



INSTITUT DE FRANCE
Académie des sciences

Comptes Rendus

Biologies

Mathilde Causse, Juliette Bénéjam, Estelle Bineau, Frédérique Bitton, Marie Brault, Yolande Carretero, Henri Desaint, Alexandre Hereil, Karine Pellegrino, Esther Pelpoir and Jiantao Zhao

Genetic control of tomato fruit quality: from QTL mapping to Genome Wide Association studies and breeding

Volume 345, issue 4 (2022), p. 3-13


Published online: 24 November 2022

Issue date: 16 February 2023

<https://doi.org/10.5802/crbio.99>

Part of Special Issue: Insights onto Plant Biology

Guest editors: Michel Delseny (Laboratoire Génome et Développement des plantes, UMR5096 CNRS-Université de Perpignan Via Domitia, Université de Perpignan 66860 Perpignan, France) and Georges Pelletier (Directeur de recherche émérite à l'Institut national de la recherche agronomique, Versailles, France)

 This article is licensed under the
CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE.
<http://creativecommons.org/licenses/by/4.0/>



*Les Comptes Rendus. Biologies sont membres du
Centre Mersenne pour l'édition scientifique ouverte*

www.centre-mersenne.org

e-ISSN : 1768-3238



Insights onto Plant Biology / *Regards sur la biologie végétale*

Genetic control of tomato fruit quality: from QTL mapping to Genome Wide Association studies and breeding

Contrôle génétique de la qualité des fruits de tomate : de la cartographie de QTL aux études d'associations et à la sélection

Mathilde Causse^{®*, a}, Juliette Bénéjam^{® a}, Estelle Bineau^a, Frédérique Bitton^{® a},
Marie Brault^{® a}, Yolande Carretero^a, Henri Desaint^{® a}, Alexandre Hereil^a,
Karine Pellegrino^a, Esther Pelpoir^a and Jiantao Zhao^{® a}

^a INRAE, UR1052, Genetics and Breeding of Fruit and Vegetables, 67 Allée des Chênes,
Centre de Recherche PACA, Domaine Saint Maurice, CS60094, 84143 Montfavet,
France

E-mails: mathilde.causse@inrae.fr (M. Causse), juliette.benejam@inrae.fr
(J. Bénéjam), estelle.bineau@gautiersemences.com (E. Bineau),
frederique.bitton@inrae.fr (F. Bitton), marie.brault@inrae.fr (M. Brault),
yolande.carretero@inrae.fr (Y. Carretero), henridesaint@gmail.com (H. Desaint),
alexandre.hereil@inrae.fr (A. Hereil), karine.pellegrino@inrae.fr (K. Pellegrino),
esther.pelpoir@inrae.fr (E. Pelpoir), jz426@cornell.edu (J. Zhao)

Abstract. Consumers began to complain about the taste of tomato varieties in the late 1990's. Although tomato taste is influenced by environmental and post-harvest conditions, varieties show a large diversity for fruit quality traits. We herein review our past and present research work intended to improve tomato fruit quality. First, results from sensory analysis allowed identifying important traits for consumer preferences. Then, we dissected the genetic control of flavor related traits by mapping several QTL in the last 20 years, and identified the genes corresponding to a few major QTL. Since the availability of the tomato genome sequence, genome-wide association studies were performed on several panels of tomato accessions. We discovered a large number of associations for fruit composition and identified relevant allele combinations for breeding. We then performed a meta-analysis combining the results of several studies. We also checked the inheritance of quality traits at the hybrid level and assessed how genomic prediction could help selecting better tomato varieties.

Résumé. Les consommateurs ont commencé à se plaindre du goût des variétés de tomates à la fin des années 1990. Bien que le goût de la tomate soit influencé par les conditions de culture et de post-récolte, les variétés présentent une grande diversité pour les caractéristiques de qualité des fruits. Nous passons ici en revue nos travaux de recherche passés et présents destinés à comprendre la diversité génétique et améliorer la qualité des fruits de tomate. Les résultats d'analyses sensorielles

* Corresponding author.

ont tout d'abord permis d'identifier les traits importants pour les préférences des consommateurs. Ensuite, nous avons disséqué le contrôle génétique des caractères liés à ces traits, cartographié de nombreux QTL depuis 20 ans et identifié les gènes correspondant à quelques QTL majeurs. Depuis la disponibilité de la séquence du génome de la tomate, des études d'association à l'échelle du génome ont été réalisées sur plusieurs panels d'accessions de tomate. Nous avons découvert un grand nombre d'associations pour la composition des fruits, et identifié les combinaisons d'allèles pertinentes pour la sélection. Nous avons ensuite réalisé une méta-analyse combinant les résultats de plusieurs études. Nous avons également vérifié l'hérédité des caractères de qualité au niveau hybride, et évalué comment la prédiction génomique pourrait aider à sélectionner de meilleures variétés de tomates.

Keywords. Tomato, Fruit quality, Breeding, Genome-wide association, Genomic selection.

Mots-clés. Tomate, Qualité des fruits, Sélection, Génétique d'association, Génétique quantitative.

Published online: 24 November 2022, Issue date: 16 February 2023

La version française de l'article est disponible à la suite de la version anglaise

1. Introduction

Tomato flavor is a key issue for tomato breeders. Over the last century, breeders have strongly improved tomato yield, yield stability and plant adaptation to diverse growth conditions. They have introgressed many disease resistance genes from wild tomato relatives. Fruit quality has been improved mainly for fruit shelf life and fruit homogeneity, but consumers started complaining about tomato flavor in the 1990's. Although diversification in size, shape and color of fruits is proposed on the market, taste is still often criticised. Improving tomato taste is complex for several reasons:

- Sensory quality is a composite characteristic involving many components: sweetness and sourness (related to sugar and acid contents), aromas (involving several volatile organic compounds) and texture (linked to firmness, meltiness, mealiness).
- The measurement of these components may be difficult, as some are only measured by sensory analyses, but most of the components can be related to fruit composition. Hundreds of volatile organic compounds (VOCs) have been identified but the list of those important in tomato aroma is limited to around 30 VOCs [1,2].
- Most of these components are influenced by the environmental conditions during plant and fruit development [3], by the harvest stage but also by postharvest conditions [4].
- Some important components are negatively correlated together, like fruit size and sugar

content or fruit shelf life and meltiness, hampering the simultaneous improvement on both traits.

