Plant Genetics and Gene Study

Lead Guest Editor: Yuri Shavrukov
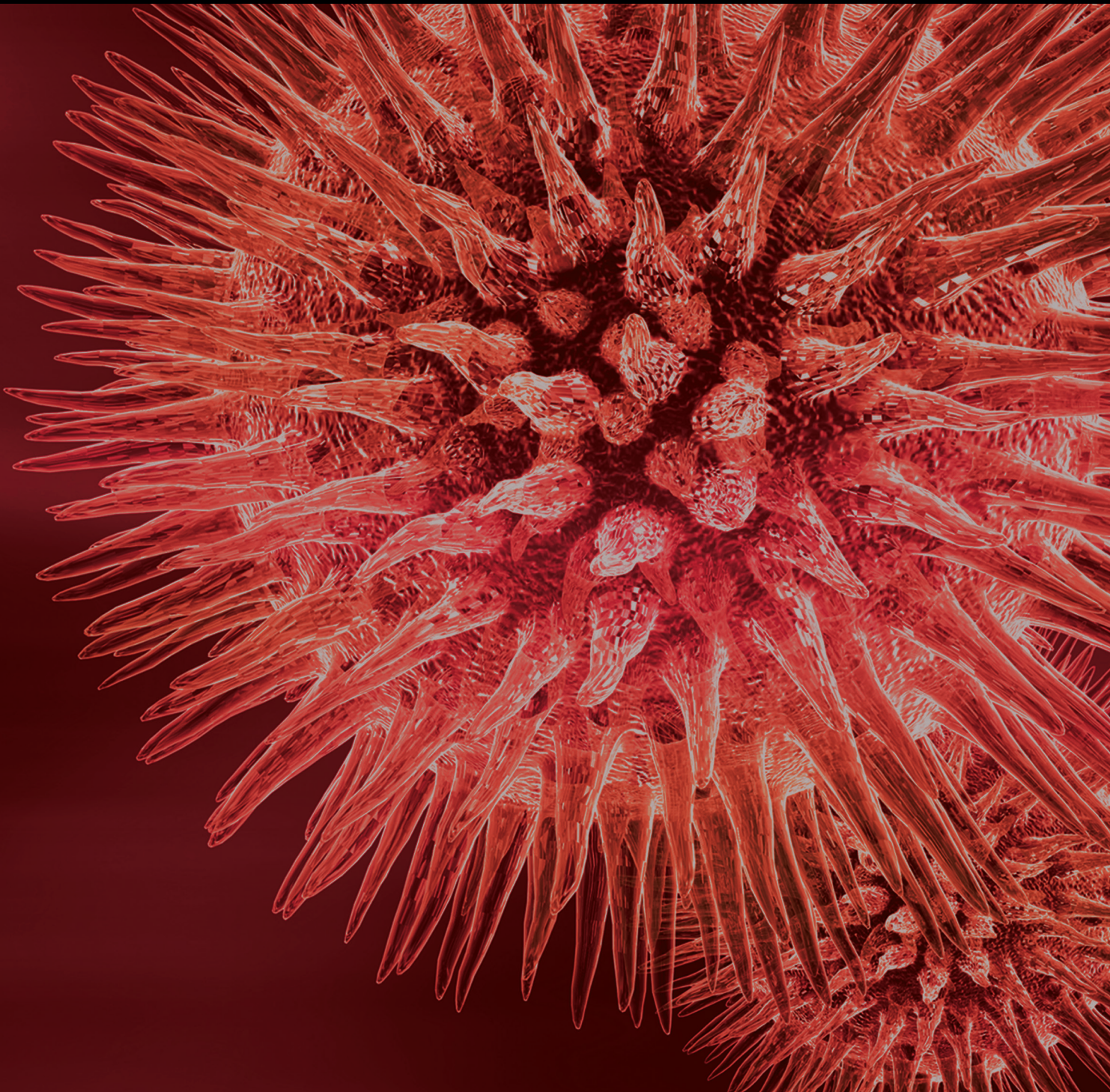Guest Editors: Nikolai Borisjuk and Narendra Gupta

Plant Genetics and Gene Study

Plant Genetics and Gene Study

Lead Guest Editor: Yuri Shavrukov
Guest Editors: Nikolai Borisjuk and Narendra Gupta

Contents

Editorial
Plant Genetics and Gene Study

Yuri Shavrukov [1], Nikolai Borisjuk [2], and Narendra K. Gupta[3]

[1]*Flinders University, College of Science and Engineering, Biological Sciences, Adelaide, Australia*
[2]*School of Life Science, Huaiyin Normal University, Huaian, China*
[3]*Rajasthan Agricultural Research Institute, SKN Agriculture University, Durgapura, Jaipur, India*

Correspondence should be addressed to Yuri Shavrukov; yuri.shavrukov@flinders.edu.au

Plants are the primary source of human food and animal feed and also form the basis of numerous industrial and pharmaceutical products. This special issue reflects the diversity of modern research in the field of plant molecular genetics. It covers a wide range of modern technologies and scientific approaches that aim to achieve a better understanding of the various aspects of molecular mechanisms underpinning the key traits in major crops and other commercially important plant species.

Three papers in this issue deal with the development, study, and application of molecular markers in plant breeding. N. Kim et al. in "Development of Clustered Resistance Gene Analogs-Based Markers of Resistance to *Phytophthora capsici* in Chili Pepper" reported on 11 novel molecular markers targeting resistance to the soil-borne pathogen *Phytophthora capsici* in chili pepper. The markers developed through high-resolution melting analysis, represent an excellent tool for marker-assisted selection. Another type of molecular markers, SSR, was used for the study of genetic diversity in cultivated and wild melon (J. Hu et al. "Microsatellite Markers Reveal Genetic Diversity and Relationships within a Melon Collection Mainly Comprising Asian Cultivated and Wild Germplasms"). The authors found enormous genetic variability within the collection, and deployment of the well-known SSR markers enabled the generation of a phylogenetic tree showing the relationships between melon accessions. An investigation presented by R. Ben Ayed and A. Rebai ("Tunisian Table Olive Oil Traceability and Quality Using SNP Genotyping and Bioinformatics Tools") revealed a significant link between the five SNP markers analysed and the biochemical composition and quality of olive fruits.

Several of the presented reports are related to the genetic control of plant response to abiotic stresses. Drought is one of the major challenges being faced in agriculture. The paper presented by N. M. Kamal et al. ("Stay-Green QTLs Response in Adaptation to Post-Flowering Drought Depends on the Drought Severity") provides important findings on QTLs identified in sorghum grown in Sudan (Africa) under drought conditions. The presented data, generated in cooperation with the International Atomic Energy Agency (IAEA), Austria, can be used for the improvement of grain yield in drought-prone environments using a "stay-green" approach through the application of QTL analysis. Tolerance to aluminum toxicity represents an entirely different type of abiotic stress, but this too is an important agricultural problem, especially in acidic soils. The paper "Aluminum Responsive Genes in Flax (*Linum usitatissimum* L.)" presented by G. S. Krasnov et al. reported valuable results from RNAseq analysis and candidate gene identification for flax genotypes tolerant or sensitive to a high concentration of Al. The authors' conclusions on glutathione metabolism, oxidoreductase, and transmembrane transporters can be further applied both in academic study and in practical breeding. Very few investigations have been conducted on plant growth in the absence of gravity during space-shuttle orbit around our planet. The paper presented by O. Yu. Yurkevich et al. ("Molecular Cytogenetics of *Pisum sativum* L. Grown under Spaceflight-Related Stress") describes a chromosome analysis of pea progenies derived from plants grown in space using a novel FISH approach. Minor chromosome rearrangements were observed in response to "spaceflight-related stress," which could lead to better guidelines for

biological experiments with plants during future space-shuttle missions.

IAEA also supported a study on the generation of mutant bread wheat with improved micronutrient content in the grain. S. Kenzhebayeva et al. presented their paper "Mutant Lines of Spring Wheat with Increased Iron, Zinc, and Micronutrients in Grains and Enhanced Bioavailability for Human Health," where stable breeding lines of wheat were produced from the M_7 generation. The authors reported a two-fold increase of Fe, Zn, and Ca, and reduced levels of the anti-nutrient phytic acid in grains, which can be directly used in practical wheat breeding. In addition to radiation or chemical mutagenesis, inherited changes could be induced by "genomic shock" in plants generated through interspecies hybridization. This phenomenon was studied in rye-wheat hybrids named "secalotriticum" by Y. A. Lipikhina et al. in their paper "Dynamics of the Centromeric Histone *CENH3* Structure in Rye-Wheat Amphidiploids (Secalotriticum)." This is an extremely intricate cytogenetic study of the centromeric nucleosomes in dividing cells. The authors presented some critical results from *CENH3* gene analysis, a central player in this very complex cytological trait, offering promising potential outcomes.

A remarkable level of genetic polymorphism was found in a germplasm collection of *Phaseolus* beans using diverse characteristics derived from the seed morphology of these species. L. Sinkovič et al. presented their paper "Morphological Seed Characterization of Common (*Phaseolus vulgaris* L.) and Runner (*Phaseolus coccineus* L.) Bean Germplasm: A Slovenian Gene Bank Example," which contained an interesting comparison of seeds, accompanied by color images. The presented findings are essential for a better understanding of the wide diversity of bean species, which can be directly used for genetics and breeding of these nutritionally important crop varieties. In contrast, another paper "*In Silico* Genome-Wide Analysis of the ATP-Binding Cassette Transporter Gene Family in Soybean (*Glycine max* L.) and Their Expression Profiling," presented by A. K. Mishra et al., reported on computer analyses of available databases concerning the important gene family of the *ATP-binding cassette transporter*, in soybean. The authors provided a very accurate and detailed analysis of all identified genes and protein isoforms, arranging these new findings within the context of other pertinent information and comparisons.

P. Soundararajan et al. presented a detailed review entitled "Insight on Rosaceae Family with Genome Sequencing and Functional Genomics Perspective." The Rosaceae family is one of the most significant plant families, and the fruits, berries, flowers, and many other important parts of these plants are familiar to each of us from early childhood. The authors have provided a comprehensive overview on the modern level of genetic knowledge and technology employed in comparative genomics and functional genome analysis for many Rosaceae species. Whole genome sequences in this plant family are so important for genetics, breeding, and agriculture that the presented review will surely be of broad general interest to future investigations in this area.

In conclusion, the review presented by N. Borisjuk et al., entitled "Genetic Modification for Wheat Improvement: From Transgenesis to Genome Editing," summarizes the attempts and results of wheat improvement using various genetic engineering approaches. The authors reflect the progress achieved since the dawn of genetic transformation up until the recent emergence of precise modern genome editing using CRISPR/Cas9 and related systems in wheat. This comprehensive review can provide a strong stimulus for both current and future researchers to pursue work on wheat improvement, and deploy all available modern technologies to secure a sustainable supply of quality agricultural produce for our growing global population.

Yuri Shavrukov
Nikolai Borisjuk
Narendra K. Gupta

Conflicts of Interest

The authors declare that the research was conducted in the absence of any potential conflicts of interest.

Research Article

Mutant Lines of Spring Wheat with Increased Iron, Zinc, and Micronutrients in Grains and Enhanced Bioavailability for Human Health

Saule Kenzhebayeva [iD],[1] **Alfia Abekova,**[2] **Saule Atabayeva** [iD],[1] **Gulzira Yernazarova,**[1] **Nargul Omirbekova,**[1] **Guoping Zhang,**[3] **Svetlana Turasheva,**[1] **Saltanat Asrandina,**[1] **Fatma Sarsu,**[4] **and Yarong Wang**[1]

[1] *Department of Biotechnology, Kazakh National University, Named After Al-Farabi, Almaty, Kazakhstan*
[2] *Kazakh Institute of Agricultural and Breeding, Almaty Region, Kazakhstan*
[3] *Agronomy Department, Zhejiang University, Zijinggang Campus, China*
[4] *The Plant Breeding and Genetics Section, Joint FAO/IAEA Division, International Atomic Energy Agency, Vienna, Austria*

Correspondence should be addressed to Saule Kenzhebayeva; kenzhebss@gmail.com

Deficiency of metals, primarily Fe and Zn, affects over half of the world's population. Human diets dominated by cereal products cause micronutrient malnutrition, which is common in many developing countries where populations depend heavily on staple grain crops such as wheat, maize, and rice. Biofortification is one of the most effective approaches to alleviate malnutrition. Genetically stable mutant spring wheat lines (M_7 generation) produced via 100 or 200 Gy gamma treatments to broaden genetic variation for grain nutrients were analyzed for nutritionally important minerals (Ca, Fe, and Zn), their bioavailability, and grain protein content (GPC). Variation was 172.3–883.0 mg/kg for Ca, 40.9–89.0 mg/kg for Fe, and 22.2–89.6 mg/kg for Zn. In mutant lines, among the investigated minerals, the highest increases in concentrations were observed in Fe, Zn, and Ca when compared to the parental cultivar Zhenis. Some mutant lines, mostly in the 100 Gy-derived germplasm, had more than two-fold higher Fe, Zn, and Ca concentrations, lower phytic acid concentration (1.4–2.1-fold), and 6.5–7% higher grain protein content compared to the parent. Variation was detected for the molar ratios of Ca:Phy, Phy:Fe, and Phy:Zn (1.27–10.41, 1.40–5.32, and 1.78–11.78, respectively). The results of this study show how genetic variation generated through radiation can be useful to achieve nutrient biofortification of crops to overcome human malnutrition.

1. Introduction

Nutrient malnutrition represents one of the major health challenges worldwide and is characterized by an increasing number of people manifesting the condition in its varying forms [1–3]. Nearly 30% of humanity—infants, children, adolescents, adults, and older persons in the developing world—are currently suffering from one or more of the multiple forms of malnutrition [2]. Undernutrition and low dietary diversity are by far the biggest risk factors for this global disease and each country faces a serious public health challenge from malnutrition [3]. Iron (Fe) is a biologically essential element for humans, participating in a wide variety of metabolic processes, including oxygen transport, deoxyribonucleic acid (DNA) synthesis, and electron transport [4]. About 70% of the body's iron is bound to hemoglobin in red blood cells and to myoglobin in muscle cells [5]. The remaining iron is bound to other proteins (transferrin in blood or ferritin in bone marrow) or stored in other body tissues. Zinc (Zn) is another essential micronutrient for all living organisms, as it performs both catalytic and structural roles in a wide variety of proteins. One-tenth of the proteome, which is about 3000 human proteins, binds Zn [6, 7]. Disorders of Fe metabolism are among the most common

nutritionally linked diseases of humans and encompass a broad spectrum of diseases with diverse clinical manifestations, ranging from anemia to iron overload, and possibly to neurodegenerative diseases. In terms of global scale and incidence of disease, iron deficiency ranks first and deficiency of zinc is third [8]. Many human disorders are related to Zn deficiency, such as impairment of development and growth, reduced immunity, and disorders of the nervous system [9].

Iron and zinc malnutrition, widely recognized as a major health problem, is predominantly caused by cereal-based diets that are deficient in micronutrients (Zn and Fe) and is prevalent in low-income and middle-income countries [10]. The consequences of malnutrition and nutrition-related diseases include impeded intrauterine growth, which affects 23.8% of all births per year, and protein-energy malnutrition (underweight) in 26.7% of children under-five worldwide, and over 60% of the world's people are Fe deficient and over 25% are Zn deficient [11]. Since Fe and Zn are often derived from the same nutrient-dense food sources in the human diet, lacking these foods generally results in a deficiency of both metals simultaneously.

The populations of the 22 developing countries of the world receive 43-78% of dietary Fe and 56-88% Zn from grains of C_3-species and legumes [10]. Wheat (*Triticum aestivum* L.) is consumed as one of the major human foods and is a source of essential nutrients and protein for nutrition, with demand increasing due to the growing population [2]. Currently, wheat provides 28% of the world's edible dry matter and up to 60% of the daily calorie intake in developing countries [1, 12]. However, genetic gains in grain yield have not changed over recent years and nutritional value and particularly in the Fe and Zn grain protein content have been difficult to improve through traditional breeding [13, 14]. It is generally assumed that selective breeding narrowly focused on increasing yields has resulted in grains with a lower concentration of metals as result of a dilution effect [14, 15]. Furthermore, such a trend may worsen since some reports have shown that the edible portions of food crops grown in fields under elevated atmospheric CO_2 have significantly decreased Fe contents by 4-10% [16].

Biofortification, or the process of genetic enhancement directly targeting the mineral status in the grains of staple food crops through breeding, is one of the most cost-effective and environmentally safe approaches to prevent and alleviate nutrient malnutrition in humans [17–20]. It was demonstrated, for example, that, with the inclusion of biofortified wheat in the human diet, Zn consumption was substantially higher relative to the nonbiofortified diet [21]. In addition, the biofortification of crops through breeding has multiplicative advantages such as long-term and sustainable means of delivering more micronutrients, maintaining improved nutritional status of malnourished people, and rise of the benefits of the initial investment [18].

Along with increasing the concentration of nutrients in food crops, a high bioavailability is also important for human nutrition [22]. Wheat foods are rich in antinutrients, especially phytic acid (Phy), which interferes with the absorption or utilization of nutrients in the digestive system [23]. In general, staple food crops and grains contain very low bioavailable Fe and Zn (i.e., about 5% of the total grain Fe and about 25% of the total Zn are bioavailable). To increase the Fe bioavailability from 5 to 20% it is roughly equivalent to increasing the total Fe four-fold [24]. It has been noted that it is genetically much easier to significantly improve the bioavailability of Fe and Zn in comparison to increasing their total concentration by this magnitude through conventional breeding [23]. Measuring mineral bioavailability in the human diet, their molar ratio with Phy has been widely used [4]. The regulation of Fe status in the human body is controlled through absorption, whereas Zn homeostasis is achieved through absorption but also gastrointestinal secretion and excretion of endogenous Zn [25].

Breeding for low phytic acid concentration is considered to be a reasonable objective to enhance the nutrient bioavailability of crop products. To reduce phytic acid, essential efforts have been made to mutagenize crops. Low phytate mutants (*lpa*) have been reported for several cereals using chemical and physical mutagenesis [26–29]. Successful breeding for yield-associated traits and increases in the nutritional value of cereal foods require genetic variation, which must be distinguishable from environmental effects. The genetic diversity of crops has decreased primarily as a consequence of breeding, including the repeated use of local germplasms and the adoption of breeding schemes that do not favour genetic recombination [27, 28]. Mutagenesis is a powerful tool to broaden genetic variation and has been used for yield increase but has been studied less for the improvement of grain nutritional value [28, 29]. To date, over 3275 mutant varieties in more than 220 plant species have been officially released worldwide [28]. Mutagenesis is especially valuable for inducing novel genetic variation in major crops that have limited genetic variability [29]. Importantly, mutant resources developed in crop breeding are not recognised as genetically modified organisms (GMO) and are freely distributed in all countries without restriction or public concern.

Medical studies have indicated that Phy inhibits Ca absorption, but its effect on Ca bioavailability seems to be less pronounced when compared to that of the bioavailability of Fe and particularly Zn [30]. This is possibly linked to the relatively high Ca concentration in cereal foods, the capability of the bacterial flora in the colon to dephosphorylate Phy, and the intake of Ca from the colon [31].

The nutritional value of crops is highly dependent on grain protein content (GPC), which has a significant impact on the end products [32]. Breeding for improvement of GPC is difficult due to the restricted range of GPC variation in available cultivars.

This study was undertaken with the following objectives: (1) to evaluate the variation in Ca, Fe, Zn, and Phy concentrations and grain protein content in spring wheat parental cv. Zhenis and advanced mutant lines (M_7) produced via 100- and 200 Gy-gamma dose treatments; (2) to estimate the bioavailability of metallic nutrients; (3) to evaluate the

correlations between thousand grain weight (TGW) and quality parameters. The comparison of mineral concentration in cv. Zhenis, 100- and 200 Gy-derived mutant lines, with recommended uptake from flour consumption was also determined.

2. Materials and Methods

2.1. Plant Material. Grains of the spring bread wheat (*Triticum aestivum* L.) cv. Zhenis were irradiated with 100 Gy and 200 Gy doses from a ^{60}Co source at the Kazakh Nuclear Centre, Almaty. After irradiation, seeds were sown to raise M_1 plants [33]. The M_1 generation was grown in the experimental field of the Kazakh Institute of Agriculture and Breeding, Almaty district (43°15′N, 76°54′E, 550 m above sea level). Single spikes from each plant for the M_2 generation were harvested, and the best lines were selected based on the yield of individual plants to continue to the M_7 generation. The number of tillers and spikes per plant varied, but seeds were gathered only from a single main spike. Seeds from the best yielding mutant lines were individually selected in each generation. The selection criteria for these lines included grain weight per main spike (GWS) and per plant (GWP) and these were measured in the M_3 and M_4 generations (2011 and 2012), and compared to the values for the parental cv. Zhenis grown in the same trial conditions. In 2011, the parent had a mean grain weight per main spike of 1.20 ± 0.51 g and grain weight per plant of 1.85 ± 0.61 g. The threshold criteria for selection in the M_4 generation were GWS > 1.4 g and GWP > 2.3 g for the mutant lines. The initial number of lines in the M_1 generation was 2000 each for the 100 Gy and 200 Gy radiation doses. In the M_3 generation, 61 lines (20%) were selected from the 100 Gy irradiation dose population and 48 lines (16%) were selected from the 200 Gy dose. The same numbers of lines for each radiation dose were selected for the M_4-M_6 generations. After harvesting the M_7 plants, 23 lines and 8 lines from the original 100- and 200 Gy-derived germplasm were selected, respectively. The 100 Gy-dosed lines were numbered as follows: 6/9, 10/15, 11/6, 13/9, 15/1, 16/4, 17/7, 18/2, 20/10, 24/21, 26/5, 29/8, 36/13, 37/4, 39/2, 42/4, 43/43, 45/1, 47/2, 49/2, 52/1, 53/5, and 55/10; and 200 Gy-treated lines were numbered 57/4, 58/8, 59/2/, 61/2, 62/2, 63/2, 64/2, and 65/3. These mutant lines, selected from the two different levels of irradiation doses, were then used for further analysis for nutritional quality. Grain samples from each mutant line and the parent cv. Zhenis were sown together in a field trial and plants were grown in three replicates of three row plots, 2 m long x 1.2 m wide, and 20 cm between rows with 30 seeds planted per row for further evaluation. The trial was managed according to locally recommended agronomic practices. Applied fertilizers and their time of use, and soil conditions were as described earlier [33]. No additional fertilizer containing Fe and Zn or other inputs of these metals was applied. Ten randomly selected lines were taken for analysis (five samples per row). To record the three yield-associated traits, thousand-grain weight (TGW) was measured as the mean weight of three sets of 100 grains per line multiplied by 10.

2.2. Determination of Grain Protein Content. Grain protein content was determined with near-infrared reflectance (NIR) spectroscopy on whole grains (ZX50 Portable Grain 174 Analyzer, USA) using proprietary calibration software provided (Zeltex Hagerstown, 175 Ma USA). The measurements of 25 grains per line were repeated in triplicate.

2.3. Analysis of Grain Calcium, Iron and Zinc Concentrations. Grain samples (advanced M_7 mutant lines and parent, cv. Zhenis) were washed with sodium dodecyl sulphate (0.1%), rinsed in deionized water, dried to a constant weight at 65-70°C, and then ground with a mixer mill (Retsch MM400 GmbH). The digestion and extraction of the sample (0.2 g) were as described [33]. Calcium, iron, and zinc concentrations were measured using flame atomic absorption spectroscopies Model NovAA350, AnalytikJena, Jena, Germany. Measurements of all minerals were checked against the certified reference values from the state standard samples LLC "HromLab", Ca-7475-184 98, Fe-7254-96, and Zn-7256-96 diluted by 0.3% HNO_3. Three extracts for analysis were performed.

2.4. Phytic Acid Extraction and Determination, and Molar Ratios of Phy:Metals and Ca:Phy. The extraction of Phy from the milled grain samples (0.3 g) was performed as described in [34]. A volume of 2.5 mL of the supernatant was treated with 2 mL 0.2% $FeCl_3$, and the mixture was boiled for 30 min with further centrifugation after cooling. The residue was washed twice with deionized water. A total of 1.5 M NaOH were added to the precipitation, shaken and the solution was centrifuged. Three mL of 0.5 M HCl was added to the precipitation and then shaken until the precipitation dissolved. The solution was diluted to 25 mL to measure Fe residue by atomic absorption spectrophotometer (AAS, Shimadzu AA6300, Japan). Phy sodium (Sigma St Louis, Missouri, USA) was used to test the Phy recovery rate. The Phy test results suggested that the recoveries were between 96 and 99%. The determination of Phy was based on the precipitation of ferric phytate and measurement of Fe residue in the supernatant. The grain Phy concentration was calculated by multiplying Fe content by a factor of 4.2. To calculate the molar ratios of Ca:Phy, Fe, and Zn, the concentrations of Phy and the metals were converted into moles by dividing by their respective molar masses and atomic weights. The [Ca][Phy]/[Zn] (mol/kg) was also calculated.

2.5. Statistical Analysis. One-way ANOVA was used for comparisons; all data were evaluated in R 3.0.2 (R Core Development Team 2013). Simultaneous tests of general linear hypotheses and Dunnett's contrasts were used for multiple comparisons of the means. Summary data were reported as mean values ± standard deviations. Correlation coefficients between TGW and grain quality parameters and Probability *p*-values were calculated using GenStat software (10th edition). A *p*-value of less than 0.05 was considered statistically significant.

TABLE 1: Comparison between trait means and ranges for spring wheat cv. Zhenis (parent) and M_7 derived from 100 Gy- and 200 Gy-irradiated mutant lines. Data are shown as mean and range (n=93).

Trait	cv. Zhenis		100 Gy-dosed lines		200 Gy-dosed lines	
	Mean	Range	Mean	Range	Mean	Range
TGW (g)	40.75	40.35–41.45	42.46	36.25–57.4	41.15	33.25–52.15
GPC, %	13.00	13.0–13.1	13.85	13.6–14.3	13.34	13.2–13.7
CaC (mg/kg)	357.07	350.6–365.1	568.13	172.30–883.0	530.80	203.4–856.5
FeC (mg/kg)	33.20	31.10–35.10	57.20	40.95–88.83	65.09	44.57–89.03
ZnC (mg/kg)	36.1	33.1–39.3	56.03	22.2–89.60	56.40	22.2–88.10
PhyC (mg/g)	2.59	2.58–2.63	2.23	1.17–2.66	2.47	2.02–2.65
Ca:Phy	2.26	2.24–2.28	4.34	1.39–10.41	3.57	1.27–5.95
Phy:Fe	6.65	6.21–7.19	3.49	1.40–5.32	3.39	2.18–4.90
Phy:Zn	7.16	6.66–7.71	4.48	1.78–11.78	5.59	2.60–11.19
[Ca][Phy]/[Zn] (mol/kg)	0.97	0.92–1.05	1.01	0.21–3.01	1.15	0.30–3.32

Note. Each line was analysed by three replicates. TGW was measured for 15 replicates.

TABLE 2: Comparison of Ca, Fe, Zn, and phytic acid concentrations and Phy:microelements and Ca:Phy molar ratios of advanced M_7 mutant lines and parent grain of spring wheat cv. Zhenis.

Source of variation	cv. Zhenis x 100 Gy-dosed lines	cv. Zhenis x 200 Gy-dosed lines	100 Gy- x 200 Gy-dosed lines
Df	83	38	92
CaC	23.13 * * *	28.89 * *	0.74
FeC	64.97 * * *	203.83 * *	7.15 *
ZnC	21.87 * * *	26.98 * *	0.01
GPC	647.78 * * *	149.29 * * *	244.41 * * *
PhyC	16.20 * * *	19.62 *	9.79 * * *
Ca:Phy	27.18 * * *	36.50 * * *	4.49 *
Phy Fe	169.05 * * *	468.55 * * *	0.20
Phy:Zn	29.50 * * *	10.08	4.29

Note. Data are presented as a percentage of the total sum from ANOVA analysis. The lines were significantly different from the parental line. Asterisks, *, **, and * * *, denote significance at the $P < 0.05$, 0.01, and 0.001 level, respectively.

3. Results

3.1. Variability in Ca, Fe, and Zn Concentrations in Grain of Spring Wheat M_7 Mutant Lines and Parent. Significant differences in Ca, Fe, and Zn concentrations were found among the new spring wheat M7 mutant lines developed using dose radiation of 100 and 200 Gy and the parent cv. Zhenis. Table 1 and Supplementary Table S1 show the means and ranges of the parameters. The CaC varied from 172.3-883.0 mg/kg in mutant lines (n = 93). Significantly enhanced CaC exceeded the parent by 1.23 to 2.47-fold, with the highest mean recorded in 100 Gy-treated lines identified in 15 M_7 lines (48.4%).

Significant variation was also found for the microelements Fe and Zn between mutant lines derived from 100 and 200 Gy-irradiated lines, with means of 40.95-89.03 mg/kg and 22.2-89.6 mg/kg (n = 93), respectively. Significantly higher FeC and ZnC than the parent by 1.23 to 2.66- and 1.45-2.42-fold, respectively, were identified in 16 (51.6%) and 17 (54.8%) M_7 lines. The GPC varied from 13.1 to 14.3% with a mean of 13.72 ± 0.26% (n = 93). Seventeen genotypes (54.8%), from the 100 Gy-dosed lines, had a significantly 6.5 to 7.6% higher GPC relative to the parent. The mutant lines exhibited

wide variations in PhyC, from 1.17 to 2.66 mg/g (Table 1 and Supplementary Table S1). When compared to the parent, significantly lower PhyC by 1.25 to 2.02-fold was detected in 9 mutant lines (29%).

Analysis of variance (ANOVA) for differences in all nutrient concentrations among cv. Zhenis and mutant lines is shown in Table 2. These results revealed significant differences between the cv. Zhenis and 100 Gy- and 200 Gy-mutant lines for all traits except that of Phy:Zn for cv. Zhenis and the 200 Gy-treated lines (Table 2). However, the interactions between the 100 Gy- and 200 Gy-dosed lines were significant in terms of nutrient concentrations of FeC, GPC, and PhyC. For metals bioavailability, significant correlation among the 100 Gy- and 200 Gy-dosed lines was revealed for the Ca:Phy molar ratios. The radiation effect of 100 Gy was highest in GPC, indicating its increased efficiency to generate mutations in the genome associated with this trait. Traits such as FeC and therefore Phy:Fe showed greater variation in 200 Gy than 100 Gy treatments, showing that the most influence for their improvement was through the higher level of radiation.

Significant variations in the molar ratios of Phy:metals (Ca, Fe, and Zn) were also noted between the parent and mutant lines (Table 2). Variation was detected for the Ca:Phy,

TABLE 3: R^2 correlation coefficients between nutrient concentration, phytic acid, and thousand grains weight, with p-values denoted by asterisks.

Parameters	TGW	GPC	FeC	ZnC	CaC
cv. Zhenis					
GPC, %	0.024				
FeC (mg/kg)	0.002	0.000			
ZnC (mg/kg)	0.622∗ ∗ ∗	0.017	0.34∗		
CaC (mg/kg)	0.011	0.000	0.98∗ ∗ ∗	0.052	
PhyC (mg/g)	0.454∗∗	0.0128	0.504∗∗	0.972∗ ∗ ∗	0.646∗ ∗ ∗
100 Gy-derived lines					
GPC, %	0.01				
FeC (mg/kg)	0.081∗	0.014			
ZnC (mg/kg)	0.084∗	0.000	0.11∗∗		
CaC (mg/kg)	0.005	0.025	0.045	0.063∗	
PhyC (mg/g)	0.092∗∗	0.012	0.080∗	0.043	0.021
200 Gy-derived lines					
GPC, %	0.005				
FeC (mg/kg)	0.220∗	0.011			
ZnC (mg/kg)	0.009	0.012	0.35∗∗		
CaC (mg/kg)	0.03	0.014	0.036	0.004	
PhyC (mg/g)	0.323∗∗	0.254∗	0.273∗∗	0.027	0.011

Note. The lines were significantly different from the parent line. Each line was analyzed by three replicates. Asterisks, ∗, ∗∗, and ∗ ∗ ∗, denote significance at the $P < 0.05$, 0.01, and 0.001 probability level, respectively.

Phy:Fe, and Phy:Zn molar ratios, (1.27-10.41, 1.40-5.32, and 1.78-11.78, respectively) (Table 1 and Supplementary Table S1). The lowest means of these characteristics and therefore their highest bioavailability were present in the 100-Gy-derived mutant lines. The most significant noticeable variation in metal bioavailability between the parent and mutant lines was found for the Phy:Fe molar ratios, followed by Ca:Phy and Phy:Zn (Table 2). It was also observed that the 100 Gy-treated lines are significantly differed from the ones developed by the 200 Gy treatment in the Ca:Phy molar ratios (Table 2).

3.2. Correlations between Nutrient Concentration, Phytic Acid, and Thousand-Grain Weight. To examine the relationships between nutrient concentrations, Phy content and yield, a correlation analysis (R^2) was conducted; data are presented in Table 3. In the parent, several traits showed consistent correlations between each other. For instance, among the metals, a highly significant relationship was detected between CaC and FeC and to a lesser degree between FeC and ZnC. The correlations between PhyC and the metal concentrations were highly significant for ZnC, CaC, and FeC.

Generally, the correlations between the investigated traits were found to be lower for the 100 Gy- and 200 Gy-derived lines than the cv. Zhenis. There was a significant and positive association between FeC and ZnC (r^2 = 0.11-0.35, respectively, $p < 0.01$) (Table 3). We revealed significant and positive associations between FeC and TGW in the 100 Gy- and 200 Gy-derived lines, but not with the parent, with the highest mean in the 200 Gy-dosed lines. The distinctive significant feature in the 200 Gy-dosed lines was the correlation between PhyC with TGW, GPC, and also FeC, but not with ZnC.

TABLE 4: Association between Phy:nutrients and Phy:Ca molar ratios and thousand grain weight and grain protein content for cv. Zhenis spring wheat and advanced mutant lines (100 Gy- and 200 Gy-derived).

Trait	Phy:Fe	Phy:Zn	Ca:Phy
cv. Zhenis			
TGW, g	0.007	0.759∗ ∗ ∗	0.126
GPC, %	0.00	0.013	0.002
100 Gy-derived lines			
TGW, g	0.086∗	0.02	0.03
GPC, %	0.025	0.005	0.00
200 Gy-derived lines			
TGW, g	0.222∗	0.033	0.003
GPC, %	0.0012	0.141∗	0.014

Note. The lines were significantly different from the parent line. Each line was analyzed by three replicates. Asterisks, ∗, ∗∗, and ∗ ∗ ∗, denote significance at the $p < 0.05$, 0.01, and 0.001 probability level, respectively.

The relations between Phy:microelements and Ca:Phy molar ratios and TGW and GPC were also analysed (Table 4). In the parent cv. Zhenis, there was only a significant relation between TGW and Phy:Zn. The 100 Gy- and 200 Gy-derived lines presented significant correlations between TGW and the Phy:Fe molar ratio, with the mean more than two times higher in 200 Gy-dosed lines. These results indicate highly possible simultaneous improvement of Fe bioavailability with the spring wheat productivity component. A significant relationship between the Phy:Zn molar ratios and GPC was only detected in the 200 Gy-treated mutant lines.

3.3. *Estimated Nutrient Bioavailability in Wheat Flours.* To determine whether the highest means of mineral concentrations from the parent and mutant lines provided the required daily intake of Fe, Zn, and Ca, we calculated the ratio of grain mineral concentration from 200 g flour consumption to the percentage of the recommended uptake (Table 5). These calculations were based on the statistics from the FAO [35], where the mean consumption of wheat flour is about 200 g per person per day, and on the values for the recommended intake for adults according to the DGE (German Nutrition Society) [36]. The results obtained in this study indicated that, in the case of the parent cv. Zhenis, Fe deficiency was highly manifested. The highest concentration of these minerals from the mutant lines that produced whole grain flour was revealed to provide 1.28-1.32-fold more than the required daily intake of Fe, Zn, and Mg. The Ca concentrations in the grain of the parent and mutant lines supplied were 5.05-12.5 and 1.46-2.44-fold higher, respectively, when compared to the required daily consumption of these minerals.

4. Discussion

This study reported the production of genetically stable advanced (M_7) mutant lines of spring wheat derived from 100 Gy and 200 Gy irradiation treatment showing exceptionally high concentrations of nutritionally important nutrients (Ca, Fe, Zn, and GPC) with accompanying analysis of their bioavailability (Table 1 and Supplementary Table S1). The variation in grain nutrient content was 172.3-883.0 mg/kg for CaC, 31.1-89.0 mg/kg for FeC, and 22.2–89.6 mg/kg for ZnC. Of the minerals investigated, the greatest increase in concentration in the mutant lines compared to the parent was found in FeC, followed by CaC and ZnC with means of 2.66, 2.47, and 2.42, respectively. 17 genotypes (54.8%), of the 100 Gy-derived lines had a significant 6.5 to 7.6% higher GPC relative to the parent.

The concentrations of minerals in the wheat mutant lines exceeded those already reported earlier for hexaploid wheat by 2.35-2.96-fold for FeC [14, 15, 37–39]. Although environmental factors can influence grain metal concentrations, in this work, all the mutant lines and the parent were grown under the same field conditions and were treated equally, with no specific fertilizer supplementation or other inputs of these metals. Studies that compared historical and modern wheat cultivars for the evaluation of grain yield and concentration of Ca, Cu, Fe, Zn, Mg, Mn, P, and Se reported that, over time, the concentrations of all minerals, except Ca, decreased, while grain yield increased [14, 15]. Therefore, this situation suggests that a greater consumption of wheat bread from modern cultivars is required to achieve the same percentage of recommended dietary allowance levels as was provided by older cultivars with lower yield.

Micronutrient intake from wheat is essentially determined by the amount available for human absorption. High micronutrient bioavailability can be achieved by the reduction of antinutritional agents. The Phy level is considered to be one of the most important causative factors limiting metal bioavailability through chelation [17, 20, 22]. However, it was recently suggested that sulphur containing peptides rather than Phy bind to Zn in barley [40]. The function of Phy is as a phosphorus and energy store, and a source of cations and myoinositol could be improved by decreasing PhyC.

Breeding for low PhyC is a reasonable objective to enhance nutrient bioavailability in grain. To reduce PhyC content, essential efforts have been made to mutagenize crops. Low phytate cereal mutants (*lpa*) have been reported using chemical and physical mutagenesis [26, 41, 42]. In wheat, the *lpa* mutant was isolated by chemical mutagenesis [43]. At the same time, *lpa* mutations in several crops usually lead to pleiotropic effects on plant and seed performance, such as reduced germination and emergence rate, lower seed filling, and susceptibility to stress [42]. Our study showed that low Phy wheat lines generated from cv. Zhenis by 100 Gy and 200 Gy radiation did not display any differences in seed viability or shoot and root growth when compared with the parent (data not shown).

The variation found for PhyC ranged from 1.17 to 2.66 mg/g (2.3-fold variation) in all of the mutant lines (Table 1 and Supplementary Table S1) with the lowest PhyC recorded in the 100 Gy-derived line numbered 24(21), which was two times lower relative to that of the parent. There were significant differences between the 100 Gy- and the 200 Gy-derived mutant lines (Table 2). In addition, a search for low Phy lines among the M_7 mutant lines generated by 100 Gy on the genetic background of cv. Erythrospermum 35 allowed us to identify the lowest PhyC, which was 3.5-fold lower than the parent [33].

Studies of natural wheat variation revealed huge differences in PhyC content. The range of variation found was of 5.9-45.4 mg/g [44–47]. However, it is possible that these findings are inconsistent due to differences in the methodology employed for determination of Phy levels, as indicated by Gibson et al. [48], stressing that selection of the most appropriate method for Phy analysis is critical.

The potential bioavailability of nutrients for human consumption is estimated by Phy:metal molar ratios, or *vice versa* for the microelements. In general, low molar ratio means high mineral bioavailability and the same conversely. In the current study, significant variability for the Ca:Phy, Phy:Fe, and Phy:Zn molar ratios between the parent and mutant lines was detected (Table 1 and Supplementary Table S1). Among them, the most pronounced variation was for Ca:Phy (more than 7-fold) and for Phy:Zn (around 7-fold) in the 100 Gy-dosed mutant lines. A high level of variation for Phy:Fe (1.40-5.32) with a mean of about a 4-fold difference was also detected. Therefore, it seems clear that the genetic variability available in mutant lines of spring wheat is enough to warrant their use as resources in breeding. To significantly increase Fe and Zn absorption Phy:Fe molar ratios was estimated at <1 or preferably <0.4 [49] and for the Phy:Zn molar ratio, <5 was considered high Zn bioavailability, corresponding to approximately 50% of the total Zn [50]. Phy:Fe ratios with a reported range similar to that obtained in the current study (1.96-3.86) were reported in 12 bread wheat varieties [51]. Higher Phy:Fe molar ratios have been reported in the literature, such as means of around 12 in two bread wheat cultivars [52] and of 15.5-31.3 in a set of nine bread wheat

TABLE 5: Comparison of mineral concentration in cv. Zhenis, 100 Gy-, and 200 Gy-derived lines with the recommended uptake of minerals (mg/day) according to DCE, percentage of recommended uptake from flour consumption of 200 g/person/day (mg/day), mg minerals from 200 g flour consumption obtained from max values of genotypes, and ratio of grain mineral concentration from percentage of recommended uptake.

