll fluorescence measurement, we determined the chlorophyll content index using a SPAD-502 chlorophyll meter (Konica Minolta, Inc., Tokyo, Japan) (Markwell *et al.*, 1995) on the same leaves.

After the SPAD measurement we clamped the leaf in a measuring head of a Hex-PAM for the chlorophyll fluorescence measurements the next day. During the night, leaf was kept

under dark (no supplementary light in the greenhouse). All chlorophyll fluorescence measurements, F_v/F_m, induction curve (IC), rapid light response curve (rLRC) were carried out automatically using a batch-file. One hour before sunrise, Hex-PAM measured minimum (F_o) and maximum (F_m) dark-adapted chlorophyll fluorescence yields (Figure 1), and F_v/F_m (maximum quantum efficiency of PSII photochemistry) was calculated by the following equation (Krause and Weis, 1991; Baker, 2008).

$$F_v/F_m = (F_m - F_o)/F_m \qquad Eq.\ 1$$

The IC and rLRC were measured starting 45 minutes after sunrise, under optimal conditions (vapor pressure deficit (VPD) expected to be small and ambient CO_2 concentration high). First the photosynthesis induction curve (IC) for electron transport rate (ETR) was measured during 10 minutes actinic illumination (mean photosynthetic photon flux density (PPFD) was 443 µmol m^{-2} s^{-1}). After the induction curve, non-photochemical quenching relaxation curve (NPQ-RC) was measured during 10 minutes. Rapid light response curve (rLRC) was measured at 75 minutes after sunrise, during 20 minutes by 11 time steps with 2 minutes interval (Figure 1). The leaf received not only LED light but also solar light, the sum of which was measured by the Hex-PAM PAR sensor positioned close to the chlorophyll fluorescence measuring spot on the leaf.

5.2.4. Curve fitting

A total of 2182 fluorescence measurements (on 2182 leaves) were conducted and for 2174 measurements also chlorophyll content index (SPAD) data were available (892 for RILs×TP1; 1011 for RILs×TP2; 28 for E1×TP1; 36 for E1×TP2; 39 for E2×TP1; 37 for E2×TP2; 131 for 5 commercial cultivars). After removing some incorrect measurements (i.e., missing time points during chlorophyll fluorescence measurements, PPFD showing a mild outlier (lower than Q1 - 1.5 × IQ or higher than Q3 + 1.5 × IQ; abbreviation, Q, quartile; IQ, interquartile range), at least at one point during IC and rLRC), 1654 measurements remained for analysis.

5.2.4.1. Rapid light response curve (rLRC)

Rapid Light Response curves (rLRC) were measured as a proxy for proper LRC (Ralph and Gademann, 2005). The main reason for this was to i) allow for the rLRC to be measured rapidly during the early morning when light intensity in the greenhouse is relatively low and ii) to obtain a light response to relatively rapid changing PPFD as occurring in a greenhouse. The linear relationship between photosystem II (PSII) operating efficiency (Y(II)) and linear electron flux allows the use of Y(II) to estimate the noncyclic electron transport rate (ETR) through PSII (Baker, 2008):

$$ETR = I_{inc} \times A_{leaf} \times fraction_{PSII} \times Y(II) \qquad Eq.\ 2$$

where I_{inc} is the incident PPFD (µmol m^{-2} s^{-1}) on the leaf, A_{leaf} is the proportion of incident PPFD that is absorbed by the leaf, $fraction_{PSII}$ is the fraction of absorbed PPFD that is received by PSII. A_{leaf} was assumed to be 0.84 and $fraction_{PSII}$ was assumed to be 0.5 (Baker, 2008).

A non-rectangular hyperbola (Thornley, 1998) was fitted to electron transport rate (ETR) of light response data for each measurement:

$$ETR = \frac{(\alpha I_{inc} + J_{max} - \sqrt{(\alpha I_{inc} + J_{max})^2 - 4\theta J_{max} \alpha I_{inc}})}{2\theta} \quad \textit{Eq. 3}$$

where I_{inc} is incident PPFD (µmol m^{-2} s^{-1}), α is the conversion efficiency of incident light into *ETR* at strictly limiting light (Slope_of_rLRC), J_{max} is the asymptotic maximum value of *ETR*, θ is a convexity factor for response of *ETR* to I_{inc} (Convexity_of_rLRC). In addition to these parameters, fitted ETR at PPFD = 1500 µmol m^{-2} s^{-1}, observed ETR, observed Y(II), observed F_v'/F_m' (PSII maximum efficiency), and observed qP (photochemical quenching) at PPFD = 1500 µmol m^{-2} s^{-1} were used for further analyses. Y(II), F_v'/F_m', and qP are related as $Y(II) = F_v'/F_m' \times qP$. Curve fitting was conducted by nls function in R 4.0.2 (R Core Team, 2020). Equation 3 was successfully fitted to 1479 out of the 1654 measurements. After removing inappropriate data, for which RMSE showed a mild outlier or a parameter showed an extreme outlier (lower than Q1 - 3 × IQ or higher than Q3 + 3 × IQ; abbreviation, Q, quartile; IQ, interquartile range), 1315 fitted rLRC remained for further analysis (540 for RILs×TP1; 619 for RILs×TP2; 17 for E1×TP1; 23 for E1×TP2; 19 for E2×TP1; 27 for E2×TP2; 70 for 5 commercial cultivars). The mean Root Mean Square Error (RMSE) for these rLRC was 1.46.

5.2.4.2. Induction curve (IC)

For the 1315 measurements, an induction curve (IC) was fitted (*Eq. 5*) only for measurements where electron transport rate (ETR) gradually increased during IC measurement. IC was fitted for ETR induction (%), which was calculated as the ratio between ETR and fitted ETR in rLRC at inducing irradiance (saturated ETR of induction curve (ETR_IC_saturated)).

$$ETR\ induction\ (\%) = \frac{ETR}{ETR\_IC\_saturated} \times 100 \quad \textit{Eq. 4}$$

A sigmoid function (Holling type III) was fitted for ETR induction:

$$ETR\ induction\ (\%) = \frac{IC_{max} \times (t - c)^2}{b^2 + (t - c)^2} \quad \textit{Eq. 5}$$

where t is the time in inducing irradiance (minutes), IC_{max} is the asymptotic maximum value of ETR induction, c is a shift on the x-axis, b is the time when the half-maximum of ETR

induction is achieved (in Holling type III equation with c=0). Curve fitting was conducted by nls function in R 4.0.2 (R Core Team, 2020). *Eq. 5* was successfully fitted to 1109 measurements. Time required to reach 50% of induction (t50_ETR_induction) was calculated as follows.

$$t50\_ETR\_induction = c + b\sqrt{(50 \,/\, (IC_{max} - 50))} \qquad \textit{Eq. 6}$$

After removing suspicious data, for which RMSE showed a mild outlier or a parameter showed an extreme outlier, 1044 plots remained for further analysis (426 for RILs×TP1; 497 for RILs×TP2; 10 for E1×TP1; 21 for E1×TP2; 16 for E2×TP1; 24 for E2×TP2; 50 for 5 commercial cultivars). The mean RMSE for IC was 1.74. We also evaluated the observed ETR at the last point of IC (ETR_IC_last) in this data set of 1044 measurements.

5.2.4.3. NPQ relaxation curve (NPQ-RC)

For the 1315 measurements where rLRC was successfully fitted, NPQ relaxation curve (NPQ-RC) was evaluated only for those measurements, where ETR gradually decreased following IC measurement. A hyperbolic function was fitted for NPQ (*Eq. 7*):

$$NPQ = \frac{d \times \left(NPQ_{IC\_last} - c_{asym}\right)}{(d + t)} + c_{asym} \qquad \textit{Eq. 7}$$

where t is the time after switching off actinic light ($t = 0$ at the last point of IC), c_{asym} is the asymptotic minimum value of NPQ relaxation, d is the time when the half-minimum of NPQ relaxation is achieved ($0.5 \times (NPQ_{IC\_last} - c_{asym})$), and NPQ_{IC_last} is the observed NPQ at the last point of IC. Curve fitting was conducted by nls function in R 4.0.2 (R Core Team, 2020). Time required to reach 50% of NPQ at the last point of IC (t50_NPQ _relaxation) was calculated as follows:

$$t50\_NPQ\_relaxation = \frac{d \times \left(NPQ_{IC\_last}\right)}{(NPQ_{IC\_last} - 2c_{asym})} \qquad \textit{Eq. 8}$$

After removing suspicious data (e.g., normalized RMSE (NRMSE) showed a mild outlier, t50_NPQ_relaxation was negative, or a parameter or t50_NPQ_relaxation showed an extreme outlier), 759 measurements remained for further analysis (285 for RILs×TP1; 376 for RILs×TP2; 12 for E1×TP1; 17 for E1×TP2; 14 for E2×TP1; 14 for E2×TP2; 41 for 5 cultivars). The mean NRMSE for NPQ-RC was 0.039. In addition, we also analyzed observed NPQ at the last point of rLRC (NPQ1500) in this data set of 759 measurements.

5.2.5. Phenotypic analyses

Spatially adjusted genotypic means (BLUEs: the best linear unbiased estimates) were computed by fitting a two-dimensional Penalized spline (P-spline) mixed model by a R-package SpATS (Spatial Analysis of Field Trials with Splines) (Velazco *et al.*, 2017; Rodríguez-Álvarez *et al.*, 2018). In addition to the Row and Column effects, the following cofactor and covariates were included into the mixed model to compute BLUEs: test parent (1 or 2) and measurement date for all traits as a fixed term; Hex-PAM measuring-head number (1-24) for all traits except for chlorophyll content as a random term; photosynthetic photon flux density (PPFD) at the last point of rLRC as a fixed term for J_{max} of rLRC, ETR1500 (fit), ETR1500 (obs), Y(II)1500, F_v'/F_m'1500, qP1500, and NPQ1500.

The same R-package was used to calculate a generalized heritability (Cullis *et al.*, 2006; Rodríguez-Álvarez *et al.*, 2018). The duplicates of the elite-parent genotypes E1×TP1, E2×TP1, E1×TP2, E2×TP2 and five commercial cultivars included in the experiment functioned as reference and facilitated the spatial corrections.

BLUEs were further used for phenotypic analyses. Histograms of BLUEs for each trait were made and correlation analyses were conducted by R version 4.0.2 (R Core Team, 2020). Stepwise regression for yield by all photosynthesis-related traits was conducted by using normalized BLUEs per RIL by a R-package of MASS.

5.2.6. QTL analyses

Single-QTL analyses followed by multi-QTL analyses were conducted by testing the association between BLUEs (phenotype) and identity by descent (IBD) probabilities (genotype), in the same way as in previous studies (Tsutsumi-Morita *et al.*, 2021; Li *et al.*, 2021, 2022). In summary, first, single-QTL analysis conducted by a loglikelihood ratio test at each marker position by using the R-package ASReml (Butler *et al.*, 2017) in the R-3.6.1 environment (R Core Team, 2020). The following models for the observation y were compared in a deviance test, where we use a scalar notation for simplicity: H_0, $y = \mu + \varepsilon$ versus H_a, $y = \mu + \mathbf{z}^t\mathbf{u} + \varepsilon$, with μ a fixed intercept, ε a normally distributed error with mean 0 and variance σ^2, $\mathbf{z}$ a design vector containing (functions of) IBD probabilities between the four parents and the RILs, $\mathbf{z}^t$ the transpose of $\mathbf{z}$ and $\mathbf{u}$ a coefficients vector of random QTL allele substitution effects with mean zero and variance σ^2_{QTL}. A threshold for significance was put at $-\log_{10}P \geq 2.9$ as the same as in a previous study (Tsutsumi-Morita *et al.*, 2021). Two adjacent peaks were recognized as different QTLs if the distance between two peaks was larger than 20 cM, or $-\log_{10}P$ value dropped by more than 0.5 between two adjacent peaks. Next, the significant QTLs from a genome scan were subsequently inserted in a multi-QTL mixed model: $y = \mu + \sum(\mathbf{z}^t\mathbf{u}) + \varepsilon$, where the summation, $\sum$, indicates that multiple QTLs were included into the model. When the variance component for a single-QTL was equal to zero, the single-QTL was removed from the final multi-QTL model. Final estimates

for QTL effects and predicted values for traits were obtained from the multi-QTL model fit for the set of QTLs with non-zero variances.

To evaluate whether QTLs for different traits were "located at the same position", a window of 20 cM on either side of a QTL was defined within which another QTL for a different trait should be located. Two QTLs at the same position were further considered to be "colocalized" if the parental origins of the QTL alleles (haplotypes) with the largest magnitude (towards higher yield) coincided for the two QTLs.

Prediction accuracies (Pearson correlation coefficient between observation and prediction) were evaluated in two ways. In the first way (r_{fitted}), we evaluated the prediction accuracy in a fitted multi-QTL model, where all RILs were used to make the multi-QTL model to predict (fit) all RILs. Another way was cross-validation (r_{cv}); we evaluated the median of prediction accuracies in five-fold cross-validation, where the data were partitioned into 5-folds, with 4-folds being included in a training set to make a multi-QTL model and 1-fold serving as validation set. This five-fold cross-validation was repeated 20 times for different random partitionings of the data.

5.2.7. Yield prediction by QTLs for photosynthesis-related traits

Because we observed QTL colocalization between yield and photosynthesis-related traits, we attempted to predict yield from QTLs for a photosynthesis-related trait. The same multi-QTL mixed model was implemented here as in the previous section. Response variable was yield (BLUE per RIL), and predictors were multiple markers related to QTLs for a photosynthesis-related trait. After fitting, markers with zero variance component were removed from the final model. Prediction accuracy was the Pearson correlation coefficient (r) between observed yield (BLUEs) and predicted yield by the multiple QTLs, or multi-QTLs, for the photosynthesis-related trait. We compared QTL allele effects between yield prediction and photosynthesis-related-trait prediction, when the same set of multi-QTLs for the photosynthesis-related trait were used for both predictions. QTLs allele effects were expressed as percentages of the trait mean: 100*(QTL effect divided by μ), with μ the intercept in the multi-QTL model). Five-fold cross-validation was conducted in the same way as the previous section.

To assess the credibility of yield prediction by QTLs for a photosynthesis-related trait, yield was predicted by using the same number of randomly selected markers. For example, yield was predicted by 9 QTLs for chlorophyll content. Then, yield was predicted by 9 randomly selected markers, which was repeated 100 times. The 95th percentile of prediction accuracy by random markers was compared with the accuracy of yield prediction by QTLs for chlorophyll content.

In addition, we tried to predict yield from the full set of single-QTLs detected for all traits measured by chlorophyll fluorescence (except for slope of rLRC, convexity of rLRC, qP1500, ETR IC last, for which no single-QTL were detected) with or without single-QTLs for chlorophyll content index. QTLs with no effects were removed from the final multi-QTL model. For five-fold cross-validation, the data were partitioned into 5-folds, with 4-folds being included in a training set to make a multi-QTL model and 1-fold serving as validation set. This five-fold cross-validation was repeated 20 times by different random partitionings of data (5 × 20 = 100 times).

To assess the credibility of yield prediction by the full set of single-QTLs for photosynthesis-related traits, yield was predicted by randomly selected 1 − 30 markers, which was iterated 100 times. The 95th percentile of prediction accuracy by random markers was compared with accuracy of yield prediction by the full set of QTLs for photosynthesis-related traits.

5.3. Results

5.3.1. No phenotypic correlation between yield and photosynthesis-related traits

Phenotypic analyses (by using mean phenotypic value per recombinant inbred line (RIL) after spatial and temporal corrections) were conducted to investigate the relationship between traits. All photosynthesis-related traits showed normal types of distribution, with elite and wild parents located in the centre of the distribution (Supplementary Figure 1). Therefore, substantial transgression occurred, suggesting that the traits were controlled by multiple genes, with all parents contributing positive as well as negative QTL effects. None of the photosynthesis-related traits correlated with total fruit fresh yield (yield), total plant biomass (total biomass), or adapted light use efficiency (LUE_{adp}; total biomass divided by fraction of intercepted light, which was based on leaf area index.) (Figure 2). In addition, RILs showing a higher yield than each of the elite parents, all of which showed higher total biomass than the elite parents, did not always show an improved photosynthetic characteristic — higher chlorophyll content index, higher F_v/F_m, higher slope of rapid light response curve (slope of rLRC), electron transport rate at photosynthetic photon flux density (PPFD) = 1500 µmol m^{-2} s^{-1} (ETR1500); or smaller t50 for ETR induction curve, smaller t50 for NPQ relaxation curve — compared to the elite parents (Supplementary Figure 2). We further investigated whether all photosynthesis-related traits together did show a relationship with yield, total biomass, or LUE_{adp} in multiple regression models. However, only 5.7%, 2.9%, or 3.9% of the phenotypic variance was explained for yield, total biomass, or LUE_{adp}, respectively (not shown). Therefore, phenotypic data of photosynthesis-related traits, measured in this study, was hardly useful to predict yield.

On the other hand, phenotypic relationships were confirmed among photosynthesis-related traits measured by chlorophyll fluorescence (Figure 2). J_{max} of rLRC, ETR1500(fit), ETR1500(obs), Y(II)1500 showed strong positive correlations with each other as expected (J_{max} is the asymptotic maximum ETR value of rLRC; ETR1500(fit) is the fitted ETR value of rLRC at PPFD = 1500 µmol m^{-2} s^{-1}; ETR1500(obs) is the observed ETR at the last point of rLRC (≈ 1500 µmol m^{-2} s^{-1}, which depended on the background solar radiation); Y(II)1500 is observed photosystem II operating efficiency, which determines ETR1500(obs) by *Eq. 2*). F_v/F_m, J_{max} of rLRC, slope of rLRC, and (-) t50 ETR induction (sign change of t50 ETR induction) all positively correlated with each other. There was no trade-off between J_{max} of rLRC and slope of rLRC at the population level, thus, no clear separation in two categories: sun-leaf (high J_{max} of rLRC, low slope of rLCR) or shade-leaf type (low J_{max} of rLRC, high slope of rLCR). Many RILs with higher J_{max} tended to show steeper slope in rLRC, and more quickly responded to the inducing irradiance. The positive correlation of these parameters with F_v/F_m (correlation (r) = 0.23 with slope of rLRC, 0.42 with J_{max}) was rather low for F_v/F_m as a proxy for photosynthetic electron transport rate. These RILs with high J_{max} and steeper slope in rLRC showed a lower NPQ at the last point of the induction curve (IC); however, NPQ showed slower relaxation after reducing irradiance (from on average 443 µmol m^{-2} s^{-1} to ambient light (mean 29.2 µmol m^{-2} s^{-1}) at one hour after sunrise). Although photosynthesis-related traits measured by chlorophyll fluorescence were correlated with each other, the chlorophyll content (index) did not show any correlation with these traits measured by chlorophyll fluorescence.

5.3.2. Many QTLs were at the same position for yield and photosynthesis-related traits but parental haplotype effects were not always consistent

Estimated heritability was zero for most traits, except for chlorophyll content and t50 for NPQ relaxation (Table 1a). Zero estimated heritability does not imply there is no genetic variation. We detected QTLs for most traits, except for slope of rLRC, convexity of rLRC, coefficient of photochemical quenching at PPFD = 1500 µmol m^{-2} s^{-1} (qP1500), and observed ETR at the last point of induction curve (ETR_IC_last) (Table 1b; Figure 3; Supplementary Figure 3). To assess the reliability of QTL detection for individual traits, as a benchmark criterion we predicted traits from sets of random markers. Prediction accuracy of the fitted model (all RILs were used for developing the multi-QTL model) by proper QTLs for each trait was higher than the 95th percentile of prediction accuracy by random markers (Supplementary Figure 4). Prediction accuracy was higher for chlorophyll content and t50 NPQ relaxation, for which estimated heritability was not zero, than for other traits (Table 1b). Wild parent(s) brought some positive effects for most traits, except for the saturated ETR of induction curve (ETR_IC_saturated) (Figure 3; Supplementary Figure 5).

Many QTLs for yield, biomass, and LUE_{adp} were at the same position (within 20cM) with QTLs for photosynthesis-related traits (grey and red arrows in Figure 3); however, only for some of these QTLs at the same position, the parental haplotype effect was consistent between yield, biomass, LUE_{adp} and photosynthesis-related traits (red arrows in Figure 3). Among 16 (multi-)QTLs for yield, 9 QTLs (56%) were at the same position as photosynthesis-related-traits (multi-)QTLs, among which 5 QTLs (31%) colocalized between yield and photosynthesis-related traits (colocalization within 20cM with the same haplotype showing the highest positive effect). Among 12 (multi-)QTLs for biomass, 9 QTLs (75%) were at the same position as photosynthesis-related-traits (multi-)QTLs, among which 4 QTLs (33%) colocalized with the same haplotype showing the highest positive effect. Among 9 (multi-)QTLs for LUE_{adp}, 7 QTLs (78%) were at the same position as photosynthesis-related-traits (multi-)QTLs, among which only 1 QTL (11%) colocalized with the same haplotype (wild parent 1) showing the highest positive effect. Y1.3 (the third (multi-)QTL for yield on chromosome 1), Y1.4, B1.1, LUE_{adp}1.2 colocalized with J_{max} of rLRC, ETR1500(fit), ETR1500(obs), Y(II)1500; Y4.1 colocalized with F_v/F_m; B6.4 colocalized with (-)Chlorophyll content (wild parent 1 showed the highest positive effect on biomass, while showing the strongest negative effect on Chlorophyll content); Y7.1 colocalized with (-)Chlorophyll content; B8.1 colocalized with (+)Chlorophyll content; Y10.2 colocalized with (-)Chlorophyll content; B10.1 colocalized with F_v/F_m.

Two out of 7 (multi-)QTLs for F_v/F_m colocalized with (multi-)QTLs for other photosynthetic parameters (Figure 3): F_v/F_m5.2 colocalized with a QTL for ETR_IC_saturated; F_v/F_m12.1 colocalized with a QTL for Y(II)1500 and a QTL for F_v'/F_m'1500.

In spite of the lack of any phenotypic correlation between chlorophyll content and photosynthesis-related traits measured by chlorophyll fluorescence (Figure 2), 4 out of 10 (multi-)QTLs for chlorophyll content (ChlC) colocalized (with the same haplotype showing the highest positive allele effect) with QTLs for photosynthesis-related traits measured by chlorophyll fluorescence (Figure 3): (+)ChlC1.1 colocalized with a QTL for (-)t50 ETR induction; (+)ChlC5.2 colocalized with a QTL for F_v/F_m and a QTL for ETR_IC_ saturated; (-)ChlC5.2 colocalized with a QTL for (-)t50 ETR induction and a QTL for (-)t50 NPQ relaxation; (+)ChlC7.1 colocalized with a QTL for (-)t50 ETR induction; (-)ChlC8.2 colocalized with a QTL for (-)t50 NPQ relaxation.

5.3.3. Yield prediction by QTLs for a photosynthesis-related trait

Because some yield QTLs colocalized with QTLs for photosynthesis-related traits (Figure 3), we investigated whether we could predict yield from QTLs for photosynthesis-related traits. High prediction accuracy was observed in yield prediction by QTLs for chlorophyll content and by QTLs for t50 ETR induction (Table 1c; Supplementary Figure 6). To confirm the reliability of yield prediction by QTLs for photosynthesis-related traits, we predicted yield by sets of randomly selected markers. A higher number of random markers predicted yield

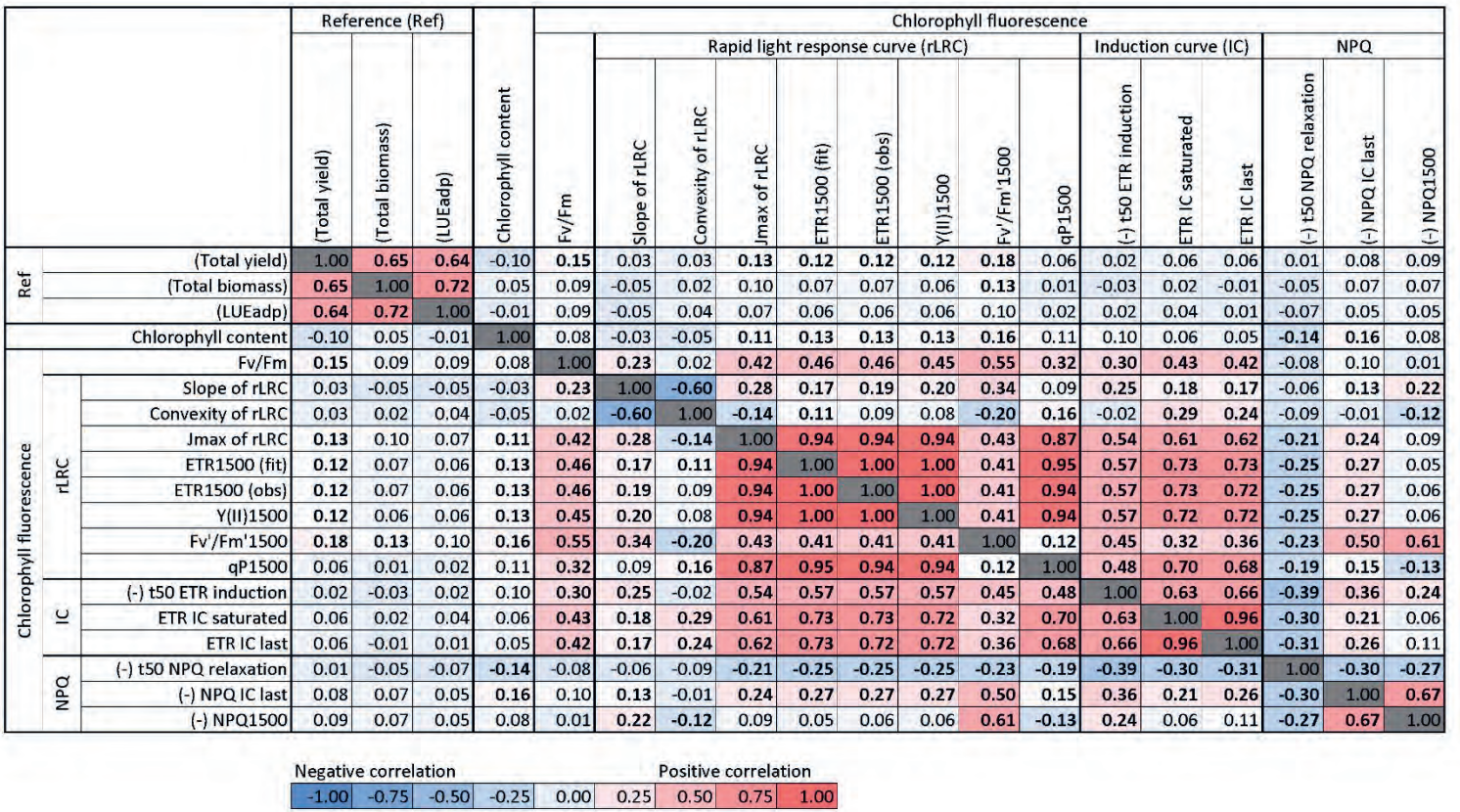
with higher prediction accuracy (Supplementary Figure 7). The prediction accuracies (r_{fitted}) of fitted models (all RILs were used to make multi-QTL models for yield prediction) by QTL markers for each of these two traits were higher than the 95th percentile of accuracy for yield prediction by sets of randomly selected markers, iterated 100 times (r_{fitted} = 0.59 by 9 QTLs for Chlorophyll content > 0.55 at the 95th percentile by 9 random markers; r_{fitted} = 0.52 by 5 QTLs for t50 ETR induction > 0.48 at the 95th percentile by 5 random markers) (Table 1c; Supplementary Figure 7). In addition, yield prediction by QTLs for F_v/F_m also showed a higher prediction accuracy than yield prediction by sets of randomly selected markers (r_{fitted} = 0.42 by 2 QTLs for F_v/F_m > 0.37 at the 95th percentile by 2 random markers), where only 2 out of 7 QTLs for F_v/F_m were retained in the yield prediction model (other 5 QTLs showed zero variance component). We also predicted total biomass and LUE_{adp} by using QTLs for each photosynthesis-related trait; however, the prediction accuracies for total biomass and LUE_{adp} were lower than that for yield (Table 1c; Supplementary Table 2).

For traits that showed the higher accuracy of yield prediction by their proper QTLs (chlorophyll content, t50 ETR induction, and F_v/F_m), parental-allele effects (haplotype effects) were not always consistent between prediction of yield and that of other traits (Figure 4). For some QTLs for chlorophyll content, positive effects on chlorophyll content were accompanied by negative effects on yield (e.g., chromosome 7, 10); however, for other QTLs, the opposite was shown (e.g., chromosome 9) (Figure 4a). For (-)t50 ETR induction, positive effects of elite haplotypes on yield were not accompanied by faster induction in (-)t50 ETR induction7.1 (Figure 4c). On the other hand, for most of the other QTLs, positive haplotype effects on yield were accompanied by faster induction, although the increasing effect on a trait was accompanied by a small effect on the other trait in most cases. For F_v/F_m, in most cases except for one haplotype (elite parent 1) on chromosome 12, the haplotype that showed a positive effect on F_v/F_m also showed a positive effect on yield, and vice versa (Figure 4b). For a few QTLs, wild haplotypes (pink and orange numbers in Figure 4) showed positive effects on both a photosynthesis-related trait and yield, which demonstrated the benefits of using wild genetic resources to improve yield via photosynthesis.

5.3.4. High yield prediction accuracy when using QTLs for all photosynthesis-related traits together

Because QTLs for individual photosynthesis-related traits could predict yield to some extent (Table 1c), we tested yield prediction based on combining all QTLs for all photosynthesis-related traits (Figure 5). All QTLs for all traits measured by chlorophyll fluorescence (ChlF) predicted yield with a high prediction accuracy (for fitted model r_{fitted} = 0.67; median in five-fold cross-validation r_{cv} = 0.55; Figure 5c), which was the same as for yield prediction by QTLs for chlorophyll content (r_{fitted} = 0.65; r_{cv} = 0.55; Figure 5b). Combining QTLs for chlorophyll content and QTLs for traits measured by chlorophyll fluorescence (all QTLs for all photosynthesis-related traits) resulted in an even higher prediction accuracy (r_{fitted} = 0.74;

r_{cv} = 0.60; Figure 5e), which was comparable with yield prediction by QTLs for yield itself (r_{fitted} = 0.72; r_{cv} = 0.61; Figure 5a). Yield prediction using all QTLs for yield, biomass, and LUE_{adp} showed a slightly higher prediction accuracy (r_{fitted} = 0.76; r_{cv} = 0.63; Figure 5d). Finally, yield prediction by all QTLs for all traits achieved the highest prediction accuracy (r_{fitted} = 0.79; r_{cv} = 0.66; Figure 5f). In this final multi-QTL model to predict yield, 26 QTLs were used, whereas QTLs for chlorophyll content achieved a slightly lower prediction accuracy with only 10 QTLs. In fitted models (all RILs were used to make multi-QTL models for yield), these sets of QTLs achieved higher prediction accuracies than the 95th percentile of accuracy for yield prediction by sets of randomly selected markers (Figure 5; Supplementary Figure 7): r_{fitted} = 0.67 using 16 QTLs for all traits measured by chlorophyll fluorescence (Figure 5c) which is slightly larger than 0.65 being the 95th percentile for prediction by 16 random markers; r_{fitted} = 0.65 using 10 QTLs for chlorophyll content (Figure 5b) which is larger than 0.58 being the 95th percentile for prediction by 10 random markers; r_{fitted} = 0.74 using 20 QTLs for all photosynthesis-related traits (Figure 5e) which is larger than 0.68 being the 95th percentile for prediction by 20 random markers; r_{fitted} = 0.72 using 16 QTLs for yield (Figure 5a) which is larger than 0.65 being the 95th percentile for prediction by 16 random markers; r_{fitted} = 0.76 using 23 QTLs for yield, biomass and LUE_{adp} (Figure 5d) which is larger than 0.69 being the 95th percentile for prediction by 23 random markers; r_{fitted} = 0.79 using 26 QTLs for all traits (Figure 5f) which is larger than 0.70 being the 95th percentile for prediction by 26 random markers. The haplotype effects showed a very similar pattern among yield predictions by different sets of QTLs (Supplementary Figure 8). In conclusion, yield was successfully predicted by QTLs for photosynthesis-related traits.