- Quality has a subjective component as every consumer preferences [5, 6].
- Quality is only important at the last stage of the production chain, and many actors participate and can damage the flavor of a variety; furthermore, economic factors restrain quality valorisation, as quality is not directly valorized on the market, making it difficult to request higher price for better varieties [7].

Nevertheless in tomato cultivars, there is a large genetic variability for fruit quality [8, 9] and the variety choice is the first step in the chain impacting tomato flavor. The advance in molecular markers, as well as the recent availability of the tomato genome sequence [10] and of many resequenced accessions [11] have paved the way towards a better understanding of genetic factors involved in fruit quality. Herein we will present the most recent advances to identify the genetic control of fruit quality parameters, and how the results can be used for genomic selection.

2. Identification of consumer expectations and major traits

Several experiments were performed to identify the expectations of consumers facing the diversity of tomato varieties. We first showed that both flavor

and texture traits are important in consumer preferences [3], and we detected the potential of hybrids between traditional varieties and modern lines, probably to reach an intermediate fruit firmness and juiciness. In Sinesio *et al.* [12], a sensory panel and a large panel of consumers from France, Netherland and Italy tested 16 tomato cultivars grown in the 3 countries. Four clusters of consumers were identified in each country with different descriptors corresponding to their preferences. The same segments were identified in each country. This study underlined the importance to offer different segments (more or less tasty and firm) on the market. Finally Sinesio *et al.* [6] showed that for a specific type of tomato (cherry, large or long fruit) both traditional and modern varieties could be appreciated by consumers, without a specific trend for one or another group of varieties.

3. QTL and GWAS to unravel the genetic control of fruit quality traits

Most tomato fruit quality traits are quantitatively inherited. Since the first QTL studies in the 1990's, several QTL mapping experiments were performed in order to identify loci to be used for marker-assisted selection [13, 14]. A few major genes involved in fruit weight, fruit shape and composition have been cloned [15]. Martina *et al.* [16] built a consensus map that summarizes QTLs reported for tomato aroma from 16 studies. The authors defined genomic regions where QTL partially overlap for similar metabolic pathways between studies, and looked into lists of genes within the regions of interest, as a way of harnessing the present knowledge without missing causal variants that may fall out of the confidence interval of a single QTL identified in only one study.

With the sequence and SNP availability, genome-wide association experiments then offered another approach. Sauvage *et al.* [17] genotyped 163 tomato accessions composed of large-fruit, cherry and wild tomato accessions with the SolCAP SNP array, generating a total of 5995 high quality SNPs. Then they performed GWAS using a multi-locus mixed model (MLMM [18]) for 36 fruit metabolites that were highly correlated between the two growth periods and identified 44 candidate loci associated for different fruit metabolites [17]. Among the candidate loci, they

identified a gene with unknown function on chromosome 6 that was strongly associated with malate content. This association was further identified in different GWAS and meta-analysis of GWAS based on different populations [19–22] and was further validated as an *Al-Activated Malate Transporter 9* (*Sl-ALMT9*) [21].

Bauchet *et al.* [19] genotyped 300 tomato accessions with both the SolCAP and CBSG arrays, generating a total of 11,012 high quality SNPs, which were used for GWAS using both MLMM and multi-trait mixed model (MTMM) [23]. A total of 79 significant associations were identified for 13 primary metabolites and 19 VOCs in tomato fruits. Among these, two associations involving fruit acidity and phenylpropanoid content were particularly investigated [19]. The same population was also characterized for agronomic traits and many QTLs were identified, such as *fw2.2* and *fw3.2* for fruit weight. GWAS for similar quality traits were also performed in other collections [24, 25]. With the fast development of whole-genome-sequencing technology and the reduction of cost per genome, it is possible to sequence hundreds of diverse tomato collections. For instance, Tieman *et al.* [20] sequenced 231 new accessions and combined these data with 245 previously sequenced genomes, generating a total of 476 genome sequences. These data were then used for GWAS for diverse flavor-related metabolites, including 27 VOCs, total soluble solids, glucose, fructose, citric acid, and malic acid. A total of 251 significant associations were detected for 20 traits. Two loci were significantly associated with both glucose and fructose, corresponding to two major QTL *Lin5* and *SSC11.1*. By combining with selection analysis, it was further shown that the negative correlation between sugar content and fruit weight was likely caused by the loss of high-sugar alleles during domestication and improvement towards larger tomato fruits [20]. In addition, some good candidate genes involved in tomato volatile contents were also identified, such as Solyc09g089580 for guaiacol and methylsalicylate. By combining the three significant associated loci for geranylacetone and 6-methyl-5-hepten-2-one, it was shown that several allelic combinations conferring favorable aromas were progressively lost during domestication and breeding [20]. A few other QTLs for volatiles were identified following positional cloning [26, 27].

4. Meta-analysis of GWAS

However, with the results of several GWAS in tomato for the same trait, only some significant associations could be identified consistently across different studies, indicating strong cross-study heterogeneity, which refers to the non-random variance in the genetic effects between different GWASs. The main sources of heterogeneity include population structure, linkage disequilibrium, phenotyping measurement methods, environmental factors, genotyping methods, $G \times E$ interactions. Meta-analysis of GWAS is a new approach to combine different GWAS studies, while properly handling the heterogeneity. For instance Zhao *et al.* [25] reported the meta-analysis of GWAS from three tomato GWAS populations [17, 19, 20]. Following genotype imputation, a total of 775 tomato accessions and 2,316,117 SNPs were used in the meta-analysis, and a total of 305 significant associations were identified for sugars, organic acids, amino acids and flavor-related volatiles. By looking at the five loci associated with both fructose and glucose, they showed that sugar content significantly increased with the number of wild alleles. The authors also demonstrated that domestication and improvement have had an impact on citrate and malate content. In particular, the major QTL *Al-Activated Malate Transporter 9* of malate was significantly associated with citrate and another malate transporter was identified for citrate content on chromosome 1. This study also identified many new significant associations for flavor-related volatiles. By targeting six significant associations, it was further demonstrated that modern tomato accessions had a limited flavor due to a lower content of pleasant volatiles but also a higher content of unpleasant volatiles compared to cherry tomatoes [22].