Mineral concentration (mg/kg)	Genotypes		Recommended uptake of minerals	Percentage of recommended uptake	mg minerals from 200 g flour consumption			Minerals mg from percentage of recommended uptake	Ratio of grain minerals concentration		
	cv. Zhenis	100 Gy-derived lines			cv. Zhenis	100 Gy-derived lines	200 Gy-derived lines		cv. Zhenis	100 Gy-derived lines	200 Gy-derived lines
Fe	31.3	88.8	10	76	6.3	17.8	17.8	13.5	0.35	1.32	1.32
Zn	36.1	89.6	10	78	7.2	17.9	17.62	14.0	0.40	1.28	1.26
Ca	357.1	883.0	1000	8	71.4	176.6	171.3	14.1	5.05	12.50	12.12

varieties [53]. Concerning the Phy:Zn molar ratio, a range of 23.9-41.4 was found in 65 bread wheat varieties from Pakistan [54], and higher means of 29-178 were revealed in bread wheat [55].

In medical studies, the molar ratio of [Ca][Phy]/[Zn] is a better indicator of Zn bioavailability, as Ca strengthens the effect of Phy on Zn absorption due to the existence of a kinetic synergism between Ca and Zn ions that results in the formation of a more insoluble Ca:Zn:Phy complex when compared to the Phy complexes formed by either alone. Therefore, the [Ca][Phytate]/[Zn] molar ratio is a better index for predicting Zn bioavailability than the Phy:Zn ratio because of this Ca-Phy interaction.

A [Ca][Phy]/[Zn] molar ratio greater than 0.5 mol/kg may reduce Zn bioavailability [56]. The considerable variation found for the [Ca][Phytate]/[Zn] molar ratio (0.21-3.32) in the current study in mutant lines (Table 1) was below the critical level. This means that there was a 11.0-14.3-fold variation across the lines, and the ones with the lowest values for [Ca][Phy]/[Zn] fall in the category of high Zn bioavailability according to the designation suggested by the authors [56].

5. Conclusions

The data generated in the present study has shown considerable variation in nutrients (Ca, Fe, Zn, and GPC) and Phy in new spring wheat mutant lines (M_7) that were derived from 100 and 200 Gy treatment of the parent cv. Zhenis. The results showed that among these mutant resources a great number of lines (8 genotypes) have a significantly higher Phy:Zn and Ca concentration than that of the parent. Of these, 4 lines also recorded simultaneously high bioavailability of Zn, Fe, and Ca. In addition to these valuable characteristics of grain nutritional quality, the line 6/9 also recorded a high GPC. The only correlation between TGW and nutrient content was found for concentrations of Fe in the mutant lines; in addition, Ca and Phy contents were correlated, and a significant correlation existed between GPC and the concentration of Phy. Thus, consumption of whole wheat bread produced from these new mutant lines could contribute a higher percentage of recommended dietary allowance levels of these essential nutrients. The promising mutant lines identified could be useful to generate mutant varieties with appropriate levels of bioavailable metals, which can lead to the development of variety-based products rich in the desired minerals to overcome deficiencies in human intake.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Authors' Contributions

Saule Kenzhebayeva, Fatma Sarsu, and Guoping Zhang conceived and designed the experiments; Alfia Abekova, Nargul Omirbekova, Svetlana Turasheva, and Gulzira Yernazarova performed the experiments; Saule Kenzhebayeva, Fatma Sarsu, and Saltanat Asrandina analyzed the data and wrote the paper.

Acknowledgments

Supplementary Materials

Brief description of Supplementary Table S1: variability of grain traits such as concentrations of Fe, Zn, Ca and phytic acid (Phy), grain protein content (GPC), molar ratios of phytic acid to Fe (Phy:Fe), Zn (Phy:Zn), Ca (Ca:Phy), and grain thousand weight (TGW) in spring wheat parental cv. Zhenis and 100- and 200-Gy derived M7 mutant lines. The traits were measured in triplicates in parent and each mutant line. *(Supplementary Materials)*

References

[1] FAO, *Faostat*, 2014, http://www.fao.org/home/en/.

[2] *Global Nutrition Report. from Promise to Impact. Ending Malnutrition by 2030*, 2016.

[3] *Nutrition for Health and Development: Progress Report (NHD). A Global Agenda for Combating Malnutrition*, 2000.

[4] N. Abbaspour, R. Hurrell, and R. Kelishadi, "Review on iron and its importance for human health," *Journal of Research in Medical Sciences*, vol. 19, no. 2, pp. 164–174, 2014.

[5] D. C. Gupta, "Role of Iron (Fe) in Body," *IOSR Journal of Applied Chemistry*, vol. 7, no. 11, pp. 38–46, 2014.

[6] C. Andreini, L. Banci, I. Bertini, and A. Rosato, "Counting the zinc-proteins encoded in the human genome," *Journal of Proteome Research*, vol. 5, no. 1, pp. 196–201, 2006.

[7] W. Maret, "New perspectives of zinc coordination environments in proteins," *Journal of Inorganic Biochemistry*, vol. 111, pp. 110–116, 2012.

[8] WHO, *Research for Universal Health Coverage, World Health Report, no. 199*, 2013, http://www.who.int/whr/en.

[9] WHO, *Childhood and Maternal under Nutrition*, 2002, http://www.who.int/whr/2002/chapter4/en/index3.html.

[10] R. E. Black, C. G. Victora, S. P. Walker et al., "Maternal and child undernutrition and overweight in low-income and middle-income countries," *The Lancet*, vol. 382, no. 9890, pp. 427–451, 2013.

[11] W. Maret and H. H. Sandstead, "Zinc requirements and the risks and benefits of zinc supplementation," *Journal of Trace Elements in Medicine and Biology*, vol. 20, no. 1, pp. 3–18, 2006.

[12] *Faostat*, 2008, http://faostat.fao.org.

[13] H. C. J. Godfray, J. R. Beddington, I. R. Crute et al., "Food security: the challenge of feeding 9 billion people," *Science*, vol. 327, no. 5967, pp. 812–818, 2010.

[14] K. M. Murphy, P. G. Reeves, and S. S. Jones, "Relationship between yield and mineral nutrient concentrations in historical and modern spring wheat cultivars," *Euphytica*, vol. 163, no. 3, pp. 381–390, 2008.

[15] M.-S. Fan, F.-J. Zhao, S. J. Fairweather-Tait, P. R. Poulton, S. J. Dunham, and S. P. McGrath, "Evidence of decreasing mineral density in wheat grain over the last 160 years," *Journal of Trace Elements in Medicine and Biology*, vol. 22, no. 4, pp. 315–324, 2008.

[16] P. J. White and M. R. Broadley, "Biofortification of crops with seven mineral elements often lacking in human diets—iron, zinc, copper, calcium, magnesium, selenium and iodine," *New Phytologist*, vol. 182, no. 1, pp. 49–84, 2009.

[17] S. S. Myers, A. Zanobetti, I. Kloog et al., "Increasing CO2 threatens human nutrition," *Nature*, vol. 510, no. 7503, pp. 139–142, 2014.

[18] R. M. Welch and R. D. Graham, "Breeding for micronutrients in staple food crops from a human nutrition perspective," *Journal of Experimental Botany*, vol. 55, no. 396, pp. 353–364, 2004.

[19] P. Borrill, J. M. Connorton, J. Balk, A. J. Miller, D. Sanders, and C. Uauy, "Biofortification of wheat grain with iron and zinc: integrating novel genomic resources and knowledge from model crops," *Frontiers in Plant Science*, vol. 5, no. 53, pp. 1–8, 2014.

[20] H. E. Bouis and A. Saltzman, "Improving nutrition through biofortification: a review of evidence from harvestplus, 2003 through 2016," *Global Food Security*, vol. 12, pp. 49–58, 2017.

[21] J. L. Rosado, K. M. Hambidge, L. V. Miller et al., "The quantity of zinc absorbed from wheat in adult women is enhanced by biofortification," *Journal of Nutrition*, vol. 139, no. 10, pp. 1920–1925, 2009.

[22] U. Schlemmer, W. Frølich, R. M. Prieto, and F. Grases, "Phytate in foods and significance for humans: Food sources, intake, processing, bioavailability, protective role and analysis," *Molecular Nutrition & Food Research*, vol. 53, no. 2, pp. 330–375, 2009.

[23] B. Lönnerdal, "Phytic acid-trace element (Zn, Cu, Mn) interactions," *International Journal of Food Science & Technology*, vol. 37, no. 7, pp. 749–758, 2002.

[24] H. E. Bouis and R. M. Welch, "Biofortification—a sustainable agricultural strategy for reducing micronutrient malnutrition in the global south," *Crop Science*, vol. 50, no. S1, pp. S21–S32, 2010.

[25] D. Oberleas, "The role of phytate in zinc bioavailability and homeostasis," in *Nutritional Bioavailability of Zinc*, vol. 210 of *ACS Symposium Series*, pp. 145–158, American Chemical Society, Washington, D.C., 1983.

[26] F. Cerino Badone, M. Amelotti, E. Cassani, and R. Pilu, "Study of low phytic acid1-7 (lpa1-7), a new ZmMRP4 mutation in maize," *Journal of Heredity*, vol. 103, no. 4, pp. 598–605, 2012.

[27] F. Senturk Akfirat and A. A. Uncuoglu, "Genetic diversity of winter wheat (Triticum aestivum L.) revealed by SSR markers," *Biochemical Genetics*, vol. 51, no. 3-4, pp. 223–229, 2013.

[28] http://mvgs.iaea.org .

[29] M. A. J. Parry, P. J. Madgwick, C. Bayon et al., "Mutation discovery for crop improvement," *Journal of Experimental Botany*, vol. 60, no. 10, pp. 2817–2825, 2009.

[30] R. Greiner, U. Konietzny, and K. D. Jany, "Phytate - an undesirable constituent of plant-based foods?" *Journal für Ernährungsmedizin*, vol. 8, no. 3, pp. 18–28, 2006.

[31] B. Sandstrom, A. Cederblad, B. Stenquist, and H. Andersson, "Effect of inositol hexaphosphate on retention of zinc and calcium from the human colon," *European Journal of Clinical Nutrition*, vol. 44, no. 10, pp. 705–708, 1990.

[32] H. S. Balyan, P. K. Gupta, S. Kumar et al., "Genetic improvement of grain protein content and other health-related constituents of wheat grain," *Plant Breeding*, vol. 132, no. 5, pp. 446–457, 2013.

[33] S. S. Kenzhebayeva, G. Doktyrbay, N. M. Capstaff et al., "Searching a spring wheat mutation resource for correlations between yield, grain size, and quality parameters," *Journal of Crop Improvement*, vol. 31, no. 2, pp. 209–228, 2017.

[34] F. Dai, J. Wang, S. Zhang, Z. Xu, and G. Zhang, "Genotypic and environmental variation in phytic acid content and its relation to protein content and malt quality in barley," *Food Chemistry*, vol. 105, no. 2, pp. 606–611, 2007.

[35] FAO (Food and Agriculture Organization), *Food Supply*, 2005, http://faostat.fao.org/site/609/DesktopDefault.aspx.PageID=609.

[36] DGE (German Nutrition Society), *Referenzwerte fur die Nährstoffzufuhr, 1. Auf-lage; Hrs. DGE, ÖGE, SGE und SVE*, Umschau/Braus, Frankfurt, Germany, 1st edition, 2001.

[37] M. H. Ryan, J. W. Derrick, and P. R. Dann, "Grain mineral concentrations and yield of wheat grown under organic and conventional management," *Journal of the Science of Food and Agriculture*, vol. 84, no. 3, pp. 207–216, 2004.

[38] F. J. Zhao, Y. H. Su, S. J. Dunham et al., "Variation in mineral micronutrient concentrations in grain of wheat lines of diverse origin," *Journal of Cereal Science*, vol. 49, no. 2, pp. 290–295, 2009.

[39] H. Kirchmann, L. Mattsson, and J. Eriksson, "Trace element concentration in wheat grain: Results from the Swedish long-term soil fertility experiments and national monitoring program," *Environmental Geochemistry and Health*, vol. 31, no. 5, pp. 561–571, 2009.

[40] D. P. Persson, T. H. Hansen, K. H. Laursen, J. K. Schjoerring, and S. Husted, "Simultaneous iron, zinc, sulfur and phosphorus speciation analysis of barley grain tissues using SEC-ICP-MS and IP-ICP-MS," *Metallomics*, vol. 1, no. 5, pp. 418–426, 2009.

[41] V. Raboy, P. F. Gerbasi, K. A. Young et al., "Origin and seed phenotype of maize low phytic acid 1-1 and low phytic acid 2-1," *Plant Physiology*, vol. 124, no. 1, pp. 355–368, 2000.

[42] Q.-L. Liu, X.-H. Xu, X.-L. Ren, H.-W. Fu, D.-X. Wu, and Q.-Y. Shu, "Generation and characterization of low phytic acid germplasm in rice (Oryza sativa L.)," *Theoretical and Applied Genetics*, vol. 114, no. 5, pp. 803–814, 2007.

[43] M. Guttieri, D. Bowen, J. A. Dorsch, V. Raboy, and E. Souza, "Identification and characterization of a low phytic acid wheat," *Crop Science*, vol. 44, no. 2, pp. 418–424, 2004.

[44] R. K. Gupta, S. S. Gangoliya, and N. K. Singh, "Screening and characterization of wheat germplasms for phytic acid and iron content," *Journal of Agricultural Science and Technology*, vol. 17, no. 3, pp. 747–756, 2015.

[45] I. Ahmad, F. Mohammad, A. Zeb, I. R. Noorka, . Farhatullah, and S. A. Jadoon, "Determination and inheritance of phytic acid as marker in diverse genetic group of bread wheat," *American Journal of Molecular Biology*, vol. 03, no. 03, pp. 158–164, 2013.

[46] C. Frontela, G. Ros, and C. Martínez, "Phytic acid content and "in vitro" iron, calcium and zinc bioavailability in bakery products: The effect of processing," *Journal of Cereal Science*, vol. 54, no. 1, pp. 173–179, 2011.

[47] R. M. García-Estepa, E. Guerra-Hernández, and B. García-Villanova, "Phytic acid content in milled cereal products and breads," *Food Research International*, vol. 32, no. 3, pp. 217–221, 1999.

[48] R. S. Gibson, K. B. Bailey, M. Gibbs, and E. L. Ferguson, "A review of phytate, iron, zinc, and calcium concentrations in plant-based complementary foods used in low-income countries and implications for bioavailability," *Food and Nutrition Bulletin*, vol. 31, no. 2, pp. S134–S146, 2010.

[49] R. Hurrell and I. Egli, "Iron bioavailability and dietary reference values," *American Journal of Clinical Nutrition*, vol. 91, no. 5, pp. 1461S–1467S, 2010.

[50] R. S. Gibson, "Zinc: The missing link in combating micronutrient malnutrition in developing countries," *Proceedings of the Nutrition Society*, vol. 65, no. 1, pp. 51–60, 2006.

[51] S. Akhter, A. Saeed, M. Irfan, and K. A. Malik, "In vitro dephytinization and bioavailability of essential minerals in several wheat varieties," *Journal of Cereal Science*, vol. 56, no. 3, pp. 741–746, 2012.

[52] T. Eagling, A. A. Wawer, P. R. Shewry, F.-J. Zhao, and S. J. Fairweather-Tait, ""Iron bioavailability in two commercial cultivars of wheat: Comparison between whole grain and white flour and the effects of nicotianamine and 20-deoxymugineic acid on iron uptake into Caco – 2 cells," *Journal of Agricultural and Food Chemistry*, vol. 62, no. 42, pp. 10320–10325, 2014.

[53] R. Salunke, N. Rawat, K. Neelam et al., "Effect of grain hardness on bioavailability of iron in wheat as determined using the coupled invitro digestion/Caco-2 model," *LWT- Food Science and Technology*, vol. 59, no. 1, pp. 433–438, 2014.

[54] S. Hussain, M. Maqsood, and L. Miller, "Bioavailable zinc in grains of bread wheat varieties of Pakistan," *Cereal Research Communications*, vol. 40, no. 1, pp. 62–73, 2012.

[55] I. Erdal, A. Yilmaz, S. Taban, S. Eker, B. Torun, and I. Cakmak, "Phytic acid and phosphorus concentrations in seeds of wheat cultivars grown with and without Zinc fertilization," *Journal of Plant Nutrition*, vol. 25, no. 1, pp. 113–127, 2002.

[56] N. T. Davies and S. Warrington, "The phytic acid mineral, trace element, protein and moisture content of UK Asian immigrant foods," *Human nutrition. Applied nutrition*, vol. 40, no. 1, pp. 49–59, 1986.

Review Article

Genetic Modification for Wheat Improvement: From Transgenesis to Genome Editing

Nikolai Borisjuk,[1] **Olena Kishchenko,**[1,2] **Serik Eliby,**[3] **Carly Schramm,**[3] **Peter Anderson,**[3] **Satyvaldy Jatayev,**[4] **Akhylbek Kurishbayev,**[4] **and Yuri Shavrukov** ⬤[3]

[1]*School of Life Science, Huaiyin Normal University, Huaian, China*
[2]*Institute of Cell Biology and Genetic Engineering, Kyiv, Ukraine*
[3]*College of Science and Engineering, Biological Sciences, Flinders University, Bedford Park, SA, Australia*
[4]*Faculty of Agronomy, S. Seifullin Kazakh AgroTechnical University, Astana, Kazakhstan*

Correspondence should be addressed to Yuri Shavrukov; yuri.shavrukov@flinders.edu.au

To feed the growing human population, global wheat yields should increase to approximately 5 tonnes per ha from the current 3.3 tonnes by 2050. To reach this goal, existing breeding practices must be complemented with new techniques built upon recent gains from wheat genome sequencing, and the accumulated knowledge of genetic determinants underlying the agricultural traits responsible for crop yield and quality. In this review we primarily focus on the tools and techniques available for accessing gene functions which lead to clear phenotypes in wheat. We provide a view of the development of wheat transformation techniques from a historical perspective, and summarize how techniques have been adapted to obtain gain-of-function phenotypes by gene overexpression, loss-of-function phenotypes by expressing antisense RNAs (RNA interference or RNAi), and most recently the manipulation of gene structure and expression using site-specific nucleases, such as CRISPR/Cas9, for genome editing. The review summarizes recent successes in the application of wheat genetic manipulation to increase yield, improve nutritional and health-promoting qualities in wheat, and enhance the crop's resistance to various biotic and abiotic stresses.

1. Introduction

Cereals are a key component of human diets, providing a significant proportion of the protein and calories consumed worldwide. While rice and maize dominate global cereal production, wheat is another vital crop consumed by humans, contributing to approximately 20% of our energy needs (calories) and 25% of our dietary protein. The Green Revolution of the 1970s achieved enormous yield gains via the introduction of disease resistant semidwarf high yielding wheat varieties developed by Dr. N.E.Borlaug and colleagues. Since that time, however, global wheat production has stagnated, and current trends show that yields will not be sufficient to meet growing market demands.

According to the United Nations' Food and Agriculture Organization (FAO), over 756 million tonnes of wheat grain was harvested from over 220 million ha of arable land

in 2016/2017 (www.fao.org/faostat). Despite this, wheat lags behind other major cereals such as maize and rice, both in terms of yield, and the application of genomic tools for its improvement [1]. While the average worldwide yield grew almost 3-fold during the Green Revolution, driven by the expansion of irrigation, intensive use of fertilisers and advanced breeding [2]; the current average global wheat yield of ~3 tonnes per hectare is far below the crop's potential [3]. In order to feed the population of 9 billion people predicted for 2050, wheat yield should grow by over 60% while still maintaining and/or improving its nutritional characteristics [3, 4]. To achieve this goal without increasing the area of cultivated land, which is simply not available, emphasis must be concentrated on key traits related to plant productivity and adaptation to environmental challenges. A deficit in this key staple crop could present a serious threat to global food security, so improved molecular-based breeding and genetic

engineering techniques are necessary to break through the current yield ceiling.

Existing modern breeding efforts now need to be complemented with advanced crop functional genomics, which can provide insights into the functioning of wheat genetic determinants. The available tools for wheat genetic modification provide the experimental means to functionally characterize genetic determinants by suppressing or enhancing gene activities. This knowledge can then be used for targeted improvements tailored to the specific needs of the diverse and changing environments in which wheat is grown across the world. This offers the potential to tackle yield gaps wherever they exist, for a variety of causes, enabling this global crop to finally reach its full potential.

2. Progress in Wheat Genetic Transformation

Bread wheat (*Triticum aestivum* L.), the most widespread of all wheat species, is an annual herb belonging to the family Gramineae or Poaceae. Wheat was domesticated around 8,000 years ago [29] and has since undergone hybridization and genome duplication events to generate its hexaploid genome (2n = 6x = 42, AABBDD), which is more than five times larger than the human genome. It was previously estimated that the genome of common wheat contained over 128,000 genes [30], with over 80% of the genome consisting of repetitive sequences of DNA [31]. However, more recent estimates suggested a total of 107,891 high-confidence genes with over 85% repetitive DNA sequences, representing a threefold redundancy due to its hexaploid genome [32].

Genetic transformation, the fundamental tool of genetic engineering, allows the introduction and expression of various genes of interest in the cells of living organisms, bypassing, when desirable, the barriers of sexual incompatibility that exist in nature. Despite the considerable efforts of the international research community, development of wheat genetic engineering lags behind that of the other key agricultural crops like rice and maize. This may be attributed to the genetic characteristics of wheat, including its very large (17,000 Mbp) and highly redundant complex genome, as well as the relative recalcitrance of most varieties to *in vitro* culture and regeneration (reviewed recently in [33]).

The first successful genetic transformation of common wheat was conducted at Florida University, USA [34], using biolistics and financed by a research grant from Monsanto. Scientists from Monsanto were also the first to report the generation of transgenic wheat using *Agrobacterium*-mediated transformation [35].

2.1. Wheat Transformation Methods. Presently, biolistics and *Agrobacterium*-mediated transformation using immature embryos as explants remain the main methods for genetic engineering of wheat. Each method has its own advantages and drawbacks (Table 1). The main advantages of *Agrobacterium* transformation are the relatively high ratio of single copy gene inserts and relative simplicity of the transformation procedure. In contrast, biolistics offer benefits in their capacity to transform organelles and deliver RNA, proteins, nanoparticles, dyes, and complexes into cells. Utilization of linear Minimal Expression Cassettes (MECs) in biolistic transformation enables the production of plants carrying much simpler patterns of transgene integration compared to plasmid bombardment, with a higher proportion of single copy inserts [36–38].

In addition, biolistic delivery of MEC simplifies the simultaneous cotransformation of several genes and, in contrast to *Agrobacterium*, does not introduce vector backbone DNA or repetitive border sequences flanking the T-DNA into the transformed plant cells. In experiments with the wheat cv. EM12, transformation with MECs instead of plasmids improved transformation frequency (TF) almost threefold from 0.4% to 1.1% [39]. A simplified method for DNA/gold coating was described by Ismagul et al. [38] for the high-throughput biolistic production of single copy transgenic wheat utilizing diluted MECs. This method involves the application of PM solution (42% PEG 2000 and 560 mM $MgCl_2$) instead of the spermidine and $CaCl_2$ used in the standard Bio-Rad procedure.

Biolistics allow for the transfer into wheat of relatively large DNA fragments. Partier et al. [40] conducted successful biolistic transformation of wheat with a 53 Kb linear cassette which contained a 44 Kb fragment of an *Arabidopsis* gene flanked by selection and reporter genes. The intact cassette was detected in T_1 and T_2 generation plants. The main disadvantages of both *Agrobacterium* and biolistic transformation methods in wheat are genotype dependency, and the requirement for lengthy periods of aseptic tissue culture.

2.2. Wheat Transformation Frequencies. Until recently, the TF for most tested genotypes of common wheat remained quite low at a level mostly below 5% [8]. Many research groups use model wheat varieties such as Bobwhite SH98-26 and Fielder in their experiments due to their amenability to transformation via published protocols [8, 9]. The cultivar Bobwhite SH98-26 was among 129 sister lines made from crosses of cultivars Aurora//Kalyan/Bluebird 3/Woodpecker at CIMMYT, and selected for its high TF of over 70% by biolistic transformation [9]. Such high transformation efficiency is yet to be reproduced by other researchers, and although the reason for this remains unclear, it is possible that not all of the finer details of the transformation protocol could be described in the report. The success rate of this high TF may be explained by the particular hybrid genotype used, or simply the advanced skills and experience of the technicians at CIMMYT. Since publishing [9], A. Pellegrineschi has been employed by two big Biotech companies, Pioneer Hi Breed and Du Pont, which indirectly confirms that he has developed a reliable wheat transformation protocol that allowed him to produce the published results. The authors of the review are therefore not sceptical; however the high TF presented in [9] must be reproduced by other researchers in future if this protocol is to be widely adopted. The cultivar Fielder, which is the model variety in *Agrobacterium* transformation, was preselected by the researchers of the Japan Tobacco Company, where the detailed protocol "PureWheat" was developed [8] that allows TFs of 40–90%. In this protocol, positive selection through the application of phosphomannose isomerase

TABLE 1: Comparison of parameters of agrobacterial and biolistic transformation.

Parameters	Agrobacterium-mediated transformation	Biolistic transformation	References
Genotype dependency	High	Less	[5]
Stability of expression of transgenes	High	High for Minimal Expression Cassettes (MEC), lower for plasmids	[6, 7]
Copy number of inserts	Around 50% single copy; Depends on the strain and transformation conditions	<50%; Depends on the amount of DNA/shot. More multicopy inserts	[7, 8]
Integration of the new genes	Random. More than one locus	Random. Often many at the same locus	[6]
Maximum transformation frequency (TF) for wheat (per 100 embryos treated)	Up to 90%	>70%	[8, 9]
Complexity of the transformation procedure in wheat	Simpler. Usually requires aseptic conditions	More complex. Requires aseptic conditions and a Biolistic Gun	[8, 10]
Main explants in wheat	Immature embryos	Immature embryos	[8, 10]
Complexity of vector construct preparation, co-transformation	More complex	Simpler	[11]
Maximum sizes of transferred inserts published	Up to 200 Kb	150–164 Kb	[12–14]
Transfer of T-DNA borders	Yes	No (for MEC)	[15]
Transfer of vector DNA	Possible	No (for MEC); Yes (for plasmids)	[15]
Transfer of bacterial chromosomal DNA	Possible	No (for MEC)	[16]
Marker free transformation in wheat	Possible	Possible	[17–19]
In planta transformation in wheat	Possible	Possible	[20–22]
Delivery of RNA, proteins, nanoparticles and dyes	No	Possible	[23]
Transformation of chloroplasts and mitochondria	No	Possible	[24, 25]
Transient gene expression in different tissues and organs of plants	Efficient for limited number of plant species	Efficient	[26–28]

(PMI), which converts mannose-6-phosphate into fructose-6-phosphate, was found to facilitate a relatively high TF (20%) in biolistic transformation of the spring wheat line UC703 [41].

2.3. Agrobacterium-Mediated Transformation. In recent years several groups have reported efficient *Agrobacterium* transformation of a number of wheat cultivars [8, 42–44]. Hensel et al. [43] developed a protocol to transform the model genotype Bobwhite SH98-26 with TFs up to 15%. In the experiments of Richardson et al. [42] using the PureWheat technology with the cv. Fielder, selection on 5-10 mg/l of phosphinothricin resulted in a TF of 41.0%. Without selective pressure, the TF was only around 2.3%. Similar results were obtained for Chinese commercial cultivars of wheat by Wang et al. [44], who demonstrated TF of 37.7% for cv. CB037, 22.7% for cv. Kenong 199, and 45.3% for the model cv. Fielder. It was shown by Ishida et al. [8] that centrifugation of immature embryos before infection with *Agrobacterium* was one of the critical requirements for successful transformation, while heat shock, contrary to findings in other cereals, was not efficient. At present, *Agrobacterium* strains EHA101, EHA105 [8], AGL0, AGL1 [43, 45], GV3101 [46], C58C1 [47, 48], and LBA4404 [49] are the most popular in wheat transformation.

Transformation using mature wheat embryos is currently characterized by relatively low TFs. Wang et al. [48] transformed longitudinally cut mature embryos and observed TFs of 0.06%, 0.67%, and 0.89% for the cultivars Bobwhite, Yumai 66, and Lunxuan 208, respectively. Aadel et al. [50] found that, with the application of 200 μM acetosyringone, the TFs for the genotypes Rajae and Amal were 0.66% and 1.00%. The protocol of Medvecka and Harwood [51], using Bobwhite SH98-56, allows production of transformants at a TF of 2.2%. More information and a simplified protocol can be found in [52].

In contrast to *in vitro* methods, the *in planta* approach has the potential to overcome the problem of high genotype dependency seen with the existing transformation methods. There have been several publications on the *in planta* production of transgenic wheat, most involving the direct injection of *Agrobacterium*. Supartana et al. [49] reported agrobacterial transformation of wheat cv. Shiranekomugi using seeds soaked overnight in water. Transformation was achieved by double piercing the area where a shoot would later emerge with a needle dipped into *Agrobacterium* inoculum. This method used the *Agrobacterium* strains LBA4404 and an M-21 mutant strain, and no tissue culture steps were used at any stage. The plants obtained were analysed for antibiotic resistance, and by PCR, Southern hybridization and plasmid rescue to confirm their transgenic status. Zhao et al. [53] produced *Agrobacterium*-mediated transgenic wheat by adding inoculum to an incision made at the base of wheat seedlings. Their tissue culture-free method was reportedly successful in transferring a powdery mildew-resistance gene, with a TF of up to 9.82%. Similarly, in the experiments of Razzaq et al. [54] with the wheat cv. GA-2002, the apical meristems of imbibed wheat seeds were wounded and inoculated with the *Agrobacterium* strain LBA4404 containing the binary vector pBI121 (35S-GUS, pNOS-NPTII). The

kanamycin resistant plants produced were analysed by PCR and GUS histochemical staining of the embryos. Risacher et al. [20] developed an efficient semi-*in planta* transformation protocol (US patent 7803988 B2, 2010) that involves *in planta* agrobacterial infection of immature embryos within developing seeds. Spikes of the spring wheat line NB$_1$ were harvested 16 to 18 days after anthesis and *Agrobacterium* injected into the base of each spikelet. The spikes, with their flag leaves still attached, were incubated in low light for 2 to 3 days before embryos were isolated and cultured *in vitro*. The protocol achieved an average TF of 5% with 30-50% of plants carrying a single transgene insertion.

The practice of dipping floral buds in a suspension of *Agrobacterium* (Floral dip) is a routine and highly efficient method of transformation in *Arabidopsis,* but it is rarely successful in other plant species. However, Zale et al. [21] were able to generate three independent transgenic lines of the spring wheat line, Crocus, by utilizing the floral dip approach. The transformants were studied thoroughly at the molecular level for three to six generations. In their most recent publication, Hamada et al. [22] trialled a biolistic method for *in planta* transformation. They found that bombarding the exposed shoot apical meristems of the wheat cultivars Fielder and Haruyokoi, using 0.6 μm gold particles and 1350 psi pressure, resulted in the integration of the GFP reporter gene into the germline cells in 62% of regenerated plants (transformants), including possible chimeric individuals. The full potential of the wheat *in planta* procedures published to date is yet to be fully realized.

Microspore transformation using immature pollen grains is a method that generates doubled haploid homozygous wheat plants in a single generation. The great advantage of this technique is that it by-passes the several years required by other transformation methods to develop true-breeding transformant lines. Current protocols for microspore-based transgenic wheat production through electroporation, biolistics, and *Agrobacterium*-mediated transformation are presented in [55–57]. Various wheat genotypes have been used as donors. Express, Chris, Farnum, Hollis, Louise, Perigee, and WestBred 926 have all performed successfully in microspore transformation. Four genes were targeted for microspore transformation, including β-glucuronidase (GUS), *uidA*; bialaphos resistance, *Bar*; *Trichoderma harzianum* endochitinase gene, *ThEn42*; and *Bacillus subtilis* 1,4-β-xylanase gene [55–57]. In general, *Agrobacterium*-mediated microspore transformation produces the best outcome, generating stable homozygous plants and fewer chimeric plants. Given the value of doubled haploid lines in simplifying and accelerating wheat breeding, it is likely that microspore methods will be optimized in the future.

Recent advances in high-throughput sequencing technologies have allowed for more detailed information to be obtained about the composition and fate of transgenes introduced into plant cells through *Agrobacterium*-mediated transformation. In one report, close to 0.4% of transgenic *Arabidopsis* lines examined contained insertions of *Agrobacterium* chromosomal DNA [16]. Singer et al. [58] showed, for the first time, the presence in transgenic plant cells of the extra-chromosomal circular T-DNA that they named

"T-circles". In rice, 26% of 331 analysed transgenic lines contained the transposon Tn5393, which was transferred from the *Agrobacterium* strain LBA4404 [59]. This transposon was not detected in the strains EHA105 and GV3101. These data reflect the importance of detailed molecular analysis of transgenic lines, and careful selection of the appropriate agrobacterial strains for transformation. A better understanding of the biological mechanisms of the transfer and integration of transgenes during bacterial infection will also permit the creation of new more efficient strains of *Agrobacterium* [5, 60] or utilization of bacterial species that are not pathogenic in nature [61].

2.4. Optimization of Tissue Culture Conditions. Tissue culture conditions are important factors that impact the TF of all plants, including wheat. Below we present information reflecting the current trends in wheat tissue culture that might be of interest to researchers working in the field. Changing the composition of the macrosalts in the growth medium can positively affect the TF. Callus induction on 6x DSEM medium with an increased concentration of ammonium nitrate (62.56 mM) as the sole nitrogen source, resulted in the sevenfold improvement of TF during biolistic transformation of the elite wheat cv. Superb [62]. The authors point out that this modification of the medium brought about an increase in the number of somatic embryoids and possibly also reduced stress during the bombardment of the cells due to the elevated osmolarity. Ishida et al. [8] identified some critical points in the transformation process and developed a detailed protocol for *Agrobacterium*-mediated transformation for both immature and mature wheat embryos from amenable genotypes. The authors however did not discuss any work on media improvements.

Antioxidants such as ascorbic acid, glutathione, lipoic acid, selenite, and cysteine were found to decrease necrosis and darkening of the tissues, improving plant regeneration and TF during genetic transformation [63]. Lipoic acid at 25–50 μM led to a twofold improvement in agrobacterial transformation of the wheat cv. Bobwhite [64].

Arabinogalactan-proteins at a concentration of 5 mg/l improved regeneration of bread (cv. Ikizce-96) and durum (cv. Mirzabey) wheat from 77.77% and 72.11%, to 94.86% and 89.73%, respectively [65]. Kumar et al. [66] were able to achieve close to 100% plant regeneration from calli induced from wheat mature and immature embryos on 2.0 mg/l picloram, using a regeneration medium with 0.1 mg/l 2,4-D, 5.0 mg/l zeatin, and 15 mg/l $CuSO_4$. Miroshnichenko et al. [67] developed a protocol for the regeneration of einkorn wheat (T. *monococcum* L.), in which a combination of 3.0 mg/l dicamba, 50.0 mg/l daminozide (an inhibitor of ethylene synthesis), and 0.25 mg/l thidiazuron in the regeneration medium was the most efficient. The authors presume that this protocol may be useful for other wheat species as well.

2.5. Transient Transformation. Transient transformation with the use of reporter genes such as GUS [68] and GFP [69] is helpful for optimization of transformation conditions, as well as the analysis of promoters and protein expression. Transgene expression that is tightly targeted to specific tissues and developmental stages is often desired for directed modification of morphogenic traits. It can also be beneficial for avoiding feedback mechanisms, transgene silencing, or other unforeseen effects that can arise from constitutive transgene expression. Transient assays using protoplasts can also facilitate efficient analysis of gene regulatory mechanisms through cotransformation of enhancer/represser elements and promoter-reporter gene constructs [11]. Viral induced gene silencing (VIGS) offers a fast and rapid transient assay for silencing of gene expression. VIGS can be achieved through a simple vacuum-aided cocultivation of germinated wheat seeds with *Agrobacterium* carrying an appropriate VIGS construct [70]. However, results from VIGS are not comparable to results from stable transformation experiments, but this does not mean that the VIGS findings are false or of no value. Like RNAi, VIGS findings are useful for increasing our understanding of gene function and may lead to the development of a strategy that is more successful than simple constitutive transgene overexpression.

2.6. Future Directions: Gene Stacking and Plant Artificial Chromosomes. Gene stacking or pyramiding (defined as the stacking of multiple transgenes at a single chromosomal location) greatly expands the potential for genetic engineering of traits in wheat [71]. Applications of this technology include modifying complex and multigenic traits, or inserting entire biosynthetic pathways, with integration at a single locus. This significantly simplifies the subsequent breeding process. Alternatively, for disease resistance traits, transgene stacks can aid in generating broad spectrum resistance to multiple threats, or more long-lasting defence to a single pathogen via genes with differing modes-of-action. Construction of a transgene stack is most often achieved using type II restriction enzymes in the Golden-Gate cloning system, or by Gibson Assembly. The integration of longer DNA fragments carries certain risks, however, such as reduced TF, fragmentation leading to only partial integration, and transgene silencing if promoters or other regulatory elements are used repeatedly. The magnitude of these risks differs according to the nature of the vectors or transformation cassettes, the transgene products, and the recipient genotype. Much of the more recent work involving stacked transgenes uses site-specific nucleases so that the transgene stack can be directed to a favourable locus in the genome. Site-specific nucleases will be covered in a later section of this review.

Plant artificial chromosomes or minichromosomes were also developed for multigene transfer [72]. The main advantage is that they generate a new genetic locus that segregates independently of endogenous chromosomes, thus lessening the disruption of existing genomic regions. Minichromosomes have been developed in mammalian cells through a process named "Telomere-mediated chromosome truncation" that targets highly conserved telomeric repeat sequences. Analogous sequences have been identified in *Arabidopsis*, but progress in developing plant minichromosomes still lags behind the work in mammalian cells. Minichromosomes have the potential to act as "super-vector platforms" for the organization and expression of foreign genes and may even be designed with Cre/lox recombination sites to accept

the introduction of new genes. This presents the option to "add on" additional alleles or genetic elements at a later date. Currently, the reliable transmission of minichromosomes to the progeny of primary transformants presents the greatest barrier to the application of this technology. However, initial work has suggested that wheat's allopolyploid nature makes it more tolerant to chromosomal truncation than that of diploid species, making this a very promising future prospect for wheat genetic engineering [73].

3. Manipulating Agronomic Traits by Transgene Expression in Wheat

Built on the progress made on techniques for genetic manipulation of wheat, overexpression of endogenous and foreign genes has tremendously enriched our understanding of the functionality of numerous genes, contributing to the generation of many new and improved agronomically important traits. This improved understanding of gene function has led to the development of specific promoters to drive gene expression in response to environmental stimuli and/or in a tissue, organ, or developmental stage-specific manner [74, 75]. Various signals for directing expressed proteins to particular cellular compartments have also been used [76]. Also, the field received another major stimulus through the use of transcription factors that function as major regulators of gene networks involved in numerous metabolic/physiological pathways. Transcription factors are often highly conserved between different plant species, so the information obtained from studying model plants like *Arabidopsis* or rice can be applied to other plants such as wheat [77–79].

In pursuit of improved crop performance, wheat genetic manipulations made by transgenic gene expression have targeted all major agronomic traits including yield, grain quality, and tolerance to abiotic and biotic stresses. The last 10 years have been characterized by acceleration in this field and an increasing number of publications. In this review we will focus primarily on attempts to improve yield and grain quality, the two areas that have not been sufficiently reviewed by other authors. For readers interested in improving wheat stress responses, we refer to recent comprehensive reviews on the topics of pathogen resistance [80–82] and abiotic stress tolerance [83, 84], as well as a recent review of transgenic wheat engineered with various genes of agronomic importance [33].

3.1. Yield. Yield is determined by the number and size of grains produced by the crop. The major genes controlling yield-related traits in wheat have been identified by genetic and genomics techniques, as reported in recent reviews [85–87].

