



Review

# Molecular Approaches to Genetically Improve the Accumulation of Health-Promoting Secondary Metabolites in Staple Crops—A Case Study: The *Lipoxygenase-B1* Genes and Regulation of the Carotenoid Content in Pasta Products

Grazia M. Borrelli and Daniela Trono \*

Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, Centro di Ricerca per la Cerealicoltura, S.S. 673 Km 25,200, 71122 Foggia, Italy; graziamaria.borrelli@crea.gov.it

\* Correspondence: daniela.trono@crea.gov.it; Tel.: +39-0881-742-972

Academic Editor: Marcello Iriti

Received: 5 June 2016; Accepted: 11 July 2016; Published: 21 July 2016

**Abstract:** Secondary metabolites, also known as phytochemicals, represent a large subset of plant molecules that include compounds with health-promoting effects. Indeed, a number of epidemiological studies have shown that, when taken regularly and in adequate amounts, these molecules can have long-term beneficial effects on human health, through reduction of the incidence of degenerative diseases, such as cardiovascular diseases, obesity, diabetes, and cancer. As the dietary intake of these phytochemicals is often inadequate, various strategies are in use to improve their content in staple crops, and the end-products thereof. One of the most effective strategies is crop improvement through genetic approaches, as this is the only way to generate new cultivars in which the high accumulation of a given phytochemical is stably fixed. Efforts to genetically improve quality traits are rapidly evolving, from classical breeding to molecular-assisted approaches; these require sound understanding of the molecular bases underlying the traits, to identify the genes/alleles that control them. This can be achieved through global analysis of the metabolic pathway responsible for phytochemical accumulation, to identify the link between phytochemical content and the activities of key enzymes that regulate the metabolic pathway, and between the key enzymes and their encoding genes/alleles. Once these have been identified, they can be used as markers for selection of new improved genotypes through biotechnological approaches. This review provides an overview of the major health-promoting properties shown to be associated with the dietary intake of phytochemicals, and describes how molecular approaches provide means for improving the health quality of edible crops. Finally, a case study is illustrated, of the identification in durum wheat of the *Lipoxygenase-B1* genes that control the final carotenoid content in semolina-based foods, such as pasta products.

**Keywords:** phytochemicals; human health; metabolic pathway; metabolite analysis; enzymatic activity; transcript profile; candidate gene; biotechnological breeding; lipoxygenase; pasta-products

## 1. Introduction

The secondary metabolites from plants, which are known as phytochemicals, represent a large subset of plant chemicals, but their importance in both plant metabolism and human health has only been recognized more recently. This is because, unlike the primary metabolites, carbohydrates, lipids and proteins, which are ubiquitously present at high concentrations in all living organisms, in which they share the same essential functions (i.e., growth, development and reproduction), secondary metabolites are present at much lower concentrations, are restricted to specific groups of organisms

and are produced in response to particular conditions [1]. Likewise, in human nutrition attention has been focused for years on the energy demand that needs to be satisfied through the correct combination of carbohydrates, lipids, and proteins in the diet, whereas the impact of secondary metabolites on human health has been ignored for a long time.

In the last decades, evidence has accumulated of the direct involvement of secondary metabolites in various processes related to plant fitness, such as defense against biotic and abiotic stress, formation of symbiosis at the root level, and reproductive processes through attraction of animals involved in pollination or seed dispersal [2]. In the same manner, the recognition of the biological properties of a number of secondary metabolites has led to the revaluation of these compounds also from the health point of view. Although phytochemicals do not have roles as nutrients, increasing epidemiological evidence has associated the consumption of foods rich in phytochemicals, such as fruit, vegetables, and grains to reduced incidence of cardiovascular diseases (CVD), obesity, diabetes, cancer, and other chronic degenerative diseases [3].

Nowadays, the dietary intake of bioactive compounds is often inadequate, and worldwide, people are becoming even more susceptible to the development of degenerative diseases [4]. Therefore, the demand for new effective strategies that are aimed at prevention of these diseases has increased considerably in the last few decades. One approach is to increase the content of health-promoting compounds in staple crops and their end products. This can be effectively achieved by genetic improvement of crops through proper breeding programs, which represent the only certain way to generate new cultivars in which a desirable phytochemical content is stably fixed.

From the late 19th century, plant breeding relied on phenotypic selection. Thus the improvement of a quality trait, such as the accumulation of a desirable compound, was achieved through screening of hundreds, or even thousands, of plants. Laboratory tests were then used to evaluate the phytochemical content and to select the plants of superior quality. The entire process was relatively long (>10 years) and laborious. Fortunately, over the years, with the advent of molecular genetics, crop improvement has rapidly evolved from classical breeding approaches to biotechnological breeding approaches, such as marker-assisted selection (MAS) [5], genetic transformation [6], and very recently, genome editing [7]. These advanced technologies require sound understanding of the molecular bases underlying the accumulation of a bioactive compound in a given crop, to be able to identify the candidate gene that controls its levels. The challenge is then to introduce this candidate gene into an elite cultivar, while retaining its performance attributes.

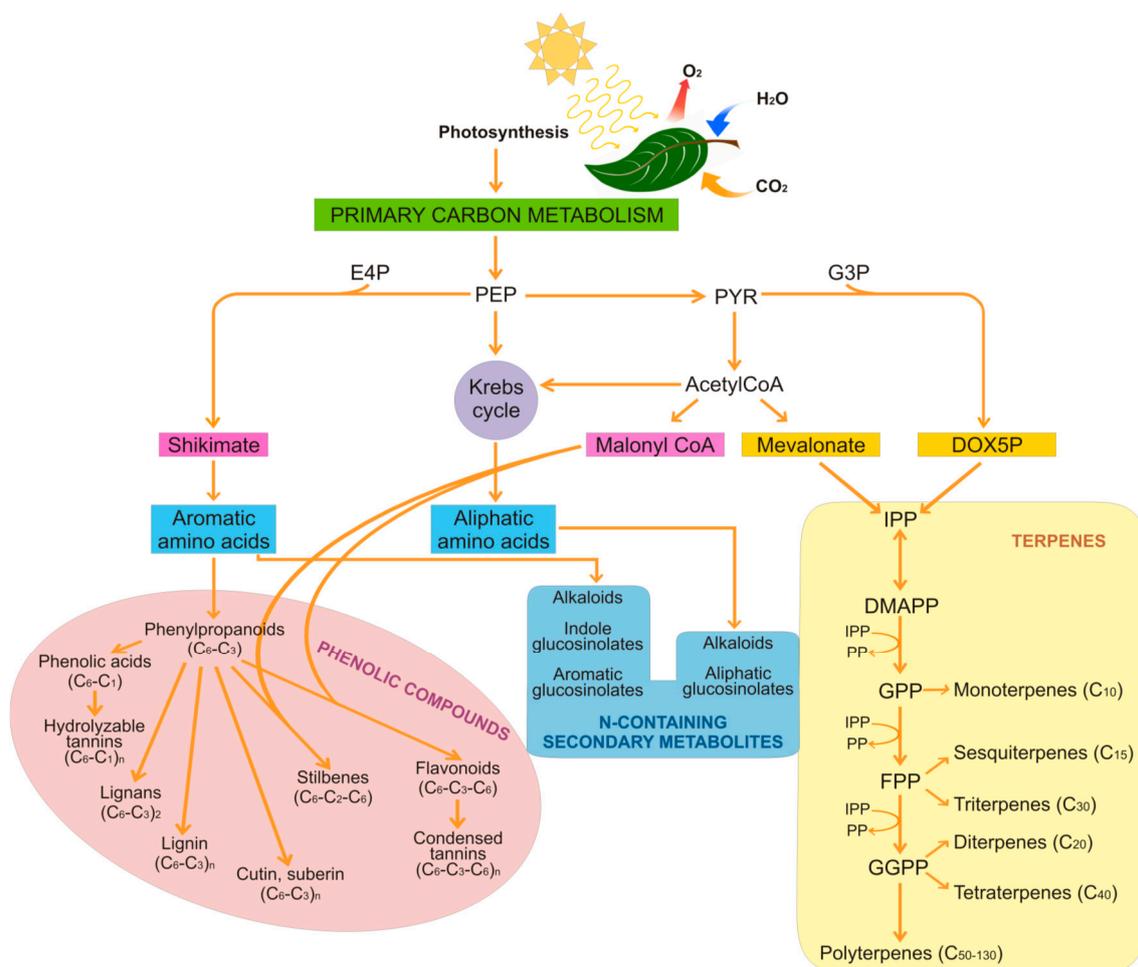
Two approaches can be used to identify and validate such a candidate gene: (i) the positional approach; and (ii) the functional approach [8]. The positional approach is based on the identification of a potential candidate gene among the genes that map around the quantitative trait locus (QTL) that is associated with the trait. This is followed by its validation through the demonstration that polymorphic variants of the candidate gene (i.e., alleles) can explain the variability of the trait. Conversely, the functional approach bypasses the QTL step, as it relies on identification of the candidate gene through its functional relation to the trait, and in particular through the association between its allelic variation and the variation in the flux rate through the pathway that is responsible for the metabolite accumulation. The metabolic pathway is often a multi-step event that involves different enzyme-catalyzed reactions, although the rate of the overall pathway is generally regulated by the rate of one, or a few, reactions that are catalyzed by the so-called "rate-limiting" enzymes. Allelic variation at the gene that encodes a key regulatory enzyme of the pathway might have the potential to modify the functioning of the enzyme, and consequently, the accumulation of the desired phytochemical. Therefore, the combination of target metabolite determination with enzymatic assays as well as gene-sequence and expression analyses might represent a valuable multi-disciplinary approach for the identification of favorable alleles that, when used as markers, can help a breeder to successfully achieve the goal of biofortification of staple crops.

Here, we provide an overview of the main classes of plant phytochemicals, and we summarize the current knowledge of their health-promoting effects. Then, we focus on the molecular strategies

and tools that can be applied to effectively manipulate the phytochemical levels in agronomically important crops. Finally, we describe a functional study of the *lipoxygenase-B1* gene family in durum wheat that controls the final carotenoid content in pasta products.

## 2. General Overview of Plant Secondary Metabolites

Secondary metabolites can essentially be divided into three main groups that are based on their structures and the metabolic pathways responsible for their biosynthesis (Figure 1): phenolic compounds; terpenoids; and nitrogen-containing compounds, i.e., alkaloids and glucosinolates [9].



**Figure 1.** Schematic representation of the biosynthesis of secondary metabolites. E4P, erythrose 4-phosphate; G3P, glyceraldehyde 3-phosphate; PEP, phosphoenolpyruvate; PYR, pyruvate; DOX5P, deoxyxylulose 5-phosphate; IPP, isopentenyl pyrophosphate; DMAPP, dimethylallyl pyrophosphate; PP, pyrophosphate; GPP, geranyl pyrophosphate; FPP, farnesyl pyrophosphate; GGPP, geranylgeranyl pyrophosphate.

### 2.1. Phenolic Compounds

Phenolic compounds represent one of the largest groups of secondary metabolites produced in plants (i.e., >8000 different structures), and particularly in edible plants (e.g., vegetables, fruit, cereals). These include several structurally different compounds that have at least one aromatic hydroxyl-substituted ring in common. Phenolic compounds are produced primarily via the shikimate pathway that gives rise to phenylpropanoids (Figure 1).

The combination of the shikimate pathway with the mevalonate pathway leads to the generation of the flavonoids [10]. Condensation and polymerization reactions produce hydrolysable and condensed tannins, lignans, lignin, cutin, and suberin. Phenolic compounds are involved in a number of physiological processes in plants, such as pollination, resistance to pathogens and predators, and resistance to abiotic stress [10]. Although they can exist in their free forms, phenolic compounds are commonly conjugated to sugars or proteins, and can also occur as esters and methyl esters, probably because most of them are toxic compounds that are at least in part neutralized in their bound form. Due to the large number and heterogeneity of structures, phenolic compounds can be classified in different ways. On the basis of their carbon chain, they can be distinguished in different classes that range from simple phenols to highly polymerized compounds (Table 1).

**Table 1.** The main classes of phenolic compounds according to their carbon chains.

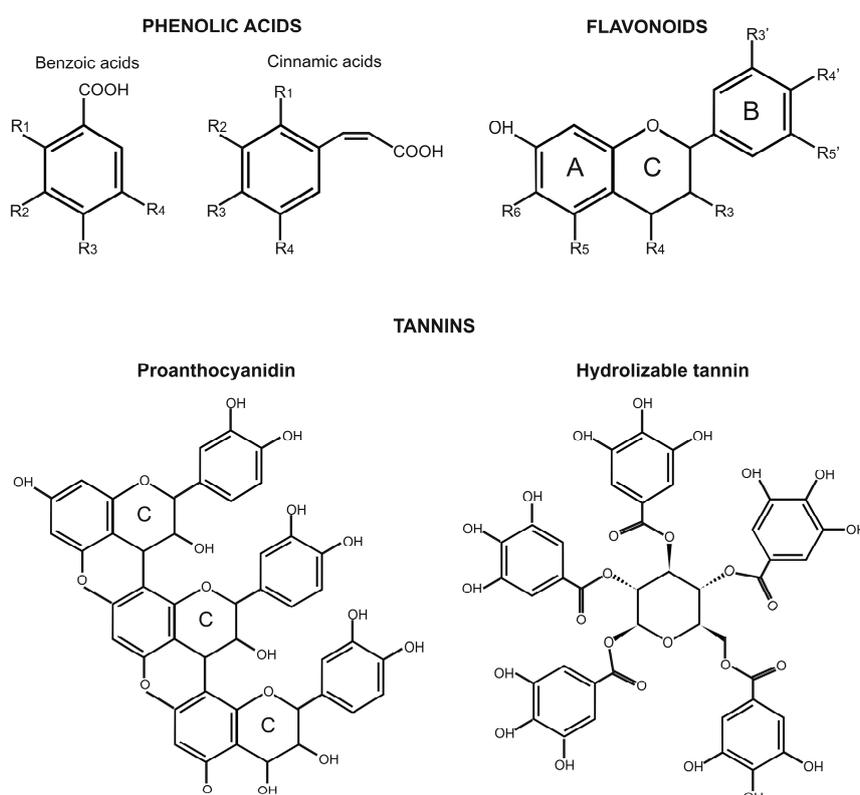
Class	Carbon Skeleton	Examples
C <sub>6</sub>	Simple phenols and benzoquinones	Catechol, 2,6-dimethoxybenzoquinone, hydroquinone, phenol, pirocatechol
C <sub>6</sub> -C <sub>1</sub>	Benzoic acids	Gallic, <i>p</i> -hydroxybenzoic, salicylic, siringic and vanillic acids
C <sub>6</sub> -C <sub>2</sub>	Phenylacetic acids and acetophenones	3-Acetyl-6-methoxybenzaldehyde, tyrosol, <i>p</i> -hydroxyphenylacetic acid
C <sub>6</sub> -C <sub>3</sub>	Cinnamic acids and phenylpropanoids	Caffeic, ferulic, <i>p</i> -coumaric, sinapic acids, aesculetin, bergenon, eugenin, myristicin
C <sub>6</sub> -C <sub>4</sub>	Naphtochinones	Juglone, plumbagin
C <sub>6</sub> -C <sub>1</sub> -C <sub>6</sub>	Xanthones	Mangiferin, mangostin
C <sub>6</sub> -C <sub>2</sub> -C <sub>6</sub>	Stilbenes and anthraquinones	Astringin, resveratrol, viniferin, chrysophanol, emodin
C <sub>6</sub> -C <sub>3</sub> -C <sub>6</sub>	Flavonoids and isoflavonoids	Catechin, cyanidin, epigallocatechin-3-gallate, naringenin, quercetin, myricetin, genistein
(C <sub>6</sub> -C <sub>3</sub> ) <sub>2</sub>	Lignans and neolignans	Lariciresinol, pinoresinol, eusiderin, magnolianin
(C <sub>6</sub> ) <sub>n</sub>	Catechol melanins	
(C <sub>6</sub> -C <sub>1</sub> ) <sub>n</sub>	Hydrolysable tannins	
(C <sub>6</sub> -C <sub>2</sub> ) <sub>n</sub>	Lignin, cutin and suberin	
(C <sub>6</sub> -C <sub>3</sub> -C <sub>6</sub> ) <sub>n</sub>	Condensed tannins (proanthocyanidins)	

The three most important classes of phenolic compounds for human health are phenolic acids, flavonoids, and tannins. Phenolic acids are the simplest phenolic compounds found in nature, and they account for one-third of all of the phenolic compounds in the human diet. Phenolic acids include two major groups, benzoic acids and cinnamic acids, which have seven (C<sub>6</sub>-C<sub>1</sub>) and nine (C<sub>6</sub>-C<sub>3</sub>) carbon atoms, respectively (Figure 2). These compounds are characterized by an aromatic ring, a carboxylic group, and one or more hydroxyl and/or methoxyl groups. Cinnamic acids are more commonly found in nature than benzoic acids. Among the cinnamic acids, caffeic acid and its esterified derivatives are the most abundant in fruit, whereas ferulic acid and its derivatives are the most abundant in cereal grains [11].