5. GWAS in hybrids

As most of the modern varieties are F1 hybrids, but GWAS analyses were performed on lines, Bineau *et al.* [28] extended the GWAS approach to hybrid genomes. The authors quantified 46 VOCs in a GWAS panel of 121 small fruited lines—cherry type and wild relative *S. pimpinellifolium* accessions—and in a test cross panel of 165 F1 hybrids (the previous panel plus 44 elite cherry tomato lines crossed with a common big fruited tester line). They performed GWAS on the

two panels separately, along with a third GWAS on the test cross subset carrying only F1 hybrids corresponding to the line panel and identified several genomic regions that could help breeders recovering flavor in modern varieties, while accounting for fruit weight trade-offs that may be at stake for improvement of big fruited plant material.

6. Impact of environmental conditions on fruit quality

Tomato is grown all over the world in many different environments. Its yield and fruit quality is impacted by environmental conditions (temperature, light, irrigation, nutrition [29–31]). Over the last 10 years we have studied the impact of the three main stresses in Mediterranean conditions (water, salt and heat stress) in two panels of tomato varieties. We showed a large variability in each panel for most of the traits relative to plant architecture, phenology or fruit composition and also for their response to specific stresses. For instance a mild water stress may improve fruit sweetness but a stronger one will simultaneously reduce yield and fruit weight [32, 33]. Hot conditions hamper the pollination and thus impact strongly yield and accelerate flowering time [34]. The study of plant plasticity allowed to identify some accessions which were less impacted by the environment variations than others, but also to characterise the genetic control of plasticity components [31]. Plasticity QTLs were detected, but in distinct regions than the average phenotype QTLs for a given trait, suggesting the possibility of independent selection for plasticity. Candidate genes potentially involved in the occurrence of genotype by environment interaction in tomato have been proposed, paving the way to the functional characterization of stress response genes in tomato.

7. Genomic prediction

Both linkage mapping and GWAS have limitations in identifying and quantifying small effect loci, as well as rare QTLs or associations that are highly susceptible to environmental conditions [35]. In contrast, genomic selection (GS), which has been proposed about two decades ago [36] uses all the genetic information from markers spread over the

whole genome, such as SNPs and phenotypic data, in a training population, in order to predict the genetic estimated breeding values (GEBVs) of unphenotyped individuals in a test population. The main advantages of GS include cost reduction and time saving compared to phenotype-based selection [35].

Several factors influence the accuracy of genomic prediction (GP), including the size, structure and genetic diversity of the training population, trait heritability, the number and distribution of molecular markers, linkage disequilibrium, prediction method and number of QTLs [37–39]. In order to improve the prediction accuracy, complex GS models were developed in order to handle different factors, such as the multi-trait and multi-environment $G \times E$ interactions [40, 41]. Many models for GS are available and the prediction accuracy vary according to traits and conditions [38, 39, 42].

The first GS test in tomato was focused on a simulation-based breeding design and phenotypic prediction, where a theoretical method was proposed to apply GS to actual breeding schemes of simultaneous improvement of yield and flavor [38]. Briefly, 96 big-fruited tomato varieties were selected and 20 agronomic traits were measured, with broad-sense heritability ranging from 0.10 to 1.00. Seven GP models were compared. The highest prediction accuracy for different traits varied and the accuracy of Bayes C was highest for up to eight traits, ranking the best among all models. Some individuals with high GEBV of total fruit weight and soluble solid contents were selected as parents to simulate later generations. Simulations demonstrated that after five generations, the simulated GEBVs were comparable with parental varieties.

Yamamoto *et al.* [39] then used a big-fruited F_1 population to construct the GS models to assess its potential for the improvement of total fruit weight and soluble solid content in a practical experiment. By testing six GS models and 10-fold cross-validation, the prediction accuracy for soluble solid content was higher than for total fruit weight.

Duangjit *et al.* [37] investigated the impacts of some key factors on the efficiency of GP, including the size of training population, the number and density of SNPs and individual relatedness. Based on the analysis of 163 tomato accessions studied by Sauvage *et al.* [17], the optimal size of the training popula-

tion was 122. The prediction accuracy also increased with marker density and number, but weakly. Individual relatedness also influenced the prediction accuracy, and predictions were better in closer individual relatedness. However, there are some limitations in this study: (1) it only tested the ridge regression best linear unbiased prediction (rrBLUP) statistical model [43]; (2) the number of SNPs was relatively small and the genomic coverage in certain genomic regions was quite limited; (3) Population structure existed and the number of wild accessions was quite small compared to cherry and large-fruited tomato accessions.

Most GS models rely on marker-based information and are unable to exploit local epistatic interactions among markers. Molecular markers can also be combined into haplotypes by combining linkage disequilibrium and linkage analysis to improve prediction accuracy as shown in [44].

8. Conclusion

In the last twenty years, knowledge about fruit development and quality has strongly increased. Many data are available for breeding new tomato varieties with improved fruit taste. The need for the involvement of the whole production chain (from variety production to marketing) and economic recognition of fruit quality remain limiting factors. At the fundamental level many important genes may still be discovered. Genomic tools and data are also available to speed up their discovery in the near future.

Conflicts of interest

Authors have no conflict of interest to declare.

Acknowledgements

This research has been supported by many National and International research projects. Among those presently supporting our work, we can mention the projects H2020 “HARNESSTOM”, grant number No. 101000716, SUscrop Era-NET No 47 2019 “SOLNUE”, PRIMA 2018-11 within the project “VEGADAPT” and ANR “3C-FruitGrowth”.