			Reference (Ref)				Chlorophyll fluorescence													
								Rapid light response curve (rLRC)								Induction curve (IC)			NPQ		
			(Total yield)	(Total biomass)	(LUEadp)	Chlorophyll content	Fv/Fm	Slope of rLRC	Convexity of rLRC	Jmax of rLRC	ETR1500 (fit)	ETR1500 (obs)	Y(II)1500	Fv'/Fm'1500	qP1500	(-) t50 ETR induction	ETR IC saturated	ETR IC last	(-) t50 NPQ relaxation	(-) NPQ IC last	(-) NPQ1500
Ref		(Total yield)	1.00	**0.65**	**0.64**	-0.10	**0.15**	0.03	0.03	**0.13**	**0.12**	**0.12**	**0.12**	**0.18**	0.06	0.02	0.06	0.06	0.01	0.08	0.09
		(Total biomass)	**0.65**	1.00	**0.72**	0.05	0.09	-0.05	0.02	0.10	0.07	0.07	0.06	**0.13**	0.01	-0.03	0.02	-0.01	-0.05	0.07	0.07
		(LUEadp)	**0.64**	**0.72**	1.00	-0.01	0.09	-0.05	0.04	0.07	0.06	0.06	0.06	0.10	0.02	0.02	0.04	0.01	-0.07	0.05	0.05
		Chlorophyll content	-0.10	0.05	-0.01	1.00	0.08	-0.03	-0.05	**0.11**	**0.13**	**0.13**	**0.13**	**0.16**	0.11	0.10	0.06	0.05	**-0.14**	**0.16**	0.08
Chlorophyll fluorescence		Fv/Fm	**0.15**	0.09	0.09	0.08	1.00	**0.23**	0.02	**0.42**	**0.46**	**0.46**	**0.45**	**0.55**	**0.32**	**0.30**	**0.43**	**0.42**	-0.08	0.10	0.01
	rLRC	Slope of rLRC	0.03	-0.05	-0.05	-0.03	**0.23**	1.00	**-0.60**	**0.28**	**0.17**	**0.19**	**0.20**	**0.34**	0.09	**0.25**	**0.18**	**0.17**	-0.06	**0.13**	**0.22**
		Convexity of rLRC	0.03	0.02	0.04	-0.05	0.02	**-0.60**	1.00	**-0.14**	**0.11**	0.09	0.08	**-0.20**	**0.16**	-0.02	**0.29**	**0.24**	-0.09	-0.01	**-0.12**
		Jmax of rLRC	**0.13**	0.10	0.07	**0.11**	**0.42**	**0.28**	**-0.14**	1.00	**0.94**	**0.94**	**0.94**	**0.43**	**0.87**	**0.54**	**0.61**	**0.62**	**-0.21**	**0.24**	0.09
		ETR1500 (fit)	**0.12**	0.07	0.06	**0.13**	**0.46**	**0.17**	**0.11**	**0.94**	1.00	**1.00**	**1.00**	**0.41**	**0.95**	**0.57**	**0.73**	**0.73**	**-0.25**	**0.27**	0.05
		ETR1500 (obs)	**0.12**	0.07	0.06	**0.13**	**0.46**	**0.19**	0.09	**0.94**	**1.00**	1.00	**1.00**	**0.41**	**0.94**	**0.57**	**0.73**	**0.72**	**-0.25**	**0.27**	0.06
		Y(II)1500	**0.12**	0.06	0.06	**0.13**	**0.45**	**0.20**	0.08	**0.94**	**1.00**	**1.00**	1.00	**0.41**	**0.94**	**0.57**	**0.72**	**0.72**	**-0.25**	**0.27**	0.06
		Fv'/Fm'1500	**0.18**	**0.13**	0.10	**0.16**	**0.55**	**0.34**	**-0.20**	**0.43**	**0.41**	**0.41**	**0.41**	1.00	0.12	**0.45**	**0.32**	**0.36**	**-0.23**	**0.50**	**0.61**
		qP1500	0.06	0.01	0.02	0.11	**0.32**	0.09	**0.16**	**0.87**	**0.95**	**0.94**	**0.94**	0.12	1.00	**0.48**	**0.70**	**0.68**	**-0.19**	**0.15**	**-0.13**
	IC	(-) t50 ETR induction	0.02	-0.03	0.02	0.10	**0.30**	**0.25**	-0.02	**0.54**	**0.57**	**0.57**	**0.57**	**0.45**	**0.48**	1.00	**0.63**	**0.66**	**-0.39**	**0.36**	**0.24**
		ETR IC saturated	0.06	0.02	0.04	0.06	**0.43**	**0.18**	**0.29**	**0.61**	**0.73**	**0.73**	**0.72**	**0.32**	**0.70**	**0.63**	1.00	**0.96**	**-0.30**	**0.21**	0.06
		ETR IC last	0.06	-0.01	0.01	0.05	**0.42**	**0.17**	**0.24**	**0.62**	**0.73**	**0.72**	**0.72**	**0.36**	**0.68**	**0.66**	**0.96**	1.00	**-0.31**	**0.26**	0.11
	NPQ	(-) t50 NPQ relaxation	0.01	-0.05	-0.07	**-0.14**	-0.08	-0.06	-0.09	**-0.21**	**-0.25**	**-0.25**	**-0.25**	**-0.23**	**-0.19**	**-0.39**	**-0.30**	**-0.31**	1.00	**-0.30**	**-0.27**
		(-) NPQ IC last	0.08	0.07	0.05	**0.16**	0.10	**0.13**	-0.01	**0.24**	**0.27**	**0.27**	**0.27**	**0.50**	**0.15**	**0.36**	**0.21**	**0.26**	**-0.30**	1.00	**0.67**
		(-) NPQ1500	0.09	0.07	0.05	0.08	0.01	**0.22**	**-0.12**	0.09	0.05	0.06	0.06	**0.61**	**-0.13**	**0.24**	0.06	0.11	**-0.27**	**0.67**	1.00

Negative correlation								Positive correlation
-1.00	-0.75	-0.50	-0.25	0.00	0.25	0.50	0.75	1.00

Figure 2 Pearson correlations between traits. Significant correlation coefficients ($p < 0.05$) are indicated in bold. Red indicates positive, blue indicates negative correlation. A minus sign (-) is placed in front of some traits to align the trait with positive yield. Traits were explained in Table 1.

Table 1 Trait name, trait abbreviation, (a) estimated heritability, (b) number of QTLs and accuracy for prediction of trait by its proper QTLs, and (c) accuracy of yield prediction by secondary traits.

			(b) QTLs for own trait				(c) Yield prediction by multi-QTLs for a secondary trait		
			Number of QTLs		Prediction accuracy (r)			Prediction accuracy (r)	
	Abbreviation	(a) Heritability	Peaks in single-QTL analysis	Multi-QTLs	Fitted	Median in cross-validation	Remained number of multi-QTLs for yield prediction	Fitted	Median in cross-validation
	(Yield)	(0.62)	*(21)	(16)	(0.72)	(0.61)	16	0.72	0.61
	(Biomass)	(0.48)	*(17)	(12)	(0.60)	(0.49)	10	0.60	0.50
	(LUEadp)	(0.41)	(11)	(9)	(0.51)	(0.40)	7	0.59	0.53
Photosynthesis-related traits									
	Chlorophyll content, (ChlC)	0.47	13	10	0.65	0.57	9	0.59	0.49
Traits measured by chlorophyll fluorescence (ChlF)									
	F_v/F_m	0	10	7	0.41	0.31	2	0.42	0.39
Rapid light response curve (rLRC) related									
	Slope of rLRC	0	0	-	-	-			
	Convexity of rLRC	0	0	-	-	-			
	J_{max} of rLRC	0	1	1	0.23	0.20	1	0.21	0.13
	ETR1500 (fit)	0	3	3	0.36	0.27	3	0.37	0.33
	ETR1500 (obs)	0	3	3	0.36	0.27	3	0.37	0.33
	Y(II)1500	0	4	4	0.36	0.26	3	0.37	0.33
	Fv'/Fm'1500	0	5	4	0.39	0.30	4	0.29	0.12

qP1500	0	0	-	-	-			
Induction curve (IC) related								
t50 ETR induction	0	6	6	0.38	0.28	5	0.52	0.46
ETR_IC_saturated	0	1	1	0.22	0.18	1	0.20	0.20
ETR_IC_last	0	0	-	-	-			
Non-photochemical quenching (NPQ) related								
t50 NPQ relaxation	0.21	7	6	0.55	0.43	5	0.41	0.35
NPQ_IC_last	0	2	2	0.31	0.25	1	0.23	0.19
NPQ1500	0	1	1	0.24	0.20	1	0.11	0.10

* The single-QTLs were not evaluated in Tsutsumi-Morita et al., 2021. All peaks in the single-QTL scan were included in the multi-QTL model. If the single-QTLs would be evaluated in the same way as this study (20cM window or $-\log_{10}P$ value dropped by more than 0.5 between adjacent peaks), the number of single-QTLs will be 20 and 15 for total fruit yield and total dry biomass, respectively.

Figure 3 QTL locations and effects. Horizontal axis indicates chromosome number (1 – 12) and genetic position (cM) within chromosome. Single-QTL positions are indicated by empty circles, where size of empty circle indicates -log10(p) value. QTL effects in multi-QTL models for each trait are indicated by dots, where the colour of dot shows the haplotype with the highest positive effect and the size of dot represents the proportion (%) of largest haplotype effect when compared to the mean for the trait (in the multi-QTL model). Red arrows and letters indicate the colocalizations between multi-QTLs for a photosynthesis-related trait and multi-QTLs for total yield (Y), total biomass (B), or light use efficiency (L) when the same haplotype showed the strongest effect. Caption continued to the next page,

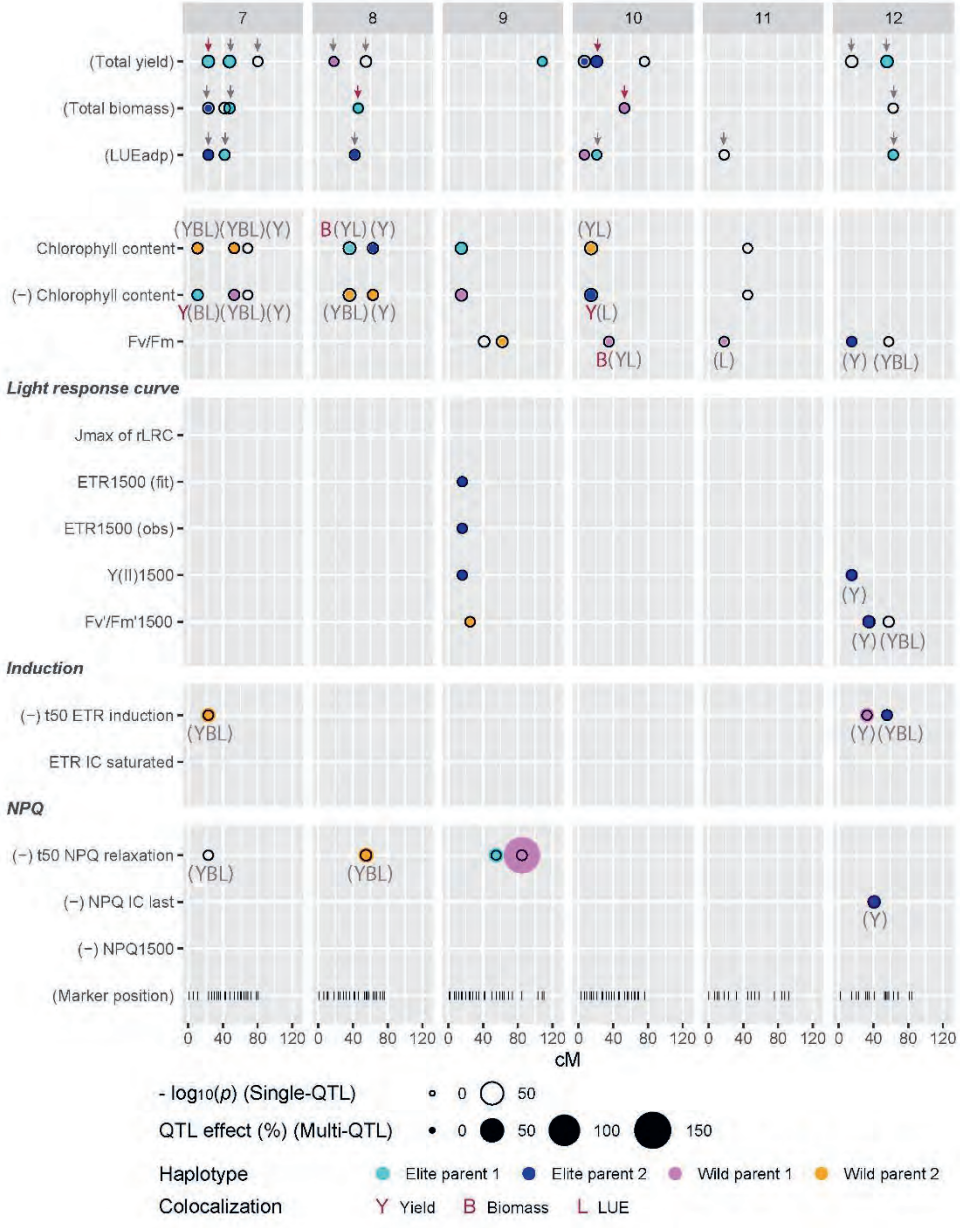


Figure 3 Caption continued. Negative signs in front of trait names indicate that negative values of the trait aligned positively with yield. Grey arrows and letters within brackets indicate that single-QTLs were at the same position between single-QTLs for a photosynthesis-related trait and single-QTLs for total yield (Y), total biomass (B), or adapted light use efficiency (L), although these single-QTLs were not retained in multi-QTL models (empty circles without colour) or different haplotype showed the strongest effect in multi-QTL models. Traits were explained in Table 1. (Results for total yield, total biomass, and adapted light use efficiency were taken from the previous study (Tsutsumi-Morita et al., 2021; Chapter 4).)

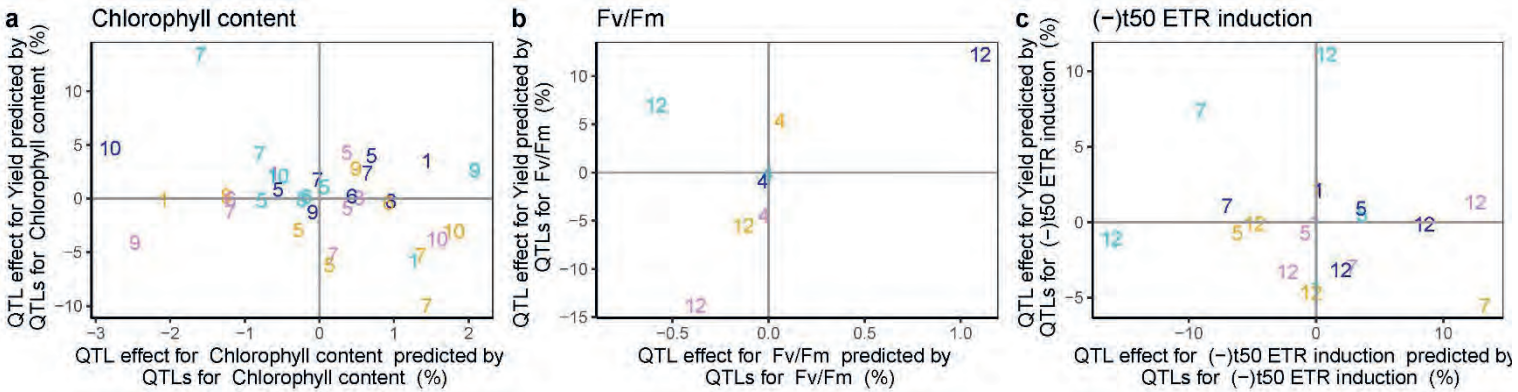


Figure 4 QTL parental allele effect comparison between yield prediction and a photosynthesis-related trait prediction by the same set of QTLs for the photosynthesis-related trait. The x-axis shows the QTL parental allele effect of a photosynthesis-related trait prediction by QTLs for the photosynthesis-related trait. The y-axis shows the QTL parental allele effect of yield prediction by QTLs for the photosynthesis-related trait. QTL parental allele effect is expressed as the proportion (%) compared to the mean for the trait (in the multi-QTL model). The numerical character indicates the chromosome in which the QTL locates. The colour of the number represents parent: light blue, elite parent 1; blue, elite parent 2; pink, wild parent 1, yellow, wild parent 2.

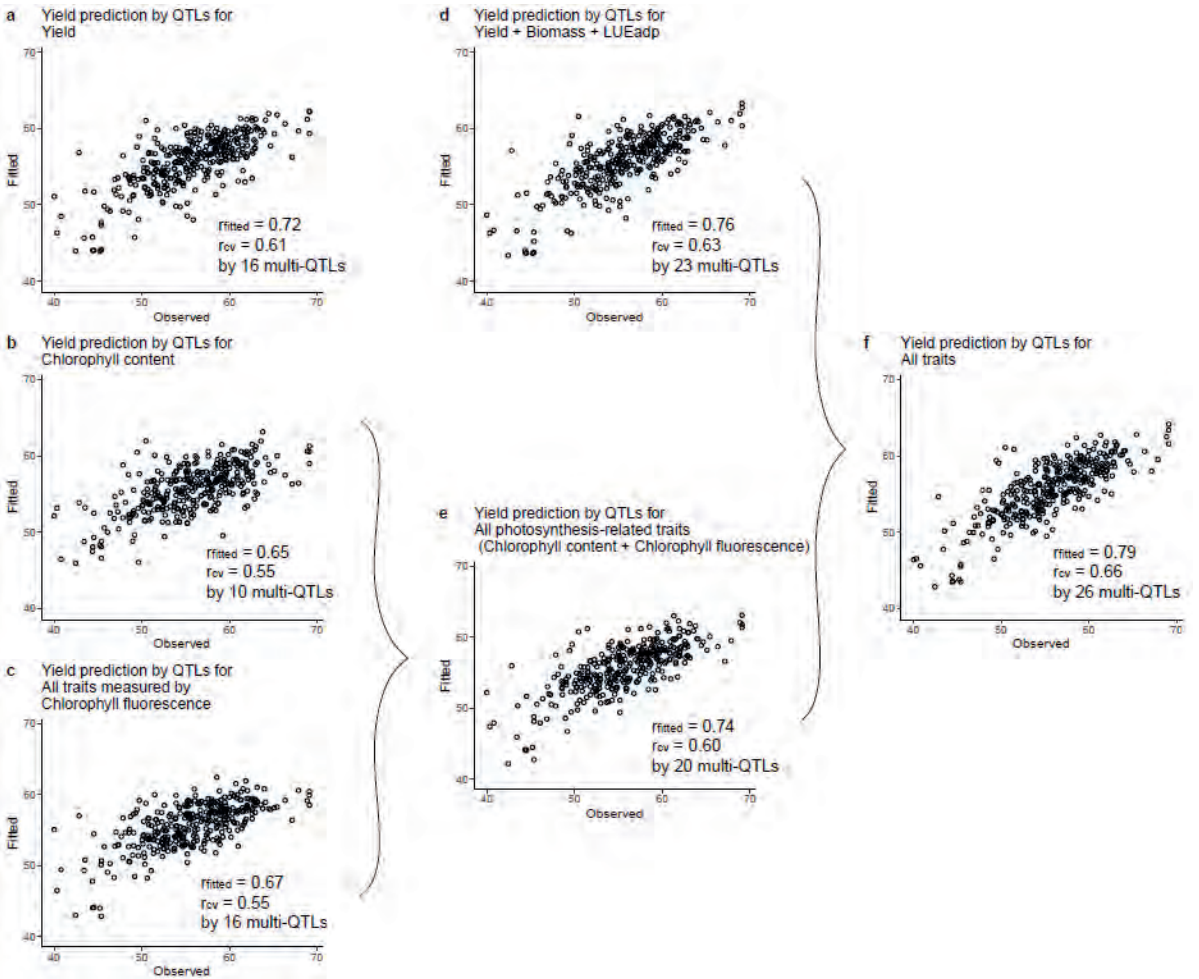


Figure 5 Yield prediction by a set of QTL-markers for secondary traits. Scatter plots present the correlation between observed yield and fitted yield by multi-QTL model using QTLs for (a) yield; (b)chlorophyll content index; (c) all traits measured by chlorophyll fluorescence; (d) yield, biomass and adapted light use efficiency; (e) all photosynthesis-related traits (chlorophyll content + traits measured by chlorophyll fluorescence); (f) all traits together. Prediction accuracy of the fitted model (r_{fitted}) is the Pearson correlation coefficient between observed yield and fitted yield in the multi-QTL model by using all RILs for fitting; prediction accuracy of cross-validation (r_{cv}) is the median of Pearson correlation coefficients between observed yield and predicted yield in five-fold cross-validation.

5.4. Discussion

Our results demonstrated that tomato yield can be predicted with high prediction accuracy by QTLs for photosynthesis-related traits providing a new tool to guide breeding efforts.

5.4.1. Most adapted-light-use-efficiency (LUE_{adp}) QTLs were underlain by photosynthesis-related QTLs but the allele effects were mostly different

We detected QTLs for photosynthesis-related traits, and many of these QTLs were at the same position as yield QTLs, total-biomass QTLs, and adapted-light-use-efficiency (LUE_{adp}) QTLs. In our previous study (Chapter 4), LUE_{adp} was calculated by dividing total biomass by light interception, which was estimated by leaf area index (LAI) and a light extinction coefficient *k,* where *k* was assumed to be the same for all genotypes. Parameter *k* represents plant architecture (light extinction following Lambert-Beer's law); at the same LAI, a higher *k* means a higher light interception. Because *k* is likely to be different among genotypes, assuming a fixed *k* made that the effect of plant architecture was included in LUE_{adp}, which makes that LUE_{adp} does not only depend on pure photosynthetic effects. In the previous study (Chapter 4), we detected 9 LUE_{adp} QTLs, of which 2 LUE_{adp} QTLs appeared at the same positions as leaf-morphology-related QTLs (QTL alleles for narrower-shaped leaves or smaller-width leaves increased LUE_{adp}). In this study, we found that the other 7 LUE_{adp} QTLs were at the same position as photosynthesis-related QTLs. However, contrary to our expectation, only for 1 LUE_{adp} QTL ($LUE_{adp}1.2$), the same parental haplotype (wild parent 1) showed the highest positive effect on both LUE_{adp} and the photosynthesis-related traits (J_{max}, ETR1500(fit), ETR1500(obs), Y(II)1500). The $LUE_{adp}1.2$ also colocalized with Yield1.3 and Biomass1.1 with also wild parent 1 showing the highest positive effect. Thus, the genome from wild parent 1 at $LUE_{adp}1.2$ increases J_{max}, ETR1500(fit), ETR1500(obs), and Y(II)1500, which increases LUE_{adp}, total biomass, and yield. For the other 6 LUE_{adp} QTLs parental haplotype effects were different between LUE_{adp} and photosynthesis-related traits. Therefore, at these loci, an improvement of photosynthesis-related traits at leaf level did not lead to LUE_{adp} improvement at canopy level.

5.4.2. Photosynthesis-related QTLs predicted yield, despite no correlation between photosynthesis-related traits and yield was observed

Chlorophyll-content (ChlC) QTLs, (-)t50 ETR induction QTLs, and F_v/F_m QTLs predicted yield with a higher prediction accuracy than the 95th percentile of prediction accuracy by the same number of randomly selected markers (Table 1c; Supplementary Figure 7). Thus, these three traits seem useful for plant breeding to improve yield. However, for ChlC7.1, ChlC7.2, ChlC10.1, and (-)t50 ETR induction7.1 (which colocalized with ChlC7.1) opposite parental

allele (haplotype) effects were observed between yield prediction and each trait prediction (Figure 4). On the other hand, for some other QTLs, parental allele effects were similar between yield prediction and prediction of each photosynthesis-related trait. Thus, for some QTLs, a positive allele effect on a photosynthesis-related trait showed a negative allele effect on yield, while for some other QTLs, a positive allele effect on a photosynthesis-related trait also had a positive effect on yield. These inconsistent allele-effect relationships between yield prediction and a photosynthesis-related trait prediction by the same set of QTLs for the photosynthesis-related trait, were consistent with the observations in the original QTL-parental-allele effects (Figure 3). Thus, the positive and negative QTL-parental-allele effects may cancel each other out at the phenotypic level. This may explain why we did not find correlations between yield and photosynthesis-related traits (Figure 2). In the history of plant breeding, the early assumption that selection on photosynthesis would lead to increased productivity has resulted in failure and frustration (Foyer *et al.*, 2017). Many studies have observed no correlation between leaf photosynthesis and yield in a breeding population or cultivar comparison (Richards, 2000; Gu *et al.*, 2014*a*), which has posed a question on whether leaf photosynthesis defines yield (Richards, 2000). Our results may provide one of the possible answers to this long-lasting question: some photosynthesis-related QTLs can contribute to yield, whereas some others can decrease yield.

5.4.3. Opposite allele effects between yield and leaf photosynthesis-related traits might reflect canopy photosynthesis

The inconsistent parental-allele effects between photosynthesis-related traits and yield might represent the discrepancy between leaf photosynthesis and canopy photosynthesis (whole plant photosynthesis). Leaf photosynthesis responds nonlinearly to increases in irradiance. Thus, leaves exposed to high light intensity cannot efficiently use it all and the excess light is wastefully dissipated or even harmful to the photosynthetic apparatus. On the other hand, leaves inside the canopy are shaded, and their photosynthesis is light-limited. Therefore, greater light penetration into the canopy mitigates efficiency losses associated with light saturation (Ort *et al.*, 2011). Thus, lowering the chlorophyll content of photosynthetic tissue might increase light penetration into canopies and therefore improve canopy photosynthesis (Melis, 2009; Zhu *et al.*, 2010; Ort *et al.*, 2011; Song *et al.*, 2017). At least, chlorophyll-content QTLs on chromosome 7 (10.8 cM) and 10 might support this theory, reducing greenness was associated with increasing yield. Therefore, in some QTLs, reductions in leaf-photosynthesis-related traits might increase canopy photosynthesis, whereas for other QTLs (e.g., J_{max}1.1), increasing leaf-photosynthesis-related traits might simply improve canopy photosynthesis. This complicated relationship may contribute to the absence of phenotypic correlation between leaf photosynthesis-related traits and yield (Figure 2).

One caution is that stronger greenness is often associated with higher photosynthesis and higher biomass production. Indeed, under nitrogen shortage or in some cases in field

conditions, greenness showed a correlation with photosynthesis and biomass production (Honda *et al.*, 2021). Thus, greenness has also been used as a secondary trait to predict yield in genomic prediction (Rutkoski *et al.*, 2016; Sun *et al.*, 2017). In these studies, greenness showed a correlation with yield. Our experiment was conducted under the commercial management of tomato production in the Netherlands. The nutrient solution was controlled and nutrient composition in the root medium was frequently monitored. Thus, there was no shortage of nutrition and water. There were no yellow leaves, and greenness is not expected to reflect the differences in photosynthesis.

5.4.4. Why did photosynthesis-related QTLs predict yield in our population?

Canopy photosynthesis (whole plant photosynthesis) is different from leaf photosynthesis. Therefore, several studies have tried to bridge the gap between leaf photosynthesis and canopy photosynthesis via crop growth models (Gu *et al.*, 2014*a*; Yin *et al.*, 2022). Why did the combination of photosynthesis-related QTLs predict yield with a high prediction accuracy in this study? The discrepancy between canopy photosynthesis and leaf photosynthesis (per unit leaf area) may arise from the difference in leaf area or leaf area index (LAI). In this population, we confirmed that LAI or the fraction of intercepted light had a positive effect on total biomass, but a smaller effect on yield (Chapter 4; the correlation between yield and the fraction of intercepted light = 0.26). On the other hand, LUE_{adp} positively correlated with total biomass and also with yield (Chapter 4; the correlation between yield and LUE_{adp} = 0.64). Therefore, yield was mainly determined by LUE_{adp} but not so much by LAI or light interception, which might make it possible for photosynthesis-related QTLs to predict yield in this study.

Pruning experiments and carbohydrate measurements (Li *et al.*, 2015) have shown that indeterminate fruit-bearing tomato plants in greenhouse production are source limited. In this experiment, we observed a trade-off between the number of fruits and the individual fruit fresh weight at population level (Chapter 2), which agrees with source limitation at the population level. We also observed that total biomass highly correlated with yield (correlation (r) = 0.65), whereas harvest index did not (r = 0.03 not significant) (Chapters 2 and 3). Total biomass was determined by LUE_{adp} and light interception, but LUE_{adp} had much more influence on yield than light interception due to the negative correlation between harvest index and light interception (r = −0.71) (Chapter 4). Thus, our population is source-limited, and LUE_{adp} is a strong determinant for yield. LUE_{adp} could include the photosynthesis effect and light-distribution effect (e.g., plant morphology). Therefore, the breeding experiment in this study made it possible to seek the relationship between photosynthesis and yield.

Photosynthesis consists of three main components: light reactions (electron transport), CO_2 fixation, and photorespiration (Sharwood *et al.*, 2022). Photorespiration competes with CO_2

assimilation for the products of electron transport. In the absence of photorespiration (e.g., under high CO_2 and low O_2 concentrations), electron transport measured by chlorophyll fluorescence can be a proxy for CO_2 assimilation (Genty *et al.*, 1989; Baker, 2008). Recently, modelling studies have suggested that electron transport capacity has a greater potential for increasing the productivity of crops than commonly studied maximum Rubisco activity, which largely determines the light-saturated photosynthetic rate (Gu *et al.*, 2014*a*; Yin *et al.*, 2022). Electron transport is important because most leaves are in a canopy and thus are in light-limiting condition (electron transport limited) during most hours (Yin *et al.*, 2022). Bioengineering studies also showed that improving photosynthetic electron transport can boost biomass production (reviewed by Walter and Kromdijk, 2022 (Walter and Kromdijk, 2022)). In this study, yield was predicted with a high prediction accuracy from all QTLs for all traits measured by chlorophyll fluorescence together (Figure 5c). Therefore, our results indicate that electron transport is not just a proxy for photosynthesis, but also useful for plant breeding to improve yield.

5.4.5. Difficulty in QTL detection for photosynthesis

Capturing the genetic variation in photosynthesis is challenging because of the nonlinear responses of photosynthesis to the environmental factors (e.g., light, CO_2 concentration, temperature), innate variation within a leaf (Xiong *et al.*, 2015), and measurement error (Gu *et al.*, 2012*b*). We observed zero estimated heritability for most of the traits measured by chlorophyll fluorescence. However, estimated heritability equal to zero does not necessarily mean that we cannot detect QTLs for these traits. Estimated heritability is calculated on the genotype basis (in this study, 342 recombinant inbred lines (RILs) basis), whereas QTL analysis quantifies the contrast between alleles at single locus. That is, at each marker position, 342 RILs form four parent groups for the statistical analysis. Thus, genetic variance at each marker position could be detected. The low estimated heritability (based on raw measured phenotypic values) indicates high environmental variance. A part of the environmental variance could be removed by a correction of raw phenotypic values for spatial, temporal and some environmental factors. This procedure might have improved the chance of QTL detection. We also measured a large number of plots ((342 RILs × Test parent 1 + 342 RILs × Test parent 2), repeated three times), which increased the statistical power for QTL detection. QTLs predicted their own trait with a higher prediction accuracy than the 95th percentile of prediction accuracy by the same number of randomly selected markers (Supplementary Figure 4). The difference between prediction accuracy by QTL markers and the median of prediction accuracy by random markers is the added value by QTLs. In addition, most of LUE_{adp} QTLs (7 out of 9) were at the same position as QTLs for photosynthesis-related traits. Especially $LUE_{adp}1.2$ colocalized with J_{max}, ETR1500(fit), ETR1500(obs), and Y(II)1500 (wild parent 1 showed the highest positive effect on all these traits), which was also at the same position as SLA1.2 with wild parent 1 decreasing SLA (Chapter 4). Decreasing SLA (therefore increasing leaf thickness) is often considered to increase

photosynthesis (light-saturated photosynthesis capacity (A_{max}) and J_{max}) (Yin *et al.*, 2022). Because LUE_{adp} and SLA were independently measured from photosynthesis-related traits, these QTL coincidences support the credibility of QTL analysis in this study. Thus, despite of the difficulties in QTL detection for photosynthesis, this study detected QTLs for photosynthesis-related traits.