Grain size (GS) has long been a focus of wheat selection and modern breeding [88]. Progress in comparative wheat genomics has led to the identification of *TaGW2* [89], a wheat homolog of a rice gene that negatively influences GS through regulation of cell division within the spikelet hull [90]. The first experiments aimed at downregulating *TaGW2* in wheat have shown some controversial results. Bednarek et al. [91] demonstrated that expression of a specific

RNAi to suppress three *TaGW2* homologs A, B, and D, of bread wheat cv. Recital, led to a decrease in grain size and weight. In contrast, Hong et al. [92], in a similar series of experiments, demonstrated a significant *increase* in the grain width and weight of a Chinese bread wheat cv. Shi 4185 that is typically characterized by small grains. Such differences in the phenotypes obtained in these two studies may be explained by a cultivar-specific reaction to experimental intervention, and/or by the application of different transformation protocols, as has been previously reported for barley [93].

Plant growth and productivity to a large extent relies on the uptake of carbon, nitrogen, and phosphorus. Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo), the most abundant protein on Earth, is the major enzyme assimilating atmospheric carbon in the form of CO_2 into organic compounds in photosynthetic organisms. However, the enzyme has relatively low efficiency and a slow turnover rate [94]. Despite significant progress in characterizing the enzyme's properties [95] and improving RuBisCo efficiency in a range of plant species [96, 97], little success has been achieved so far in commercial crops, including wheat. Another area for optimizing carbon assimilation in crops is to introduce components of the C_4 photosynthetic apparatus into C_3 plants like wheat and rice. C_4 plants, such as maize, display definite advantages over plants with C_3 photosynthesis, especially under high temperature and limited water conditions [98]. These advantages rely on a number of specific enzymes associated with RuBisCo, which provide higher photosynthetic efficiency and CO_2 assimilation for the C_4 plants. Transferring genes encoding the C_4-specific enzymes, phosphoenolpyruvate carboxylase (PEPC), and pyruvate orthophosphate dikinase (PPDK) has led to promising results in rice [99] and wheat [100, 101]. As an example, Zhang et al. [101] generated transgenic wheat plants overexpressing PEPC and PPDK both separately and simultaneously in the same transgenic lines. The authors showed a positive synergistic effect on wheat photosynthetic characteristics and yield in the latter design. The follow-up study on wheat expressing maize PEPC demonstrated that, in addition to increased yield, the transgenic lines had improved drought tolerance linked to elevated expression of proteins involved in photosynthesis and protein metabolism [102].

The maize gene encoding the transcription factor Dof1, known to upregulate the expression of PEPC, was introduced into wheat by *Agrobacterium*-mediated transformation [103]. Expression of *ZmDof1* under the control of the light-inducible RuBisCo promoter, led to increased biomass and yield components in transgenic wheat, while constitutive expression resulted in the downregulation of photosynthetic genes and a corresponding negative impact on crop productivity.

The Nuclear Factor Y (NF-Y) transcription factors are recognized as important regulators of many plant developmental and physiological processes [104]. NF-Ys are composed of protein subunits from three distinct transcription factor families (NF-YA, NF-YB, and NF-YC), with each of them represented by multiple members [105]. *TaNFY-A-B1*, a low-nitrogen- and low-phosphorus-inducible NF-YA transcription factor, was overexpressed in wheat. This led to a significant increase in nitrogen and phosphorus

uptake and grain yield in a field experiment [106]. The authors suggested that increased nutrient uptake resulted from stimulated root development and upregulation of both nitrate and phosphate transporters in the roots of the transgenic wheat. Our own work showed the positive role of the second wheat gene, *TaNF-YB4*, in wheat grain yield [107]. Constitutive overexpression of the gene under the control of the maize ubiquitin promoter in transgenic wheat cv. Gladius, resulted in the development of significantly more spikes, and a 20–30% increase in grain yield compared to the untransformed control plants.

Overexpression of another wheat transcription factor, *TaNAC2-5A,* which plays a role in nitrogen signalling, enhanced root growth and the nitrate influx rate, consequently increasing the root's ability to acquire nitrate from the soil. Transgenic wheat lines revealed higher grain yield and higher nitrogen accumulation in aerial parts of the plant, which was subsequently allocated to grains [108].

Recently, a US research group reported on the expression of a rice gene encoding a soluble starch synthase gene (*OsSS-I*) with increased heat stability. This lead to a significant, 21-34%, yield increase in T_2 and T_3 generations of transgenic wheat under heat stress conditions [109]. The expression of *OsSS-I* also prolonged the duration of the photosynthetic growth period in bread wheat. Similarly, overexpression of an endogenous gene coding for the chloroplastic glutamine synthase gene (*TaGS2*) in wheat led to prolonged leaf photosynthesis, and an increased rate of nitrogen remobilization into grains, which translated to higher spike number, grain number per spike, and total yield [110]. Taken together, these reports on the modulation of carbon and nitrogen pathways once again illustrate the tight link between nitrogen assimilation and carbon metabolism [111] and resulting crop productivity.

3.2. Grain Quality. Grain is the harvested part of the wheat plant and its nutritional and health properties are determined by its biochemical composition. Starch, making up 55–75% of total dry grain weight, and storage protein 10–15%, are the main reserves of the wheat seed. Therefore, starch and protein greatly affect the quality of the products made from wheat flour. Along with optimized starch and protein, an adequate level of essential elements like iron, zinc, calcium, phosphorus, and antioxidants is also essential for healthy and balanced wheat products. Many of those quality traits have been addressed in recent years by means of transgenic interventions.

A number of studies have focused on the biosynthetic pathways and composition of starch, as reviewed by Sonnewald and Kossmann [112] and more recently by Kumar et al. [113]. Based on the insights gained, a number of biotechnological interventions have been undertaken in order to both increase the amount of starch, and modulate its quality, with mixed outcomes. In an attempt to increase the level of precursors for starch synthesis in wheat, Weichert et al. [114] overexpressed the barley sucrose transporter gene (*HvSUT1*), which led to enhanced sucrose uptake and protein content in wheat grains, but no significant modification

to starch levels. Expression of an optimized maize ADP-glucose pyrophosphorylase (*ZmAGPase*) resulted in elevated yield and enhanced photosynthetic rates in the transgenic wheat lines [115]. Downregulation of the transcription factor *TaRSR1*, a wheat homolog of Rice Starch Regulator (*OsRSR1*), which was shown to negatively regulate the gene expression of some starch synthesis-related enzymes in wheat grains [116], resulted in a significant 30% increase in starch content, and also a ~20% increase in yield in terms of 1000-kernel weight [117]. The increases in starch and yield were underpinned by the marked induction of expression of many key-genes in sugar metabolism and starch biosynthesis.

Starch consumer quality depends mostly on the ratio of amylose to amylopectin, the two main macromolecules forming starch. Starch with increased amylose has attracted much interest because of its contribution to resistant starch (RS) in food, which confers beneficial effects on human health. There is evidence that RS can provide protection from several health conditions such as diabetes, obesity, and cardiovascular diseases [118]. A number of experiments focused on downregulation of starch branching enzymes, SBEIIa, and SBEIIb, which led to substantially increased amylose levels in wheat [119, 120]. High amylose starch has demonstrated positive health-related effects in rats [119] and more recently in a study involving obesity in humans [121].

Nitrogen is not only the most important plant nutrient contributing to crop yield, but also plays a significant role in defining the accumulation and, to some extent, the composition of storage protein in wheat grains [122]. Nitrogen is supplied to the grain by two major pathways: remobilization from the canopy (leaves and stems), and root uptake from the soil. In their experiments, Zhao et al. [123] identified a novel wheat gene, *TaNAC-S*, a member of the NAC transcription factor family, that showed decreased expression during leaf senescence but significant expression in a stay-green phenotype. Transgenic overexpression of the gene in wheat resulted in delayed leaf senescence and increased protein concentration in grains, while the crop's biomass and grain yield remained unaffected. Another group of Chinese researchers overexpressed a tobacco nitrate reductase gene (*NtNR*) in two commercial winter wheat cultivars, ND146 and JM6358, following *Agrobacterium*-mediated transformation [124]. Constitutive overexpression of the gene remarkably enhanced foliar nitrate reductase activity and resulted in significantly augmented seed protein content and 1000-grain weight in the majority of the T_1 offspring analysed.

The main component of wheat storage proteins, gluten, primarily determines the viscoelastic properties of wheat dough [125]. Glutens consist of gliadins and glutenins, which together comprise 70-80% of the total flour protein. Therefore, genes encoding different classes of storage proteins have been targeted in efforts to improve both the nutritional value and bread-making quality of wheat. In fact, the genes coding for high-molecular glutenin subunits (HMW-GS) were among the first introduced into wheat [126–129], with the aim of improving dough functions following the first reports of a method for wheat transformation. Specifically, introduction of the subunits 1Ax1 and 1Dx5 into several common wheat cultivars by genetic transformation has demonstrated the

potential of the transgenes to enhance dough quality to various extents [130–132]. In follow-up studies, the HMW-GS genes were introduced into selected wheat cultivars, mostly via the easily transformed cv. Bobwhite and then transferred into selected elite commercial varieties. Improvements in dough properties were obtained, demonstrating the feasibility of utilizing transgenic lines in wheat breeding programs [133, 134].

In contrast to HMW glutenins, which form complex polymer structures and are strongly correlated to dough elasticity, gliadins are monomeric components that contribute mainly to the extensibility and viscosity of the dough [135]. Based on their electrophoretic mobility, gliadins are grouped into three structural types: α-, γ-, and ω-gliadins, each encoded by tightly-linked multigene clusters. The interest in genetic modification of gliadins has been stimulated, not only by their contribution to dough quality, but also because they include the majority of immunogenic epitopes related to immune conditions such as wheat-dependent exercise-induced anaphylaxis (WDEIA) and coeliac disease [136, 137].

Iron and zinc are essential micronutrients for human nutrition. According to the World Health Organization, over a billion people suffer from iron-deficiency anaemia, and Zn-deficiency is estimated to be associated with an annual death rate of almost half a million children under the age of five. Modern elite wheat cultivars usually contain suboptimal quantities of micronutrients [138], and because most of it is accumulated in the outer husk, aleurone, and embryo, the micronutrients are lost during milling and polishing [139]. Another problem is that phytic acid, a major antinutritional factor for iron and zinc uptake in the human digestive tract, is codeposited with the minerals in aleurone storage vacuoles.

Biofortification, the enhancing of crop nutritional quality, is considered a promising approach to alleviate micronutrient deficiency. Transgenic studies have made significant headway in the development of strategies aimed at improving the available content of micronutrients such as iron in wheat grains. Plants store iron primarily in the form of ferritin structures accumulated mainly in nongreen plastids, etioplasts, and amyloplasts. Expression of a gene coding for an *Aspergillus niger* phytase, a phytic acid degrading enzyme targeted to the wheat aleurone [140], and endogenous [141] or soybean [142] ferritin genes in wheat endosperm, were the first successful attempts to transgenically biofortify wheat grains with iron. Recently, Connorton et al. [143] have isolated, characterized, and overexpressed two wheat Vacuolar Iron Transporter (*TaVIT*) genes under the control of an endosperm-specific promoter in wheat and barley. They reported that the introduction of one of the genes, *TaVIT2*, resulted in a greater than twofold increase in iron in the flour prepared from transgenic wheat grains without other detected changes.

4. RNA Interference Applications in Wheat

RNA interference (RNAi) is a common regulatory mechanism of gene expression in eukaryotic cells that has become a powerful tool for functional gene analysis and the engineering of novel phenotypes. The technique is based on the expression of antisense or hairpin RNAi constructs, or other forms of short interfering RNA molecules to direct posttranscriptional gene silencing in a sequence specific manner.

The vernalisation gene, *TaVRN2*, was the first wheat gene to be targeted by RNAi in transgenic wheat plants [144]. *TaVRN1* mRNA in transgenic plants was reduced by 60%, which led to much earlier flowering. In another study, Loukoianov et al. [145] suppressed expression of *TaVRN1* by up to 80%, delaying flowering time by 14 to 19 days and increasing the number of leaves relative to the nontransgenic controls. These two breakthrough studies provided essential evidence for understanding the molecular mechanisms of flowering timing and vernalisation requirements in wheat, which may assist in diversifying the environments in which wheat can be grown.

The application of RNAi has made a solid contribution to manipulating wheat grain size [146–148] and quality [119, 149–151]. For example, Alterbach and Allen [152] used RNAi silencing to suppress the expression of ω-gliadins associated with WDEIA in the US wheat cv. Butte 86. They later demonstrated that transgenic wheat lines deficient in ω-gliadins had no changes in patterns of other grain proteins, and even showed increased protein stability with improved dough properties under various growth conditions [153]. Similarly, Gil-Humanes et al. [154] downregulated the γ-gliadin genes in the wheat cv. Bobwhite and then transferred the trait into three commercial wheat cultivars by conventional crossing.

The reduction of γ-gliadins in wheat grains was compensated by an increased amount of glutenins, which led to stronger dough with better over-mixing-resistance. In a recent comprehensive study, Barro et al. [149] used a combination of seven RNAi expressing plasmids to selectively target α-, γ-, and ω-gliadins, and Low Molecular Weight (LMW) glutenins in the wheat cv. Bobwhite, with the goal to reduce gluten epitopes related to coeliac disease (CD). The protein analyses showed that three RNAi plasmid combinations resulted in total absence of CD epitopes from the most immunogenic α- and ω-gliadins in the transgenic wheat lines. These very promising results pave the way to developing wheat varieties with nonallergenic properties.

However, the major application of RNAi technology for wheat has been in pathogen and pest control using virus- and host-induced gene silencing platforms [155]. A recent strategy, referred to as host-induced gene silencing (HIGS), has been developed to silence pest or pathogen genes by plant-mediated RNAi during their feeding or attempted infection, thereby reducing disease levels. This strategy relies on the host-plant's ability to produce interfering RNA molecules complementary to targeted pest/pathogen genes. These molecules are then transferred to the invader, causing silencing of the targeted gene. In wheat, HIGS has been most widely applied to control fungal and insect diseases. As an example, silencing of fungal glucanosyltransferase genes (*GTF1* and *GTF2*), and the virulence effector gene *Avra10*, affects wheat resistance to the powdery mildew fungus *Blumeria graminis* [156]. Three hairpin RNAi constructs corresponding to different regions of the *Fusarium graminearum* chitin synthase gene (*Chs3b*) were found to silence *Chs3b* in transgenic *F. graminearum* strains. Coexpression

of these three RNAi constructs in two independent elite wheat cultivar transgenic lines, conferred high levels of stable and durable resistance to both *Fusarium* head blight and *Fusarium* seedling blight [157]. Stable transgenic wheat plants carrying an RNAi hairpin construct against the *β*-1, 3-glucan synthase gene *FcGls1* of *F. culmorum,* or a triple combination of *FcGls1* with mitogen-activated protein kinase (*FcFmk1*) and chitin synthase V myosin-motor domain (*FcChsV*), also showed enhanced resistance in leaf and spike inoculation assays under greenhouse and near-field conditions, respectively [158].

In other examples of RNAi, targeting the myosin-5 gene (*FaMyo5*) from *F. asiaticum* provided disease resistance in wheat [159]. In further examples, wheat transformed with a vector expressing a double-stranded RNA, targeting the mitogen-activated protein kinase kinase gene (*PsFUZ7*) from *Puccinia striiformis* f. sp. *tritici* exhibited strong resistance to stripe rust [160]. Transgenic wheat plants expressing siRNAs targeting the catalytic subunit protein kinase A gene from *P. striiformis* f. sp. *Tritici,* displayed high levels of stable and durable resistance throughout the T_3 to T_4 generations [161]. Stable expression of hairpin RNAi constructs with a sequence homologous to mitogen-activated protein-kinase from *P. triticina* (*PtMAPK1*), or a cyclophilin (*PtCYC1*) encoding gene, in susceptible wheat plants showed efficient silencing of the corresponding genes in the interacting fungus, resulting in disease resistance throughout the T_2 generation [162].

The grain aphid (*Sitobion avenae*) chitin synthase 1 gene (*CHS1*) was also targeted with HIGS. After feeding on the representative T_3 transgenic wheat lines, *CHS1* expression levels in grain aphid decreased by half, and both total and moulting aphid numbers reduced significantly [163]. Other target genes used for the control of grain aphids by RNAi in transgenic wheat were lipase maturation factor-like2 gene from pea aphid *Acyrthosiphon pisum* [164], carboxylesterase gene [165], *Hpa1* [166], and a gene encoding a salivary sheath protein [167]. Feeding on these transgenic plants resulted in significant reductions in survival and reproduction of aphids. It seems that the scope of RNAi-induced pest and disease resistance is as broad as the number of different pathogen and pest species that infect and cause damage to wheat.

5. Site-Specific Nucleases for Targeted Genome Modifications in Wheat

Targeted genome engineering using nucleases, such as zinc finger nucleases (ZFNs) and TAL effector nucleases (TAL-ENs), were developed in the late 20th century as innovative tools to generate mutations at specific genetic loci [168]. Nuclease-based mutagenesis relies on induced site-specific DNA double-strand breaks (DSBs), which are either repaired by error-prone nonhomologous end joining (NHEJ), or high-fidelity homologous recombination (HR). The former often results in insertions or deletions (InDels) at the cleavage site leading to loss-of-function gene knockouts, whereas the latter leads to precise genome modification. In 2012, the field of eukaryotic genome editing was revolutionized by the introduction of CRISPR/Cas9 (bacterial Clustered Regularly Interspaced Short Palindromic Repeats) technology [169]. This technology confers targeted gene mutagenesis by a Cas9 nuclease that is guided by small RNAs (sgRNAs) to the target gene through base pairing. This is in contrast to the DNA-recognition protein domain that must be specifically tailored for each DNA target in the case of ZFNs and TAL-ENs. Because of its universality and operational simplicity compared to ZFNs and TALEN genome editing systems, CRISPR/Cas9 has rapidly superseded these earlier editing systems and been adopted by the majority of the scientific community [170, 171].

In wheat, the principal applicability of CRISPR/Cas9 was demonstrated in protoplasts and suspension cultures, where multiple genes were successfully targeted in the year following the publication of the original CRISPR/Cas9 principle [172–174]. Original methods for plant genome editing rely on the delivery of plasmids carrying cassettes for the coexpression of Cas9 and sgRNA, either by *Agrobacterium* or particle bombardment. For gene editing in wheat, a Cas9 protein containing one or more signals for nuclear localization is expressed from a codon optimized gene under the control of RNA polymerase II promoters such as CaMV35S or ZmUbi, while the sgRNA is usually expressed from a polymerase III promoter (most commonly, rice or wheat U6 and U3 promoters).

One of the additional advantages of the CRISPR/Cas9 system is its potential for multiplexing, i.e., the simultaneous targeting of several genes with a single molecular construct. Multiple sgRNAs can be introduced either as separate transcription units, or in polycistronic form [175]. In bread wheat, editing has been reported in PEG-transfected protoplasts [172, 173, 176–178], electroporated microspores [179], and cell suspension cultures transformed by *Agrobacterium* [174]. Edited wheat plants have been regenerated from immature embryos, immature embryo-derived callus, or shoot apical meristems transformed via particle bombardment [176, 177, 180–184] or *Agrobacterium* [185, 186].

Recently, protocols for DNA-free editing of wheat by delivering *in vitro* transcripts, or ribonucleoprotein complexes (RNPs) of CRISPR/Cas9 by particle bombardment, have been developed [176, 177]. The authors claim that these methods not only eliminate random integration of the CRISPR/Cas9 coding DNA elements into the targeted genome, but also reduce off-target effects. Thus, these advances allow for the production of completely transgene-free mutants in bread wheat with high precision. The main limitation of these transgene-free protocols is the lack of selection in the transformation and regeneration process. Another optimized delivery system has been developed by Gil-Humanes et al. [177]. Here the authors used replicated vectors based on the wheat dwarf virus (Geminiviridae) for cereal genome engineering. It was shown that, due to increased copy number of the system components, virus-derived replicons increase gene targeting efficiency greater than 10-fold in wheat callus cells and protoplasts, compared to the nonreplicating control. The virus-based CRISPR/Cas9 system also promoted multiplexed gene targeted integration in different loci of the polyploid wheat genome by homologous recombination.

Since the advent of the principle of RNA guided nuclease genome editing, a number of additional tools for genome modifications and functional genomics studies have been developed [187]. The DNA binding ability of Cas9 and Cas12 has been used to develop tools for various applications, such as transcriptional regulation and fluorescence-based imaging of specific chromosomal loci in plant genomes. Another nuclease, Cas13, has been applied to degrade mRNAs and combat viral RNA replication [188]. Orthologues of Cas9 found in other bacterial species such as *Neisseria meningitides* [189], *Staphylococcus aureus* [190], and *Campylobacter jejuni* [191] have different and more complex Protospacer adjacent motif (PAM) sequences. These PAM sequences function in their native bacterial hosts to direct the CRISPR/Cas9 complex to the target sequence and not to the CRISPR/Cas9 locus. Although the use of these orthologous Cas9 proteins does limit the available target sequences for genome editing, it also reduces off-target edits. Most recently, systems for targeted base editing in wheat have been established by fusing a cytidine deaminase [181] or adenosine deaminase [192], to the Cas9 nickase for C/G to T/A or A/T to G/C conversion. In these systems, the efficiency of base editing was enhanced by using a Cas9-based nickase instead of an inactive Cas9. As an example, the base editor Cas9-APOBEC3A was used to edit *TaMTL* (*MATRILINEAL*) encoding a sperm-specific phospholipase [182]. Loss of function of *MTL* triggers haploid induction in maize [193]. Ten base-edited wheat mutants with *TaMTL* knock-out were identified at a frequency of 16.7%, with three being homozygous for all six alleles without InDels. Functional analysis of wheat mutants with *TaMTL* knock-out is still to be completed. Other nucleases with similar editing functions to that of Cas9 have been identified [194]. Most notably, Cpf1 possesses both DNase and RNase activity and cleaves DNA to generate four to five bp $5'$-overhangs, potentially enhancing insertion of DNA sequences by homologous recombination. The Cpf1-based editing system has been successfully applied in plants, but not in wheat at this stage [195, 196].

Although many agriculturally important traits of wheat have been targeted by genome editing, some of the main ones include the following: (i) resistance/tolerance to biotic and abiotic stresses, (ii) yield and grain quality, and (iii) male sterility.

(i) The first successful experiment using the CRISPR/ Cas9 system in wheat was editing of *TaMLO*, a powdery mildew-resistance locus. Powdery mildew diseases caused by *Blumeria graminis* f. sp. *tritici* result in significant wheat yield losses, and knock-out of the *TaMLO* leads to disease resistance. The mutation frequency of *TaMLO* in protoplasts was 28.5% [172]. Further, Wang et al. [180] described editing of the *TaMLO-A1* allele by the CRISPR/Cas9 system and simultaneous editing of three homoeoalleles of *TaMLO* in hexaploid bread wheat using the TALEN nuclease. The mutation frequency of regenerated *TaMLO*-edited wheat (5.6%) was similar for both editing methods. More recently, Zhang et al. [174] used CRISPR/Cas9 technology to generate *Taedr1* wheat lines by simultaneous knock-down of the three homologs of wheat *TaEDR1*, a negative regulator of powdery mildew resistance. The mutated plants were resistant to

powdery mildew and did not show mildew-induced cell death [174].

The lipoxygenase genes, *TaLpx1* and *TaLox2*, attracted attention as potential subjects for gene editing in relation to resistance to *Fusarium*, one of the most devastating fungal diseases in wheat. Lipoxygenases hydrolyze polyunsaturated fatty acids and initiate biosynthesis of oxylipins, playing a role in the activation of jasmonic acid-mediated defence responses in plants. Silencing of the *TaLpx*-1 gene has resulted in resistance to *Fusarium graminearum* in wheat [197]. *TaLpx1* and *TaLox2* genes were edited in protoplasts with a mutation frequency of 9% and 45%, respectively [173, 183]. Wheat plants with mutated *TaLOX2* were obtained with a frequency of 9.5%, of which homozygous mutants accounted for 44.7% [198].

The CRISPR/Cas9 system was also used for editing a wheat homolog of *TaCer9* (*ECERIFERUM9*) with the goal to improve drought tolerance and water use efficiency [199]. Mutation of the *AtCer9* gene in *A. thaliana*, encoding an E3 ubiquitin ligase, causes elevated amounts of 18-carbon-length cutin monomers and very-long-chain free fatty acids (C24, C26) in cuticular wax, both of which are associated with elevated cuticle membrane thickness and drought tolerance [200].

The Cas9 nickase fused to a human cytidine deaminase, APOBEC3A, was used to produce herbicide resistant wheat plants through editing of *TaALS* [182]. ALS encodes acetolactate synthase, the first enzyme of the branched-chain amino acid biosynthesis. Amino acid substitutions in *TaALS* confer resistance to the sulfonylurea class of herbicides. Wheat plants with a mutated *TaALS* gene were obtained in high frequency (22.5%, 27/120). Among them, two plants had six alleles simultaneously edited and were nicosulfuron resistant.

(ii) With the aim of enhancing grain size and yield, several genes have been edited by the CRISPR/Cas9 system: *TaGASR7* [176, 184, 198], *TaGW2* [198, 199, 201], and *TaDEP1* [198]. *TaGASR7*, a member of the *Snakin/GASA* gene family, has been associated with grain length in wheat. A CRISPR/Cas9 vector designed to target *TaGASR7* was delivered by particle bombardment into shoot apical meristems. Eleven (5.2%) of the 210 bombarded plants carried mutant alleles, and the mutations of three (1.4%) of these were inherited in the next generation [184]. Transiently expressing the CRISPR/Cas9 DNA and using the CRISPR/Cas9 RNP-mediated method were also highly effective for *TaGASR7* editing [176, 198]. The *TaGW2* gene encodes a previously unknown RING-type E3 ubiquitin ligase that was reported to be a negative regulator of grain size and thousand grain weight in wheat [198, 199, 201]. Recent studies detailing the functionality of the allelic *TaGW2* genes through genome-specific knockouts [201], showed that the *TaGW2* gene in wheat acts by regulating the gibberellin hormone biosynthesis pathway [202], principally confirming the parallel functions of these genes in rice and wheat. T_1 plants carrying knock-out mutations in all three copies of the *TaGW2* gene (genotype *aabbdd*) showed significantly increased thousand grain weight (27.7%), grain area (17.0%), grain width (10.9%), and grain length (6.1%) compared to the wild-type cultivar [183].

Another gene editing target is *DENSE AND ERECT PANICLE 1* (*DEP1*), which encodes a G protein gamma-subunit in rice that is involved in the regulation of erect panicles, number of grains per panicle, nitrogen uptake, and stress tolerance through the G protein signalling pathway [203]. Zhang et al. [198] applied the CRISPR/Cas9 system to target *TaDEP1* and *TaNAC2* editing in wheat. CRISPR/Cas9-induced mutation of *TaNAC2* (a member of a plant specific transcription factor family) and *TaDEP1* in wheat plants was successful in 2% of cases [198]. One possible outcome of the loss of *TaNAC2* function could be increased size of grain and changes in stress responses. *TaPIN1* (*PINFORMED1*) was edited using CRISPR/Cas9 technology with a frequency of 1% [198]. The plant specific *PIN* family of efflux carriers comprises integral membrane proteins and has been associated with polar auxin transport during embryogenesis and endosperm development. It should be noted that no homozygous mutants were obtained for *TaDEP1*, *TaNAC2,* and *TaPIN1* by CRISPR/Cas9 editing [198] probably because they were not viable.

CRISPR/Cas9 technology was also successful in obtaining low immunogenic wheat. Sánchez-León et al. [204] have shown that CRISPR/Cas9 technology can be used to efficiently reduce the amount of alfa-gliadins in the seeds, providing bread and durum wheat lines with reduced immunoreactivity for consumers with coeliac disease. Twenty-one mutant lines were generated (15 bread wheat and 6 durum wheat), all showing a strong reduction in α-gliadins. Up to 35 of the 45 different genes identified in the wild-type were mutated in one of the lines, and immunoreactivity, as measured by competitive ELISA assays using two monoclonal antibodies, was reduced by 85%.

(iii) Male sterility and the induction of haploids can provide a powerful tool in cultivar breeding and genetic analysis. Generation of male-sterile and doubled haploid plants can facilitate development of hybrid seed production in wheat. In maize, the male-sterile gene 45 (*Ms45*) encodes a strictosidine synthase-like enzyme that was shown to be required for male fertility and required for exine structuring and pollen development [185]. Genetic analysis of mutated plants obtained by CRISPR/Cas9 technology demonstrated that all three wheat *Ms45* homeologues contribute to male fertility. Mutant plants, *Tams45-abd*, abort pollen development resulting in male sterility.

These examples provide an insight into the many ways in which modern genome modification technologies are being used to mine the core research findings from model plant transgenesis, and finally harness that understanding to drive essential crop. The ability to enact targeted changes to the genome has revolutionized genetic modification for polyploid crop species such as wheat. There are now some enticing indications that the nearest remaining hurdle, in the form of low transformation frequencies in favoured breeding genotypes, may soon be overcome. What remains is for political entities, and the societies they represent, to adopt a more receptive view to the very real and practical benefits these technologies may provide.

6. Conclusions and Prospective Developments

Demand for wheat is projected to rise at a rate of 1.6% annually until 2050 as a result of increased population and prosperity. Consequently, average global wheat yields on a per hectare basis will need to increase to approximately 5 tonnes per ha from the current 3.3 tonnes [205]. Bread wheat has a very complicated hexaploid genome and, therefore, further progress in breeding of this crop is strongly dependent upon knowledge of functional genomics. Thus, it is necessary to identify the most important key-genes, their structure, role, and function in the development of wheat plants and finally for higher grain yield and better quality. Based on the knowledge of functional genomics, plant biologists can alter the structures and functions of selected key-genes through "genetic manipulation". Genetic transformation is a very powerful tool for generating scientific proof of the roles and functions of key-genes. The authors of this review are not in a position to discuss the applications of GM-wheat in world breeding practice, since this is beyond the scope of this review. However, the term "genetic manipulation" is very broad and includes other molecular approaches that generate products that fall outside the traditional definition of "GM". RNA interference and CRISPR/Cas9 represent modern and very advanced GM technologies that in a growing number of countries, such as USA and Canada, result in products that attract the same level of regulation as the products of traditional breeding techniques. Such "end-product-based" rather than "process-based" regulation, presents a far more favourable environment for the progression of molecular-based breeding technologies, which can and should change the future of wheat breeding across the world. However, all advanced methods will remain simply "laboratory tools" if their application is not connected with wheat breeders currently working by traditional methods. Therefore, we see the chance for real progress and positive future prospects through effective collaborations between plant molecular geneticists and wheat breeders. The application of novel methods and analysis of genetically manipulated wheat plants for their utility in breeding can be translated to the field through the introgression of genetic traits into conventional wheat breeding programs.

Awareness and concerns are also growing regarding the huge economic gaps between developed and developing countries. Underdeveloped countries are more reliant on agriculture for their overall economies, yet they have fewer opportunities to develop or collaborate on projects involving modern technologies for genetic manipulation in wheat and most other crop plants. Therefore, researchers in developed countries must take the lead and assume responsibility for sharing and freely disseminating their results and genetically manipulated wheat germplasm accessions with breeders in the developing world. This act of "research donation" can enrich lives and communities where needs are the greatest and uphold the future security and sustainability of wheat production, a key global commodity.

Conflicts of Interest

Acknowledgments

References

[1] C. Uauy, "Plant genomics: unlocking the genome of wheat's progenitor," *Current Biology*, vol. 27, no. 20, pp. R1122–R1124, 2017.

[2] R. E. Evenson and D. Gollin, "Assessing the impact of the green revolution, 1960 to 2000," *Science*, vol. 300, no. 5620, pp. 758–762, 2003.

[3] P. Langridge, "Wheat genomics and the ambitious targets for future wheat production," *Genome*, vol. 56, no. 10, pp. 545–547, 2013.

[4] R. J. Henry, P. Rangan, and A. Furtado, "Functional cereals for production in new and variable climates," *Current Opinion in Plant Biology*, vol. 30, pp. 11–18, 2016.

[5] F. Altpeter, N. M. Springer, L. E. Bartley et al., "Advancing crop transformation in the era of genome editing," *The Plant Cell*, vol. 28, pp. 1510–1520, 2016.

[6] F. Altpeter, N. Baisakh, R. Beachy et al., "Particle bombardment and the genetic enhancement of crops: myths and realities," *Molecular Breeding*, vol. 15, no. 3, pp. 305–327, 2005.

[7] H. Wu, F. S. Awan, A. Vilarinho et al., "Transgene integration complexity and expression stability following biolistic or agrobacterium-mediated transformation of sugarcane," *In Vitro Cellular & Developmental Biology - Plant*, vol. 51, no. 6, pp. 603–611, 2015.

[8] Y. Ishida, M. Tsunashima, Y. Hiei, and T. Komari, "Wheat (triticum aestivum L.) transformation using immature embryos," in *Agrobacterium Protocols*, vol. 1223 of *Methods in Molecular Biology*, pp. 189–198, Springer New York, New York, NY, USA, 2015.

[9] A. Pellegrineschi, L. M. Noguera, B. Skovmand et al., "Identification of highly transformable wheat genotypes for mass production of fertile transgenic plants," *Genome*, vol. 45, no. 2, pp. 421–430, 2002.

[10] C. Tassy and P. Barret, "Biolistic transformation of wheat," in *Wheat Biotechnology: Methods and Protocols*, P. Bhalla and M. Singh, Eds., vol. 1679 of *Methods in Molecular Biology*, pp. 141–152, Springer New York, New York, NY, 2017.

[11] H. D. Jones, A. Doherty, and C. A. Sparks, "Transient transformation of plants," in *Plant Genomics*, J. M. Walker, Ed., vol. 513 of *Methods in Molecular Biology*, pp. 131–152, Humana Press, Totowa, NJ, USA, 2009.

[12] A. C. Brasileiro, C. Tourneur, J. Leple, V. Combes, and L. Jouanin, "Expression of the mutant arabidopsis thaliana acetolactate synthase gene confers chlorsulfuron resistance to transgenic poplar plants," *Transgenic Research*, vol. 1, no. 3, pp. 133–141, 1992.

[13] J. Mullen, G. Adam, A. Blowers, and E. Earle, "Biolistic transfer of large DNA fragments to tobacco cells using YACs retro fitted for plant transformation," *Molecular Breeding*, vol. 4, no. 5, pp. 449–457, 1998.

[14] Y. Wang, H. Zeng, X. Zhou et al., "Transformation of rice with large maize genomic DNA fragments containing high content repetitive sequences," *Plant Cell Reports*, vol. 34, no. 6, pp. 1049–1061, 2015.

[15] S. De Buck, C. De Wilde, M. Van Montagu, and A. Depicker, "T-DNA vector backbone sequences are frequently integrated into the genome of transgenic plants obtained by Agrobacterium-mediated transformation," *Molecular Breeding*, vol. 6, pp. 459–468, 2000.

[16] B. Ülker, Y. Li, M. G. Rosso, E. Logemann, I. E. Somssich, and B. Weisshaar, "T-DNA-mediated transfer of Agrobacterium tumefaciens chromosomal DNA into plants," *Nature Biotechnology*, vol. 26, no. 9, pp. 1015–1017, 2008.

[17] A. Gadaleta, A. Giancaspro, A. Blechl, and A. Blanco, "Phosphomannose isomerase, pmi, as a selectable marker gene for durum wheat transformation," *Journal of Cereal Science*, vol. 43, pp. 31–37, 2006.

[18] A. Gadaleta, A. Giancaspro, A. E. Blechl, and A. Blanco, "A transgenic durum wheat line that is free of marker genes and expresses 1Dy10," *Journal of Cereal Science*, vol. 48, no. 2, pp. 439–445, 2008.

[19] K. Wang, H. Liu, L. Du, and X. Ye, "Generation of marker-free transgenic hexaploid wheat via an agrobacterium-mediated co-transformation strategy in commercial Chinese wheat varieties," *Plant Biotechnology Journal*, vol. 15, no. 5, pp. 614–623, 2017.

[20] T. Risacher, M. Craze, S. Bowden, W. Paul, and T. Barsby, "Highly efficient agrobacterium-mediated transformation of wheat via in planta inoculation," *Methods in Molecular Biology*, vol. 478, pp. 115–124, 2009.

[21] J. M. Zale, S. Agarwal, S. Loar, and C. M. Steber, "Evidence for stable transformation of wheat by floral dip in agrobacterium tumefaciens," *Plant Cell Reports*, vol. 28, no. 6, pp. 903–913, 2009.

[22] H. Hamada, Q. Linghu, Y. Nagira, R. Miki, N. Taoka, and R. Imai, "An in planta biolistic method for stable wheat transformation," *Scientific Reports*, vol. 7, article 11443, 2017.

[23] S. Martin-Ortigosa and K. Wang, "Proteolistics: a biolistic method for intracellular delivery of proteins," *Transgenic Research*, vol. 23, no. 5, pp. 743–756, 2014.

[24] C. Remacle, P. Cardol, N. Coosemans, M. Gaisne, and N. Bonnefoy, "High-efficiency biolistic transformation of Chlamydomonas mitochondria can be used to insert mutations in complex I genes," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 103, no. 12, pp. 4771–4776, 2006.

[25] L. Cheng, H. Li, B. Qu et al., "Chloroplast transformation of rapeseed (Brassica napus) by particle bombardment of cotyledons," *Plant Cell Reports*, vol. 29, no. 4, pp. 371–381, 2010.

[26] Y. Gleba, V. Klimyuk, and S. Marillonnet, "Magnifection - a new platform for expressing recombinant vaccines in plants," *Vaccine*, vol. 23, no. 17-18, pp. 2042–2048, 2005.

[27] S. Ueki, S. Magori, B. Lacroix, and V. Citovsky, "Transient gene expression in epidermal cells of plant leaves by biolistic DNA delivery," in *Biolistic DNA Delivery: Methods and Protocols, Methods in Molecular Biology*, S. Sudowe and A. B. Reske-Kunz, Eds., vol. 940, pp. 17–26, 2013.

[28] P. Krenek, O. Samajova, I. Luptovciak, A. Doskocilova, G. Komis, and J. Samaj, "Transient plant transformation mediated by Agrobacterium tumefaciens: Principles, methods and applications," *Biotechnology Advances*, vol. 33, no. 6, pp. 1024–1042, 2015.

[29] N. Goncharov and E. Kondratenko, "Origin, domestication and evolution of wheats," *Vestnik VOGiS*, vol. 12, pp. 159–179, 2008, http://www.bionet.nsc.ru/vogis/pict_pdf/2008/t12_1_2/vogis_12_1_2_15.pdf.

[30] J. D. Montenegro, A. A. Golicz, P. E. Bayer et al., "The pangenome of hexaploid bread wheat," *The Plant Journal*, vol. 90, no. 5, pp. 1007–1013, 2017.

[31] P. L. Bhalla, A. Sharma, and M. B. Singh, "Enabling molecular technologies for trait improvement in wheat," in *Wheat Biotechnology*, vol. 1679 of *Methods in Molecular Biology*, pp. 3–24, Springer New York, New York, NY, USA, 2017.

[32] R. Appels, K. Eversole, C. Feuillet et al., "Shifting the limits in wheat research and breeding using a fully annotated reference genome," *Science*, vol. 361, no. 6403, article eaar7191, 2018.

[33] A. K. Shrawat and C. L. Armstrong, "Development and application of genetic engineering for wheat improvement," *Critical Reviews in Plant Sciences*, vol. 37, no. 5, pp. 335–421, 2018.

[34] V. Vasil, A. M. Castillo, M. E. Fromm, and I. K. Vasil, "Herbicide resistant fertile transgenic wheat plants obtained by micro-projectile bombardment of regenerable embryogenic callus," *Nature Biotechnology*, vol. 10, no. 6, pp. 667–674, 1992.

[35] M. Cheng, J. E. Fry, S. Pang et al., "Genetic transformation of wheat mediated by agrobacterium tumefaciens," *Plant Physiology*, vol. 115, no. 3, pp. 971–980, 1997.

[36] Q. Yao, L. Cong, J. L. Chang, K. X. Li, G. X. Yang, and G. Y. He, "Low copy number gene transfer and stable expression in a commercial wheat cultivar via particle bombardment," *Journal of Experimental Botany*, vol. 57, no. 14, pp. 3737–3746, 2006.

[37] C. Tassy, A. Partier, M. Beckert, C. Feuillet, and P. Barret, "Biolistic transformation of wheat: increased production of plants with simple insertions and heritable transgene expression," *Plant Cell, Tissue and Organ Culture (PCTOC)*, vol. 119, no. 1, pp. 171–181, 2014.