Version française

1. Introduction

La saveur de la tomate est une question essentielle pour les sélectionneurs de tomates. Au cours du siècle dernier, les sélectionneurs ont fortement amélioré le rendement de la tomate, sa stabilité et l'adaptation de la plante à diverses conditions de croissance. Ils ont introduit de nombreux gènes de résistance aux maladies provenant de tomates sauvages apparentées. La qualité des fruits a été améliorée principalement pour leur durée de conservation et leur homogénéité, mais les consommateurs ont commencé à se plaindre du goût des tomates dans les années 1990. Bien que la diversification de la taille, de la forme et de la couleur des fruits soit proposée sur le marché, le goût est encore souvent critiqué. L'amélioration du goût des tomates est complexe pour plusieurs raisons :

- La qualité sensorielle est une caractéristique composite qui fait intervenir de nombreuses composantes : les saveurs sucrées et acides (liées aux teneurs en sucres et en acides), les arômes (impliquant plusieurs composés organiques volatils) et la texture (liée à la fermeté, au fondant, au caractère farineux) ;
- La mesure de ces composants peut être difficile, car certains ne sont mesurés que par des analyses sensorielles, mais plusieurs composants peuvent être liés à la composition du fruit. Des centaines de composés organiques volatils (COV) ont été identifiés, mais la liste de ceux qui sont importants pour l'arôme de la tomate se limite à environ 30 COV [1, 2].
- La plupart de ces composants sont influencés par les conditions environnementales pendant le développement de la plante et du fruit [3], par le stade de la récolte mais aussi par les conditions post-récolte [4].
- Certains composants importants sont négativement corrélés entre eux, comme le calibre du fruit et la teneur en sucres ou la durée de conservation du fruit et le fondant, ce qui empêche l'amélioration simultanée des deux caractéristiques.
- La qualité a une composante subjective comme toutes les préférences des consommateurs [5, 6].

- La qualité n'est importante qu'au dernier stade de la chaîne de production, et de nombreux acteurs y participent et peuvent nuire à la qualité gustative d'une variété; en outre, les facteurs économiques freinent la valorisation de la qualité, car celle-ci n'est pas directement valorisée sur le marché, ce qui rend difficile de demander un prix plus élevé pour de meilleures variétés [7].

Néanmoins, dans les cultivars de tomates, il existe une grande variabilité génétique pour la qualité des fruits [8, 9] et le choix de la variété est la première étape de la chaîne d'impact sur la saveur de la tomate. Les progrès du marquage moléculaire, ainsi que la disponibilité de la séquence du génome de la tomate [10] et de nombreuses accessions reséquencées [11] ont ouvert la voie à une meilleure compréhension des facteurs génétiques impliqués dans la qualité des fruits. Nous présenterons ici les avancées les plus récentes de notre équipe dans l'identification du contrôle génétique des paramètres de qualité des fruits, et comment les résultats peuvent être utilisés pour la sélection génomique.

2. Identification des attentes des consommateurs et des principaux caractères de qualité

Plusieurs expériences ont été réalisées pour identifier les attentes des consommateurs face à la diversité des variétés de tomates. Nous avons d'abord montré que les caractéristiques de saveur et de texture sont importantes dans les préférences des consommateurs [3], et nous avons détecté le potentiel des hybrides entre les variétés traditionnelles et les lignées modernes, probablement pour atteindre une fermeté et une jutosité intermédiaires du fruit. Dans l'étude de Sinesio *et al.* [12], un panel d'analyse sensoriel et un large panel de consommateurs de France, des Pays-Bas et d'Italie ont testé 16 cultivars de tomates cultivés dans ces trois pays. Quatre groupes de consommateurs ont été identifiés dans chaque pays avec différents descripteurs correspondant à leurs préférences. Les mêmes segments ont été identifiés dans chaque pays. Cette étude a souligné l'importance de proposer différents segments (plus ou moins savoureux et fermes) sur le marché.

Enfin, Sinesio *et al.* [6] ont montré que pour un type spécifique de tomate (cerise, gros ou long fruit), les variétés traditionnelles et modernes pouvaient être appréciées par les consommateurs, sans qu'il y ait une tendance systématique pour l'un ou l'autre groupe de variétés.

3. QTL et GWAS pour élucider le contrôle génétique des caractéristiques de qualité des fruits

La plupart des caractéristiques de qualité des fruits de la tomate sont héritées quantitativement. Depuis les premières études de cartographie de QTL dans les années 1990, plusieurs expériences de cartographie de QTL ont été réalisées afin d'identifier les locus à utiliser pour la sélection assistée par marqueurs [13, 14]. Quelques gènes majeurs impliqués dans le poids du fruit, sa forme et sa composition ont été clonés [15]. Martina *et al.* [16] ont construit une carte consensus qui synthétise les QTLs identifiés pour l'arôme de la tomate à partir de 16 études. Les auteurs ont défini des régions génomiques où les QTL se chevauchent partiellement entre les études pour des COV de voies métaboliques similaires, et ont examiné les listes de gènes dans les régions d'intérêt, afin d'exploiter les connaissances actuelles et de ne pas manquer les variants causaux qui peuvent tomber hors de l'intervalle de confiance d'un QTL identifié dans une seule étude.

Avec la disponibilité des séquences génomiques et des SNP, les études d'associations à l'échelle du génome (GWAS) ont alors offert une autre approche. Sauvage *et al.* [17] ont génotypé 163 accessions de tomates, composées d'accessions de tomates à gros fruits, de tomates cerises et de tomates sauvages, avec une puce à SNP, générant un total de 5995 SNP de haute qualité. Ils ont ensuite réalisé une étude d'association pangénomique à l'aide d'un modèle mixte multi-locus (MLMM [18]) pour 36 métabolites des fruits qui étaient fortement corrélés entre les deux années de test et ont identifié 44 locus candidats associés à différents métabolites des fruits [17]. Parmi les locus candidats, ils ont identifié un gène de fonction inconnue sur le chromosome 6 qui était fortement associé au contenu en malate. Cette association a été identifiée dans différentes études d'association pangénomiques et méta-analyses de GWAS

basées sur différentes populations [19–22] et a été validée comme étant un transporteur de malate activé par l'aluminium (Sl-ALMT9) [21].