Flavonoids are the most widely distributed phenolic compounds in plants, as they account for more than half of the 8000 phenolic compounds found in nature [12]. These occur in various tissues and organs of edible crops, such as fruit, leaves, roots, and tubers, as well as in legumes, tea, coffee, herbs, and spices. Flavonoids have a common structure that consists of two aromatic rings (the A and B rings), joined by a C<sub>3</sub> chain that is generally in the form of a heterocyclic ring (the C ring) (Figure 2). Aromatic ring A comes from the malonate pathway, whereas ring B comes from the shikimate pathway. Substitutions to ring C result in different types of flavonoids, such as flavones, flavonols, flavanones,

flavanols, anthocyanidins, and isoflavones, whereas substitutions to rings A and B give rise to different compounds within each type of flavonoid.

Tannins are the third most important group of phenolic compounds. These include a large number of oligomers and polymers that can form complexes with starch, proteins, cellulose, and minerals. Tannins can be distinguished as two groups: hydrolysable tannins and condensed tannins, which are also known as proanthocyanidins. Hydrolysable tannins are compounds formed from gallic acid or epigallic acid units that are condensed to a central sugar molecule, while condensed tannins are polymers of catechin and epicatechin, or similar units (Figure 2). Proanthocyanidins are more widely distributed in edible plants than hydrolysable tannins, as they are widespread in fruit, legumes, nuts, and minor cereals, such as sorghum and barley, and they tend to concentrate in the peel of fruit and the bran of grains.



**Figure 2.** General structures of the three most important classes of phenolic compounds for human health. Benzoic acids: *p*-hydroxybenzoic acid,  $R_3=OH$ ; vanillic acid,  $R_2=OCH_3$  and  $R_3=OH$ ; gallic acid,  $R_2=R_3=R_4=OH$ ; syringic acid,  $R_2=R_4=OCH_3$  and  $R_3=OH$ . Cinnamic acids: *p*-coumaric acid,  $R_3=OH$ ; caffeic acid,  $R_3=R_4=OH$ ; ferulic acid,  $R_2=OCH_3$  and  $R_3=OH$ ; sinapic acid,  $R_2=R_4=OCH_3$  and  $R_3=OH$ . Flavonoids:  $R_3-R_6=H$  or other groups such as  $OH$  or oxo- groups. For most food flavonoids:  $R_4'=R_6=H$  and  $R_5=OH$ . In biochanin A,  $R_4'=OCH_3$ ; formononetin,  $R_4'=OCH_3$  and  $R_5=R_6=H$ ; glycitein,  $R_5=H$  and  $R_6=OH$ .

Phenolic compounds are considered to be powerful antioxidants due to the reactivity of their phenol moiety. Their main mode of action as antioxidants is radical scavenging via hydrogen atoms or electron donation, together with the delocalization of the unpaired electron within the aromatic ring [13]. Substituents on the aromatic rings affect their stabilization, and therefore affect the radical-quenching of these phenolic compounds. Of note, flavonoids are among the most powerful antioxidants that have been obtained from plants. Their strong antioxidant capacity is due to the presence of hydroxyl groups in positions 3' and 4' of ring B, which participate in the electron delocalization and stabilize the radical that is formed, and of a double bond between the C2 and

C3 carbons in ring C, together with a carbonyl group at the C4 carbon, which make delocalization of an electron from ring B possible [14].

Another class of compounds that contains a phenol ring is represented by the alkylresorcinols. However, compared to phenolic compounds, alkylresorcinols represent a very distinct group, as they have both polar and nonpolar properties. Indeed, in addition to the phenol ring with two hydroxyl groups, alkylresorcinols also have a hydrophobic and mostly saturated alkyl chain. The rich sources of alkylresorcinols include cereal grains (i.e., from wheat, rye, triticale, barley), in which these compounds accumulate in the cuticle layer. There is evidence that alkylresorcinols protect cellular lipids from oxidative damage [15].

## 2.2. Terpenes

Terpenes and terpenoids (i.e., terpene-like compounds) are lipid-soluble compounds that represent the most abundant and structurally diverse group of plant secondary metabolites (with >50,000 different structures) [16]. These are normally produced in vegetative tissues and flowers, and occasionally in roots [17]. They are also referred to as isoprenoids, as they are all derived from the repetitive fusion of isoprene (C<sub>5</sub>H<sub>8</sub>) units. All isoprenoids are derived from the common precursor isopentenyl pyrophosphate (IPP), which is ubiquitously synthesized through the mevalonate pathway or the deoxyxylulose-5-phosphate pathway [18] (Figure 1). IPP isomerizes to dimethylallyl pyrophosphate, and the condensation between IPP and dimethylallyl pyrophosphate gives rise to geranyl pyrophosphate, the precursor of all of the monoterpenes. The IPP additions that follow generate other prenyl diphosphates, which are the precursors of all of the other terpenes. In plants, the synthesis of monoterpenes (C<sub>10</sub>), sesquiterpenes (C<sub>15</sub>), diterpenes (C<sub>20</sub>), triterpenes (C<sub>30</sub>), tetraterpenes (C<sub>40</sub>), and polyterpenes (C<sub>50</sub>–C<sub>130</sub>) has been reported ([19], and references therein).

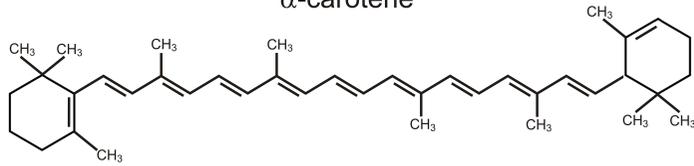
As shown in Table 2, terpenoids include compounds that are implicated in both primary and secondary metabolism. Gibberellin, cytokinins, abscisic acid, and brassinosteroids are hormones that are involved in plant growth and development, whereas carotenoids, and compounds that contain an isoprenyl side chain, such as chlorophyll, phyloquinone, plastoquinone, and tocopherols have functions in photosynthesis [16]. However, most of the terpenoids are secondary metabolites that have crucial roles as pollinator attractants, and in the defense of plants against herbivores and microbial pathogens [19].

**Table 2.** Classification of terpenes and terpenoids on the basis of their isoprene units.

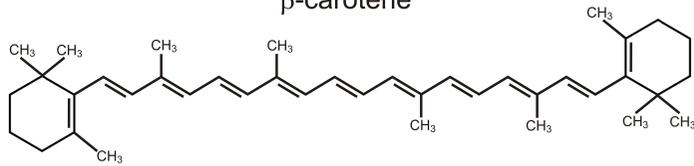
Class	Isoprene Units	Carbon Atoms	Examples
Monoterpene	2	10	Canphor, eucalyptol, geraniol, lavandulol, limonene, menthol, pinene, thymol
Sesquiterpene	3	15	Abscisic acid, bergamotene, cedrol, curcumene, patchoulol, vetivone
Diterpene	4	20	Abietic acid, cafestol, caffeol, carnosol, gibberellin, phytol
Triterpene	6	30	Betulinic acid, morolic acid, oleanolic acid, ursolic acid, brassinosteroids, saponins
Tetraterpene	8	40	Carotenoids (e.g., $\alpha$ -carotene, $\beta$ -carotene, crocin, lycopene) and xanthophylls (e.g., lutein, zeaxanthin, $\alpha$ -cryptoxanthin, $\beta$ -cryptoxanthin)
Polyterpene	>8	>40	Cytokinines, gutta-percha, rubber

### CAROTENES

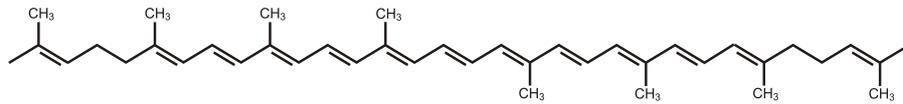
$\alpha$ -carotene



$\beta$ -carotene

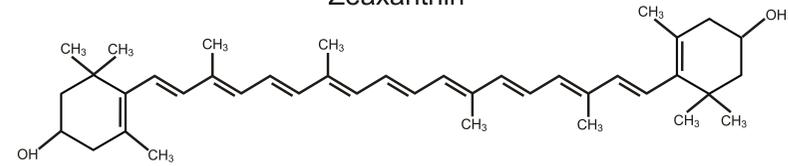


Lycopene

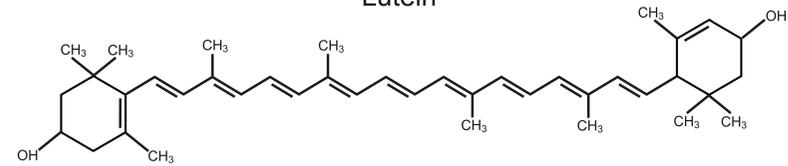


### XANTHOPHYLLS

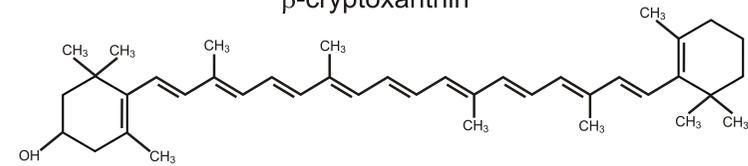
Zeaxanthin



Lutein



$\beta$ -cryptoxanthin



**Figure 3.** Structures of carotenes and xanthophylls that are important for human health.

The carotenoids are a class of terpenoids that have recognized beneficial functions for human health [20]. These are C<sub>40</sub> isoprenoids, and they represent the second most abundant pigments in nature (with >750 compounds). There are two general classes of carotenoids: carotenes, which consist exclusively of hydrogen and carbon; and xanthophylls, which also contain oxygen (Figure 3). As animals and humans cannot synthesize carotenoids *de novo*, the diet is relied on for these essential compounds. Pro-vitamin A carotenoids, such as  $\alpha$ -carotene,  $\beta$ -carotene, and  $\beta$ -cryptoxanthin (Figure 3), are important precursors of retinol (i.e., vitamin A), retinal, retinoid, and retinoic acid in animals and humans [20,21], whereas some xanthophylls, such as lutein and zeaxanthin (Figure 3), are essential components of macular pigments. In addition, carotenoids act as antioxidants. In particular, they can efficiently quench singlet oxygen. This implies energy transfer from the singlet oxygen to the carotenoid molecule, to yield ground-state oxygen and a triplet excited carotenoid molecule; the excited carotenoid can then return to the ground state by dissipating this energy as heat [22]. Chemical quenching of singlet oxygen has also been reported, which leads to oxidation of the carotenoid molecule [23].

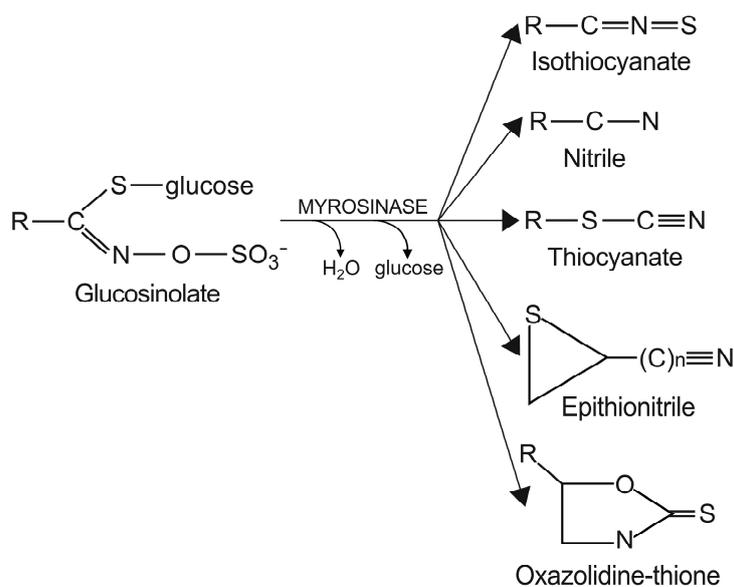
Among the various radicals, carotenoids most efficiently react with peroxy radicals generated during lipid peroxidation. For this reason, and because of their lipophilicity, carotenoids are believed to have important roles in the protection of cell membranes against oxidative damage [22]. The food sources of these compounds include various fruit and vegetables.

### 2.3. Nitrogen-Containing Compounds

Secondary metabolites that contain nitrogen include alkaloids and glucosinolates. Alkaloids represent a large class of nitrogen-containing compounds (with >12,000) where the nitrogen is derived from an amino acid, such as tyrosine, tryptophan, lysine, and aspartate (Figure 1). Compared to the other classes of secondary metabolites, alkaloids have great structural diversity, and for this reason, there is no single uniform classification for these compounds. Most of them are toxic to other organisms, and therefore they have important roles in plants in terms of their defense against pathogen and herbivore attack [24]. Once produced, alkaloids are stored in the cell vacuole and then released when mechanical damage breaks the vacuole. Alkaloids include a number of molecules of pharmacological interest, such as morphine, codeine, atropine, quinine, vincristine, and vinblastine, which have important therapeutic activities that include antibacterial, antiviral, antimalarial, antiproliferative, and antimetastatic effects in various types of cancers [25].

Conversely, there is a nutritional interest in the class of glucosinolates, which is a relatively small, but diverse, group of about 200 compounds. These are substituted esters of thio-amino acids that are mainly limited to species of the order of Brassicales, such as broccoli, cauliflower, cabbage, Brussels sprouts, red radish, horseradish, and mustard [26]. In plants, glucosinolates act as pesticides and as herbivore repellents [27]. Again, these compounds are stored in the cell vacuole, and upon cell disruption they are rapidly hydrolyzed by myrosinase and converted to thiocyanate, isothiocyanate, nitrile, epithionitrile, and oxazolidine-thione (Figure 4), which are toxic for pathogens and herbivores [28].

Based on the structure of the amino-acid precursor, glucosinolates can be defined as aliphatic, indole, and aromatic [29,30] (Table 3). Aliphatic glucosinolates are derived from alanine, leucine, isoleucine, methionine, and valine, while indole glucosinolates and aromatic glucosinolates are derived from tryptophan and phenylalanine or tyrosine, respectively. The presence of glucosinolates varies not only among plant species, organs, and tissues, but also under specific environmental conditions. Of note, nearly all of the potential health benefits of glucosinolates are due to their hydrolytic products and in particular to the isothiocyanates [31]. Mammalian tissues do not contain myrosinases, but degradation of glucosinolates still occurs in humans through the gut microflora. Indeed, several studies have identified gut bacterial strains that are associated with the formation of isothiocyanates from glucosinolates ([32], and references therein).



**Figure 4.** Hydrolysis of glucosinolates catalyzed by myrosinase. Myrosinase catalyzes the hydrolysis of the thioglucoside linkage, thus leading to the formation of different products. R represents the amino-acid-derived side-chain.

**Table 3.** Classification of glucosinolates on the basis of their amino-acid precursor.

Class	Example	Structure
Aliphatic	Sinigrin	
	Glucoraphanin	
	Dehydroerucin	
Indole	Glucobrassicin	
Aromatic	Sinalbin	
	Gluconasturtiin	

### 3. Health Benefits of Secondary Metabolites

Chronic diseases, such as CVD, diabetes, cancer, and neurodegenerative and age-related diseases are worldwide health problems that cause severe disability and even death. Epidemiological studies have shown that due to their significant content of phytochemicals, regular consumption of fruit, vegetables, and grains is strongly associated with reduced risk of these chronic diseases.

#### 3.1. Cardiovascular Diseases

Cardiovascular diseases are the main cause of death and disability in developed countries. Epidemiological studies have shown that there is a strong inverse association between intake of some phytochemicals and incidence of hypertension, myocardial infarction, stroke, and atherosclerosis. In particular, consumption of foods rich in phenols has been shown to have a protective effect upon CVD. This is particularly convincing for flavanols from cocoa-derived products and tea, the consumption of which leads to reduced blood pressure and improved endothelial function [33,34]. Although conclusive data have yet to be realized, promising findings have also been obtained for the effects of resveratrol (from berries, grapes, plums, peanuts) on ischemic heart disease [35], of anthocyanins (from berries, grapes, citrus fruit) on low-density-lipoprotein cholesterol [36], and of quercetin (from onions, apples, berries, grapes, tea) on hypertension and atherosclerosis [37]. As for carotenoids, crocin (from saffron) has proven to be a good candidate for inhibition of platelet aggregation and for protection against the apoptosis of platelets induced by oxidative stress [38], while lycopene (from tomato) appears to improve endothelial function in patients with CVD, and to reduce incidence of coronary heart disease [39]. Finally, there is evidence of the involvement of the dietary intake of isothiocyanates (from *Brassica* species) in improved post-ischemic ventricular function, in reduced myocardial infarct size, and in decreased cardiomyocyte apoptosis in mammals [40]. Allicin (from garlic, onion, shallot, leek) is another organosulfur compound, and it has been shown to protect the cardiovascular system through induction of vasorelaxation, and to alleviate various pathological conditions of CVD [41].

#### 3.2. Obesity

Obesity is associated with increased risk of CVD, type 2 diabetes, and gastrointestinal cancer. At the cellular level, obesity is characterized by increased number and size of adipocytes, which are differentiated from fibroblastic pre-adipocytes in adipose tissue. Evidence has been accumulated that the dietary intake of foods rich in phenols can prevent obesity by induction of pre-adipocyte and adipocyte apoptosis, and by inhibition of adipocyte lipid accumulation. In particular, catechins and epigallocatechin gallates (from green tea), resveratrol, and curcumin (from the curry spice turmeric) can reduce the viability of adipocytes and proliferation of pre-adipocytes, suppress adipocyte differentiation and triglyceride accumulation, stimulate lipolysis and fatty-acid  $\beta$ -oxidation, and reduce inflammation [42]. Moreover, epidemiological studies in humans have associated higher dietary intake and serum levels of carotenoids with decreased adiposity. This is probably due to the anti-adiposity and anti-inflammatory actions of these compounds and their derivatives. Indeed, cell and animal studies have demonstrated that some carotenoids, such as fucoxanthin-, astaxanthin-, and  $\beta$ -carotene-derived retinoids, can suppress adipogenesis and activate lipid oxidation and thermogenesis in mature adipocytes [43]. As for other classes of phytochemicals, animal studies have revealed that an onion extract has anti-obesity effects [44]. In the same study, sulfur-containing components, such as cycloalliin, *S*-methyl-L-cysteine, *S*-propyl-L-cysteine sulfoxide, dimethyl trisulfide, and especially *S*-methyl-L-cysteine sulfoxide, were reported to be effective for inhibition of the formation of oil droplets in cells, which suggests that these compounds are involved in the anti-obesity effects of the onion extract.