[38] A. Ismagul, N. Yang, E. Maltseva et al., "A biolistic method for high-throughput production of transgenic wheat plants with single gene insertions," *BMC Plant Biology*, vol. 18, pp. 135–143, 2018.

[39] Q. Yao, L. Cong, G. He et al., "Optimization of wheat co-transformation procedure with gene cassettes resulted in an improvement in transformation frequency," *Molecular Biology Reports*, vol. 34, pp. 61–67, 2007.

[40] A. Partier, G. Gay, C. Tassy, M. Beckert, C. Feuillet, and P. Barret, "Molecular and FISH analyses of a 53-kbp intact DNA fragment inserted by biolistics in wheat (triticum aestivum L.) genome," *Plant Cell Reports*, vol. 36, no. 10, pp. 1547–1559, 2017.

[41] M. Wright, J. Dawson, E. Dunder et al., "Efficient biolistic transformation of maize (Zea mays L.) and wheat (Triticum aestivum L.) using the phosphomannose isomerase gene, pmi, as the selectable marker," *Plant Cell Reports*, vol. 20, no. 5, pp. 429–436, 2001.

[42] T. Richardson, J. Thistleton, T. J. Higgins, C. Howitt, and M. Ayliffe, "Efficient agrobacterium transformation of elite wheat germplasm without selection," *Plant Cell, Tissue and Organ Culture (PCTOC)*, vol. 119, no. 3, pp. 647–659, 2014.

[43] G. Hensel, C. Marthe, and J. Kumlehn, "Agrobacterium-mediated transformation of wheat using immature embryos," in *Wheat Biotechnology*, P. Bhalla and M. Singh, Eds., vol. 1679 of *Methods in Molecular Biology*, pp. 129–139, Springer New York, New York, NY, USA, 2017.

[44] K. Wang, B. Riaz, and X. Ye, "Wheat genome editing expedited by efficient transformation techniques: progress and perspectives," *The Crop Journal*, vol. 6, no. 1, pp. 22–31, 2018.

[45] G. R. Lazo, P. A. Stein, and R. A. Ludwig, "A DNA transformation-competent Arabidopsis genomic library in agrobacterium," *Nature Biotechnology*, vol. 9, no. 10, pp. 963–967, 1991.

[46] N. R. Peters, S. Ackerman, and E. A. Davis, "A modular vector for Agrobacterium mediated transformation of wheat," *Plant Molecular Biology Reporter*, vol. 17, no. 4, pp. 323–331, 1999.

[47] N. V. Larebeke, G. Engler, M. Holsters et al., "Large plasmid in agrobacterium tumefaciens essential for crown gall-inducing ability," *Nature*, vol. 252, no. 5479, pp. 169-170, 1974.

[48] Y. Wang, M. Xu, G. Yin, L. Tao, D. Wang, and X. Ye, "Transgenic wheat plants derived from Agrobacterium-mediated transformation of mature embryo tissues," *Cereal Research Communications*, vol. 37, pp. 1–12, 2009.

[49] P. Supartana, T. Shimizu, M. Nogawa et al., "Development of simple and efficient in planta transformation method for wheat (triticum aestivum l.) using agrobacterium tumefaciens," *Journal of Bioscience and Bioengineering*, vol. 102, no. 3, pp. 162–170, 2006.

[50] H. Aadel, R. Abdelwahd, S. Udupa et al., "Agrobacterium-mediated transformation of mature embryo tissues of bread wheat (triticum aestivum L.) genotypes," *Cereal Research Communications*, vol. 46, pp. 10–20, 2018.

[51] E. Medvecká and W. A. Harwood, "Wheat (triticum aestivum L.) transformation using mature embryos," in *Agrobacterium Protocols: Volume 1*, K. Wang, Ed., vol. 1223 of *Methods in Molecular Biology*, pp. 199–209, Springer New York, New York, NY, USA, 2015.

[52] H. Chauhan and P. Khurana, "Wheat genetic transformation using mature embryos as explants," in *Wheat Biotechnology: Methods and Protocols*, P. Bhalla and M. Singh, Eds., vol. 1679 of *Methods in Molecular Biology*, pp. 153–167, Humana Press, New York, NY, USA, 2017.

[53] T. Zhao, S. Zhao, H. Chen et al., "Transgenic wheat progeny resistant to powdery mildew generated by agrobacterium inoculum to the basal portion of wheat seedling," *Plant Cell Reports*, vol. 25, no. 11, pp. 1199–1204, 2006.

[54] A. Razzaq, I. A. Hafiz, I. Mahmood, and A. Hussain, "Development of in planta transformation protocol for wheat," *African Journal of Biotechnology*, vol. 10, no. 5, pp. 740–750, 2011.

[55] S. Rustgi, N. O. Ankrah, R. A. T. Brew-Appiah et al., "Doubled haploid transgenic wheat lines by microspore transformation," in *Wheat Biotechnology: Methods and Protocols*, P. Bhalla and M. Singh, Eds., vol. 1679 of *Methods in Molecular Biology*, pp. 213–234, Humana Press, New York, NY, USA, 2017.

[56] L. Folling and A. Olesen, "Transformation of wheat (triticum aestivum L.) microspore-derived callus and microspores by particle bombardment," *Plant Cell Reports*, vol. 20, no. 7, pp. 629–636, 2001.

[57] R. A. Brew-Appiah, N. Ankrah, W. Liu et al., "Generation of doubled haploid transgenic wheat lines by microspore transformation," *PLoS ONE*, vol. 8, no. 11, article e80155, 2013.

[58] K. Singer, Y. M. Shiboleth, J. Li, and T. Tzfira, "Formation of complex extrachromosomal T-DNA structures in agrobacterium tumefaciens-infected plants," *Plant Physiology*, vol. 160, pp. 511–522, 2012.

[59] S. Kim and G. An, "Bacterial transposons are co-transferred with T-DNA to rice chromosomes during Agrobacterium-mediated transformation," *Molecules and Cells*, vol. 33, no. 6, pp. 583–589, 2012.

[60] A. Pitzschke, "Agrobacterium infection and plant defense-transformation success hangs by a thread," *Frontiers in Plant Science*, vol. 4, no. 519, 2013.

[61] E. Zuniga-Soto, E. Mullins, and B. Dedicova, "Ensifer-mediated transformation: an efficient non-agrobacterium protocol for the genetic modification of rice," *SpringerPlus*, vol. 4, pp. 1–10, 2015.

[62] M. S. Greer, I. Kovalchuk, and F. Eudes, "Ammonium nitrate improves direct somatic embryogenesis and biolistic transformation of Triticum aestivum," *New Biotechnology*, vol. 26, no. 1-2, pp. 44–52, 2009.

[63] Y. Dan, "Biological functions of antioxidants in plant transformation," *In Vitro Cellular & Developmental Biology - Plant*, vol. 44, no. 3, pp. 149–161, 2008.

[64] Y. Dan, C. L. Armstrong, J. Dong et al., "Lipoic acid—an unique plant transformation enhancer," *In Vitro Cellular & Developmental Biology - Plant*, vol. 45, no. 6, pp. 630–638, 2009.

[65] Y. Coskun, R. E. Duran, C. Savaskan, T. Demirci, and M. T. Hakan, "Efficient plant regeneration with arabinogalactan-proteins on various ploidy levels of cereals," *Journal of Integrative Agriculture*, vol. 12, no. 3, pp. 420–425, 2013.

[66] R. Kumar, H. M. Mamrutha, A. Kaur et al., "Development of an efficient and reproducible regeneration system in wheat (triticum aestivum L.)," *Physiology and Molecular Biology of Plants*, vol. 23, no. 4, pp. 945–954, 2017.

[67] D. Miroshnichenko, I. Chaban, M. Chernobrovkina, S. Dolgov, and R. P. Niedz, "Protocol for efficient regulation of in vitro morphogenesis in einkorn (triticum monococcum L.), a recalcitrant diploid wheat species," *PLoS ONE*, vol. 12, no. 3, Article ID e0173533, 2017.

[68] R. A. Jefferson, "Assaying chimeric genes in plants: the GUS gene fusion system," *Plant Molecular Biology Reporter*, vol. 5, no. 4, pp. 387–405, 1987.

[69] M. Hraška, S. Rakouský, and V. Čurn, "Green fluorescent protein as a vital marker for non-destructive detection of transformation events in transgenic plants," *Plant Cell, Tissue and Organ Culture*, vol. 86, no. 3, pp. 303–318, 2006.

[70] J. Zhang, D. Yu, Y. Zhang et al., "Vacuum and co-cultivation agroinfiltration of (Germinated) seeds results in tobacco rattle virus (TRV) mediated whole-plant virus-induced gene silencing (VIGS) in wheat and maize," *Frontiers in Plant Science*, vol. 8, no. 393, 2017.

[71] S. Naqvi, G. Farré, G. Sanahuja, T. Capell, C. Zhu, and P. Christou, "When more is better: multigene engineering in plants," *Trends in Plant Science*, vol. 15, no. 1, pp. 48–56, 2010.

[72] W. Yu, Y. Yau, and J. A. Birchler, "Plant artificial chromosome technology and its potential application in genetic engineering," *Plant Biotechnology Journal*, vol. 14, no. 5, pp. 1175–1182, 2016.

[73] J. Yuan, Q. Shi, X. Guo et al., "Site-specific transfer of chromosomal segments and genes in wheat engineered chromosomes," *Journal of Genetics and Genomics*, vol. 44, no. 11, pp. 531–539, 2017.

[74] A. Furtado, R. J. Henry, and A. Pellegrineschi, "Analysis of promoters in transgenic barley and wheat," *Plant Biotechnology Journal*, vol. 7, no. 3, pp. 240–253, 2009.

[75] S. Alotaibi, C. Sparks, M. Parry, A. Simkin, and C. Raines, "Identification of leaf promoters for use in transgenic wheat," *Plants*, vol. 7, no. 2, p. E27, 2018.

[76] G. Hensel, A. Himmelbach, W. Chen, D. K. Douchkov, and J. Kumlehn, "Transgene expression systems in the triticeae cereals," *Journal of Plant Physiology*, vol. 168, pp. 30–44, 2011.

[77] N. Borisjuk, M. Hrmova, and S. Lopato, "Transcriptional regulation of cuticle biosynthesis," *Biotechnology Advances*, vol. 32, no. 2, pp. 526–540, 2014.

[78] H. Wang, H. Wang, H. Shao, and X. Tang, "Recent advances in utilizing transcription factors to improve plant abiotic stress tolerance by transgenic technology," *Frontiers in Plant Science*, vol. 7, no. 67, 2016.

[79] H. Bi, J. Shi, N. Kovalchuk et al., "Overexpression of the TaSHN1 transcription factor in bread wheat leads to leaf surface modifications, improved drought tolerance, and no yield penalty under controlled growth conditions," *Plant, Cell & Environment*, vol. 41, no. 11, pp. 2549–2566, 2018.

[80] R. P. Singh, P. K. Singh, J. Rutkoski et al., "Disease impact on wheat yield potential and prospects of genetic control," *Annual Review of Phytopathology*, vol. 54, pp. 303–322, 2016.

[81] A. E. Ricroch, "What will be the benefits of biotech wheat for european agriculture?" *Methods of Molecular Biology*, vol. 1679, pp. 25–35, 2017.

[82] B. Keller, T. Wicker, and S. G. Krattinger, "Advances in wheat and pathogen genomics: implications for disease control," *Annual Review of Phytopathology*, vol. 56, pp. 67–87, 2018.

[83] H. Budak, B. Hussain, Z. Khan, N. Z. Ozturk, and N. Ullah, "From genetics to functional genomics: improvement in drought signaling and tolerance in wheat," *Frontiers in Plant Science*, vol. 19, no. 1012, 2015.

[84] S. Khan, M. Li, S. Wang, and H. Yin, "Revisiting the role of plant transcription factors in the battle against abiotic stress," *International Journal of Molecular Sciences*, vol. 19, no. 6, p. E1634, 2018.

[85] Y. Yu, D. Zhu, C. Ma et al., "Transcriptome analysis reveals key differentially expressed genes involved in wheat grain development," *The Crop Journal*, vol. 4, no. 2, pp. 92–106, 2016.

[86] A. Nadolska-Orczyk, I. K. Rajchel, W. Orczyk, and S. Gasparis, "Major genes determining yield-related traits in wheat and barley," *Theoretical and Applied Genetics*, vol. 130, no. 6, pp. 1081–1098, 2017.

[87] R. J. Henry, A. Furtado, and P. Rangan, "Wheat seed transcriptome reveals genes controlling key traits for human preference and crop adaptation," *Current Opinion in Plant Biology*, vol. 45, pp. 231–236, 2018.

[88] V. C. Gegas, A. Nazari, S. Griffiths et al., "A genetic framework for grain size and shape variation in wheat," *The Plant Cell*, vol. 22, no. 4, pp. 1046–1056, 2010.

[89] Z. Su, C. Hao, L. Wang, Y. Dong, and X. Zhang, "Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat (triticum aestivum L.)," *Theoretical and Applied Genetics*, vol. 122, pp. 211–223, 2011.

[90] X. Song, W. Huang, M. Shi, M. Zhu, and H. Lin, "A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase," *Nature Genetics*, vol. 39, no. 5, pp. 623–630, 2007.

[91] J. Bednarek, A. Boulaflous, C. Girousse et al., "Down-regulation of the TaGW2 gene by RNA interference results in decreased grain size and weight in wheat," *Journal of Experimental Botany*, vol. 63, no. 16, pp. 5945–5955, 2012.

[92] Y. Hong, L. Chen, L. Du et al., "Transcript suppression of TaGW2 increased grain width and weight in bread wheat," *Functional & Integrative Genomics*, vol. 14, no. 2, pp. 341–349, 2014.

[93] W. Zalewski, W. Orczyk, S. Gasparis, and A. Nadolska-Orczyk, "HvCKX2 gene silencing by biolistic or Agrobacterium-mediated transformation in barley leads to different phenotypes," *BMC Plant Biology*, vol. 12, no. 206, 2012.

[94] P. J. Andralojc, E. Carmo-Silva, G. E. Degen, and M. A. J. Parry, "Increasing metabolic potential: C-fixation," *Essays in Biochemistry*, vol. 62, pp. 109–118, 2018.

[95] A. Prins, D. J. Orr, P. J. Andralojc, M. P. Reynolds, E. Carmo-Silva, and M. A. Parry, "Rubisco catalytic properties of wild and domesticated relatives provide scope for improving wheat photosynthesis," *Journal of Experimental Botany*, vol. 67, no. 6, pp. 1827–1838, 2016.

[96] M. T. Lin, A. Occhialini, P. J. Andralojc, M. A. Parry, and M. R. Hanson, "A faster Rubisco with potential to increase photosynthesis in crops," *Nature*, vol. 513, no. 7519, pp. 547–550, 2014.

[97] E. Carmo-Silva, J. C. Scales, P. J. Madgwick, and M. A. Parry, "Optimizing Rubisco and its regulation for greater resource use efficiency," *Plant, Cell & Environment*, vol. 38, no. 9, pp. 1817–1832, 2015.

[98] R. A. Richards, "Selectable traits to increase crop photosynthesis and yield of grain crops," *Journal of Experimental Botany*, vol. 51, pp. 447–458, 2000.

[99] Y. Taniguchi, H. Ohkawa, C. Masumoto et al., "Overproduction of C4 photosynthetic enzymes in transgenic rice plants: an approach to introduce the C4-like photosynthetic pathway into rice," *Journal of Experimental Botany*, vol. 59, no. 7, pp. 1799–1809, 2007.

[100] L. Hu, Y. Li, W. Xu et al., "Improvement of the photosynthetic characteristics of transgenic wheat plants by transformation with the maize C4 phosphoenolpyruvate carboxylase gene," *Plant Breeding*, vol. 131, no. 3, pp. 385–391, 2012.

[101] H. Zhang, W. Xu, H. Wang et al., "Pyramiding expression of maize genes encoding phosphoenolpyruvate carboxylase (PEPC) and pyruvate orthophosphate dikinase (PPDK) synergistically improve the photosynthetic characteristics of transgenic wheat," *Protoplasma*, vol. 251, no. 5, pp. 1163–1173, 2014.

[102] N. Qin, W. Xu, L. Hu et al., "Drought tolerance and proteomics studies of transgenic wheat containing the maize C4 phosphoenolpyruvate carboxylase (PEPC) gene," *Protoplasma*, vol. 253, no. 6, pp. 1503–1512, 2016.

[103] P. A. Peña, T. Quach, S. Sato et al., "Expression of the maize Dof1 transcription factor in wheat and sorghum," *Frontiers in Plant Science*, vol. 8, no. 434, 2017.

[104] Z. A. Myers and B. F. Holt III., "Nuclear factor-Y: still complex after all these years?" *Current Opinion in Plant Biology*, vol. 45, pp. 96–102, 2018.

[105] M. E. Zanetti, C. Rípodas, and A. Niebel, "Plant NF-Y transcription factors: key players in plant-microbe interactions, root development and adaptation to stress," *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, vol. 1860, no. 5, pp. 645–654, 2017.

[106] B. Qu, X. He, J. Wang et al., "A wheat CCAAT box-binding transcription factor increases the grain yield of wheat with less fertilizer input," *Plant Physiology*, vol. 167, no. 2, pp. 411–423, 2015.

[107] D. Yadav, Y. Shavrukov, N. Bazanova et al., "Constitutive overexpression of the TaNF-YB4 gene in transgenic wheat significantly improves grain yield," *Journal of Experimental Botany*, vol. 66, no. 21, pp. 6635–6650, 2015.

[108] X. He, B. Qu, W. Li et al., "The nitrate-inducible NAC transcription factor TaNAC2-5A controls nitrate response and increases wheat yield," *Plant Physiology*, vol. 169, pp. 1991–2005, 2015.

[109] B. Tian, S. K. Talukder, J. Fu, A. K. Fritz, and H. N. Trick, "Expression of a rice soluble starch synthase gene in transgenic wheat improves the grain yield under heat stress conditions," *In Vitro Cellular & Developmental Biology - Plant*, vol. 54, no. 3, pp. 216–227, 2018.

[110] M. Hu, X. Zhao, Q. Liu et al., "Transgenic expression of plastidic glutamine synthetase increases nitrogen uptake and yield in wheat," *Plant Biotechnology Journal*, vol. 16, no. 11, pp. 1858–1867, 2018.

[111] X. Zhu, S. P. Long, and D. R. Ort, "What is the maximum efficiency with which photosynthesis can convert solar energy into biomass?" *Current Opinion in Biotechnology*, vol. 19, no. 2, pp. 153–159, 2008.

[112] U. Sonnewald and J. Kossmann, "Starches-from current models to genetic engineering," *Plant Biotechnology Journal*, vol. 11, no. 2, pp. 223–232, 2013.

[113] R. Kumar, S. Mukherjee, and B. T. Ayele, "Molecular aspects of sucrose transport and its metabolism to starch during seed development in wheat: a comprehensive review," *Biotechnology Advances*, vol. 36, no. 4, pp. 954–967, 2018.

[114] N. Weichert, I. Saalbach, H. Weichert et al., "Increasing sucrose uptake capacity of wheat grains stimulates storage protein synthesis," *Plant Physiology*, vol. 152, no. 2, pp. 698–710, 2010.

[115] E. D. Smidansky, F. D. Meyer, B. Blakeslee, T. E. Weglarz, T. W. Greene, and M. J. Giroux, "Expression of a modified ADP-glucose pyrophosphorylase large subunit in wheat seeds stimulates photosynthesis and carbon metabolism," *Planta*, vol. 225, no. 4, pp. 965–976, 2007.

[116] G. Kang, W. Xu, G. Liu, X. Peng, T. Guo, and J. Bell, "Comprehensive analysis of the transcription of starch synthesis genes and the transcription factor RSR1 in wheat (triticum aestivum) endosperm," *Genome*, vol. 56, no. 2, pp. 115–122, 2013.

[117] G. Liu, Y. Wu, M. Xu et al., "Virus-Induced gene silencing identifies an important role of the TaRSR1 transcription factor in starch synthesis in bread wheat," *International Journal of Molecular Sciences*, vol. 17, no. 10, p. 1557, 2016.

[118] M. Meenu and B. Xu, "A critical review on anti-diabetic and anti-obesity effects of dietary resistant starch," *Critical Reviews in Food Science and Nutrition*, pp. 1–13, 2018.

[119] A. Regina, A. Bird, D. Topping et al., "High-amylose wheat generated by RNA interference improves indices of large-bowel health in rats," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 103, no. 10, pp. 3546–3551, 2006.

[120] F. Sestili, M. Janni, A. Doherty et al., "Increasing the amylose content of durum wheat through silencing of the SBEIIa genes," *BMC Plant Biology*, vol. 10, no. 144, 2010.

[121] C. Vetrani, F. Sestili, M. Vitale et al., "Metabolic response to amylose-rich wheat-based rusks in overweight individuals," *European Journal of Clinical Nutrition*, vol. 72, no. 6, pp. 904–912, 2018.

[122] C. Zörb, U. Ludewig, and M. J. Hawkesford, "Perspective on wheat yield and quality with reduced nitrogen supply," *Trends in Plant Science*, vol. 23, no. 11, pp. 1029–1037, 2018.

[123] D. Zhao, A. P. Derkx, D. Liu, P. Buchner, M. J. Hawkesford, and E. Flemetakis, "Overexpression of a NAC transcription

factor delays leaf senescence and increases grain nitrogen concentration in wheat," *The Journal of Plant Biology*, vol. 17, no. 4, pp. 904–913, 2015.

[124] X. Zhao, X. Nie, X. Xiao, and H. Yang, "Over-expression of a tobacco nitrate reductase gene in wheat (triticum aestivum l.) increases seed protein content and weight without augmenting nitrogen supplying," *PLoS ONE*, vol. 8, no. 9, article e74678, 2013.

[125] F. M. Anjum, M. R. Khan, A. Din, M. Saeed, I. Pasha, and M. U. Arshad, "Wheat gluten: high molecular weight glutenin subunits?structure, genetics, and relation to dough elasticity," *Journal of Food Science*, vol. 72, no. 3, pp. R56–R63, 2007.

[126] F. Altpeter, V. Vasil, V. Srivastava, and I. K. Vasil, "Integration and expression of the high-molecular-weight glutenin subunit 1Ax1 gene into wheat," *Nature Biotechnology*, vol. 14, no. 9, pp. 1155–1159, 1996.

[127] A. E. Blechl and O. D. Anderson, "Expression of a novel high-molecular-weight glutenin subunit gene in transgenic wheat," *Nature Biotechnology*, vol. 14, no. 7, pp. 875–879, 1996.

[128] F. Barro, L. Rooke, F. Békés et al., "Transformation of wheat with high molecular weight subunit genes results in improved functional properties," *Nature Biotechnology*, vol. 15, no. 12, pp. 1295–1299, 1997.

[129] L. Rooke, F. Békés, R. Fido et al., "Overexpression of a gluten protein in transgenic wheat results in greatly increased dough strength," *Journal of Cereal Science*, vol. 30, no. 2, pp. 115–120, 1999.

[130] M. L. Alvarez, M. Gómez, J. María Carrillo, and R. H. Vallejos, "Analysis of dough functionality of flours from transgenic wheat," *Molecular Breeding*, vol. 8, pp. 103–108, 2001.

[131] P. R. Shewry, S. Powers, J. M. Field et al., "Comparative field performance over 3 years and two sites of transgenic wheat lines expressing HMW subunit transgenes," *Theoretical and Applied Genetics*, vol. 113, pp. 128–136, 2006.

[132] Y. Li, Q. Wang, X. Li et al., "Coexpression of the high molecular weight glutenin subunit 1ax1 and puroindoline improves dough mixing properties in durum wheat (triticum turgidum l. ssp. durum)," *PLoS ONE*, vol. 7, no. 11, article e50057, 2012.

[133] S. Li, N. Wang, Y. Wang et al., "Inheritance and expression of copies of transgenes 1Dx5 and 1Ax1 in elite wheat (triticum aestivum l.) varieties transferred from transgenic wheat through conventional crossing," *Acta Biochimica et Biophysica Sinica*, vol. 39, no. 5, pp. 377–383, 2007.

[134] X. Mao, Y. Li, S. Zhao et al., "The interactive effects of transgenically overexpressed 1Ax1 with various HMW-GS combinations on dough quality by introgression of exogenous subunits into an elite chinese wheat variety," *PLoS ONE*, vol. 8, no. 10, article e78451, 2013.

[135] P. I. Payne, "Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality," *Annual Review of Plant Physiology and Plant Molecular Biology*, vol. 38, pp. 141–153, 1987.

[136] K. A. Scherf, P. Koehler, and H. Wieser, "Gluten and wheat sensitivities – an overview," *Journal of Cereal Science*, vol. 67, pp. 2–11, 2016.

[137] A. Juhász, T. Belova, C. G. Florides et al., "Genome mapping of seed-borne allergens and immunoresponsive proteins in wheat," *Science Advances*, vol. 4, no. 8, article eaar8602, 2018.

[138] I. Cakmak, H. Ozkan, H. J. Braun, R. M. Welch, and V. Romheld, "Zinc and iron concentrations in seeds of wild, primitive, and modern wheats," *Food and Nutrition Bulletin*, vol. 21, no. 4, pp. 401–403, 2000.

[139] R. M. Welch and R. D. Graham, "A new paradigm for world agriculture: meeting human needs. productive, sustainable, nutritious," *Field Crops Research*, vol. 60, no. 1-2, pp. 1–10, 1999.

[140] P. B. Holm, K. N. Kristiansen, and H. B. Pedersen, "Transgenic approaches in commonly consumed cereals to improve iron and zinc content and bioavailability," *Journal of Nutrition*, vol. 132, no. 3, pp. 514S–516S, 2002.

[141] S. Borg, H. Brinch-Pedersen, B. Tauris et al., "Wheat ferritins: improving the iron content of the wheat grain," *Journal of Cereal Science*, vol. 56, no. 2, pp. 204–213, 2012.

[142] X. Sui, Y. Zhao, S. Wang et al., "Improvement Fe content of wheat (triticum aestivum) grain by soybean ferritin expression cassette without vector backbone sequence," *Journal of Agricultural Biotechnology*, vol. 20, pp. 766–773, 2012.

[143] J. M. Connorton, E. R. Jones, I. Rodríguez-Ramiro et al., "Wheat vacuolar iron transporter TaVIT2 transports Fe and Mn and is effective for biofortification," *Plant Physiology*, vol. 174, pp. 2434–2444, 2017.

[144] L. Yan, A. Loukoianov, A. Blechl et al., "The wheat VRN2 gene is a flowering repressor down-regulated by vernalization," *Science*, vol. 303, no. 5664, pp. 1640–1644, 2004.

[145] A. Loukoianov, L. Yan, and A. Blechl, "Regulation of VRN-1 vernalization genes in normal and transgenic polyploid wheat," *Plant Physiology*, vol. 138, no. 4, pp. 2364–2373, 2005.

[146] C. Uauy, A. Distelfeld, T. Fahima, A. Blechl, and J. Dubcovsky, "A NAC gene regulating senescence improves grain protein, Zinc, and iron content in wheat," *Science*, vol. 314, no. 5803, pp. 1298–1301, 2006.

[147] Y. Li, G. Song, J. Gao et al., "Enhancement of grain number per spike by RNA interference of cytokinin oxidase 2 gene in bread wheat," *Hereditas*, vol. 155, no. 33, 2018.

[148] X. Y. Zhao, P. Hong, J. Y. Wu et al., "The tae-miR408-mediated control of TaTOC1 genes transcription is required for the regulation of heading time in wheat," *Plant Physiology*, vol. 170, pp. 1578–1594, 2016.

[149] F. Barro, J. C. Iehisa, M. J. Giménez et al., "Targeting of prolamins by RNAi in bread wheat: effectiveness of seven silencing-fragment combinations for obtaining lines devoid of coeliac disease epitopes from highly immunogenic gliadins," *Plant Biotechnology Journal*, vol. 14, no. 3, pp. 986–996, 2016.

[150] J. R. Li, W. Zhao, Q. Z. Li et al., "RNA silencing of waxy gene results in low levels of amylose in the seeds of transgenic wheat (triticum aestivum L.)," *Acta Genetica Sinica*, vol. 32, no. 8, pp. 846–854, 2005.

[151] S. Yue, H. Li, Y. Li et al., "Generation of transgenic wheat lines with altered expression levels of 1Dx5 high-molecular weight glutenin subunit by RNA interference," *Journal of Cereal Science*, vol. 47, no. 2, pp. 153–161, 2008.

[152] S. B. Altenbach and P. V. Allen, "Transformation of the US bread wheat 'Butte 86' and silencing of omega-5 gliadin genes," *GM Crops*, vol. 2, pp. 66–73, 2011.

[153] S. B. Altenbach, C. K. Tanaka, and B. W. Seabourn, "Silencing of omega-5 gliadins in transgenic wheat eliminates a major source of environmental variability and improves dough mixing properties of flour," *BMC Plant Biology*, vol. 14, no. 393, 2014.

[154] J. Gil-Humanes, F. Pistón, M. J. Giménez et al., "The introgression of RNAi silencing of γ-gliadins into commercial lines of bread wheat changes the mixing and technological properties of the dough," *PLoS ONE*, vol. 7, no. 9, article e45937, 2012.

[155] W. Lee, K. E. Hammond-Kosack, and K. Kanyuka, "Barley stripe mosaic virus-mediated tools for investigating gene function in

cereal plants and their pathogens: virus-induced gene silencing, host-mediated gene silencing, and virus-mediated overexpression of heterologous protein," *Plant Physiology*, vol. 160, no. 2, pp. 582–590, 2012.

[156] D. Nowara, A. Gay, C. Lacomme et al., "HIGS: host-induced gene silencing in the obligate biotrophic fungal pathogen Blumeria graminis," *The Plant Cell*, vol. 22, no. 9, pp. 3130–3141, 2012.

[157] W. Cheng, X. Song, H. Li et al., "Host-induced gene silencing of an essential chitin synthase gene confers durable resistance to fusarium head blight and seedling blight in wheat," *Plant Biotechnology Journal*, vol. 13, no. 9, pp. 1335–1345, 2015.

[158] Y. Chen, Q. Gao, M. Huang et al., "Characterization of RNA silencing components in the plant pathogenic fungus fusarium graminearum," *Scientific Reports*, vol. 5, no. 12500, 2015.

[159] X. Song, K. Gu, X. Duan et al., "A myosin5 dsRNA that reduces the fungicide resistance and pathogenicity of fusarium asiaticum," *Pesticide Biochemistry and Physiology*, vol. 150, pp. 1–9, 2018.

[160] X. Zhu, T. Qi, Q. Yang et al., "Host-induced gene silencing of the MAPKK gene PsFUZ7 confers stable resistance to wheat stripe rust," *Plant Physiology*, vol. 175, no. 4, pp. 1853–1863, 2017.

[161] T. Qi, X. Zhu, C. Tan et al., "Host-induced gene silencing of an important pathogenicity factor PsCPK1 in Puccinia striiformis f. sp. tritici enhances resistance of wheat to stripe rust," *Plant Biotechnology Journal*, vol. 16, no. 3, pp. 797–807, 2018.

[162] V. Panwar, M. Jordan, B. McCallum, and G. Bakkeren, "Host-induced silencing of essential genes in Puccinia triticina through transgenic expression of RNAi sequences reduces severity of leaf rust infection in wheat," *Plant Biotechnology Journal*, vol. 16, no. 5, pp. 1013–1023, 2018.

[163] Y. Zhao, X. Sui, L. Xu et al., "Plant-mediated RNAi of grain aphid CHS1 gene confers common wheat resistance against aphids," *Pest Management Science*, vol. 74, no. 12, pp. 2754–2760, 2018.

[164] L. Xu, Q. Hou, Y. Zhao et al., "Silencing of a lipase maturation factor 2-like gene by wheat-mediated RNAi reduces the survivability and reproductive capacity of the grain aphid, sitobion avenae," *Archives of Insect Biochemistry and Physiology*, vol. 95, no. 3, article e21392, 2017.

[165] L. Xu, X. Duan, Y. Lv et al., "Silencing of an aphid carboxylesterase gene by use of plant-mediated RNAi impairs Sitobion avenae tolerance of Phoxim insecticides," *Transgenic Research*, vol. 23, no. 2, pp. 389–396, 2014.

[166] M. Fu, M. Xu, T. Zhou et al., "Transgenic expression of a functional fragment of harpin protein Hpa1 in wheat induces the phloem-based defence against English grain aphid," *Journal of Experimental Botany*, vol. 65, no. 6, pp. 1439–1453, 2014.

[167] E. Abdellatef, T. Will, A. Koch, J. Imani, A. Vilcinskas, and K. Kogel, "Silencing the expression of the salivary sheath protein causes transgenerational feeding suppression in the aphid Sitobion avenae," *Plant Biotechnology Journal*, vol. 13, no. 6, pp. 849–857, 2015.

[168] H. Puchta and F. Fauser, "Gene targeting in plants: 25 years later," *The International Journal of Developmental Biology*, vol. 57, no. 6-7-8, pp. 629–637, 2013.

[169] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity," *Science*, vol. 337, no. 6096, pp. 816–821, 2012.

[170] P. Mali, K. M. Esvelt, and G. M. Church, "Cas9 as a versatile tool for engineering biology," *Nature Methods*, vol. 10, no. 10, pp. 957–963, 2013.

[171] Y. Demirci, B. Zhang, and T. Unver, "CRISPR/Cas9: an RNA-guided highly precise synthetic tool for plant genome editing," *Journal of Cellular Physiology*, vol. 233, no. 3, pp. 1844–1859, 2018.

[172] Q. Shan, Y. Wang, J. Li et al., "Targeted genome modification of crop plants using a CRISPR-Cas system," *Nature Biotechnology*, vol. 31, pp. 686–688, 2013.

[173] Q. Shan, Y. Wang, J. Li, and C. Gao, "Genome editing in rice and wheat using the CRISPR/Cas system," *Nature Protocols*, vol. 9, no. 10, pp. 2395–2410, 2014.

[174] S. K. Upadhyay, J. Kumar, A. Alok, and R. Tuli, "RNA-guided genome editing for target gene mutations in wheat," *G3: Genes, Genomes, Genetics*, vol. 3, no. 12, pp. 2233–2238, 2013.

[175] T. Čermák, S. J. Curtin, J. Gil-Humanes et al., "A multi-purpose toolkit to enable advanced genome engineering in plants," *The Plant Cell*, vol. 29, pp. 1196–1217, 2017.

[176] Z. Liang, K. Chen, T. Li et al., "Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes," *Nature Communications*, vol. 8, p. 14261, 2017.

[177] J. Gil-Humanes, Y. Wang, Z. Liang et al., "High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9," *The Plant Journal*, vol. 89, no. 6, pp. 1251–1262, 2017.

[178] Y. Zhang, Y. Bai, G. Wu et al., "Simultaneous modification of three homoeologs of TaEDR1 by genome editing enhances powdery mildew resistance in wheat," *The Plant Journal*, vol. 91, no. 4, pp. 714–724, 2017.

[179] P. Bhowmik, E. Ellison, B. Polley et al., "Targeted mutagenesis in wheat microspores using CRISPR/Cas9," *Scientific Reports*, vol. 8, no. 6502, 2018.

[180] Y. Wang, X. Cheng, Q. Shan et al., "Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew," *Nature Biotechnology*, vol. 32, no. 9, pp. 947–951, 2014.

[181] Y. Zong, Y. Wang, C. Li et al., "Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion," *Nature Biotechnology*, vol. 35, no. 5, pp. 438–440, 2017.

[182] Y. Zong, Q. Song, C. Li et al., "Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A," *Nature Biotechnology*, vol. 36, pp. 950–953, 2018.

[183] W. Wang, Q. Pan, F. He et al., "Transgenerational CRISPR-Cas9 activity facilitates multiplex gene editing in allopolyploid wheat," *The CRISPR Journal*, vol. 1, pp. 65–74, 2018.

[184] H. Hamada, Y. Liu, Y. Nagira, R. Miki, N. Taoka, and R. Imai, "Biolistic-delivery-based transient CRISPR/Cas9 expression enables in planta genome editing in wheat," *Scientific Reports*, vol. 8, no. 14422, 2018.

[185] M. Singh, M. Kumar, M. C. Albertsen, J. K. Young, and A. M. Cigan, "Concurrent modifications in the three homeologs of Ms45 gene with CRISPR-Cas9 lead to rapid generation of male sterile bread wheat (Triticum aestivum L.)," *Plant Molecular Biology*, vol. 97, no. 4-5, pp. 371–383, 2018.

[186] R. M. Howells, M. Craze, S. Bowden, and E. J. Wallington, "Efficient generation of stable, heritable gene edits in wheat using CRISPR/Cas9," *BMC Plant Biology*, vol. 18, no. 215, 2018.

[187] J. Murovec, Ž. Pirc, and B. Yang, "New variants of CRISPR RNA-guided genome editing enzymes," *Plant Biotechnology Journal*, vol. 15, no. 8, pp. 917–926, 2017.

[188] O. O. Abudayyeh, J. S. Gootenberg, P. Essletzbichler et al., "RNA targeting with CRISPR–Cas13," *Nature*, vol. 550, no. 7675, pp. 280–284, 2017.

[189] C. M. Lee, T. J. Cradick, and G. Bao, "The Neisseria meningitides CRISPR-Cas9 system enables specific genome editing in mammalian cells," *Molecular Therapy*, vol. 24, no. 3, pp. 645–654, 2016.

[190] F. A. Ran, L. Cong, W. X. Yan et al., "In vivo genome editing using Staphylococcus aureus Cas9," *Nature*, vol. 520, no. 7546, pp. 186–191, 2015.

[191] E. Kim, T. Koo, S. W. Park et al., "In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni," *Nature Communications*, vol. 8, article 14500, 2017.

[192] C. Li, Y. Zong, Y. Wang et al., "Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion," *Genome Biology*, vol. 19, no. 59, 2018.

[193] T. Kelliher, D. Starr, L. Richbourg et al., "MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction," *Nature*, vol. 542, no. 7639, pp. 105–109, 2017.

[194] I. Fonfara, H. Richter, M. Bratovič, A. Le Rhun, and E. Charpentier, "The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA," *Nature*, vol. 532, no. 7600, pp. 517–521, 2016.

[195] R. Xu, R. Qin, H. Li et al., "Generation of targeted mutant rice using a CRISPR-Cpf1 system," *Plant Biotechnology Journal*, vol. 15, no. 6, pp. 713–717, 2017.

[196] X. Tang, L. G. Lowder, T. Zhang et al., "A CRISPR–Cpf1 system for efficient genome editing and transcriptional repression in plants," *Nature Plants*, vol. 3, article 17018, 2017.

[197] V. J. Nalam, S. Alam, J. Keereetaweep et al., "Facilitation of Fusarium graminearum infection by 9-Lipoxygenases in Arabidopsis and wheat," *Molecular Plant-Microbe Interactions*, vol. 28, no. 10, pp. 1142–1152, 2015.

[198] Y. Zhang, Z. Liang, Y. Zong et al., "Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA," *Nature Communications*, vol. 7, Article ID 12617, 2016.

[199] Z. Liang, K. Chen, Y. Yan, Y. Zhang, and C. Gao, "Genotyping genome-edited mutations in plants using CRISPR ribonucleoprotein complexes," *Plant Biotechnology Journal*, vol. 16, no. 12, pp. 2053–2062, 2018.

[200] S. Lu, H. Zhao, D. L. Des Marais et al., "Arabidopsis ECERIFERUM9 involvement in cuticle formation and maintenance of plant water status," *Plant Physiology*, vol. 159, no. 3, pp. 930–944, 2012.

[201] W. Wang, J. Simmonds, Q. Pan et al., "Gene editing and mutagenesis reveal inter-cultivar differences and additivity in the contribution of TaGW2 homoeologues to grain size and weight in wheat," *Theoretical and Applied Genetics*, vol. 131, no. 11, pp. 2463–2475, 2018.

[202] Q. Li, L. Li, Y. Liu et al., "Influence of TaGW2-6A on seed development in wheat by negatively regulating gibberellin synthesis," *Journal of Plant Sciences*, vol. 263, pp. 226–235, 2017.

[203] H. Xu, M. Zhao, Q. Zhang, Z. Xu, and Q. Xu, "The dense and erect panicle 1 (DEP1) gene offering the potential in the breeding of high-yielding rice," *Breeding Science*, vol. 66, no. 5, pp. 659–667, 2016.

[204] S. Sánchez-León, J. Gil-Humanes, C. V. Ozuna et al., "Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9," *Plant Biotechnology Journal*, vol. 16, no. 4, pp. 902–910, 2018.

[205] "Wheat: vital grain of civilization and food security," CGIAR Research Program Wheat Annual Report, CGIAR, Mexico, D.F., 2013.