Bauchet *et al.* [19] ont génotypé 300 accessions de tomates avec un total de 11 012 SNP de haute qualité, qui ont été utilisés pour l'étude d'association pangénomique en utilisant à la fois le modèle MLMM et le modèle mixte multi-trait (MTMM) [23]. Un total de 79 associations significatives ont été identifiées pour 13 métabolites primaires et 19 COV dans les fruits de la tomate. Parmi celles-ci, deux associations impliquant l'acidité du fruit et la teneur en phénylpropanoïdes ont été particulièrement étudiées [19]. La même population a également été caractérisée pour des traits agronomiques et de nombreux QTL ont été identifiés, tels que fw2.2 et fw3.2 pour le poids des fruits. Des études de GWAS pour des traits de qualité similaires ont également été réalisées dans d'autres collections [24, 25]. Avec le développement rapide de la technologie de séquençage du génome entier et la réduction du coût par génome, il est possible de séquencer des centaines d'accessions de tomates. Par exemple, Tieman *et al.* [20] ont séquencé 231 nouvelles accessions et ont combiné ces données avec 245 génomes précédemment séquencés, générant un total de 476 séquences génomiques. Ces données ont ensuite été utilisées pour l'étude d'association pangénomique de divers métabolites liés à l'arôme, dont les teneurs en 27 COV, en solides solubles totaux, en glucose, fructose, acide citrique et acide malique. Un total de 251 associations significatives ont été détectées pour 20 traits. Deux locus ont été significativement associés à la fois au glucose et au fructose, correspondant à deux QTL majeurs Lin5 et SSC11.1. En combinant cette étude et la recherche de traces de sélection, il a été démontré que la corrélation négative entre la teneur en sucre et le poids du fruit était probablement causée par la perte d'allèles à haute teneur en sucre pendant la domestication et la sélection initiale de fruits de tomate plus gros [20]. En outre, quelques gènes candidats impliqués dans les contenus en volatils ont également été identifiés, tels que Solyc09g089580 pour le guaiacol et le méthylsalicylate. En combinant les trois locus associés significativement pour la géranylacétone et la 6-méthyl-5-heptène-2-one, il a été montré que plusieurs combinaisons alléliques conférant des arômes favorables ont été progressivement perdues au cours de la domestication et

de la sélection [20]. Quelques autres QTLs pour les volatils ont été identifiés suite à un clonage positionnel [26,27].

4. Méta-analyse des GWAS

Cependant, avec les résultats de plusieurs études d'association pangénomiques sur la tomate pour le même caractère, seules quelques associations significatives ont pu être identifiées de manière cohérente entre les différentes études, ce qui indique une forte hétérogénéité inter-études, qui fait référence à la variance non aléatoire des effets génétiques entre les différentes études d'association pangénomiques. Les principales sources d'hétérogénéité entre études sont la structure de la population, le déséquilibre de liaison, les méthodes de mesure du phénotype, les facteurs environnementaux, les méthodes de génotypage, les interactions $G \times E$. La méta-analyse des études d'association pangénomique est une nouvelle approche permettant de combiner différentes études d'association pangénomique, tout en gérant correctement l'hétérogénéité. Ainsi, Zhao *et al.* [22] ont réalisé la méta-analyse des GWAS de trois populations de GWAS de tomates [17,19,20]. Après imputation du génotype, un total de 775 accessions de tomates et 2 316 117 SNP ont été utilisés dans la méta-analyse, et un total de 305 associations significatives ont été identifiées pour les sucres, les acides organiques, les acides aminés et les COV liés à la qualité. En examinant les cinq locus associés à la fois au fructose et au glucose, ils ont montré que la teneur en sucres augmentait significativement avec le nombre d'allèles sauvages. Les auteurs ont également démontré que la domestication et la sélection ont eu un impact sur la teneur en citrate et en malate. En particulier, le QTL majeur *Al-Activated Malate Transporter 9* du malate a été significativement associé au citrate et un autre transporteur de malate a été identifié pour la teneur en citrate sur le chromosome 1. Cette étude a également identifié de nombreuses nouvelles associations significatives pour les COV liés aux arômes du fruit. En ciblant six associations significatives, il a été démontré que les accessions de tomates modernes avaient un arôme limité en raison d'une teneur plus faible en COV agréables mais aussi d'une teneur plus élevée en COV jugés désagréables par rapport aux tomates cerises [22].

5. GWAS dans les hybrides

Comme la plupart des variétés modernes sont des hybrides F1, mais que les analyses GWAS ont été effectuées sur des lignées, Bineau *et al.* [28] ont étendu l'approche GWAS aux génomes hybrides. Les auteurs ont quantifié 46 COV dans un panel de GWAS de 121 lignées à petits fruits — type cerise et accessions *S. pimpinellifolium* sauvages — et dans un panel de 165 hybrides F1 (le panel précédent plus 44 lignées élites de tomates cerises croisées avec une lignée testeur commune à gros fruits). Ils ont réalisé une analyse d'association pangénomique sur les deux panels séparément, ainsi qu'une troisième analyse d'association pangénomique sur le sous-ensemble de croisements de test portant uniquement les hybrides F1 correspondant au panel de lignées. Ils ont identifié plusieurs régions génomiques qui pourraient aider les sélectionneurs à récupérer la saveur dans les variétés modernes, tout en tenant compte des compromis sur le poids des fruits qui peuvent être en jeu pour l'amélioration du matériel végétal à gros fruits.

6. Impact des conditions environnementales sur la qualité des fruits

La tomate est cultivée dans le monde entier dans des environnements très différents. Son rendement et la qualité de ses fruits sont influencés par les conditions environnementales (température, lumière, irrigation, nutrition [29–31]). Au cours des 10 dernières années, nous avons étudié l'impact des trois principaux stress des conditions méditerranéennes (stress hydrique, salin et thermique) sur deux panels de variétés de tomates. Nous avons montré une grande variabilité dans chaque panel pour la plupart des traits relatifs à l'architecture de la plante, la phénologie ou la composition des fruits et aussi pour leur réponse à des stress spécifiques. Par exemple, un léger stress hydrique peut améliorer la saveur sucrée des fruits mais un stress plus fort réduira simultanément le rendement et le poids des fruits [32, 33]. Les conditions de chaleur entravent la pollinisation et ont donc un impact important sur le rendement et accélèrent la vitesse de floraison [34]. L'étude de la plasticité des plantes a permis d'identifier certaines accessions moins impactées que d'autres par les variations de l'environnement, mais aussi de caractériser le contrôle génétique des composantes de la

plasticité [31]. Des QTL de plasticité ont été détectés, mais en général dans des régions distinctes de celles des QTL de moyenne d'un caractère donné, suggérant la possibilité d'une sélection indépendante pour la plasticité et la valeur du trait. Des gènes candidats potentiellement impliqués dans l'apparition de l'interaction entre génotype et environnement ont été proposés, ouvrant la voie à la caractérisation fonctionnelle des gènes de réponse au stress chez la tomate.