### 3.3. Diabetes

Phytochemicals prevent type 2 diabetes through different mechanisms. Experimental and clinical evidence has shown that dietary phenols modulate carbohydrate metabolism through the regulation of  $\alpha$ -amylase and  $\alpha$ -glucosidase activities. This thus protects pancreatic  $\beta$ -cell integrity and physiology against glucose toxicity, reduces postprandial glycemic levels, and enhances insulin releasing activity [45–47]. Due to the high content of ferulic acid, flavonoids, and alkylresorcinols in cereal wholegrain foods, these have been demonstrated to delay or even prevent the development of type 2 diabetes and its complications [48]. Fruit, such as berries, grapes, and pomegranates, vegetables such as artichoke and common bean, and popular drinks like cocoa, coffee, and green tea, are all rich in phenolic compounds, and these have also been shown to have beneficial antidiabetic effects [46,49–51]. For the other phytochemicals, there is evidence that organo-sulfur compounds in onion extracts can decrease lipid levels and increase glucose tolerance [44]. Finally, an association between the intake of foods rich in  $\alpha$ -carotene and  $\beta$ -carotene and reduced risk of diabetes 2 has been reported recently [52].

### 3.4. Cancer

Dietary phytochemicals can have complementary and overlapping protective mechanisms of action against cancers. As for phenolic compounds, phytochemicals have protective effects mainly by acting against UV-radiation-induced inflammation, oxidative stress, DNA damage, and suppression of immune responses. Catechin and proanthocyanidins (from grape seeds, cocoa, apples, peanuts, almonds, cranberries, blueberries) have been shown to be particularly good for protection of the skin from the adverse effects of the UV radiation that causes skin cancers [53]. Anthocyanidins (from berries, cherries) and genistein (from soy) can have anticancer activities through down-regulation of the expression of cancer-related genes [54,55]. Potential beneficial effects of quercetin (from onions, apples, tea) in adjuvant chemotherapy in prostate, breast, colon, and lung cancers have been demonstrated both in vitro and in vivo [56]. Numerous preclinical findings based on human and animal studies have also suggested that resveratrol is a promising tool for cancer prevention and treatment, through its antioxidant activity [57]. As for carotenoids, there is evidence that a diet rich in  $\beta$ -carotene, lycopene, lutein, and zeaxanthin can prevent the incidence of colorectal cancer at an early stage of disease progression [58]. Finally, epidemiological studies have suggested that isothiocyanates are protective against cancers of the lung and alimentary tract [59].

### 3.5. Alzheimer's Disease

Neurodegenerative disorders, and especially Alzheimer's disease (AD), represent a common cause of cognitive decline and dementia, particularly in older people. There is evidence that the dietary intake of foods rich in flavonoids, such as wine, tea, and chocolate, are associated with improved performance for several cognitive abilities [60,61]. Animal studies have demonstrated that at the cellular level, phenolic compounds, such as ferulic acid, myricetin, nordihydroguaiaretic acid, and rosmarinic acid, can prevent the development of AD pathology through reduction of the aggregation of amyloid- $\beta$  protein in the brain parenchyma, which is responsible for the loss of synapses and impairment of neuronal functions [62]. In addition, other studies have suggested that the protective effects of some phenolic compounds against neurodegenerative diseases, such as quercetin, curcumin, resveratrol, and naringenin, are related to their inhibition of acetylcholine esterase activity, which enhances cholinergic transmission [63]. In vitro and in vivo studies have demonstrated that similar to phenolic compounds,  $\beta$ -carotene, lycopene, retinol, and retinoic acid can also prevent AD symptoms, primarily through inhibition of amyloid fibril formation [64–66].

### 3.6. Eye Diseases

Lutein and zeaxanthin constitute the main pigments in the yellow spot of the human retina, which protect the macula from damage by blue light and harmful reactive oxygen species. Lutein

and zeaxanthin are the most common xanthophylls in green leafy vegetables (e.g., broccoli, spinach, collard green, kale, corn, orange pepper, zucchini, and squash), fruit (e.g., kiwi, grapes), and egg yolk, although these are also found at relatively high levels in maize, einkorn, and durum wheat grains and their food products. In the last decade, several epidemiological and clinical studies have demonstrated that consumption of such foods is related to reduced risk of two major age-related eye diseases: macular degeneration and cataracts [67].  $\beta$ -carotene is abundant in vegetables (e.g., carrots, asparagus, sweet potatoes, spinach, kale, pumpkin, squash, peppers, oregano, paprika, parsley) and fruit (e.g., cantaloupes, grapes, plums, apricots), and after its ingestion, it is converted into retinol, or vitamin A, which serves as the precursor of the light sensor molecules in the retina. An adequate intake of  $\beta$ -carotene can prevent degenerative eye damage, such as night blindness, xerophthalmia, Bitot's spot, keratitis, and keratomalacia [68].

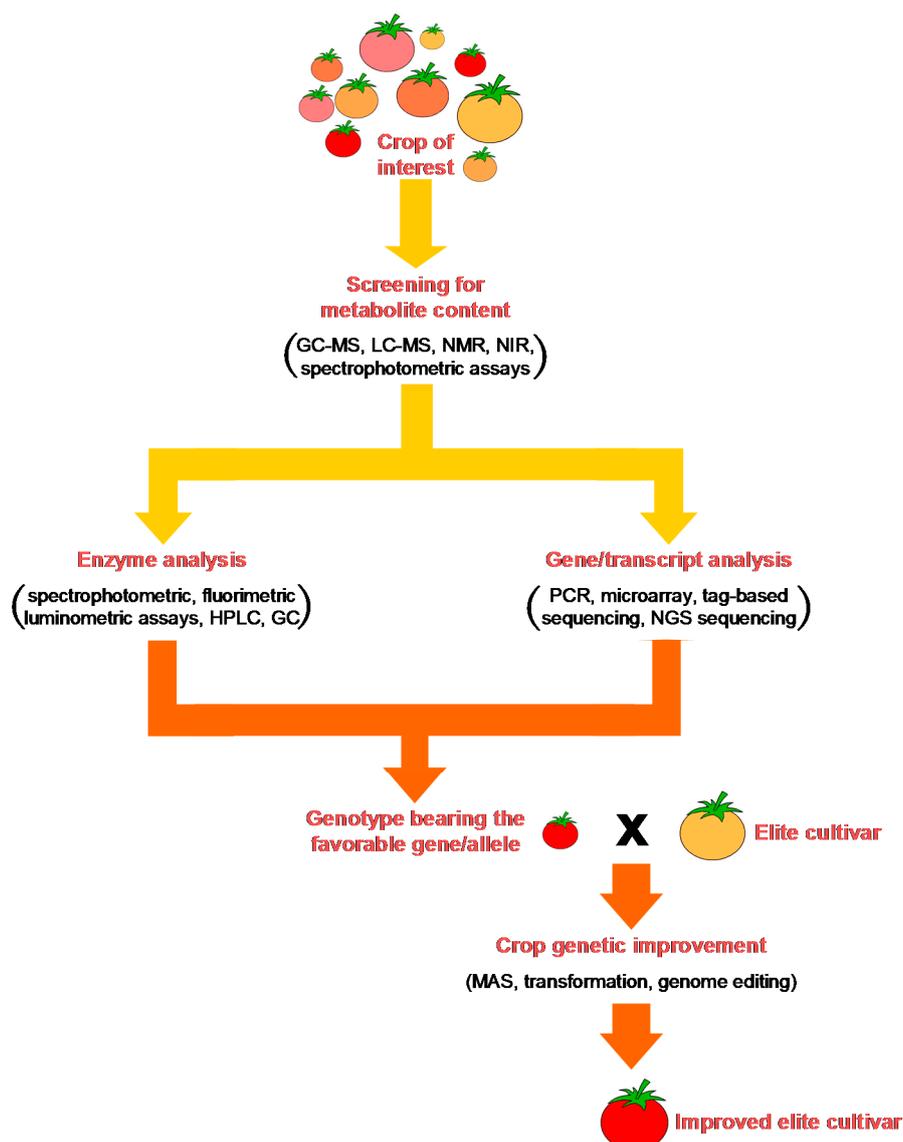
#### 4. Molecular Tools for the Identification of Candidate Genes

A possible strategy to improve the content of a desired phytochemical in an agronomically relevant crop is based on the knowledge of the genes that control its accumulation. Through proper biotechnological breeding approaches, these genes can be used as tools for the selection of new high-quality genotypes in which the trait is stably fixed. Candidate gene can be identified through the general knowledge of the biochemical and molecular regulation that governs the metabolic pathway responsible for accumulation of the desired phytochemical, together with an understanding of the mechanisms that underlie the individual enzymatic reactions that make up the pathway. Indeed, the rationale for function-dependent strategy states that metabolite accumulation represents the outcome of the functioning of key enzymes that catalyze the rate-limiting reactions of the pathway and that regulate its flux rate. Differences in the functioning of a key enzyme can be related to differences in the amount of the enzyme and/or to different variants of the enzyme (known as allozymes), which are characterized by differences in their amino-acid sequences, and consequently, in their kinetic behavior. This is related, in turn, to the presence of different alleles of the gene that encode the key enzyme, which differ from each other in terms of polymorphisms in their regulatory and/or coding regions, such as insertions/deletions (Indels) or single nucleotide polymorphisms (SNPs). A favorable allele that is associated to a more active/efficient allozyme in a biosynthetic pathway or to a less active/efficient allozyme in a catalytic pathway, might represent a valuable candidate that is associated to greater accumulation of a desired phytochemical, and thus should be useful in the selection of new, biofortified genotypes.

In the light of this, it is evident that an understanding of all of the aspects of a metabolic pathway requires multidisciplinary expertise, which includes: (i) analytical skills for identification of the molecules of interest, their precursors, and their degradation products; (ii) biochemical skills for characterization of the enzymes, understanding of their biochemical properties, and identification of allozymes with the desirable performance; (iii) molecular skills for identification and characterization of their encoding genes/alleles and understanding how their expression is regulated. A possible multidisciplinary approach to enhance the production of secondary metabolites in crops is outlined in Figure 5.

Nowadays, the functional approach for identification of candidate genes/alleles can be advanced by taking advantage of publicly available databases and repositories. In particular, there are databases that contain metabolite (ChEBI, GMD, KEGG compound, PubChem) [69–72], reaction (BRENDA, ExPASy-enzyme, KEGG enzyme/KEGG reaction, Rhea BioPAX, Sabio-RK) [73–77] and plant-specific metabolic pathways (e.g., KEGG pathway, MetaCrop, MetaCyc, PANTHER pathway, PlantCyc, Reactome) [78–83] that are available and under continuous updates [84]. In addition, with the recently emerged cheap and immense next-generation sequencing (NGS) technologies, rapid whole transcriptome and genome sequencing that was previously restricted to model plants has been extended to an increasing number of crops. Consequently, plant transcriptome-based databases are available (e.g., PlantransDB, PLEXdb) [85,86] that provide relevant information about genes

expressed in certain plant tissues or organs at any given growth stage and under any particular conditions. As for plant genomes, there are a number of sequencing projects where cultivated species and their wild relatives have been completely sequenced, or are being assembled or are in progress (e.g., leuks.cgi, PlantGDB, Phytozome, and other species-specific databases) [87–90]. All of this heterogeneous information can now be integrated using in-silico approaches to establish a ranking of candidate genes on the basis of their relevance in a given pathway in the species of interest, and/or in closely related species. From these genes, the most promising can be selected for experimental validation through biochemical and molecular approaches.



**Figure 5.** Outline of how a functional approach can contribute to identification of a candidate gene associated with high levels of a health-promoting phytochemical in a staple crop. A germplasm collection is screened to determine the variability in nature of the trait. Genotypes characterized by contrasting metabolite accumulation are subjected to enzymatic analysis to identify key enzyme variants (i.e., allozymes) that are characterized by contrasting activities/efficiencies, and molecular analysis to identify alleles that encode the different allozymes and to evaluate their expression levels. Then, on the basis of both the enzymatic and molecular information, a candidate gene/allele can be identified that can then be efficiently used in advanced breeding programs that are aimed at the introgression of the trait of interest into an elite cultivar.

#### 4.1. Metabolite Analysis

To choose the breeding strategy that is more adapted for improvement of the phytochemical content in a crop of interest, a metabolite analysis has to be performed that determines the natural variability of the trait in the genotypes of a large germplasm collection. To date, the most frequently used approach in plant breeding has been based on targeted analyses that have focused on a restricted group of metabolites. In contrast, the broad, untargeted approaches that allow simultaneous determination of hundreds of compounds is still poorly applied for plant breeding purposes.

Metabolite analyses can be performed using different techniques, such as gas chromatography (GC), liquid chromatography (LC), also as high performance (HPLC) and ultra performance, capillary electrophoresis, near-infrared (NIR) spectroscopy, and nuclear magnetic resonance (NMR). In plant metabolite analyses, GC and LC are often interfaced with mass spectrometry (MS) as the detector, as the most used system at present. Both GC-MS and LC-MS involve separation of metabolites using a chromatographic approach followed by mass separation by MS. For the reproducibility, ease of use, and sensitivity in the nanomolar to picomolar range, GC-MS represents the most widely used technique for determination of plant metabolites [84]. As metabolite volatility is absolutely required for GC, this is the technique of choice for the determination of volatile compounds. GC is also widely used for determination of nonvolatile metabolites, including plant secondary metabolites such as terpenoids and phenolic compounds, which have to be chemically derivatized to become volatile [84]. LC-MS represents a valuable complement to GC-MS. LC is not influenced by volatility and thermal stability of the samples, and combined with MS it can detect different plant secondary metabolites, such as terpenoids, phenolic compounds, glucosinolates, and alkaloids, with a sensitivity in the picomolar to femtomolar range when it is used in combination with electron spray ionization MS. LC-based separations use two different approaches: C18 reverse-phase separation of hydrophobic metabolites; and hydrophilic interaction chromatography separation of hydrophilic metabolites [91]. To improve the separation of complex samples, different parameters can be varied, such as column chemistry, and programmed temperature method for GC and elution method for LC. Recently, an improvement to chromatographic separation was achieved with the use of multi-dimensional separation systems, such as two-dimensional GC (GC  $\times$  GC) [92] and two-dimensional LC (LC  $\times$  LC) [93]. These are based on the use of columns with different chemistries which minimize analyte co-elution and peak overlap in complex samples.

In both GC-MS and LC-MS, the eluent from the chromatographic step is transferred into an ion source where the dissolved metabolites are converted into ions. Different ionization techniques are available, with electron ionization as the most widely used in GC-MS, while the most common technique in LC-MS is electron spray ionization [94]. The mass spectrometer determines the size of the ionized molecule through the mass-to-charge ratio of each detected ion, as well as the number of the individual ions detected. The mass-to-charge ratio is used to identify the metabolite, whereas its quantification is obtained through the number of ions detected, which can then be related back to the metabolite concentration in the plant extract [94]. An additional advantage in MS is that precursor ions of particular mass-to-charge ratios obtained after the first round of MS can be fragmented in a collision gas cell, and the resulting daughter ions can then be separated and detected in a subsequent round of MS. This can be repeated one (MS/MS) or more (MS<sup>n</sup>) times, with the advantage of greatly enhanced specificity of the analysis.

Unlike the chromatographic techniques, NMR spectroscopy is a nondestructive technique that requires minimal sample preparation and no separation processes before metabolite detection. In NMR analysis, samples in deuterated water are placed in a magnetic field and irradiated with a radio frequency. The energy absorbed causes the transition of the hydrogen nuclei from a low-energy to a high-energy state. The subsequent emission of radiation due to the relaxation of the nucleus generates the resonance frequencies or signals, which are recorded and processed to yield an NMR spectrum. This can be related to the molecule structure, whereas the signal intensity is related to the molar concentration of the molecule [95]. Although time-saving, the use of NMR spectroscopy has some

limitations. In particular, this technique requires large amounts of sample, its sensitivity is low (in the micromolar range), and the analysis of complex samples is difficult to achieve, as multiple signals from different metabolites frequently overlap, thus making their identification difficult.

Compared to chromatographic and NMR techniques, NIR spectroscopy is more rapid and cost effective. NIR is a vibrational spectroscopy that uses radiation in the 750 nm to 2500 nm wavelength range to obtain absorption spectra. The qualitative and quantitative information are then acquired through correct instrument calibration. Although NIR spectroscopy is not as sensitive as the chromatographic techniques, this simple method allows rapid, nondestructive analysis of high numbers of samples, and it is thus particularly suitable for plant breeding purposes [96].