Research Article

Aluminum Responsive Genes in Flax (*Linum usitatissimum* L.)

George S. Krasnov [iD],[1] **Alexey A. Dmitriev** [iD],[1] **Alexander V. Zyablitsin,**[1]
Tatiana A. Rozhmina,[1,2] **Alexander A. Zhuchenko,**[2,3] **Parfait Kezimana,**[1,4]
Anastasiya V. Snezhkina [iD],[1] **Maria S. Fedorova** [iD],[1] **Roman O. Novakovskiy,**[1]
Elena N. Pushkova,[1] **Liubov V. Povkhova,**[1,5] **Nadezhda L. Bolsheva** [iD],[1]
Anna V. Kudryavtseva [iD],[1] **and Nataliya V. Melnikova** [iD][1]

[1]*Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow 119991, Russia*
[2]*Federal Research Center for Bast Fiber Crops, Torzhok 172002, Russia*
[3]*All-Russian Horticultural Institute for Breeding, Agrotechnology and Nursery, Moscow 115598, Russia*
[4]*Peoples' Friendship University of Russia (RUDN University), Moscow 117198, Russia*
[5]*Moscow Institute of Physics and Technology, Dolgoprudny 141701, Russia*

Correspondence should be addressed to Nataliya V. Melnikova; mnv-4529264@yandex.ru

Flax (*Linum usitatissimum* L.) is a multipurpose crop which is used for the production of textile, oils, composite materials, pharmaceuticals, *etc*. Soil acidity results in a loss of seed and fiber production of flax, and aluminum toxicity is a major factor that depresses plant growth and development in acid conditions. In the present work, we evaluated gene expression alterations in four flax genotypes with diverse tolerance to aluminum exposure. Using RNA-Seq approach, we revealed genes that are differentially expressed under aluminum stress in resistant (Hermes, TMP1919) and sensitive (Lira, Orshanskiy) cultivars and selectively confirmed the identified alterations using qPCR. To search for differences in response to aluminum between resistant and sensitive genotypes, we developed the scoring that allowed us to suggest the involvement of MADS-box and NAC transcription factors regulating plant growth and development and enzymes participating in cell wall modifications in aluminum tolerance in flax. Using Gene Ontology (GO) enrichment analysis, we revealed that glutathione metabolism, oxidoreductase, and transmembrane transporter activities are the most affected by the studied stress in flax. Thus, we identified genes that are involved in aluminum response in resistant and sensitive genotypes and suggested genes that contribute to flax tolerance to the aluminum stress.

1. Introduction

Among abiotic stresses, aluminum (Al) toxicity is a major constraint for crop production in acid soils worldwide [1]. In acidic conditions, the mineral form of Al dissolves to release the soluble Al^{3+} form, which is capable of crossing the plant membranes and is highly toxic to plants that even micro concentrations can inhibit root growth within minutes or hours in many agricultural plant species [2–5]. Al negatively affects cell elongation and division, uptake and transport of nutrients, and Ca^{2+} homeostasis and disturbs the structure and function of the plasma membrane, cell wall, and chromatin [6–10].

The mechanisms of resistance to Al are diverse in plants and could be divided into exclusion, which decreases the amount of phytotoxic Al^{3+} in the cells and internal tolerance, which reduces Al toxicity in root and shoot symplast [6, 11–19]. Excretion of Al detoxifying organic acids (OAs) to the apoplast or rhizosphere is the most common variant of exclusion, in which genes encoding the OA transporters, including aluminum-activated malate transporter 1 (ALMT1) and members of the multidrug and toxic compound extrusion (MATE) citrate transporter gene family, are involved [13, 20]. Exudation of mucilage also binds Al ions and results in the exclusion of Al and its detoxification [21–23]. Mechanism of Al tolerance encompasses processes that result in chelation

of cytosolic Al^{3+} with organic ligands, sequestration into the vacuole, or transport of Al^{3+} to less-sensitive regions of the plant [24–26]. Recovery from damages following exposure to Al toxicity is mediated by the detoxification of the reactive oxygen species (ROS) through ROS-detoxifying enzymes, such as glutathione S-transferases, peroxidases, and superoxide dismutases that confer Al tolerance [20, 27, 28]. Significant genetic diversity in Al resistance or tolerance was found in crops that enable development of improved cultivars, which maintain yield on acid soil [29, 30].

For flax (*Linum usitatissimum* L.), a crop grown worldwide for fiber and seeds and also known for its health benefits [31, 32], the mechanisms for Al tolerance are little known, although Al toxicity in acid soils is a serious problem for cultivation and rich harvest of flax [33]. Therefore, it is imperative to understand the molecular mechanisms underlying Al tolerance in flax. High-throughput sequencing methods are extensively used for evaluation of expression alterations of protein-coding genes or miRNAs in flax under diverse stresses [34–41]. We have previously shown that UDP-glycosyltransferases, glutathione S-transferases, and Ca^{2+}/H^{+}-exchanger CAX3 are involved in flax tolerance to aluminum [42, 43]. miR319, miR390, and miR393 also participate in aluminum response *via* regulation of growth processes [44]. However, Al tolerance mechanisms are complex and could involve diverse strategies. In the present work, we conducted comprehensive gene expression analysis of flax cultivars grown under control and Al-treated conditions to reveal genes that participate in Al response and to identify differences in gene expression alterations between resistant and sensitive to aluminum genotypes.

2. Materials and Methods

2.1. Plant Material. A laboratory method for evaluation of flax tolerance to aluminum, whose results have high correlation with field assessment, was taken as the basis for creating aluminum treatment conditions [45]. Resistant and sensitive cultivars for the study were chosen on the basis of our field and laboratory experiments, in which productivity indexes and root length were used for assessment of flax genotype tolerance to soil acidity and aluminum [33]. For the present study of gene expression alterations under short-term (4 h) aluminum exposure (500 μM $AlCl_3$) at low pH (4.5), resistant (Hermes and TMP1919) and sensitive (Lira and Orshanskiy) to aluminum stress flax cultivars were used. Flax seeds were germinated in Petri dishes on filter paper and then were transferred into 50 mL Falcon tubes with filter paper soaked in 0.5 mM $CaCl_2$ pH 4.5 for 24 h. Seedlings in control conditions were then grown for 24 h more, while seedlings in stress conditions were grown for 20 h followed by addition of 500 μM $AlCl_3$ for 4 h. After that, root tips, 8-10 mm in length, were collected and immediately frozen in liquid nitrogen. The samples were stored at -70°C. RNA was extracted using RNA MiniPrep kit (Zymo Research, USA) and then used for transcriptome sequencing on HiSeq2500 (Illumina, USA) with paired-end 100-nucleotide reads in two biological replicates as described in our previous work [42].

2.2. Identification of Differentially Expressed Genes. Transcriptome assembly and annotation were performed using Trinity and Trinotate as described earlier [42]. The reads of each cultivar (Hermes, TMP1919, Lira, and Orshanskiy) were mapped to the assembled transcripts of Hermes using bowtie2 [46] and quantified using RSEM [47]. Then, read count data were analyzed using edgeR package [48]. Genes with read counts per million (CPM) below 1.5 were filtered out. Normalization using the TMM method was performed and genes with expression alterations under aluminum treatment were identified within the following groups:

(1) pool of all analyzed flax genotypes;

(2) pool of resistant cultivars (Hermes and TMP1919);

(3) pool of sensitive cultivars (Lira and Orshanskiy);

(4) individual genotypes (Hermes, TMP1919, Lira, or Orshanskiy).

To evaluate expression alterations for each transcript, expression level fold change (*FC*) was calculated for each gene as the ratio of CPM under Al^{3+} exposure to CPM under control conditions.

To identify genes with diverse expression alterations in resistant and sensitive to aluminum genotypes, we calculated delta-score (S_Δ) that takes into account (1) consistency of gene expression changes within each group (resistant or sensitive cultivars), (2) magnitude of these changes, and (3) differences between groups. To proceed this way, we derived consistency scores for resistant and sensitive cultivars, which correspond to the first component:

$$C_{\text{res.}} = \frac{\left(\sum_{\text{res.}} \sqrt[3]{\log FC_i} \right)^3}{\max \left(|\log FC_i| \right)} \tag{1}$$

where $\log FC$ is the binary logarithm of *FC*, and i = 1, 2 (number of resistant genotypes). Hence, if aluminum-induced gene expression changes are unidirectional and *FC* values are close to each other, $C_{\text{res.}}$ would have the greatest value. In the similar way, we calculated the consistency score for sensitive cultivars ($C_{\text{sens.}}$). Finally, delta-scores (S_Δ) were calculated as

$$S_\Delta = (C_{\text{res.}} \times C_{\text{sens.}}) \times \left(\max_{\text{res.,sens.}} \left(\text{mean} \left(|\log FC_i| \right) \right) \right) \\ \times \left(\underset{\text{res.}}{\text{mean}} \left(\log FC_i \right) - \underset{\text{sens.}}{\text{mean}} \left(\log FC_i \right) \right)^2 \tag{2}$$

Here, the 1st multiplier reflects the consistency of gene expression changes, the 2nd one—their maximum magnitude, and the 3rd one—differences between groups. We introduced 2-fold greater weight for the 3rd component because of its prime importance.

2.3. Quantitative PCR (qPCR) Analysis. For validation of high-throughput sequencing gene expression data, qPCR was used. The same RNA samples of Hermes, TMP1919, Lira, and Orshanskiy that were used for high-throughput sequencing were analyzed. RevertAid Reverse Transcriptase (Thermo

Fisher Scientific, USA) was used to generate first-strand cDNA. Transcripts for qPCR analysis were chosen on the basis of transcriptome sequencing data. The selected transcripts encode the following proteins: LHY (TR32133|c1_g1), transcription factor (TF) LUX (TR44570|c0_g2), putative lysine-specific demethylase JMJ30 (TR19190|c0_g1), high mobility group B protein 1 (TR12175|c1_g1), zinc finger protein CONSTANS-like 10 (TR53996|c0_g1), high-affinity nitrate transporter 3.1 (TR26619|c0_g1), and two-component response regulator-like APRR1 (TR32104|c0_g2). Primers and probes were designed using ProbeFinder Software (Roche, Switzerland) (Table 1). qPCR was performed on 7500 Real-Time PCR System (Applied Biosystems, USA). Reaction mix contained 1x PCR buffer (GenLab, Russia), 250 nM of dNTPs (Fermentas, Lithuania), 300 nM of each primer, 200 nM of short hydrolysis probes from Universal ProbeLibrary (Roche), 1 U of Taq polymerase (GenLab), and cDNA. The following qPCR program was used: 95°C for 10 min, 50 cycles of 95°C for 15 s, and 60°C for 60 s. *ETIF3E* was used as a reference gene [35, 49]. All reactions were performed in three technical replicates. All the calculations were carried out using our ATG tool [35, 50]. $\Delta\Delta C_t$ values were calculated for the assessment of gene expression level [35, 37]. Correlation between qPCR ($\Delta\Delta C_t$) and high-throughput sequencing (log *FC*) data was evaluated by Spearman's correlation coefficient.

2.4. Gene Ontology Analysis. The gene set enrichment analysis (GSEA) based on Gene Ontology (GO) data was performed using Goseq package (http://bioconductor.org/packages/release/bioc/html/goseq.html). Analysis of top lists of up- and downregulated genes was carried out for the pool of all analyzed genotypes, the pool of resistant cultivars, the pool of sensitive cultivars, and individual genotypes for identification of enriched GO terms. Heatmaps illustrating gene expression alterations in selected GO terms were created using R package pheatmap. Weighted gene correlation network analysis was carried out using R package WGCNA [51] for genes with CPM > 20 in at least 6 samples.

3. Results and Discussion

3.1. The Effects of Aluminum Exposure at Low pH on Flax Plants. Soil acidity and Al^{3+} ions result in depression of flax plants, inhibition of growth, changes in height, fiber mass, and seed productivity both in resistant and sensitive cultivars. However, the degree of changes varied between genotypes [33]. For the present study, we used flax cultivars with contrast extent of phenotypic changes in acid soils with high aluminum content (pH 4.0, Al content – 11.07 mg/100 g of soil). Results of field experiments for Hermes, TMP1919, Lira, and Orshanskiy cultivars under control conditions and low pH with Al treatment are presented in Table 2. Cultivars Hermes and TMP1919 showed greater tolerance compared to Lira and Orshanskiy that was especially noticeable for fiber mass and a number of seed pods.

In laboratory experiments, inhibition of root growth was observed in flax plants under aluminum exposure at low

pH (pH 4.5, 500 μM $AlCl_3$). Negative effects were more pronounced in sensitive cultivars, where 50-60% reduction in root length was revealed after 5 days, while in resistant cultivars reduction was less than 30% [45].

3.2. Genes with Differential Expression under Al^{3+} Exposure. As shown in our previous works, even after 4 h of Al^{3+} exposure when phenotype changes were not yet noticeable, significant gene expression alterations occurred in flax plants [42, 44]. These results are consistent with the studies on other plant species, for many of which short response time to aluminum was revealed [19]. Therefore, in the present work, we focused on the effects of short-term aluminum treatment and evaluated gene expression alterations after 4 h of aluminum exposure in resistant and sensitive to the studied stress flax genotypes based on transcriptome sequencing data. In this regard, in contrast to our previous study [42], we compared gene expression levels in flax plants under control conditions and after 4 h of Al^{3+} exposure in the pool of all studied genotypes, groups of resistant and sensitive plants, or individual cultivars. Results of our analysis are presented in Supplementary Materials for the pool of all studied genotypes (S1 table), pool of resistant (Hermes and TMP1919, S2 Table) and pool of sensitive (Lira and Orshanskiy, S3 Table) genotypes, and individual cultivars (S4, S5, S6, and S7 Tables for Hermes, TMP1919, Lira, and Orshanskiy respectively), where the genes are listed in the order of decreasing statistical significance of expression alterations.

To validate high-throughput sequencing data, we evaluated mRNA level of transcripts with significant expression alterations under the studied stress in flax cultivars grown under control conditions and Al^{3+} exposure using qPCR. Expression of genes encoding protein LHY (TR32133|c1_g1), TF LUX (TR44570|c0_g2), putative lysine-specific demethylase JMJ30 (TR19190|c0_g1), high mobility group B protein 1 (TR12175|c1_g1), zinc finger protein CONSTANS-like 10 (TR53996|c0_g1), high-affinity nitrate transporter 3.1 (TR26619|c0_g1), and two-component response regulator-like APRR1 (TR32104|c0_g2) was analyzed. Reaction efficiencies were 95% or higher, and Ct (cycle threshold) values varied from 23 to 32. As seen from Figure 1, the expression data obtained by high-throughput sequencing (Figure 1(a)) and qPCR (Figure 1(b)) were highly consistent that indicates the reliability of the sequencing data. Spearman's correlation coefficient was 0.89 ($p < 0.01$).

We performed the search for differences in transcriptomic response to aluminum stress between resistant and sensitive genotypes to identify genes potentially involved in aluminum tolerance. We developed the scoring that takes into account gene expression changes within resistant and sensitive groups of genotypes, their consistency, and differences between groups. This allowed us to find out transcripts with diverse expression alterations between groups of resistant and sensitive to aluminum flax genotypes but similar alterations within resistant or sensitive groups. Results of the analysis are presented in S8 Table. Transcripts were sorted by decreasing delta-score that reflects differences in Al^{3+}-induced transcriptomic response between resistant and

TABLE 1: Primers and probes for qPCR.

Primer name	Primer sequence	Probe number from Roche Universal ProbeLibrary
LHY-F	CAGGAATCGAAGTTGGGAGA	12
LHY-R	CGCTGCTTCAAATCCTCTCTAA	
LUX-F	GGGGAGTGGATGCAAAGAG	69
LUX-R	CGACTTTACCTCAAGAGGATGC	
JMJ-F	GAACCATCTTTGCCCTGAATC	48
JMJ-R	AGGAGTGGAAGCAAGAGCTG	
HMGB-F	GCTTTCTCTGCACTAGACAAAGATT	7
HMGB-R	AACAAAGCTGTCTCCGCTGT	
COL-F	CGTATGAATTCAATGCAGCAG	66
COL-R	CAAACAGCTGGTTCGGTTTTA	
NRT-F	ATTTTGAACTGGCAGGTCTTG	69
NRT-R	CGGTGAGCCAGAAGGACA	
APRR-F	CGCTGAAGTTGATCTTCCAAT	89
APRR-R	GCAAATCCTTATGCCGAGTT	
ETIF3E-F	TTACTGTCGCATCCATCAGC∗	53
ETIF3E-R	GGAGTTGCGGATGAGGTTTA∗	

Note. Primers were designed for genes encoding protein LHY (LHY, TR32133|c1_g1), TF LUX (LUX, TR44570|c0_g2), putative lysine-specific demethylase JMJ30 (JMJ, TR19190|c0_g1), high mobility group B protein 1 (HMGB, TR12175|c1_g1), zinc finger protein CONSTANS-LIKE 10 (COL, TR53996|c0_g1), high-affinity nitrate transporter 3.1 (NRT, TR26619|c0_g1), and two-component response regulator-like APRR1 (APRR, TR32104|c0_g2). ∗: Primer sequences for *ETIF3E* are from Huis et al. article [47].

TABLE 2: Phenotype changes in flax plants under aluminum treatment at low pH.

| Cultivar	Condition	Plant height, cm	Decline, %	Fiber mass, mg	Decline, %	Number of seed pods	Decline, %
Hermes	Stress	60.6±0.9	85.9	115.6±5.6	85.3	2.9±0.4	96.6
	Control	70.5±0.9		135.6±8.2		2.8±0.4	
TMP1919	Stress	69.4±1.6	94.2	93.7±3.8	86.0	3.5±0.3	81.4
	Control	73.7±1.7		108.9±5.4		4.3±0.5	
Lira	Stress	56.2±1.2	75.3	54.4±5.8	41.3	2.5±0.3	56.8
	Control	74.6±1.5		131.8±7.4		4.4±0.3	
Orshanskiy	Stress	61.2±1.4	86.2	52.8±4.4	63.1	3.2±0.3	59.3
	Control	71.0±0.6		83.7±8.3		5.4±0.4	

Note. Stress: aluminum exposure at pH 4.5.
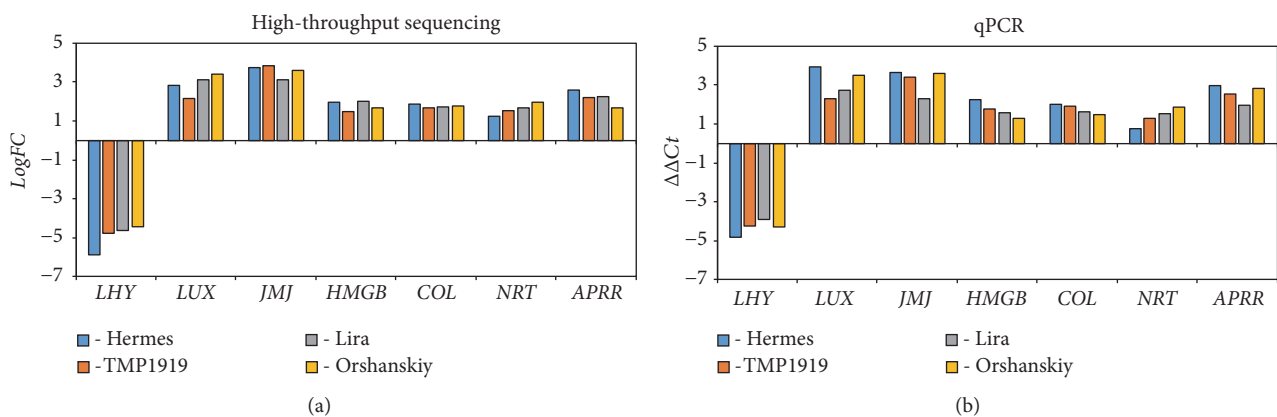
FIGURE 1: Expression alterations of 7 genes in flax cultivars Hermes, TMP1919, Lira, and Orshanskiy under aluminum exposure at low pH evaluated by high-throughput sequencing (a) and qPCR (b).

sensitive groups of genotypes. Within upregulated transcripts in resistant genotypes compared to sensitive ones, transcripts encoding Agamous-like MADS-box protein AGL62, polygalacturonase, NAC domain-containing protein 100, protein OSB1, GDSL esterase/lipase, beta-glucosidase 11, (+)-neomenthol dehydrogenase, osmotin-like protein OSML13, and 1-aminocyclopropane-1-carboxylate oxidase homolog 1 were in the top.

First of all, our attention was attracted by genes that were upregulated only in resistant cultivars under Al^{3+} exposure. In the top of S8 Table, two TFs are located: Agamous-like MADS-box protein AGL62 and NAC domain-containing protein 100. MADS-box genes control numerous aspects of plant development and are involved in stress response [52–55]. NAC TFs also participate in stress response and their overexpression in transgenic plants enhances tolerance to abiotic stresses and promotes development of lateral roots [56–61]. Role of NAC in response to aluminum was revealed in maize and rice; these TFs were involved in phytohormone signaling and growth regulation [62–64]. In our study, expression of transcript TR35721|c0_g1, which encodes Agamous-like MADS-box protein AGL62, was induced by Al only in resistant flax genotypes (the binary logarithm of fold change, $\log FC$, was equal to 2.3 and 3.0 in Hermes and TMP1919, respectively). The same was true for transcript TR25300|c0_g1 encoding NAC domain-containing protein 100: significant upregulation was revealed under aluminum stress in resistant cultivars ($\log FC$ was equal to 1.0 and 1.1 in Hermes and TMP1919, respectively), while some decrease of the expression—in sensitive ones. Thus, overexpression of genes encoding NAC domain-containing protein 100 and Agamous-like MADS-box protein AGL62 only in resistant flax genotypes under Al^{3+} exposure indicates that these TFs could contribute to flax tolerance to aluminum. We also analyzed mRNA level alterations of other TFs, whose role in response to Al is known. The role of WRKY TFs in Al tolerance in plants was shown to be ambiguous: in *Arabidopsis*, WRKY46 was identified as a negative regulator of ALMT1 and mutations in WRKY46 lead to increased malate secretion and higher Al resistance [65]; on the contrary, WRKY22 promoted aluminum tolerance in rice through activation of FRDL4 (Ferric reductase defective 4) and increased citrate secretion [66]. In flax, we revealed slight upregulation of a number of WRKY-encoding genes but did not observe differences between resistant and sensitive cultivars. For genes encoding sensitive to proton rhizotoxicity 1 (STOP1) and Abscisic acid stress ripening (ASR), which are also known TFs associated with aluminum response [17, 20, 67–69], no significant expression alterations were revealed by us in flax under aluminum stress. Thus, it can be suggested that Agamous-like MADS-box protein AGL62 and NAC domain-containing protein 100 are the leading TFs involved in Al tolerance in flax *via* regulation of plant growth and development.

Besides TFs, we observed more pronounced upregulation of genes encoding polygalacturonase, cellulose synthase, pectinesterase, beta-glucosidase, and GDSL esterase in resistant to aluminum flax cultivars compared to sensitive ones.

These enzymes are involved in cell wall metabolism, which plays important role in plant response to heavy metals and other abiotic stresses [70]. Modification of cell wall and changes in binding properties of the apoplast are known to contribute to Al tolerance in plants [70, 71]. In flax, upregulation of genes encoding cell wall-related proteins was observed in resistant genotypes; therefore, cell wall modification could be one of the mechanisms of flax tolerance to Al ions.

In the top of the list of transcripts that were differentially expressed between resistant and sensitive genotypes (S8 Table), we also observed peroxidase and ABC transporter-encoding genes, whose role in plant tolerance to aluminum stress is known [20]. Peroxidases are involved in detoxification of ROS and their overexpression in some plant species is associated with Al tolerance [20]. Moreover, overexpression of *Arabidopsis* peroxidase in tobacco plants improved their tolerance to aluminum [72]. At the same time, peroxidase expression was decreased under aluminum treatment in *Camellia sinensis* [73]. In wheat, peroxidase expression was induced by Al stress; however, their activity was lower in resistant genotypes compared to sensitive ones [74, 75], while in chickpea, peroxidase activity was almost similar in tolerant and sensitive genotypes [76]. In our study on flax, for TR33816|c0_g1 transcript encoding peroxidase 5, a significant expression decrease was revealed in resistant genotypes ($\log FC$ was equal to -1.7 for both Hermes and TMP1919), while slight decrease or retention was identified in sensitive ones ($\log FC = -0.2$ for Lira and $\log FC = -0.5$ for Orshanskiy). The role of ABC transporters in plant response to aluminum stress is well characterized [8, 77–81]. In flax, ABC transporter-encoding transcript TR4576|c0_g2 had diverse expression changes between resistant and sensitive to aluminum genotypes: strong expression decrease was revealed in resistant cultivars ($\log FC$ was equal to -1.7 and -1.2 for Hermes and TMP1919, respectively) and slight expression decrease was observed in sensitive ones ($\log FC = -0.3$ for Lira and $\log FC = -0.4$ for Orshanskiy). For ABC transporter, we did not find associations of flax tolerance to Al with high expression levels of transcripts both under stress and control conditions. Considering the fact that expression of aluminum tolerance genes is usually higher in resistant genotypes and often increased by Al treatment [8], we suggested that peroxidase- and ABC transporter-encoding genes are not the aluminum tolerance gene in flax.

The scoring for identification of differentially expressed genes in resistant genotypes compared to sensitive ones under Al^{3+} exposure was shown to be the promising tool for revelation flax genes that are involved in tolerance to aluminum. Obtained results allowed us to suggest that some genes (including STOP1-, ASR-, peroxidase-, and ABC transporter-encoding ones), whose role in aluminum response was revealed for several plant species, do not play the key role in flax tolerance to the stress. At the same time, MADS-box and NAC TFs, which regulate plant growth and development, and the enzymes that are involved in cell wall modifications are likely important for flax tolerance to aluminum.
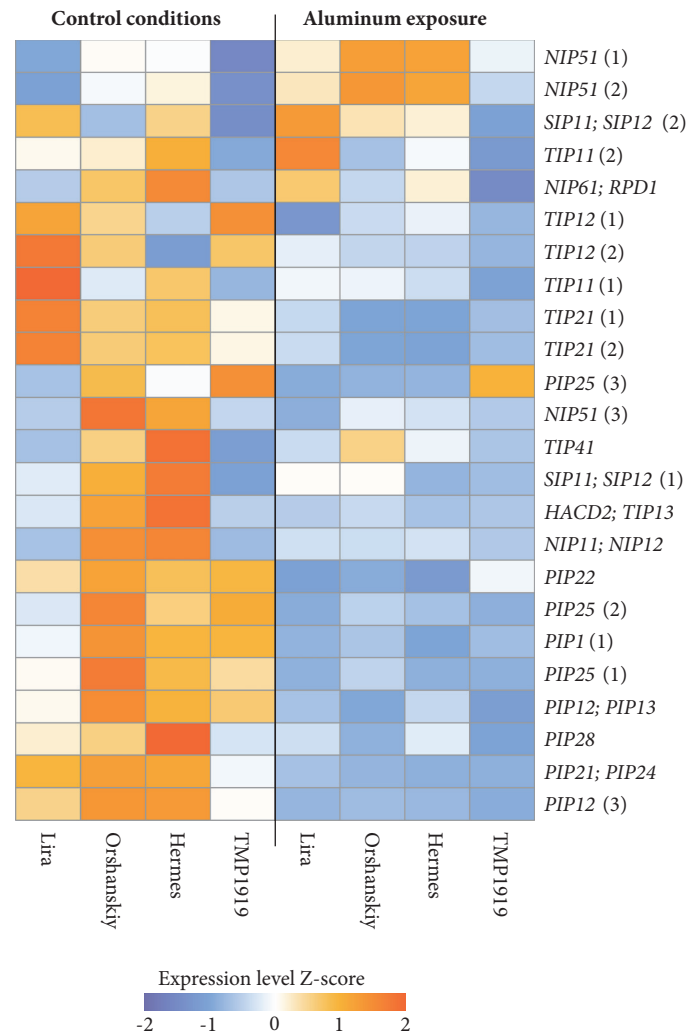
FIGURE 2: **Expression patterns for genes participating in water transmembrane transporter activity (GO 0005372) in flax cultivars Lira, Orshanskiy, Hermes, and TMP1919 under control conditions and aluminum exposure.** The heatmap represents Z-scores of normalized read counts per million (CPM) for each gene: from blue (low expression level) to orange (high expression level). Row names show top BLAST hits of the assembled gene transcripts.

3.3. GO Analysis and Gene Expression Profiles. For deeper understanding of the mechanisms of flax response to aluminum, we performed GO enrichment analysis to identify the processes in which genes with significant expression alterations under Al^{3+} exposure are involved. Overrepresented GO terms were assessed for top lists (tops) of up- and downregulated genes in flax plants under aluminum exposure. Tops of 50, 100, 300, and more than 300 differentially expressed genes (maximum number of genes was limited by the statistical significance of expression alterations corresponding to $p\text{-}value < 0.05$ in S1, S2, and S3 Tables) were used in the analysis. For the pool of all studied cultivars, overrepresented GO terms for the tops of 50 and 100 upregulated genes were related to circadian rhythm, multicellular organismal movement, and potassium: sodium symporter activity, while for the top of 300 upregulated genes, transferase activity had also appeared (S9 Table). For all the tops of downregulated genes for the pool of all studied

genotypes, GO terms related to channel activity and transport were the most abundant. Heatmap for "water transmembrane transporter activity" GO term is presented in Figure 2 as an example. In the pool of sensitive to aluminum flax genotypes, we identified less number of significantly overrepresented GO terms compared to the pool of resistant ones for the tops of upregulated genes (S10 and S11 Tables). These GO terms were associated with apoplast, rhythmic processes, peptidase and oxidoreductase activities. In resistant cultivars, overrepresented GO terms included oxidoreductase activity and rhythmic processes too, but unlike sensitive genotypes, GO terms related to glutathione metabolism, glucosyltransferase activity, and response to stimulus were also overrepresented.

We also performed GO analysis for individual cultivars to evaluate the similarity in response to aluminum of genotypes of the same tolerance group (S12, S13, S14, and S15 Tables). For the top of 50 upregulated genes, glutathione transferase activity term was overrepresented in both resistant cultivars

Lira Orshanskiy Hermes TMP1919

Expression level, *logFC*

-2   -1   0   1   2
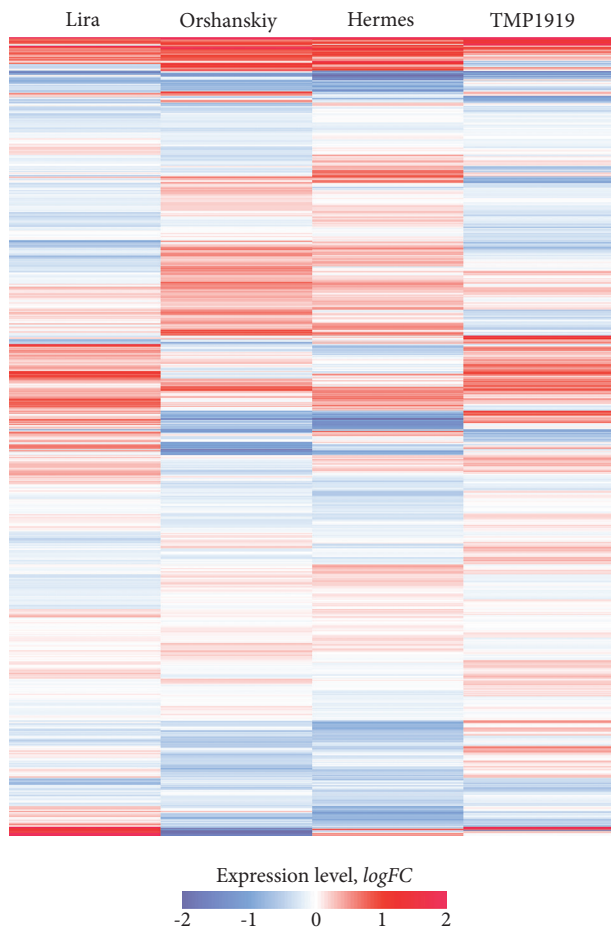
FIGURE 3: **Expression alterations in flax plants under aluminum exposure in Lira, Orshanskiy, Hermes, and TMP1919 cultivars for genes involved in oxidoreductase activity (GO 0016491).** Each heatmap row corresponds to one gene (total 21 000 genes). Color scale represents the binary logarithm of expression level fold change (aluminum exposure/control conditions) from -2 (i.e., 4-fold downregulation, blue) to +2 (4-fold upregulation, red).

(Hermes and TMP1919). However, for TMP1919, unlike Hermes, oxidoreductase activity term was also overrepresented and, in this, it was similar to sensitive cultivar Orshanskiy. When the top of 300 upregulated genes was used for analysis, much more GO terms were overrepresented. Glutathione transferase activity was in the top of the list not only in resistant genotypes but also in sensitive cultivar Orshanskiy, while oxidoreductase activity term was overrepresented in all studied genotypes. Gene expression profiles of particular GO terms had no relation to the degree of tolerance to aluminum. For example, expression alterations of genes related to oxidoreductase activity were more similar for pairs of cultivars TMP1919/Lira and Hermes/Orshanskiy (Figure 3).

GO enrichment analysis allowed us to reveal biological processes that are the most affected by aluminum stress in flax. Summarizing the results of this analysis, it could be concluded that Al^{3+} exposure had negative effects on transmembrane transport in flax. It is not surprising since aluminum disturbs ion transport in plants [19]. It is known that oxidase

overexpression confers aluminum tolerance of plants [82, 83] and genes related to oxidoreductase activity GO term were upregulated in response to aluminum in flax. For glutathione transferase activity GO term, alterations in gene expression were more pronounced in resistant cultivars. Their role in Al response was discussed in our previous study [42].

To reveal coexpression gene networks in flax under aluminum stress, weighted gene coexpression network analysis (WGCNA) was performed for 8986 genes with CPM more than 20 in 6 or more samples. We identified 23 clusters (modules) of coexpressed genes based on similarity in gene expression (S16 Figure). Modules included from 46 to 1831 genes: module 1 was formed by 1831 genes enriched for RNA metabolic process and proton antiporter activity; module 2—organelle membrane and oxidation-reduction process; 3—ribosome and macromolecule biosynthetic process; 6—transporter activity and rhythmic process; 7—regulation of cellular process and macromolecule biosynthesis; 8—phosphorus metabolic process, ion binding, and oxidation-reduction process; 9—nucleolus, ribosome, and RNA metabolic process; 13—mitochondrial part; 14—ion transport; 16—lipid metabolic processes. Genes from the same module are probably coregulated, and the processes, in which they are involved, may be related to each other. For other modules, we did not reveal significant enrichment in genes of particular GO terms.

4. Conclusions

We analyzed transcriptomes of flax cultivars with diverse tolerance to aluminum under control conditions and Al stress and identified genes that were significantly up- or downregulated. A number of genes whose expression alterations differed between resistant and sensitive cultivars were revealed, including those encoding MADS-box and NAC TFs and cell wall biogenesis related enzymes, suggesting them to be involved in aluminum tolerance in flax. These results indicate that, for flax, the internal tolerance to aluminum *via* reduction of Al ions toxicity is inherited. GO enrichment analysis led to the identification of processes that are affected by aluminum in both resistant and sensitive flax genotypes. The genotype effect on the response to Al^{3+} exposure was strong enough; although resistant and sensitive cultivars had diverse gene expression changes under the stress, significant impact of a particular genotype on stress response was also observed. Understanding the mechanisms of flax response to Al and identification of tolerance genes is the basis for the development of improved cultivars which will retain high productivity on acid soils that is more preferable strategy than liming the soils.

Data Availability

Sequencing data used to support the findings of this study have been deposited in Sequence Read Archive–SRP089959. Transcript sequences and annotations could be downloaded from the following link: http://194.226.21.15/. Flax.Aluminium/. Gene expression data used to support the

findings of this study are included in the Supplementary Files.

Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Authors' Contributions

George S. Krasnov, Alexey A. Dmitriev, and Alexander V. Zyablitsin contributed equally to this work.

Acknowledgments

Supplementary Materials

S1 Table. Gene expression alterations in all analyzed flax genotypes under aluminum stress. S2 Table. Gene expression alterations in resistant flax genotypes under aluminum stress. S3 Table. Gene expression alterations in sensitive flax genotypes under aluminum stress. S4 Table. Gene expression alterations in resistant flax cultivar Hermes under aluminum stress. S5 Table. Gene expression alterations in resistant flax cultivar TMP1919 under aluminum stress. S6 Table. Gene expression alterations in sensitive flax cultivar Lira under aluminum stress. S7 Table. Gene expression alterations in sensitive flax cultivar Orshanskiy under aluminum stress. S8 Table. Diverse expression alterations between groups of resistant and sensitive to aluminum flax genotypes. S9 Table. GO enrichment analysis for tops of up- and downregulated genes in the pool of all studied flax cultivars after aluminum exposure. S10 Table. GO enrichment analysis for tops of up- and downregulated genes in the pool of resistant flax cultivars after aluminum exposure. S11 Table. GO enrichment analysis for tops of up- and downregulated genes in the pool of sensitive flax cultivars after aluminum exposure. S12 Table. GO enrichment analysis for tops of up- and downregulated genes in resistant flax cultivar Hermes after aluminum exposure. S13 Table. GO enrichment analysis for tops of up- and downregulated genes in resistant flax cultivar TMP1919 after aluminum exposure. S14 Table. GO enrichment analysis for tops of up- and downregulated genes in sensitive flax cultivar Lira after aluminum exposure. S15 Table. GO enrichment analysis for tops of up- and downregulated genes in sensitive flax cultivar Orshanskiy after aluminum exposure. S16 Figure. Hierarchical cluster dendrogram showing coexpression modules identified using weighted gene coexpression network analysis (WGCNA) for flax under aluminum stress. **Symbols in Supplementary Tables:** Gene-gene identifier (generated by Trinity). Transcript lengths: comma-separated lengths of all predicted transcripts. Blast symbol: gene symbol of top BLAST hit(s). Blast name: gene name of top BLAST hit(s). Blast species: organism to which top BLAST hit(s) belongs. KEGG Orthology: KEGG BRITE gene/protein identifier (pan-organismal) of the top BLAST hit. KEGG A.Thaliana: KEGG gene/protein identifier (*Arabidopsis thaliana*) of the top BLAST hit. Gene Ontology: Gene Ontology annotation of the top BLAST hit. logFC: binary logarithm of gene expression level fold change between intact and Al-treated plants (an average across all four cultivars). logCPM: binary logarithm of read counts per million (average across all the samples). LR: likelihood ratio from edgeR's likelihood ratio test (LRT). PValue: LRT p-value. FDR: adjusted LRT p-value. Hermes-C: CPM (Hermes, control). Hermes-Al4: CPM (Hermes, Al-treated for 4 hours). TMP-C: CPM (TMP1919, control). TMP-Al4: CPM (TMP1919, Al-treated for 4 hours). Lira-C: read counts per million (CPM) per gene (Lira, control). Lira-Al4: CPM (Lira, Al-treated for 4 hours). Orshanskiy-C: CPM (Orshanskiy, control). Orshanskiy-Al4: CPM (Orshanskiy, Al-treated for 4 hours). *(Supplementary Materials)*

References

[1] S. K. Panda, F. Baluska, and H. Matsumoto, "Aluminum stress signaling in plants," *Plant Signaling and Behavior*, vol. 4, no. 7, pp. 592–597, 2009.

[2] T. Eekhout, P. Larsen, and L. De Veylder, "Modification of DNA Checkpoints to Confer Aluminum Tolerance," *Trends in Plant Science*, vol. 22, no. 2, pp. 102–105, 2017.

[3] T. Grevenstuk and A. Romano, "Aluminium speciation and internal detoxification mechanisms in plants: Where do we stand?" *Metallomics*, vol. 5, no. 12, pp. 1584–1594, 2013.

[4] N. Gupta, S. S. Gaurav, and A. Kumar, "Molecular basis of aluminium toxicity in plants: a review," *American Journal of Plant Sciences*, vol. 4, no. 12, pp. 21–37, 2013.

[5] Y. Shavrukov and Y. Hirai, "Good and bad protons: Genetic aspects of acidity stress responses in plants," *Journal of Experimental Botany*, vol. 67, no. 1, pp. 15–30, 2016.

[6] L. V. Kochian, M. A. Piñeros, and O. A. Hoekenga, "The physiology, genetics and molecular biology of plant aluminum resistance and toxicity," *Plant and Soil*, vol. 274, no. 1-2, pp. 175–195, 2005.