5.4.6. Which photosynthesis-related trait is important for yield? — Many photosynthesis-related genes near interesting QTLs

The light-saturated photosynthesis capacity (A_{max}), related J_{max} and the maximum carboxylation capacity of Rubisco (V_{cmax}) have been commonly studied (Yin *et al.*, 2022). However, the natural light environment fluctuates. Fluctuations in a natural environment, make it necessary that regulatory mechanisms within the leaves allow for a rapid shift from a low to a high photosynthetic rate (Long et al., 2006; Lawson et al., 2012). In vivo analysis combined with a canopy ray tracing model showed that over 20% of productivity is lost in wheat by the slow induction of photosynthesis during shade-to-sun transitions (Taylor and Long, 2017). On the other hand, when shifting from high to low light intensity, enhancing the rate of NPQ relaxation boosts photosynthesis and crop growth in the field (Kaiser *et al.*, 2018). Bioengineering of an accelerated response to natural shading events (fast NPQ relaxation) in tobacco increased leaf CO_2 uptake and dry matter productivity by about 15% in fluctuating light (Kromdijk *et al.*, 2016). Thus, physiological studies (on the same genotype) or bioengineering studies (altering a limited number of genes without changing other genes) show the importance of each photosynthetic trait. But what happens in natural genetic variation within a breeding population? Which trait is more related to yield in our breeding population? Yield was predicted with a high prediction accuracy (in fitted model r_{fitted} = 0.52; in cross-validation r_{cv} = 0.46; Table 1c) by five of the 6 QTLs for (-)t50 ETR induction. This might support the importance of fast induction of photosynthesis for yield. However, the haplotype effects were opposite between (-)t50 ETR induction and yield in chromosome 7. It is not clear whether a gene improving (-)t50 ETR induction has a detrimental effect on yield, or the opposite effects were merely caused by the closely located ChlC7.1. Many other QTLs for t50 ETR induction have similar haplotype effects between (-)t50 ETR induction and yield, therefore these QTLs are useful to improve (-)t50 ETR induction, which improves also yield. On the other hand, we detected 6 QTLs for (-)t50 NPQ relaxation, which predicted the own trait with high prediction accuracy (r_{fitted} = 0.55; r_{cv} = 0.43 ; Table 1b). However, QTLs for (-)t50 NPQ relaxation did not predict yield better than random markers (Table 1c; Supplementary Figure 7). Thus, a quick relaxation of NPQ might not be important for yield in this experiment. One consideration to take is that we evaluated NPQ relaxation from around 443 µmol m^{-2} s^{-1} to ambient light intensity at dawn (mean 29 µmol m^{-2} s^{-1}). NPQ might need to be evaluated under higher light intensity, which was not possible in our study. Regarding J_{max}, we detected only one QTL. This QTL is interesting because it colocalized with QTLs for LUE_{adp}, total biomass, and yield, for all of which wild

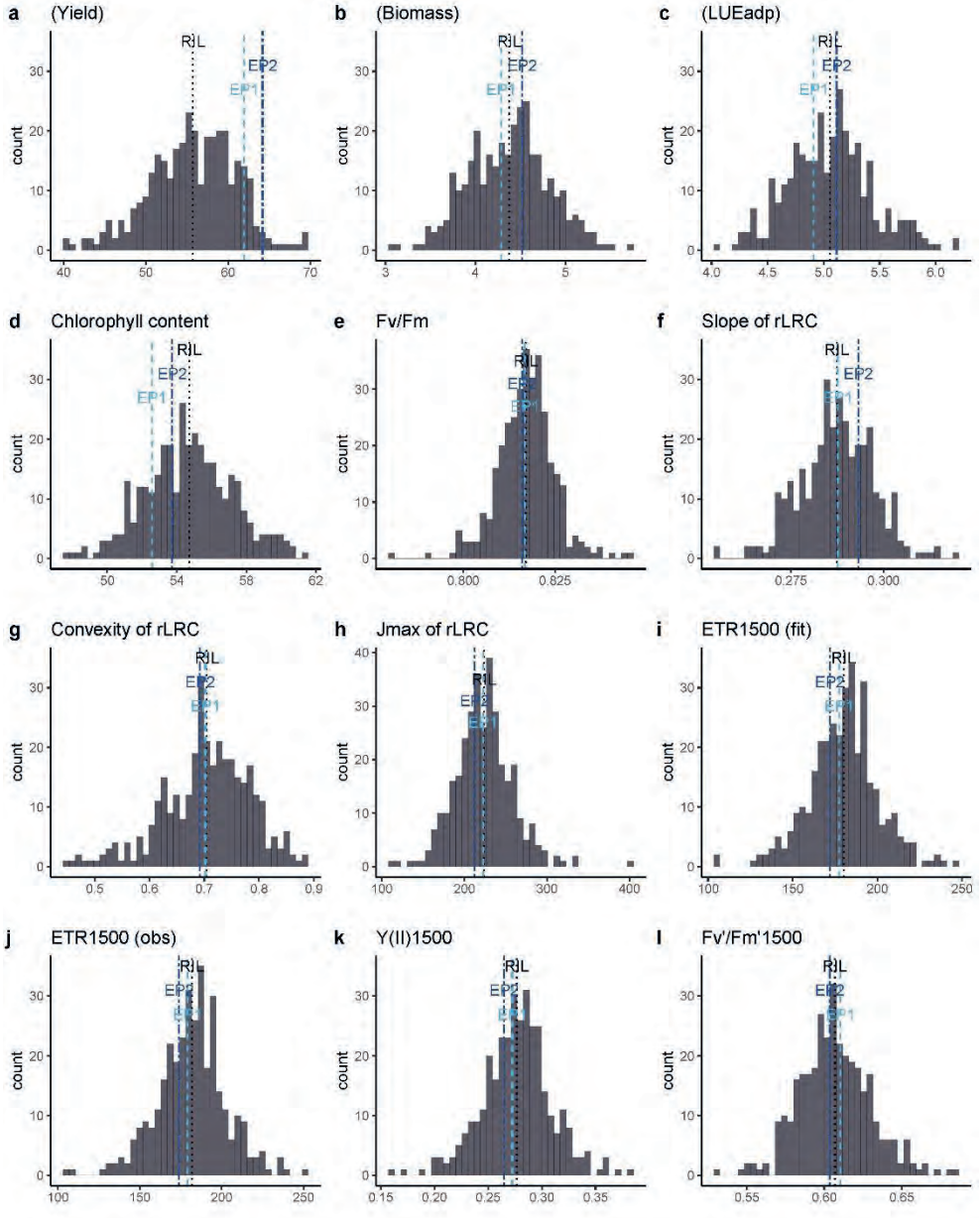
parent 1 showed the highest positive effect. This result suggests that wild germplasm harbours useful genes, which improve photosynthesis-related traits and yield in the elite genetic background.

Candidate genes were suggested for some interesting QTLs (within 2 Mbp, which was a narrower region than 20 cM) (Supplementary Table 3; Supplementary Table 4). The candidate genes were manually searched on a website of Solanaceae Genomics Network by using ITAG4.1 and ITAG2.4 gene models (https://solgenomics.net/jbrowse_solgenomics/). We found several candidate genes related to electron transport, chlorophyll or photosynthesis. For J_{max}-related QTLs on chromosome 1 (QTLs for J_{max}, ETR1500(fit), ETR1500(obs), Y(II)1500), for which the allele effects were the same between yield and photosynthesis-related traits, genes related to photosystem II (Photosystem II reaction center protein Z), as well as photosystem I (Chlororespiratory reduction, Ferredoxin), were detected. On chromosome 7, two chlorophyll-content QTLs and 1 (-)t50-ETR-induction QTL showed opposite allele effects between yield and photosynthesis-related traits. That is, elite parent 1 reduced chlorophyll content whereas increased yield; wild parent 2 speeded up ETR induction, however, did not increase yield. In this region, 1 photosystem II CP43 chlorophyll apoprotein was found, which is an antenna protein in the light-harvesting complexes (Walter and Kromdijk, 2022). This supports the hypothesis of reducing greenness or antenna protein increases canopy photosynthesis (Walter and Kromdijk, 2022; Yin *et al.*, 2022). On the other hand, four Plastocyanin-like genes were detected near (-)t50 ETR induction 7.1. Plastocyanin is the long-range electron carrier between photosystem II and photosystem I (Höhner *et al.*, 2020). Genes coding Photosystem II D2 protein and Photosystem II family protein, Translocase of chloroplast, Protochlorophyllide reductase-like protein, and many Cytochromes P450 were also detected near QTLs on chromosome 7. Some of these genes might be involved in (-)t50 ETR induction. A faster ETR induction is generally considered to be beneficial for yield. Thus, it is not clear whether the negative QTL-effect relationship between yield and (-)t50 ETR induction in (-)t50 ETR induction 7.1 might have been caused by near Chlorophyll-content QTLs (potentially antenna protein) or the same genes influenced both chlorophyll content index and ETR induction. In the latter case, the allele of wild parent 2 increases the chlorophyll content index and speeds up ETR induction, resulting in lower yield compared to elite parent 1. Because many photosynthesis-related genes were located in this region, positive parental-haplotype alleles are likely different among genes. To utilize beneficial variation from wild parents, further research is necessary to dissect these genes' effects. Eventually, it is possible to combine positive alleles for these genes from elite as well as wild parents to improve photosynthesis and yield.

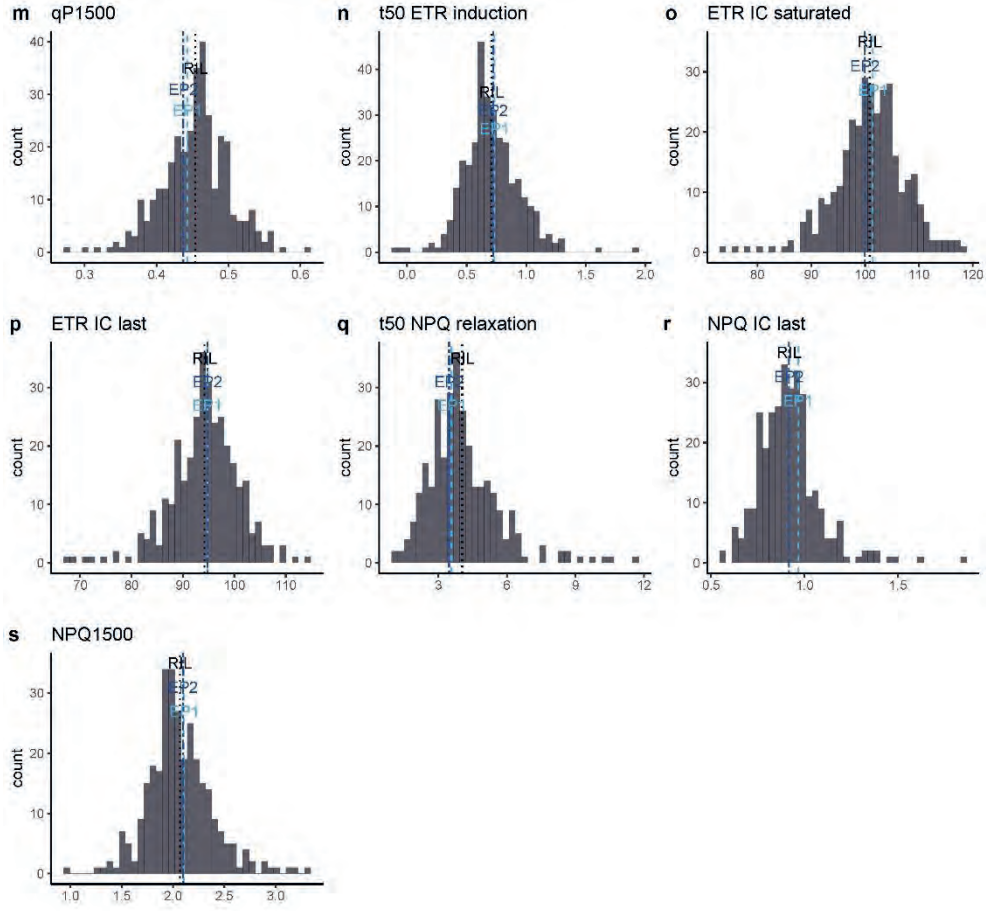
5.5. Conclusion

No correlation of leaf photosynthesis with yield in many studies and the failure of high-yielding breeding by targeting photosynthesis has caused a long-standing discussion whether photosynthesis is a target trait for plant breeding to increase yield. In this study, we detected QTLs for photosynthesis-related traits and these QTLs together successfully predicted yield (the prediction accuracy of the fitted model was 0.74 using the 20 QTLs for all photosynthesis-related traits, which was larger than the 95th percentile of prediction accuracy of the fitted model using 20 random markers (0.68)). However, counterintuitively, the positive allele effects of QTL for photosynthesis-related traits do not necessarily result in a high yield. Some of the opposite effects between photosynthesis-related traits and yield might represent the discrepancy between leaf photosynthesis and whole-plant photosynthesis. That is, reducing light absorption (leaf photosynthesis) at upper leaves may reduce leaf photosynthesis of upper leaves but allows deeper light penetration into the canopy, improving whole-plant photosynthesis. The difficulty of connecting leaf photosynthesis and yield also arose from other traits. The success of yield prediction by photosynthesis-related QTLs in this study might be due to the characteristics of this breeding population: source limited, thus biomass production is the determinant for yield; higher importance of light use efficiency rather than light interception for yield. Our results indicate that electron transport (light reactions), which can be estimated from chlorophyll fluorescence measurements, is useful for high-yielding breeding, and photosynthesis plays a role to determine yield. Natural genetic variation from wild relatives can improve photosynthesis and yield in the genetic background of the elite parent. To utilize this natural variation, it is necessary to consider the relationship with other traits and the relationship between leaf photosynthesis and whole-plant photosynthesis. Our results demonstrate that photosynthesis-related QTLs provide a new tool to guide breeding efforts, which will become even more powerful if the genetic effects of photosynthesis-related traits are dissected in more detail.

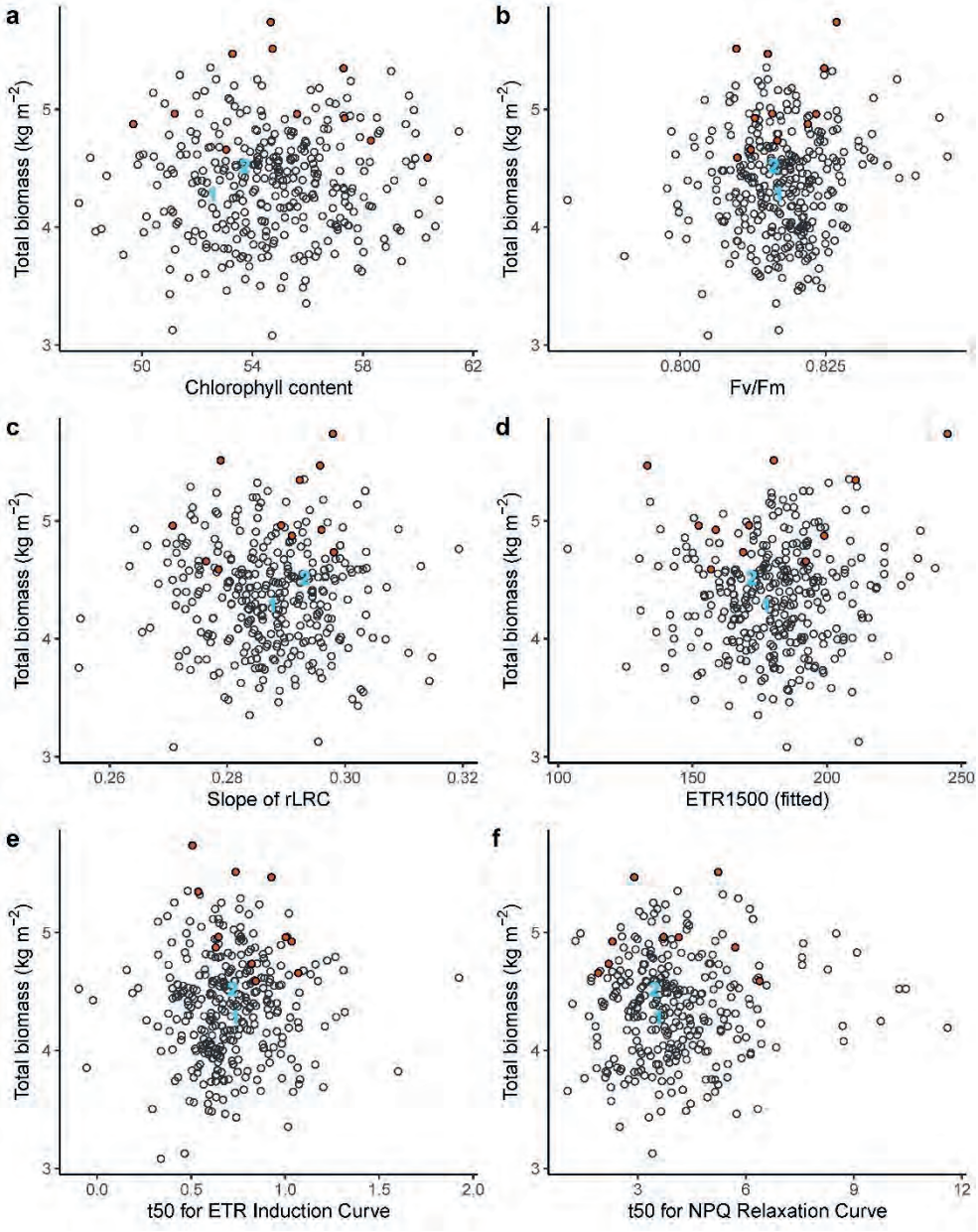
Supporting information



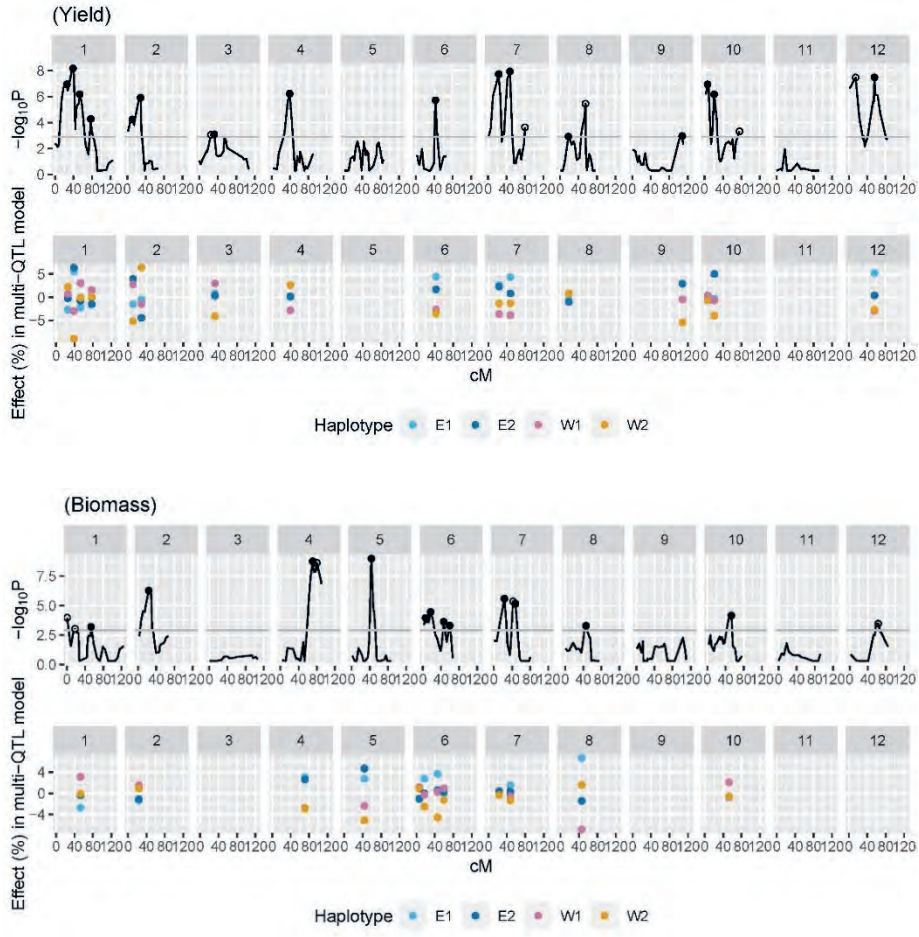
Supplementary Figure 1 Caption next page.



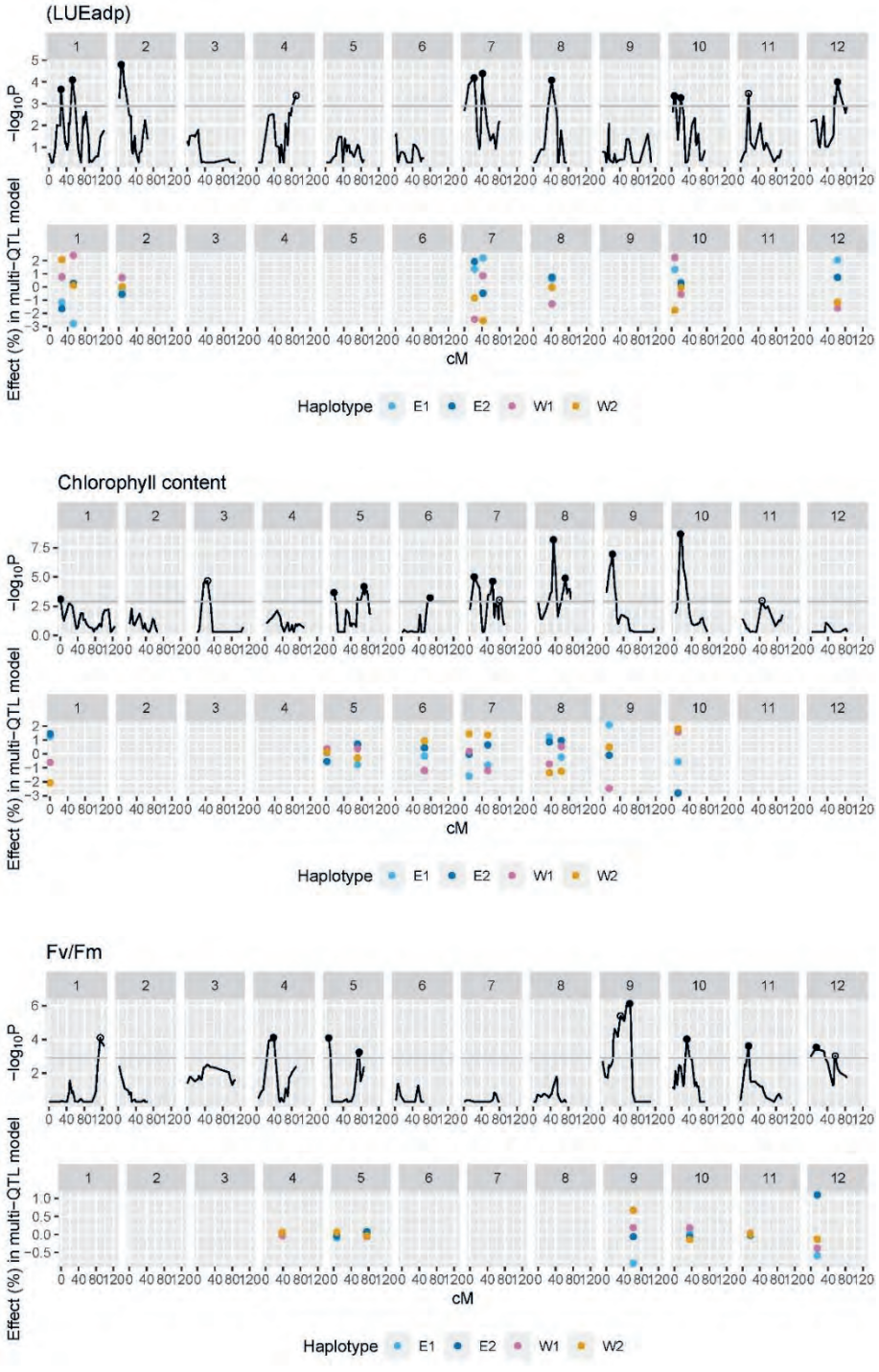
Supplementary Figure 1 Distributions of the best linear unbiased estimates (BLUEs) per recombinant inbred line (RIL). The mean of RILs is indicated by the vertical black dashed line, the performance of the two elite parents by vertical blue dashed lines (the light blue colour for the elite parent 1 (EP1) and the blue colour for the elite parent 2 (EP2)). Traits were explained in Table 1.



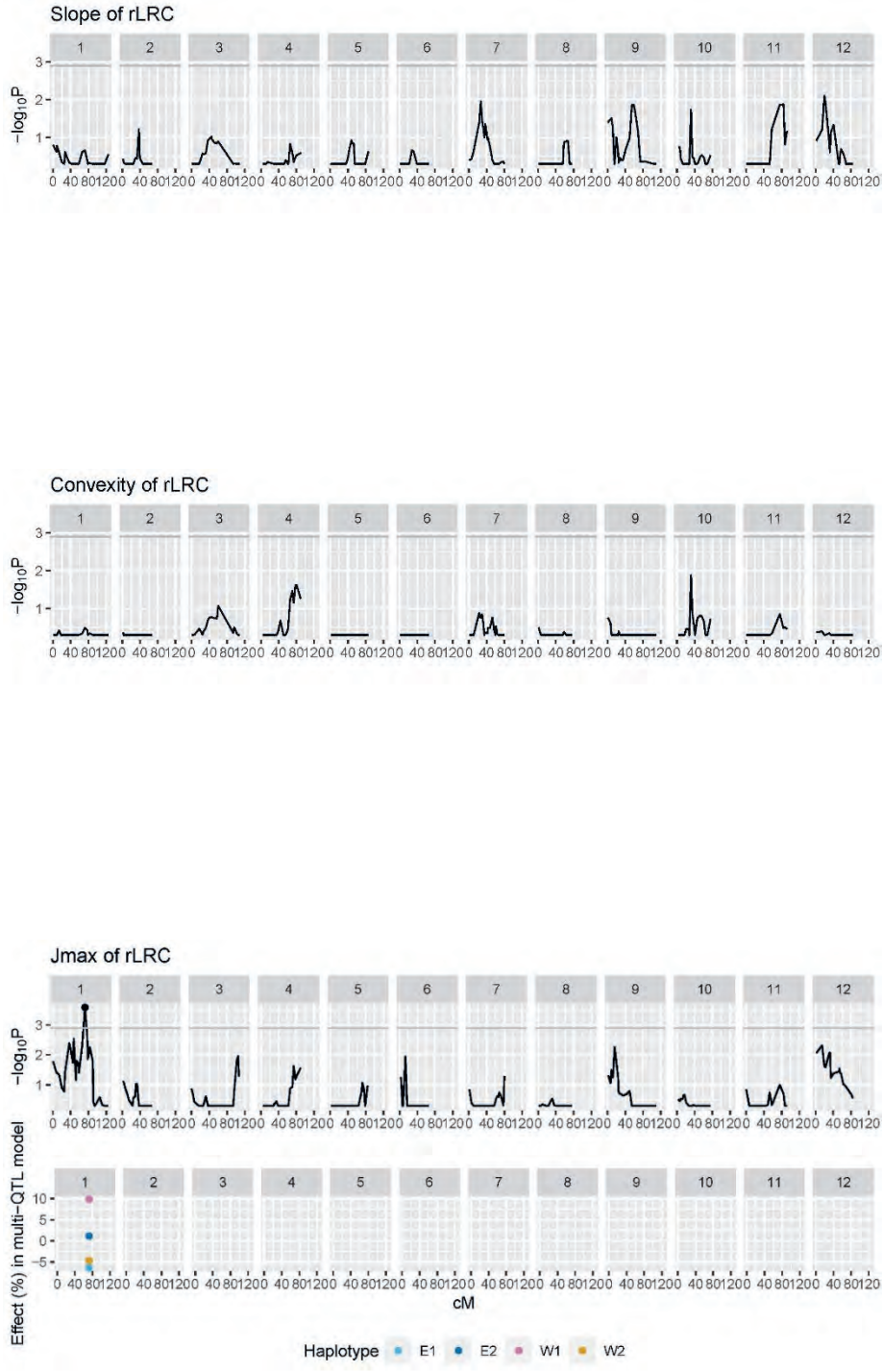
Supplementary Figure 2 Phenotypic correlation between total (plant) biomass and a photosynthesis-related trait. Red dots indicate RILs showing higher yield than elite parents. Blue numbers indicate two elite parents. Traits were explained in Table 1.



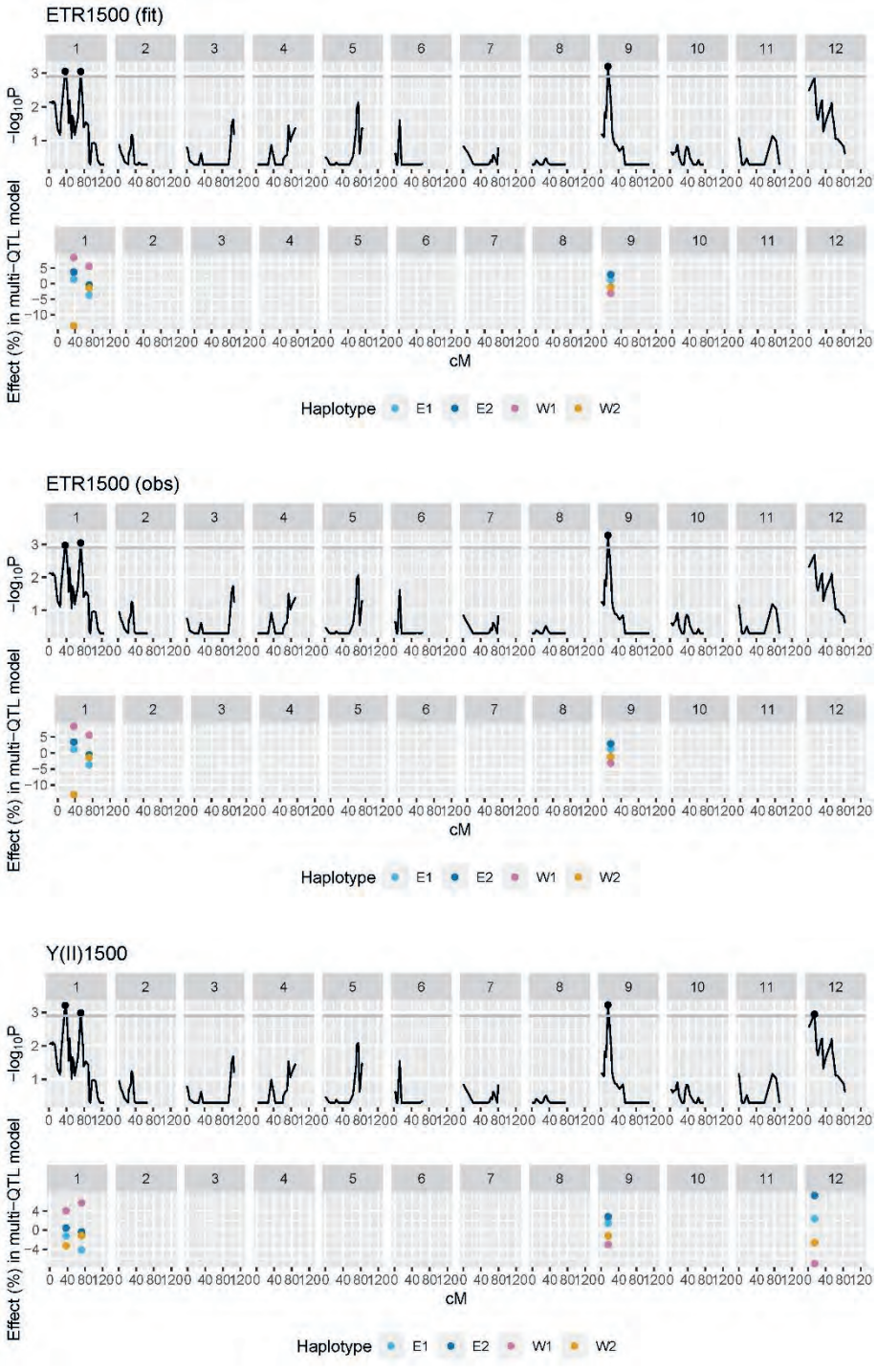
Supplementary Figure 3 Line plots of single-QTL analyses and effects in multi-QTL models. On the upper panels for each trait, line plots indicate -log10(p) values in single-QTL analyses, where dots indicate multi-QTLs, and empty circles indicate single-QTLs that were not retained in the final multi-QTL models. Lower panels for each trait indicate multi-QTL effects (percentage; proportion of the effect to the trait mean (intercept in the multi-QTL model)). Traits were explained in Table 1. Data of total yield (a), total biomass (b), and adapted light use efficiency (c) were taken from the previous study (Tsutsumi-Morita et al., 2021; Chapter 4) as references.



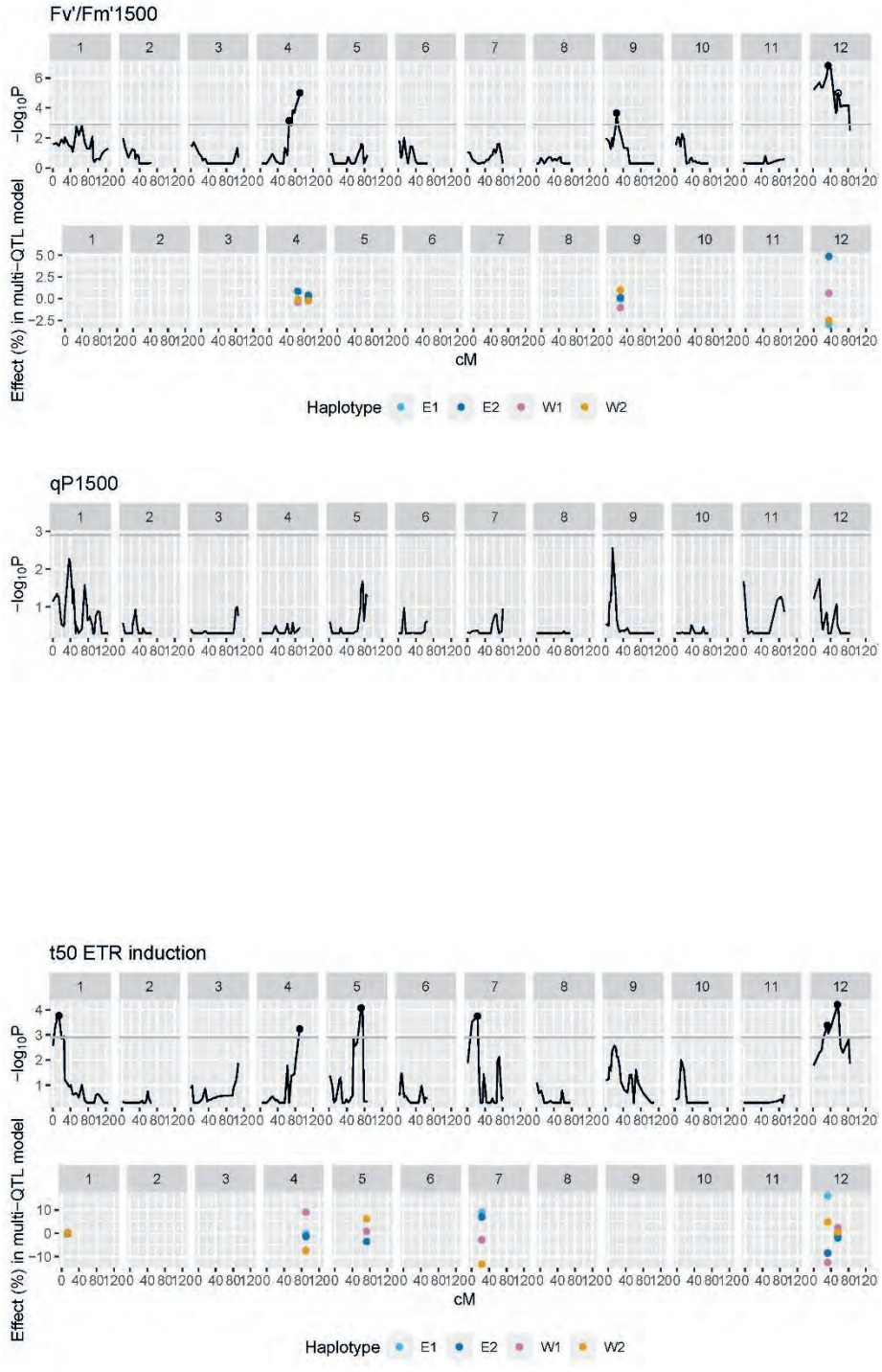
Supplementary Figure 3 Continued.