7. Prédiction génomique

La cartographie de QTL et la GWAS ont des limites dans l'identification et la quantification des locus à faible effet, ainsi que des QTL ou associations sensibles aux conditions environnementales [35]. En revanche, la sélection génomique (SG), qui a été proposée il y a une vingtaine d'années [36], utilise toutes les informations génétiques provenant de marqueurs répartis sur l'ensemble du génome. En utilisant les SNP et les données phénotypiques dans une population d'apprentissage, on peut prédire les valeurs génétiques d'individus non phénotypés dans une population de test. Les principaux avantages de la SG sont la réduction des coûts et le gain de temps par rapport à la sélection basée sur le phénotype [35].

Plusieurs facteurs influencent la précision de la prédiction génomique (PG), notamment la taille, la structure et la diversité génétique de la population d'entraînement, l'héritabilité des caractères, le nombre et la distribution des marqueurs moléculaires, le déséquilibre de liaison, la méthode de prédiction et le nombre de QTL [37–39]. Afin d'améliorer la précision de la prédiction, des modèles complexes de PG ont été développés afin de traiter différents facteurs, tels que les interactions $G \times E$ et la prédiction multi-traits et multi-environnements [40, 41]. De nombreux modèles de PG sont disponibles et la précision de la prédiction varie en fonction des caractères et des conditions [38, 39, 42].

Le premier test de PG sur la tomate était axé sur un plan de sélection et une prédiction basés sur la simulation, où une méthode théorique a été proposée pour appliquer la SG à des plans de sélection réels visant à améliorer simultanément le rendement et la saveur [38]. En pratique, 96 variétés de tomates à gros fruits ont été sélectionnées et 20 caractères

agronomiques ont été mesurés, avec une héritabilité au sens large allant de 0,10 à 1,00. Sept modèles de PG ont été comparés. La précision de prédiction la plus élevée pour différents traits a varié et la précision du modèle Bayes C a été la plus élevée pour un maximum de huit traits, se classant au premier rang parmi tous les modèles. Certains individus présentant une prédiction élevée pour le poids total des fruits et la teneur en solides solubles ont été sélectionnés comme parents pour simuler les générations suivantes. Les simulations ont démontré qu'après cinq générations, les valeurs simulées étaient comparables aux variétés parentales.

Yamamoto *et al.* [39] ont ensuite utilisé une population d'hybrides F1 à gros fruits pour construire les modèles de PG afin d'évaluer leur potentiel pour l'amélioration du poids total des fruits et de la teneur en solides solubles dans une expérience. En testant six modèles de PG et en effectuant une validation croisée 10 fois, la précision de la prédiction de la teneur en solides solubles était plus élevée que celle du poids total du fruit. Duangjit *et al.* [37] ont étudié l'impact de certains facteurs clés sur l'efficacité de la PG, notamment la taille de la population d'entraînement, le nombre et la densité des SNP et la parenté entre individus. Sur la base de l'analyse de 163 accessions de tomates étudiées par Sauvage *et al.* [17], la taille optimale de la population d'entraînement était de 122. La précision de la prédiction a également augmenté avec la densité et le nombre de marqueurs, mais faiblement. L'appareillement entre individus a également influencé la précision de la prédiction, et les prédictions étaient meilleures lorsque l'appareillement des individus était plus fort. Cependant, cette étude présente certaines limites : (1) elle n'a testé que le modèle statistique *ridge regression best linear unbiased prediction* (rrBLUP) [43]; (2) le nombre de SNP était relativement faible et la couverture génomique dans certaines régions génomiques était assez limitée; (3) il existait une structure de population et le nombre d'accessions sauvages était assez faible par rapport aux accessions de tomates cerises et à gros fruits.

La plupart des modèles de PG reposent sur des informations basées sur les marqueurs et sont incapables d'exploiter les interactions épistatiques locales entre les marqueurs. Les marqueurs moléculaires peuvent également être combinés en haplotypes en associant le déséquilibre de liaison et

l'analyse de liaison pour améliorer la précision de la prédiction, comme le montre [44].

8. Conclusion

Au cours des vingt dernières années, les connaissances sur le développement et la qualité des fruits ont fortement augmenté. De nombreuses données sont disponibles pour la sélection de nouvelles variétés de tomates avec un goût de fruit amélioré. La nécessité d'impliquer l'ensemble de la chaîne de production (de la production de la variété à la commercialisation) et la reconnaissance économique de la qualité des fruits restent des facteurs limitants. Au niveau fondamental, de nombreux gènes importants peuvent encore être découverts. Les outils et les données génomiques sont également disponibles pour accélérer leur découverte dans un avenir proche.

Conflit d'intérêt

Les auteurs n'ont aucun conflit d'intérêt à déclarer.

Remerciements

Cette recherche a été soutenue par de nombreux projets de recherche nationaux et internationaux. Parmi ceux qui soutiennent actuellement notre travail, nous pouvons mentionner les projets H2020 "HARNESSTOM", subvention no 101000716, SUscrop Era-NET No 47 2019 "SOLNUE", PRIMA 2018-11 "VE-GADAPT" et ANR "3C-FruitGrowth".