[7] B. Du, H. Nian, Z. Zhang, and C. Yang, "Effects of aluminum on superoxide dismutase and peroxidase activities, and lipid peroxidation in the roots and calluses of soybeans differing in aluminum tolerance," *Acta Physiologiae Plantarum*, vol. 32, no. 5, pp. 883–890, 2010.

[8] L. V. Kochian, M. A. Piñeros, J. Liu, and J. V. Magalhaes, "Plant adaptation to acid soils: the molecular basis for crop aluminum resistance," *Annual Review of Plant Biology*, vol. 66, pp. 571–598, 2015.

[9] H. Matsumoto, D. E. Riechers, A. V. Lygin, F. Baluška, and M. Sivaguru, "Aluminum signaling and potential links with safener-induced detoxification in plants," in *Aluminum Stress Adaptation in Plants*, S. K. Panda and F. Baluška, Eds., vol. 24 of *Signaling and Communication in Plants*, pp. 1–35, Springer International Publishing, Cham, Switzerland, 2015.

[10] J. Bose, O. Babourina, Y. Ma, M. Zhou, S. Shabala, and Z. Rengel, "Specificity of ion uptake and homeostasis maintenance during acid and aluminium stresses," in *Aluminum Stress Adaptation in Plants*, S. K. Panda and F. Baluška, Eds., vol. 24 of *Signaling and Communication in Plants*, pp. 229–251, Springer International Publishing, Cham, Switzerland, 2015.

[11] G. J. Taylor, "Exclusion of metals from the symplasm: A possible mechanism of metal tolerance in higher plants," *Journal of Plant Nutrition*, vol. 10, no. 9-16, pp. 1213–1222, 1987.

[12] G. J. Taylor, "The physiology of aluminum phytotoxicity," *Metal Ions in Biological Systems*, vol. 24, pp. 123–163, 1988.

[13] J. Liu, J. V. Magalhaes, J. Shaff, and L. V. Kochian, "Aluminum-activated citrate and malate transporters from the MATE and ALMT families function independently to confer Arabidopsis aluminum tolerance," *The Plant Journal*, vol. 57, no. 3, pp. 389–399, 2009.

[14] A. Mangeon, R. Pardal, A. D. Menezes-Salgueiro et al., "AtGRP3 is implicated in root size and aluminum response pathways in arabidopsis," *PLoS ONE*, vol. 11, no. 3, Article ID e0150583, 2016.

[15] J. Moustaka, G. Ouzounidou, G. Bayçu, and M. Moustakas, "Aluminum resistance in wheat involves maintenance of leaf Ca2+ and Mg2+ content, decreased lipid peroxidation and Al accumulation, and low photosystem II excitation pressure," *BioMetals*, vol. 29, no. 4, pp. 611–623, 2016.

[16] L. G. Maron, C. T. Guimarães, M. Kirst et al., "Aluminum tolerance in maize is associated with higher MATE1 gene copy number," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 110, no. 13, pp. 5241–5246, 2013.

[17] R. A. Arenhart, Y. Bai, L. F. Valter De Oliveira et al., "New insights into aluminum tolerance in rice: The ASR5 protein binds the STAR1 promoter and other aluminum-responsive genes," *Molecular Plant*, vol. 7, no. 4, pp. 709–721, 2014.

[18] J. Xia, N. Yamaji, and J. F. Ma, "A plasma membrane-localized small peptide is involved in rice aluminum tolerance," *The Plant Journal*, vol. 76, no. 2, pp. 345–355, 2013.

[19] E. Bojórquez-Quintal, C. Escalante-Magaña, I. Echevarría-Machado, and M. Martínez-Estévez, "Aluminum, a friend or foe of higher plants in acid soils," *Frontiers in Plant Science*, vol. 8, article no. 1767, 2017.

[20] A. A. Daspute, A. Sadhukhan, M. Tokizawa, Y. Kobayashi, S. K. Panda, and H. Koyama, "Transcriptional regulation of aluminum-tolerance genes in higher plants: Clarifying the underlying molecular mechanisms," *Frontiers in Plant Science*, vol. 8, article no. 1358, 2017.

[21] M. Cai, N. Wang, C. Xing, F. Wang, K. Wu, and X. Du, "Immobilization of aluminum with mucilage secreted by root cap and root border cells is related to aluminum resistance in Glycine max L," *Environmental Science and Pollution Research*, vol. 20, no. 12, pp. 8924–8933, 2013.

[22] S. C. Miyasaka and M. C. Hawes, "Possible role of root border cells in detection and avoidance of aluminum toxicity," *Plant Physiology*, vol. 125, no. 4, pp. 1978–1987, 2001.

[23] G. E. Alves Silva, F. Toledo Ramos, A. P. de Faria, and M. G. Costa França, "Seeds' physicochemical traits and mucilage protection against aluminum effect during germination and root elongation as important factors in a biofuel seed crop (Ricinus communis)," *Environmental Science and Pollution Research*, vol. 21, no. 19, pp. 11572–11579, 2014.

[24] J. F. Ma, P. R. Ryan, and E. Delhaize, "Aluminium tolerance in plants and the complexing role of organic acids," *Trends in Plant Science*, vol. 6, no. 6, pp. 273–278, 2001.

[25] J. Xia, N. Yamaji, T. Kasai, and J. F. Ma, "Plasma membrane-localized transporter for aluminum in rice," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 107, no. 43, pp. 18381–18385, 2010.

[26] M. Reyes-Díaz, C. Inostroza-Blancheteau, and Z. Rengel, "Physiological and molecular regulation of aluminum resistance in woody plant species," in *Aluminum Stress Adaptation in Plants*, S. K. Panda and F. Baluška, Eds., vol. 24 of *Signaling and Communication in Plants*, pp. 187–202, Springer International Publishing, Cham, Switzerland, 2015.

[27] K. D. Richards, E. J. Schott, Y. K. Sharma, K. R. Davis, and R. C. Gardner, "Aluminum induces oxidative stress genes in *Arabidopsis thaliana*," *Plant Physiology*, vol. 116, no. 1, pp. 409–418, 1998.

[28] B. Ezaki, M. Katsuhara, M. Kawamura, and H. Matsumoto, "Different mechanisms of four aluminum (Al)-resistant transgenes for Al toxicity in Arabidopsis," *Plant Physiology*, vol. 127, no. 3, pp. 918–927, 2001.

[29] A. L. Garcia-Oliveira, C. Poschenrieder, J. Barceló, and P. Martins-Lopes, "Breeding for al tolerance by unravelling genetic diversity in bread wheat," in *Aluminum Stress Adaptation in Plants*, S. K. Panda and F. Baluška, Eds., vol. 24 of *Signaling and Communication in Plants*, pp. 125–153, Springer International Publishing, Cham, Switzerland, 2015.

[30] J. Mammadov, R. Buyyarapu, S. K. Guttikonda, K. Parliament, I. Y. Abdurakhmonov, and S. P. Kumpatla, "Wild relatives of maize, rice, cotton, and soybean: treasure troves for tolerance to biotic and abiotic stresses," *Frontiers in Plant Science*, vol. 9, article no. 886, 2018.

[31] J. L. Adolphe, S. J. Whiting, B. H. J. Juurlink, L. U. Thorpe, and J. Alcorn, "Health effects with consumption of the flax lignan secoisolariciresinol diglucoside," *British Journal of Nutrition*, vol. 103, no. 7, pp. 929–938, 2010.

[32] A. Touré and X. Xueming, "Flaxseed lignans: Source, biosynthesis, metabolism, antioxidant activity, Bio-active components, and health benefits," *Comprehensive Reviews in Food Science and Food Safety*, vol. 9, no. 3, pp. 261–269, 2010.

[33] N. V. Kishlyan and T. A. Rozhmina, "Ocenka genofonda lna kulturnogo (Linum usitatissimum L.) po kislotoustojchivosti (on Russian)," *Agricultural Biology*, vol. 1, pp. 96–103, 2010.

[34] L. Galindo-Gonzalez and M. K. Deyholos, "RNA-seq transcriptome response of flax (Linum usitatissimum L.) to the pathogenic fungus fusarium oxysporum f. sp. lini," *Frontiers in Plant Science*, vol. 7, article no. 1766, 2016.

[35] N. V. Melnikova, A. A. Dmitriev, M. S. Belenikin et al., "Identification, expression analysis, and target prediction of flax genotroph microRNAs under normal and nutrient stress conditions," *Frontiers in Plant Science*, vol. 7, article 399, 2016.

[36] Y. Yu, G. Wu, H. Yuan et al., "Identification and characterization of miRNAs and targets in flax (Linum usitatissimum) under saline, alkaline, and saline-alkaline stresses," *BMC Plant Biology*, vol. 16, no. 1, article 124, 2016.

[37] N. V. Melnikova, A. A. Dmitriev, M. S. Belenikin et al., "Excess fertilizer responsive miRNAs revealed in Linum usitatissimum L," *Biochimie*, vol. 109, pp. 36–41, 2015.

[38] Y. Yu, W. Huang, H. Chen et al., "Identification of differentially expressed genes in flax (Linum usitatissimum L.) under saline-alkaline stress by digital gene expression," *Gene*, vol. 549, no. 1, pp. 113–122, 2014.

[39] P. K. Dash, R. Rai, A. K. Mahato, K. Gaikwad, and N. K. Singh, "Transcriptome landscape at different developmental stages

of a drought tolerant cultivar of flax (Linum usitatissimum)," *Frontiers in Chemistry*, vol. 5, article no. 82, 2017.

[40] A. A. Dmitriev, G. S. Krasnov, T. A. Rozhmina et al., "Differential gene expression in response to Fusarium oxysporum infection in resistant and susceptible genotypes of flax (Linum usitatissimum L.)," *BMC Plant Biology*, vol. 17, article no. 253, 2017.

[41] A. A. Dmitriev, G. S. Krasnov, T. A. Rozhmina et al., "Flax (Linum usitatissimum L.) response to non-optimal soil acidity and zinc deficiency," *BMC Plant Biology*, vol. 19, article no. 54, 2019.

[42] A. A. Dmitriev, G. S. Krasnov, T. A. Rozhmina et al., "Glutathione S-transferases and UDP-glycosyltransferases are involved in response to aluminum stress in flax," *Frontiers in Plant Science*, vol. 7, article no. 1920, 2016.

[43] A. V. Zyablitsin, A. A. Dmitriev, G. S. Krasnov et al., "CAX3 gene is involved in flax response to high soil acidity and aluminum exposure," *Journal of Molecular Biology*, vol. 52, no. 4, pp. 514–519, 2018.

[44] A. A. Dmitriev, A. V. Kudryavtseva, N. L. Bolsheva et al., "MiR319, miR390, and miR393 are involved in aluminum response in flax (Linum usitatissimum L.)," *BioMed Research International*, vol. 2017, Article ID 4975146, 6 pages, 2017.

[45] N. V. Kishlyan and T. A. Rozhmina, "Laboratornyj metod ocenki genotipov lna dolgunca na alyumoustojchivost," in *Innovacionnye Razrabotki Lnovodstvu*, p. 11, 2011.

[46] B. Langmead and S. L. Salzberg, "Fast gapped-read alignment with Bowtie 2," *Nature Methods*, vol. 9, no. 4, pp. 357–359, 2012.

[47] B. Li and C. N. Dewey, "RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome," *BMC Bioinformatics*, vol. 12, no. 1, article 323, 2011.

[48] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.," *Bioinformatics*, vol. 26, no. 1, pp. 139-140, 2010.

[49] R. Huis, S. Hawkins, and G. Neutelings, "Selection of reference genes for quantitative gene expression normalization in flax (Linum usitatissimum L.)," *BMC Plant Biology*, vol. 10, article no. 71, 2010.

[50] G. S. Krasnov, N. Y. Oparina, A. A. Dmitriev et al., "RPN1, a new reference gene for quantitative data normalization in lung and kidney cancer," *Journal of Molecular Biology*, vol. 45, no. 2, pp. 211–220, 2011.

[51] P. Langfelder and S. Horvath, "WGCNA: an R package for weighted correlation network analysis," *BMC Bioinformatics*, vol. 9, article 559, 2008.

[52] L. Gramzow and G. Theißen, "Phylogenomics of MADS-box genes in plants - Two opposing life styles in one gene family," *Biology*, vol. 2, no. 3, pp. 1150–1164, 2013.

[53] J. Jia, P. Zhao, L. Cheng et al., "MADS-box family genes in sheepgrass and their involvement in abiotic stress responses," *BMC Plant Biology*, vol. 18, no. 1, p. 42, 2018.

[54] K. Duan, L. Li, P. Hu, S.-P. Xu, Z.-H. Xu, and H.-W. Xue, "A brassinolide-suppressed rice MADS-box transcription factor, OsMDP1, has a negative regulatory role in BR signaling," *The Plant Journal*, vol. 47, no. 4, pp. 519–531, 2006.

[55] G. N. Khong, P. K. Pati, F. Richaud et al., "OsMADS26 negatively regulates resistance to pathogens and drought tolerance in rice," *Plant Physiology*, vol. 169, no. 4, pp. 2935–2949, 2015.

[56] H. Shao, H. Wang, and X. Tang, "NAC transcription factors in plant multiple abiotic stress responses: progress and prospects," *Frontiers in Plant Science*, vol. 6, article no. 902, 2015.

[57] M. Nuruzzaman, A. M. Sharoni, and S. Kikuchi, "Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants," *Frontiers in Microbiology*, vol. 4, 2013.

[58] M. Fujita, Y. Fujita, K. Maruyama et al., "A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway," *The Plant Journal*, vol. 39, no. 6, pp. 863–876, 2004.

[59] X.-J. He, R.-L. Mu, W.-H. Cao, Z.-G. Zhang, J.-S. Zhang, and S.-Y. Chen, "AtNAC2, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development," *The Plant Journal*, vol. 44, no. 6, pp. 903–916, 2005.

[60] Y. Zhu, J. Yan, W. Liu et al., "Phosphorylation of a NAC transcription factor by a calcium/calmodulin-dependent protein kinase regulates abscisic acid-induced antioxidant defense in maize," *Plant Physiology*, vol. 171, no. 3, pp. 1651–1664, 2016.

[61] S. Balazadeh, M. Kwasniewski, C. Caldana et al., "ORS1, an H2O2-responsive NAC transcription factor, controls senescence in arabidopsis thaliana," *Molecular Plant*, vol. 4, no. 2, pp. 346–360, 2011.

[62] G. M. A. Cançado, F. T. S. Nogueira, S. R. Camargo, R. D. Drummond, R. A. Jorge, and M. Menossi, "Gene expression profiling in maize roots under aluminum stress," *Biologia Plantarum*, vol. 52, no. 3, pp. 475–485, 2008.

[63] H. F. Escobar-Sepúlveda, L. I. Trejo-Téllez, S. García-Morales, F. C. Gómez-Merino, and K. Wu, "Expression patterns and promoter analyses of aluminum-responsive NAC genes suggest a possible growth regulation of rice mediated by aluminum, hormones and NAC transcription factors," *PLoS ONE*, vol. 12, no. 10, Article ID e0186084, 2017.

[64] M. Moreno-Alvarado, S. García-Morales, L. I. Trejo-Téllez, J. V. Hidalgo-Contreras, and F. C. Gómez-Merino, "Aluminum enhances growth and sugar concentration, alters macronutrient status and regulates the expression of nac transcription factors in rice," *Frontiers in Plant Science*, vol. 8, article no. 73, 2017.

[65] Z. J. Ding, J. Y. Yan, X. Y. Xu, G. X. Li, and S. J. Zheng, "WRKY46 functions as a transcriptional repressor of *ALMT1*, regulating aluminum-induced malate secretion in Arabidopsis," *The Plant Journal*, vol. 76, no. 5, pp. 825–835, 2013.

[66] G. Z. Li, Z. Q. Wang, K. Yokosho et al., "Transcription factor WRKY22 promotes aluminum tolerance via activation of OsFRDL4 expression and enhancement of citrate secretion in rice (Oryza sativa)," *New Phytologist*, vol. 219, no. 1, pp. 149–162, 2018.

[67] W. Fan, H. Q. Lou, J. L. Yang, and S. J. Zheng, "The roles of STOP1-like transcription factors in aluminum and proton tolerance," *Plant Signaling and Behavior*, vol. 11, no. 2, Article ID e1131371, 2016.

[68] R. A. Arenhart, J. C. De Lima, M. Pedron et al., "Involvement of ASR genes in aluminium tolerance mechanisms in rice," *Plant, Cell & Environment*, vol. 36, no. 1, pp. 52–67, 2013.

[69] R. A. Arenhart, M. Schunemann, L. Bucker Neto, R. Margis, Z.-Y. Wang, and M. Margis-Pinheiro, "Rice ASR1 and ASR5 are complementary transcription factors regulating aluminium responsive genes," *Plant, Cell & Environment*, vol. 39, no. 3, pp. 645–651, 2016.

[70] H. Gall, F. Philippe, J. Domon, F. Gillet, J. Pelloux, and C. Rayon, "Cell wall metabolism in response to abiotic stress," *Plants*, vol. 4, no. 1, pp. 112–166, 2015.

[71] W. J. Horst, Y. Wang, and D. Eticha, "The role of the root apoplast in aluminium-induced inhibition of root elongation

and in aluminium resistance of plants: a review," *Annals of Botany*, vol. 106, no. 1, pp. 185–197, 2010.

[72] Y. Wu, Z. Yang, J. How, H. Xu, L. Chen, and K. Li, "Overexpression of a peroxidase gene (AtPrx64) of Arabidopsis thaliana in tobacco improves plant's tolerance to aluminum stress," *Plant Molecular Biology*, vol. 95, no. 1-2, pp. 157–168, 2017.

[73] M. Safari, F. Ghanati, M. R. Safarnejad, and N. A. Chashmi, "The contribution of cell wall composition in the expansion of Camellia sinensis seedlings roots in response to aluminum," *Planta*, vol. 247, no. 2, pp. 381–392, 2018.

[74] W. Liu, F. Xu, T. Lv et al., "Spatial responses of antioxidative system to aluminum stress in roots of wheat (Triticum aestivum L.) plants," *Science of the Total Environment*, vol. 627, pp. 462–469, 2018.

[75] C. Sun, L. Liu, W. Zhou, L. Lu, C. Jin, and X. Lin, "Aluminum induces distinct changes in the metabolism of reactive oxygen and nitrogen species in the roots of two wheat genotypes with different aluminum resistance," *Journal of Agricultural and Food Chemistry*, vol. 65, no. 43, pp. 9419–9427, 2017.

[76] M. Sharma, V. Sharma, and B. N. Tripathi, "Rapid activation of catalase followed by citrate efflux effectively improves aluminum tolerance in the roots of chick pea (Cicer arietinum)," *Protoplasma*, vol. 253, no. 3, pp. 709–718, 2016.

[77] T. Sasaki, B. Ezaki, and H. Matsumoto, "A gene encoding multidrug resistance (mdr)-like protein is induced by aluminum and inhibitors of calcium flux in wheat," *Plant & Cell Physiology (PCP)*, vol. 43, no. 2, pp. 177–185, 2002.

[78] P. B. Larsen, M. J. B. Geisler, C. A. Jones, K. M. Williams, and J. D. Cancel, "*ALS3* encodes a phloem-localized ABC transporter-like protein that is required for aluminum tolerance in Arabidopsis," *The Plant Journal*, vol. 41, no. 3, pp. 353–363, 2005.

[79] C.-F. Huang, N. Yamaji, Z. Chen, and J. F. Ma, "A tonoplast-localized half-size ABC transporter is required for internal detoxification of aluminum in rice," *The Plant Journal*, vol. 69, no. 5, pp. 857–867, 2012.

[80] H. Chen, C. Lu, H. Jiang, and J. Peng, "Global transcriptome analysis reveals distinct aluminum-tolerance pathways in the Al-accumulating species hydrangea macrophylla and marker identification," *Plos One*, vol. 10, no. 12, Article ID e0144927, 2015.

[81] H. Sade, B. Meriga, V. Surapu et al., "Toxicity and tolerance of aluminum in plants: tailoring plants to suit to acid soils," *BioMetals*, vol. 29, no. 2, pp. 187–210, 2016.

[82] H. Q. Lou, Y. L. Gong, W. Fan et al., "A formate dehydrogenase confers tolerance to aluminum and low pH," *Plant Physiology*, vol. 171, no. 1, pp. 294–305, 2016.

[83] S. K. Panda, L. Sahoo, M. Katsuhara, and H. Matsumoto, "Overexpression of alternative oxidase gene confers aluminum tolerance by altering the respiratory capacity and the response to oxidative stress in tobacco cells," *Molecular Biotechnology*, vol. 54, no. 2, pp. 551–563, 2013.

Review Article

Insight on Rosaceae Family with Genome Sequencing and Functional Genomics Perspective

Prabhakaran Soundararajan, So Youn Won, and Jung Sun Kim [ID]

Department of Agricultural Biotechnology, National Institute of Agricultural Sciences, RDA, Jeonju 54874, Republic of Korea

Correspondence should be addressed to Jung Sun Kim; jsnkim@korea.kr

Rosaceae is one of the important families possessing a variety of diversified plant species. It includes many economically valuable crops that provide nutritional and health benefits for the human. Whole genome sequences of valuable crop plants were released in recent years. Understanding of genomics helps to decipher the plant physiology and developmental process. With the information of cultivating species and its wild relative genomes, genome sequence-based molecular markers and mapping loci for economically important traits can be used to accelerate the genome assisted breeding. Identification and characterization of disease resistant capacities and abiotic stress tolerance related genes are feasible to study across species with genome information. Further breeding studies based on the identification of gene loci for aesthetic values, flowering molecular circuit controls, fruit firmness, nonacid fruits, etc. is required for producing new cultivars with valuable traits. This review discusses the whole genome sequencing reports of *Malus*, *Pyrus*, *Fragaria*, *Prunus*, and *Rosa* and status of functional genomics of representative traits in individual crops.

1. Introduction

Rosaceae consists of 100 genera and 3,000 species. It is one of the most economically important families which comprised the fruit, nut, ornamental, aroma, herb, and woody plants. Edible crops domesticated for human consumption in Rosaceae include apple, strawberry, pear, peach, plum, almond, raspberry, sour cherry, and sweet cherry. Though most of the choices are dietary based, some of the vital phytochemicals and antioxidants in fruits of Rosaceae have potential to inhibit cancer. For instance, ellagic acid abundant in strawberry, red raspberry, and arctic bramble was shown to prevent cell proliferation and induce apoptosis of cancer cells [1, 2].

Rosaceae consist of highly distinctive fruit types such as drupe, pome, drupelet, and achene. Conventionally, Rosaceae has been divided into four subfamilies based on the fruit types such as Rosoideae (several apocarpous pistils mature into achenes), Amygdaloideae/Prunoideae (single monocarpellate pistil mature into a drupe), Spiraeaoideae, (gynoecium consists of two or more apocarpous pistils mature into follicles), and Maloideae/Pomoideae (ovary is compound and

inferior where floral receptacle is fleshy edible tissues) [3]. Recently, the phylogeny of Rosaceae has been divided into three basal groups based on nuclear and chloroplast loci, namely, Amygdaloideae, Rosoideae, and Dryadoideae [1]. Amygdaloideae has included the other subfamilies such as former Amygdaloideae (n=8) (plum, cherry, apricot peach, almond, etc.), Spiraeaoideae (n=9) (*Spiraea*, *Aruncus*, *Sorbaria*, etc.), and Maloideae (n=17) (apple, pear, cotoneaster, etc.). Rosoideae (n=7) includes *Fragaria*, *Potentilla*, *Rosa*, and *Rubus*. Dryadoideae (n=9) includes *Cercocarpus*, *Chamaebatia*, *Dryas*, and *Purshia*.

Exhaustive breeding on fruit trees offered different types of variety with variant alleles of genes controlling the key traits. To produce the sustainable cultivars we need to extend functional genomics studies in Rosaceae. As Rosaceae consists of highly distinctive types of fruits and diversified growth patterns, multiple genome models are required to improve the agronomic practices, produce high-yield and disease resistance varieties, overcome self-incompatibility, and reduce juvenile period, long-lasting postharvest self-life, tolerant to chilling (storage), firmness against transportation damage, and higher nutritional content and health benefit

values. An emergence of next-generation sequencing (NGS) technologies revolutionize biological field with its feasibility to assemble and annotate any size and number of the genome(s) [4]. High-throughput genome sequencing offers the substantial advantages for the explicit understanding of genetics and genomics [5]. Recent breakthrough in the sequencing technologies and the availability of tools improve the accuracy of *de novo* genome sequencing. Unveiling the genome information gives us an invaluable insight into the epigenetic characteristics [6]. Genes responsible for traits of agronomic importance are rapidly identified and characterized with the forward and reverse genetics studies on many plants [4]. Genome-wide association studies (GWAS) characterize the functional role(s) of gene [5]. Genotyping-by-sequencing (GBS) and marker assisted selection (MAS) helps the precise breeding program [4]. Genomics provides huge amount of information in convenient manner for evolutional studies. Comparative analysis among diverse plant families helps to know about the evolutionary details of the gene(s)/plant(s) [7]. Candidate gene mapping in one species serves as a substrate for comparative analysis of other related species [5].

Therefore, this review will cover the progress of NGS of important commercial and model plants in Rosaceae such as apple, pear, strawberry, peach, sweet cherry, apricot, and rose. Brief information about the functional genomics studies conducted on critical key traits of the above-mentioned plants are also covered in this review.

2. Genome Assembly and Annotation

Genome-scale study gives rich candidate genetic resource to deciphering the functional and regulatory networks for growth and development. NGS is the perfect platform to know about the genomic information which has wide application in crop improvement and evolutionary studies. Genome sequencing details of apple, pear, strawberry, peach, and rose have been given in Table 1. Desirable key traits will be discussed in functional genomics section.

2.1. Apple. Apple fruit has higher nutritional values. For several centuries, humans consumed apple-based beverages such as ciders [8]. *Malus* x *domestica* or *M. pumila* is the widely growing apple tree. Ancestor of domesticated *M. domestica* is *M. sieversii*. It is originated in Central Asia (Southern China). Wild *M. pumila* tree bearing smaller sized fruits is still covered 80% of Tian Shan Mountains. Microsatellite markers study showed that *M. domestica* is genetically similar to European crabapple *M. sylvestris* than to the Asian wild apple *M. sieversii* [9, 10].

So far three genomes have been released in apple. Firstly, Velasco et al. (2010) covered 81.3% (603.9 Mb) of *M* x *domestica* Borkh "Golden Delicious" genome. In that, 57,386 genes were identified. Almost 67.4% of *M* x *domestica* genome consists of repetitive sequences [11]. Secondly, Li et al. (2016) covered about 90% (632.4 Mb) of *M.* x *domestica* Borkh "Golden Delicious" genome. A total number of identified protein-coding and noncoding genes were 53,922 and 2,765, respectively [12]. Thirdly, Daccord et al. (2017) assembled

genome of *M* x *domestica* Borkh "Golden Delicious doubled-haploid" line (GDDH13). Estimated genome size of GDDH13 is 651 Mb, from which 649.7 Mb (99.8%) was assembled. However, only 42,140 protein-coding genes and 1,965 non-protein coding genes were identified in GDDH13 genome [6]. Major burst of transposable elements (TEs) happening around 21 MYA was correlated with the uplift of the Tian Shan mountains as well as the diversification of apple and pear [11]. Study on structural and functional evolution of genome cannot be completed without characterizing the TEs. Around 59.5% of the GDDH13 genome was covered with the TE elements. Most interestingly, *HODOR* (High-Copy Golden Delicious repeat), TE consensus sequences are present at about 22.3 Mb (3.6% of genome) [6].

2.2. Pear. Pear is one of the most important temperate fruit. It is originated in Western China. In spite of thousands of cultivars, based on the habituation, *Pyrus* species are divided into two major groups such as Occidental pears or European pears (*P. communis*) and Oriental pears or Asiatic pears (*P. bretschneideri*) [13]. Nevertheless commercially important cultivars were domesticated from the wide range of wild species; still pear cultivation faces challenges such as susceptibility to the pear scab, black spot disease, self-incompatibility, early ripening, short shelf life, firmness, sucrose content, grit/stone cells, color and odor of fruit, and inbreeding depression [14].

Recently 97.1% of *P. bretschneideri* Rehd. (Chinese pear) genome, i.e., 512.0 Mb (42,812 genes), has been annotated by Wu et al. [15]. Following it, 577.3 Mb of *P. communis* (European pear) was sequenced. It covered around 98.4% of genome containing 43,419 genes [16]. Pear is phylogenetically closer towards the apple [1]. Hence higher collinearity was existed between the chromosomes of pear and apple. Pear and apple divergence could have happened only 5.4-21.5 MYA [15]. Presence of repetitive sequence about 53.1% in *P. bretschneideri* [15] and 34.5% in *P. communis* [16] hampered the investigation of uncharacterized regions.

2.3. Strawberry. Strawberry comes under the category of soft fruit. It is widely attracted for its aroma, bright red color, texture, and taste. Preserved/processed strawberries are largely used for ice creams, milkshakes, chocolates, etc. It is considered to be difficult to propagate.

2.3.1. Fragaria vesca. *Fragaria vesca* is a diploid species generally called woodland strawberry. It has unique characteristics such as day neutrality, nonrunning, and yellow colored fruits. It is self-compatible and has short generation time. It is indigenous to northern Eurasia and North America [17].

Small genome (240.0 Mb) of strawberry (*Fragaria vesca* "Hawaai4") showed the absence of whole genome duplications. Though all members of rosids shared the ancient triplication, no evidence of whole genome duplication was found in *F. vesca*. About 99.8% (239.5 Mb) of genome was covered with identification of 33,264 genes [17]. Later, Darwish et al. done the reference based reannotation and assembly of woodland strawberry *F. vesca* "YW5AF7" genome [18]. Similar

TABLE 1: Genome sequencing of important commercial plants belongs to the Rosaceae family.

Common name	Sample name	Chr number	Genome size Estimated (Mb)	Genome size Assembled (Mb)	Coverage (%)	Platform	Number of genes	Repetitive sequences (Mb)	Reference
Apple	Mallus x domestica "Golden Delicous"		742.3	603.9	81.3	BAC + 454	57,386	362.3	Velasco et al., 2010
	Mallus x domestica "Golden Delicous" (Heterologous)	2n=2x=34	701.0	632.4	90.2	Illumina+ PacBio	53,922	382.0	Li et al., 2016
	Mallus x domestica "Golden Delicous" doubled-haploid"		651.0	649.7	99.8	Illumina+ PacBio	42,140	372.2	Daccord et al., 2017
Pear	Pyrus bretschneideri "Dangshansuli"	2n=2x=34	512.0	501.3	97.9	BAC-by-BAC + Illumina	42,812	240.2	Wu et al., 2013
	Pyrus communis "Bartlett"		600.0	577.3	96.2	454	43,419	197.7	Chagné et al., 2014
Strawberry	Fragaria vesca ssp vesca acc. Hawaii 4	2n=2x=14	240.0	239.5	99.8	Illumina + 454 + SOLiD	33,264	49.8	Shulaev et al., 2010
	Fragaria x ananassa "Reikou"	2n=8x=56	692.0	697.7	100.8*		64,947	328.3	Hirakawa et al. 2014
	Fragaria iinumae		221.0	199.6	90.3		26,411	63.2	
	Fragaria nipponica	2n=2x=14	208.0	206.5	99.3	454 + Illumina	21,540	52.5	
	Fragaria nubicola		202.0	203.7	100.8*		21,053	49.9	
	Fragaria orientalis		349.3	214.2	61.3		17,239	56.2	
Chinese plum and Japanese apricot	Prunus mume "Mei"	2n=2x=16	280.0	237	84.6	Illumina	31,390	106.8	Zhang et al., 2012
Peach	Prunus persica "Lovell" v1.0	2n=2x=16	265.0	224.6	84.7	BAC-by-BAC	27,852	84.41	Verde et al., 2013
	Prunus persica "Lovell" v2.0			227.4	85.8	Illumina	26,873	-	Verde et al., 2017
Sweet cherry	Prunus avium "Santonishiki"	2n=2x=16	380.0	272.4	77.8	Illumina	43,349	119.4	Shirasawa et al., 2017
Rose	Rosa chinensis "Old Blush"		560.0	503.0	97.7	Illumina+ PacBio	36,377	341.5	Raymond et al. 2018
	Rosa chinensis "Old Blush" (doubled haploid – "HapOB")	2n=2x=14	568.0±9.0	512.0	90.1 ~ 96.1	Illumina+ PacBio	44,481	279.6	Saint-Oyant et al., 2018
	Rosa multiflora		750	711	94.8	Illumina	67,380	417.2	Nakamura et al., 2018

* The higher size of genome assembled than the estimated could be either due to limitation in the *kmer* abundance analysis or duplication occurring during the genome assembly of highly repetitive region.

to the macrosyntenic relationships between pear and apple, *Fragaria* shared the synteny with *Prunus*. Lesser genome size of *F. vesca* could be mainly due to the lack of highly abundant LTR retrotransposons (< 2,100 copies). Based on the obtained genome sequences, 389 rosaceous conserved orthologous set (RosCOS) markers were developed in Rosaceae [19].

2.3.2. Fragaria x ananassa.

F. x ananassa is commonly cultivated species that play an important role in the strawberry production worldwide. Interestingly, *F. x ananassa* was reported as an accidental hybrid rose in France during mid-1700 between *F. chiloensis* (Chile) and *F. virginiana* (North American cultivar) [17].

Genome size of this octoploid species *F. x ananassa* was estimated between 708 Mb and 720 Mb. *F. x ananassa* shared the genome information with wild diploids such as *F. iinumae*, *F. nipponica*, *F. nubicola*, and *F. orientalis* and their genome size is 221 Mb, 208 Mb, 202 Mb, and 349.3 Mb, respectively. The octaploid genome, *F. x ananassa*, was assembled about 697.7 Mb, and its wild relatives are as follows: *F. iinumae* 199.6 Mb (90.3%), *F. nipponica* 206.4 Mb (99.2%), *F. nubicola* 203.6 Mb, and *F. orientalis* 214.2 Mb (61.3%) [20]. In total, the number of genes identified from *F. x ananassa* was 230,838. Protein-coding genes identified in wild relatives are 76,760 in *F. iinumae*, 87,803 in *F. nipponica*, 85,062 in *F. nubicola*, and 99674 in *F. orientalis*. About 47.1% (328.2 Mb) of *F. x ananassa* genome consists of repeats. In case of wild relatives, 31.7% (63.3 Mb) in *F. iinumae*, 25.5% (52.6 Mb) in *F. nipponica*, 24.5% (49.9 Mb) in *F. nubicola*, and 26.3% (56.2 Mb) in *F. orientalis* are repeat regions in genome [20].

2.4. Prunus.

Prunus fruit has attractive bright shiny skin color, subtle flavor, and sweetness. It has long generation time and bigger plant size. It needs 3-5 years for flowering/fruit production from planting. Processed cherry product is sold worldwide.

2.4.1. Chinese Plum and Japanese Apricot (Prunus mume).

Prunus mume is the first plant in Prunoideae subfamily to be sequenced. Domestication of *P. mume* could have started 3,000 years ago in China [21]. This woody perennial is considered as the first tree to be bloomed during the transition from winter to spring at lesser than 0°C [22].

Out of 280 Mb of the genome size, 237.0 Mb (84.6%) was sequenced. Totally 31,390 protein-coding genes were characterized in the *P. mume*. Genome of *P. mume* provides information about the 1,154 candidate genes involved in flower aroma, flowering time, and disease resistance. Assembled genome contains 106.8 Mb (45.0%) of repetitive sequences. Investigation of *P. mume* genome with the *Vitis vinifera*, paleohexaploid ancestor showed that 27,819 gene models aligned with its seven ancestral chromosomes. It is noteworthy that 2,772 orthologs' (78.1%) collinearity blocks were present in the *P. mume* genome (Table 1). Comparative analysis of *P. mume* chromosome with the Rosaceae ancestral chromosome showed that 4, 5, and 7 chromosomes of *P. mume* does not undergo any changes and they are direct Rosaceae ancient chromosomes such as III, VII, and VI, respectively [23].

2.4.2. Peach (Prunus persica).

Peach is one of great fruit that provides vitamins, minerals, fiber, and antioxidant compounds. Peach fruit is also called nectarine due to smooth skin without fuzz or short hairs. Selection and domestication of peach could have started in Yangzi River valley, China, around 7,500 years ago [24].

Whole genome analysis of *P. persica* L. "Lovell" covered 224.6 Mb (84.7%) of genome (estimated total size 265 Mb) and represented 27,852 protein-coding genes. Repetitive sequences present in peach were estimated as 84.41 Mb (37.14%) which is lesser than the apple (42.4%) and grape (44.5%). 67.26 Mb (29.60%), 20.56 Mb (9.05%), and 17.14 Mb (7.54%) appeared as TEs, DNA transposons, and unclassified repeats, respectively [25]. Recently, *P. persica* "Lovell" double haploid genome version 2.0 was released with deep resequencing approach. Assembled genome of 227.4 Mb (85.8%) contains 26,873 genes [26].

2.4.3. Sweet Cherry (Prunus avium).

Prunus avium generally called sweet cherry is an important drupe fruit in the Rosacea family. Sweet cherry is used for human consumption and wild cherry trees for wood which is also called mazzards. Sweet cherry and sour cherry are the most commercial and edible crops in *Prunus* genus [27].

Genome size of *P. avium* is about approximately 350 Mb. Shirasawa et al. (2017) assembled about 77.8% (272.4 Mb) of the *P. avium* "Satonishiki". About 43.8% (119.4 Mb) of the *P. avium* genome were covered with the repetitive sequences. Among the 119.4 Mb of repeats, 85.1 Mb of repeats are unique to *P. avium* "Satonishiki". Identified genes clustered with the *P. persica*, *P. mume*, *M. domestica*, and *F. vesca*. 75,627 genes clusters are formed. 3,459 clusters (4,535 genes) from *P. avium* are present in all the investigated species and 16,151 clusters (21,642 genes) were found only in the *P. avium* with the absence of 869 clusters [28].

2.5. Rose.

Roses are one of the most essential ornamental plants worldwide. Ornamental value of rose enjoyed since the dawn of civilization. Cultivation of roses traced back to 3000 years ago. It consists of 200 species and most of them are polyploid. It has also been cultivated for its cosmetic values such as perfumes and antiques and also some of the phytochemicals of roses have high therapeutic values. Rose hips can be used to cure osteoarthritis [29].

2.5.1. Rosa chinensis.

Rosa chinensis is one of the important pot-type rose cultivars. Recently, Raymond et al. (2018) sequenced the whole genome of *R. chinensis* "Old Blush" and resequenced the major genotypes contributed for rose domestication. Totally, 503 Mb (97.7%) of the genome was assembled. Genome results comprised 36,377 protein-coding genes, 3,971 long noncoding RNAs, and 207 miRNAs. In the genome TEs were present about 67.9%. From that, 50.6% were identified as long-terminal-repeat retrotransposons [30]. From the doubled-haploid rose line of *R. chinensis* "Old Blush" ("HapOB") about 90.1 to 96.1% (512 Mb) of genome was assembled. About 466 Mb was anchored to seven pseudo-chromosomes and the remaining were assigned to the chromosome 0 (Chr0). Totally 44,481 genes were

identified including 39,669 protein-coding and 4,812 noncoding genes. Repeats covered about 279.6 Mb of genome [31]. *Rosa* and *Fragaria* genomes shared the eight chromosomes of ancestral Rosaceae with one chromosome fission and two fusions. Divergence of *Rosa*, *Fragaria*, and *Rubus* could have occurred within a short period [30]. Synteny analysis showed that chromosomes 1, 4, 5, 6, and 7 of *R. chinensis* have higher collinearity with chromosomes 7, 4, 3, 2, and 5 of *F. vesca*. Interestingly, chromosomes 2 and 3 of *R. chinensis* were detected as reciprocal translocation with chromosomes 6 and 1 of *F. vesca* [31].

2.5.2. Rosa multiflora. *Rosa multiflora* is a five-petal plant belongs to the section *Synstylae*. It is native to the eastern Asian regions [32]. *R. multiflora* was used for breeding purpose to the modern roses. Especially, its resistance locus (*Rdr1*), tolerance against powdery mildew was introgressed with the *R. hybrida* [33].

Genome size of *R. multiflora* was estimated as 750 Mb and about 711 Mb was sequenced. Assembled genome was characterized with 67,380 genes (54,893 complete genes and 12,487 partial genes). Repeat regions covered 56.4% (417.2 Mb) of assembled genome. Out of 18,956 gene clusters in *R. multiflora* 1,287, 904, and 241 clusters were shared with the *F. vesca*, *P. persica*, and *M* x *domestica*, respectively. *R. multiflora* shared more number of gene clusters with the *F. vesca* than the other two plants of Rosaceae. However, unique gene clusters and genes of *R. multiflora* are 2.5 (3,482 of *R. multiflora* and 1,397 of *F. vesca*) and 3.3 (14,663 of *R. multiflora* and 4,482 of *F. vesca*) times higher than the *F. vesca*, respectively [34].