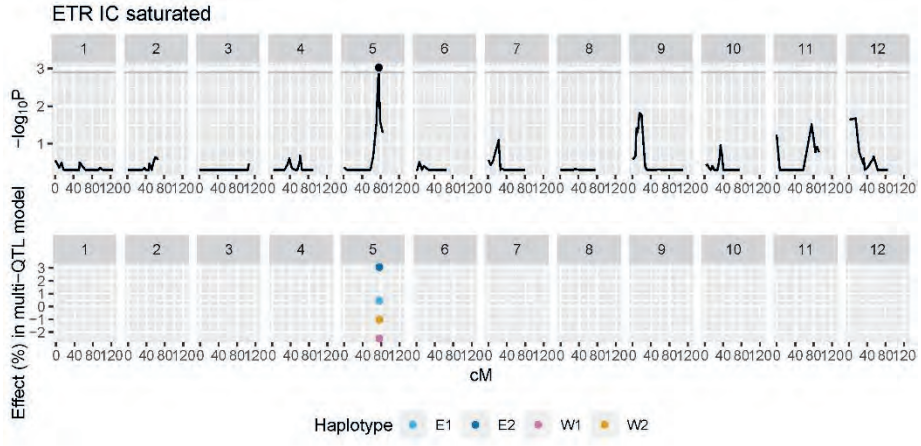
Supplementary Figure 3 Continued.

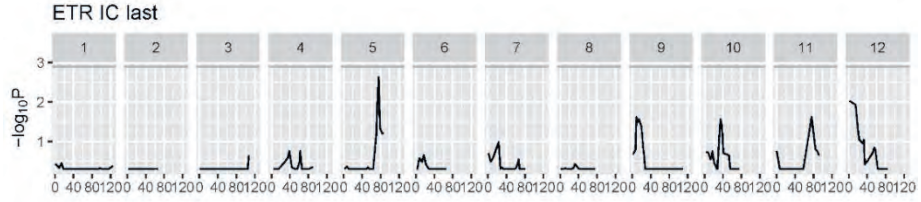


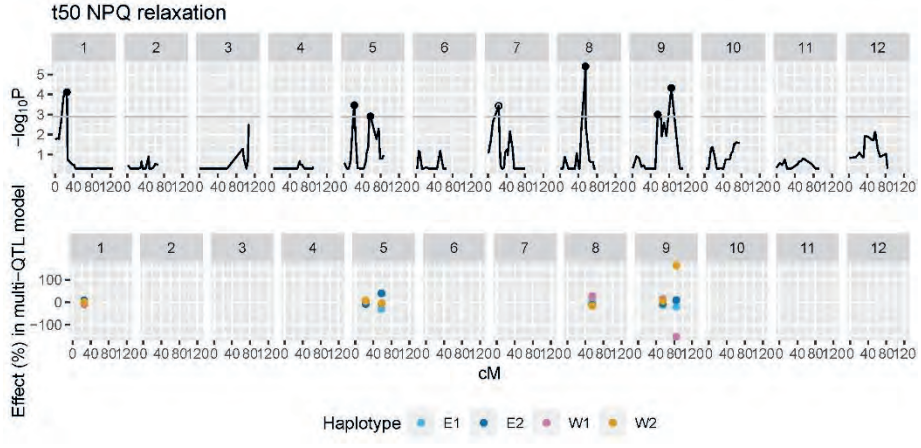
Supplementary Figure 3 Continued.



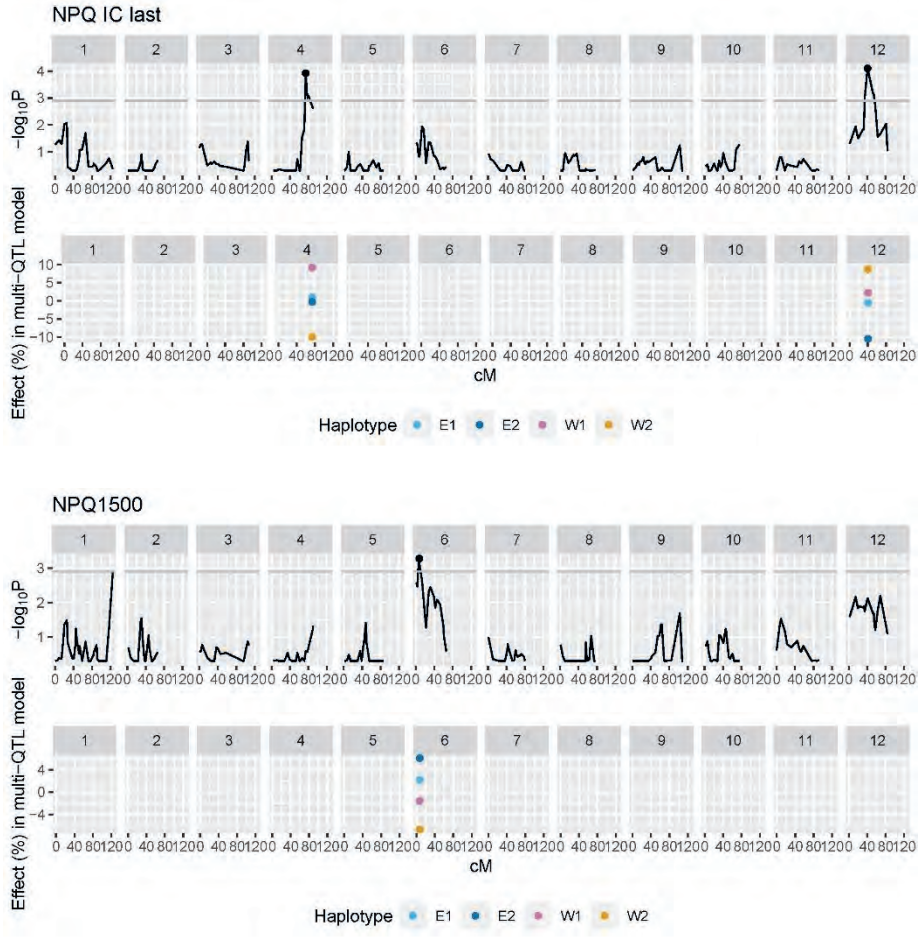
Supplementary Figure 3 Continued.



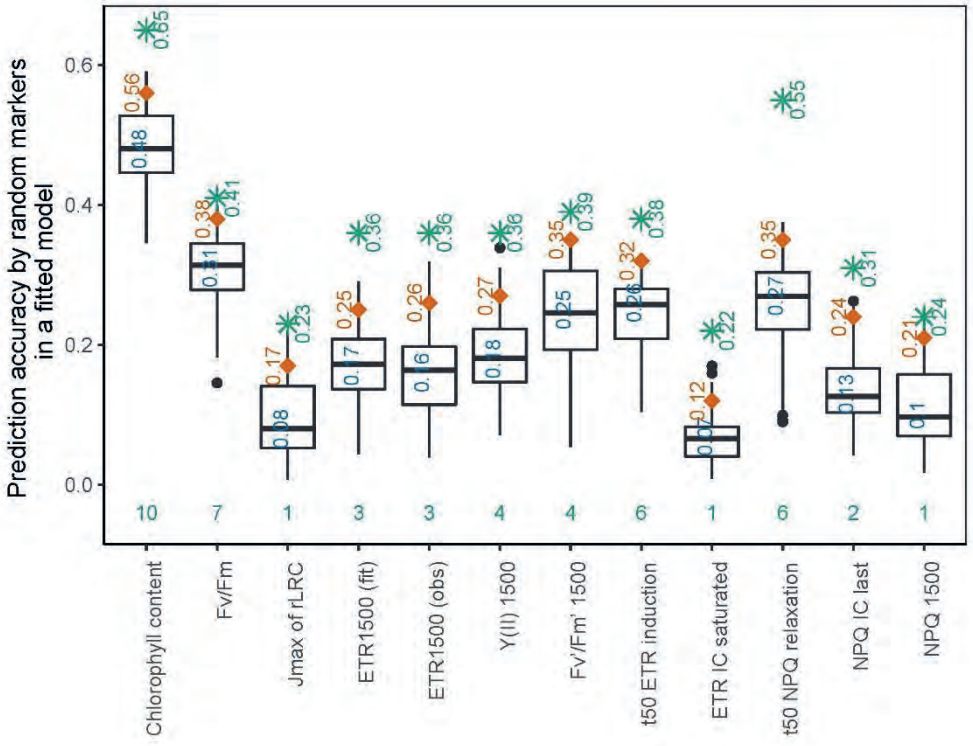




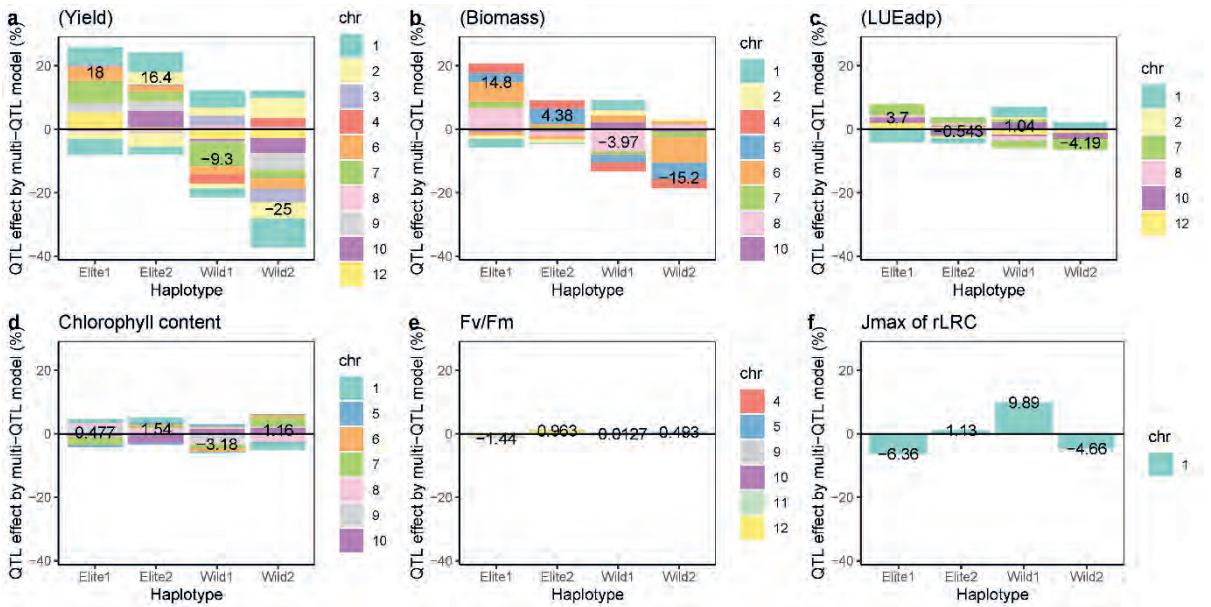
Supplementary Figure 3 Continued.



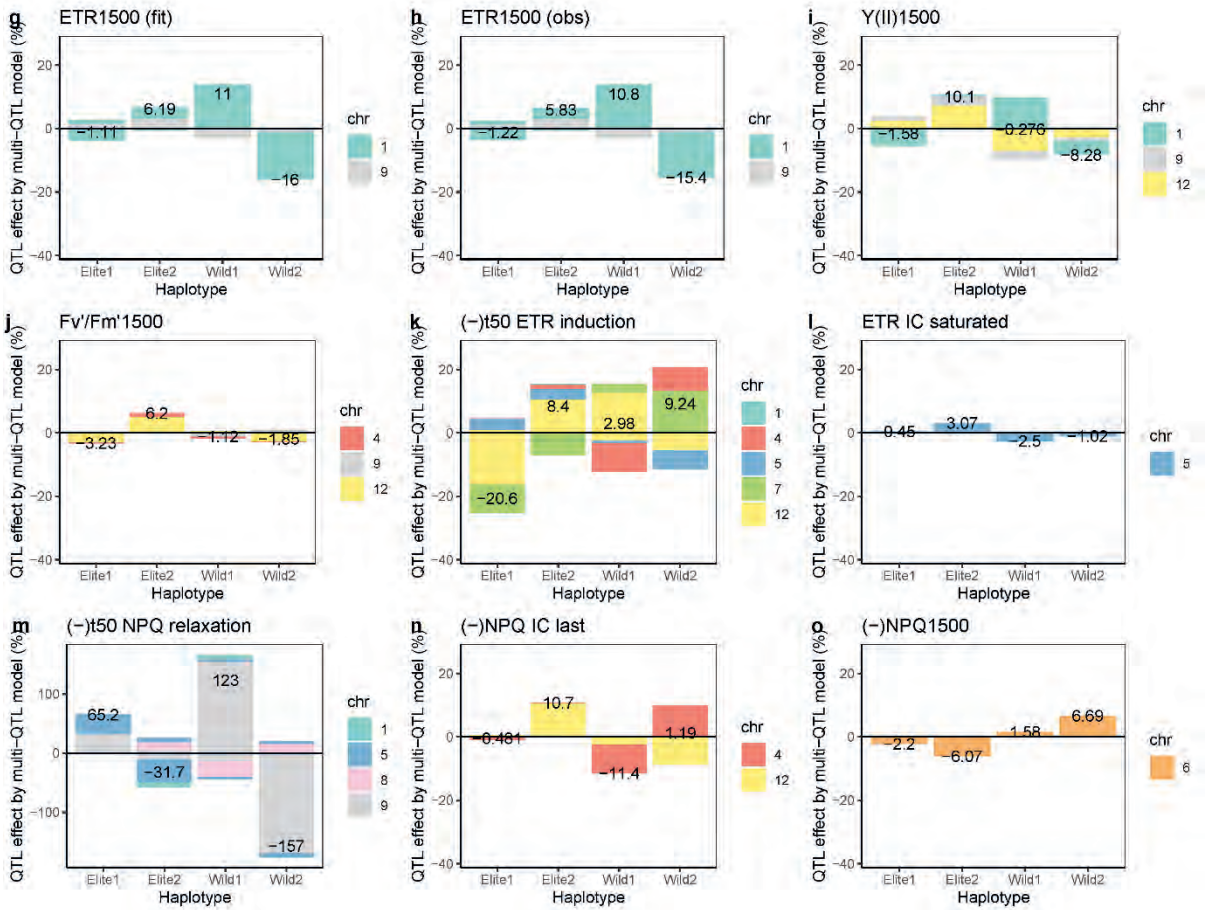
Supplementary Figure 3 Continued.



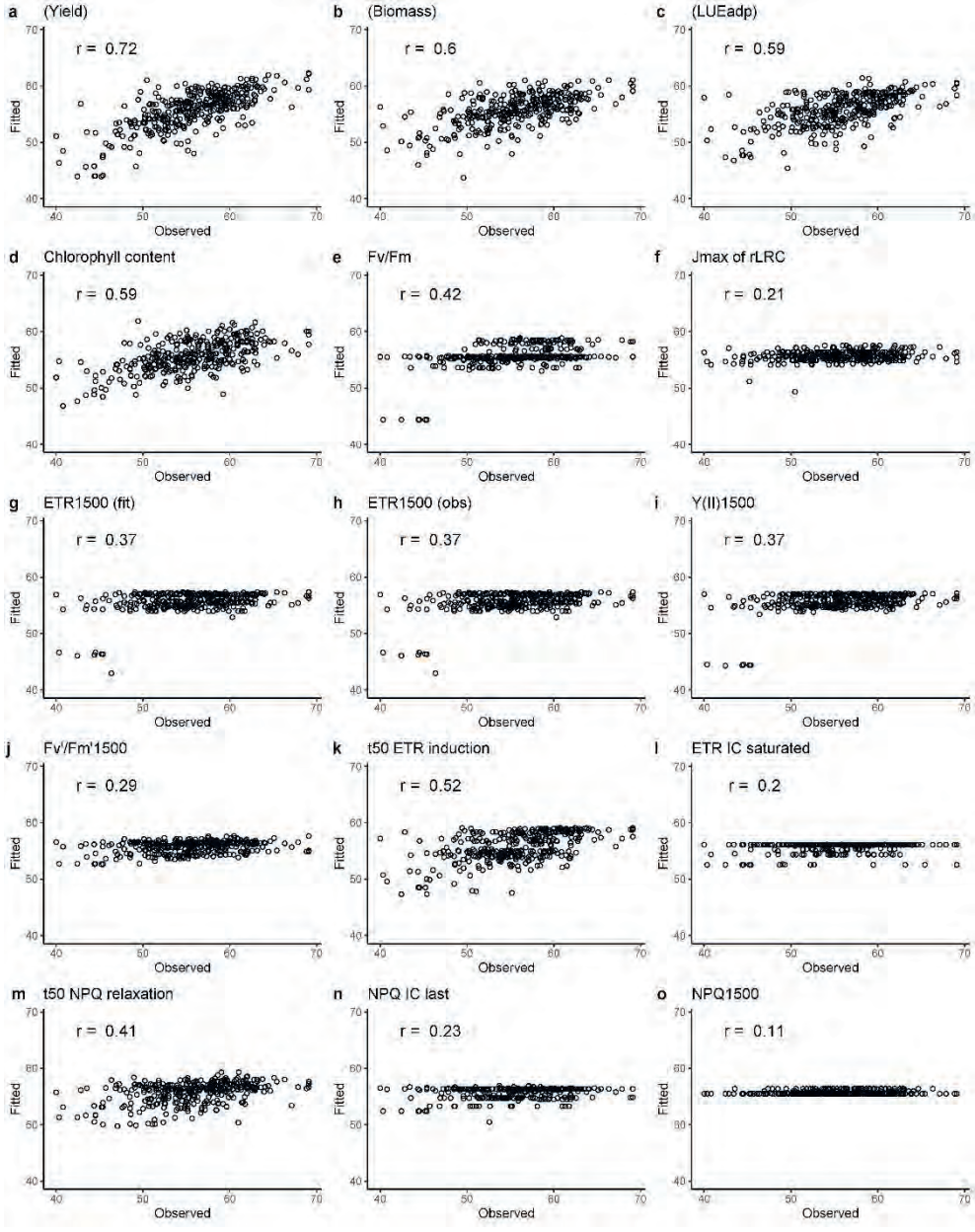
Supplementary Figure 4 Prediction accuracies of photosynthesis-related traits predictions. Box plots indicate accuracies of predictions for each trait by a set of randomly selected markers, which was iterated 100 times. The number of random markers, which is the same as the number of multi-QTLs for each trait, is shown at the bottom of the figure with green colour. The median of the box plot is shown by the blue number, and the 95th percentile of the box plot is shown by the red diamond with the red number. The prediction accuracy by QTLs for each trait is indicated by the asterisk with the green number. Traits were explained in Table 1.



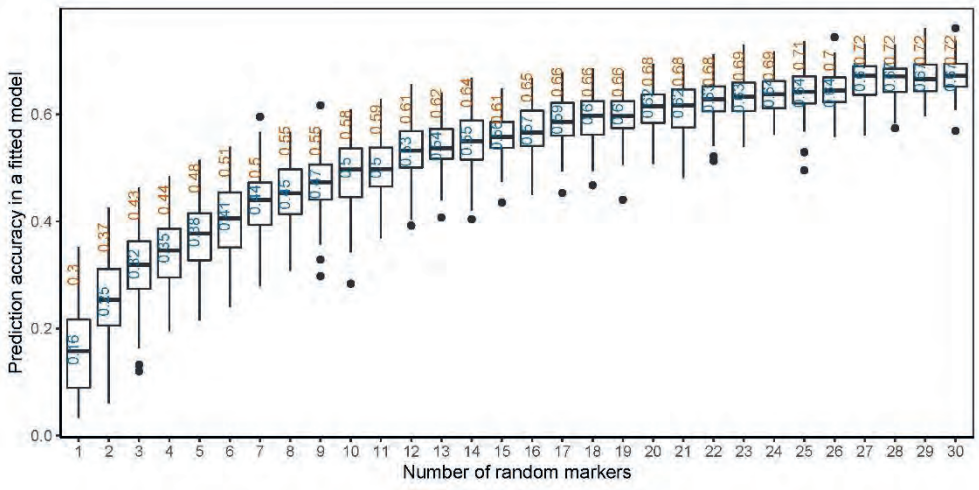
Supplementary Figure 5 Haplotype effects per chromosome and multi-locus predictions for the parents. Colours haplotype effects correspond to chromosomes, size corresponds to the proportion of the effect to the trait mean (intercept in the multi-QTL model). The numbers in the figure indicate multi-locus predictions. Traits were explained in Table 1. Data of total yield (a), total biomass (b), and adapted light use efficiency (c) were taken from the previous study (Tsutsumi-Morita et al., 2021; Chapter 4).



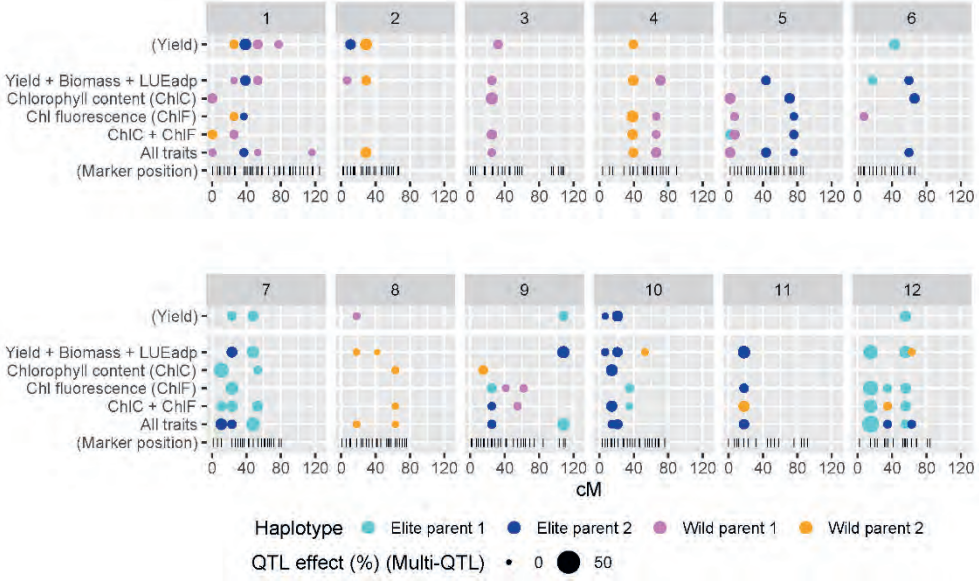
Supplementary Figure 5 Continued.



Supplementary Figure 6 Comparison between observed yield and fitted yield predicted by QTLs for each photosynthesis-related trait. Each dot represents a recombinant inbred line (RIL). The r indicates prediction accuracy (the correlation coefficient between observed and fitted values). Traits were explained in Table 1.



Supplementary Figure 7. Prediction accuracy (correlation coefficient between observation and prediction) in a fitted model by using sets of markers selected at random. Blue numbers indicate the median, and red numbers represent the 95th percentile of 100 iterations.



Supplementary Figure 8 Multi-QTL locations and effects of yield predictions by a set of QTL-markers for secondary traits. The horizontal axis indicates chromosome number (1 – 12) and genetic position (cM) within the chromosome. The colour of dots indicates haplotype showing the highest positive effect. The size of dots corresponds to the size of the QTL effect (proportion of trait mean). Yield is predicted by single-QTLs for some traits, the names of which are indicated on the left of the figure. Please see also Figure5. Abbreviation: LUE$_{adp}$, adapted light use efficiency; ChlC, Chlorophyll content index; ChlF, traits measured by chlorophyll fluorescence.

Supplementary Table 1 QTLs list. QTL positions, SolCAP marker names, -log$_{10}$(p) values for test on variance component of haplotype effects, and haplotype effects.

						QTL effect from multi-QTL model			
Trait	Chromosome	cM	Position (bp)	SolCAP.marker	-log$_{10}$(*p*)	Elite parent 1	Elite parent 2	Wild parent 1	Wild parent 2
Chlorophyll content									
	1	0.5	310,894	CL004303-0524	3.10	0.690	0.791	-0.337	-1.144
	3	25	2,704,993	solcap_snp_sl_14355	4.67	NA	NA	NA	NA
	5	1.8	558,180	solcap_snp_sl_69104	3.66	0.032	-0.304	0.200	0.072
	5	71	63,161,791	solcap_snp_sl_12244	4.17	-0.425	0.382	0.203	-0.160
	6	66	48,691,982	solcap_snp_sl_24255	3.22	-0.094	0.239	-0.654	0.509
	7	10.8	1,557,551	solcap_snp_sl_68398	5.00	-0.873	-0.014	0.098	0.789
	7	53	62,093,144	solcap_snp_sl_53425	4.63	-0.438	0.352	-0.657	0.743
	7	68.6	65,232,683	solcap_snp_sl_71085	3.03	NA	NA	NA	NA
	8	36	56,253,744	solcap_snp_sl_18195	8.19	0.663	0.466	-0.391	-0.739
	8	63	62,865,804	solcap_snp_sl_15436	4.90	-0.130	0.526	0.291	-0.686
	9	15	1,514,198	solcap_snp_sl_7731	6.95	1.142	-0.052	-1.358	0.268
	10	14.5	2,335,513	solcap_snp_sl_17859	8.67	-0.306	-1.541	0.863	0.984
	11	45	6,343,916	solcap_snp_sl_34336	2.96	NA	NA	NA	NA
Fv/Fm									
	1	116.2	95,068,045	solcap_snp_sl_54704	4.10	NA	NA	NA	NA
	4	38.5	7,186,344	SGN-U574577_snp856_solcap_snp_sl_41764	4.10	-0.000046	-0.000239	-0.000207	0.000491
	5	7	1,311,252	SGN-U583136_snp1228_solcap_snp_sl_52714	4.08	-0.000723	-0.000276	0.000493	0.000505
	5	76	63,803,770	solcap_snp_sl_185	3.23	0.000175	0.000630	-0.000507	-0.000299
	9	41.2	4,543,446	solcap_snp_sl_39525	5.39	NA	NA	NA	NA

9	62	7,156,689	solcap_snp_sl_45141	6.12	-0.006513	-0.000506	0.001542	0.005478
10	35	59,632,727	solcap_snp_sl_2938	4.01	0.000213	-0.000555	0.001487	-0.001144
11	18	2,443,617	solcap_snp_sl_10596	3.61	-0.000098	-0.000173	0.000206	0.000065
12	15	1,971,483	solcap_snp_sl_12656	3.53	-0.004774	0.008965	-0.003117	-0.001076
12	57.6	63,580,692	solcap_snp_sl_31953	3.01	NA	NA	NA	NA
Jmax of Light Response Curve								
1	72.3	83,829,584	SL10943_357	3.59	-14.3	2.5	22.2	-10.5
ETR at PPFD = 1500 (fitted)								
1	37	4,597,945	solcap_snp_sl_59890	3.05	2.50	6.46	14.73	-23.70
1	72.3	83,829,584	SL10943_357	3.04	-6.57	-0.79	9.77	-2.42
9	16	1,692,427	solcap_snp_sl_57946	3.20	2.13	5.14	-5.37	-1.90
ETR at PPFD = 1500 (observed)								
1	37	4,597,945	solcap_snp_sl_59890	2.97	2.03	6.12	14.78	-22.94
1	72.3	83,829,584	SL10943_357	3.04	-6.50	-1.00	9.92	-2.41
9	16	1,692,427	solcap_snp_sl_57946	3.27	2.31	5.20	-5.57	-1.94
1	37	4,597,945	solcap_snp_sl_59890	3.21	0.00	0.00	0.01	-0.01
1	72.3	83,829,584	SL10943_357	2.99	-0.01	0.00	0.01	0.00
9	16	1,692,427	solcap_snp_sl_57946	3.22	0.00	0.01	-0.01	0.00
12	15	1,971,483	solcap_snp_sl_12656	2.95	0.01	0.02	-0.02	-0.01
Fv'/Fm' at PPFD = 1500 (observed)								
4	66	61,100,680	solcap_snp_sl_17921	3.16	-0.0017	0.0050	-0.0025	-0.0008
4	90	65,402,987	solcap_snp_sl_3894	5.01	0.0004	0.0023	-0.0016	-0.0012
9	25	2,673,326	solcap_snp_sl_39806	3.66	-0.0002	0.0006	-0.0062	0.0059
12	34.8	4,951,011	solcap_snp_sl_14762	6.84	-0.0177	0.0289	0.0037	-0.0149
12	57.6	63,580,692	solcap_snp_sl_31953	5.00	NA	NA	NA	NA

t50 for ETR Induction Curve								
1	14	1,809,372	solcap_snp_sl_60276	3.77	0.000	-0.001	0.001	0.000
4	90	65,402,987	solcap_snp_sl_3894	3.24	-0.002	-0.009	0.063	-0.052
5	72.4	63,351,010	solcap_snp_sl_12274	4.07	-0.025	-0.025	0.006	0.043
7	23.2	2,881,365	solcap_snp_sl_15786	3.75	0.063	0.049	-0.020	-0.092
12	32.6	4,385,199	solcap_snp_sl_32638	3.38	0.112	-0.059	-0.087	0.034
12	55.7	63,393,430	solcap_snp_sl_55530	4.21	-0.005	-0.013	0.016	0.003
Saturated ETR of Induction Curve								
5	76	63,803,770	solcap_snp_sl_185	3.03	0.45	3.05	-2.48	-1.02
t50 for NPQ Relaxation Curve								
1	25.6	2,879,156	solcap_snp_sl_60089	4.11	-0.028	0.354	-0.345	0.019
5	23	3,012,808	solcap_snp_sl_49009	3.46	-0.152	-0.317	0.169	0.300
5	58	61,885,368	solcap_snp_sl_22622	2.90	-1.223	1.599	-0.197	-0.179
7	23.2	2,881,365	solcap_snp_sl_15786	3.43	NA	NA	NA	NA
8	55	60,545,538	solcap_snp_sl_64736	5.40	0.043	-0.494	1.071	-0.620
9	55	5,120,587	CL016057-0229	3.00	-0.497	-0.246	0.541	0.201
9	84.7	67,350,015	CL016636-0429	4.33	-0.790	0.391	-6.233	6.635
NPQ at the last point of IC								
4	74	62,550,134	CL016270-0427_solcap_snp_sl_47056	3.92	0.0098	-0.0028	0.0872	-0.0942
12	40.8	38,735,374	solcap_snp_sl_38512	4.10	-0.0053	-0.0991	0.0216	0.0829
NPQ at PPFD = 1500								
6	7.4	3,135,509	solcap_snp_sl_35001	3.28	0.044	0.122	-0.032	-0.134

Supplementary Table 2 Biomass and adapted light use efficiency (LUE_{adp}) predictions by QTLs for a photosynthesis-related trait.

				Prediction accuracy (r)	
Predicted trait	by multi-QTLs for	Initial number of QTL markers	Final number of multi-QTLs	Fitted	Median in cross-validation
Biomass					
	(Yield)	16	8	0.52	0.40
	(Biomass)	12	12	0.60	0.49
	(LUEadp)	9	6	0.46	0.39
	Chlorophyll content	10	6	0.42	0.28
	Fv/Fm	7	6	0.33	0.20
	Jmax of rLRC	1	0	-	-
	ETR1500 (fit)	3	1	0.12	0.07
	ETR1500 (obs)	3	1	0.12	0.07
	Y(II)1500	4	2	0.12	0.02
	Fv'/Fm'1500	4	2	0.38	0.35
	t50 ETR induction	6	5	0.42	0.34
	ETR IC saturated	1	1	0.12	0.08
	t50 NPQ relaxation	6	4	0.31	0.22
	NPQ IC last	2	2	0.36	0.33
	NPQ1500	1	1	0.24	0.23
Adapted light use efficiency					
	(Yield)	16	10	0.52	0.38
	(Biomass)	12	10	0.45	0.33
	(LUEadp)	9	9	0.51	0.39
	Chlorophyll content	10	8	0.41	0.29
	Fv/Fm	7	3	0.30	0.25
	Jmax of rLRC	1	0	-	-
	ETR1500 (fit)	3	1	0.19	0.10
	ETR1500 (obs)	3	1	0.19	0.10
	Y(II)1500	4	2	0.20	0.16
	Fv'/Fm'1500	4	2	0.27	0.19
	t50 ETR induction	6	5	0.39	0.32
	ETR IC saturated	1	1	0.13	0.12
	t50 NPQ relaxation	6	4	0.33	0.22
	NPQ IC last	2	1	0.20	0.18
	NPQ1500	1	1	0.07	0.06

Supplementary Table 3 Categorization of candidate genes from Supplementary Table 4.