References

- [1] E. Baldwin, J. Scott, C. Shewmaker, W. Schuch, "Flavor trivia and tomato aroma: biochemistry and possible mechanisms for control of important aroma components", *HortScience* **35** (2000), p. 1013-1022.
- [2] H. J. Klee, D. M. Tieman, "The genetics of fruit flavour preferences", *Nat. Rev. Genet.* **19** (2018), p. 347-356.
- [3] M. Causse, M. Buret, K. Robini, P. Verschave, "Inheritance of nutritional and sensory quality traits in fresh market tomato and relation to consumer preferences", *J. Food Sci.* **68** (2003), p. 2342-2350.
- [4] B. D. Whitaker, "Postharvest flavor deployment and degradation in fruits and vegetables", in *Fruit and Vegetable Flavour* (B. Bruckner, S. Grant Willie, eds.), CRC Press, Cambridge, 2008, p. 103-131.
- [5] M. Causse, C. Friguet, C. Coiret, M. Lépicier, B. Navez, M. Lee, N. Holthuysen, F. Sinesio, E. Moneta, S. Grandillo, "Consumer preferences for fresh tomato at the European scale: a common segmentation on taste and firmness", *J. Food Sci.* **79** (2010), p. S531-S541.
- [6] F. Sinesio, M. Cammareri, V. Cottet, L. Fontanet, M. Jost, E. Moneta, S. Palombieri, M. Peparaio, R. Romero del Castillo, E. Saggia Civitelli, P. Spigno, A. Vitiello, B. Navez, J. Casals, M. Causse, A. Granell, S. Grandillo, "Sensory traits and consumer's perceived quality of traditional and modern fresh market tomato varieties: a study in three european countries", *Foods* **10** (2021), article no. 2521.
- [7] A. Bellec-Gauche, Y. Chiffolleau, C. Maffezzoli, "Case study: multidimensional comparison of local and global fresh tomato supply chains", 2015, GLAMUR Project Report.
- [8] M. A. Stevens, "Inheritance of tomato fruit quality components", *Plant Breed. Rev.* **4** (1986), p. 274-231.
- [9] J. L. Rambla, Y. M. Tikunov, A. J. Monforte, A. G. Bovy, A. Granell, "The expanded tomato fruit volatile landscape", *J. Exp. Bot.* **65** (2014), p. 4613-4623.
- [10] Tomato Genome Consortium, "The tomato genome sequence provides insights into fleshy fruit evolution", *Nature* **485** (2012), p. 635-641.
- [11] T. Lin, G. Zhu, J. Zhang, X. Xu, Q. Yu, Z. Zheng, Z. Zhang, Y. Lun, S. Li, X. Wang, Z. Huang, J. Li, C. Zhang, T. Wang, Y. Zhang, A. Wang, Y. Zhang, K. Lin, C. Li, G. Xiong, Y. Xue, A. Mazzucato, M. Causse, Z. Fei, J. J. Giovannoni, R. T. Chetelat, T. Städler, J. Li, Z. Ye, Y. Du, S. Huang, "Genomic analyses provide insights into the history of tomato breeding", *Nat. Genet.* **46** (2014), p. 1220-1226.
- [12] F. Sinesio, M. Cammareri, E. Moneta, B. Navez, M. Peparaio, M. Causse, S. Grandillo, "Sensory quality of fresh market tomatoes: an Italian preference mapping study", *J. Food Sci.* **75** (2009), p. S55-S67.
- [13] M. Causse, V. Saliba-Colombani, L. LeComte, P. Duffé, P. Rouselle, M. Buret, "QTL analysis of fruit quality in fresh market tomato: a few chromosome regions control the variation of sensory and instrumental traits", *J. Exp. Bot.* **53** (2002), p. 2089-2098.
- [14] L. Lecomte, P. Duffé, M. Buret, B. Servin, F. Hospital, M. Causse, "Marker-assisted introgression of five QTLs controlling fruit quality traits into three tomato lines revealed interactions between QTLs and genetic backgrounds", *Theor. Appl. Genet.* **109** (2004), p. 658-668.
- [15] C. Rothan, I. Diouf, M. Causse, "Trait discovery and editing in tomato", *Plant J.* **97** (2019), p. 73-90.
- [16] M. Martina, Y. Tikunov, E. Portis, A. G. Bovy, "The genetic basis of tomato aroma", *Genes* **12** (2021), article no. 226.
- [17] C. Sauvage, V. Segura, G. Bauchet, R. Stevens, P. T. Do, Z. Nikoloski, A. R. Fernie, M. Causse, "Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits", *Plant Physiol.* **165** (2014), p. 1120-1132.
- [18] V. Segura, B. J. Vilhjálmsón, A. Platt, A. Korte, Ü. Seren, Q. Long, M. Nordborg, "An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations", *Nat. Genet.* **44** (2012), p. 825-830.
- [19] G. Bauchet, S. Grenier, N. Samson, V. Segura, A. Kende, J. Beekwilder, K. Cankar, J. L. Gallois, J. Gricourt, J. Bonnet, C. Baxter, L. Grivet, M. Causse, "Identification of major loci