Spectrophotometric assays have also been developed to determine the total content of given classes of phytochemicals ([97], and references therein). Although these approaches are much less specific and sensitive than the analytical methods described above, spectrophotometric assays provide a rapid, simple, and inexpensive tool for large-scale screening associated to plant breeding programs. After the extraction of the metabolites using the relevant solvent (e.g., ethanol, methanol, *n*-butanol, acetone), the assays are carried out spectrophotometrically at a wavelength specific for a given class of phytochemicals. Here, total phenolics can be assayed at 725 nm, total anthocyanins at 535 nm, total flavonols at 374 nm, total proanthocyanidins at 500 nm, and total carotenoids at 450 nm.

#### 4.2. Enzyme Analysis

Genotypes with contrasting metabolite contents can be analyzed with respect to the key regulatory enzymes of a pathway. This approach determines their catalytic properties and whether the differences observed in metabolite accumulation are related to the presence in these genotypes of enzymes with contrasting behaviors that differentially influence the flux rate of the pathway. Differences can be observed at two different levels: in the amount of the enzyme, and in its catalytic properties. Differences in the amount of an enzyme are the consequence of differences in the expression of its encoding gene, which, in turn, can be related to the existence of different alleles that are polymorphic in their regulatory regions. Conversely, differences in the catalytic properties are due to the existence of different allozymes, which are encoded by alleles that are polymorphic in their coding regions and that, for this reason, differ in their structure and function. All of the catalytic properties of an enzyme can be assessed through enzymatic assays.

The aim of any enzymatic assay is to follow the time-dependent formation of a product or the consumption of a reagent of the catalyzed reaction. In the simplest case, an enzyme reaction can be observed using a spectrophotometer to monitor the appearance or disappearance of a compound that absorbs in the UV/visible range. Other optical approaches include fluorimetric assays, which are more sensitive than absorbance assays, but which have the disadvantage that only a few compounds emit fluorescence, such as NADH and some artificial substrate analogs [98]. In addition, whereas for spectrophotometric assays the substrate or product concentration can be directly calculated from the absorbance signal through the Lambert-Beer law, in fluorimetric assays calibration curves at different substrate or product concentrations are strictly required. Turbidimetry can also be used to follow the enzymatic degradation of particles [99], while luminometry can be applied to monitor ATP-dependent reactions [100,101]. Enzymatic reactions without any optically active substrate or product cannot be followed directly by means of spectrophotometric or fluorimetric measurements. To render these enzymatic reactions detectable, an enzyme-coupled assay must be carried out. This consists in further converting a reaction product by a second enzyme to form a second product, and so on, until an optically active product is generated that thus produces a detectable signal. As the reaction rate measured must correspond to the rate of the first reaction, the coupled reactions should never become limiting; to achieve this, the enzymes for the coupled reactions should be added in excess compared to the first enzyme to be assayed [98].

The methods described above allow continuous, time-dependent monitoring of enzymatic reactions, which is a fundamental prerequisite for the correct determination of reaction rates. At the

beginning, the signal varies linearly with varying substrate or product concentrations; then, the progress curve deviates from this linearity. Only the linear part of the progress curve should be taken into consideration for the determination of the reaction rate, which is calculated from the slope of the linear part of the progress curve, and is expressed as the amount of substrate consumed or product formed in a given time interval. Alternatively, there are also discontinuous assays (referred to as 'end-point' assays), in which aliquots of the reaction mixture are removed at various times and substrate depletion or product formation is determined by analytical techniques, such as HPLC or GC [102]. To avoid underestimation of the reaction rate, the time points selected for the determinations must fall within the linear portion of the progress curve of the enzymatic assay.

According to the Michaelis-Menten equation, the rate of an enzymatic reaction is hyperbolically related to the substrate concentration [103]. The binding of a substrate to the active sites of an enzyme occurs more efficiently at low concentrations, while with progressive occupation of the active sites, increasing substrate concentrations are required. When all of the active sites are occupied, the enzyme is saturated and reaches its maximal activity ( $V_{\max}$ ), such that any further increase in the substrate concentration will not produce further increases in the reaction rate, which at saturation depends exclusively on the amount of enzyme. According to the SI system, the unit of the enzymatic activity is the katal, whereby 1 katal is defined as the amount of enzyme that converts 1 mol of substrate (or forms 1 mol of product) in 1 s. However, most frequently, enzymatic activity is expressed in international units (IU); 1 IU is the amount of enzyme that converts 1  $\mu\text{mol}$  of substrate (or forms 1  $\mu\text{mol}$  of product) in 1 min [104]. Enzyme activity is strictly influenced by pH, and almost all enzymes have a pH optimum, at which they show their highest activity. This depends on the protonation of the functional groups of the amino acids and on the three-dimensional structure of the enzyme. The turnover number,  $k_{\text{cat}} = V_{\max}/[E_T]$ , can be derived from the enzyme activity, which defines the maximum number of substrate molecules converted to product per unit of time by a single catalytic site. Through the Michaelis-Menten equation, it is also possible to determine the Michaelis constant,  $K_m$ , which corresponds to the substrate concentration that results in half saturation of the enzyme, and which represents the affinity of the enzyme for the substrate: the higher the  $K_m$ , the lower the enzyme affinity for the substrate.

Differences in the amounts of a key enzyme in genotypes with contrasting metabolite contents can lead to different enzymatic activities that can then be detected through differences in  $V_{\max}$ . Here, to unequivocally attribute a different enzyme activity to a different enzyme content, differences in the protein level (e.g., as determined by Western blotting), as well as in the transcript abundance (e.g., determined by expression analysis; see Section 4.3) of its encoding gene should also be detected. In addition, differences in enzyme abundance should have no effect on  $k_{\text{cat}}$ , as this parameter is normalized to the amount of enzyme ( $E_T$ ). When the different metabolite content in contrasting genotypes is due to the presence of different allozymes of a key enzyme, differences in the catalytic parameters are detected that are the consequence of their different amino acid sequences and protein structures. In this case, differences in the  $V_{\max}$  can be observed that are not accompanied by parallel differences in protein and transcript abundances, but rather by differences in  $k_{\text{cat}}$ . In addition, a different substrate affinity can be observed when the amino-acid substitutions affect the catalytic site. Amino-acid substitutions at different protein sites can also affect the pH optimum of an enzyme.

### 4.3. Molecular Analysis

Together with enzymatic analysis, allele mining and expression analysis have to be exploited in contrasting genotypes to identify candidate genes that are potentially responsible for variations in a trait. Once a key regulatory enzyme has been identified and characterized, a search is carried out in the public databases to identify its encoding gene. If the genome of the species of interest has been sequenced and annotated, the gene can be searched for using gene annotation information (e.g., GO, Interpro). Alternatively, non-annotated genes can be identified through a BLAST search by comparing them with known genes or proteins from other plant species. Then, the gene can be

isolated from the contrasting genotypes by polymerase chain reaction (PCR) strategies and sequenced to identify possible allelic variants. Specific PCRs can be set up to discriminate between different alleles. An InDel polymorphism can be detected by amplifying the DNA region surrounding the polymorphism and comparing the size of the PCR product, whereas a SNP can be detected by using allele-specific PCR (AS-PCR), which includes the competitive allele specific PCR (KASP), the cleaved amplified polymorphic sequences (CAPS) and the pyrosequencing. KASP assay is based on a competitive allele-specific primer and a fluorescence-based reporting system that allow the bi-allelic discrimination of polymorphism [105]. CAPS assays rely on differences in the restriction-enzyme digestion patterns of the PCR fragments between the genotypes caused by a SNP [106]. Pyrosequencing is a sequencing-by-synthesis method that is based on the luminometric detection of the pyrophosphate released upon nucleotide incorporation [107].

In addition to gene-sequence isolation and allele discrimination, PCR can also be used to determine gene expression levels. The discovery of reverse transcriptase led to the development of reverse transcription-PCR (RT-PCR), which is the most widely used method for measuring the number of copies of specific transcripts. Starting with a very small amount of material (usually total RNA), it is possible to copy the RNA by reverse transcription to produce the complementary, single-stranded DNA (cDNA), which is much less susceptible to degradation than RNA. The cDNA can then be amplified by PCR, with a semi-quantitative determination of the amount of transcript obtained by measuring the intensity of the bands of the generated amplicons on agarose gels; in the exponential phase this is directly related to the number of copies of the target cDNA in the sample. To ensure that quantification yields reliable results, normalization is generally achieved by comparing the expression levels of the gene of interest with those of a constitutive control gene that is assumed to have invariant levels of expression. This allows correction for the measured transcript levels in terms of variable starting cDNA amounts and differences in the reverse transcriptase efficiency [108]. RT-PCR is widely used because it is neither difficult nor expensive to perform; moreover, it is characterized by high sequence specificity, high sensitivity, and wide dynamic range of detection.

An important development of RT-PCR was that of real-time quantitative RT-PCR (qRT-PCR; often qPCR), which is based on the use of fluorophore-labeled primers that allow the detection of the amplification products as they accumulate [109]. The main advantage of qRT-PCR over traditional RT-PCR assays is that the starting cDNA amount can be determined quantitatively. Moreover, the data are analyzed in a closed-tube system, thus eliminating the need for post-amplification manipulation. Quantification is most commonly achieved using internal control genes (reference genes), which help to reduce possible errors generated by comparisons between the expression levels of the gene of interest and those of the constitutive control gene [110].

The advent of microarray technology enabled SNP detection and gene expression analysis to be expanded from investigations of a few genes to wide-scale analyses, in which the analysis of many genes can be monitored simultaneously. Different microarrays are commercially available; in particular, microarrays manufactured by Affimetrix are the most commonly used for plant studies [111]. Measurement of transcript levels is based on the hybridization events on a flat surface to which specific probes (i.e., cDNA sequences or oligonucleotides) that are complementary to the transcripts of interest have been fixed and immobilized. For SNP detection, the probes are oligonucleotides that distinguish between different alleles. Microarray technology enables large-scale analyses, although it can have some limitations compared to PCR, as it has lower specificity due to cross-hybridization events, and lower dynamic range of detection.

Alternatively, tag-based sequencing methods have been developed, which include serial analysis of gene expression (SAGE), cap analysis of gene expression (CAGE), and massively parallel signature sequencing (MPSS) [112]. Compared to PCR and microarray, tag-based sequencing approaches have the advantage of not depending on genome annotation for probe selection, although they are based on the more expensive Sanger sequencing technology, and isoforms are generally not distinguishable from each other. These disadvantages have been overcome by the recently defined NGS-based sequencing

technologies, such as the Illumina IG, the Applied Biosystems SOLiD, and the Roche 454 Life Science. These platforms can provide information about gene expression, allele-specific expression, novel transcripts, alternative splice sites, and SNPs [113].

These high-throughput approaches are expected to facilitate molecular analysis and expression of genes. However, they are not likely to replace the current RT-PCR, qRT-PCR, and AS-PCR methods, which remain the first choices due to their high sensitivities and specificities, and their low cost. This is particularly the case for researchers who are not equipped for, or do not have expertise to perform, high-throughput analyses, or who are specifically interested in low-throughput applications, which might include detection and quantification of the expression profiles of a few selected genes, such as those that encode enzymes involved in a specific metabolic pathway. Anyway, as indicated above, data obtained using high-throughput gene expression platforms are generally stored in publicly accessible databases, and their consultation has the potential to accelerate candidate gene discoveries.

## 5. Genetics Approaches to Improve Phytochemical Contents in Crops

Marker-assisted selection and genetic transformation are the major tools of the molecular breeding strategies of today. Genome editing is a new technology that promises to contribute greatly to future crop improvements. The choice of one or the other method depends strictly on the genetic variation of the desired trait in the genetic material that is available. MAS is used when the genes/alleles required for the development of new, superior genotypes are already present within the gene pool of the species of interest or in a sexually compatible species. Genetic transformation and genome editing can be used both to introduce native genes/alleles from the gene pool of the crop itself, and to extend the gene pool of the recipient species by introducing genes from non-sexually compatible species. The main advantage of the genetic transformation and genome editing approaches compared to MAS is that they can achieve the same alterations much quicker. Moreover, these advanced approaches allow the targeted improvement of a desired trait without altering the genetic constitution of the recipient genotype, as they do not transfer unwanted genetic material along with the desired gene or allele (i.e., linkage drag).

### 5.1. Marker-Assisted Selection

A DNA marker is a small region of DNA that is tightly linked to the gene of interest and that reveals a mutation in the gene; this can therefore be used to detect polymorphisms between different individuals in a population [114]. Most polymorphisms are in close proximity to the gene of interest, although as indicated above, the polymorphism can also reside within the gene itself, thus generating different alleles at the same locus. As the reliability of the DNA marker to predict the phenotype increases as its distance from the target gene decreases, the use of the gene itself as the DNA marker enables 100% reliable prediction. As indicated above, sequence polymorphisms between different alleles can be due to InDels, or more frequently, to SNPs. Due to their high abundance in the genome, their co-dominant inheritance, and the low cost of their assay, SNPs represent a powerful tool for application to MAS. It is possible to exploit a SNP using AS-PCR, and thus to specifically amplify the allele that controls the trait. In this way, the polymorphism can be converted into a genetic marker that can be easily assayed. The advantages of MAS over conventional breeding are as follows: (i) genotype screening is more objective and often simpler than phenotype screening; (ii) selection is carried out at the early stages of plant growth, thus significantly reducing the number of lines to be tested in the stages that follow; (iii) selection can be made on a single plant basis, rather than on plant families or plots, as necessary for phenotype selection; and (iv) homozygous and heterozygous plants can be distinguished, thus allowing recessive genes to be maintained without the need for progeny tests for each generation.

With regard to the use of MAS in breeding programs that are aimed at improving the phytochemical content in edible crops, most of the applications relate to increases in carotenoid levels in cereal grains. In particular, in maize, PCR-based markers have been developed for the

selection of favorable alleles of *lycopene epsilon cyclase* (*lcyE*), which alters the flux from the  $\alpha$ -carotene versus  $\beta$ -carotene branch of the pathway, and *carotene  $\beta$ -hydroxylase 1* (*crtB1*), which is responsible for  $\beta$ -carotene hydroxylation [115,116]. Four natural polymorphisms of the *lcyE* gene were found to explain more than half of the variation in the two branches, and a three-fold difference in the levels of provitamin A compounds [115]. For the *crtB1* gene, the selection of the most favorable allele in maize elite inbred parents characterized by low kernel  $\beta$ -carotene (at 1.4 mg/g) resulted in a 12-fold increase in  $\beta$ -carotene content in the *crtB1*-introgressed inbreds. In addition, the reconstituted hybrids developed from improved parental inbreds also showed enhanced kernel  $\beta$ -carotene content, compared to the original hybrids (21.7 vs. 2.6 mg/g) [116]. This increase is due to the reduction in transcript expression of the *crtRB1* allele, which decreases the hydroxylation of  $\beta$ -carotene [117]. Allele mining and marker development are also underway for other genes of the carotenoid biosynthetic pathway in maize, including those that encode phytoene synthase and carotenoid cleavage dioxygenases [118].

Promising findings have also been obtained in the identification of genes suitable for MAS of fruit with anthocyanin accumulation. In particular, allele mining has been carried out in apple for the gene that encodes the transcription factor MdMYB1, which regulates anthocyanin accumulation in the fruit peel. The results obtained have highlighted the presence of the *MdMYB1-1* allele that is associated with the red peel color of the fruit, and a CAPS marker was developed for this allele [119]. This was then successfully tested in different segregating populations, in which it provided approximately 80% predictability [120]. A potentially useful marker has also been developed from the InDel in the *MdMYB10* gene, which is responsible for anthocyanin accumulation in the fruit flesh, which is only present in the red fruit core genotypes [121].

## 5.2. Genetic Transformation

Genetic transformation is a process by which a crop species is altered by the incorporation into its genome of foreign genes and regulatory elements. These genetic elements can be manipulated before they are combined, and transferred into a new background, to produce a so-called genetically modified organisms. Only one or a few genes can be transferred by any single transformation event; moreover, neither the number of copies nor the position of the foreign gene in the recipient genome can be controlled. Genes can be moved between any species. Cisgenesis provides for the transformation of a crop species with genes isolated from cross-compatible species, which also include their own introns and promoter and terminator regions, in a sense orientation. In intragenesis, the coding regions and the regulatory elements derive from different genes of a cross-compatible species that have been differently combined. In transgenesis, the genes and their regulatory elements are transferred between non-crossable crop species.

At present, there are two main plant transformation methods: the *Agrobacterium*-mediated method, and particle bombardment. The *Agrobacterium*-mediated method uses *Agrobacterium tumefaciens* to deliver the foreign gene into the host genome. In nature, this soil bacterium causes tumor formation (i.e., crown galls) on a large number of dicotyledonous species, as well as on some monocotyledonous species and gymnosperms, by transferring a specific DNA fragment (the transferred DNA; T-DNA) from its tumor-inducing plasmid to the plant cells [122]. Recombinant *Agrobacterium* strains in which the native T-DNA is removed and replaced with the gene of interest are thus used for the introduction of foreign genes into plants. However, monocotyledons, and in particular cereal crops, are not susceptible to *Agrobacterium*, and therefore they cannot be transformed using this method. Particle bombardment can overcome the host-dependency of the *Agrobacterium*-mediated method. This technique provides for the delivery of the genetic material using highly accelerated gold or tungsten particles, such that they can penetrate through the cell wall. The foreign DNA coated on their surface will then be freed and will recombine with the chromosomal DNA of the target cells [123].