3. Functional Genomics

3.1. Fruit Development and Sucrose Metabolism in Apple. Pome is a unique nature of false fruit formation from the basal part of sepals and receptacles. Velasco et al. (2010) suggest that pome could have evolved recently from Maleae specific WGD which could be a major factor contributing to apple development and its specificity [11]. Genes encoding for like-hetero chromatin protein 1 (LHP1) such as *MdLHP1a* and *MdLHP1b* regulate the flowering time of apple [35]. *Flowering locus T1* (*MdFT1*) can promote flowering whereas *terminal flower* (*MdTFL1* and *MdTFL2*) expressed in the vegetative part could repress flowering and maintains the vegetative meristem identity [36]. Soon after fertilization, higher expression of two *cyclin-dependent kinase* (*CDK*)*b* genes and one *cyclin-dependent kinase regulatory subunit* (*CKS*) *1* indicates the active cell division of fruits [37]. Transcription factors such as Agamous (AG), Fruitfull (FUL), AG-like (AGL)1/AGL5, Spatula (SPT), Crabs Claw (CRC), and Ettin (ETT) regulate the carpel identity and differentiation [38]. Microarray data on apple reported that SPT, ETT/Auxin Response Factor (ARF) 3, FUL/AGL8, and CRC transcripts were abundant during the fruit enlargement stage. However, most of their expressions are downregulated in cell division stage [39]. In apple, fruit development-related gene families such as *MADS-box* genes, carbohydrate metabolism, sorbitol assimilation, and transportation were expanded more than

the cucumber, soybean, poplar, *A. thaliana*, grape, rice, *Brachypodium*, sorghum, and maize [11]. Expression of α-*expansin* (α-EXP) was detected only during the cell expansion phase of apple [39]. MdMADS2.1 and MdMADS2.2, orthologous to FUL-like genes in *A. thaliana*, were progressively involved in the fruit developmental process. Among two candidate genes, MdMADS2.1 was closely associated with fruit flesh firmness [40]. ARF106 gene expressed during cell division and cell expansion stages is consistent with a potential role in the control of fruit size [41]. Methylation of DNA plays an essential role in the fruit size [12]. Comparative study between the bigger size apple (Golden Delicious) and smaller size apple (GDDH18) showed that twenty-two genes found as responsible for small size have lesser methylation in the promoter region [6].

After pollination, the small amount of starch present in the floral buds starts to metabolize. Conversion of carbon to sucrose was mediated by the tonoplast monosaccharide transporters (TMTs), *MdTMT1* and *MdTMT2*. Expansion of fruit cells is associated with the starch accumulation. Higher expression of *sorbitol dehydrogenase* (*SHD*), *cell wall invertase* (*CIN*), *neutral invertase* (*NIN*), *sucrose synthase* (*SS*), *fructokinase* (*FRK*), and *hexokinase* (*HK*) indicates the metabolization of sorbitol and sucrose [42]. In the early period of cell expansion, starch accumulation was higher, and it starts to decline in the later phase [11]. Transcript of SS genes in apple is correlated with the starch accumulation [39]. *Sorbitol dehydrogenase* (*SDH*) converts carbohydrate into fructose. Nine *SDH* genes were identified in apple fruit [43]. In young fruit, *MdSDH1* expression was higher than in mature fruit [42]. Other genes significantly upregulated during ripening stage are *isopentenyl pyrophosphate* (IPP) *isomerase*, *catalase* (CAT), *histone 2B* (*H2B*), and the *ripening-inhibitor* (RIN) *MADS-box gene* [39]. During the ripening process, a decrease of starch synthesis is vice versa with the sugar level [44]. Expression profiles of *sucrose-phosphatase phosphatase* (*SPP*) and *sucrose-phosphate synthase* (*SPS*) were active in the ripening stage [42], suggesting that these enzymes may be involved in starch degradation pathway. *Polygalacturonase 1* (*MdPG1*) and aminocyclopropane-1-carboxylate oxidase (*MdACO1*) were involved in the fruit softening and ethylene biosynthesis in apple, respectively [45]. Decrease in the expression of *PG1* alters the firmness, tensile strength, and water loss in apple M x domestica fruit [46]. Meanwhile, *MdFT1*, *MdACS1* (*1-aminocyclopropane-1-carboxylic acid synthase*), *MdACO1*, and *MdExp7* are regulating the fruit softening. Among them, *MdExp7* and *MdACO1* control firmness in apple [45]. Gene coding for MYB TF in apple, *MdMyb1*, increases the anthocyanin content and is responsible for the red skin color [47]. Acidity in apple is due to the malic acid, and *mama* recessive gene is responsible for low acidity [48].

In apple, fruit size, sugar content, and palatability are essential qualities determining its marketability. Knowledge of genes governing the fruit quality could be essential for screening better lines/genotypes for breeding.

3.2. Lignin Metabolism and Stone Cell Formation in Pear. Stone cell content is the main quality determinant of pear fruit. Deposition of lignin on the primary cell wall of
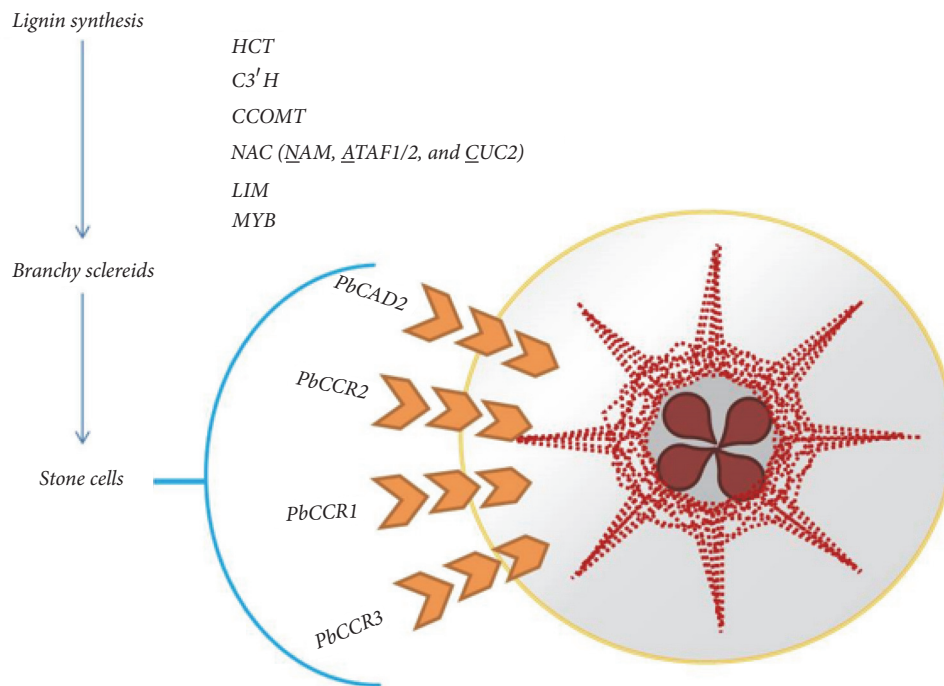
FIGURE 1: Simple heuristic representation of genes/transcription factors involved in lignin synthesis and stone cell formation in pear fruit. *Pb*, *Pyrus bretschneideri*; *hydroxycinnamoyl transferase*, *HCT*; *p*-coumaroyl-shikimate/quinate $3'$-hydroxylases, $C3'H$; *caffeoyl-CoA O-methyltransferase*, *CCOMT*; N̲AM, A̲TAF1/2, and C̲UC2, NAC; L̲in11/I̲sl1/M̲ec3, LIM; m̲yelob̲lastosis, MYB; *cinnamyl alcohol dehydrogenase*, *CAD;* and *cinnamoyl-CoA reductase*, *CCR*. Red colored dots represent the stone cells.

parenchyma cell followed by the secondary sedimentation on a sclerenchyma cell forms the stone cells. Majority of stone cells present in pear is branchy sclereids comprised lignin and cellulose. Lignins are synthesized by two ways, one starts with p-coumaric acid and second with phenylalanine precursor to cinnamic acid and then p-coumaric acid. Other forms of lignin monomers are caffeic acid, ferulic acid, 5-hydroxy-ferulic acid, and sinapinic acid [49–51]. Finally, monomers are polymerized to form lignin products. Monomers of lignin are categorized into three types, syringly lignin (S-lignin), guaiacyl lignin (G-lignin), and hydroxyphenyl lignin (H-lignin). From the gnome analysis, a total of 66 lignin synthesis-related gene families were characterized in *P. bretschneideri*. It signifies the high demand for lignin synthesis in pear [15]. In "Dangshan Su" pulp, milled wood lignin was identified as guaiacyl-syringyl-lignin. It was observed that "Dangshan Su" lignin possesses more guaiacyl units than the syringyl units [49]. *Hydroxycinnamoyl transferases* (*HCT*) play a significant role in the lignin synthesis [52]. Accumulation of G-lignin and S-lignin is interrelated with the *HCT* expression especially at early fruit developmental stage [15].

Cinnamoyl-CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD), belonging to medium-or-short-chain dehydrogenase/reductase, are key enzymes for lignin monomer synthesis [53]. Totally 31 *CCR*s and 26 *CAD*s genes were identified in *P. bretschneideri* "Dangshan Su". All members of *CCR* and *CAD* identified in *P. bretschneideri* are not involved in the lignin biosynthesis [54]. Among them,

PbCAD2, *PbCCR1*, *PbCCR2*, and *PbCCR3* were identified to participate in the lignin synthesis of stone cells [15]. NAC (N̲AM, A̲TAF1/2, and C̲UC2) and LIM (L̲in11/I̲sl1/M̲ec3) are an important TF influencing the lignin pathway [55, 56]. Most of the *CCR* and *CAD* members present in the pear possess SPL (squamosal promoter binding-like) light-responsive element on their upstream. Functions of *PbCCR* and *PbCAD* are related to the light signaling. Presence of MYB-binding AC *cis* elements in some promoter of the *PbCCR*s suggested that phenylpropanoid metabolism of lignin synthesis was regulated by MYB transcription factors. Similarly, TGACG-motif on some PbCCRs and all PbCAD's promoter regions revealed their involvement in the abscisic acid, jasmonic acid, and methyl jasmonic acid metabolism [15]. The pictorial illustration of genes/TFs required for the lignin synthesis as well as stone cell formation is mentioned in Figure 1.

There are many internal and external factors involved in the stone cell formation of pear. Identification of candidate genes of lignin biosynthesis and stone cell formation will be very much useful to improve the cultural practices for producing pear fruits with different palatable level of stone cells.

3.3. Fruit Aroma and Softness in Strawberry. Strawberry is widely appreciated for its delicate flavor, aroma, and nutritional value. Aroma of strawberry is due to esters, alcohols, aldehydes, and sulfur compounds. Hundreds of volatile esters have been correlated with strawberry ripening and aroma

[57]. Volatile esters are the major constituents of floral scent. Wild species such as *F. vesca* and *F. virginiana* have much stronger aroma than the cultivated types. Compared to the regular octoploid strawberry, unique phenolic compounds were found in *F. vesca* fruits, such as taxifolin 3-*O*-arabinoside and peonidin 3-*O*-malonylglucoside [58]. *Pinene synthase (PINS)* is primarily expressed in wild strawberry while insertional mutation reduced its expression in cultivated species. *F. vesca* contains high amounts of ethyl-acetate and lower methyl-butyrate, ethyl-butyrate, and furanone levels. *F. nilgerrensis* possesses higher ethyl-acetate and furanone but lower methyl-butyrate and ethyl-butyrate. Hybrids between *F. vesca* and *F. ananassa* have intermediate contents of fragrance and aroma related compounds while crosses between *F. nilgerrensis* and *F. ananassa* more closely resemble *F. nilgerrensis* [17]. Volatile compounds found to be responsible for general strawberry smell are 2, 5-dimethyl-4-hydroxy-3(2H)-furanone, linalool, and ethyl hexanoate. Nevertheless, ethyl butanoate, methyl butanoate, γ-decalactone, and 2-heptanone are represented as cultivar specific aroma compounds [59]. O-methyltransferase of strawberry (*FaOMT*) is vital for the biosynthesis of vanillin and furaneol [60]. *Alcohol acyltransferase (AATs)* in strawberry (*SAAT*) is involved in the last step of volatile esters synthesis and vital for flavor biogenesis in ripening fruit. *SAAT* catalyzes esterification of an acyl moiety from actyl-CoA to alcohol [61]. Strawberry *quinone oxidoreductase (FaQR)* is required for the biosynthesis of furaneol. Furaneol and its methoxy derivative (methoxyfuraneol and mesifuran) are catalyzed by OMT. All three furaneol compounds are highly accumulated during fruit ripening stage [62]. Though two types of *pyruvate decarboxylase (PDC)* were identified in strawberry, only *FaPDC1* was induced during fruit ripening [63].

Strawberry is highly perishable even with controlled atmospheric storage. Higher proportion of fruit last occurred due to its softness and sensitivity to fungal disease. Red colored strawberry showed the higher level of anthocyanin-related transcripts [64]. *FcMYB1* could regulate branching-point of the anthocyanin/proanthocyanidin biosynthesis. *FaWRKY1* mediate defense response and *FaPE1*, encoded for pectin methyl esterase, are conferred at least with a partial resistance of ripened fruit against Botrytis cinerea [65, 66]. *Polygalacturonase 1 of F x ananassa (FaPG1)* is critical for fruit softening [67]. In strawberry fruits *beta-D-glucosyltransferase (FaGT)* correlated with the relevant phenylpropanoid glucosides [68]. *D-xylose reductase (FaXyl1)* and *beta-xylosidase* activity were higher in "Toyonaka" (soft) than in the "Camarosa" (firm) showing the correlation between *FaXyl1* expression and fruit softening [69]. Fruit-specific *rhamnogalacturonate lyase 1 (FaRGLyase)* is involved in the firmness and postharvest life [70]. A lesser activity of beta-galactosidase (*βGal*) and *βXyl* activity were correlated with decreased fruit firmness in *F. chiloensis* and *F. $\times$ ananassa*, respectively [71]. Expression of *FaCCR* is higher in soft fruit cultivar (Gorella) whereas *FaCAD* is higher in firm fruit cultivar (Holiday) [72]. Expression of five *expansin* genes (*FaEXP1, FaEXP2, FaEXP4, FaEXP5,* and *FaEXP6*) was studied in cultivars with different firmness "Selva" (hard), "Camarosa" (medium), and "Toyonaka" (soft). Higher level

of *FaEXP1, FaEXP2,* and *FaEXP5* expression was found in fruit with less firmness ("Toyonaka") than the other two cultivars ("Selva" and "Camarosa"). Fruit firmness is identified to be associated with pectate lyase (*FaPel1*) identified. Expansin activity was characterized by cell wall modification [73]. Polysaccharides were modified by five different genes such as *FasPG, FaPG*-like, *FaPel1, FaPel2,* and *FaEXP2* [74]. *Sorbitol dehydrogenase (FaSDH)* and *sorbitol-6-phosphate dehydrogenase (FaS6PDH)* genes are involved with the sorbitol synthesis in leaves, fruits, and shoot tips [75]. *SEPALLATA (SEP)4-like* gene *FaMADS9* is responsible for the fruit ripening [76].

Apart from the aesthetic and taste, mechanisms of flowering and its response to the light signaling in strawberry need to be studied in detail. Cultivars with continuous flowering and growing under minimal light energy are beneficial for the strawberry growers as most of the commercial cultivation is carried out in the controlled greenhouse.

3.4. Early Blooming and Fruit Ripening in Prunus. Prunus is the first plant to bloom in later winter/early spring. So, it is the best model plant to study early flowering as well as chilling tolerance. Dehydrins are known as 2 or D-11 family late-embryogenesis-abundant (LEA) proteins. They play a vital role in plant growth and cold tolerance [77]. In *P. mume*, 30 LEA genes were characterized and classified into eight groups LEA1, LEA2, LEA3, LEA4, LEA5, PvLEA18, dehydrin, and seed maturation protein. Out of 30 identified genes, 22 were expressed in flowers, and 19 were induced by *abscisic acid* (ABA) treatments [78]. Molecular cloning of *PmLEA8, PmLEA10, PmLEA19, PmLEA20,* and *PmLEA29* showed that, except *PmLEA8*, all other genes enhanced the freezing-tolerance. Interestingly, among all cold-resistant *LEA* gene members studied, only *PmLEA19* were upregulated four times when the branches of *P. mume* were exposed to 4°C [79]. Downregulation in *P. mume dormancy associated MADS (PmDAM) 4, PmDAM5,* and *PmDAM6* expression releases the endodormancy [80]. Among the six *DAM* genes, except *PmDAM3*, all other genes are responsive to the photoperiod and seasonal (cold) responses [81]. In *P. persica* from six identified *DAM* genes, *PpDAM5* and *PpDAM6* were characterized to be involved in the lateral bud dormancy breakage [82]. *DAM5* and *DAM6* were identified as homologous to *Short Vegetative Phase* (SVP)/*AGL 24* in *A. thaliana*. Both *SVP* and *AGL-24* are required for floral meristem identity [83]. AGL24 is well known for promoting early flowering and floral transition in plants [58]. Transcriptome analysis between cold sensitive ("Morettini") and cold tolerant ("Royal Glory") cultivars in *P. persica* showed that *β-D-xylosidase (BXL)* and pathogen-related protein 4b *(PR-4B)* were significantly expressed only in resistant variety [84]. Other candidate genes identified as required to control flowering time are *suppressor of phyA (SPA), COP1 interacting protein8 (CIP8), phytochrome A (phyA),* and *phytochrome interacting factor 3 (PIF3)* [85]. Figure 2 demonstrates the important key genes involved in cold tolerance, early blooming, and flowering time control of *P. mume*. Higher number of aesthetic properties related genes such as *benzyl alcohol acetyltransferase (BEAT)* (34) are identified in *P. mume*. Only, 16 in *Malus* x
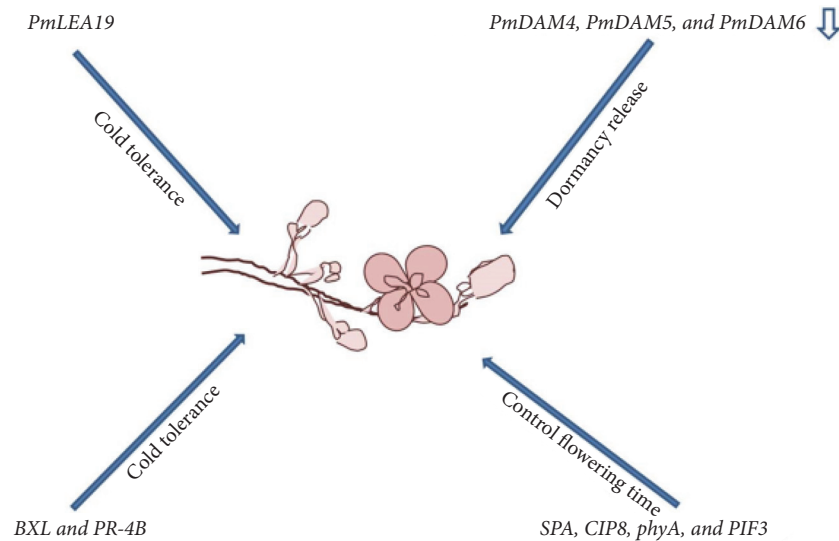
PmLEA19

PmDAM4, PmDAM5, and PmDAM6 ⬇

Cold tolerance

Dormancy release

Cold tolerance

Control flowering time

BXL and PR-4B

SPA, CIP8, phyA, and PIF3

FIGURE 2: Factors involved in the early blooming of Chinese plum/Japanese apricot. *Prunus meme, Pm*; *late-embryogenesis-abundant, LEA*; *dormancy associated MADS, DAM*; *β-D-xylosidase, BXL*; *pathogen-related protein 4b, PR-4B*; *suppressor of phyA, SPA*; *COP1 interacting protein8, CIP8*; *phytochrome A, phyA*; *and phytochrome interacting factor3, PIF3*. The lower arrow represents downregulation.

domestica, 14 in *F. vesca*, 4 in *Vitis vinifera*, 17 in *P. trichocarpa*, and 3 in *A. thaliana* of *BEAT* genes were identified. Therefore, in *P. mume BEAT* genes are considered as key factor to determine its exclusive floral fragrance [23].

In Prunus, *UDP-glucose-flavonoid-3-O-glucosyltransferase (UFGT)* expression was higher during the initial period and it is reduced on the developmental process. During the ripening process, *MYB10, MYB123*, and basic-helix-loop-helix *(bHLH3)* were upregulated whereas *MYB16* and *MYB111* were downregulated. Higher anthocyanin and Proanthocyanidin levels were correlated with the *MYB10* and *MYBPA1*, respectively. Stimulation of TFs is responsive for the development and external stimuli [86]. Gene encoding *ethylene-responsive transcription factor (ERF)4* is necessary for the fruit maturity [87]. Though 74 *EFR* genes were predicted in the peach genome, only one copy of *ERF4* has existed. Therefore, *ERF4* is vital to control the fruit maturity and ripening in peach [85]. An initial stage of fruit has higher aldehyde and alcohol production whereas later stages have lesser content which is correlated with the ester production. Abundance of alcohol dehydrogenase *(ADH)* and *lipoxygenase (LOX)* gene is constant in the fruit development stages. Expression of *AAT* was sharply increased in the later stage of harvest [88]. Rapid softening of fruits was related to the ethylene production in *P. persica*. It is correlated with expression of *Pp*ACS1 (*1-aminocyclopropane-1-carboxylic acid synthase*) [89]. Ripened sweet cherry *P. avium* has unique fragrance. In the sweet cherry ("Hongdeng", "Hongyan", and "Rainier") 97 volatile compounds were identified. Alcohols and terpenes were the predominant components of bound volatiles. Benzyl alcohol, geraniol, and 2-phenylethanol are the major bound volatile constituents. Free volatile compounds majorly present in sweet cherry are hexanal, 2-hexenal, 2-hexen-1-ol, benzyl alcohol, and benzaldehyde. Free volatiles are responsible for floral aroma and bound volatiles involved in fruit freshness. Depending on the level of free and bound volatiles, aroma and glycosidically bound compounds aroma and fruit firmness were varied between the cultivars of sweet cherry [90].

In *Prunus*, apricot, peach, sweet cherry, and sour cherry are widely used for human consumption. Comparative genomics study between the species offer the candidate gene to produce hybrids with more preferable qualities.

3.5. Blooming and Scent Pathways in Rose. Continuous flowering (CF)/recurrent blooming (RB) genotypes flowers in all favorable seasons, whereas once-flowering (OF) genotypes only flowers in spring. Recurrent blooming is an important trait required by breeders. *R x hybrida* "La France" was the first hybrid combined with the growth vigor of European species and recurrent blooming of Chinese species. It has the complex genetic pool combination of three ancestral genotypes such as Cinnamomeae, Synstylae, and Chinenses. Insertion of TE in the *TFL1* encodes gene *Ksncopia* (*KSN*) was found as responsible for the recurrent blooming of "La France" [30]. Previously, Horibe et al. (2013) reported that *KSN* gene regulates the CF behavior of *R. rugosa* [91]. Wang et al. (2012) studied recurrent flowering character and the expression patterns of TFL1 homologs in *R. multiflora, R. rugosa, R. chinensis*, and other species/cultivars. Among the three orthologs, RTFL1c was highly expressed at all four flowering stages in *R. multiflora* and *R. rugosa* (nonrecurrent flowering species) and barely detected in *R. chinensis* (a recurrent flowering species) at any stage. Therefore, it can be considered that lower expression of *RTFL1c* is required for recurrent flowering of roses [92]. Iwata et al. (2012) elucidate that occurrence of TE insertion and point mutation in the

TFL1 ortholog on rose and strawberry correlate with recurrent blooming [93]. Higher expression of TFL1 in seasonal flower is associated with the repression of *LEAFY* (*LFY*) and *activating protein-1* (*AP1*), a downstream gene of FT [94]. Expression of *flowering locus T* of rose (*RoFT*) was progressively increased after floral bud formation. CONSTANS TF induces the *FT*, and, upon induction, *FT* was suggested to move from leaves to shoot apical meristem (SAM) via phloem [95–97]. Additionally, *suppressor of Ty* (*SPT*) and *delay of germination* (*DOG)1* are other important candidates determining recurrent blooming in roses [31].

Rose scent is a complex trait involved with hundreds of volatile molecules. Rose floral scent contains phenolic derivatives, terpenoids, and fatty acid derivatives. Several genes have been identified to be related to rose scent production. Floral scent of roses contains higher germacrene D synthase. Cyanidin and germacrene D were identified to be involved in the color and scent pathways. Sesquiterpene synthase catalyzes the production of germacrene D [98]. *Phenylacetaldehyde synthase* (*PAAS*) and *phenylacetaldehyde reductase* (*PAR*) are responsible for the synthesis of 2-phenylethyl alcohol, a typical rose scent compound [99]. Anthocyanin synthesis on rose was linked with the pigmentation and volatile (scent) compounds related pathways. Anthocyanin and volatile compound have been generated by enabling the formation of MYB-bHLH-WD40 protein complex. *Orcinol-o-methyltransferase* (*RhOOMT) 1* and *2* is responsible for synthesis of 2OMT. Alcohol acyltransferase of *R. hybrida* (*RhAAT1*) gene converts alcohol geraniol into geranyl acetate [62]. Major scent compound of European roses is 2-phenylethanol and monoterpenes [100, 101]. *RhOOMTs* catalyze the orcinol to synthesize two important volatiles such as 3,5-dimethoxy toluene (DMT) and 1,3,5-metoxy benzene (TMB) biosynthesis in *R. hybrida* [62]. From the study of interhybrid cultivars of rose, DMT was concluded to come from Chinese rose, as ancient European roses such as *R. damascena* and *R. gallica* do not produce DMT [102]. Geraniol, a hydrolyzed product and its downstream monoterpene volatile metabolites, are responsible for the aroma of rose petals. *Nudix hydrolase* (*NUDX*)1 is involved in synthesis of geraniol and other geraniol-derived monoterpenes [103]. Still, there are many pathways rose scent need to be elucidated.

Genome released in rose will be helpful for decoding the metabolic networks of scent pathway, floral transition, and flowering pattern. Therefore, irrespective of complex and cumbersome heterozygous nature, interspecific hybridization can be accelerated to produce hybrid with valuable traits in rose.

4. Conclusions

Complete genome information of plant reduces effort and time required for conventional MAS approach. Identification and characterization of genes controlling important traits and tagging molecular markers for introgression to produce a new variety are feasible with available genome information. Along with abiotic and biotic stress resistance, several fruit quality traits can be improved with genomics-based studies. Fruit firmness is one of the desirable quality traits. It depends on the postharvest shelf life, cell turgor pressure, and intrinsic characteristics of the cell wall. Modification and turnover of the primary cell wall are required for both size and softness of fruits. New varieties/cultivars with small/larger size, good-flavored fruits, attractive color, sugar and acid levels, reduced juvenile phase, massive and constant yields, reduced susceptibility to fruit cracking, self-compatibility, and improved resistance or tolerance to disease are now feasible with the completion of the whole genome sequence.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] D. Potter, T. Eriksson, R. C. Evans et al., "Phylogeny and classification of Rosaceae," *Plant Systematics and Evolution*, vol. 266, no. 1-2, pp. 5–43, 2007.

[2] V. Shulaev, S. S. Korban, B. Sosinski et al., "Multiple models for Rosaceae genomics," *Plant Physiology*, vol. 147, no. 3, pp. 985–1003, 2008.

[3] D. Potter, F. Gao, P. E. Bortiri, S.-H. Oh, and S. Baggett, "Phylogenetic relationships in Rosaceae inferred from chloroplast matK and trnL-trnF nucleotide sequence data," *Plant Systematics and Evolution*, vol. 231, no. 1–4, pp. 77–89, 2002.

[4] C. S. Pareek, R. Smoczynski, and A. Tretyn, "Sequencing technologies and genome sequencing," *Journal of Applied Genetics*, vol. 52, no. 4, pp. 413–435, 2011.

[5] A. G. Day-Williams and E. Zeggini, "The effect of next-generation sequencing technology on complex trait research," *European Journal of Clinical Investigation*, vol. 41, no. 5, pp. 561–567, 2011.

[6] N. Daccord, J.-M. Celton, G. Linsmith et al., "High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development," *Nature Genetics*, vol. 49, no. 7, pp. 1099–1106, 2017.

[7] J. Stapley, J. Reger, P. G. D. Feulner et al., "Adaptation genomics: the next generation," *Trends in Ecology and Evolution*, vol. 25, no. 12, pp. 705–712, 2010.

[8] M. Kellerhals, "Introduction to Apple (*Malus* x *domestica*)," in *Genetics and Genomics of Rosaceae*, S. E. G. Kevin and M. Folta, Eds., pp. 73–84, 2009.

[9] A. Cornille, P. Gladieux, M. J. M. Smulders et al., "New insight into the history of domesticated apple: Secondary contribution of the European wild apple to the genome of cultivated varieties," *PLoS Genetics*, vol. 8, article e1002703, 2012.

[10] A. Cornille, T. Giraud, M. J. M. Smulders, I. Roldán-Ruiz, and P. Gladieux, "The domestication and evolutionary ecology of apples," *Trends in Genetics*, vol. 30, no. 2, pp. 57–65, 2014.

[11] R. Velasco, A. Zharkikh, and J. Affourtit, "The genome of the domesticated apple (*Malus* × *domestica* Borkh.)," *Nature Genetics*, vol. 42, no. 10, pp. 833–839, 2010.

[12] X. Li, L. Kui, J. Zhang et al., "Improved hybrid de novo genome assembly of domesticated apple (Malus x domestica)," *GigaScience*, vol. 5, article 35, 2016.

[13] G. Rubtsov, "Geographical distribution of the genus Pyrus and trends and factors in its evolution," *The American Naturalist*, vol. 78, pp. 358–366, 1944.

[14] T. Saito, "Advances in Japanese pear breeding in Japan," *Breeding Science*, vol. 66, no. 1, pp. 46–59, 2016.

[15] J. Wu, Z. Wang, Z. Shi et al., "The genome of the pear (Pyrus bretschneideri Rehd.)," *Genome Research*, vol. 23, no. 2, pp. 396–408, 2013.

[16] D. Chagné, R. N. Crowhurst, M. Pindo et al., "The draft genome sequence of European pear (Pyrus communis L. 'Bartlett')," *PLoS ONE*, vol. 9, article e92644, 2014.

[17] K. E. Hummer and J. Hancock, "Strawberry genomics: botanical history, cultivation, traditional breeding, and new technologies," in *Genetics and Genomics of Rosaceae*, pp. 413–435, Springer, 2009.

[18] O. Darwish, J. P. Slovin, C. Kang et al., "SGR: an online genomic resource for the woodland strawberry," *BMC Plant Biology*, vol. 13, article 223, 2013.

[19] V. Shulaev, D. J. Sargent, R. N. Crowhurst et al., "The genome of woodland strawberry (Fragaria vesca)," *Nature Genetics*, vol. 43, no. 2, pp. 109–116, 2011.

[20] H. Hirakawa, K. Shirasawa, S. Kosugi et al., "Dissection of the octoploid strawberry genome by deep sequencing of the genomes of fragaria species," *DNA Research*, vol. 21, no. 2, pp. 169–181, 2014.

[21] J. Shi, J. Gong, J. Liu, X. Wu, and Y. Zhang, "Antioxidant capacity of extract from edible flowers of Prunus mume in China and its active components," *LWT- Food Science and Technology*, vol. 42, no. 2, pp. 477–482, 2009.

[22] S. Fan, D. G. Bielenberg, T. N. Zhebentyayeva et al., "Mapping quantitative trait loci associated with chilling requirement, heat requirement and bloom date in peach (Prunus persica)," *New Phytologist*, vol. 185, no. 4, pp. 917–930, 2010.

[23] Q. Zhang, W. Chen, L. Sun, Z. Danyang, and W. Zeng, "The genome of Prunus mume," *Nature*, vol. 3, article 1318, 2012.

[24] K. G. D. Bielenberg and J. X. Chaparro, "An introduction to peach (Prunus persica)," in *Genetics and Genomics of Rosaceae*, S. E. G. Kevin and M. Folta, Eds., pp. 223–234, 2009.

[25] I. Verde, A. G. Abbott, S. Scalabrin et al., "The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution," *Nature Genetics*, vol. 45, no. 5, pp. 487–494, 2013.

[26] I. Verde, J. Jenkins, L. Dondini et al., "The Peach v2.0 release: high-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity," *BMC Genomics*, vol. 18, article 225, 2017.

[27] J. C. E. Dirlewanger and A. F. Iezzoni, "Ana wiinsch sweet and sour cherries: linkage maps," in *Genetics and Genomics of Rosaceae*, S. E. G. Kevin and M. Folta, Eds., pp. 291–314, 2009.

[28] K. Shirasawa, K. Isuzugawa, M. Ikenaga et al., "The genome sequence of sweet cherry (Prunus avium) for use in genomics-assisted breeding," *DNA Research*, vol. 24, no. 5, pp. 499–508, 2017.

[29] H. Nybom, "Introduction to rosa," in *Genetics and Genomics of Rosaceae*, pp. 339–351, Springer, 2009.

[30] O. Raymond, J. Gouzy, J. Just et al., "The Rosa genome provides new insights into the domestication of modern roses," *Nature Genetics*, vol. 50, pp. 772–777, 2018.

[31] L. H. Saint-Oyant, T. Ruttink, L. Hamama et al., "A high-quality genome sequence of Rosa chinensis to elucidate ornamental traits," *Nature plants*, article 1, 2018.

[32] C. Hurst, "Notes on the origin and evolution of our garden roses," *Journal of the Horticultural Society*, vol. 66, pp. 282–289, 1941.

[33] D. Terefe-Ayana, A. Yasmin, T. L. Le et al., "Mining disease-resistance genes in roses: functional and molecular characterization of the rdr1 locus," *Frontiers in Plant Science*, vol. 2, article 35, 2011.

[34] N. Nakamura, H. Hirakawa, S. Sato et al., "Genome structure of Rosa multiflora, a wild ancestor of cultivated roses," *DNA Research*, vol. 25, no. 2, pp. 113–121, 2017.

[35] N. Mimida, S.-I. Kidou, and N. Kotoda, "Constitutive expression of two apple (Malus x domestica Borkh.) homolog genes of LIKE HETEROCHROMATIN PROTEIN1 affects flowering time and whole-plant growth in transgenic Arabidopsis," *Molecular Genetics and Genomics*, vol. 278, no. 3, pp. 295–305, 2007.

[36] N. Kotoda, H. Hayashi, M. Suzuki et al., "Molecular characterization of flowering LOCUS t-like genes of apple (malus × domestica borkh.)," *Plant & Cell Physiology (PCP)*, vol. 51, no. 4, pp. 561–575, 2010.

[37] C. Spruck, H. Strohmaier, M. Watson et al., "A CDK-independent function of mammalian Cks1: targeting of SCFSkp2 to the CDK inhibitor p27Kip1," *Molecular Cell*, vol. 7, no. 3, pp. 639–650, 2001.

[38] C. Ferrándiz, S. Pelaz, and M. F. Yanofsky, "Control of carpel and fruit development in Arabidopsis," *Annual Review of Biochemistry*, vol. 68, pp. 321–354, 1999.

[39] B. J. Janssen, K. Thodey, R. J. Schaffer et al., "Global gene expression analysis of apple fruit development from the floral bud to ripe fruit," *BMC Plant Biology*, vol. 8, article 16, 2008.

[40] V. Cevik, C. D. Ryder, A. Popovich, K. Manning, G. J. King, and G. B. Seymour, "A FRUITFULL-like gene is associated with genetic variation for fruit flesh firmness in apple (Malus domestica Borkh.)," *Tree Genetics and Genomes*, vol. 6, no. 2, pp. 271–279, 2010.

[41] F. Devoghalaere, T. Doucen, B. Guitton et al., "A genomics approach to understanding the role of auxin in apple (Malus x domestica) fruit size control," *BMC Plant Biology*, vol. 12, article 7, 2012.

[42] M. Li, F. Feng, and L. Cheng, "Expression patterns of genes involved in sugar metabolism and accumulation during apple fruit development," *PLoS ONE*, vol. 7, article e33055, 2012.

[43] M. Nosarzewski and D. D. Archbold, "Tissue-specific expression of sorbitol dehydrogenase in apple fruit during early development," *Journal of Experimental Botany*, vol. 58, no. 7, pp. 1863–1872, 2007.

[44] P. Brookfield, P. Murphy, R. Harker, and E. MacRae, "Starch degradation and starch pattern indices; interpretation and relationship to maturity," *Postharvest Biology and Technology*, vol. 11, no. 1, pp. 23–30, 1997.

[45] F. Costa, C. P. Peace, S. Stella et al., "QTL dynamics for fruit firmness and softening around an ethylene-dependent polygalacturonase gene in apple (MalusXdomestica Borkh.)," *Journal of Experimental Botany*, vol. 61, no. 11, pp. 3029–3039, 2010.

[46] R. G. Atkinson, P. W. Sutherland, S. L. Johnston et al., "Down-regulation of POLYGALACTURONASE1 alters firmness, tensile strength and water loss in apple (Malus x domestica) fruit," *BMC Plant Biology*, vol. 12, article 129, 2012.

[47] A. M. Takos, F. W. Jaffé, S. R. Jacob, J. Bogs, S. P. Robinson, and A. R. Walker, "Light-induced expression of a *MYB* gene regulates anthocyanin biosynthesis in red apples," *Plant Physiology*, vol. 142, no. 3, pp. 1216–1232, 2006.

[48] C. Maliepaard, F. H. Alston, G. Van Arkel et al., "Aligning male and female linkage maps of apple (Malus pumila Mill.) using

multi-allelic markers," *Theoretical and Applied Genetics*, vol. 97, no. 1-2, pp. 60–73, 1998.

[49] Y. Cai, G. Li, J. Nie et al., "Study of the structure and biosynthetic pathway of lignin in stone cells of pear," *Scientia Horticulturae*, vol. 125, no. 3, pp. 374–379, 2010.

[50] H. Meyermans, K. Morreel, C. Lapierre et al., "Modifications in lignin and accumulation of phenolic glucosides in poplar xylem upon down-regulation of caffeoyl-coenzyme A O-methyltransferase, an enzyme involved in lignin biosynthesis," *The Journal of Biological Chemistry*, vol. 275, no. 47, pp. 36899–36909, 2000.

[51] J. M. Humphreys and C. Chapple, "Rewriting the lignin roadmap," *Current Opinion in Plant Biology*, vol. 5, no. 3, pp. 224–229, 2002.

[52] L. Hoffmann, S. Besseau, P. Geoffroy et al., "Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis," *The Plant Cell*, vol. 16, no. 6, pp. 1446–1465, 2004.

[53] X. Cheng, M. Li, D. Li et al., "Characterization and analysis of CCR and CAD gene families at the whole-genome level for lignin synthesis of stone cells in pear (Pyrus bretschneideri) fruit," *Biology Open*, vol. 6, no. 11, pp. 1602–1613, 2017.

[54] J. Thévenin, B. Pollet, B. Letarnec et al., "The simultaneous repression of CCR and CAD, two enzymes of the lignin biosynthetic pathway, results in sterility and dwarfism in Arabidopsis thaliana," *Molecular Plant*, vol. 4, no. 1, pp. 70–82, 2011.

[55] R. Zhong, T. Demura, and Z.-H. Ye, "SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of Arabidopsis," *The Plant Cell*, vol. 18, no. 11, pp. 3158–3170, 2006.

[56] A. Kawaoka, P. Kaothien, K. Yoshida, S. Endo, K. Yamada, and H. Ebinuma, "Functional analysis of tobacco LIM protein Ntlim1 involved in lignin biosynthesis," *The Plant Journal*, vol. 22, no. 4, pp. 289–301, 2000.

[57] D. J. Sargent, T. M. Davis, and D. W. Simpson, "Strawberry (Fragaria spp.) structural genomics," in *Genetics and genomics of Rosaceae*, pp. 437–456, Springer, 2009.

[58] J. Sun, X. Liu, T. Yang, J. Slovin, and P. Chen, "Profiling polyphenols of two diploid strawberry (Fragaria vesca) inbred lines using UHPLC-HRMSn," *Food Chemistry*, vol. 146, pp. 289–298, 2014.

[59] M. Larsen, L. Poll, and C. E. Olsen, "Evaluation of the aroma composition of some strawberry (Fragaria ananassa Duch) cultivars by use of odour threshold values," *Zeitschrift für Lebensmittel-Untersuchung und -Forschung*, vol. 195, no. 6, pp. 536–539, 1992.