QTL	Chromosome	Position (cM)	Position (bp)	Antenna protein	Photosystem II (PSII)	Electron transport from PSII to PSI	Photosystem I (PSI)	Chlorophyll related / Photosynthesis related	Number of Cytochromes P450
ETR1500(fit) 1.1 ETR1500(obs) 1.1 Y(II)1500 1.1	1	37	4,597,945				Chlororespiratory reduction 42		4
Jmax 1.1 ETR1500(fit) 1.2 ETR1500(obs) 1.2 Y(II)1500 1.2	1	72.3	83,829,584		Photosystem II reaction center protein Z		Ferredoxin	2 Chlorophyll a-b binding protein	0
Chlorophyll content 7.1	7	10.8	1,557,551			4 Blue copper protein (Plastocyanin-like); 1 Blue copper-like protein (Cupredoxin)			
(-) t50 ETR Induction 7.1	7	23.2	2,881,365	Photosystem II CP43 chlorophyll apoprotein (photosystem antenna protein-like)	Photosystem II D2 protein (Photosynthetic reaction centre)			Translocase of chloroplast 90	3
Chlorophyll content 7.2	7	53	62,093,144		Photosystem II family protein	Blue copper protein (Plastocyanin-like)		Protochlorophyllide reductase like protein	14
Chlorophyll content 8.1	8	36	56,253,744		PsbP			2 Chlorophyll a-b binding protein	1
Chlorophyll content 10.1	10	14.5	2,335,513					2 Chlorophyll a-b binding protein Protochlorophyllide reductase Thioredoxin	7

Supplementary Table 4 Candidate genes searched from Solanaceae Genomics Network (https://solgenomics.net/jbrowse_solgenomics/)

Search method	ITAG4.1 gene models / ITAG2.4 gene models	Position	Description
ETR1500(fit), ETR1500(obs), and Y(II) at 1500; chr1 37cM, 4597945bp, solcap_snp_sl_59890			
	Solyc01g009245.1.1	SL4.0ch01:3218469..3220311 (- strand)	Chlororespiratory reduction 42 (AHRD V3.3 *** A0A0F7J401_9ROSI)
	Solyc01g009370.2.1	SL4.0ch01:3501372..3507803 (- strand)	Cytochrome P450 (AHRD V3.3 *** A0A2U1MC34_ARTAN)
ETR, Y(II) at 1500; chr1 37cM	solcap_snp_sl_59890	4597945bp	
	Solyc01g010260.3.1	SL4.0ch01:5084760..5086957 (+ strand)	Cytochrome 93A3 (AHRD V3.3 *** A0A1U8H5Z8_CAPAN)
	Solyc01g010490.4.1	SL4.0ch01:5403494..5405729 (- strand)	Cytochrome P450 (AHRD V3.11 *** tr|A0A2U1NH79|A0A2U1NH79_ARTAN)
	Solyc01g010640.3.1	SL4.0ch01:5682950..5683902 (+ strand)	NAD(P)-linked oxidoreductase superfamily protein (AHRD V3.3 *** Q949S6_ARATH)
	Solyc01g010900.3.1	SL4.0ch01:6386385..6391425 (- strand)	Cytochrome P450 (AHRD V3.3 *** A0A200QEP7_9MAGN)
Jmax chr1 72.3cM, 83829584 bp, SL10943_357			
	Solyc01g101050.3.1	SL4.0ch01:83184635..83185777 (+ strand)	Peroxidase (AHRD V3.3 *** K4B1D5_SOLLC)
	Solyc01g102490.3.1	SL4.0ch01:83551025..83561110 (+ strand)	NADPH:adrenodoxin oxidoreductase
	Solyc01g102610.3.1	SL4.0ch01:83645044..83649566 (- strand)	Ferric reduction oxidase 6 (AHRD V3.3 *** A0A2G3BUN6_CAPCH)
	Solyc01g102770.1.1	SL4.0ch01:83768878..83769066 (- strand)	Photosystem II reaction center protein Z (AHRD V3.3 *** A0A1U8QD60_CAPAN)
Jmax chr1 72.3cM	SL10943_357	83829584 bp	

		Solyc01g10 3920.3.1	SL4.0ch01:84715169..84 715923 (- strand)	Ferredoxin (AHRD V3.3 *** K4B1W7_SOLLC)
		Solyc01g10 4860.3.1	SL4.0ch01:85492687..85 494954 (+ strand)	Peroxidase (AHRD V3.11 *** tr|K4B258|K4B258_SOLLC)
		Solyc01g10 5030.3.1	SL4.0ch01:85633660..85 635047 (+ strand)	Chlorophyll a-b binding protein
		Solyc01g10 5050.3.1	SL4.0ch01:85639523..85 640691 (+ strand)	Chlorophyll a-b binding protein
		Solyc01g10 5070.3.1	SL4.0ch01:85661129..85 662918 (+ strand)	LECEVI16G peroxidase precursor
	Chlorophyll content (ChlC) chr7 10.8cM; 1,557,551bp; solcap_snp_sl_68398			
	t50 ETR Induction chr7 23.2cM; 2,881,365 bp; solcap_snp_sl_15786			
	browse search	Solyc07g00 6140.2	SL2.50ch07:984510..988 395 (- strand)	Cytochrome P450
	ChlC chr7 10.8cM	solcap_snp_ sl_68398	1,557,551bp	
	browse search	Solyc07g00 6890.1.1	SL4.0ch07:1718992..172 0500 (+ strand)	Cytochrome P450 (AHRD V3.3 *** A0A200Q4B5_9MAGN)
	browse search	Solyc07g00 7460.1	SL2.50ch07:2165949..21 67526 (+ strand)	Cytochrome P450
	browse search	Solyc07g00 7650.1	SL2.50ch07:2303928..23 05559 (+ strand)	Translocase of chloroplast 90, chloroplastic (AHRD V1 ***- TOC90_ARATH)
	browse search	Solyc07g00 8110.2	SL2.50ch07:2831138..28 32794 (- strand)	Blue copper protein (Fragment) (AHRD V1 *--- O82576_MAIZE); contains Interpro domain(s) IPR003245 Plastocyanin-like
	browse search	Solyc07g00 8120.2	SL2.50ch07:2836185..28 37343 (+ strand)	Blue copper protein (Fragment) (AHRD V1 *--- O82576_MAIZE); contains Interpro domain(s) IPR003245 Plastocyanin-like
	browse search	Solyc07g00 8130.2	SL2.50ch07:2838443..28 40766 (- strand)	Blue copper protein (Fragment) (AHRD V1 *--- O82576_MAIZE); contains Interpro domain(s) IPR003245 Plastocyanin-like
	browse search	Solyc07g00 8140.2	SL2.50ch07:2848064..28 49548 (- strand)	Blue copper protein (Fragment) (AHRD V1 *--- O82576_MAIZE); contains Interpro domain(s) IPR003245 Plastocyanin-like
	t50 ETR Induction chr7 23.2cM	solcap_snp_ sl_15786	2,881,365 bp	
	browse search	Solyc07g00 8420.2	SL2.50ch07:3213591..32 15016 (- strand)	Blue copper-like protein (AHRD V1 **-- A3F8V0_GOSHI); contains Interpro domain(s) IPR008972 Cupredoxin

browse search	Solyc07g008990.1	SL2.50ch07:4007484..4008490 (+ strand)	Photosystem II D2 protein (AHRD V1 ***- D3WCL6_PHOSE); contains Interpro domain(s) IPR000484 Photosynthetic reaction centre, L/M
browse search	Solyc07g009000.1	SL2.50ch07:4008533..4008962 (+ strand)	Photosystem II CP43 chlorophyll apoprotein (AHRD V1 ***- C8XU03_COILA); contains Interpro domain(s) IPR000932 Photosystem antenna protein-like
Chlorophyll content (ChlC) chr7 53cM; 62,093,144 bp; solcap_snp_sl_53425			
literature: Ding et al., 2016 (Solyc05g02600)	Solyc07g049720		10kDa polypeptide precursor of photosystem 2
literature → browse	Solyc07g049720.2	SL2.50ch07:60062710..60072674 (- strand)	DNA topoisomerase I (AHRD V1 **** Q9XGL1_DAUCA); contains Interpro domain(s) IPR013499 DNA topoisomerase I, C-terminal, eukaryotic-type
browse search	Solyc07g052070.1	SL2.50ch07:60623496..60627130 (+ strand)	Cytochrome P450
browse search	Solyc07g052370.2	SL2.50ch07:60883858..60885735 (+ strand)	Cytochrome P450
browse search	Solyc07g052660.1	SL2.50ch07:61114478..61115399 (+ strand)	Blue copper protein (AHRD V1 *--- B6U7P6_MAIZE); contains Interpro domain(s) IPR003245 Plastocyanin-like
ChlC chr7 53cM	solcap_snp_sl_53425	SL2.50ch07:62093144..62093144 (+ strand)	
browse search	Solyc07g054210.2	SL2.50ch07:62565942..62569145 (- strand)	Protochlorophyllide reductase like protein (AHRD V1 ***- Q0WVW0_ARATH); contains Interpro domain(s) IPR005979 Light-dependent protochlorophyllide reductase
browse search	Solyc07g054290.1	SL2.50ch07:62640921..62641514 (- strand)	Photosystem II family protein (AHRD V1 *-*- D7KCG4_ARALY)
browse search	Solyc07g055350.2	SL2.50ch07:63412672..63415080 (- strand)	Cytochrome P450
browse search	Solyc07g055450.1	SL2.50ch07:63477422..63480076 (- strand)	Cytochrome P450 (AHRD V1 ***- B9NEQ9_POPTR); contains Interpro domain(s) IPR002401 Cytochrome P450, E-class, group I
browse search	Solyc07g055440.1	SL2.50ch07:63472670..63474986 (- strand)	Cytochrome P450
browse search	Solyc07g055460.2	SL2.50ch07:63482507..63484648 (- strand)	Cytochrome P450
browse search	Solyc07g055470.2	SL2.50ch07:63485489..63488308 (- strand)	Cytochrome P450
browse search	Solyc07g055480.1	SL2.50ch07:63494989..63497156 (+ strand)	Cytochrome P450

browse search	Solyc07g055490.2	SL2.50ch07:63504319..63506645 (- strand)	Cytochrome P450
browse search	Solyc07g055530.2	SL2.50ch07:63523522..63527175 (- strand)	Cytochrome P450
browse search	Solyc07g055540.2	SL2.50ch07:63528600..63532474 (- strand)	Cytochrome P450
browse search	Solyc07g055550.1	SL2.50ch07:63533657..63537235 (- strand)	Cytochrome P450
browse search	Solyc07g055970.1	SL2.50ch07:63877377..63879048 (+ strand)	Cytochrome P450
browse search	Solyc07g056160.2	SL2.50ch07:64018080..64022506 (+ strand)	Cytochrome P450
Chlorophyll content (ChlC) chr8 36cM, 56253744bp, solcap_snp_sl_18195			
ChlC chr8 36cM	solcap_snp_sl_18195	56253744 bp	
browse search	Solyc08g067320.1	SL2.50ch08:56387815..56388090 (- strand)	Chlorophyll a/b binding protein (AHRD V1 ***- Q41422_SOLTU); contains Interpro domain(s) IPR001344 Chlorophyll A-B binding protein
browse search	Solyc08g067330.1	SL2.50ch08:56388147..56388443 (- strand)	Chlorophyll a-b binding protein 3C-like (AHRD V1 ***- Q2XTE0_SOLTU); contains Interpro domain(s) IPR001344 Chlorophyll A-B binding protein
browse search	Solyc08g083450.1	SL2.50ch08:56813657..56815352 (+ strand)	Cytochrome P450
browse search	Solyc08g067840.2	SL2.50ch08:56873120..56881232 (+ strand)	PsbP domain-containing protein 5, chloroplastic (AHRD V1 **-- PPD5_ARATH); contains Interpro domain(s) IPR016123 Mog1/PsbP, alpha/beta/alpha sandwich
Chlorophyll content (ChlC) chr10 14.5cM; 2,335,513 bp; solcap_snp_sl_17859			
browse search	Solyc10g006230.2	SL2.50ch10:920027..923670 (+ strand)	Chlorophyll a-b binding protein 7, chloroplastic (AHRD V1 ***- CB12_SOLLC); contains Interpro domain(s) IPR001344 Chlorophyll A-B binding protein
browse search	Solyc10g006900.2	SL2.50ch10:1335152..1338824 (+ strand)	Protochlorophyllide reductase (AHRD V1 ***- Q8LAV9_ARATH); contains Interpro domain(s) IPR005979 Light-dependent protochlorophyllide reductase
browse search	Solyc10g007690.2	SL2.50ch10:1950571..1952635 (+ strand)	Chlorophyll a-b binding protein 8, chloroplastic (AHRD V1 ***- CB13_SOLLC); contains Interpro domain(s) IPR001344 Chlorophyll A-B binding protein
browse search	Solyc10g007860.2	SL2.50ch10:2067866..2070928 (+ strand)	Cytochrome P450 ent-kaurenoic acid oxidase (AHRD V1 **** B9IB42_POPTR); contains Interpro domain(s) IPR002401 Cytochrome P450, E-class, group I
browse search	Solyc10g007890.2	SL2.50ch10:2086394..2090624 (+ strand)	Cytochrome P450 (AHRD V1 ***- A9ZT56_COPJA); contains Interpro domain(s) IPR002401 Cytochrome P450, E-class, group I

browse search	Solyc10g007880.2	SL2.50ch10:2080596..2083797 (+ strand)	Cytochrome P450
browse search	Solyc10g007900.2	SL2.50ch10:2091468..2095251 (- strand)	Cytochrome P450
browse search	Solyc10g007930.2	SL2.50ch10:2099361..2102837 (- strand)	Cytochrome P450
ChlC chr10 14.5cM	solcap_snp_sl_17859	2,335,513 bp	
browse search	Solyc10g008390.2	SL2.50ch10:2524682..2529183 (+ strand)	Thioredoxin (AHRD V1 ***- A7LNX7_9CARY); contains Interpro domain(s) IPR005746 Thioredoxin
browse search	Solyc10g009310.2	SL2.50ch10:3349934..3352637 (+ strand)	Cytochrome P450
browse search	Solyc10g009390.2	SL2.50ch10:3453235..3456027 (+ strand)	Cytochrome P450

Chapter 6

General discussion

6.1. Usefulness of yield dissection models combined with quantitative genetics

In this thesis, I demonstrated how the combination of yield dissection models and component traits with quantitative genetics (QTLs and prediction) provides benefits for plant breeding to improve yield. Several studies combined crop models with quantitative genetics, however, there was no clear methodology on how to find an influential component trait to represent the variation in yield. In addition, to my knowledge, the number of studies of this methodology for plant breeding to improve yield is still limited. To fill this knowledge gap, I used a recombinant inbred line (RIL) population of tomato derived from multiple parents to assess (1) which physiological traits share their genetic basis with fruit yield, and (2) via which component traits, yield of the elite parent can be improved. I discuss how QTLs for component traits could contribute to the selection for high-yielding genotypes, where positive alleles for multiple traits are aggregated.

6.1.1. Contribution of wild germplasms to improve yield

The use of introgression of wild germplasms in my study generated a large genetic variation in both populations for Dutch and Spanish production systems. Because wild species show a low yield, it is expected that positive allele effects mainly arise from the elite parents. Still, I found that wild alleles also provided positive effects in some QTLs. Thus, wild germplasms possess beneficial genes to improve yield of the elite genotypes.

The positive effects of wild germplasms were especially large in the Spanish population, for which around half of RILs showed a higher yield than the elite parent. Thus, around half of positive allele effects arose from the wild parents. Therefore, the Spanish population has substantial room to improve the yield of the elite recurrent parent by introgression of wild germplasm.

Gur and Zamir (2004) reported on the contribution of wild germplasm to improve elite materials or cultivars. It is widely accepted that natural genetic diversity from wild germplasms is necessary to improve yield and other traits in plant breeding (Tanksley and McCouch, 1997; Huang and Han, 2014). However, it is still challenging to evaluate the contribution of wild germplasm for complex traits, such as yield (Singh *et al.*, 2021). In this thesis, I demonstrated that the evaluation of the contribution from wild parents is possible by using identity-by-descent probabilities (IBD) as genetic information in QTL analysis for a multi-parent population, for which the normal genetic marker (identity by state: IBS) cannot show the parental origin.

6.1.2. Which traits underlie the variation in yield?

The introgressions of wild parents caused genetic and phenotypic variation of offspring (RIL population) in both Dutch and Spanish breeding populations. Variation in yield must have

been caused by variation in one or more component traits. Phenotypic variation in yield was explained by phenotypic variation in many traits — except for harvest index in Dutch and Spanish (both greenhouses) experiments, photosynthesis-related traits in Dutch experiment, the number of fruits per truss (FNtruss) in both greenhouses in Spain, and fruit fresh dry weight ratio (Fresh_Dry_ratio) in the Fer greenhouse in Spain. The component traits that explain the variation in yield should be included in a crop model to predict yield if a breeder wants to predict yield by combining a crop model with QTLs.

6.1.3. What is the phenotypic ideotype for a high yield?

Although it is important to determine which component traits explain the variation in yield, if all offspring (RILs) show a lower yield than the elite recurrent parent, the population will not be useful to breed a high-yielding line. For plant breeding, it is crucial to generate a high-yielding line compared to the elite recurrent parent and reveal the characteristics of component traits of the high-yielding line, that is, how the phenotypic ideotype looks like.

The high-yielding lines (RILs) can be developed by different compositions of component traits (so-called "equifinality") (Lamsal *et al.*, 2018). For example, the same high yield can be achieved by increasing the number of fruits (FN) or by increasing individual fruit fresh weight (FFW). An important consideration here is the relationship between component traits. If two components can independently improve yield, a breeder can increase yield by simply increasing one or both of these component traits. However, if there is a trade-off between component traits, an increment in a component trait may cause a reduction in another component trait, not resulting in yield improvement. Therefore, it is important to look at these component traits simultaneously. Yield dissection models, for which the upper trait is the product of two underlying component traits, are a useful tool to consider these relationships between component traits, as demonstrated in this thesis (Chapter 2, 3, and 4).

Interestingly, the number of fruits (FN) and individual fruit fresh weight (FFW) showed a trade-off in both Dutch and Spanish breeding populations. Despite this trade-off, some RILs showed a higher yield than the elite recurrent parent. More high-yielding lines arose from a higher FN rather than from a higher FFW compared to the elite recurrent parent. This is an interesting observation because it was reported that an increase in FFW was the driving force to improve yield during tomato domestication (Lin *et al.*, 2014). My results showed that further yield improvement can better be achieved through FN than FFW by introducing wild germplasm.

The relationship between total fruit dry weight (FDWtotal) and fruit fresh-dry weight ratio (Fresh_Dry_ratio) was not consistent between Dutch and Spanish experiments. Nevertheless, more high-yielding lines arose from higher FDWtotal in both Dutch and Spanish experiments. This is useful information for breeding companies. High-yielding fresh tomatoes were used to be criticized as a "Water bomb" (Hendriks, 2016; Schouten *et al.*, 2019), which means that

a high-yielding fresh tomato is watery and less tasty. Fortunately, my results indicate that breeders can increase yield by increasing FDWtotal regardless of Fresh_Dry_ratio, that is, there is a way to increase yield without sacrificing taste.

The relationship between harvest index (HI) and total biomass was not consistent among experiments. Nevertheless, more high-FDWtotal lines arose from higher total biomass. HI contributed to yield improvement in the past (Evans, 1993; Hay, 1995). However, in our population, HI is less important to breed a high-yielding line.

Furthermore, in the Dutch experiment, I consider how to breed a high-total-biomass line. Around half of RILs showed higher total biomass than the elite parents, which arose from a higher adapted light use efficiency (LUE_{adp}) and/or a higher fraction of intercepted light (F_{int}). Among high-total-biomass lines, only 10 lines showed a higher yield than the elite parents. Most high-yielding lines possessed a higher LUE_{adp} and a higher F_{int} than the elite parents; three high-yielding lines possessed a higher LUE_{adp} but a lower F_{int} than the elite parents. Because of no trade-off between LUE_{adp} and F_{int} (Chapter 4), both traits can be improved simultaneously to improve total biomass. However, due to the negative correlation between F_{int} and harvest index, LUE_{adp} is more straightforward to produce a high-yielding line.

LUE_{adp} is thought to be primarily determined by photosynthesis. However, leaf photosynthesis-related traits neither correlated with LUE_{adp}, total biomass, nor yield. Thus, the phenotypic ideotype of photosynthesis-related traits for a high yield was not identified in this thesis. The absence of correlation between leaf photosynthesis and yield was reported by multiple studies and caused doubt about whether photosynthesis is an important trait to improve yield (Evans, 1993; Richards, 2000; Long et al., 2006; Gu et al., 2014a).

6.1.4. Design of genotypic ideotype from yield QTLs

Phenotypic variations in the RIL population were caused by genotypic variation in the RIL population and environmental effects. Statistical tests of association between phenotype and genotype enabled to detect QTLs for yield and component traits. In this thesis, many yield QTLs were detected. By aggregating positive QTL alleles for yield, the genotypic ideotype for a high yield can be developed.

As expected, in the Dutch breeding population, overall the elite parents had positive, and the wild parents had negative allelic effects. However, at each QTL, wild parents also possessed some positive alleles for some QTLs. Surprisingly, in the Spanish breeding population, wild parents possessed a similar amount of positive allelic effects as the elite parent. Therefore, it is possible to combine positive QTL alleles for yield from both elite and wild parents to produce high-yielding lines.

6.1.5. Design of genotypic ideotype from component-traits QTLs

If yield QTLs are detected, it is possible to consider the genotypic ideotype for high yield from yield QTLs. However, yield QTLs are not always detected (Foolad, 2007; Yamamoto *et al.*, 2016). In such a case, the genotypic ideotype for high yield can be designed by component-traits QTLs. However, even if yield QTLs are detected as in this thesis, it is still very useful to consider component-trait QTLs. First, more QTLs would be available to aggregate positive alleles towards high yield. I detected more component-traits QTLs in total, compared to yield QTLs. This was more evident in the Spanish experiment than in the Dutch experiment (Chapter 3). This might be because the environment was less stable in the Spanish experiment, which made QTL detection difficult for complex traits, such as yield. Second, component-trait QTLs provide underlying reasons for those yield QTLs that colocalize with component-traits QTLs. Then, breeders can design genotypic ideotypes to develop preferred phenotypic ideotypes. For example, if a breeder wants to develop a high-yielding line without losing good taste, the breeder should combine positive alleles for high yield, however, without touching QTLs for fruit fresh-dry weight ratio. Third, if we consider component-trait QTLs, we can take into account the trait-by-trait relationship: trade-off, synergy, or no correlation. The trade-off or synergy might be caused by (1) Physiological connection between component traits (physiological pleiotropy) (Boote *et al.*, 2021), (2) Genetic pleiotropy, or (3) Genetic linkage. Physiological connections between traits can be considered by carbon or nitrogen balance, which is implemented in crop models (Boote *et al.*, 2021). For example, the trade-off between the number of fruits and individual fruit fresh weight can be considered as resulting from carbon availability. Genetic pleiotropy can occur if the same gene has an effect on both traits. Genetic linkage can occur if two genes for different traits are located close to each other, and therefore recombination hardly occurs between these two genes (Schulthess *et al.*, 2017). For all cases of physiological pleiotropy, genetic pleiotropy and genetic linkages, QTLs might be detected and colocalized between these traits. If a component trait did not show a trade-off with other traits, it is simple to improve yield by using QTLs for this component trait (Sadras and Richards, 2014). On the other hand, if two traits show a trade-off, positive alleles from these two component traits should be carefully selected (Reynolds *et al.*, 2012). In my results, the number of fruits (FN) and individual fruit fresh weight (FFW) showed a trade-off (Chapter 2 and 3). Several QTLs were at the same position between FN and FFW, but most of them showed different parental effects between FN and FFW. Thus, positive alleles for these traits should be carefully selected to improve yield. After consideration of all of these, finally, breeders can design a genotypic ideotype to achieve a preferred phenotypic ideotype. A high yield can be achieved in different ways. For example, by increasing fruit fresh-dry weight ratio (including more water into fruits); or by increasing light use efficiency to increase total biomass. Designing phenotypic and associated genotypic ideotypes cannot be achieved by considering yield alone, but can be achieved by considering component-trait QTLs, which are connected to yield via yield dissection models. Breeders can consider a QTL effect of a component trait on yield.

6.2. Which trait should a breeder look at to breed a high-yielding line?

6.2.1. Sink strength vs source strength

In this thesis, I dissected yield from two perspectives. One is from the sink-organ capacity ("Harvest dissection"). In the harvest dissection model, yield is the product of the number of fruits and individual fruit fresh weight. Such a dissection is often used for cereal yield by crop physiologists and agronomists (Yin and Struik, 2008). Sink-organ capacity is not identical to sink strength, which can be defined as the potential demand or potential capacity of an organ for assimilate accumulation and can be quantified by the potential growth rate of the organ (Marcelis, 1996). Because it is not possible to measure fruit-sink strength (potential capacity of fruit-sink organs) for more than 700 plots, I analyzed fruit-sink capacity (weight) instead of fruit-sink strength. For yield prediction studies, only a few authors employed this method (Yin *et al.*, 2002; Millet *et al.*, 2019). An alternative perspective looks at biomass ("Biomass dissection"). The biomass dissection model in this thesis is the simplified version of crop growth models. Crop growth models formulate yield from underlying physiological processes and are commonly photosynthesis-driven, focusing on biomass production and its allocation (Marcelis *et al.*, 1998). Most yield prediction studies, which combine crop models with quantitative genetics (QTLs or genomic prediction) used crop growth models, therefore focus on biomass production (source strength). The question is which model is more suitable — "harvest dissection" or "biomass dissection" — for plant breeding to improve yield.

One important consideration is "which trait determines yield?" In other words, "what is the limiting factor to improve yield?" For example, if the maximum photosynthetic rate (P_{max}) is the limiting factor, then it is necessary to improve P_{max} to improve yield. However, if P_{max} has enough capacity but fruit-sink capacity is limiting yield, then improving P_{max} does not lead to a higher yield. Crop modellers and photosynthesis researchers often focus on source strength. In crop growth models, yield is mainly determined by source availability (Yin *et al.*, 2000*b*). Photosynthesis researchers often claimed that light use efficiency (LUE) is the only remaining factor to improve yield because harvest index and light interception are close to their theoretical maxima, and photosynthesis is the main determinant of LUE (Zhu *et al.*, 2010; Foyer *et al.*, 2017). For indeterminate tomatoes grown in the greenhouse, source limitation was shown by pruning experiments and carbohydrate analysis (Cockshull and Ho, 1995*b*; Iglesias *et al.*, 2002; Li *et al.*, 2015). On the other hand, some studies argued that yield is more sink than source limited during seed filling in major field crops like wheat, maize and soybean (Borrás *et al.*, 2004; FISCHER, 2007; Smith *et al.*, 2018). A perennial debate about whether plant growth is limited by source or sink suggests that both are important and need to be considered (Evans, 2013). Importantly, sink or source limitation is not fixed. For example, a sink-limited crop in most growing conditions can become a source-limited crop if resource availability is strongly reduced during seed filling (Borrás *et al.*, 2004). For

another example, free-air CO_2 enrichment (FACE) experiments showed that enhanced photosynthetic rate under elevated CO_2 concentration could lead to higher productivity only if sink size is large. Thus, the improvement of source strength can cause the sink limitation (Hasegawa *et al.*, 2013).

Another consideration is that genetic determinism of "source strength" and "sink strength" could be partly common (Welcker *et al.*, 2007). In maize, leaf growth and anthesis-silking interval — the main determinants of source and sink strengths, respectively — shared some QTLs under drought stress (Welcker *et al.*, 2007). In this thesis (Chapter 2 and 3), several QTLs were common between component traits in harvest dissection and component traits in biomass dissection. Thus, the same gene could affect both "source strength" and "fruit-sink capacity", making it difficult to disentangle which is the limiting factor for yield. Because both sink strength and source strength are important for yield, and sink-limited or source-limited is changeable depending on species, population, developmental stage, and growing conditions, I recommend breeders to always use both yield dissection models. By using both yield dissection models, plant breeders know which direction they should choose to improve yield of the elite parent: by increasing fruit-sink capacity or by increasing source strength.

6.2.2. On which traits did other studies focus to breed for high yield?

Domestication studies focus on the harvestable-sink capacity

Crops were domesticated from wild ancestors. Domestication studies (Lin *et al.*, 2014; Stitzer and Ross-Ibarra, 2018) compared modern crops with their ancestral wild species to reveal which trait contributes to the transformation from the wild ancestor to the domesticated crop species. In these studies, morphology, especially the harvestable-sink capacity is the main focus. For example, *Zea mays* (maize) is domesticated from teosinte (specifically *Z. mays* subsp. *Parviglumi*). Because vegetative morphology of teosinte is largely indistinguishable from maize, ear phenotypes were the major focus to define the difference between teosinte and maize (Stitzer and Ross-Ibarra, 2018). In tomato, it was proposed that a two-step evolution (domestication followed by improvement by breeding) resulted in a fruit size more than 100 times larger than the wild ancestor of tomato, *Solanum pimpinellifolium* (Lin *et al.*, 2014). Domestication studies focus on harvestable-sink capacity, therefore, it is not clear how traits related to source strength were altered during domestication.

Early breeding

Increases in yield potential over the last century in cereals have been achieved through an increase in *harvest index* (Hay, 1995), additional nitrogen fertilizers and an increase in light interception (Long et al., 2006). Although the harvest index is famous for its contribution to the green revolution (Miflin, 2000; Hedden, 2003), so far, no study focused on the harvest index to predict yield by combining crop models with genetics.

Recent studies combining crop models with quantitative genetics

Phenology, especially *flowering time*, is the most popular trait to be predicted by QTLs or genes (White and Hoogenboom, 1996; Hoogenboom *et al.*, 1997, 2004; Nakagawa *et al.*, 2005; Yin *et al.*, 2005*a*,*b*; Messina *et al.*, 2006; Uptmoor *et al.*, 2008; Zheng *et al.*, 2013; Bogard *et al.*, 2014; Technow *et al.*, 2015; Bhakta *et al.*, 2017; Hwang *et al.*, 2017; Parent *et al.*, 2018; Lamsal *et al.*, 2018; Toda *et al.*, 2020; Oliveira *et al.*, 2021). Some of these studies connected the genetic effects on phenology to total biomass (Toda *et al.*, 2020), or yield (White and Hoogenboom, 1996; Hoogenboom *et al.*, 1997; Messina *et al.*, 2006; Parent *et al.*, 2018; Oliveira *et al.*, 2021; Bustos-Korts *et al.*, 2021). Phenology is relatively simple to apply the methodology of combining crop models with quantitative genetics because phenology is controlled by a relatively low number of genes (Boote *et al.*, 2021). Phenology is important to fit cultivars to their target environment or target growing season, and as Messina *et al.* (2006) showed, phenology QTL effects can account for up to 54% of yield variation among cultivars (Boote *et al.*, 2021).

Light use efficiency (LUE) is often thought to be the last remaining factor to improve yield and *photosynthesis* to be the main determinant of LUE (Zhu *et al.*, 2010; Foyer *et al.*, 2017). However, the application of LUE and photosynthesis to improve yield by combining crop models with genetics is limited. A study predicted yield by using five traits, one of which is LUE (Technow *et al.*, 2015). Another study simulated the effects of leaf-photosynthetic-rate QTLs on total biomass (Gu *et al.*, 2014*a*).

Light interception was used to predict yield in some studies by combining crop models with genetics. Two studies predicted yield via five component traits, among which total leaf number and the area of the largest leaf were related to the light interception (Technow *et al.*, 2015; Cooper *et al.*, 2016). In studies on maize under drought stress (Reymond *et al.*, 2003, 2004; Chenu *et al.*, 2008, 2009), leaf area expansion was an important factor to predict yield. Leaf elongation rate was predicted by genetic effects of sensitivity to temperature, vapor pressure deficit, and predawn leaf water potential (Reymond *et al.*, 2003, 2004). The predicted leaf elongation rate was used to predict yield (Chenu *et al.*, 2008, 2009).

Only a few studies used *harvestable-sink organ capacity* to predict yield by combining crop models with genetics (Yin *et al.*, 2002; Millet *et al.*, 2019). This is probably because most crop models simulated yield from source availability (Yin *et al.*, 2000*b*). One of these studies considered phenology for the prediction of grain number, which was subsequently used for yield prediction. That is, the grain number was further dissected via a factorial regression, which includes three genotype-dependent sensitivities to environmental indices. These environmental indices were calculated for certain developmental stages, respectively. Therefore, for this model, physiology related to source strength was also taken into account (Millet *et al.*, 2019).

Many studies focus on some influential traits on yield. Some of these traits might be less influential for indeterminate tomato production in the greenhouse. For example, flowering time is important for determinate field crops to find a best-fit genotype to a target environment among existing genotypes. However, for indeterminate tomatoes, flowering time is hardly of importance for yield because a plant continually produces new trusses with flowers. Or another example, leaf elongation rate is important under water deficit conditions. However, for greenhouse tomatoes, irrigation is well controlled. Therefore, in this thesis, I focused on increasing potential yield, for which water and nutrient availability are not limiting, and the environment is relatively well controlled.

6.2.3. A trait important for management is not necessarily important for breeding

What impressed me is that some traits usually important for horticultural studies on a genotype — due to high sensitivity to a management or environmental factor — were not important to breed a high-yielding line in this thesis. That is, the harvest index was reported to be the key trait to improving yield in a study applying far-red light (Ji *et al.*, 2020). Another example is specific leaf area, which becomes larger under a low red/far-red light ratio, similar to the light condition under the canopy (Shibuya *et al.*, 2011). These traits showed large variation but did not correlate with yield in this thesis. My results indicate that we cannot conclude the importance of a trait for high-yielding breeding based on a management experiment on a genotype. Even if a component trait shows G×E, this trait might not contribute to the variation in yield within a breeding population and therefore might not be useful to breed a high-yielding line.

6.3. Implication for high-throughput phenotyping

High-throughput phenotyping (HTP) (in situ or platform) is beneficial for breeding if a trait can be measured more precisely, high-throughput (faster, easier, cheaply), or earlier. If the HTP platform can keep a stable environment, the heritability of a trait might be improved. If more genotypes are phenotyped, a higher number of QTLs can be detected (higher statistical power). If a trait can be measured earlier than yield, the breeding cycle can be shortened (van Eeuwijk *et al.*, 2019).

If HTP can be applied to measure yield, it would be very useful. However, tomato yield cannot be determined by HTP until now. Instead, HTP might be able to measure some component traits. In which cases is HTP useful for tomato breeding to improve yield? In this thesis, I examined the number of fruits, individual fruit fresh weight, fruit fresh-dry weight ratio, total fruit dry weight, harvest index, total biomass, adapted light use efficiency, fraction of intercepted light, leaf area index, number of leaves, leaf width, leaf area, leaf dry weight, specific leaf area, chlorophyll content index (SPAD), and chlorophyll fluorescence. Among

them, chlorophyll content index (SPAD) and chlorophyll fluorescence were reported to be measured in HTP (Flood *et al.*, 2016; Rutkoski *et al.*, 2016; Montesinos-López *et al.*, 2017; McAusland *et al.*, 2019). Some studies reported measurements of leaf-related traits in HTP in maize (Cabrera-Bosquet *et al.*, 2016; Lacube *et al.*, 2017; Zhang *et al.*, 2017). However, it remains challenging to measure leaf-related traits of indeterminate tomatoes because the canopy is about 2 m high, a plant is guided by a rope, and the leaves are curving. Total biomass can be measured in HTP (for example, in rice (Tanger *et al.*, 2017)). If total biomass shows a high correlation between young and mature plants (as is the case for wheat (Rebetzke and Richards, 1999)), it is possible to shorten the breeding cycle by measurement of total biomass in young plants (van Eeuwijk *et al.*, 2019). One caution is that total biomass is often estimated by volume in HTP (Junker *et al.*, 2015; Muraya *et al.*, 2017). It is necessary to check whether there is a strict correlation between total biomass and total volume, that is, a genotype of higher total biomass shows a higher total volume, meaning that fresh-dry weight ratio is the same for all genotypes. But at the same time, total volume may not necessarily be a proxy to the total biomass, but total volume itself might be useful to predict yield.