Genetic transformation has been mainly used to improve staple crops with respect to their carotenoid and flavonoid contents. As for carotenoids, enrichment of  $\beta$ -carotene has been achieved in rice, potato, white maize, and cassava, while enrichment in xanthophylls has been achieved in

carrot, tomato, potato, and maize [124]. In particular, the genes that encode the daffodil phytoene synthase (*psy*) and lycopene  $\beta$ -cyclase (*lcy*), together with the gene that encodes bacterial phytoene desaturase (*crtI*), have been transferred to increase the  $\beta$ -carotene content in rice grains to 1.6  $\mu\text{g/g}$ , thus generating the so-called “Golden rice” [125]. Successively, a “Golden Rice 2” was developed by introducing a more efficient maize *phytoene synthase* gene in combination with the bacterial *crtI* used to generate the original “Golden Rice”, thus obtaining an increase in the total carotenoids, as mainly  $\beta$ -carotene, of up to 23-fold, compared to the original “Golden Rice” [126]. Another example is carotenoid improvement in tubers of potatoes transformed with a mini-pathway of bacterial origin, which has given rise to “Golden potatoes”, which have an approximately 20-fold increase in their total carotenoids, and a 3600-fold increase in  $\beta$ -carotene [127]. Alternatively, an increase in carotenoids (of up to 10-fold) has been obtained in tubers by transforming potato plants with the *Or* gene, which encodes a DnaJ cysteine-rich-domain-containing protein that can trigger differentiation of proplastids and/or non-colored plastids into chromoplasts, thus providing a metabolic sink for carotenoid accumulation [128].

For flavonoids, enrichment by transformation has been achieved mainly in tomato, using different approaches. For instance, an up to 78-fold increase in total flavonols in tomato peel was obtained through ectopic expression of petunia *chalcone isomerase*, while in the fruit flesh, the concomitant expression of the sequences that encode chalcone synthase and flavonol synthase from petunia induced increases in flavonol accumulation [129]. Alternatively, ectopic expression of two maize transcription factors, Lc and C1, together with the biosynthetic gene *chalcone isomerase*, resulted in a similar phenotype [129]. Transgenic tomato fruit were also obtained by overexpression of a grapevine gene that encodes the enzyme stilbene synthase, which provides accumulation in the tomato peel of stilbenes, and particularly resveratrol, and other naturally occurring polyphenols [130].

### 5.3. Genome Editing

Site-specific DNA editing using engineered endonucleases has rapidly become a powerful tool for genome editing in diverse eukaryotic organisms, including plants ([131] and references therein). The widespread application of this technology has been largely fueled by the development of the clustered regularly interspaced short palindromic repeats (CRISPR)-associated 9 (Cas9) endonuclease system [132]. CRISPR-Cas are immune systems in bacteria and archaea for protection against foreign nucleic acids [133]. Type II CRISPR-Cas systems contain short direct repeats that are interspaced at regular intervals by sequences incorporated from foreign DNA (protospacers). Transcripts from CRISPRs are processed into short CRISPR RNAs (crRNAs), with each including a protospacer and part of the CRISPR repeat. This crRNA hybridizes with a trans-activating crRNA (tracrRNA) to form a crRNA:tracrRNA duplex, which is, in turn, complexed to Cas9. The protospacer portion of the crRNA guides Cas9 to cleave the complementary foreign DNA sequence if this is adjacent to a short sequence known as a protospacer-adjacent motif.

This system has been adapted to easily and rapidly induce modifications to endogenous genes [132]. In the simplest form of CRISPR-Cas9 technology, the eukaryotic cell is transfected to heterologously express Cas9 and a single guide RNA (sgRNA); this latter consists of a synthetic fusion of crRNA and tracrRNA that is designed to direct Cas9 to cleave a complementary 20-bp sequence that is flanked by the NGG protospacer-adjacent motif sequence. The Cas9-induced double-strand break is repaired via non-homologous end joining. As this non-homologous end joining is an error-prone repair, small insertions and/or deletions can be introduced into the double-strand break, which interrupts the frame, and inactivates the gene. A more precise repair can be achieved by inducing homology-directed repair. With this strategy, the double-strand break is repaired using a donor DNA template that has homology to sequences that flank the target DNA. This enables specific mutations to be introduced, thus generating favorable alleles associated with traits of interest.

Genome editing using the CRISPR-Cas9 system is a new technology, but its use in the genetic improvement of food crops appears to be very promising. Indeed, the first studies have demonstrated

that the CRISPR/Cas9 system is highly efficient in rice for inducing target gene editing, with several genes with diverse functions being tested, including *phytoene desaturase* (*OsPDS*), which is important in carotenoid biosynthesis. Homozygotes of the edited target genes were readily found in the T0 plants, and the gene mutations were passed on to the next generation following classic Mendelian law, without any detectable new mutation or reversion [134]. Similarly, the CRISPR/Cas9 system has been used successfully for targeted modification of the *CsPDS* gene into sweet orange [135]; also in this case, no off-target mutations were observed.

## 6. A Case Study: The *Lipoxygenase-B1* (*Lpx-B1*) Genes and the Regulation of the Final Carotenoid Content in Pasta Products

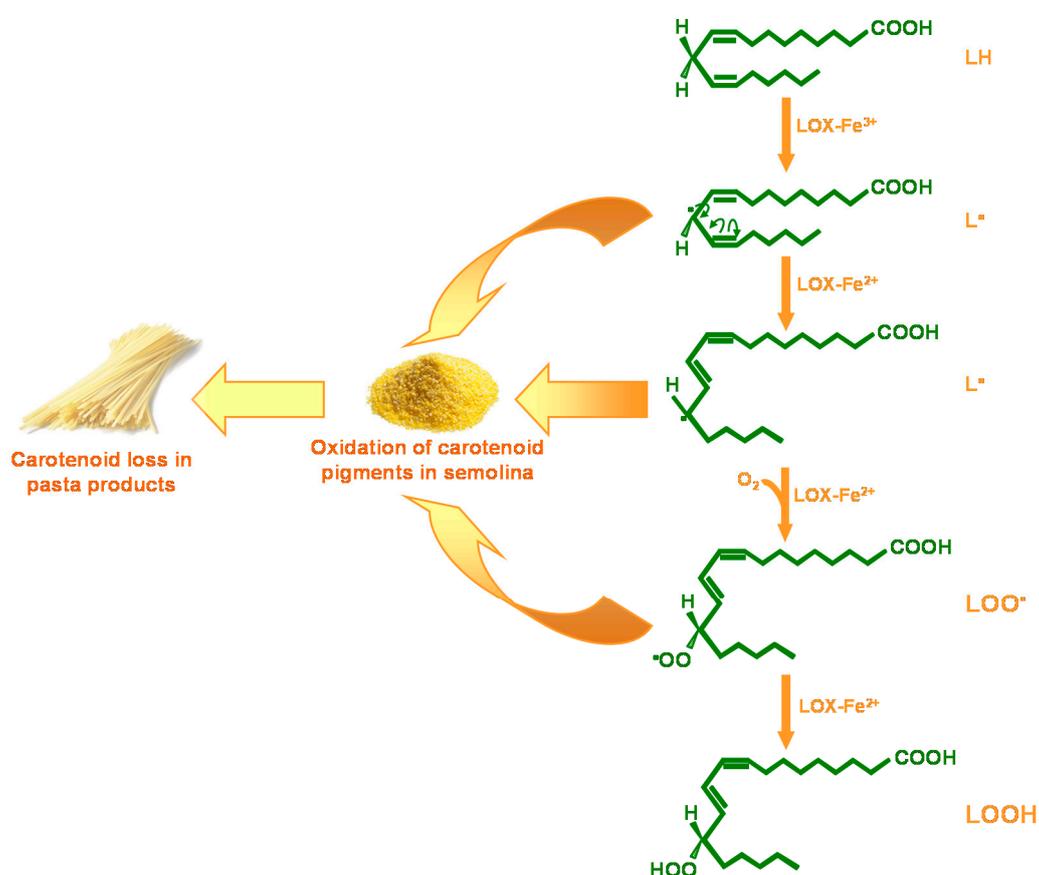
A yellow-amber color represents an important parameter in the assessment of pasta quality. This is due to the presence in the durum wheat endosperm of the carotenoid pigments, which are mainly represented by lutein. Therefore, as well as its role as an esthetic parameter that is strongly appreciated by consumers (who prefer a yellow to amber color, rather than a brown color), the lutein content also represents an important nutritional parameter. Indeed, several findings support the association between dietary intake of foods rich in lutein and lower incidence of degenerative diseases, such as macular degeneration and cancers (see Section 3). Pasta products have a large *per capita* consumption, and they can help in the meeting of the daily recommended dose of lutein (which is still too low) better than other foods. Therefore, genetic improvement of the carotenoid content in pasta products represents a valuable goal to be pursued.

Different analytical procedures have been developed to determine the amounts of carotenoid pigments in durum wheat grain. Sensitive, selective, and accurate determination has been achieved through HPLC techniques. This chromatographic separation can be achieved using either normal or reverse-phase HPLC after extraction of the carotenoids with the appropriate organic solvents (e.g., ethanol, butylated hydroxytoluene, methanol/tetrahydrofuran) [136–140]. This approach was used by Digesù and coworkers [139] to screen a large durum wheat germplasm collection for carotenoid content in the endosperm. Great variability of this trait has been observed in this crop, with a total carotenoid concentration that ranges from 1.178 to 4.416  $\mu\text{g/g}$ .

However, although it is sensitive and accurate, HPLC analysis is costly and time-consuming, and is therefore not suited for large-scale screening. For this reason, other methods based on light spectrophotometry and reflectance measurements have been developed. These methods do not give information about the composition of the carotenoids, but give overall information on the total yellow pigment content. A colorimetric method for determination of the yellow pigment concentration (YPC) has been developed that is based on extraction in water-saturated *n*-butanol and subsequent spectrophotometric measurement of the absorbance of the alcoholic extract at 435.8 nm [141]. Recently, this method was revised and improved to allow carotenoid content measurements for a single seed, and thus to make the screening of large populations easier during breeding programs when the material available is limited [142,143]. Reflectance measurements are obtained using a chromameter (Minolta CR-300; Konica Minolta Pty Ltd., Macquarie Park, NSW, Australia) equipped with a pulsed xenon arc lamp. This provides absolute measures of the  $L^*$  (lightness),  $a^*$  (red-green chromaticity), and  $b^*$  (yellow-blue chromaticity) coordinates of the Munsell color system, using D65 lightning [144]. The  $b^*$  value represents the variation in the yellow intensity, and it is also known as the Yellow index (YI). NIR spectroscopy is also used for yellow color measurements for durum wheat grain. However, although this technique is rapid, non-destructive and cost-effective, the instrument calibration requires a very large sample set, and for this reason it has not been greatly used in durum wheat breeding [145,146]. Conversely, both YPC and YI represent good tests for assessing grain and semolina color, as they are highly correlated with the total carotenoid concentration determined by HPLC analysis (0.94,  $p \leq 0.001$ ; 0.88,  $p \leq 0.001$ ; respectively) [139]. Consistent with this, and in line with HPLC measurements, large variability has also been observed in durum wheat grain for both YPC (from 3.7 to 12.3  $\mu\text{g/g}$ ) [139,147] and YI (from 12.0 to 19.1) [139]. Interestingly, regardless of the method used, modern commercial

genotypes show significantly higher carotenoid content compared to the older genotypes [139,147]. These findings confirm that semolina color is a trait worthy of attention, and attests to the focus of the breeders to improve it.

However, high carotenoid content in the endosperm does not guarantee an equally high content in pasta products. This is because during pasta processing, oxidative degradation of carotenoids can occur, which leads to carotenoid loss and discoloration (bleaching) of the end product. The main enzymes responsible for this degradative phenomenon are the lipoxygenases (LOX, linoleate:oxygen oxidoreductase; EC 1.13.11.12), which are a class of non-heme iron-containing dioxygenases that catalyze the positional and specific dioxygenation of polyunsaturated fatty acids that contain one or more 1,4-*cis*, *cis*-pentadiene structures (e.g., linoleic, linolenic, arachidonic acids), to give the corresponding hydroperoxide (Figure 6). The radicals produced in the intermediate steps of this reaction can cause oxidation of carotenoid pigments in semolina, and consequently, pigment loss in the pasta products.



**Figure 6.** Reaction catalyzed by lipoxygenase (LOX) during pasta processing. LOX catalyzes hydroperoxidation of free linoleate (LH), to give the corresponding hydroperoxide (LOOH). During pasta processing the pentadienyl (L•) and peroxy radicals (LOO•) of linoleate oxidize the semolina carotenoid pigments, thus leading to carotenoid loss and discoloration of the end product.

Both the LOX-catalyzed linoleate hydroperoxidation and carotenoid bleaching reactions can be monitored spectrophotometrically. Linoleate hydroperoxidation activity can be determined by using this fatty acid as the substrate, and following the absorbance increase at 234 nm [148]. This is due to conversion of linoleate into the corresponding hydroperoxide, which has a conjugated diene with maximum absorbance at this wavelength. The bleaching activity is determined in the presence of linoleate and  $\beta$ -carotene as substrates, by monitoring the absorbance decrease at 460 nm that is due to oxidation of  $\beta$ -carotene [148]. In durum wheat semolina, there is large variability with respect to the

LOX activity (see below). This suggests that the durum wheat germplasm has different genes/alleles that contribute to the natural variation of this trait. Among these, the favorable gene/allele is the one associated with very low LOX activity in durum wheat endosperm. As described below, this has been identified and is now used to select superior genotypes in which the carotenoid loss during pasta processing is strongly reduced, thus ensuring high carotenoid levels in the end products.

In cereals, the LOX genes have been isolated from different species, including rice [149,150], maize [151,152], barley [153–155], common wheat [156–159], and durum wheat [160–162]. In particular, the existence of different LOXs has been intensively investigated in barley, which is a species closely related to durum wheat. Three cDNA sequences, *LoxA*, *LoxB*, and *LoxC*, that encode the LOX proteins have been isolated from barley grains [153–155]. The *LoxA* and *LoxC* genes are located on chromosome 4HS and encode, respectively, the LOX-1 and LOX-2 isoforms [163], whereas *LoxB* is located on chromosome 5HL and encodes a LOX isoform that has not been identified to date. LOX-1 is the predominant isoform in mature and quiescent grains, in which it accounts for most of the total LOX activity [154,155].

In durum wheat, the LOX genes are located in regions that are collinear with barley: two *Lpx-2* sequences orthologous to the barley *LoxC* gene have been identified on the group 5 chromosomes, and two *Lpx-3* sequences orthologous to the barley *LoxB* gene have been mapped to the group 4 chromosomes [160]. For the durum wheat genes that are orthologous to barley *LoxA*, a partially deleted copy of *Lpx-1* (*Lpx-A1-like*) has been found on the A genome [158–161], whereas five *Lpx-1* sequences have been identified on the short arm of chromosome 4B [162]. Three sequences correspond to different alleles of the *Lpx-B1.1* gene, referred to as *Lpx-B1.1a*, *Lpx-B1.1b*, and *Lpx-B1.1c*, while the other two sequences correspond to the *Lpx-B1.2* and *Lpx-B1.3* genes. There is also evidence that in durum wheat there is a major QTL that has its peak at the *Lpx-B1* locus; this is associated with most of the LOX activity in semolina [164] and explains a large portion of the variation in pasta color [165]. Similarly, in common wheat, a QTL for LOX activity has been mapped to chromosome 4BS [166,167]. These findings are in line with previous observations on barley LOXs, and they strongly suggest that, in durum wheat, LOXs-1 encoded by genes/alleles at the *Lpx-B1* locus contribute almost exclusively to the total activity detected in semolina, and thus these have a major role in the oxidation of carotenoid pigments that occurs during pasta processing.

The *Lpx-B1* genes and alleles share a common structure, which comprises six exons and seven introns [162]. It is of note that there is a length polymorphism in the last intron of the *Lpx-B1* genes and alleles, due to the presence of a miniature inverted-repeat transposable element (MITE) of the *Stowaway* class [162–164]. This MITE is completely absent in the *Lpx-B1.2* and *Lpx-B1.3* genes, whereas, for the *Lpx-B1.1* gene, there is a complete MITE in the *Lpx-B1.1a* allele that is followed by a partial MITE, and in the *Lpx-B1.1b* allele, there is a partially deleted MITE; in contrast, the *Lpx-B1.1c* allele has a large deletion that covers the region comprised between the second intron and the first part of the last exon [162–164]. These differences are probably due to different insertions/deletions events, as well as large-scale rearrangements, that have affected the *Stowaway*-type MITE and its adjacent sequences, and that in the *Lpx-B1.1c* allele have caused a large deletion, with the consequent loss of function. Based on the distribution of the *Lpx-B1* genes and alleles in the durum wheat germplasm, three different groups of genotypes can be distinguished: haplotype I, which carries the *Lpx-B1.1b* allele associated with the *Lpx-B1.3* gene; and haplotypes II and III that comprise the genotypes that carry the *Lpx-B1.1a* and *Lpx-B1.1c* alleles, respectively, associated with the *Lpx-B1.2* gene [162] (Table 4). As a consequence, haplotypes I and II have two functional LOX-1 isoforms, whereas there is only one functional LOX-1 isoform in haplotype III. Interestingly, only those genotypes released before 1970 included all of these three haplotypes, whereas the genotypes released in the following years included only haplotypes II and III [162] (Table 4). The *Lpx-B1* locus maps in the short arm of the chromosome 4B, near the *Rht-B1b* dwarfing gene [160,162,164], which was introgressed in durum wheat in the 1970s from the common wheat cultivar "Norin 10" [168]. Therefore, the distribution of the *Lpx-B1* genes and

alleles in durum wheat germplasm could reflect the selection carried out on the *RhtB1b* locus from the 1970s onwards.