- and genomic regions controlling acid and volatile content in tomato fruit: implications for flavor improvement”, *New Phytol.* **215** (2017), p. 624-641.
- [20] D. Tieman, G. Zhu, M. F. R. Resende, T. Lin, C. Nguyen, D. Bies, J. L. Rambla, K. S. O. Beltran, M. Taylor, B. Zhang, H. Ikeda, Z. Liu, J. Fisher, I. Zemach, A. Monforte, D. Zamir, A. Granell, M. Kirst, S. Huang, H. Klee, “A chemical genetic roadmap to improved tomato flavor”, *Science* **35** (2017), p. 391-394.
- [21] J. Ye, X. Wang, T. Hu, F. Zhang, B. Wang, C. Li, T. Yang, H. Li, Y. Lu, J. J. Giovannoni, Y. Zhang, Z. Ye, “An InDel in the promoter of *AI-ACTIVATED MALATE TRANSPORTER9* selected during tomato domestication determines fruit malate contents and aluminum tolerance”, *Plant Cell* **29** (2017), p. 2249-2268.
- [22] J. T. Zhao, C. Sauvage, J. Zhao, F. Bitton, G. Bauchet, D. Liu, S. Huang, D. Tieman, H. J. Klee, M. Causse, “Meta-analysis of genome-wide association studies provides insights into genetic control of tomato flavor”, *Nat. Commun.* **10** (2019), p. 1-12.
- [23] A. Korte, B. J. Vilhjálmsson, V. Segura, A. Platt, Q. Long, M. Nordborg, “A mixed-model approach for genome-wide association studies of correlated traits in structured populations”, *Nat. Genet.* **44** (2012), p. 1066-1071.
- [24] V. Ruggieri, G. Francese, A. Sacco, A. D’Alessandro, M. M. Rignano, M. Parisi, M. Milone, T. Cardi, G. Mennella, A. Barone, “An association mapping approach to identify favourable alleles for tomato fruit quality breeding”, *BMC Plant Biol.* **14** (2014), article no. 337.
- [25] J. Zhang, J. Zhao, Y. Liang, Z. Zou, “Genome-wide association-mapping for fruit quality traits in tomato”, *Euphytica* **207** (2016), p. 439-451.
- [26] Y. M. Tikunov, J. Molthoff, R. C. H. de Vos, J. Beekwilder, A. van Houwelingen, J. van der Hooft, M. Nijenhuis-de Vries, C. W. Labrie, W. Verkerke, H. van de Geest, M. V. Zamora, S. Presa, J. L. Rambla, A. Granell, R. D. Hall, A. G. Bovy, “NON-SMOKY GLYCOSYLTRANSFERASE1 prevents the release of smoky aroma from tomato fruit”, *Plant Cell* **25** (2013), p. 3067-3078.
- [27] Y. M. Tikunov, R. Roohanitaziani, F. Meijer-Dekens, J. Molthoff, J. Paulo, R. Finkers, I. Capel, F. C. Moreno, C. Maliapaard, M. Nijenhuis-de Vries, C. W. Labrie, W. Verkerke, A. W. van Heusden, F. van Eeuwijk, R. G. F. Visser, A. G. Bovy, “The genetic and functional analysis of flavor in commercial tomato: the *FLORAL4* gene underlies a QTL for floral aroma volatiles in tomato fruit”, *Plant J.* **103** (2020), p. 1189-1204.
- [28] E. Bineau, J. L. Rambla, S. Priego-Cubero, A. Hereil, F. Bitton, C. Plissonneau, A. Granell, M. Causse, “Breeding tomato hybrids for flavour: comparison of GWAS results obtained on lines and F1 hybrids”, *Genes* **12** (2021), article no. 1443.
- [29] S. Saadi, M. Todorovic, L. S. Pereira, *Climate Change and Mediterranean Agriculture: 2. Impacts on Wheat and Tomato Yields and Water Productivity*, vol. 147, Elsevier, 2010, 1-14 pages.
- [30] S. Fragkostefanakis, S. Simm, P. Paul, D. Bublak, K.-D. Scharf, E. Schleiff, “Chaperone network composition in *Solanum lycopersicum* explored by transcriptome profiling and microarray meta-analysis”, *Plant Cell Environ.* **38** (2015), p. 693-709.
- [31] I. Diouf, L. Derivot, S. Koussevitzky, Y. Carretero, F. Bitton, L. Moreau, M. Causse, “Genetic basis of phenotypic plasticity and genotype × environment interactions in a multi-parental tomato population”, *J. Exp. Bot.* **71** (2020), p. 5365-5376.
- [32] I. A. Diouf, L. Derivot, F. Bitton, L. Pascual, M. Causse, “Water deficit and salinity stress reveal many specific QTL for plant growth and fruit quality traits in tomato”, *Front. Plant Sci.* **9** (2018), p. 1-13.
- [33] E. Albert, R. Duboscq, M. Latreille, S. Santoni, M. Beukers, J. P. Bouchet, F. Bitton, J. Gricourt, C. Poncet, V. Gautier, J. M. Jiménez-Gómez, G. Rigail, M. Causse, “Allele-specific expression and genetic determinants of transcriptomic variations in response to mild water deficit in tomato”, *Plant J.* **96** (2018), p. 635-650.
- [34] E. Bineau, I. Diouf, Y. Carretero, R. Duboscq, F. Bitton, A. Djari, M. Zouine, M. Causse, “Genetic diversity of tomato response to heat stress at the QTL and transcriptome levels”, *Plant J.* **107** (2021), p. 1213-1227.
- [35] J. Crossa, P. Pérez-Rodríguez, J. Cuevas, O. Montesinos-López, D. Jarquín, G. de los Campos, J. Burgueño, J. M. Camacho-González, S. Pérez-Elizalde, Y. Beyene, S. Dreisigacker, R. Singh, X. Zhang, M. Gowda, M. Rorkkial, J. Rutkoski, R. K. Varshney, “Genomic selection in plant breeding: methods, models, and perspectives”, *Trends Plant Sci.* **22** (2017), p. 961-975.
- [36] T. H. E. Meuwissen, B. J. Hayes, M. E. Goddard, “Prediction of total genetic value using genome-wide dense marker maps”, *Genetics* **157** (2001), p. 1819-1829.
- [37] J. Duangjit, M. Causse, C. Sauvage, “Efficiency of genomic selection for tomato fruit quality”, *Mol. Breed.* **36** (2016), article no. 29.
- [38] E. Yamamoto, H. Matsunaga, A. Onogi, H. Kajiya-Kanegae, M. Minamikawa, A. Suzuki, K. Shirasawa, H. Hirakawa, T. Nunome, H. Yamaguchi, K. Miyatake, A. Ohyama, H. Iwata, H. Fukuoka, “A simulation-based breeding design that uses whole-genome prediction in tomato”, *Sci. Rep.* **6** (2016), article no. 19454.
- [39] E. Yamamoto, H. Matsunaga, A. Onogi, A. Ohyama, K. Miyatake, H. Yamaguchi, T. Nunome, H. Iwata, H. Fukuoka, “Efficiency of genomic selection for breeding population design and phenotype prediction in tomato”, *Heredity (Edinb)* **118** (2017), p. 202-209.
- [40] S. B. Fernandes, K. O. G. Dias, D. F. Ferreira, P. J. Brown, “Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum”, *Theor. Appl. Genet.* **131** (2018), p. 747-755.
- [41] O. A. Montesinos-López, A. Montesinos-López, J. Crossa, F. H. Toledo, O. Pérez-Hernández, K. M. Eskridge, J. Rutkoski, “A genomic bayesian multi-trait and multi-environment model”, *G3-Genes Genome Genet.* **6** (2016), p. 2725-2744.
- [42] N. Heslot, H.-P. Yang, M. E. Sorrells, J.-L. Jannink, “Genomic selection in plant breeding: a comparison of models”, *Crop Sci.* **52** (2012), p. 146-160.
- [43] J. B. Endelman, “Ridge regression and other kernels for genomic selection with R package rrBLUP”, *Plant Genome J.* **4** (2011), article no. 250.
- [44] J. Zhao, C. Sauvage, F. Bitton, M. Causse, “Multiple haplotype-based analyses provide genetic and evolutionary insights into tomato fruit weight and composition”, *Hortic. Res.* **9** (2022), article no. uhab009.