It is argued that a component trait can be a proxy for yield selection only when it has a high correlation with yield (van Eeuwijk *et al.*, 2019). In this thesis, only total fruit dry weight in NL (0.82), total biomass in NL (0.65), and adapted light use efficiency in NL (0.64) showed a correlation larger than 0.6 with yield. Thus, only total fruit dry weight in NL showed a very high correlation with yield and can be a proxy for yield selection; however, this trait is difficult to measure.

If a component trait can increase the accuracy of yield prediction when it is used in a multi-trait yield prediction model together with yield (as a secondary trait), this component trait is useful for high-yielding breeding. I did not try this type of yield prediction model in this thesis. However, QTLs for photosynthesis-related traits predicted yield with a high prediction accuracy even without QTLs for yield itself. Photosynthesis-related traits can be measured in HTP platforms, where the environmental conditions are controlled and a large number of genotypes can be measured quickly. Thus, at the HTP platform, more photosynthesis-related QTLs can be detected. Due to no correlation between yield and photosynthesis, yield still need to be phenotyped if we want to use QTLs for photosynthesis-related traits to breed high-yielding lines. Otherwise, it is necessary to unravel photosynthesis in more detail to improve yield. For example, if we know that the antenna size of chlorophyll should be smaller but some enzymes need to be more enriched to improve yield, phenotyping of photosynthesis-related traits in HTP could be immediately useful for breeding to improve yield.

6.4. Model structure

Yield can be predicted by several approaches:

(1) By yield QTLs
(2) By component-trait QTLs (+ yield QTLs)
(3) By a statistical model or AI, where all or some of the component traits are predicted by own QTLs
(4) By a yield dissection model, where all or some of the component traits are predicted by own QTLs
(5) By a crop growth model, where all or some of the component traits are predicted by own QTLs

In this thesis, I tried approach (1), (2) (Yield = f(photosynthesis-related-traits QTLs)), and (4). Approaches (1) and (2) are similar to the genomic prediction, but the number of genetic markers is limited as only single-QTL markers were included in the model. Approaches (3) (4), and (5) formulate yield from component traits. Approach (3)(statistical model or AI) does not consider, but approach (4)(yield dissection model) and (5)(crop growth model) consider the underlying physiological relationship among component traits. Among approaches (3), (4), and (5), (5)(crop growth model) is the most popular for yield prediction with genetics (Chenu *et al.*, 2009; Cooper *et al.*, 2016; Bustos-Korts *et al.*, 2021).

Compared to approach (5)(crop growth model), the advantage of approach (4)(yield dissection model) is that uncertainty does not arise from model architecture and associated input coefficients. For approach (5)(crop growth models), the model does not necessarily show 100% precision even when it predicts yield from observed component traits. Thus, it is not clear whether the difference between observation and prediction of yield arises from the poor prediction of component traits or from model architecture with associated input coefficients, for which genetic variation is not included (Yin *et al.*, 1999*b*,*a*, 2000*b*; Hammer *et al.*, 2002). Therefore, when using a crop growth model to predict yield from QTLs for component traits, first, target component traits should be defined. Then, it is necessary to upgrade or re-develop the crop growth model to ensure that the crop growth model predicts yield precisely from the targeted component traits.

6.5. Methodology of QTL analysis in tomato population

6.5.1. I detected yield QTLs

Yield QTLs are often claimed for their difficulty or instability of detection. In this thesis, I detected many yield QTLs, especially for the Dutch experiment. Why could I detect many yield QTLs? This might be because our population was suitable for QTL detection. Our population showed a large genetic variation from multiple parents including wild parents. In addition, our breeding population has a relatively uniform background (around 50% from the elite parent 2 in NL, around 75% from the elite parent in Spain), where the effects of introduced alleles are examined in the genetic background of the recurrent elite parent

(Tanksley and Nelson, 1996; Tanksley and McCouch, 1997). Furthermore, I conducted phenotyping on hybrids. If phenotyping would be conducted on RIL itself, novel allelic effects could be masked by major deleterious alleles and the absence of important agronomic genes (Voss-Fels *et al.*, 2019). To test the wild allelic effects in the modern cultivar's background, it is proposed to cross the wild "donor" accession to an elite tester to produce F1 hybrids, where the dominant action of many agronomically important genes are active (Longin and Reif, 2014). Similarly, I tested on hybrids, which might have increased the possibility to detect novel allelic effects from wild parents.

Another reason for success for yield-QTLs detection could be that I experimented in the greenhouse, where the environmental conditions were relatively stable, especially in the Netherlands. This might have provided similar advantages to high-throughput phenotyping platforms. A stable environment reduces the environmental variance and increases the power of QTL detection.

6.5.2. Correlations between markers

One caution is that there are correlations between markers because the tomato genome duplicated during its evolution (Tomato Genome Consortium *et al.*, 2012). Due to these correlations between markers, if I use composite interval mapping (CIM: a marker with the highest significance is fixed in the multi-QTL model in forward selection) or include the kinship matrix in the QTL model, the number of QTL peaks would reduce. One might say that multi-QTL analysis in this thesis includes some false-positive QTLs. However, CIM or including kinship matrix might lose some real QTLs, which correlate with other QTLs. CIM employs forward selection, whereas my method is similar to backward elimination (single-QTL markers with zero variance in the multi-QTL model were removed from the final multi-QTL model). Forward selection and backward elimination are two different methods, and both are useful. In addition, our method can be considered as a similar method to genomic prediction, but markers were pre-selected by a single-QTL scan.

6.6. Recommendations for future research

6.6.1. G×E

Although I confirmed that some component traits showed less G×E than yield itself (Chapter 3, Spanish experiments), one missing point of this thesis is that environmental effects were not included in the model. A typical way to evaluate the environmental effect is to use a crop growth model. Crop growth models formulate the environmental responses of morphological and physiological processes. Combining crop growth models with genetics can be used to assess crop performance in different environments and evaluate G×E (Parent *et al.*, 2018; Bustos-Korts *et al.*, 2021).

Yield dissection models, however, can also incorporate environmental effects. Millet et al. (2019) formulated yield as the product of grain weight and grain number. Grain number was further dissected via a factorial regression, where genotype-dependent sensitivities to environmental indices were incorporated.

In my project, each population was tested in one country only for one year. Therefore, it was not possible to incorporate environmental effects into my yield dissection models, while this would be possible if the same population is tested in other environments someday.

6.6.2. Dynamics (time course)

In this thesis, I focused on the final yield. Therefore, a dynamic change (time course) was not considered. The dynamic change can for example be evaluated by applying the yield dissection models for each season separately. However, if someone wants to know, for example, which genotype produces more tomato fruits in the very early season or which genotype shows a high yield in a specific environmental condition, a dynamic crop growth model with QTLs for environmental sensitivities would be more convenient. In addition, the dynamic crop growth model could be a bridge between plant breeding and production management. The same dynamic crop growth model can be used for breeding purposes and for decision support for farmers.

Even though I did not assess the dynamic change, the accuracies of yield predictions via yield dissection models were relatively high in this thesis. This might be because I studied indeterminate tomato. The indeterminate tomato produces a unit continually and fruits were harvested every week for a long production period. Thus, the situation is more static compared to determinate crops. In addition, tomatoes were grown in greenhouses, where the environment was controlled especially well in the Netherlands. Thus, the dynamics of the environment are less important in this thesis compared to the field condition.

6.6.3. Response to CO_2 concentration

One big difference between field and greenhouse is carbon dioxide concentration ($[CO_2]$). In Dutch greenhouses, farmers inject CO_2 into the greenhouses to keep high $[CO_2]$ to stimulate photosynthesis. Therefore, $[CO_2]$ is higher than outside during the daytime, except for the hot time when the ventilation is high for cooling (Rosenqvist, 2005; de Gelder *et al.*, 2012). On the other hand, in Spanish greenhouses, farmers do not input CO_2. During the daytime, plants conduct photosynthesis, which reduces $[CO_2]$ in the greenhouse. Ventilation capacity is often not enough to keep $[CO_2]$ at the same level as outside (Stanghellini *et al.*; Lorenzo *et al.*, 1990). $[CO_2]$ has a great impact on photosynthesis (Rosenqvist, 2000). Therefore, it is necessary to investigate whether there is an interaction between genotype and $[CO_2]$.

6.6.4. Root

In this thesis, I investigated shoots (above-ground plant parts) but not roots. It is expected that the preferred characteristics of the root is different between Dutch and Spanish production systems. In the Dutch system, stone wool is a common medium, where water and air are ideally distributed, and nutrition is very well controlled (Costa and Heuvelink, 2018). Thus, this is an easy environment for roots and plant. On the other hand, in the Spanish system, plants are transplanted on soil (Costa and Heuvelink, 2018). Therefore, nutrition control is limited. Our experiments were conducted in Almeria, where salinity level is high and plants might be under salt stress (Luis Caparrós-Martínez *et al.*, 2020). Therefore, favourable characteristics of roots must be different between Dutch and Spanish production systems. There is a possibility that some QTLs for high yielding might originate from root characteristics. Although some root QTLs might appear as QTLs for some traits above ground as reported in maize (Dignat *et al.*, 2013), it is worthwhile to investigate roots.

6.7. Concluding remarks

When I started this PhD project, I heard that it would be difficult to detect QTLs for yield; therefore, I would use QTLs for component traits to predict yield. Unexpectedly, I detected many QTLs for yield in the Dutch experiment. Detecting many QTLs for yield is not always the case, especially in field crops. Indeed, in the Spanish experiment, yield showed a low heritability, and I detected a smaller number of QTLs, especially for the Fer greenhouse. Unexpected many QTLs detection for yield, however, made me possible to compare yield QTLs with component-trait QTLs. Thus, in this thesis, I succeeded in showing which component-trait QTLs underly yield QTLs.

Because I detected yield QTLs, someone might think why investigate component-trait QTLs? In this thesis, I demonstrated that

(1) Some component traits show higher heritability, a higher number of detected QTLs, higher prediction accuracy, and less G×E (higher correlation between two greenhouses in Spain; higher colocalization of QTLs between two greenhouses in Spain) than yield. Therefore, these component-trait QTLs are more stable than yield and can be used for selection for a higher yield.
(2) If a plant breeder looks at yield alone, a high-yielding line might possess some unfavourable characteristics. For example, the fruit fresh-dry weight ratio showed less G×E in the Spanish experiment. In the Netherlands, the fruit fresh-dry weight ratio positively correlates with yield. Thus, selection for higher yield might end up in selection for higher fruit fresh-dry weight ratio (less tasty) because the fruit fresh-dry weight ratio is stable in the selection over the years (less G×E). Thus, a high-yielding selection might end up in less tasty tomatoes in the Netherlands.

Interestingly, this seems unlikely in the Spanish breeding population; yield did not correlate with fruit fresh-dry weight ratio, and selecting for a higher yield does not result in less tasty fruits.

(3) Most importantly, a plant breeder can know in which direction the yield of an elite parent can be improved. Yield dissection models show the relationship between traits: synergy, independence, and trade-off. In this thesis, a trade-off happened between the number of fruits (FN) and individual fruit fresh weight (FFW). In this case, positive alleles for these two traits should be carefully combined to produce a high-yielding line. On the other hand, for the total plant biomass production, adapted light use efficiency (LUE_{adp}) and the fraction of intercepted light (F_{int}) showed an independent relationship, and most QTLs for these two traits were separately located. Thus, it is relatively easy to combine positive alleles for these two traits to produce a high-biomass line.

Therefore, component-trait QTLs are useful for plant breeding to select a high-yielding line.

In this thesis, I also showed the connection between photosynthesis-related traits and yield. To produce a high-yielding line, LUE_{adp} is more powerful than F_{int} due to the negative correlation between F_{int} and harvest index. Nine LUE_{adp} QTLs were found, two of which colocalized with leaf-morphology QTLs. The rest of the 7 LUE_{adp} QTLs were at the same positions as QTLs for photosynthesis-related traits. Among them, only QTLs for J_{max}-related traits in chromosome 1 showed similar parental effects between photosynthesis-related traits and yield. Nevertheless, photosynthesis-related QTLs together succeeded in the prediction of yield with high prediction accuracy. Thus, this thesis showed the usefulness of photosynthesis-related QTLs for plant breeding to improve yield. But at the same time, it showed the difficulty of using photosynthesis-related traits for high-yielding breeding because the QTL parental effects are not always consistent between photosynthesis-related traits and yield.

To increase the rate of yield improvement, beneficial genes need to be introduced into cultivars from wild germplasm. As a consequence, yield variation occurs. This thesis was a journey to explore this genetic variation in yield. I demonstrated a systematic approach to finding genetic variation in which component traits produce genetic variation in yield, and to finding the superiorities of which component traits lead to a higher yielding line compared to the elite parent. By using the methodology of this thesis — combining yield dissection models with QTLs — a plant breeder can find in which direction yield can be improved: what is the phenotypic ideotype (line with higher yield than the elite parent with preferable characteristics) and how it can be achieved by the genotypic ideotype — the combination of positive alleles for component traits.

List of references

Acreche MM, Slafer GA. 2009. Grain weight, radiation interception and use efficiency as affected by sink-strength in Mediterranean wheats released from 1940 to 2005. Field Crops Research **110**, 98–105.

Ainsworth EA, Bush DR. 2011. Carbohydrate export from the leaf: A highly regulated process and target to enhance photosynthesis and productivity. Plant Physiology **155**, 64–69.

Ainsworth EA, Long SP. 2005. What have we learned from 15 years of free-air CO_2 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO_2. New Phytologist **165**, 351–372.

Albert E, Gricourt J, Bertin N, Bonnefoi J, Pateyron S, Tamby JP, Bitton F, Causse M. 2016. Genotype by watering regime interaction in cultivated tomato: lessons from linkage mapping and gene expression. Theoretical and Applied Genetics **129**, 395–418.

Albrecht E, Escobar M, Chetelat RT. 2010. Genetic diversity and population structure in the tomato-like nightshades *Solanum lycopersicoides* and *S. sitiens*. Annals of Botany **105**, 535–554.

Alimi NA, Bink MCAM, Dieleman JA, *et al.* 2013*a*. Genetic and QTL analyses of yield and a set of physiological traits in pepper. Euphytica **190**, 181–201.

Alimi NA, Bink MCAM, Dieleman JA, Magán JJ, Wubs AM, Palloix A, van Eeuwijk FA. 2013*b*. Multi-trait and multi-environment QTL analyses of yield and a set of physiological traits in pepper. Theoretical and Applied Genetics **126**, 1–29.

Araus JL, Cairns JE. 2014. Field high-throughput phenotyping: The new crop breeding frontier. Trends in Plant Science **19**, 52–61.

Araus JL, Kefauver SC, Zaman-Allah M, Olsen MS, Cairns JE. 2018. Translating high-throughput phenotyping into genetic gain. Trends in Plant Science **23**, 451–466.

Arrones A, Vilanova S, Plazas M, Mangino G, Pascual L, Díez MJ, Prohens J, Gramazio P. 2020. The dawn of the age of multi-parent MAGIC populations in plant breeding: novel powerful next-generation resources for genetic analysis and selection of recombinant elite material. Biology **9**, 229.

Austin RB. 1989. Genetic variation in photosynthesis. The Journal of Agricultural Science **112**, 287–294.

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

Baker NR. 2008. Chlorophyll fluorescence: a probe of photosynthesis in vivo. Annual review of plant biology **59**, 89–113.

van Bezouw RFHM, Keurentjes JJB, Harbinson J, Aarts MGM. 2019. Converging phenomics and genomics to study natural variation in plant photosynthetic efficiency. The Plant Journal **97**, 112–133.

Bhakta MS, Gezan SA, Clavijo Michelangeli JA, *et al.* 2017. A predictive model for time-to-flowering in the common bean based on QTL and environmental variables. G3: Genes, Genomes, Genetics **7**, 3901–3912.

Bineau E, Diouf I, Carretero Y, Duboscq R, Bitton F, Djari A, Zouine M, Causse M. 2021. Genetic diversity of tomato response to heat stress at the QTL and transcriptome levels. The Plant Journal **107**, 1213–1227.

Blanca J, Cañizares J, Cordero L, Pascual L, Diez MJ, Nuez F. 2012. Variation revealed by SNP genotyping and morphology provides insight into the origin of the tomato. PLoS ONE **7**.

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

Boer MP, Wright D, Feng L, Podlich DW, Luo L, Cooper M, van Eeuwijk FA. 2007. A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. Genetics **177**, 1801–1813.

Bogard M, Ravel C, Paux E, Bordes J, Balfourier F, Chapman SC, Le Gouis J, Allard V. 2014. Predictions of heading date in bread wheat (*Triticum aestivum L.*) using QTL-based parameters of an ecophysiological model. Journal of Experimental Botany **65**, 5849–5865.

Bolker BM. 2008. Ecological Models and Data in R. Princeton University Press.

Boote KJ, Jones JW, Hoogenboom G. 2021. Incorporating realistic trait physiology into crop growth models to support genetic improvement (K Chenu and SP Long, Eds.). in silico Plants **3**, 2.

Boote KJ, Kropff MJ, Bindraban PS. 2001. Physiology and modelling of traits in crop plants: Implications for genetic improvement. Agricultural Systems **70**, 395–420.

Borrás L, Slafer GA, Otegui ME. 2004. Seed dry weight response to source-sink manipulations in wheat, maize and soybean: A quantitative reappraisal. Field Crops Research **86**, 131–146.

Bouman BAM, Van Keulen H, Van Laar HH, Rabbinge R. 1996. The 'School of de Wit' crop growth simulation models: A pedigree and historical overview. Agricultural Systems **52**, 171–198.

Bustos D v., Hasan AK, Reynolds MP, Calderini DF. 2013. Combining high grain number and weight through a DH-population to improve grain yield potential of wheat in high-yielding environments. Field Crops Research **145**, 106–115.

Bustos-Korts D, Boer MP, Chenu K, Zheng B, Chapman S, Van Eeuwijk FA. 2021. Genotype-specific P-spline response surfaces assist interpretation of regional wheat adaptation to climate change. in silico Plants **3**, 1–23.

Bustos-Korts D, Malosetti M, Chapman S, Biddulph B, van Eeuwijk F. 2016. Improvement of predictive ability by uniform coverage of the target genetic space. G3: Genes, Genomes, Genetics **6**, 3733–3747.

Bustos-Korts D, Malosetti M, Chenu K, Chapman S, Boer MP, Zheng B, van Eeuwijk FA. 2019. From QTLs to adaptation landscapes: using genotype-to-phenotype models to characterize G×E over time. Frontiers in Plant Science **10**, 1–23.

Butler DG, Cullis BR, Gilmour AR, Gogel BJ, Thompson R. 2017. ASReml-R reference manual version 4. , 176.

Cabrera-Bosquet L, Fournier C, Brichet N, Welcker C, Suard B, Tardieu F. 2016. High-throughput estimation of incident light, light interception and radiation-use efficiency of thousands of plants in a phenotyping platform. The New phytologist **212**, 269–281.

Campanelli G, Sestili S, Acciarri N, Montemurro F, Palma D, Leteo F, Beretta M. 2019. Multi-parental Advances Generation Inter-Cross population, to develop organic tomato genotypes by participatory plant breeding. Agronomy 2019, Vol. 9, Page 119 **9**, 119.

Causse M, Grandillo S. 2016. Gene mapping in tomato.23–37.

Causse M, Zhao J, Diouf I, Wang J, Lefebvre V, Caromel B, Génard M, Bertin N. 2020. Genomic designing for climate-smart tomato. Genomic designing of climate-smart vegetable crops. Springer International Publishing, 47–159.

Cavanagh C, Morell M, Mackay I, Powell W. 2008. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Current Opinion in Plant Biology **11**, 215–221.

Chakrabarti M, Zhang N, Sauvage C, *et al.* 2013. A cytochrome P450 regulates a domestication trait in cultivated tomato. Proceedings of the National Academy of Sciences of the United States of America **110**, 17125–17130.

Chandler RF, Eastin JD, Haskins FA, Sullivan CY, Van Bavel CHM, Dinauer RC. 1969. Plant morphology and stand geometry in relation to nitrogen plant morphology and stand geometry in relation to nitrogen plant morphology and stand geometry in relation to nitrogen. Agronomy & Horticulture -- Faculty Publications **196**.

Chang TG, Zhu XG, Raines C. 2017. Source-sink interaction: a century old concept under the light of modern molecular systems biology. Journal of experimental botany **68**, 4417–4431.

Chapman S, Cooper M, Podlich D, Hammer G. 2003. Evaluating plant breeding strategies by simulating gene action and dryland environment effects. Agronomy Journal **95**, 99–113.

Chenu K, Chapman SC, Hammer GL, McLean G, Salah HBH, Tardieu F. 2008. Short-term responses of leaf growth rate to water deficit scale up to whole-plant and crop levels: An integrated modelling approach in maize. Plant, Cell and Environment **31**, 378–391.

Chenu K, Chapman SC, Tardieu F, McLean G, Welcker C, Hammer GL. 2009. Simulating the yield impacts of organ-level quantitative trait loci associated with drought response in maize: A 'gene-to-phenotype' modeling approach. Genetics **183**, 1507–1523.

Chenu K, van Oosterom EJ, McLean G, *et al.* 2018. Integrating modelling and phenotyping approaches to identify and screen complex traits: transpiration efficiency in cereals. Journal of Experimental Botany **69**, 3181–3194.

Chenu K, Porter JR, Martre P, Basso B, Chapman SC, Ewert F, Bindi M, Asseng S. 2017. Contribution of crop models to adaptation in wheat. Trends in Plant Science **22**, 472–490.

Chitwood DH, Kumar R, Headland LR, *et al.* 2013. A quantitative genetic basis for leaf morphology in a set of precisely defined tomato introgression lines. Plant Cell **25**, 2465–2481.

Cockshull KE, Ho LC. 1995. Regulation of tomato fruit size by plant density and truss thinning. Journal of Horticultural Science **70**, 395–407.

Cooper M, van Eeuwijk FA, Hammer GL, Podlich DW, Messina C. 2009. Modeling QTL for complex traits: detection and context for plant breeding. Current Opinion in Plant Biology **12**, 231–240.

Cooper M, Hammer GL. 1996. Plant Adaptation and Crop Improvement. Wallingford, UK: CAB International.

Cooper M, Messina CD, Podlich D, Totir LR, Baumgarten A, Hausmann NJ, Wright D, Graham G. 2014. Predicting the future of plant breeding: Complementing empirical evaluation with genetic prediction. Crop and Pasture Science **65**, 311–336.

Cooper M, Powell O, Voss-Fels KP, *et al.* 2021. Modelling selection response in plant-breeding programs using crop models as mechanistic gene-to-phenotype (CGM-G2P) multi-trait link functions (V Vadez, Ed.). in silico Plants **3**, 1–21.

Cooper M, Technow F, Messina C, Gho C, Radu Totir L. 2016. Use of crop growth models with whole-genome prediction: Application to a maize multienvironment trial. Crop Science **56**, 2141–2156.

Costa JM, Heuvelink E. 2018. The global tomato industry. Tomatoes. Wallingford: CABI, 1–26.

Crossa J, Pérez-Rodríguez P, Cuevas J, *et al.* 2017. Genomic selection in plant breeding: methods, models, and perspectives. Trends in Plant Science **22**, 961–975.

Cuartero J, Bolarín MC, Asíns MJ, Moreno V. 2006. Increasing salt tolerance in the tomato. Journal of Experimental Botany **57**, 1045–1058.

Cullis BR, Smith AB, Coombes NE. 2006. On the design of early generation variety trials with correlated data. Journal of Agricultural, Biological, and Environmental Statistics **11**, 381–393.

Darvasi A, Soller M. 1997. A simple method to calculate resolving power and confidence interval of QTL map location. Behavior Genetics **27**, 125–132.

Dekkers JCM, Hospital F. 2002. The use of molecular genetics in the improvement of agricultural populations. Nature Reviews Genetics 2002 3:1 **3**, 22–32.

Dignat G, Welcker C, Sawkins M, Ribaut JM, Tardieu F. 2013. The growths of leaves, shoots, roots and reproductive organs partly share their genetic control in maize plants. Plant, Cell & Environment **36**, 1105–1119.

Diouf IA, Derivot L, Bitton F, Pascual L, Causse M. 2018. Water deficit and salinity stress reveal many specific QTL for plant growth and fruit quality traits in tomato. Frontiers in Plant Science **9**, 1–13.

Donald CM. 1968. The breeding of crop ideotypes. Euphytica **17**, 385–403.

Driever SM, Lawson T, Andralojc PJ, Raines CA, Parry MAJ. 2014. Natural variation in photosynthetic capacity, growth, and yield in 64 field-grown wheat genotypes. Journal of Experimental Botany **65**, 4959–4973.

Driever SM, Simkin AJ, Alotaibi S, Fisk SJ, Madgwick PJ, Sparks CA, Jones HD, Lawson T, Parry MAJ, Raines CA. 2017. Increased SBPase activity improves photosynthesis and grain yield in wheat grown in greenhouse conditions. Philosophical Transactions of the Royal Society B: Biological Sciences **372**.

Duvick DN, Smith JSC, Cooper M. 2004. Long-term selection in a commercial hybrid maize breeding program. Plant Breeding Reviews. Oxford, UK: John Wiley & Sons, Inc., 109–151.

Eeuwijk FA van, Bink MCAM, Chenu K, Chapman SC. 2010. Detection and use of QTL for complex traits in multiple environments. Current Opinion in Plant Biology **13**, 193–205.

Eeuwijk FA van, Bustos-Korts D, Malosetti M. 2016. What should students in plant breeding know about the statistical aspects of genotype × environment interactions? Crop Science **56**, 2119–2140.

Eeuwijk FA van, Bustos-Korts D, Millet EJ, *et al.* 2019. Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. Plant Science **282**, 23–39.

Evans LT. 1993. Crop evolution, adaptation and yield. Cambridge University Press.

Evans LT. 1998. Greater crop production whence and whither? In: Feeding a world population of more than eight million people, J.C.Waterlow, D.G Armstrong, L. Fowden and R. Riley (eds) (Oxford University Press) pp 89-97.

Evans JR. 2013. Improving photosynthesis. Plant physiology **162**, 1780–1793.

Fischer RA. 2007. Paper presented at international workshop on increasing wheat yield potential, CIMMYT, Obregon, Mexico, 20–24 March 2006 Understanding the physiological basis of yield potential in wheat. The Journal of Agricultural Science **145**, 99.

Fischer RA, Rees D, Sayre KD, Lu ZM, Condon AG, Larque Saavedra A. 1998. Wheat yield progress associated with higher stomatal conductance and photosynthetic rate, and cooler canopies. Crop Science **38**, 1467–1475.

Flood PJ, Harbinson J, Aarts MGM. 2011. Natural genetic variation in plant photosynthesis. Trends in Plant Science **16**, 327–335.

Flood PJ, Kruijer W, Schnabel SK, van der Schoor R, Jalink H, Snel JFH, Harbinson J, Aarts MGM. 2016. Phenomics for photosynthesis, growth and reflectance in *Arabidopsis thaliana* reveals circadian and long-term fluctuations in heritability. Plant Methods **12**, 14.

Foolad MR. 2007. Genome mapping and molecular breeding of tomato. International Journal of Plant Genomics **2007**.

Foolad MR, Panthee DR. 2012. Marker-assisted selection in tomato breeding. Critical Reviews in Plant Sciences **31**, 93–123.

Foulkes MJ, Slafer GA, Davies WJ, Berry PM, Sylvester-Bradley R, Martre P, Calderini DF, Griffiths S, Reynolds MP. 2011. Raising yield potential of wheat. III. Optimizing partitioning to grain while maintaining lodging resistance. Journal of Experimental Botany **62**, 469–486.

Foyer CH, Ruban A V., Nixon PJ. 2017. Photosynthesis solutions to enhance productivity. Philosophical Transactions of the Royal Society B: Biological Sciences **372**.

Frary A, Fulton TM, Zamir D, Tanksley SD. 2004. Advanced backcross QTL analysis of a *Lycopersicon esculentum* x *L. pennellii* cross and identification of possible orthologs in the Solanaceae. Theoretical and Applied Genetics **108**, 485–496.

Fukai S, Inthapanya P, Blamey FPC, Khunthasuvon S. 1999. Genotypic variation in rice grown in low fertile soils and drought-prone, rainfed lowland environments. Field Crops Research **64**, 121–130.

Fulop D, Ranjan A, Ofner I, *et al.* 2016. A new advanced backcross tomato population enables high resolution leaf QTL mapping and gene identification. G3: Genes, Genomes, Genetics **6**, 3169–3184.

Furbank RT, Sharwood R, Estavillo GM, Silva-Perez V, Condon AG. 2020. Photons to food: genetic improvement of cereal crop photosynthesis. Journal of Experimental Botany **71**, 2226–2238.

Gambín BL, Borrás L. 2010. Resource distribution and the trade-off between seed number and seed weight: A comparison across crop species. Annals of Applied Biology **156**, 91–102.

de Gelder A, Dieleman JA, Bot GPA, Marcelis LFM. 2012. An overview of climate and crop yield in closed greenhouses. Journal of Horticultural Science and Biotechnology **87**, 193–202.

Génard M, Memmah M-M, Quilot-Turion B, Vercambre G, Baldazzi V, Le Bot J, Bertin N, Gautier H, Lescourret F, Pagès L. 2016. Process-based simulation models are essential tools for virtual profiling and design of ideotypes: example of fruit and root. Crop Systems Biology. Cham: Springer International Publishing, 83–104.

Genty B, Briantais J-M, Baker NR. 1989. The relationship between the quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. Biochimica et Biophysica Acta (BBA) - General Subjects **990**, 87–92.

Giraud H, Bauland C, Falque M, *et al.* 2017. Reciprocal genetics: Identifying QTL for general and specific combining abilities in hybrids between multiparental populations from two maize (*Zea mays L.*) heterotic groups. Genetics **207**, 1167–1180.

Grandillo S, Cammareri M. 2016. Molecular mapping of quantitative trait loci in tomato. The Tomato Genome. Springer, Berlin, Heidelberg, 39–73.

Grandillo S, Chetelat R, Knapp S, *et al.* 2011. *Solanum sect. Lycopersicon*. Wild crop relatives: genomic and breeding resources. Springer Berlin Heidelberg, 129–215.

Griffiths S, Wingen L, Pietragalla J, *et al.* 2015. Genetic dissection of grain size and grain number trade-offs in CIMMYT wheat germplasm. PLoS ONE **10**, 1–18.

Gu J, Yin X, Stomph TJ, Struik PC. 2014*a*. Can exploiting natural genetic variation in leaf photosynthesis contribute to increasing rice productivity? A simulation analysis. Plant, Cell and Environment **37**, 22–34.

Gu J, Yin X, Stomph TJ, Wang H, Struik PC. 2012*a*. Physiological basis of genetic variation in leaf photosynthesis among rice (*Oryza sativa L.*) introgression lines under drought and well-watered conditions. Journal of Experimental Botany **63**, 5137–5153.

Gu J, Yin X, Struik PC, Stomph TJ, Wang H. 2012*b*. Using chromosome introgression lines to map quantitative trait loci for photosynthesis parameters in rice (*Oryza sativa L.*) leaves under drought and well-watered field conditions. Journal of Experimental Botany **63**, 455–469.

Gu J, Yin X, Zhang C, Wang H, Struik PC. 2014*b*. Linking ecophysiological modelling with quantitative genetics to support marker-assisted crop design for improved yields of rice (*Oryza sativa*) under drought stress. Annals of Botany **114**, 499–511.

Gur A, Osorio S, Fridman E, Fernie AR, Zamir D. 2010. hi2-1, A QTL which improves harvest index, earliness and alters metabolite accumulation of processing tomatoes. Theoretical and Applied Genetics **121**, 1587–1599.

Gur A, Zamir D. 2004. Unused natural variation can lift yield barriers in plant breeding. PLoS Biology **2**.

Hammer GL, Chapman S, van Oosterom E, Podlich DW. 2005. Trait physiology and crop modelling as a framework to link phenotypic complexity to underlying genetic systems. Australian Journal of Agricultural Research **56**, 947.

Hammer G, Cooper M, Tardieu F, Welch S, Walsh B, van Eeuwijk F, Chapman S, Podlich D. 2006. Models for navigating biological complexity in breeding improved crop plants. Trends in Plant Science **11**, 587–593.

Hammer GL, Kropff MJ, Sinclair TR, Porter JR. 2002. Future contributions of crop modelling - From heuristics and supporting decision making to understanding genetic regulation and aiding crop improvement. European Journal of Agronomy **18**, 15–31.

Hammer GL, Sinclair TR, Chapman SC, van Oosterom E. 2004. On systems thinking, systems biology, and the in silico plant. Plant Physiology **134**, 909–911.

Hasegawa T, Sakai H, Tokida T, *et al.* 2013. Rice cultivar responses to elevated CO_2 at two free-air CO_2 enrichment (FACE) sites in Japan. Functional Plant Biology **40**, 148–159.

Hay RKM. 1995. Harvest index: a review of its use in plant breeding and crop physiology. Annals of Applied Biology **126**, 197–216.

Hedden P. 2003. The genes of the Green Revolution. Trends in Genetics **19**, 5–9.

Heffner EL, Sorrells ME, Jannink JL. 2009. Genomic selection for crop improvement. Crop Science **49**, 1–12.