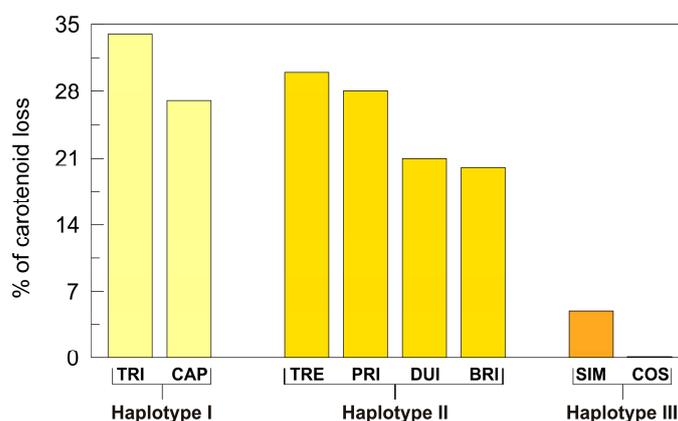
**Table 4.** Molecular and biochemical characteristics of the three *Lpx-B1* haplotypes in durum wheat germplasm.

Haplotype <sup>a</sup>	Year of Release	Gene/Allele <sup>a</sup>	Transcript Abundance <sup>a</sup>	Linoleate Hydroperoxidation Activity (IU/g) <sup>a,b</sup>	$\beta$ -Carotene Bleaching Activity (IU/g) <sup>a,b</sup>
I	Pre-1971	<i>Lpx-B1.1b</i> <i>Lpx-B1.3</i>	High High	2.30–7.91 (4.60)	0.074–0.037 (0.058)
II	1971–1990	<i>Lpx-B1.1a</i> <i>Lpx-B1.2</i>	High Very low	1.32–4.52 (2.59)	0.026–0.043 (0.034)
III	1991–2005	<i>Lpx-B1.1c</i> <i>Lpx-B1.2</i>	High Very low	0.22–0.02 (0.12)	0.007–0.003 (0.005)

<sup>a</sup> Data from [162]; <sup>b</sup> Means are given in brackets.

The *Lpx-B1* genes and alleles differ in their expression levels. All three of the *Lpx-B1.1* alleles and the *Lpx-B1.3* gene are expressed at high levels, whereas the *Lpx-B1.2* gene is expressed at very low levels [162] (Table 4). As a consequence, and taking into consideration only those transcripts that encode functional LOXs, haplotype I is characterized by the highest overall transcript levels, with intermediate and very low expression levels for haplotypes II and III, respectively. This trend is reflected in the LOX activity of the three haplotypes. Indeed, genotypes belonging to haplotype I have the highest LOX activity, when monitored both as linoleate hydroperoxidation and  $\beta$ -carotene bleaching; further, genotypes of haplotype II are characterized by intermediate LOX activity, while haplotype III is characterized by low and very low activity [147,162] (Table 4).

In line with both the molecular and biochemical findings, the loss of carotenoids that occurs during pasta processing is very low in haplotype III genotypes, which have the deleted *Lpx-B1.1c* allele and have very low LOX activity in the endosperm (Figure 7) [169].



**Figure 7.** Carotenoid loss during pasta processing. Cultivars: TRI, Trinakria; CAP, Capeiti 8; TRE, Tresor; PRI, Primadur; BRI, Brindur; SIM, Simeto; COS, Cosmodur. Data from [169].

Although the high LOX activity haplotype I has disappeared as a consequence of the selection for the short straw, there remains large variability in LOX activity in the modern genotypes, with values ranging from 4.22 to 8.97 IU/g semolina [147,162]. Therefore, the selection of haplotype III, which has the deleted *Lpx-B1.1c* allele, might be a useful tool to fix low LOX activity in the semolina of all new genotypes, consequently providing high carotenoid content in the end products. Consistent with this, simple PCR-based marker systems have been set up to test the presence/absence of the deletion and differentiate *Lpx-B1* alleles in MAS programs. For instance, deletion associated at the *Lpx-B1.1c* allele

has been included in the list of markers used in the WheatCAP project (a consortium of more than 20 US breeders), the main objective of which is to use modern selection technologies to increase the competitiveness of public wheat breeding programs [170]. Currently, this marker is included in a MAS program that is aimed at the pyramiding of the genes responsible for disease resistance and qualitative traits into an elite cultivar of durum wheat (PR22D89) [171–173].

## 7. Conclusions

More and more evidence has become available that regular consumption of fruit, vegetables, and whole grains is associated with reduced risk of developing chronic diseases, and that the health benefits of plant-based foods are due to their phytochemical content. However, there is a huge gap between the average intake of these compounds and the recommended doses. For this reason there is strong interest in the nutritional improvement of this plant-based food supply. The combination of the new tools for metabolic and gene expression analyses with classical techniques of enzymatic analysis can be put in use to address the functional identification of the key genes in plant genomes that regulate the accumulation of a desired phytochemical. Although, to date, the whole genome sequence is available for a limited number of plant species, the recently emerged cheap and massive DNA sequencing technologies can be exploited to meet the challenge of genome sequencing also in non-model plants. Together with integration of the increasing databases relating to transcript, metabolite and protein levels, this will allow the decoding of the gene functions, and thus the unravelling of the causal links between specific molecular traits and the level of health-promoting compounds in crops. These insights will aid breeders in the formulation of advanced improvement approaches that are more closely targeted to nutrition-related issues.

**Acknowledgments:** The authors wish to thank Chris Berrie for scientific English language editorial assistance.

**Author Contributions:** Grazia M. Borrelli carried out the analysis of the literature and co-wrote the manuscript; Daniela Trono reviewed the relevant literature, and wrote and supervised the manuscript preparation.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

AD	Alzheimer's disease
AS-PCR	Allele specific PCR
CAGE	Cap analysis of gene expression
CAPS	Cleaved amplified polymorphic sequences
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CVD	Cardiovascular diseases
GC	Gas chromatography
HPLC	High-performance liquid chromatography
InDel	Insertion/deletion
IPP	Isopentenyl pyrophosphate
KASP	Kompetitive allele specific PCR
LC	Liquid chromatography
LOX	Lipoxygenase
<i>Lpx-B1</i>	<i>Lipoxygenase-B1</i>
MAS	Marker-assisted selection
MITE	Miniature inverted-repeat transposable element
MPSS	Massively parallel signature sequencing
MS	Mass spectrometry
NGS	Next-generation sequencing
NIR	Near-infrared
NMR	Nuclear magnetic resonance
PCR	Polymerase chain reaction
RT-PCR	Reverse transcription-PCR
qRT-PCR	Real time quantitative RT-PCR
QTL	Quantitative trait locus
SAGE	Serial analysis of gene expression
sgRNA	Single guide RNA
SNP	Single nucleotide polymorphism
T-DNA	Transferred DNA
tracrRNA	Trans-activating crRNA
YI	Yellow index
YPC	Yellow pigment concentration

## References

1. Hussain, S.; Fareed, S.; Ansari, S.; Rahman, A.; Iffat-Zareen, A.; Saeed, M. Current approaches toward production of secondary plant metabolites. *J. Pharm. Bioall. Sci.* **2012**, *4*, 10–20. [[CrossRef](#)] [[PubMed](#)]
2. Verpoorte, R. Secondary metabolism. In *Metabolic Engineering of Plant Secondary Metabolism*; Verpoorte, R., Alfermann, A.W., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2000; pp. 1–29.
3. Zhang, Y.-J.; Gan, R.-Y.; Li, S.; Zhou, Y.; Li, A.-N.; Xu, D.-P.; Li, H.-B. Antioxidant phytochemicals for the prevention and treatment of chronic diseases. *Molecules* **2015**, *20*, 21138–21156. [[CrossRef](#)] [[PubMed](#)]
4. Tennant, D.R.; Davidson, J.; Day, A.J. Phytonutrient intakes in relation to European fruit and vegetable consumption patterns observed in different food surveys. *Br. J. Nutr.* **2014**, *112*, 1214–1225. [[CrossRef](#)] [[PubMed](#)]
5. Xu, Y.; Crouch, J.H. Marker-assisted selection in plant breeding: From publications to practice. *Crop Sci.* **2008**, *48*, 391–407. [[CrossRef](#)]
6. Newell, C.A. Plant transformation technology. Developments and applications. *Mol. Biotechnol.* **2000**, *16*, 53–65. [[CrossRef](#)]
7. Khatodia, S.; Bhatotia, K.; Passricha, N.; Khurana, S.M.P.; Tuteja, N. The CRISPR/Cas genome-editing tool: Application in improvement of crops. *Front. Plant Sci.* **2016**, *7*, 506. [[CrossRef](#)] [[PubMed](#)]
8. Pflieger, S.; Lefebvre, V.; Causse, M. The candidate gene approach in plant genetics: A review. *Mol. Breed.* **2001**, *7*, 275–291. [[CrossRef](#)]
9. Kabera, J.N.; Semana, E.; Mussa, A.R.; He, X. Plant secondary metabolites: Biosynthesis, classification, function and pharmacological properties. *J. Pharm. Pharmacol.* **2014**, *2*, 377–392.
10. Cheynier, V.; Comte, G.; Davies, K.M.; Lattanzio, V.; Martens, S. Plant phenolics: Recent advances on their biosynthesis, genetics, and ecophysiology. *Plant Physiol. Biochem.* **2013**, *72*, 1–20. [[CrossRef](#)] [[PubMed](#)]
11. Shahidi, F.; Ho, C.-T. Phenolics in food and natural health products: An overview. *ACS Symp. Ser.* **2005**, *909*, 1–8.
12. Cook, N.C.; Samman, S. Flavonoids—Chemistry, metabolism, cardioprotective effects, and dietary sources. *J. Nutr. Biochem.* **1996**, *7*, 66–76. [[CrossRef](#)]
13. Fernandez-Panchon, M.S.; Villano, D.; Troncoso, A.M.; Garcia-Parrilla, M.C. Antioxidant activity of phenolic compounds: From in vitro results to in vivo evidence. *Crit. Rev. Food Sci. Nutr.* **2008**, *48*, 649–671. [[CrossRef](#)] [[PubMed](#)]
14. Han, R.-M.; Zhang, J.-P.; Skibsted, L.H. Reaction dynamics of flavonoids and carotenoids as antioxidants. *Molecules* **2012**, *17*, 2140–2160. [[CrossRef](#)] [[PubMed](#)]
15. Kamal-Eldin, A.; Pours, A.; Eliasson, C.; Åman, P. Alkylresorcinols as antioxidants: Hydrogen donation and peroxy radical-scavenging effects. *J. Sci. Food Agric.* **2000**, *81*, 353–356. [[CrossRef](#)]
16. Vranová, E.; Coman, D.; Grussem, W. Structure and dynamics of the isoprenoid pathway network. *Mol. Plant* **2012**, *5*, 318–333. [[CrossRef](#)] [[PubMed](#)]
17. Dudareva, N.; Pichersky, E.; Gershenzon, J. Biochemistry of plant volatiles. *Plant Physiol.* **2004**, *135*, 1893–1902. [[CrossRef](#)] [[PubMed](#)]
18. Rohmer, M. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat. Prod. Rep.* **1999**, *16*, 565–574. [[CrossRef](#)] [[PubMed](#)]
19. Gershenzon, J.; Dudareva, N. The function of terpene natural products in the natural world. *Nat. Chem. Biol.* **2007**, *3*, 408–414. [[CrossRef](#)] [[PubMed](#)]
20. Fraser, P.D.; Bramley, P. The biosynthesis and nutritional uses of carotenoids. *Prog. Lipid Res.* **2004**, *43*, 228–265. [[CrossRef](#)] [[PubMed](#)]
21. Krinsky, N.I.; Johnson, E.J. Carotenoid actions and their relation to health and disease. *Mol. Asp. Med.* **2005**, *26*, 459–516. [[CrossRef](#)] [[PubMed](#)]
22. Stahl, W.; Sies, H. Antioxidant activity of carotenoids. *Mol. Asp. Med.* **2003**, *24*, 345–351. [[CrossRef](#)]
23. Ramel, F.; Birtic, S.; Cuiné, S.; Triantaphylidès, C.; Ravanat, J.L.; Havaux, M. Chemical quenching of singlet oxygen by carotenoids in plants. *Plant Physiol.* **2012**, *158*, 1267–1278. [[CrossRef](#)] [[PubMed](#)]
24. Bohinc, T.; Ban, S.G.; Ban, D.; Trdan, S. Glucosinolates in plant protection strategies: A review. *Arch. Biol. Sci.* **2012**, *64*, 821–828. [[CrossRef](#)]
25. Schläger, S.; Dräger, B. Exploiting plant alkaloids. *Curr. Opin. Biotechnol.* **2016**, *37*, 155–164. [[CrossRef](#)] [[PubMed](#)]

26. Ishida, M.; Hara, M.; Fukino, N.; Kakizaki, T.; Morimitsu, Y. Glucosinolate metabolism, functionality and breeding for the improvement of *Brassicaceae* vegetables. *Breed. Sci.* **2014**, *64*, 48–59. [[CrossRef](#)] [[PubMed](#)]
27. Agerbirk, N.; Olsen, C.E. Glucosinolate structures in evolution. *Phytochemistry* **2012**, *77*, 16–45. [[CrossRef](#)] [[PubMed](#)]
28. Bones, A.M.; Rossiter, J.T. The enzymic and chemically induced decomposition of glucosinolates. *Phytochemistry* **2006**, *67*, 1053–1067. [[CrossRef](#)] [[PubMed](#)]
29. Cartea, M.E.; Velasco, P.; Obregon, S.; Padilla, G.; de Haro, A. Seasonal variation in glucosinolate content in *Brassica oleracea* crops grown in northwestern Spain. *Phytochemistry* **2008**, *69*, 403–410. [[CrossRef](#)] [[PubMed](#)]
30. Van Eylen, D.; Bellostas, N.; Strobel, B.W.; Oey, I.; Hendricks, M.; van Loey, A.; Sørensen, H.; Sørensen, J.C. Influence of pressure/temperature treatments on glucosinolate conversion in broccoli (*Brassica oleracea* L. cv. *italica*) heads. *Food Chem.* **2009**, *112*, 646–653. [[CrossRef](#)]
31. Dinkova-Kostova, A.T.; Kostov, R.V. Glucosinolates and isothiocyanates in health and disease. *Trends Mol. Med.* **2012**, *18*, 337–347. [[CrossRef](#)] [[PubMed](#)]
32. Li, F.; Hullar, M.A.J.; Beresford, S.A.A.; Lampe, J.A. Variation of glucoraphanin metabolism in vivo and ex vivo by human gut bacteria. *Br. J. Nutr.* **2011**, *106*, 408–416. [[CrossRef](#)] [[PubMed](#)]
33. Habauzit, V.; Morand, C. Evidence for a protective effect of polyphenols-containing foods on cardiovascular health: An update for clinicians. *Ther. Adv. Chronic Dis.* **2012**, *3*, 87–106. [[CrossRef](#)] [[PubMed](#)]
34. De Paula Nogueira, L.; Knibel, M.P.; Torres, M.R.; Nogueira Neto, J.F.; Sanjuliani, A.F. Consumption of high-polyphenol dark chocolate improves endothelial function in individuals with stage 1 hypertension and excess body weight. *Int. J. Hypertens.* **2012**, *2012*, 147321. [[CrossRef](#)]
35. Raj, P.; Zieroth, S.; Netticadan, T. An overview of the efficacy of resveratrol in the management of ischemic heart disease. *Ann. N. Y. Acad. Sci.* **2015**, *1348*, 55–67. [[CrossRef](#)] [[PubMed](#)]
36. Wallace, T.C.; Slavin, M.; Frankenfeld, C.L. Systematic review of anthocyanins and markers of cardiovascular disease. *Nutrients* **2016**, *8*, 32. [[CrossRef](#)] [[PubMed](#)]
37. Gormaz, J.G.; Quintremil, S.; Rodrigo, R. Cardiovascular disease: A target for the pharmacological effects of quercetin. *Curr. Top. Med. Chem.* **2015**, *15*, 1735–1742. [[CrossRef](#)] [[PubMed](#)]
38. Thushara, R.M.; Hemshekhar, M.; Santhosh, M.S.; Jnaneshwari, S.; Nayaka, S.C.; Naveen, S.; Kemparaju, K.; Girish, K.S. Crocin, a dietary additive protects platelets from oxidative stress-induced apoptosis and inhibits platelet aggregation. *Mol. Cell. Biochem.* **2013**, *373*, 73–83. [[CrossRef](#)] [[PubMed](#)]
39. Jacques, P.F.; Lyass, A.; Massaro, J.M.; Vasan, R.S.; D'Agostino, R.B. Relation of lycopene intake and consumption of tomato products to incident cardiovascular disease. *Br. J. Nutr.* **2013**, *110*, 545–551. [[CrossRef](#)] [[PubMed](#)]
40. Mukherjee, S.; Gangopadhyay, H.; Das, D.K. Broccoli: A unique vegetable that protects mammalian hearts through the redox cycling of the thioredoxin superfamily. *J. Agric. Food Chem.* **2008**, *56*, 609–617. [[CrossRef](#)] [[PubMed](#)]
41. Chan, J.; Yuen, A.; Chan, R.; Chan, S.W. A review of the cardiovascular benefits and antioxidant properties of allicin. *Phytother. Res.* **2013**, *27*, 637–646. [[CrossRef](#)] [[PubMed](#)]
42. Wang, S.; Moustaid-Moussa, N.; Chen, L.; Mo, H.; Shastri, A.; Su, R.; Bapat, P.; Kwun, I.; Shen, C.L. Novel insights of dietary polyphenols and obesity. *J. Nutr. Biochem.* **2014**, *25*, 1–18. [[CrossRef](#)] [[PubMed](#)]
43. Bonet, M.L.; Canas, J.A.; Ribot, J.; Palou, A. Carotenoids and their conversion products in the control of adipocytedefunction, adiposity and obesity. *Arch. Biochem. Biophys.* **2015**, *572*, 112–125. [[CrossRef](#)] [[PubMed](#)]
44. Yoshinari, O.; Shiojima, Y.; Igarashi, K. Anti-obesity effects of onion extract in Zucker diabetic fatty rats. *Nutrients* **2012**, *4*, 1518–1526. [[CrossRef](#)] [[PubMed](#)]
45. Chuang, C.C.; McIntosh, M.K. Potential mechanisms by which polyphenol-rich grapes prevent obesity-mediated inflammation and metabolic diseases. *Annu. Rev. Nutr.* **2011**, *31*, 155–176. [[CrossRef](#)] [[PubMed](#)]
46. Tiwari, A.K.; Jyothi, A.L.; Tejeswini, V.B.; Madhusudana, K.; Kumar, D.A.; Zehra, A.; Agawane, S.B. Mitigation of starch and glucose-induced postprandial glycemic excursion in rats by antioxidant-rich green-leafy vegetables juice. *Pharmacogn. Mag.* **2013**, *9S*, 66–73. [[CrossRef](#)] [[PubMed](#)]
47. Dragan, S.; Andrica, F.; Serban, M.C.; Timar, R. Polyphenols-rich natural products for treatment of diabetes. *Curr. Med. Chem.* **2015**, *22*, 14–22. [[CrossRef](#)] [[PubMed](#)]

48. Nettleton, J.A.; McKeown, N.M.; Kanoni, S.; Lemaitre, R.N.; Hivert, M.F.; Ngwa, J.; van Rooij, F.J.A.; Sonestedt, E.; Wojczynski, M.K.; Ye, Z.; et al. Interactions of dietary whole-grain intake with fasting glucose- and insulin-related genetic loci in individuals of European descent: A meta-analysis of 14 cohort studies. *Diabetes Care* **2010**, *33*, 2684–2691. [[CrossRef](#)] [[PubMed](#)]
49. Johnston, K.L.; Clifford, M.N.; Morgan, L.M. Coffee acutely modifies gastrointestinal hormone secretion and glucose tolerance in humans: Glycemic effects of chlorogenic acid and caffeine. *Am. J. Clin. Nutr.* **2003**, *78*, 728–733. [[PubMed](#)]
50. Mellor, D.D.; Madden, L.A.; Smith, K.A.; Kilpatrick, E.S.; Atkin, S.L. High-polyphenol chocolate reduces endothelial dysfunction and oxidative stress during acute transient hyperglycaemia in Type 2 diabetes: A pilot randomized controlled trial. *Diabet. Med.* **2012**, *30*, 478–483. [[CrossRef](#)] [[PubMed](#)]
51. Hokayem, M.; Blond, E.; Vidal, H.; Lambert, K.; Meugnier, E.; Feillet-Coudray, C.; Coudray, C.; Pesenti, S.; Luyton, C.; Lambert-Porcheron, S.; et al. Grape polyphenols prevent fructose induced oxidative stress and insulin resistance in first-degree relatives of type 2 diabetic patients. *Diabetes Care* **2013**, *36*, 1454–1461. [[CrossRef](#)] [[PubMed](#)]
52. Sluijs, I.; Cadier, E.; Beulens, J.W.; van der A, D.L.; Spijkerman, A.M.; van der Schouw, Y.T. Dietary intake of carotenoids and risk of type 2 diabetes. *Nutr. Metab. Cardiovasc. Dis.* **2015**, *25*, 376–381. [[CrossRef](#)] [[PubMed](#)]
53. Nichols, J.A.; Katiyar, S.K. Skin photoprotection by natural polyphenols: Anti-inflammatory, antioxidant and DNA repair mechanisms. *Arch. Dermatol. Res.* **2010**, *302*, 71–83. [[CrossRef](#)] [[PubMed](#)]
54. Kumar, M.; Kumar, S.; Kaur, S. Role of ROS and COX-2/iNOS inhibition in cancer chemoprevention: A review. *Phytochem. Rev.* **2012**, *11*, 309–337. [[CrossRef](#)]
55. Taylor, C.K.; Levy, R.M.; Elliott, J.C.; Burnett, B.P. The effect of genistein aglycone on cancer and cancer risk: A review of in vitro, preclinical, and clinical studies. *Nutr. Rev.* **2009**, *67*, 398–415. [[CrossRef](#)] [[PubMed](#)]
56. Sak, K. Site-Specific anticancer effects of dietary flavonoid quercetin. *Nutr. Cancer* **2014**, *66*, 177–193. [[CrossRef](#)] [[PubMed](#)]
57. Carter, L.G.; D’Orazio, J.A.; Pearson, K.J. Resveratrol and cancer: Focus on in vivo evidence. *Endocr. Relat. Cancer* **2014**, *21*, R209–R225. [[CrossRef](#)] [[PubMed](#)]
58. Jung, S.; Wu, K.; Giovannucci, E.; Spiegelman, D.; Willett, W.C.; Smith-Warner, S.A. Carotenoid intake and risk of colorectal adenomas in a cohort of male health professionals. *Cancer Causes Control* **2013**, *24*, 705–717. [[CrossRef](#)] [[PubMed](#)]
59. Singh, S.V.; Singh, K. Cancer chemoprevention with dietary isothiocyanates mature for clinical translational research. *Carcinogenesis* **2012**, *33*, 1833–1842. [[CrossRef](#)] [[PubMed](#)]
60. Nurk, E.; Refsum, H.; Drevon, C.A.; Tell, G.S.; Nygaard, H.A.; Engedal, K.; Smith, A.D. Intake of flavonoid-rich wine, tea, and chocolate by elderly men and women is associated with better cognitive test performance. *J. Nutr.* **2009**, *139*, 120–127. [[CrossRef](#)] [[PubMed](#)]
61. Beking, K.; Vieira, A. Flavonoid intake and disability-adjusted life years due to Alzheimer’s and related dementias: A population-based study involving twenty-three developed countries. *Public Health Nutr.* **2010**, *13*, 1403–1409. [[CrossRef](#)] [[PubMed](#)]
62. Hamaguchi, T.; Ono, K.; Murase, A.; Yamada, M. Phenolic compounds prevent Alzheimer’s pathology through different effects on the amyloid- $\beta$  aggregation pathway. *Am. J. Pathol.* **2009**, *75*, 2557–2565. [[CrossRef](#)] [[PubMed](#)]
63. Roseiro, L.B.; Rauter, A.P.; Serralheiro, M.L.M. Polyphenols as acetylcholinesterase inhibitors: Structural specificity and impact on human disease. *Nutr. Aging* **2012**, *1*, 99–111.
64. Obulesu, M.; Venu, R.; Somashekhar, R. Tau mediated neurodegeneration: An insight into Alzheimer’s disease pathology. *Neurochem. Res.* **2011**, *36*, 1329–1335. [[CrossRef](#)] [[PubMed](#)]
65. Lerner, A.J.; Gustaw-Rothenberg, K.; Smyth, S.; Casadesus, G. Retinoids for treatment of Alzheimer’s disease. *Biofactors* **2012**, *38*, 84–89. [[CrossRef](#)] [[PubMed](#)]
66. Ono, K.; Yamada, M. Vitamin A and Alzheimer’s disease. *Geriatr. Gerontol. Int.* **2012**, *12*, 180–188. [[CrossRef](#)] [[PubMed](#)]
67. Koushan, K.; Rusovici, R.; Li, W.; Ferguson, L.R.; Chalam, K.V. The role of lutein in eye-related disease. *Nutrients* **2013**, *5*, 1823–1839. [[CrossRef](#)] [[PubMed](#)]
68. Faustino, J.F.; Ribeiro-Silva, A.; Dalto, R.F.; Souza, M.M.; Furtado, J.M.; de Melo Rocha, G.; Alves, M.; Rocha, E.M. Vitamin A and the eye: An old tale for modern times. *Arq. Bras. Oftalmol.* **2016**, *79*, 56–61. [[CrossRef](#)] [[PubMed](#)]

69. ChEBI. Available online: <http://www.ebi.ac.uk/chebi/> (accessed on 25 May 2016).
70. GMD. Available online: <http://gmd.mpimp-golm.mpg.de/> (accessed on 4 May 2016).
71. KEGG Compound. Available online: <http://www.genome.jp/kegg/compound/> (accessed on 25 May 2016).
72. PubChem. Available online: <http://www.ncbi.nlm.nih.gov/pccompound> (accessed on 11 May 2016).
73. BRENDA. Available online: <http://www.brenda-enzymes.org/> (accessed on 29 April 2016).
74. ExPASy-enzyme. Available online: <http://enzyme.expasy.org/> (accessed on 29 April 2016).
75. KEGG enzyme/KEGG Reaction. Available online: <http://www.genome.jp/kegg/reaction/> (accessed on 4 May 2016).
76. Rhea BioPAX. Available online: <http://www.ebi.ac.uk/rhea/> (accessed on 4 May 2016).
77. Sabio-RK. Available online: <http://sabio.villa-bosch.de/> (accessed on 25 May 2016).
78. KEGG pathway. Available online: <http://www.genome.jp/kegg/pathway.html> (accessed on 29 April 2016).
79. MetaCrop. Available online: <http://metacrop.ipk-gatersleben.de/> (accessed on 25 May 2016).
80. MetaCyc. Available online: <http://metacyc.org/> (accessed on 11 May 2016).
81. PANTHER Pathway. Available online: <http://www.pantherdb.org/pathway/> (accessed on 11 May 2016).
82. PlantCyc. Available online: <http://www.plantcyc.org/> (accessed on 4 May 2016).
83. Reactome. Available online: <http://www.reactome.org/> (accessed on 25 May 2016).
84. Hill, C.B.; Roessner, U. Metabolic profiling of plants by GC-MS. In *The Handbook of Plant Metabolomics*; Weckwerth, W., Kahl, G., Eds.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2013; pp. 1–23.
85. PlantransDB. Available online: <http://lifecenter.sgst.cn/plantransdb/index.do> (accessed on 29 April 2016).
86. PLEXdb. Available online: <http://www.plexdb.org/index.php> (accessed on 4 May 2016).
87. leuks.cgi. Available online: <http://www.ncbi.nlm.nih.gov/genomes/> (accessed on 4 May 2016).
88. Plant GDB. Available online: <http://www.plantgdb.org> (accessed on 11 May 2016).
89. Phytozome. Available online: <http://www.phytozome.net> (accessed on 11 May 2016).
90. Pérez-de-Castro, A.M.; Vilanova, S.; Cañizares, J.; Pascual, L.; Blanca, J.M.; Díez, M.J.; Prohens, J.; Picó, B. Application of genomic tools in plant breeding. *Curr. Genom.* **2012**, *13*, 179–195. [[CrossRef](#)] [[PubMed](#)]
91. Hill, C.B.; Roessner, U. Advances in high-throughput LC-MS analysis for plant metabolomics. In *Advanced LC-MS Applications for Metabolomics*; de Vooght-Johnson, R., Ed.; Future Science Ltd.: London, UK, 2015; pp. 58–71.
92. Murray, J.A. Qualitative and quantitative approaches in comprehensive two-dimensional gas chromatography. *J. Chromatogr. A* **2012**, *1261*, 58–68. [[CrossRef](#)] [[PubMed](#)]
93. Jandera, P. Programmed elution in comprehensive two-dimensional liquid chromatography. *J. Chromatogr. A* **2012**, *1255*, 112–129. [[PubMed](#)]
94. De Hoffmann, E.; Stroobant, V. *Mass Spectrometry—Principles and Applications*, 3rd ed.; John Wiley & Sons Ltd.: West Sussex, UK, 2007.
95. Eisenreich, W.; Bacher, A. Advances of high-resolution NMR techniques in the structural and metabolic analysis of plant biochemistry. *Phytochemistry* **2007**, *68*, 2799–2815. [[CrossRef](#)] [[PubMed](#)]
96. Rinnan, Å.; Engelsen, S.B. The shining future of near infrared spectroscopy in plant phenomics, crop sorting and biofuel production. *NIR News* **2016**, *27*, 20–23. [[CrossRef](#)]
97. Mélo, E.A.; Lima, V.L.A.G.; Maciel, M.I.S.; Caetano, A.C.S.; Leal, F.L.L. Polyphenol, ascorbic acid, total carotenoid contents in common fruits and vegetables. *Braz. J. Food Technol.* **2006**, *9*, 89–94.
98. Reymond, J.-L.; Fluxà, V.S.; Maillard, N. Enzyme assays. *Chem. Commun.* **2009**, *1*, 34–46.
99. Bock, R. *Methoden der Analytischen Chemie*; Verlag Chemie: Weinheim, Germany, 1980; Volume 2, pp. 232–238.
100. Campbell, A.K. Living light: Biochemistry, function and biomedical applications. *Essays Biochem.* **1989**, *24*, 41–81. [[PubMed](#)]
101. DeLuca, M.; McElroy, W.D. Purification and properties of firefly luciferase. *Methods Enzymol.* **1978**, *57*, 3–15.
102. Rossomando, E.F. *HPLC in Enzymatic Analysis*, 2nd ed.; John Wiley & Sons, Inc.: New York, NY, USA, 1998; pp. 1–12.
103. Michaelis, L.; Menten, M.L. Kinetik der invertinwirkung. *Biochem. Zeitung* **1913**, *49*, 333–369.
104. Nomenclature Committee of the International Union of Biochemistry. Symbolism and terminology in enzyme kinetics: Recommendations 1981. *Arch. Biochem. Biophys.* **1983**, *224*, 732–740.
105. He, C.; Holme, J.; Anthony, J. SNP genotyping: The KASP assay. *Methods Mol. Biol.* **2014**, *1145*, 75–86. [[PubMed](#)]