Hendriks A. 2016. De tomaat en de bizarre wereld van vers voedsel. Nieuw Amsterdam.

Heuvelink E. 1996. Tomato growth and yield : quantitative analysis and synthesis. Doctoral dissertation, Wageningen University.

Heuvelink E (Ed.). 2018. Tomatoes. Wallingford: CABI.

Heuvelink E, Bertin N. 1994. Dry-matter partitioning in a tomato crop: Comparison of two simulation models. Journal of Horticultural Science **69**, 885–903.

Heuvelink E, Buiskool RPM. 1995. Influence of sink-source interaction on dry matter production in tomato. Annals of Botany **75**, 381–389.

Heuvelink E, Okello RCO, Peet M, Giovannoni JJ, Dorais M. 2020. Tomato. In The physiology of vegetable crops, 2nd ed.; Wien, H.C., Stützel, H., Eds.; CABI: Boston, MA, USA; pp. 138–178.

Hickey JM, Chiurugwi T, Mackay I, Powell W. 2017. Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. Nature Genetics **49**, 1297–1303.

Hickey LT, N. Hafeez A, Robinson H, Jackson SA, Leal-Bertioli SCM, Tester M, Gao C, Godwin ID, Hayes BJ, Wulff BBH. 2019. Breeding crops to feed 10 billion. Nature Biotechnology **37**, 744–754.

Higashide T, Heuvelink E. 2009. Physiological and morphological changes over the past 50 years in yield components in tomato. Journal of the American Society for Horticultural Science **134**, 460–465.

Higashide T, Yasuba K ichiro, Suzuki K, Nakano A, Ohmori H. 2012. Yield of Japanese tomato cultivars has been hampered by a breeding focus on flavor. HortScience **47**, 1408–1411.

Hirose T. 2005. Development of the Monsi-Saeki theory on canopy structure and function. Annals of Botany **95**, 483–494.

Höhner R, Pribil M, Herbstová M, *et al.* 2020. Plastocyanin is the long-range electron carrier between photosystem II and photosystem i in plants. Proceedings of the National Academy of Sciences of the United States of America **117**, 15354–15362.

Holtan HEE, Hake S. 2003. quantitative trait locus analysis of leaf dissection in tomato using *Lycopersicon pennellii* segmental introgression lines. Genetics **165**, 1541–1550.

Honda S, Ohkubo S, San NS, Nakkasame A, Tomisawa K, Katsura K, Ookawa T, Nagano AJ, Adachi S. 2021. Maintaining higher leaf photosynthesis after heading stage could promote biomass accumulation in rice. Scientific Reports 2021 11:1 **11**, 1–11.

Hoogenboom G, White JW, Acosta-Gallegos J, Gaudiel RG, Myers JR, Silbernagel MJ. 1997. Evaluation of a crop simulation model that incorporates gene action. Agronomy Journal **89**, 613–620.

Hoogenboom G, White JW, Messina CD. 2004. From genome to crop: Integration through simulation modeling. Field Crops Research **90**, 145–163.

Huang X, Han B. 2014. Natural variations and Genome-Wide Association Studies in crop plants. Annual Review of Plant Biology **65**, 531–551.

Huang BE, Verbyla KL, Verbyla AP, Raghavan C, Singh VK, Gaur P, Leung H, Varshney RK, Cavanagh CR. 2015. MAGIC populations in crops: current status and future prospects. Theoretical and Applied Genetics **128**, 999–1017.

Hwang C, Correll MJ, Gezan SA, Zhang L, Bhakta MS, Vallejos CE, Boote KJ, Clavijo-Michelangeli JA, Jones JW. 2017. Next generation crop models: A modular approach to model early vegetative and reproductive development of the common bean (*Phaseolus vulgaris L*). Agricultural Systems **155**, 225–239.

Iglesias DJ, Lliso I, Tadeo FR, Talon M. 2002. Regulation of photosynthesis through source: sink imbalance in citrus is mediated by carbohydrate content in leaves. Physiologia Plantarum **116**, 563–572.

Jackson P, Robertson M, Cooper M, Hammer G. 1996. The role of physiological understanding in plant breeding; From a breeding perspective. Field Crops Research **49**, 11–37.

Ji Y, Nuñez Ocaña D, Choe D, Larsen DH, Marcelis LFM, Heuvelink E. 2020. Far-red radiation stimulates dry mass partitioning to fruits by increasing fruit sink strength in tomato. New Phytologist, nph.16805.

Jones JW, Dayan E, Allen LH, Van Keulen H, Challa H. 1991. A dynamic tomato growth and yield model (TOMGRO). Transactions of the ASAE **34**, 0663–0672.

Junker A, Muraya MM, Weigelt-Fischer K, Arana-Ceballos F, Klukas C, Melchinger AE, Meyer RC, Riewe D, Altmann T. 2015. Optimizing experimental procedures for quantitative evaluation of crop plant performance in high throughput phenotyping systems. Frontiers in Plant Science **5**, 1–21.

Kaiser E, Morales A, Harbinson J. 2018. Fluctuating light takes crop photosynthesis on a rollercoaster ride. Plant Physiology **176**, 977–989.

Khan MS, Struik PC, van der Putten PEL, Jansen HJ, van Eck HJ, van Eeuwijk FA, Yin X. 2019. A model-based approach to analyse genetic variation in potato using standard cultivars and a segregating population. I. Canopy cover dynamics. Field Crops Research **242**, 107581.

De Koning ANM. 1994. Development and dry matter distribution in glasshouse tomato: a quantitative approach.

Krause GH, Weis E. 1991. chlorophyll fluorescence and photosynthesis: the basics. Annual Review of Plant Physiology and Plant Molecular Biology **42**, 313–349.

Kromdijk J, Głowacka K, Leonelli L, Gabilly ST, Iwai M, Niyogi KK, Long SP. 2016. Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. Science **354**, 857–861.

Kropff MJ, Haverkort AJ, Aggarwal PK, Kooman PL. 1995. Using systems approaches to design and evaluate ideotypes for specific environments.417–435.

Labate JA, Grandillo S, Fulton T, *et al.* 2007. Tomato. In: Kole C, ed. Vegetables. Berlin, Heidelberg: Springer Berlin Heidelberg, 1–125.

Lacube S, Fournier C, Palaffre C, Millet EJ, Tardieu F, Parent B. 2017. Distinct controls of leaf widening and elongation by light and evaporative demand in maize. Plant Cell and Environment **40**, 2017–2028.

Lacube S, Manceau L, Welcker C, Millet EJ, Gouesnard B, Palaffre C, Ribaut JM, Hammer G, Parent B, Tardieu F. 2020. Simulating the effect of flowering time on maize individual leaf area in contrasting environmental scenarios. Journal of Experimental Botany **71**, 5577–5588.

Lamsal A, Welch SM, White JW, Thorp KR, Bello NM. 2018. Estimating parametric phenotypes that determine anthesis date in *Zea mays*: Challenges in combining ecophysiological models with genetics (FA Feltus, Ed.). PLOS ONE **13**, e0195841.

Lawson T, Kramer DM, Raines CA. 2012. Improving yield by exploiting mechanisms underlying natural variation of photosynthesis. Current Opinion in Biotechnology **23**, 215–220.

Li W, Boer MP, van Rossum B-J, Zheng C, Joosen RVL, van Eeuwijk FA. 2022. StatgenMPP: an R package implementing an IBD-based mixed model approach for QTL mapping in a wide range of multi-parent populations. Bioinformatics.

Li W, Boer MP, Zheng C, Joosen RVL, van Eeuwijk FA. 2021. An IBD-based mixed model approach for QTL mapping in multiparental populations. Theoretical and Applied Genetics **134**, 3643–3660.

Li T, Heuvelink E, Marcelis LFM. 2015. Quantifying the source-sink balance and carbohydrate content in three tomato cultivars. Frontiers in Plant Science **6**, 1–10.

Lin T, Zhu G, Zhang J, *et al.* 2014. Genomic analyses provide insights into the history of tomato breeding. Nature Genetics **46**, 1220–1226.

Lippman ZB, Semel Y, Zamir D. 2007. An integrated view of quantitative trait variation using tomato interspecific introgression lines. Current Opinion in Genetics and Development **17**, 545–552.

Long SP, Zhu X-G, Naidu SL, Ort DR. 2006. Can improvement in photosynthesis increase crop yields? Plant, Cell and Environment **29**, 315–330.

Longin CFH, Reif JC. 2014. Redesigning the exploitation of wheat genetic resources. Trends in Plant Science **19**, 631–636.

López-Calcagno PE, Brown KL, Simkin AJ, Fisk SJ, Vialet-Chabrand S, Lawson T, Raines CA. 2020. Stimulating photosynthetic processes increases productivity and water-use efficiency in the field. Nature Plants 2020 6:8 **6**, 1054–1063.

Lorenzo P, Maroto C, Castilla N. 1990. CO_2 in plastic greenhouse in Almeria (Spain) - refereed paper -. Acta Horticulturae, 165–170.

Luis Caparrós-Martínez J, Rueda-Lópe N, Milán-García J, de Pablo Valenciano J. 2020. Public policies for sustainability and water security: The case of Almeria (Spain). Global Ecology and Conservation **23**, e01037.

Maccaferri M, Sanguineti MC, Corneti S, *et al.* 2008. Quantitative trait loci for grain yield and adaptation of Durum Wheat (*Triticum durum Desf.*) across a wide range of water availability. Genetics **178**, 489–511.

Marcelis LFM. 1996. Sink strength as a determinant of dry matter partitioning in the whole plant. Journal of Experimental Botany **47**, 1281–1291.

Marcelis LFM, Heuvelink E, Goudriaan J. 1998. Modelling biomass production and yield of horticultural crops: A review. Scientia Horticulturae **74**, 83–111.

Markwell J, Osterman JC, Mitchell JL. 1995. Calibration of the Minolta SPAD-502 leaf chlorophyll meter. Photosynthesis Research 1995 46:3 **46**, 467–472.

Martre P, He J, Le Gouis J, Semenov MA. 2015. In silico system analysis of physiological traits determining grain yield and protein concentration for wheat as influenced by climate and crop management. Journal of Experimental Botany **66**, 3581–3598.

Mathan J, Singh A, Jathar V, Ranjan A. 2021. High photosynthesis rate in two wild rice species is driven by leaf anatomy mediating high Rubisco activity and electron transport rate. Journal of Experimental Botany **72**, 7119–7135.

McAusland L, Atkinson JA, Lawson T, Murchie EH. 2019. High throughput procedure utilising chlorophyll fluorescence imaging to phenotype dynamic photosynthesis and photoprotection in leaves under controlled gaseous conditions. Plant Methods **15**, 1–15.

McAusland L, Vialet-Chabrand S, Jauregui I, *et al.* 2020. Variation in key leaf photosynthetic traits across wheat wild relatives is accession dependent not species dependent. New Phytologist **228**, 1767–1780.

McMullen MD, Kresovich S, Villeda HS, *et al.* 2009. Genetic properties of the maize nested association mapping population. Science **325**, 737–740.

Melis A. 2009. Solar energy conversion efficiencies in photosynthesis: Minimizing the chlorophyll antennae to maximize efficiency. Plant Science **177**, 272–280.

Messina CD, Jones JW, Boote KJ, Vallejos CE. 2006. A gene-based model to simulate soybean development and yield responses to environment. Crop Science **46**, 456–466.

Messina CD, Podlich D, Dong Z, Samples M, Cooper M. 2011. Yield–trait performance landscapes: from theory to application in breeding maize for drought tolerance. Journal of Experimental Botany **62**, 855–868.

Messina CD, Technow F, Tang T, Totir R, Gho C, Cooper M. 2018. Leveraging biological insight and environmental variation to improve phenotypic prediction: Integrating crop growth models (CGM) with whole genome prediction (WGP). European Journal of Agronomy **100**, 151–162.

Meuwissen THE, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics **157**, 1819–1829.

Miflin B. 2000. Crop improvement in the 21st century. Journal of Experimental Botany **51**, 1–8.

Miller JC, Tanksley SD. 1990. RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon. Theoretical and Applied Genetics **80**, 437–448.

Millet EJ, Kruijer W, Coupel-Ledru A, Alvarez Prado S, Cabrera-Bosquet L, Lacube S, Charcosset A, Welcker C, van Eeuwijk F, Tardieu F. 2019. Genomic prediction of maize yield across European environmental conditions. Nature Genetics **51**, 952–956.

Millet E, Welcker C, Kruijer W, *et al.* 2016. Genome-wide analysis of yield in Europe: allelic effects as functions of drought and heat scenarios. Plant Physiology **172**, pp.00621.2016.

Molero G, Joynson R, Pinera-Chavez FJ, Gardiner LJ, Rivera-Amado C, Hall A, Reynolds MP. 2019. Elucidating the genetic basis of biomass accumulation and radiation use efficiency in spring wheat and its role in yield potential. Plant Biotechnology Journal **17**, 1276–1288.

Monsi M, Saeki T. 1953. Uber den Lichtfaktor in den Pflanzengesellschaften und seine Bedeutung fur die Stoffproduktion. Japanese Journal of Botany **14**, 22–52.

Monteith JL. 1977. Climate and the efficiency of crop production in Britain. Philosophical Transactions of the Royal Society of London. B, Biological Sciences **281**, 277–294.

Monteith JL. 1981. Climatic variation and the growth of crops. Quarterly Journal of the Royal Meteorological Society **107**, 749–774.

Montesinos-López OA, Montesinos-López A, Crossa J, los Campos G, Alvarado G, Suchismita M, Rutkoski J, González-Pérez L, Burgueño J. 2017. Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. Plant Methods **13**.

Muir CD, Pease JB, Moyle LC. 2014. Quantitative genetic analysis indicates natural selection on leaf phenotypes across wild tomato species (*Solanum sect. Lycopersicon* ; Solanaceae). Genetics **198**, 1629–1643.

Mujeeb-Kazi A, Kazi AG, Dundas I, *et al.* 2013. Genetic diversity for wheat improvement as a conduit to food security. Elsevier Inc.

Muraya MM, Chu J, Zhao Y, Junker A, Klukas C, Reif JC, Altmann T. 2017. Genetic variation of growth dynamics in maize (*Zea mays L.*) revealed through automated non-invasive phenotyping. The Plant Journal **89**, 366–380.

Murchie EH, Kefauver S, Araus JL, Muller O, Rascher U, Flood PJ, Lawson T. 2018. Measuring the dynamic photosynthome. Annals of Botany **122**, 207–220.

Murchie EH, Lawson T. 2013. Chlorophyll fluorescence analysis: A guide to good practice and understanding some new applications. Journal of Experimental Botany **64**, 3983–3998.

Nakagawa H, Yamagishi J, Miyamoto N, Motoyama M, Yano M, Nemoto K. 2005. Flowering response of rice to photoperiod and temperature: A QTL analysis using a phenological model. Theoretical and Applied Genetics **110**, 778–786.

Nakano H, Sasaki K, Mine Y, Takahata K, Lee O, Sugiyama N. 2016. Quantitative trait loci (QTL) controlling plant architecture traits in a *Solanum lycopersicum* × *S. pimpinellifolium* cross. Euphytica **211**, 353–367.

Navarro JM, Flores P, Carvajal M, Martinez V. 2005. Changes in quality and yield of tomato fruit with ammonium, bicarbonate and calcium fertilisation under saline conditions. The Journal of Horticultural Science and Biotechnology **80**, 351–357.

Ohyama A, Shirasawa K, Matsunaga H, Negoro S, Miyatake K, Yamaguchi H, Nunome T, Iwata H, Fukuoka H, Hayashi T. 2017. Bayesian QTL mapping using genome-wide SSR markers and segregating population derived from a cross of two commercial F1 hybrids of tomato. Theoretical and Applied Genetics **130**, 1601–1616.

Oliveira FAA, Jones JW, Pavan W, Bhakta M, Vallejos CE, Correll MJ, Boote KJ, Fernandes JMC, Hölbig CA, Hoogenboom G. 2021. Incorporating a dynamic gene-based process module into a crop simulation model (X Yin, Ed.). in silico Plants **3**, 1–18.

Ort DR, Merchant SS, Alric J, *et al.* 2015. Redesigning photosynthesis to sustainably meet global food and bioenergy demand. Proceedings of the National Academy of Sciences **112**, 8529–8536.

Ort DR, Zhu X, Melis A. 2011. Optimizing antenna size to maximize photosynthetic efficiency. Plant Physiology **155**, 79–85.

Ortiz R, Braun HJ, Crossa J, *et al.* 2008. Wheat genetic resources enhancement by the International Maize and Wheat Improvement Center (CIMMYT). Genetic Resources and Crop Evolution **55**, 1095–1140.

Parent B, Leclere M, Lacube S, Semenov MA, Welcker C, Martre P, Tardieu F. 2018. Maize yields over Europe may increase in spite of climate change, with an appropriate use of the genetic variability of flowering time. Proceedings of the National Academy of Sciences of the United States of America **115**, 10642–10647.

Parent B, Tardieu F. 2014. Can current crop models be used in the phenotyping era for predicting the genetic variability of yield of plants subjected to drought or high temperature? Journal of Experimental Botany **65**, 6179–6189.

Parry MAJ, Andralojc PJ, Scales JC, Salvucci ME, Carmo-Silva AE, Alonso H, Whitney SM. 2013. Rubisco activity and regulation as targets for crop improvement. Journal of Experimental Botany **64**, 717–730.

Pascual L, Albert E, Sauvage C, *et al.* 2016. Dissecting quantitative trait variation in the resequencing era: Complementarity of bi-parental, multi-parental and association panels. Plant Science **242**, 120–130.

Pascual L, Desplat N, Huang BE, Desgroux A, Bruguier L, Bouchet JP, Le QH, Chauchard B, Verschave P, Causse M. 2015. Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. Plant Biotechnology Journal **13**, 565–577.

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature **335**, 721–726.

Peng S, Khush GS, Virk P, Tang Q, Zou Y. 2008. Progress in ideotype breeding to increase rice yield potential. Field Crops Research **108**, 32–38.

Peng B, Li Y, Wang Y, *et al.* 2011. QTL analysis for yield components and kernel-related traits in maize across multi-environments. Theoretical and Applied Genetics **122**, 1305–1320.

Pertuzé RA, Ji Y, Chetelat RT. 2002. Comparative linkage map of the *Solanum lycopersicoides* and *S. sitiens* genomes and their differentiation from tomato. Genome **45**, 1003–1012.

Phan NT, Trinh LT, Rho M-Y, Park T-S, Kim O-R, Zhao J, Kim H-M, Sim S-C. 2019. Identification of loci associated with fruit traits using genome-wide single nucleotide polymorphisms in a core collection of tomato (*Solanum lycopersicum L.*). Scientia Horticulturae **243**, 567–574.

Van Der Ploeg A, Van Der Meer M, Heuvelink E. 2007. Breeding for a more energy efficient greenhouse tomato: Past and future perspectives. Euphytica **158**, 129–138.

Pnueli L, Carmel-Goren L, Hareven D, Gutfinger T, Alvarez J, Ganal M, Zamir D, Lifschitz E. 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.

Porcar-Castell A, Pfündel E, Korhonen JFJ, Juurola E. 2008. A new monitoring PAM fluorometer (MONI-PAM) to study the short- and long-term acclimation of photosystem II in field conditions. Photosynthesis Research **96**, 173–179.

Price A, Courtois B. 1999. Mapping QTLs associated with drought resistance in rice: Progress, problems and prospects. Plant Growth Regulation **29**, 123–133.

Quilot-Turion B, Génard M, Valsesia P, Memmah MM. 2016. Optimization of allelic combinations controlling parameters of a peach quality model. Frontiers in Plant Science **7**, 1–14.

R Core Team. 2020. R: A language and environment for statistical computing. Vienna, Austria.

Ralph PJ, Gademann R. 2005. Rapid light curves: A powerful tool to assess photosynthetic activity. Aquatic Botany **82**, 222–237.

Ramstein GP, Jensen SE, Buckler ES. 2019. Breaking the curse of dimensionality to identify causal variants in Breeding 4. Theoretical and Applied Genetics **132**, 559–567.

Ray DK, Mueller ND, West PC, Foley JA. 2013. Yield trends are insufficient to double global crop production by 2050. PLoS ONE **8**.

Ray DK, Ramankutty N, Mueller ND, West PC, Foley JA. 2012. Recent patterns of crop yield growth and stagnation. Nature Communications **3**, 1–7.

Rebetzke GJ, Richards RA. 1999. Genetic improvement of early vigour in wheat. Australian Journal of Agricultural Research **50**, 291–302.

Reymond M, Muller B, Leonardi A, Charcosset A, Tardieu F. 2003. Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. Plant Physiol. **131**, 664–675.

Reymond M, Muller B, Tardieu F. 2004. Dealing with the genotype x environment interaction via a modelling approach: A comparison of QTLs of maize leaf length or width with QTLs of model parameters. Journal of Experimental Botany **55**, 2461–2472.

Reynolds M, Bonnett D, Chapman SC, Furbank RT, Man??s Y, Mather DE, Parry MAJ. 2011. Raising yield potential of wheat. I. Overview of a consortium approach and breeding strategies. Journal of Experimental Botany **62**, 439–452.

Reynolds M, Foulkes J, Furbank R, Griffiths S, King J, Murchie E, Parry M, Slafer G. 2012. Achieving yield gains in wheat. Plant, Cell and Environment **35**, 1799–1823.

Reynolds M, Langridge P. 2016. Physiological breeding. Current Opinion in Plant Biology **31**, 162–171.

Richards RA. 2000. Selectable traits to increase crop photosynthesis and yield of grain crops. Journal of Experimental Botany **51**, 447–458.

Rick CM. 1978. The Tomato. Scientific American **239**, 76–89.

Rick CM, Chetelat RT. 1995. Utilization of related wild species for tomato improvement. Acta Horticulturae, 21–38.

Rodríguez-Álvarez MX, Boer MP, van Eeuwijk FA, Eilers PHC. 2018. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. Spatial Statistics **23**, 52–71.

Ronga D, Francia E, Rizza F, Badeck F-W, Caradonia F, Montevecchi G, Pecchioni N. 2019. Changes in yield components, morphological, physiological and fruit quality traits in processing tomato cultivated in Italy since the 1930's. Scientia Horticulturae **257**, 108726.

Ronga D, Zaccardelli M, Lovelli S, Perrone D, Francia E, Milc J, Ulrici A, Pecchioni N. 2017. Biomass production and dry matter partitioning of processing tomato under organic vs conventional cropping systems in a Mediterranean environment. Scientia Horticulturae **224**, 163–170.

van Rooijen R, Kruijer W, Boesten R, van Eeuwijk FA, Harbinson J, Aarts MGM. 2017. Natural variation of *YELLOW SEEDLING1* affects photosynthetic acclimation of *Arabidopsis thaliana*. Nature Communications **8**, 1421.

Rosenqvist E. 2000. IntelliGrow-a new climate control concept. Grøn Viden, 1–8.

Rosenqvist E. 2005. Optimisation of CO_2 and temperature in terms of crop growth and energy use. Acta Horticulturae, 149–154.

Rothan C, Diouf I, Causse M. 2019. Trait discovery and editing in tomato. Plant Journal **97**, 73–90.

Ruggieri V, Francese G, Sacco A, D'Alessandro A, Rigano MM, Parisi M, Milone M, Cardi T, Mennella G, Barone A. 2014. An association mapping approach to identify favourable alleles for tomato fruit quality breeding. BMC Plant Biology **14**, 337.

Rutkoski J, Poland J, Mondal S, Autrique E, Pérez LG, Crossa J, Reynolds M, Singh R. 2016. Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. G3: Genes, Genomes, Genetics **6**, 2799–2808.

Sacco A, Ruggieri V, Parisi M, Festa G, Rigano MM, Picarella ME, Mazzucato A, Barone A. 2015. Exploring a tomato landraces collection for fruit-related traits by the aid of a high-throughput genomic platform (H Ezura, Ed.). PLOS ONE **10**, e0137139.

Sadras VO. 2007. Evolutionary aspects of the trade-off between seed size and number in crops. Field Crops Research **100**, 125–138.

Sadras VO, Lawson C. 2011. Genetic gain in yield and associated changes in phenotype, trait plasticity and competitive ability of South Australian wheat varieties released between 1958 and 2007. Crop and Pasture Science **62**, 533–549.

Sadras VO, Richards RA. 2014. Improvement of crop yield in dry environments: Benchmarks, levels of organisation and the role of nitrogen. Journal of Experimental Botany **65**, 1981–1995.

Saito T, Mochizuki Y, Kawasaki Y, Ohyama A, Higashide T. 2020. Estimation of leaf area and light-use efficiency by non-destructive measurements for growth modeling and recommended leaf area index in greenhouse tomatoes. The Horticulture Journal **89**, 445–453.

Sarlikioti V, De Visser PHB, Buck-Sorlin GH, Marcelis LFM. 2011. How plant architecture affects light absorption and photosynthesis in tomato: Towards an ideotype for plant architecture using a functionalstructural plant model. Annals of Botany **108**, 1065–1073.

Sauvage C, Segura V, Bauchet G, Stevens R, Do PT, Nikoloski Z, Fernie AR, Causse M. 2014. Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. Plant Physiology **165**, 1120–1132.

Savvides A, van Ieperen W, Dieleman JA, Marcelis LFM. 2013. Meristem temperature substantially deviates from air temperature even in moderate environments: Is the magnitude of this deviation species-specific? Plant, Cell and Environment **36**, 1950–1960.

Scholberg JMS, Boote KJ, Jones JW, McNeal BL. 1997. Adaptation of the CROPGRO model to simulate the growth of field-grown tomato. Springer, Dordrecht, 135–151.

Schouten HJ, Tikunov Y, Verkerke W, Finkers R, Bovy A, Bai Y, Visser RG. 2019. Breeding has increased the diversity of cultivated tomato in The Netherlands. Frontiers in Plant Science **10**.

Schulthess AW, Kale SM, Liu F, *et al.* 2022. Genomics-informed prebreeding unlocks the diversity in genebanks for wheat improvement. Nature Genetics **54**, 1544–1552.

Schulthess AW, Reif JC, Ling J, *et al.* 2017. The roles of pleiotropy and close linkage as revealed by association mapping of yield and correlated traits of wheat (*Triticum aestivum L.*). Journal of Experimental Botany **68**, 4089–4101.

Schwarz D, Kläring HP. 2001. Allometry to estimate leaf area of tomato. Journal of Plant Nutrition **24**, 1291–1309.

Scott MF, Ladejobi O, Amer S, *et al.* 2020. Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. Heredity 2020 125:6 **125**, 396–416.

Sharwood RE, Quick WP, Sargent D, Estavillo GM, Silva-Perez V, Furbank RT. 2022. Mining for allelic gold: finding genetic variation in photosynthetic traits in crops and wild relatives. Journal of Experimental Botany **73**, 3085–3108.

Shibuya T, Itagaki K, Tojo M, Endo R, Kitaya Y. 2011. Fluorescent illumination with high red-to-far-red ratio improves resistance of cucumber seedlings to powdery mildew. HortScience **46**, 429–431.

Shirasawa K, Fukuoka H, Matsunaga H, Kobayashi Y, Kobayashi I, Hirakawa H, Isobe S, Tabata S. 2013. Genome-wide association studies using single nucleotide polymorphism markers developed by re-sequencing of the genomes of cultivated tomato. DNA research : an international journal for rapid publication of reports on genes and genomes **20**, 593–603.

Silva FM de O, Lichtenstein G, Alseekh S, *et al.* 2018. The genetic architecture of photosynthesis and plant growth-related traits in tomato. Plant, Cell & Environment **41**, 327–341.

Sim S-C, van Deynze A, Stoffel K, *et al.* 2012. High-density SNP genotyping of tomato (*Solanum lycopersicum L.*) reveals patterns of genetic variation due to breeding (S-B Wu, Ed.). PLoS ONE **7**, e45520.

Sim SC, Robbins MD, Chilcott C, Zhu T, Francis DM. 2009. Oligonucleotide array discovery of polymorphisms in cultivated tomato (*Solanum lycopersicum L.*) reveals patterns of SNP variation associated with breeding. BMC Genomics **10**, 466.

Simkin AJ, López-Calcagno PE, Raines CA. 2019. Feeding the world: improving photosynthetic efficiency for sustainable crop production. Journal of Experimental Botany **70**, 1119–1140.

Singh S, Jighly A, Sehgal D, *et al.* 2021. Direct introgression of untapped diversity into elite wheat lines. Nature Food **2**, 819–827.

Sivasakthi K, Thudi M, Tharanya M, *et al.* 2018. Plant vigour QTLs co-map with an earlier reported QTL hotspot for drought tolerance while water saving QTLs map in other regions of the chickpea genome. BMC Plant Biology **18**, 1–18.

Slafer GA, Savin R, Sadras VO. 2014. Coarse and fine regulation of wheat yield components in response to genotype and environment. Field Crops Research **157**, 71–83.

Smith MR, Rao IM, Merchant A. 2018. Source-sink relationships in crop plants and their influence on yield development and nutritional quality. Frontiers in Plant Science **9**, 1–10.

Song Q, Wang Y, Qu M, Ort DR, Zhu XG. 2017. The impact of modifying photosystem antenna size on canopy photosynthetic efficiency—Development of a new canopy photosynthesis model scaling from metabolism to canopy level processes. Plant Cell and Environment **40**, 2946–2957.

Stanghellini C, Incrocci L, Carlos Gázquez J, Dimauro B. Carbon dioxide concentration in Mediterranean greenhouses: how much lost production? Acta Horticulturae, GreenSys.

Stitzer MC, Ross-Ibarra J. 2018. Maize domestication and gene interaction. New Phytologist **220**, 395–408.

Sun J, Rutkoski JE, Poland JA, Crossa J, Jannink J-L, Sorrells ME. 2017. Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. The Plant Genome **10**, plantgenome2016.11.0111.

Sun J, Yang L, Wang Y, Ort DR. 2009. FACE-ing the global change: Opportunities for improvement in photosynthetic radiation use efficiency and crop yield. Plant Science **177**, 511–522.

Suriharn B, Patanothai A, Boote KJ, Hoogenboom G. 2011. Designing a peanut ideotype for a target environment using the CSM-CROPGRO-Peanut model. Crop Science **51**, 1887–1902.

Suriharn B, Patanothai A, Pannangpetch K, Jogloy S, Hoogenboom G. 2007. Determination of cultivar coefficients of peanut lines for breeding applications of the CSM-CROPGRO-Peanut model. Crop Science **47**, 607–619.

Tanger P, Klassen S, Mojica JP, *et al.* 2017. Field-based high throughput phenotyping rapidly identifies genomic regions controlling yield components in rice. Scientific Reports 2017 7:1 **7**, 1–8.

Tanksley SD, Ganal MW, Prince JP, *et al.* 1992. High density molecular linkage maps of the tomato and potato genomes. Genetics **132**, 1141–1160.

Tanksley SD, McCouch SR. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. Science **277**, 1063–1066.

Tanksley SD, Nelson JC. 1996. Advanced backcross QTL analysis: A method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theoretical and Applied Genetics **92**, 191–203.

Tardieu F. 2003. Virtual plants: Modelling as a tool for the genomics of tolerance to water deficit. Trends in Plant Science **8**, 9–14.

Tardieu F. 2012. Any trait or trait-related allele can confer drought tolerance: Just design the right drought scenario. Journal of Experimental Botany **63**, 25–31.

Tardieu F, Tuberosa R. 2010. Dissection and modelling of abiotic stress tolerance in plants. Current Opinion in Plant Biology **13**, 206–212.

Taylor SH, Long SP. 2017. Slow induction of photosynthesis on shade to sun transitions in wheat may cost at least 21% of productivity. Philosophical Transactions of the Royal Society B: Biological Sciences **372**.

Technow F, Messina CD, Totir LR, Cooper M. 2015. Integrating crop growth models with whole genome prediction through approximate Bayesian computation (I de Smet, Ed.). PLOS ONE **10**, e0130855.

Tester M, Langridge P. 2010. Breeding technologies to increase crop production in a changing world. Science **327**, 818–822.

Thornley JHM. 1998. Dynamic model of leaf photosynthesis with acclimation to light and nitrogen. Annals of Botany **81**, 421–430.

Timmerman-Vaughan GM, Mills A, Whitfield C, Frew T, Butler R, Murray S, Lakeman M, McCallum J, Russell A, Wilson D. 2005. Linkage mapping of QTL for seed yield, yield components, and developmental traits in pea. Crop Science **45**, 1336–1344.

Toda Y, Wakatsuki H, Aoike T, *et al.* 2020. Predicting biomass of rice with intermediate traits: Modeling method combining crop growth models and genomic prediction models (L Lukens, Ed.). PLOS ONE **15**, e0233951.

Tomato Genome Consortium T, DNA Research Institute K, Sciences L, *et al.* 2012. The tomato genome sequence provides insights into fleshy fruit evolution.

Trethowan RM, Mujeeb-Kazi A. 2008. Novel germplasm resources for improving environmental stress tolerance of hexaploid wheat. Crop Science **48**, 1255–1265.

Tripodi P. 2021. Methods of development of biparental mapping populations in horticultural crops. Methods in Molecular Biology **2264**, 1–12.

Tsutsumi-Morita Y, Heuvelink E, Khaleghi S, Bustos-Korts D, Marcelis LFM, Vermeer KMCA, van Dijk H, Millenaar FF, van Voorn GAK, van Eeuwijk FA. 2021. Yield dissection models to improve yield: a case study in tomato (C Messina and SP Long, Eds.). in silico Plants **3**, 1–18.

Turner SD, Maurizio PL, Valdar W, Yandell BS, Simon PW. 2018. Dissecting the genetic architecture of shoot growth in carrot (*Daucus carota L.*) using a diallel mating design. G3: Genes, Genomes, Genetics **8**, 411–426.

Unkovich M, Baldock J, Forbes M. 2010. Variability in harvest index of grain crops and potential significance for carbon accounting. Elsevier Inc.