106. Konieczny, A.; Ausubel, F.M. A procedure for mapping *Arabidopsis* mutations using codominant ecotype-specific PCR-based markers. *Plant J.* **1993**, *4*, 403–410. [[CrossRef](#)] [[PubMed](#)]
107. Ahmadian, A.; Gharizadeh, B.; Gustafsson, A.C.; Sterky, F.; Nyrén, P.; Uhlén, M.; Lundeberg, J. Single-nucleotide polymorphism analysis by pyrosequencing. *Anal. Biochem.* **2000**, *280*, 103–110. [[CrossRef](#)] [[PubMed](#)]
108. Joice, C. Quantitative RT-PCR. In *RT-PCR Protocols*; O'Connell, J., Ed.; Humana Press Inc.: Totowa, NJ, USA, 2002; Volume 193, pp. 83–92.
109. VanGuilder, H.D.; Vrana, K.E.; Freeman, W.M. Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques* **2008**, *44*, 619–626. [[CrossRef](#)] [[PubMed](#)]
110. Guenin, S.; Mauriat, M.; Pelloux, J.; Van Wuytswinkel, O.; Bellini, C.; Gutierrez, L. Normalization of qRT-PCR data: The necessity of adopting a systematic, experimental conditions-specific, validation of references. *J. Exp. Bot.* **2009**, *60*, 487–493. [[CrossRef](#)] [[PubMed](#)]
111. Galbraith, D.W.; Edwards, J. Applications of microarrays for crop improvement: Here, there, and everywhere. *BioScience* **2010**, *60*, 337–348. [[CrossRef](#)]
112. Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **2009**, *10*, 57–63. [[CrossRef](#)] [[PubMed](#)]
113. Unamba, C.I.; Nag, A.; Sharma, R.K. Next generation sequencing technologies: The doorway to the unexplored genomics of non-model plants. *Front. Plant Sci.* **2015**, *6*, 1074. [[CrossRef](#)] [[PubMed](#)]
114. Collard, B.C.Y.; Mackill, D.J. Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. B* **2008**, *363*, 557–572. [[CrossRef](#)] [[PubMed](#)]
115. Harjes, C.E.; Rocheford, T.R.; Bai, L.; Brutnell, T.P.; Kandianis, C.B.; Sowinski, S.G.; Stapleton, A.E.; Vallabhaneni, R.; Williams, M.; Wurtzel, E.T.; et al. Natural genetic variation in *lycopene epsilon cyclase* tapped for maize biofortification. *Science* **2008**, *319*, 330–333. [[CrossRef](#)] [[PubMed](#)]
116. Muthusamy, V.; Hossain, F.; Thirunavukkarasu, N.; Choudhary, M.; Saha, S.; Bhat, J.S.; Prasanna, B.M.; Gupta, H.S. Development of  $\beta$ -carotene rich maize hybrids through marker-assisted introgression of  *$\beta$ -carotene hydroxylase* allele. *PLoS ONE* **2014**, *9*, e113583. [[CrossRef](#)] [[PubMed](#)]
117. Yan, J.; Kandianis, C.B.; Harjes, C.E.; Bai, L.; Kim, E.; Yang, X.; Skinner, D.; Fu, Z.; Mitchell, S.; Li, Q.; et al. Rare genetic variation at *Zea mays crtRB1* increases  $\beta$ -carotene in maize grain. *Nat. Genet.* **2010**, *42*, 322–327. [[CrossRef](#)] [[PubMed](#)]
118. Xu, Y.; Xie, C.; Wan, J.; He, Z.; Prasanna, B.M. Marker-assisted selection in cereals: Platforms, strategies and examples. In *Cereal Genomics II*; Gupta, P.K., Varshney, R.K., Eds.; Springer Science+Business Media: Dordrecht, The Netherlands, 2013; Chapter 14; pp. 375–411.
119. Takos, A.M.; Jaffé, F.W.; Jacob, S.R.; Bogs, J.; Robinson, S.P.; Walker, A.R. Light-induced expression of a *MYB* gene regulates anthocyanin biosynthesis in red apples. *Plant Physiol.* **2006**, *142*, 1216–1232. [[CrossRef](#)] [[PubMed](#)]
120. Zhu, Y.; Evans, K.; Peace, C. Utility testing of an apple skin color MdMYB1 marker in two progenies. *Mol. Breed.* **2011**, *27*, 525–532. [[CrossRef](#)]
121. Chagné, D.; Carlisle, C.M.; Blond, C.; Volz, R.K.; Whitworth, C.J.; Oraguzie, N.C.; Crowhurst, R.N.; Allan, A.C.; Espley, R.V.; Hellens, R.P.; et al. Mapping a candidate gene (*MdMYB10*) for red flesh and foliage colour in apple. *BMC Genom.* **2007**, *8*, 212. [[CrossRef](#)] [[PubMed](#)]
122. Gelvin, S.B. *Agrobacterium* and plant genes involved in T-DNA transfer and integration. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **2000**, *51*, 223–256. [[CrossRef](#)] [[PubMed](#)]
123. Vasil, V.; Srivastava, A.M.; Castillo, A.M.; From, M.E.; Vasil, I.K. Rapid production of transgenic wheat plants by direct bombardment of cultured immature embryos. *Nat. Biotechnol.* **1993**, *11*, 1553–1558. [[CrossRef](#)]
124. Giuliano, G. Plant carotenoids: Genomics meets multi-gene engineering. *Curr. Opin. Plant Biol.* **2014**, *19*, 111–117. [[CrossRef](#)] [[PubMed](#)]
125. Ye, X.; Al-Babili, S.; Klöti, A.; Zhang, J.; Lucca, P.; Beyer, P.; Potrykus, I. Engineering the provitamin A ( $\beta$ -carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* **2000**, *287*, 303–305. [[CrossRef](#)] [[PubMed](#)]
126. Paine, J.A.; Shipton, C.A.; Chaggar, S.; Howells, R.M.; Kennedy, M.J.; Vernon, G.; Wright, S.Y.; Hinchliffe, E.; Adams, J.L.; Silverstone, A.L.; et al. Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat. Biotechnol.* **2005**, *23*, 482–487. [[CrossRef](#)] [[PubMed](#)]

127. Diretto, G.; Al-Babili, S.; Tavazza, R.; Papacchioli, V.; Beyer, P.; Giuliano, G. Metabolic engineering of potato carotenoid content through tuber-specific overexpression of a bacterial mini-pathway. *PLoS ONE* **2007**, *2*, e350q. [[CrossRef](#)] [[PubMed](#)]
128. Lopez, A.B.; van Eck, J.; Conlin, B.J.; Paolillo, D.J.; O'Neill, J.; Li, L. Effect of the cauliflower *Or* transgene on carotenoid accumulation and chromoplast formation in transgenic potato tubers. *J. Exp. Bot.* **2008**, *59*, 213–223. [[CrossRef](#)] [[PubMed](#)]
129. Verhoeyen, M.E.; Bovy, A.; Collins, G.; Muir, S.; Robinson, S.; de Vos, C.H.; Colliver, S. Increasing antioxidant levels in tomatoes through modification of the flavonoid biosynthetic pathway. *J. Exp. Bot.* **2002**, *53*, 2099–2106. [[CrossRef](#)] [[PubMed](#)]
130. Nicoletti, I.; De Rossi, A.; Giovinazzo, G.; Corradini, D. Identification and quantification of stilbenes in fruits of transgenic tomato plants (*Lycopersicon esculentum* Mill.) by reversed phase HPLC with photodiode array and mass spectrometry detection. *J. Agric. Food Chem.* **2007**, *55*, 3304–3311. [[CrossRef](#)] [[PubMed](#)]
131. Schaeffer, S.M.; Nakata, P.A. CRISPR/Cas9-mediated genome editing and gene replacement in plants: Transitioning from lab to field. *Plant Sci.* **2015**, *240*, 130–142. [[CrossRef](#)] [[PubMed](#)]
132. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J.A.; Charpentier, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **2012**, *337*, 816–821. [[CrossRef](#)] [[PubMed](#)]
133. Wiedenheft, B.; Sternberg, S.H.; Doudna, J.A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **2012**, *482*, 331–338. [[CrossRef](#)] [[PubMed](#)]
134. Zhang, H.; Zhang, J.; Wei, P.; Zhang, B.; Gou, F.; Feng, Z.; Mao, Y.; Yang, L.; Zhang, H.; Xu, N.; et al. The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol. J.* **2014**, *12*, 797–807. [[CrossRef](#)] [[PubMed](#)]
135. Jia, H.; Wang, N. Targeted genome editing of sweet orange using Cas9/sgRNA. *PLoS ONE* **2014**, *9*, e93806. [[CrossRef](#)] [[PubMed](#)]
136. Hentschel, V.; Kranl, K.; Hollmann, J.; Lindhauer, M.G.; Bohmand, V.; Bitsch, R. Spectrophotometric determination of yellow pigment content and evaluation of carotenoids by high-performance liquid chromatography in durum wheat grain. *J. Agric. Food Chem.* **2002**, *50*, 6663–6668. [[CrossRef](#)] [[PubMed](#)]
137. Panfili, G.; Fratianni, A.; Irano, M. Improved normal-phase high performance liquid chromatography procedure for the determination of carotenoids in cereals. *J. Agric. Food Chem.* **2004**, *52*, 6373–6377. [[CrossRef](#)] [[PubMed](#)]
138. Burkhardt, S.; Böhm, V. Development of a new method for the complete extraction of carotenoids from cereals with special reference to durum wheat (*Triticum durum* Desf.). *J. Agric. Food Chem.* **2007**, *55*, 8295–8301. [[CrossRef](#)] [[PubMed](#)]
139. Digesù, A.M.; Platani, C.; Cattivelli, L.; Mangini, G.; Blanco, A. Genetic variability in yellow pigment components in cultivated and wild tetraploid wheats. *J. Cereal Sci.* **2009**, *50*, 210–218. [[CrossRef](#)]
140. Van Hung, P.; Hatcher, D.W. Ultra-performance liquid chromatography (UPLC) quantification of carotenoids in durum wheat: Influence of genotype and environment in relation to the colour of yellow alkaline noodles (YAN). *Food Chem.* **2011**, *125*, 1510–1516. [[CrossRef](#)]
141. AACC International Official Method 14-50.01. *AACC International Approved Methods Analysis*, 11th ed. (Method 14-50 in the 10th ed., 2000, American Association of Cereal Chemists (AACC): St. Paul, MN, USA). 2013. Available online: <http://methods.aaccnet.org/toc.aspx> (accessed on 12 July 2016).
142. Fu, B.X.; Schlichting, L.; Pozniak, C.J.; Singh, A.K. Pigment loss from semolina to dough: Rapid measurement and relationship with pasta colour. *J. Cereal Sci.* **2013**, *57*, 560–566. [[CrossRef](#)]
143. Beleggia, R.; Platani, C.; Nigro, F.; Cattivelli, L. A micromethod for the determination of yellow pigment content in durum wheat. *J. Cereal Sci.* **2010**, *52*, 106–110. [[CrossRef](#)]
144. Commission Internationale de l'Eclairage (CIE). *Colorimetry*, 2nd ed.; Publication 15.2.; Central Bureau: Wien, Austria, 1986.
145. Clarke, F.R.; Clarke, J.M.; McCaig, T.N.; Knox, R.E.; DePauw, R.M. Inheritance of yellow pigment concentration in seven durum wheat crosses. *Can. J. Plant Sci.* **2006**, *86*, 133–141. [[CrossRef](#)]
146. Pozniak, C.J.; Knox, R.E.; Clarke, F.R.; Clarke, J.M. Identification of QTL and association of a phytoene synthase gene with endosperm colour in durum wheat. *Theor. Appl. Genet.* **2007**, *114*, 525–537. [[CrossRef](#)] [[PubMed](#)]

147. De Simone, V.; Menzo, V.; de Leonardis, A.M.; Ficco, D.B.M.; Trono, D.; Cattivelli, L.; de Vita, P. Different mechanisms control lipoxygenase activity in durum wheat kernels. *J. Cereal Sci.* **2010**, *52*, 121–128. [[CrossRef](#)]
148. Pastore, D.; Trono, D.; Padalino, L.; Simone, S.; Valenti, D.; di Fonzo, N.; Passarella, S. Inhibition by  $\alpha$ -tocopherol and L-ascorbate of linoleate hydroperoxidation and  $\beta$ -carotene bleaching activities in durum wheat semolina. *J. Cereal Sci.* **2000**, *31*, 41–54. [[CrossRef](#)]
149. Wang, R.; Shen, W.; Liu, L.; Jiang, L.; Liu, Y.; Su, N.; Wan, J. A novel lipoxygenase gene from developing rice seeds confers dual position specificity and responds to wounding and insect attack. *Plant Mol. Biol.* **2008**, *66*, 401–414. [[CrossRef](#)] [[PubMed](#)]
150. Umate, P. Genome-wide analysis of lipoxygenase gene family in Arabidopsis and rice. *Plant Signal. Behav.* **2011**, *6*, 335–338. [[CrossRef](#)] [[PubMed](#)]
151. Gao, X.; Stumpe, M.; Feussner, I.; Kolomiets, M. A novel plastidial lipoxygenase of maize (*Zea mays*) ZmLOX6 encodes for a fatty acid hydroperoxide lyase and is uniquely regulated by phytohormones and pathogen infection. *Planta* **2008**, *227*, 491–503. [[CrossRef](#)] [[PubMed](#)]
152. Park, Y.S.; Kunze, S.; Ni, X.; Feussner, I.; Kolomiets, M.V. Comparative molecular and biochemical characterization of segmentally duplicated 9-lipoxygenase genes ZmLOX4 and ZmLOX5 of maize. *Planta* **2010**, *231*, 1425–1437. [[CrossRef](#)] [[PubMed](#)]
153. Van Mechelen, J.R.; Smits, M.; Douma, A.C.; Rouster, J.; Cameron-Mills, V.; Heidekamp, F.; Valk, B.E. Primary structure of a lipoxygenase from barley grain as deduced from its cDNA sequence. *Biochim. Biophys. Acta* **1995**, *1254*, 221–225. [[CrossRef](#)]
154. Holtman, W.L.; van Duijn, G.; Sedee, N.J.A.; Douma, A.C. Differential expression of lipoxygenase isoenzymes in embryos of germinating barley. *Plant Physiol.* **1996**, *111*, 569–576. [[PubMed](#)]
155. Van Mechelen, J.R.; Schuurink, R.C.; Smits, M.; Graner, A.; Douma, A.C.; Sedee, N.J.A.; Schmitt, N.F.; Valk, B.E. Molecular characterization of two lipoxygenases from barley. *Plant Mol. Biol.* **1999**, *39*, 1283–1289. [[CrossRef](#)] [[PubMed](#)]
156. Feng, B.; Dong, Z.; Xu, Z.; An, X.; Qin, H.; Wu, N.; Wang, D.; Wang, T. Molecular analysis of lipoxygenase (LOX) genes in common wheat and phylogenetic investigation of LOX proteins from model and crop plants. *J. Cereal Sci.* **2010**, *52*, 387–394. [[CrossRef](#)]
157. Feng, B.; Dong, Z.; Xu, Z.; Wang, D.; Wang, T. Molecular characterization of a novel type of lipoxygenase (LOX) gene from common wheat (*Triticum aestivum* L.). *Mol. Breed.* **2012**, *30*, 113–124. [[CrossRef](#)]
158. Garbus, I.; Soresi, D.; Romero, J.; Echenique, V. Identification, mapping and evolutionary course of wheat *lipoxygenase-1* genes located on the A genome. *J. Cereal Sci.* **2013**, *58*, 298–304. [[CrossRef](#)]
159. Zhang, F.; Chen, F.; Wu, P.; Zhang, N.; Cui, D. Molecular characterization of lipoxygenase genes on chromosome 4BS in Chinese bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **2015**, *128*, 1467–1479. [[CrossRef](#)] [[PubMed](#)]
160. Carrera, A.; Echenique, V.; Zhang, W.; Helguera, M.; Manthey, F.; Schragger, A.; Picca, A.; Cervigni, G.; Dubcovsky, J. A deletion at the *Lpx-B1* locus is associated with low lipoxygenase activity and improved pasta color in durum wheat *Triticum turgidum* ssp. *durum*. *J. Cereal Sci.* **2007**, *45*, 67–77. [[CrossRef](#)]
161. Garbus, I.; Carrera, A.; Dubcovsky, J.; Echenique, V. Physical mapping of durum wheat lipoxygenase. *J. Cereal Sci.* **2009**, *50*, 67–73. [[CrossRef](#)]
162. Verlotta, A.; de Simone, V.; Mastrangelo, A.M.; Cattivelli, L.; Papa, R.; Trono, D. Insight into durum wheat *Lpx-B1*: A small gene family coding for the lipoxygenase responsible for carotenoid bleaching in mature grains. *BMC Plant Biol.* **2010**, *10*, 263. [[CrossRef](#)] [[PubMed](#)]
163. Yang, G.; Schwarz, P.B.; Vick, B.A. Purification and characterization of lipoxygenase isoenzymes in germinating barley. *Cereal Chem.* **1993**, *70*, 589–595.
164. Hessler, G.; Thomson, M.J.; Bensher, D.; Nachit, M.M.; Sorrells, M.E. Association of a lipoxygenase locus, *Lpx-B1*, with variation in lipoxygenase activity in durum wheat seeds. *Crop Sci.* **2002**, *42*, 1695–1700. [[CrossRef](#)]
165. Zhang, W.; Chao, S.; Manthey, F.; Chicaiza, O.; Brevis, J.C.; Echenique, V.; Dubcovsky, J. QTL analysis of pasta quality using a composite microsatellite and SNP map of durum wheat. *Theor. Appl. Genet.* **2008**, *117*, 1361–1377. [[CrossRef](#)] [[PubMed](#)]
166. Geng, H.; Zhang, Y.; He, Z.; Zhang, L.; Appels, R.; Qu, Y.; Xia, X. Molecular markers for tracking variation in lipoxygenase activity in wheat breeding. *Mol. Breed.* **2011**, *28*, 117–126. [[CrossRef](#)]

167. Pshenichnikova, T.A.; Osipova, S.V.; Permyakova, M.D.; Mitrofanova, T.N.; Trufanov, V.A.; Lohwasser, U.; Röder, M.; Börner, A. Mapping of quantitative trait loci (QTL) associated with activity of disulfide reductase and lipoxygenase in grain of bread wheat *Triticum aestivum* L. *Russ. J. Genet.* **2008**, *44*, 567–574. [[CrossRef](#)]
168. Börner, A.; Röder, M.; Korzun, V. Comparative molecular mapping of GA insensitive Rht loci on chromosome 4B and 4D of common wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **1997**, *95*, 1133–1137. [[CrossRef](#)]
169. Trono, D.; Pastore, D.; di Fonzo, N. Carotenoid dependent inhibition of durum wheat lipoxygenase. *J. Cereal Sci.* **1999**, *29*, 99–102. [[CrossRef](#)]
170. WheatCAP. Available online: <http://maswheat.ucdavis.edu/Achievements/markers.htm> (accessed on 13 May 2016).
171. Menzo, V.M. Molecular Markers Validation to Develop a Marker Assisted Selection Programme in Durum Wheat (*Triticum turgidum* L. subsp. *durum* (Desf) Husn.). Ph.D. Thesis, Tuscia University, Viterbo, Italy, 21 March 2011.
172. Menzo, V.M.; Mastrangelo, A.M.; Cattivelli, L.; Papa, R.; de Vita, P. Development of a marker assisted selection program for the improvement of durum wheat (*Triticum durum* Desf.). In Proceedings of the Joint Meeting AGI-SIBV-SIGA, Assisi, Italy, 19–22 September 2011.
173. De Simone, V.; Menzo, V.M.; di Paola, C.; Palumbo, G.; Ferragonio, P.; Mastrangelo, A.M.; Papa, R.; Rubiales, D.; Piarulli, L.; Simeone, R.; et al. Marker-assisted pyramiding of resistance genes and grain quality traits in durum wheat (*Triticum durum* Desf.). In Proceedings of the 57th Italian Society of Agricultural Genetics Annual Congress, Foggia, Italy, 16–19 September 2013.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).