Uptmoor R, Schrag T, Stützel H, Esch E. 2008. Crop model based QTL analysis across environments and QTL based estimation of time to floral induction and flowering in *Brassica oleracea*. Molecular Breeding **21**, 205–216.

Velazco JG, Jordan DR, Mace ES, Hunt CH, Malosetti M, van Eeuwijk FA. 2019. Genomic prediction of grain yield and drought-adaptation capacity in sorghum is enhanced by multi-trait analysis. Frontiers in Plant Science **10**, 997.

Velazco JG, Rodríguez-Álvarez MX, Boer MP, Jordan DR, Eilers PHC, Malosetti M, van Eeuwijk FA. 2017. Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model. Theoretical and Applied Genetics **130**, 1375–1392.

Voss-Fels KP, Cooper M, Hayes BJ. 2019. Accelerating crop genetic gains with genomic selection. Theoretical and Applied Genetics **132**, 669–686.

Walter J, Kromdijk J. 2022. Here comes the sun: How optimization of photosynthetic light reactions can boost crop yields. Journal of Integrative Plant Biology **64**, 564–591.

Wang W, Gao X, Cheng Y, Ren Y, Zhang Z, Wang R, Cao J, Geng H. 2022. QTL Mapping of leaf area index and chlorophyll content based on UAV remote sensing in wheat. Agriculture (Switzerland) **12**.

Washburn JD, Burch MB, Franco JAV. 2020. Predictive breeding for maize: Making use of molecular phenotypes, machine learning, and physiological crop models. Crop Science **60**, 622–638.

Watanabe N, Evans J, Chow W. 1994. Changes in the photosynthetic properties of Australian wheat cultivars over the last century. Functional Plant Biology **21**, 169.

Welcker C, Boussuge B, Bencivenni C, Ribaut JM, Tardieu F. 2007. Are source and sink strengths genetically linked in maize plants subjected to water deficit? A QTL study of the responses of leaf growth and of anthesis-silking interval to water deficit. Journal of Experimental Botany **58**, 339–349.

White JW, Hoogenboom G. 1996. Simulating effects of genes for physiological traits in a process-oriented crop model. Agronomy Journal **88**, 416–422.

White JW, Hoogenboom G. 2003. Gene-based approaches to crop simulation: Past experiences and future opportunities. Agronomy Journal **95**, 52–64.

Wu R, Casella G, Ma C-X. 2007. Statistical Genetics of Quantitative Traits. New York, NY: Springer New York.

Xiong D, Yu T, Liu X, Li Y, Peng S, Huang J. 2015. Heterogeneity of photosynthesis within leaves is associated with alteration of leaf structural features and leaf N content per leaf area in rice. Functional Plant Biology **42**, 687–696.

Yamamoto E, Matsunaga H, Onogi A, *et al.* 2016. A simulation-based breeding design that uses whole-genome prediction in tomato. Scientific Reports **6**, 1–11.

Yin X, Chasalow SD, Dourleijn CJ, Stam P, Kropff MJ. 2000*a*. Coupling estimated effects of QTLs for physiological traits to a crop growth model: predicting yield variation among recombinant inbred lines in barley. Heredity **85**, 539–549.

Yin X, Chasalow SD, Stam P, Kropff MJ, Dourleijn CJ, Bos I, Bindraban PS. 2002. Use of component analysis in QTL mapping of complex crop traits: A case study on yield in barley. Plant Breeding **121**, 314–319.

Yin X, Gu J, Dingkuhn M, Struik PC. 2022. A model-guided holistic review of exploiting natural variation of photosynthesis traits in crop improvement. Journal of Experimental Botany **73**, 3173–3188.

Yin X, Kropff MJ, Goudriaan J, Stam P. 2000*b*. A model analysis of yield differences among recombinant inbred lines in barley. Agronomy Journal **92**, 114–120.

Yin X, Kropff MJ, Stam P. 1999*a*. The role of ecophysiological models in QTL analysis: The example of specific leaf area in barley. Heredity **82**, 415–421.

Yin X, Stam P, Dourleijn CJ, Kropff MJ. 1999*b*. AFLP mapping of quantitative trait loci for yield-determining physiological characters in spring barley. Theoretical and Applied Genetics **99**, 244–253.

Yin X, Struik PC. 2008. Applying modelling experiences from the past to shape crop systems biology: The need to converge crop physiology and functional genomics. New Phytologist **179**, 629–642.

Yin X, Struik PC, Van Eeuwijk FA, Stam P, Tang J. 2005*a*. QTL analysis and QTL-based prediction of flowering phenology in recombinant inbred lines of barley. Journal of Experimental Botany **56**, 967–976.

Yin X, Struik PC, Gu J, Wang H. 2016. Modelling QTL-trait-crop relationships: Past experiences and future prospects. In: Yin X,, In: Struik PC, eds. Crop Systems Biology. Cham: Springer International Publishing, 193–218.

Yin X, Struik PC, Kropff MJ. 2004. Role of crop physiology in predicting gene-to-phenotype relationships. Trends in Plant Science **9**, 426–432.

Yin X, Struik PC, Tang J, Qi C, Liu T. 2005*b*. Model analysis of flowering phenology in recombinant inbred lines of barley. Journal of Experimental Botany **56**, 959–965.

Yu J, Holland JB, McMullen MD, Buckler ES. 2008. Genetic design and statistical power of nested association mapping in maize. Genetics **178**, 539–551.

Yu J, Pressoir G, Briggs WH, *et al.* 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics **38**, 203–208.

Zalapa JE, Staub JE, Mccreight JD, Chung · S M, Cuevas · H. 2007. Detection of QTL for yield-related traits using recombinant inbred lines derived from exotic and elite US Western Shipping melon germplasm. Theor Appl Genet **114**, 1185–1201.

Zamir D. 2001. Improving plant breeding with exotic genetic libraries. Nature Reviews Genetics **2**, 983–989.

Zhang X, Huang C, Wu D, *et al.* 2017. High-throughput phenotyping and QTL mapping reveals the genetic architecture of maize plant growth. Plant Physiology **173**, 1554–1564.

Zheng B, Biddulph B, Li D, Kuchel H, Chapman S. 2013. Quantification of the effects of *VRN1* and *Ppd-D1* to predict spring wheat (*Triticum aestivum*) heading time across diverse environments. **64**, 3747–3761.

Zheng C, Boer MP, van Eeuwijk FA. 2015. Reconstruction of genome ancestry blocks in multiparental populations. Genetics **200**, 1073–1087.

Zhu X-G, Long SP, Ort DR. 2010. Improving photosynthetic efficiency for greater yield. Annual Review of Plant Biology **61**, 235–261.

Zhu X, Ort DR, Whitmarsh J, Long SP. 2004. The slow reversibility of photosystem II thermal energy dissipation on transfer from high to low light may cause large losses in carbon gain by crop canopies: a theoretical analysis. Journal of Experimental Botany **55**, 1167–1175.

Summary

Yield improvement is one of the most important targets for plant breeding. To further improve crop yield, it is proposed to introduce genetic variation from wild genetic resources into elite cultivars. However, identifying the genetic loci that increase yield, and the traits that drive this improvement remains challenging. Yield is a complex trait, which is an integrated outcome of many underlying physiological processes, each of which is controlled by many genes under the influence of the environment and crop management. One strategy to disentangle this complexity is to dissect yield into component traits. Component traits are expected to be under more direct genetic control. Thus, component traits may show less genotype-by-environment interaction (G×E) and higher heritability. Furthermore, more quantitative trait loci (QTLs) for component traits than for yield may be detected. Therefore, the selection for component traits might be more stable compared to the selection for yield. Selection for high-yielding lines via component traits is pursued by identifying a phenotypic ideotype for high yield. To connect genetic variation in component traits to variation in yield, several studies have utilized crop models. Crop models integrate underlying physiological processes to compute yield. Thus, incorporating genetic variation in component traits (or model parameters) into crop models enables breeders to evaluate high-yielding lines via component traits. This thesis aimed to (1) assess which traits share their genetic basis with fruit yield in tomato, (2) identify component traits and QTLs that contribute to improving the yield of the elite parent by introducing wild germplasm, and (3) demonstrate the value of combining yield dissection models with QTLs by using identity-by-descent (IBD) probabilities. A more detailed background of the study with references to the literature is provided in **Chapter 1**.

In **Chapter 2**, two types of yield dissection models (Harvest dissection and Biomass dissection) have been employed to calculate the yield from two aspects. The first aspect is the fruit-sink capacity; in the harvest dissection model, yield is the product of the number of fruits (FN) and the individual fruit fresh weight (FFW). The second aspect is biomass production; in the biomass dissection model, yield is the product of the fruit fresh-dry weight ratio and total fruit dry weight. Total fruit dry weight is the product of total plant biomass and harvest index. Chapter 2 aimed to investigate the utility of both yield dissection models in tomato for identifying promising yield components and corresponding QTLs. We used a tomato breeding population, which consisted of hybrids obtained from crossing 342 recombinant inbred lines (RILs) derived from four parents with two testers, grown in a commercial greenhouse in the Netherlands. In the harvest dissection model, FN and FFW showed a trade-off (negative correlation between FN and FFW; mostly different QTL parental allele effects between FN QTLs and FFW QTLs at the same position). Therefore, FN QTLs and FFW QTLs need to be carefully used by confirming QTL effects on both traits to breed high-yielding lines. In the biomass dissection model, the fruit fresh-dry weight ratio and total fruit dry weight showed no correlation with each other. Thus, it is relatively easy to

improve both traits simultaneously. Total fruit dry weight was further dissected into total biomass and harvest index. Total biomass and harvest index showed a trade-off. Because total biomass correlated with yield, but the harvest index did not, total biomass is a more important target for yield improvement in this population than the harvest index. For lines with a higher yield than the elite parents, this was attributed to the number of fruits than individual fruit fresh weight, more attributed to total fruit dry weight than fruit fresh-dry weight ratio, and more attributed to total biomass than harvest index. Yield prediction via yield dissection models using QTLs for component traits achieved the equivalent prediction accuracy as yield prediction by QTLs for yield itself. This showed that the QTL allele effects of component traits were transformed into the effects on yield via yield dissection models. The yield dissection model showed how the introgression of wild germplasm created genetic variation in component traits that caused variation in yield, and how these variations were underpinned by QTLs. By using these results, plant breeders can design high-yielding lines with the preferred composition of component traits by accumulating positive alleles for component traits.

In **Chapter 3**, the methodology combining yield dissection models with QTLs as demonstrated in Chapter 2 was validated by using another breeding population, which consisted of hybrids obtained from crossing 346 recombinant inbred lines (RILs) derived from three parents with two testers, in two greenhouses in Spain. Chapter 3 aimed to (1) identify component traits and QTLs that contribute to improving yield of the elite parent in Spanish trials and validate the methodology combining yield dissection models with QTLs, (2) compare the breeding strategy for high yield between Spanish and Dutch trials and, (3) investigate whether some component traits show less G×E than yield by comparing two greenhouses in Spain. Around half of RILs showed a higher yield than the elite parent. Thus, introgression of wild germplasm brought much more genetic gain for yield in the Spanish trial compared to the Dutch trial. In the harvest dissection model, FN and FFW showed a trade-off similar to the Dutch trial. In the biomass dissection model, the fruit fresh-dry weight ratio did not correlate with yield in Spain, whereas a positive correlation was found in the Dutch trial. Nevertheless, the strategy to breed a high-yielding line will be similar between Spanish and Dutch trials; yield improvements can be more attributed to the number of fruits than individual fruit fresh weight, more attributed to total fruit dry weight than fruit fresh-dry weight ratio, and more attributed to total biomass than harvest index. Despite different populations and environments, some QTLs were at the same positions for Spanish and Dutch trials, which are more likely to be true QTLs. Some component traits showed less G×E than yield, therefore, these component traits can be more stable for selections over years. Yield prediction via the harvest dissection model using QTLs for component traits showed a similar accuracy as yield prediction by QTLs for yield itself. Therefore, the QTL allele effects of component traits can be transformed into the effects on yield via yield dissection models. Wild germplasm brought a large improvement in component traits and yield. Plant breeders

can design high-yielding lines with preferred characteristics of component traits by accumulating positive alleles from elite as well as wild parents.

In both Dutch and Spanish trials, total biomass is the important determinant for high-yielding lines. In **Chapter 4**, total biomass was further dissected into the adapted light use efficiency (LUE_{adp}) and the fraction of intercepted light (F_{int}) in the same trial as Chapter 2 in the Netherlands. F_{int} was calculated from leaf area index, and LUE_{adp} was calculated as total biomass divided by F_{int}. Chapter 4 aimed to (1) assess whether LUE_{adp} or F_{int} share their genetic basis with total biomass and yield, and (2) determine how the detected QTLs for LUE_{adp} and F_{int} can contribute to the selection for high-yielding genotypes. LUE_{adp} and F_{int} showed no correlation with each other, and most of their QTLs were unlinked. Therefore, high-biomass lines can be produced by improving both LUE_{adp} and F_{int} or each one of them separately. However, for high-yielding breeding, using LUE_{adp} is a more certain strategy than using F_{int} because of the negative correlation between F_{int} and harvest index. This chapter demonstrates that wild parents have a large potential to improve the total biomass of the elite parent. Some of the high total-biomass lines achieved a higher yield than the elite parent.

LUE_{adp} is thought to be primarily determined by photosynthesis. In **Chapter 5**, QTLs for photosynthesis-related traits were examined in the same trial as in Chapter 2 and Chapter 4 in the Netherlands. Chapter 5 aimed to (1) find QTLs for photosynthesis-related traits, and (2) assess how QTLs for photosynthesis-related traits contribute to yield. None of the photosynthesis-related traits correlated with yield, total biomass, or LUE_{adp}. On the other hand, photosynthesis-related QTLs were often at the same position as QTLs for yield, total biomass and/or LUE_{adp}. However, the QTL allele effects were not always consistent between photosynthesis-related traits and yield, which may explain the absence of correlation between photosynthesis-related traits and yield. Nevertheless, yield was predicted by QTLs for all photosynthesis-related traits with high prediction accuracy. Therefore, photosynthesis-related QTLs can be an invaluable tool for high-yielding breeding.

For further improvement of yield, introducing wild germplasm into elite cultivars is key. This thesis demonstrates that wild genetic resources possess beneficial alleles to improve the yield of the elite parent. Yield dissection models with QTLs provide a framework for evaluating the effect of introgressed wild germplasm on component traits and subsequently on yield. In **Chapter 6**, I discuss the usefulness of this methodology of combining yield dissection models with QTLs and compare my results with the literature. I recommend plant breeders to analyse both the harvest dissection (fruit-sink capacity) and the biomass dissection (source strength) to determine what is the limiting trait to improve the yield of the elite parent. I demonstrate that by using yield dissection models with QTLs, plant breeders can design high-yielding lines with the preferred composition of component traits (phenotypic ideotype) by assembling positive alleles for component traits from the elite as well as wild parents (genotypic ideotype).

Acknowledgement

I thank all people involved in this PhD project. There were a lot of challenges. Without help and support from everyone, I could not have finalized this thesis.

First of all, I thank my supervisors for bringing me to this PhD journey. When I applied for this project, I thought that this topic seemed very difficult. I was worried about whether I could manage all phenotyping and learning all statistics things. Thanks to all your support, finally I finished this project. Thank you so much. Fred van Eeuwijk, my promoter and daily supervisor. Thank you very much for giving me a lot of opportunities to learn statistics, statistical genetics and breeding. Your smart and creative idea is fundamental for this project, and it was fun to conduct this research. You gave me essential advice for presentation and writing, which gave me a lot of insights. It is very obvious that this project would not have existed without you. It is a kind of mystery in my life, how I could have gotten you as a promoter and supervisor, although I was not a statistician or mathematician. You opened several new doors for me, which I did not know at all before this project. Now I'm wondering how to use all this useful knowledge for my next career, but anyway, it is very certain that I learnt quite many things from you. Thank you so much. Leo Marcelis, my promoter. Thank you for your advice on how to make a comprehensible presentation and writing to convey scientific knowledge to a large audience. Thank you so much for the very good organization of Horticulture and Product Physiology (HPP) group and for always taking care of our working conditions and making efforts to improve our situation. Ep Heuvelink, my daily supervisor. You have a very profound knowledge of plant physiology and gave me a lot of advice and opinions from the start till the end of the project. It became a long journey. Still, you are always kind and took care of my progress. Thank you so much for encouraging me all the time. George, thank you for always being worried about my progress. Thank you for your insights from system analysis and mathematics.

This project was funded by Dutch STW TKI-U Program 'Green Genetics' project nr. 14524, BASF's vegetable seeds business (Nunhems) and Adviesbureau JFH Snel. Thank you so much for this opportunity and all your support.

I thank our user committee, Mirka Macel, Henry van der Valk, and Brigitte van Dijk from NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek), Frank Millenaar from Nunhems, Jan Snel from Adviesbureau JFH Snel, Bertrand Schuiling from Wiersum Plant Breeding, and Piet Arts from Barenbrug. Thank you all for all the insightful discussions during STW project meetings. It was fun to know each company's perspective. Also thank you for always encouraging me.

Special thanks to Nunhems for the large support for this project. Without cooperation with Nunhems, of course, this project would not have existed. This was a quite unique opportunity

to join the pre-breeding trial of 700 genotypes in a 7000 m^2 testing area in a 6 ha commercial greenhouse. Nunhems developed these breeding materials, provided marker information, and conducted phenotyping of yield. Phenotyping yield was a full-time job for 7 months. Thus, without Nunhems, a PhD candidate could have measured only yield. Frank, thank you so much for always kindly responding to our questions and requests. Kim Vermeer, thank you for all your collaboration and support during phenotyping. I thank Hannelore van Stappen-van Dijk and also other people in Nunhems for developing population and marker information.

Special thanks to Jan Snel for all your support during chlorophyll fluorescence measurements by HexPAM: setting up Hex-PAMs, developing the measurement protocol and script, conducting maintenance and repairment of Hex-PAMs. Without your support and help, chlorophyll fluorescence measurements could have never been done.

Special thanks to Mahboobeh (Sedighehsadat Khaleghi). During your stay in Wageningen, you joined my project. We worked very hard during phenotyping. It is very obvious that I could not have finished all measurements without you and Daisuke. You are always very motivated, eager to learn new things, and keep encouraging me. Thank you so much! I wish you all the best.

Thanks to student assistants in this project. Csongor Szanto, thank you for the photosynthesis measurements during the weekend. You are very motivated and very accurate for gas-exchange measurement by LiCor. You noticed very soon if something was wrong with the machine. Thank you so much! Oscar Castellanos, thank you for joining this project after Mahboobeh went back to her country. Your measurements were always precise. Alejandro Bustamante, actually you are my PhD colleague and friend. Thank you for joining the validation experiment. Since you were also good at driving, you made my life much easier. Thank you! Carmen Saez Cuesta, thank you for joining the Spanish trial. Without you, it would never have been possible to do some measurements in the Spanish trial. You're from the region, you know everything in the region, you communicated with growers very well, and your measurements were very precise. It's not an easy job at all, but you managed everything. Thank you so much! Thank you so much also to your family for kindly letting me stay at your family's place. We ate very tasty food every day. At that time, I was pregnant and somehow got a cold. All your family were so kind and took care of me very much. Thank you very much. I remember, one day, your father played the guitar with his friends. Then our daughter danced a lot in my belly;) I also thank the people of Nunhems in Spain and the University of Almeria for kindly helping with Carmen's measurements.

Thanks to De Bakker Westland: Roel, Bart and Frans de Bakker and Jim Zwinkels. You were always kind and supportive during our phenotyping. You kindly made additional desks for us to use on the car in the greenhouse, and also a supporting device for photosynthesis measurements by LiCor. Sometimes we had to continue the measurements until late at night,

but you never complained. We sometimes caused inconvenient situations for you (*e.g.,* we used three cars; you had to wait for us to finish stem harvesting before you cleaned the greenhouse), but you were always kind. Thank you so much! I will also never forget the tomatoes from your greenhouse. They were very tasty!

Thanks to Unifarm: Wim, Herman and others. We sometimes caused difficult situations for Unifarm (because our samples occupied many ovens), but you were always kind and flexible. Thank you so much for all your support and encouragement!

Thanks to Barenbrug: Jos Jager and Piet Arts. In 2019, the tomato virus pandemic happened in Europe. We had a validation experiment, but we were not allowed to bring any external tomato plants (regardless of virus contamination) to the university. Thanks to your kind offer, we used your oven and managed the dry weight measurement. Thank you so much for all your support. (I have to consider how to manage to write the paper about the validation experiment. Unfortunately, it is out of this thesis, but we know the results are good.)

Faline, sitting at the next desk to my empty desk during phenotyping, thank you for being worried about me during the phenotyping experiment. At the beginning of phenotyping, you helped with our measurement. It is a nice tradition in Horticulture and Product Physiology (HPP) to help each other's measurements. But I realized that it was not realistic to ask colleagues to help with measurements in my project's case because we had hundreds of measurements every day. But your kindness made my heart warm. Thank you!

Piri, my ex-housemate. Thank you very much for letting me stay at your shared house in Delft sometimes. Because I did not like driving, staying at your place saved my life. Thank you so much for your kindness always.

Thanks to colleagues working on photosynthesis. Elias, Wim and Jeremy, thank you for your kind advice and warning during photosynthesis measurements. Maarten, thank you for your support and repairment of LiCor. Thanks to Arjen for teaching me how to extract chlorophyll and measure chlorophyll content.

Thanks to the Centre for Crop Systems Analysis (CSA) group, Peter van der Putten for letting me borrow a chlorophyll meter (SPAD). We wanted to check the chlorophyll level during chlorophyll-fluorescence measurements because chlorophyll level might be associated with chlorophyll fluorescence. In the end, unexpectedly, we got a good result from the chlorophyll content index measurement (SPAD measurement) itself. I'm sorry for having caused inconvenient situations in CSA group sometimes, and thank you so much for your kindness.

Until here, I hope I manage to mention all people involved in and helping me during the phenotyping experiment. Now I move on to the analysis.

Special thanks to Martin Boer for teaching me how to set up the mixed model equation for correction for the spatial trend of phenotypic raw data by SpATS. Because the experimental design was complicated, it was not an easy job at all. Also thank you for all your advice about QTL analysis. I was not a breeding student and I sometimes had questions. You are always very kind and answered my questions. It was very important for me that I could meet you in Radix.

Special thanks to Daniela Bustos-Korts. You selected RILs for the subset to represent the genetic diversity in the population for my project. In addition, your insight from a plant breeding perspective helped me significantly with understanding this subject. You found positive points in my work and gave me constructive opinions. This helped me tremendously. Thank you so much for all your intelligence and all your kindness.

Thanks to Chaozhi Zheng for the calculation of Identity-by-descent (IBD) probabilities. This was not possible for me to conduct by myself. (Now I learned how to use your software "RABBIT", so I may be able to do it by myself using RABBIT?) Thank you also for your kind attitude and for explaining these difficult things to people including me, who do not know so much about mathematics.

Thanks to Wenhao Li, my PhD friend, for teaching me how to conduct QTL analysis by the mixed model with IBD. I started PhD earlier than you, but I spent the first year only on phenotyping and then spent all the time at the university together with you; before corona, during the lockdown, and after the lockdown until now. It was very difficult for us during the lockdown, but I'm very happy that both of us finally finish PhD together. And you are the paranymph of my defence. Thanks to Bader Arouisse, my PhD friend, for checking whether my cross-validation was correct before the submission of my paper (Chapter 2). It was under lockdown, so it was difficult to contact people. Although you were very busy, you are always very kind and checked my calculation for us. Thank you so much. Thanks to Julio Velazco, my PhD friend, for considering together about correction for spatial trends. Thanks to Marcos Malosetti for teaching me how to develop a multi-QTL model. Of course, it was not possible for me to develop it by myself. Thanks to Pariya Behrouzi for teaching me partial correlation and how to use your R-package "netgwas" for partial correlation and QTL analysis. Although I did not use this analysis in this thesis, I may use it in future. Thanks to Antoine Languillaume for teaching me how to include a kinship matrix in QTL analysis. In the end, I did not include the kinship matrix in the QTL model, but it helped me to understand QTL analysis.

Thanks to PhD friends in statistical genetics in Biometris. My neighbour, Georgios, Wenhao and Bader. Before corona, I spent one day per week in Biometris. I enjoyed very much eating lunch together with you on Wednesday. Your talks were very funny. I sometimes asked questions about statistics and you are all very smart and gave me useful information. Thanks

to Vincent and Stefan for helping me with R in the initial phase. James, George, Kevin, Henry, Gavin, Sanne, Tryntje, and Diana. Thank you for all your kindness and help during my PhD.

Thanks to all my colleagues in Biometris. Initially, I felt very strange that I belong to a group of mathematicians and statisticians. I did not know so much about math or statistics. I belong to Biometris because I need help from mathematicians and statisticians. I learned many things that I had not known at all before I started this PhD project. You are very open, and I could join seminars and courses. I learned AI in the deep learning discussion group, which was also open for non-experts like me. Thanks to Manya and Willem for organizing the discussion group. Thanks to Dinie and Hanneke for organizing fun events. Thanks to Maikel for solving my computer problems. Thanks to Evert-Jan and Martin for practising Dutch with me. Dank jullie wel! Thanks to Daniela, Nadia, and Dominique for visiting us when our daughter was born. Thanks to Gerrit and Bas for teaching me the mixed models and other statistical models. Thanks to Bart-Jan, Jip, Emilie, Joao, Paul, Marco, another Marco, Alessio, (PhD) Bas, Willem, Sabine, Eva, Nienke, Onno, Zhu, Carolien, Jaap, Ron, Peter, another Peter, Hilko, Bert, Elly, Lia, Rene, Patrick, Gert, ... it is not possible to mention everyone, but thank you all for encouraging me all the time.

Thanks to Ying Liu, my friend from master study to PhD. You suggested joining the Solanaceae conference in 2019. I did not know about the conference. Thanks to you, I met almost all the famous people for tomato genome / QTL / genetic studies. I had a bit of a health issue at that time, but you and your husband saved my life. You always worked very hard and encouraged me. Thank you!

Thanks to all colleagues in the Horticulture and Product Physiology (HPP) group. Thank you for your supportive discussions during the FLOP discussion group. Thanks for sharing all tips for experiments and writing. Thanks to PV for organizing all the fun events. Thanks to the management support team, Pauline, Melanie, Katharina and Linda, for making our life easier. Thanks to all staff for supporting our activity. Thanks to all PhD friends and postdocs before corona and after the corona lockdown relaxed. This is very difficult to mention everyone; what should I do? I think I spent time together with more than 70 PhDs and postdocs in HPP. Each of us worked seriously, and at the same time helped each other, and enjoyed the time together. Thank you all for accompanying me. Thanks to Wannida, my other paranymph. Thank you for encouraging me all the time. It was fun chatting in Japanese sometimes. Your smile always brings me a heartwarming atmosphere.

I also thank SKiPR, Sharing Knowledge in Photosynthesis Research group, across different research groups at Wageningen university. I was not an expert in photosynthesis, but I had to manage photosynthesis measurements and analyses. I could not join any conferences or courses about photosynthesis, but I managed photosynthesis measurements and analyses (Chapter 5) thanks to SKiPR, also several seminars organized by HPP and other groups. I had occasions in Wageningen to meet many photosynthesis researchers from Wageningen

and also from abroad. This made me possible to consider what to measure and how to analyze. Without this interaction with other researchers, it would have been much more difficult to conduct the research on photosynthesis.

Therefore, I also thank Wageningen University for this opportunity. Interaction among different research disciplines, among different stakeholders, in the Netherlands also among different countries.

I also thank my friends outside of my research. Friends from the Japanese community in Wageningen, Etsuko, Reiko, Mariko and all others. Friends from Farm Technology group, a previous neighbour of HPP. Friends from my master's study. Friends from our ex-housemates. Friends related to our daughter. Thanks to Saskia for your music class for kids. Jelle, dank je voor uw nederlands les. During the lockdown, sometimes I was wondering "which day is today?". During the partial lockdown, we could not go to the office, but your music or Dutch lessons was allowed under certain conditions. It gave me the feeling of real life. Veel bedankt.

Dank je wel Kees, onze huisbaas. Je bent altijd aardig en zorgt voor ons. We vergeten nooit de dagen bij jou thuis in Wageningen.

Last but not least, I thank all my family in Japan for letting me come to the Netherlands to study master's and then continue PhD. Thank you for your positive attitude toward my study and research, and for all your support. Our daughter, thank you for being born to us. You may think I go to the university for fun (this may be true?), but this was also serious work for me. Please eat tomatoes. Daisuke, thank you for accompanying me, a lot of actual work for phenotyping, then a lot of work for childcare. Being a partner of a PhD candidate might be a more difficult job than doing a PhD. Thank you so much.

About the author

Yutaka Tsutsumi-Morita was born in a parttime farmer family in the countryside of Himeji city in Japan in 1980. Family farming and surrounding nature stimulated her interest in plants, agriculture and ecology. Because her parents worked full-time, her grandmother, *Tomiko*, mostly took care of her and her brothers. Tomiko's bright insights into nature and people had a huge influence on Yutaka. Although Tomiko's cheerful and thoughtful characteristics always brought happiness to everyone, most of Tomiko's life was as a poor small farmer until Yutaka's father started working. This made Yutaka think about why most farmers are small and poor, and why it is difficult to make a profit from farming.

Yutaka proceeded to The University of Tokyo. She studied mainly agriculture. During her bachelor's study, she thought that molecular biology was at the forefront of science. Her bachelor's and master's research was about "Characterization of *OsPID*, the rice ortholog of *PINOID*, and its possible involvement in the control of polar auxin transport". Although the research was published, she felt that research in molecular biology is far from something useful for farmers. She understood that many researchers made tremendous efforts, and finally everyone's work together must be useful for society. But it seemed very far to go. She wanted to do something more directly useful for farmers.

Although it was difficult to find a job in agriculture at that time in Japan, she got a job in a company producing micro-vegetables in greenhouses and a plant factory. Her role was to improve growing systems, develop growing systems for new products, control quality, and control hygiene practices for food safety. It was very fun for her to see how plants respond to changes in environment or management. During her work, she had an opportunity to stay in the Netherlands for a month in a collaborative company producing micro-vegetables. She was very impressed by Dutch horticulture. Farming was conducted based on science. The climate was automatically controlled, natural enemies were introduced in greenhouses, and consultants gave advice based on science. It seems that the greenhouse and agricultural sectors achieved the position of an industry, similar to other industries. This made her want to learn horticulture and agriculture in the Netherlands.

A few years later, she decided to study master's program at Wageningen University in the Netherlands. She studied crop science and greenhouse horticulture. She noticed that crop modelling and all kinds of modelling are the keys to applying scientific knowledge to practice. After finishing her master's program, the perfect opportunity came. The PhD project "Yield

prediction based on QTLs for yield components using crop ecophysiological models". The results of her PhD research are described in this thesis. Yutaka is now thinking about how to contribute back to society by using her knowledge and ability, which has been gained thanks to support from many people.

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)

Review of literature (4.5 ECTS)

- Yield prediction by combining crop models with quantitative genetics in FLOP discussion group

Post-graduate courses (7.2 ECTS)

- Introduction to R for statistical analysis; PE&RC and SENSE (2016)
- Dynamic models in R: programming, parameter estimation and model selection; PE&RC and SENSE (2017)
- Statistical analysis for new phenotyping techniques; WUR (2018)
- Structural equation modelling; PE&RC and SENSE (2018)
- Mixed linear models; PE&RC and SENSE (2018)
- Crop physiology and climate change; WUR and University of Florida (2019)

Deficiency, refresh, brush-up courses (9 ECTS)

- R for statistics; Biometris (2017)
- Modern statistics for the life sciences; Animal Breeding and Genetics (2017)

Competence strengthening / skills courses (6.9 ECTS)

- Reviewing a scientific paper; WGS (2016)
- The essentials of scientific writing and presentation; Wageningen in'to Languages (2017)
- Ethics in plant and environmental sciences; PE&RC and others (2018)
- Scientific integrity; PE&RC and others (2018)
- Efficient writing strategies; Wageningen in'to Languages (2018)
- Brain friendly working and writing; PE&RC and others (2018)
- Scientific writing; Wageningen in'to Languages (2019)
- Posters and pitching; Wageningen in'to Languages (2019)
- Critical thinking and argumentation; WGS (2020)

PE&RC Annual meetings, seminars and the PE&RC weekend (1.8 ECTS)

- PE&RC First years weekend (2016)
- PE&RC Day (2016)
- WGS PhD Workshop Carousel (2018, 2019)

Discussion groups / local seminars or scientific meetings (9 ECTS)

- FLOP: Frontier Literature in Plant Physiology (2016-2021)
- Life congress; poster presentation (2019)
- Plant breeding and biotechnology symposium; oral and poster presentation (2019)

International symposia, workshops and conferences (10.2 ECTS)

- The 17th meeting of the EUCARPIA; Biometrics in Plant Breeding (2018)
- Breeding data; DuPont Plant Sciences (2018)
- Greensys (2019)
- Solanaceae conference (2019)
- iCROPM (2020)

This research was part of a 'TKI-U Green Genetics' project where Wageningen University cooperated with BASF's vegetable seeds business (Nunhems), Adviesbureau JFH Snel, Barenbrug, and Wiersum Plant Breeding.

This project was financially supported by the Netherlands Organization for Scientific Research (NWO), BASF's vegetable seeds business (Nunhems), and Adviesbureau JFH Snel.

Cover design and photo

By Yutaka Tsutsumi-Morita

Genotype produces phenotype under influence of the environment. The image on the back is an example of Identity-by-Descent (IBD) probabilities. Recombinant inbred lines (RILs) of a breeding population are vertically lined. Chromosome and genetic positions are horizontally lined. The genome of three parents (represented by the colour light blue, pink and orange) are introgressed into a genetic background of a recurrent parent (represented by the colour blue). The central picture shows the phenotyping experiment in the Netherlands in this thesis. The picture captures one path, which was one-sixteenth of the whole experimental area.

Printed

By ProefschriftMaken.nl