[60] M. Wein, N. Lavid, S. Lunkenbein, E. Lewinsohn, W. Schwab, and R. Kaldenhoff, "Isolation, cloning and expression of a multifunctional O-methyltransferase capable of forming 2,5-dimethyl-4-methoxy-3(2H)-furanone, one of the key aroma compounds in strawberry fruits," *The Plant Journal*, vol. 31, no. 6, pp. 755–765, 2002.

[61] A. Aharoni, L. C. P. Keizer, H. J. Bouwmeester et al., "Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays," *The Plant Cell*, vol. 12, no. 5, pp. 647–661, 2000.

[62] N. Lavid, J. Wang, M. Shalit et al., "O-methyltransferases involved in the biosynthesis of volatile phenolic derivatives in rose petals," *Plant Physiology*, vol. 129, no. 4, pp. 1899–1907, 2002.

[63] E. Moyano, S. Encinas-Villarejo, J. A. López-Ráez et al., "Comparative study between two strawberry pyruvate decarboxylase genes along fruit development and ripening, post-harvest and stress conditions," *Journal of Plant Sciences*, vol. 166, no. 4, pp. 835–845, 2004.

[64] A. Salvatierra, P. Pimentel, M. A. Moya-Leon, P. D. S. Caligari, and R. Herrera, "Comparison of transcriptional profiles of flavonoid genes and anthocyanin contents during fruit development of two botanical forms of Fragaria chiloensis ssp. chiloensis," *Phytochemistry*, vol. 71, no. 16, pp. 1839–1847, 2010.

[65] S. Encinas-Villarejo, A. M. Maldonado, F. Amil-Ruiz et al., "Evidence for a positive regulatory role of strawberry (Fragaria × ananassa) Fa WRKY1 and Arabidopsis At WRKY75 proteins in resistance," *Journal of Experimental Botany*, vol. 60, no. 11, pp. 3043–3065, 2009.

[66] S. Osorio, A. Bombarely, P. Giavalisco et al., "Demethylation of oligogalacturonides by FaPE1 in the fruits of the wild strawberry Fragaria vesca triggers metabolic and transcriptional changes associated with defence and development of the fruit," *Journal of Experimental Botany*, vol. 62, no. 8, pp. 2855–2873, 2011.

[67] M. A. Quesada, R. Blanco-Portales, S. Posé et al., "Antisense down-regulation of the FaPG1 gene reveals an unexpected central role for polygalacturonase in strawberry fruit softening," *Plant Physiology*, vol. 150, no. 2, pp. 1022–1032, 2009.

[68] A. Aharoni, C. H. R. De Vos, M. Wein et al., "The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco," *The Plant Journal*, vol. 28, no. 3, pp. 319–332, 2001.

[69] C. A. Bustamante, H. G. Rosli, M. C. Añón, P. M. Civello, and G. A. Martínez, "β-Xylosidase in strawberry fruit: isolation of a full-length gene and analysis of its expression and enzymatic activity in cultivars with contrasting firmness," *Journal of Plant Sciences*, vol. 171, no. 4, pp. 497–504, 2006.

[70] F. J. Molina-Hidalgo, A. R. Franco, C. Villatoro et al., "The strawberry (Fragaria×ananassa) fruit-specific rhamnogalacturonate lyase 1 (FaRGLyase1) gene encodes an enzyme involved in the degradation of cell-wall middle lamellae," *Journal of Experimental Botany*, vol. 64, no. 6, pp. 1471–1483, 2013.

[71] C. R. Figueroa, H. G. Rosli, P. M. Civello, G. A. Martínez, R. Herrera, and M. A. Moya-León, "Changes in cell wall polysaccharides and cell wall degrading enzymes during ripening of Fragaria chiloensis and Fragaria ×ananassa fruits," *Scientia Horticulturae*, vol. 124, no. 4, pp. 454–462, 2010.

[72] E. M. J. Salentijn, A. Aharoni, J. G. Schaart, M. J. Boone, and F. A. Krens, "Differential gene expression analysis of strawberry cultivars that differ in fruit-firmness," *Physiologia Plantarum*, vol. 118, no. 4, pp. 571–578, 2003.

[73] M. C. Dotto, G. A. Martínez, and P. M. Civello, "Expression of expansin genes in strawberry varieties with contrasting fruit firmness," *Plant Physiology and Biochemistry*, vol. 44, no. 5-6, pp. 301–307, 2006.

[74] W. Schwab, J. G. Schaart, and C. Rosati, "Functional molecular biology research in Fragaria," in *Genetics and Genomics of Rosaceae*, pp. 457–486, Springer, 2009.

[75] G. Cumplido Laso, "Functional characterization os satrawberry (Fragaria x Ananassa) fruit-specific and ripening-related genes involved in aroma and anthochyanins biosynthesis," PhD Thesis, Universidad de Córdoba, 2012.

[76] G. B. Seymour, C. D. Ryder, V. Cevik et al., "A SEPALLATA gene is involved in the development and ripening of strawberry (Fragaria×ananassa Duch.) fruit, a non-climacteric tissue," *Journal of Experimental Botany*, vol. 62, no. 3, pp. 1179–1188, 2011.

[77] A. Banerjee and A. Roychoudhury, "Group II late embryogenesis abundant (LEA) proteins: structural and functional aspects in plant abiotic stress," *Plant Growth Regulation*, vol. 79, no. 1, pp. 1–17, 2016.

[78] D. Du, Q. Zhang, T. Cheng, H. Pan, W. Yang, and L. Sun, "Genome-wide identification and analysis of late embryogenesis abundant (LEA) genes in Prunus mume," *Molecular Biology Reports*, vol. 40, no. 2, pp. 1937–1946, 2013.

[79] F. Bao, D. Du, Y. An et al., "Overexpression of Prunus mume dehydrin genes in tobacco enhances tolerance to cold and drought," *Frontiers in Plant Science*, vol. 8, article 151, 2017.

[80] R. Sasaki, H. Yamane, T. Ooka et al., "Functional and expressional analyses of PmDAM genes associated with endodormancy in Japanese apricot (Prunus mume)," *Plant Physiology*, vol. 157, no. 1, pp. 485–497, 2011.

[81] Z. Li, G. L. Reighard, A. G. Abbott, and D. G. Bielenberg, "Dormancy-associated MADS genes from the EVG locus of peach [Prunus persica (L.) Batsch] have distinct seasonal and photoperiodic expression patterns," *Journal of Experimental Botany*, vol. 60, pp. 3521–3530, 2009.

[82] H. Yamane, T. Ooka, H. Jotatsu, R. Sasaki, and R. Tao, "Expression analysis of PpDAM5 and PpDAM6 during flower bud development in peach (Prunus persica)," *Scientia Horticulturae*, vol. 129, no. 4, pp. 844–848, 2011.

[83] V. Gregis, A. Sessa, L. Colombo, and M. M. Kater, "AGAMOUS-LIKE24 and SHORT VEGETATIVE PHASE determine floral meristem identity in Arabidopsis," *The Plant Journal*, vol. 56, no. 6, pp. 891–902, 2008.

[84] V. Falara, G. A. Manganaris, F. Ziliotto et al., "A β-D-xylosidase and a PR-4B precursor identified as genes accounting for differences in peach cold storage tolerance," *Functional & Integrative Genomics*, vol. 11, no. 2, pp. 357–368, 2011.

[85] E. Dirlewanger, J. Quero-García, L. Le Dantec et al., "Comparison of the genetic determinism of two key phenological traits, flowering and maturity dates, in three Prunus species: Peach, apricot and sweet cherry," *Heredity*, vol. 109, no. 5, pp. 280–292, 2012.

[86] D. Ravaglia, R. V. Espley, R. A. Henry-Kirk et al., "Transcriptional regulation of flavonoid biosynthesis in nectarine (Prunus persica) by a set of R2R3 MYB transcription factors," *BMC Plant Biology*, vol. 13, article 68, 2013.

[87] T. Nakano, K. Suzuki, T. Fujimura, and H. Shinshi, "Genome-wide analysis of the ERF gene family in arabidopsis and rice," *Plant Physiology*, vol. 140, no. 2, pp. 411–432, 2006.

[88] M. González-Agüero, S. Troncoso, O. Gudenschwager, R. Campos-Vargas, M. A. Moya-León, and B. G. Defilippi, "Differential expression levels of aroma-related genes during ripening of apricot (Prunus armeniaca L.)," *Plant Physiology and Biochemistry*, vol. 47, no. 5, pp. 435–440, 2009.

[89] M. Tatsuki, N. Nakajima, H. Fujii et al., "Increased levels of IAA are required for system 2 ethylene synthesis causing fruit softening in peach (Prunus persica L. Batsch)," *Journal of Experimental Botany*, vol. 64, pp. 1049–1059, 2013.

[90] Y.-Q. Wen, F. He, B.-Q. Zhu et al., "Free and glycosidically bound aroma compounds in cherry (Prunus avium L.)," *Food Chemistry*, vol. 152, pp. 29–36, 2014.

[91] T. Horibe, K. Yamada, S. Otagaki et al., "Molecular genetic studies on continuous-flowering roses that do not originate from Rosa chinensis," in *Proceedings of the 6th International Symposium on Rose Research and Cultivation*, vol. 1064, pp. 185–192, 2013.

[92] L.-N. Wang, Y.-F. Liu, Y.-M. Zhang, R.-X. Fang, and Q.-L. Liu, "The expression level of Rosa Terminal Flower 1 (RTFL1) is related with recurrent flowering in roses," *Molecular Biology Reports*, vol. 39, no. 4, pp. 3737–3746, 2012.

[93] H. Iwata, A. Gaston, A. Remay et al., "The TFL1 homologue KSN is a regulator of continuous flowering in rose and strawberry," *The Plant Journal*, vol. 69, no. 1, pp. 116–125, 2012.

[94] S. Hanano and K. Goto, "Arabidopsis TERMINAL FLOWER1 is involved in the regulation of flowering time and inflorescence development through transcriptional repression," *The Plant Cell*, tpc. 111.088641, 2011.

[95] M. Notaguchi, M. Abe, T. Kimura et al., "Long-distance, graft-transmissible action of Arabidopsis FLOWERING LOCUS T protein to promote flowering," *Plant and Cell Physiology*, vol. 49, pp. 1645–1658, 2008.

[96] S. Otagaki, Y. Ogawa, L. Hibrand-Saint Oyant et al., "Genotype of FLOWERING LOCUS T homologue contributes to flowering time differences in wild and cultivated roses," *The Journal of Plant Biology*, vol. 17, no. 4, pp. 808–815, 2015.

[97] D. P. Wickland and Y. Hanzawa, "The FLOWERING LOCUS T/TERMINAL FLOWER 1 gene family: functional evolution and molecular mechanisms," *Molecular Plant*, vol. 8, no. 7, pp. 983–997, 2015.

[98] I. Guterman, M. Shalit, N. Menda et al., "Rose scent: genomics approach to discovering novel floral fragrance–related genes," *The Plant Cell*, vol. 14, pp. 2325–2338, 2002.

[99] Y. Kaminaga, J. Schnepp, G. Peel et al., "Plant phenylacetaldehyde synthase is a bifunctional homotetrameric enzyme that catalyzes phenylalanine decarboxylation and oxidation," *The Journal of Biological Chemistry*, vol. 281, no. 33, pp. 23357–23366, 2006.

[100] A. Joichi, K. Yomogida, K.-I. Awano, and Y. Ueda, "Volatile components of tea-scented modern roses and ancient Chinese roses," *Flavour and Fragrance Journal*, vol. 20, no. 2, pp. 152–157, 2005.

[101] I. Flament, C. Debonneville, and A. Furrer, "Volatile compounds of roses: characterization of cultivars based on the headspace analysis of living flower emissions," *ACS Symposium Series*, 1993.

[102] G. Scalliet, N. Journot, F. Jullien et al., "Biosynthesis of the major scent components 3,5-dimethoxytoluene and 1,3,5-trimethoxybenzene by novel rose O-methyltransferases," *FEBS Letters*, vol. 523, no. 1-3, pp. 113–118, 2002.

[103] D. Tholl and J. Gershenzon, "The flowering of a new scent pathway in rose," *Science*, vol. 349, no. 6243, pp. 28-29, 2015.

Research Article

Microsatellite Markers Reveal Genetic Diversity and Relationships within a Melon Collection Mainly Comprising Asian Cultivated and Wild Germplasms

Jianbin Hu [ID], Luyin Gao, Yanbin Xu, Qiong Li, Huayu Zhu, Luming Yang, Jianwu Li, and Shouru Sun [ID]

College of Horticulture, Henan Agricultural University, Zhengzhou 450002, China

Correspondence should be addressed to Shouru Sun; ssr365@sina.com

Melon, *Cucumis melo* L., is an important horticultural crop with abundant morphological variability, but the genetic diversity and relationships within wild and cultivated melons remain unclear to date. In this study, thick-skinned (TC) (cultivated subspecies *melo*), thin-skinned (TN) (cultivated subspecies *agrestis*), and wild accessions were analyzed for genetic diversity and relationships using 36 microsatellite markers. A total of 314 alleles were detected with a mean allelic number of 8.72 and polymorphism information content of 0.67. Cluster analysis of the accessions resulted in four distinct clusters (I, II, III, and IV) broadly matching with the TC, TN, and wild groups. Cluster I contained only two Indian wild accessions. Cluster II was consisted of 49 South Asian accessions, 34 wild accessions, and 15 TN accessions. Cluster III was a typical TC group including 51 multiorigin TC accessions and one wild accession. The remaining 88 accessions, including 75 TN accessions, 6 wild accessions, and 7 TC accessions, formed the cluster IV, and all the TN and wild accessions in this cluster were from China. These findings were also confirmed by Principal component analysis and STRUCTURE analysis. The South Asian subspecies *agrestis* accessions, wild and cultivated, had close genetic relationships with a distinctive genetic background. Chinese wild melons showed closeness to cultivated subspecies *agrestis* landraces and could be a return from the indigenous cultivated melons. The AMOVA and pairwise F statistics (F_{ST}) presented genetic differentiation among the three groups, with the strongest differentiation ($F_{ST} = 0.380$) between TC and TN melons. These results offer overall information on genetic diversity and affiliations within a variety of melon germplasms and favor efficient organization and utilization of these resources for the current breeding purpose.

1. Introduction

Melon (*Cucumis melo* L., 2n=2x=24) is an economically important horticultural crop widely distributed in tropical and subtropical areas. This species is highly diverse in morphology, particularly for the fruits, leading to its multiple applications. In China, the sweet fruits of melon are conventionally consumed as a dessert (called "Tiangua") and some medium-size nonsweet fruits as vegetables (called "Caigua"), whereas some large-size nonsweet fruits are commonly used as animal fodder. Also, the abundant diversity in this species attracts a number of studies concerning phylogenetics and taxonomy [1–5]. Although there appear to be some taxonomic methods and several of them have been in controversy or contradiction, an intraspecific taxonomy in *C.*

melo proposed firstly by Pitrat [3] is now generally accepted. This taxonomic criterion divides *C. melo* into two subspecies on the basis of ovary pubescence, *melo* and *agrestis*. Mostly, subspecies *melo* bears comose ovaries and subspecies *agrestis* has ovaries with glabrous skins or short hairs. Subspecies *melo*, conventionally known as thick-skinned (TC) melon in China, is characterized by large or medium fruits and grown widely around the world, while subspecies *agrestis*, also called thin-skinned (TN) melon, carries smaller fruits and is limited in East Asia, especially in China [6]. The TC and TN melons simply refer to the cultivated forms excluding the wild or feral accessions.

Wild melons or the feral forms are mainly found in the centers of origin, Africa and South Asia [7, 8]. These wild accessions are commonly considered as subspecies *agrestis*

but not assigned to the specific varietas. Most wild forms have small leaves and flowers and carry small and oval fruits with thin flesh and small-size seeds [8]. Morphological differences are easily distinguished between cultivated and wild melons; however, the genetic differences at DNA level between them still remain unclear. In addition, most of the available wild melons are found in the Indian subcontinent, and whether they act as the pioneers or ancestors of the modern cultivated melons lacks sufficient evidences yet. Clarification of these issues depends not only on the phenotypic statistics but also on the genotypic data of different accession types.

Genetic diversity and relationships of *C. melo* accessions have received an enormous amount of studies [9–14], most of which focused on a certain melon group (mostly subspecies *melo*) or the accessions from a certain region. Using different marker systems, several studies analyzed the genetic diversity of melon accessions including wild melons but with a limited accession size; most of the wild accessions were found to be more close to subspecies *agrestis* [15–17]. In this study, we aimed to analyze the genetic diversity and relationships of cultivated and wild melon accessions mainly from an Asian collection using a set of 36 core microsatellite markers. These melon accessions were collected from the probable origin regions of *C. melo* ensuring an exact examination from the analysis. The information from our results will favor dissecting the lineage relationship of various melon groups and promote the utilization of these diverse plant resources.

2. Materials and Methods

2.1. Plant Materials. A total of 191 melon accessions were used in this study (Supplemental Table 1), of which 90 were TN accessions, 58 were TC accessions, and 43 were wild accessions. All the accessions were classified into 133 subspecies *agrestis* accessions (TN and wild types) and 58 subspecies *melo* accessions and covered eight countries in the world including the major origin regions of melon, such as India, Iran, Turkey, and China. A majority of the accessions (137, 71.73%) were landraces and the remaining were commercial cultivars (11, 5.76%) and wild accessions (43, 22.51%). Among them, 88 were requested from the National Mid-term Genebank for watermelon and melon (Zhengzhou, China), 86 were from USDA-ARS National Plant Germplasm System, and the remaining 17 were from the Research Group of Watermelon and Melon at Henan Agricultural University.

2.2. Microsatellite Marker Genotyping. Genomic DNA was extracted from young leaf samples of all the accessions using a CTAB procedure described by Doyle and Doyle [18]. SSR markers were used for genotyping. Initially, we collected 300 SSR primer pairs from the published reports [19, 20] and then developed 70 SSR markers from the melon lines DHL92 and TopMark genome assembly to fill up the marker gaps in the melon chromosomes. These markers were screened using 10 diverse accessions (four subspecies *agrestis* accessions, three subspecies *melo* accessions, and three wild accessions). Also, this screening took into account even distribution across the melon chromosomes. Finally, a set of 36 high-polymorphism

SSR markers were obtained, with each chromosome containing three markers at the top, medium, and bottom of the chromosomes. As a result, 15 SSRs were coming from the report of Zhu *et al.* [19], 13 were from the consensus linkage map of Diaz *et al.* [20], and the remaining 8 were newly developed by our research group according to the method of Zhu *et al.* [19]. Detailed information of the 36 markers is listed in Supplementary Table 2.

PCR amplification was carried out in a Thermal Cycler (BIORAD C1000™) with the reaction system and amplification program being same to the report of Wang *et al.* [14]. The amplification products were analyzed by electrophoresis on 6% polyacrylamide gels (19: 1 acrylamide: bis). The band patterns were visualized by silver staining and recorded with a digital camera. The band sizes for each locus were estimated by reference to a DNA ladder (pUC19 DNA/*Msp*l marker, Sangon Biotech, Shanghai).

2.3. Data Analysis. All the markers were scored as codominant data according to the amplicon size. This resulted in a genotypic matrix that was used to calculate the genetic parameters with the software PowerMarker v3.51 [21], i.e., the number of observed alleles (Na) and effective alleles (Ne), Shannon's information index (I), observed (Ho), and expected (He) heterozygosity. Polymorphic information content (PIC) for each marker was calculated using an online program PICcalc [22] that adopted the formula described by Botstein *et al.* [23].

To analyze the genetic diversity and relationship of the melon accessions, the genotypic data were imported into the software MEGA6 [24] to construct a neighbor-joining (NJ) dendrogram. Also, confirmation of the genetic relationships among the accessions was performed using a principal component analysis (PCA) implemented in the software NTSYSpc 2.20e [25] and a model-based program available in STRUCTURE 2.3.1 [26]. The former resulted in a two-dimensional PCA plot showing clustering patterns of the accessions by performing the DCENTER and EIGEN modules in NTSYSpc 2.20e, and the latter offered the clusters for all the K values. The optimal K values (the number of subpopulations in the whole collection) was determined using Markov Chain Monte Carlo (MCMC) algorithm in STRUCTURE HARVESTER. Briefly, each of the probable K was run 10 times with K=1 to 10, and the length of burn-in period was separately set at 10,000 and 100,000 MCMC repeats after burn-in with an admixture and allele frequency correlated model. The optimal K was determined by the log probability of data [Lnp(D)] from the output and the Evanno's ΔK between successive K values [27].

Genetic differentiation among the different groups was measured by calculating pairwise F statistics (F_{ST}), genetic distance [28], and analysis of molecular variance (AMOVA) using GeneAlEx 6.5 [29].

3. Results

3.1. Characterization of Microsatellite Marker Polymorphism. All the 36 SSR markers produced clear band patterns, revealing single-locus variation among the melon accessions.

TABLE 1: Statistics of genetic variation as measured for 36 SSRs estimated from 191 melon accessions.

Marker	Na	Ne	Ho	He	PIC
CM07	7	5.83	0.04	0.83	0.60
CMCT505	8	3.89	0.09	0.75	0.71
SSR011330	9	3.63	0.11	0.73	0.69
SSR012562	9	3.39	0.12	0.71	0.68
gSSR4959	6	1.95	0.03	0.49	0.46
SSR013487	16	3.41	0.16	0.71	0.69
SSR014660	14	4.75	0.05	0.79	0.78
SSR015784	6	3.69	0.11	0.73	0.68
SSR016829	9	3.97	0.14	0.75	0.71
HNM33	10	4.32	0.10	0.77	0.74
HNM12	11	5.92	0.11	0.83	0.81
SSR020162	5	2.46	0.05	0.60	0.54
SSR020947	5	2.31	0.06	0.57	0.50
DE1557	12	5.25	0.03	0.81	0.79
SSR023138	8	4.20	0.16	0.76	0.74
HNM41	7	1.94	0.13	0.48	0.45
DE1103	7	2.40	0.09	0.58	0.55
CMAGN52	9	3.54	0.02	0.72	0.70
CMAGN75	13	7.67	0.05	0.87	0.86
gSSR22419	6	3.15	0.04	0.68	0.63
SSR029474	5	4.21	0.03	0.76	0.72
SSR029716	9	4.47	0.02	0.78	0.75
HNM31	8	4.04	0.09	0.75	0.71
HNM40	8	4.11	0.15	0.76	0.73
CMTC47	8	3.08	0.18	0.68	0.64
SSR033639	8	4.70	0.02	0.79	0.76
CMATN22	7	3.32	0.07	0.70	0.65
CM38	9	4.81	0.11	0.79	0.76
CMTCN8	10	2.56	0.05	0.61	0.58
HSSR010	9	5.16	0.05	0.81	0.79
DM0673	12	6.08	0.11	0.84	0.82
SSR038372	5	2.29	0.05	0.56	0.47
CMGA104	9	3.68	0.05	0.73	0.69
SSR040314	10	3.51	0.06	0.72	0.67
SSR041311	11	3.56	0.04	0.72	0.70
CMGAN80	9	2.54	0.11	0.61	0.58
Mean	8.72	3.88	0.08	0.72	0.68

Na: the number of observed alleles.
Ne: the number of effective alleles.
Ho: observed heterozygosity.
He: expected heterozygosity.
PIC: polymorphic information content.

Five genetic parameters (Na, Ne, Ho, He, and PIC) were calculated for the 36 markers estimated from the 191 melon accessions, as shown in Table 1. In total, 314 alleles were detected varying from 5 (SSR020162, SSR020947, SSR029474, and SSR038372) to 18 (SSR013487) with a mean of 8.72. Ne, an important parameter to measure genetic diversity in a finite population, averaged 3.88 ranging from 1.94 (HNM41) to 7.67 (CMAGN75). No heterozygosity deficiency was observed in the accession collection; the Ho values were quite low (<0.20) at the loci with a mean of 0.08. He means expected heterozygosity in a certain population and averaged 0.72 in the accession collection. The highest (0.87) and lowest (0.56) He values were observed for CMAGN75 and SSR038372, respectively. For each locus, He value was much higher than Ho value, revealing a high homozygosity at the given loci among the accessions. PIC is generally used for characterization of marker polymorphism and the values ranged between 0.45 (HNM41) and 0.86 (CMAGN75) (mean=0.68) in the accession collection. All the genetic parameters revealed a high level of polymorphism for the 36 markers, favoring
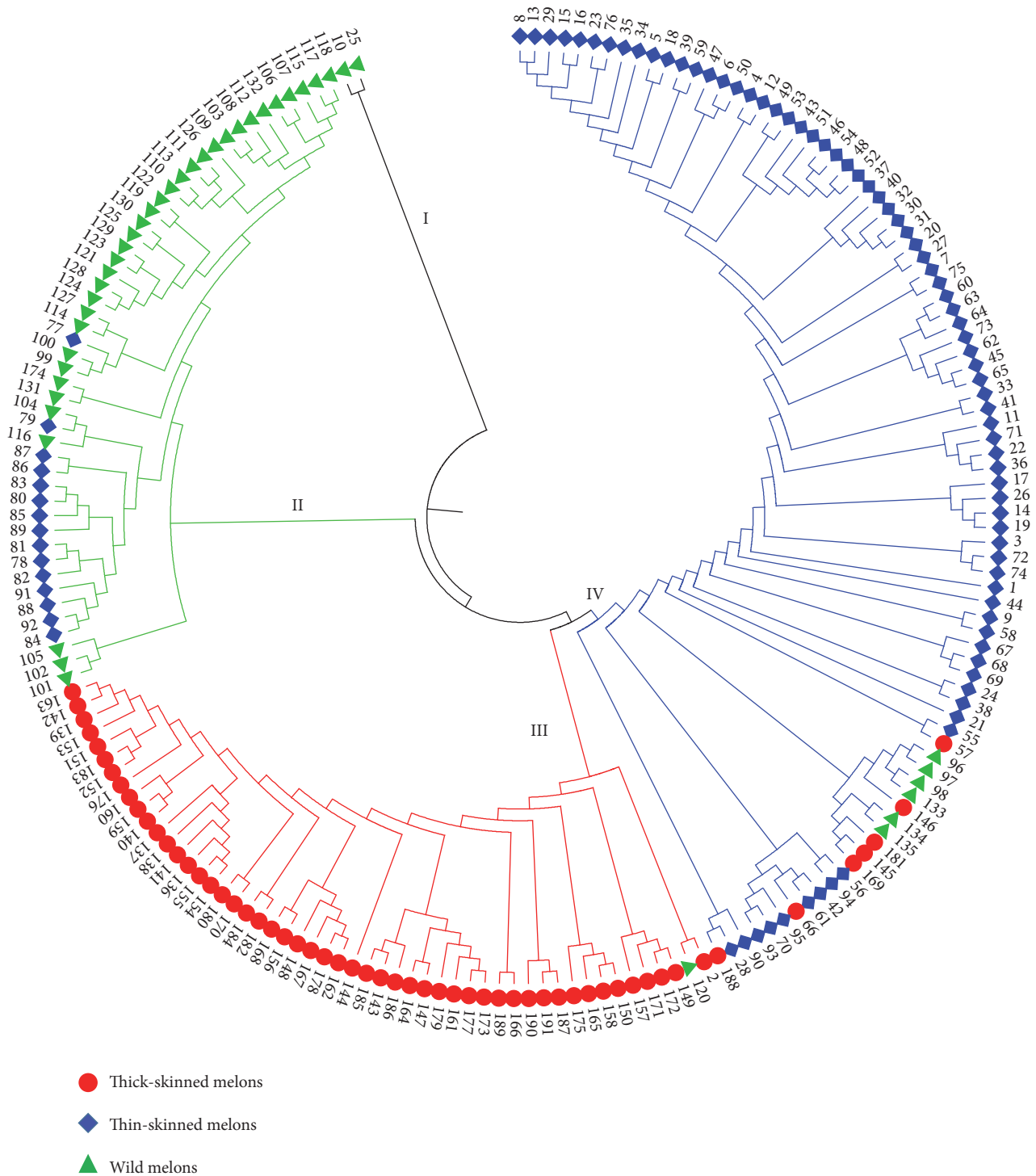
FIGURE 1: A neighbor-joining dendrogram showing the genetic affiliations of the 191 melon accessions. All the accessions were divided into four clusters (I, II, III, and IV). Numbers indicate the accession codes as listed in Supplementary Table 1.

the establishment of the genetic affiliations within the melon collection.

3.2. Establishment of Genetic Relationship for the Accession Collection. With the SSR genotypic data, a NJ dendrogram (Figure 1) was constructed based on Nei's similarity

coefficients [30] showing the genetic relationship among the accessions. The dendrogram clustered the 191 accessions into four distinct clusters (I, II, III, and IV). Cluster I contained only two Indian wild accessions (No. 10 and 25), which were highly diverse and distinguished from the other accessions. Cluster II was consisted of 34 wild accessions and 15 TN

landraces (*momordica* accessions) with the pairwise genetic distances (GDs) varying from 0.23 to 0.78 (mean = 0.63). Most of the accessions (32 wild accessions and 14 TN landraces) in this cluster derived from India and only three from Maldives, two wild accessions (No. 99 and 100) and one TN landrace (No. 77). Clearly, the wild and TN accessions in this cluster from the two adjacent countries had close lineages. Cluster III was a typical TC group (GDs ranging from 0.20 to 0.67, mean = 0.54), containing 51 TC accessions (subspecies *melo*) and one wild accession (No. 120). These accessions in this cluster were morphologically diverse belonging to at least five varietas suggested by Pitrat [3], such as *cantalupensis*, *reticulatus*, *inodorus*, *ameri*, and *chandalak*, and also, their origins covered a wide geographical distribution (China, India, Tunisia, Japan, Afghanistan, and Iran). Although these subspecies *melo* landraces in cluster III derived from different regions, they had close genetic relationships and similar genetic backgrounds, probably implying their same origin. The wild accession in this cluster III (No. 120) was an exception since it was clustered together with the subspecies *melo* accessions, and therefore it could involve in gene exchange with the subspecies *melo* plants.

The remaining 88 accessions formed cluster IV that covered both the two subspecies (TC, TN, and wild groups), mainly representing by East Asian TN melons (i.e., 75 subspecies *agrestis* accessions). The TN accessions in this cluster included 48 *chinensis* accessions, 13 *conomon* accessions, 7 *makuwa* accessions, 6 *momordica* accessions and one *acidulous* accession, most of which derived from China. Except for the TN accessions, cluster IV possessed 7 TC accessions (No. 57, 66, 145, 146, 169, 181, and 188) and 6 wild accessions (No. 96, 97, 98, 133, 134, and 135). This fact indicated close affiliations of these accessions. The seven TC accessions were clustered into this cluster, perhaps due to the introgression of the subspecies *agrestis* lineages during their domestication processes. It should be noted that six wild accessions were also included in this cluster, in contrast to most of the wild accessions assigned to cluster II. Of these wild accessions, 4 were from China and the other two from Costa Rica and the US, which were markedly different from the wild accessions in clusters I and II with South Asian origin. A lower level of variation was observed for cluster IV with the GDs ranging from 0.18 to 0.72 (mean = 0.45), indicating a comparatively narrow genetic basis in this cluster.

Furtherly, two methods, principal component analysis (PCA) and STRUCUTRE analysis, were used to offer an alternative view of the relationships within the accession collection. On the PCA dendrogram (Figure 2(a)), all the melon accessions, which were labelled with different symbols and colors according to the accession classification, tended to form three clusters, i.e., the red, blue, and green regions. Most of the TC accessions were positioned to the red region while the TN accessions were mainly to the blue region. These two regions were not separated absolutely because some TN and TC accession (e.g., No. 28, 42, 61, 66, 57, 146, 181, and 188) were mixed together. The wild accessions (green triangles) were scattered across a wide area, even some (e.g., No. 96, 97, 98, 133, 134, and 135) seeping into the blue region. This implied the complex genetic background of the wild

(a)

(b)
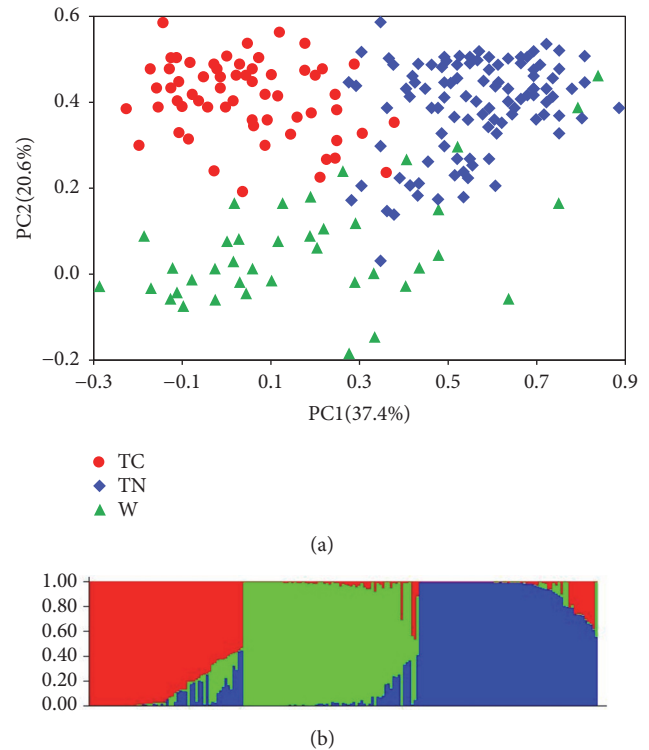
Figure 2: Genetic structure of the 191 melon accessions revealed by principal component analysis (PCA) (a) and STRUCTURE analysis (b). The symbols and colors for the accessions correspond to those of the Figure 1. TC, TN, and W represent thick-skinned, thin-skinned, and wild accessions, respectively.

forms. Similarly, the STRUCUTRE analysis positioned all the accessions into three subpopulations (Figure 2(b)), which represented TC (red), TN (blue) and wild (green) groups, respectively. Obviously, the three methods gave similar results on positioning of the melon accessions.

3.3. Comparison of Genetic Diversity for Wild, TC, and TN Groups. Since distinct divergences were found among the different groups, the six genetic parameters (Na, Ne, Ho, He, I, and PIC) were computed for each group to compare their diversity levels. As shown in Table 2, TN and wild groups had more Na than TC group demonstrating a higher level of allelic polymorphism. Wild group showed the highest values of Ne (4.08) and He (0.72) indicating a wide heterogenicity at the genome level. The sample size of wild accessions was the smallest among the three groups; however, the three parameters (He, I, and PIC) that reflect diversity level revealed the highest values in wild group and verified its abundant diversity. According to the parameter values (Ne, He, and PIC) in Table 2, TC group was more slightly diverse than TN group. In addition, the alleles that are specific to a certain group and the ones shared by the two or more groups were shown in Figure 3. Both TN and wild groups had 14 group-specific alleles, whereas TC group possessed such 11 genes. Certainly, the shared alleles among the groups accounted for the main part, 143 alleles shared by all the three

TABLE 2: Comparison of genetic diversity for the thick-skinned (TC), thin-skinned (TN), and wild (W) groups.

Accession group	Na	Ne	Ho	He	I	PIC
TC	226	2.93	0.04	0.63	1.26	0.58
TN	256	2.87	0.07	0.59	1.26	0.55
W	250	4.08	0.18	0.72	1.51	0.67

Na: the number of observed alleles.
Ne: the number of effective alleles.
Ho: observed heterozygosity.
He: expected heterozygosity.
I: Shannon's information index.
PIC: Polymorphic information content.

TABLE 3: Molecular analyses of variance (AMOVA) among the accession groups and origin regions.

Source of variation	df	Variance components	Percentage of variation	P value
Among groups	2	58.03	28.70	<0.01
Among regions	2	61.54	30.44	<0.01
Among individuals	176	75.67	37.42	<0.01
Within individuals	191	6.96	3.44	<0.01

Groups were defined by the thick-skinned, thin-skinned, and wild accessions.
Regions were defined by South Asia (India and Maldives), East Asia (China, Malaysia, and Japan), West Asia (Iran and Turkey), and Africa (Tunisia).

TABLE 4: Pairwise estimates of genetic differentiation among the three accession groups using pairwise F_{ST} (above the diagonal) and Nei's genetic distance (below the diagonal). Permutation tests confirmed that all the F_{ST} values were significant at $P < 0.01$.

Accession group	TC	TN	W
TC	—	0.380	0.293
TN	0.102	—	0.319
W	0.083	0.100	—

TC, TN, and W mean the thick-skinned, thin-skinned, and wild groups.
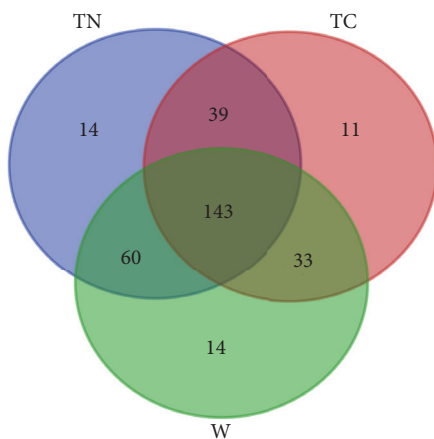
FIGURE 3: A Venn diagram showing the number of alleles specific to a certain group or shared by different groups. TC, TN, and W mean the thick-skinned, thin-skinned, and wild groups, respectively.

groups and 33–60 alleles by each two groups. Combination of the findings from the genetic parameters and group-specific alleles demonstrated the highest level of diversity present within wild accessions, following by cultivated subspecies *melo* and *agrestis* accessions.

3.4. *Examination of Genetic Differentiation for the Collection.* To analyze the genetic differentiation of the collection, AMOVA was conducted using accession groups and geographic origins as sources of variation, and showed that 28.70% of the total variation was attributed to the differentiation between groups and 30.44% was to the differentiation between geographic regions (Table 3). The highest percentage (37.42%) occurred among the accessions while the variation within accessions was quite low with the percentage of 3.44%. Given that the geographic origins were basically related to the melon classification (Supplemental Table 1), the accession type was an important factor leading to genetic differentiation. Also, the differentiation among the three groups was measured by pairwise F_{ST} and Nei's genetic distance (Table 4). Each of the pairwise F_{ST} among the three groups was higher than 0.25, a threshold for existence of a very high level of genetic differentiation suggested by Wright [31]. That indicated that each of the melon groups was clearly differentiated from the others. The highest pairwise F_{ST} value of 0.380 was found between the TC and TN groups and displayed a strong genetic differentiation. This was also confirmed by the maximum genetic distance (0.102) between the TC and TN groups. The minimum genetic distance (0.083) occurred between TC and wild groups. Of the 36 marker loci, 15 showed clear differentiation among the groups with the pairwise F_{ST} value of >0.15 (Figure 4), such as SSR013487, SSR014660, HNM33,

Figure 4: Pairwise F_{ST} values among the thick-skinned (TC), thin-skinned (TN), and wild (W) accessions at the 36 marker loci.
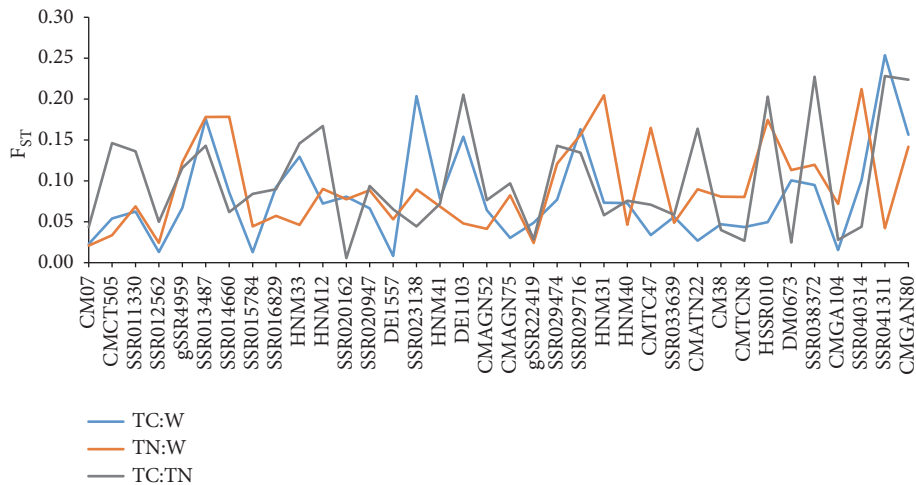
HNM12, SSR023138, DE1103, SSR029716, HNM31, CMTC47, CMATN22, HSSR010, SSR038372, SSR040314, SSR041311, and CMGAN80. These loci could reflect evolutionary forces (e.g., artificial selection) affecting domestication of cultivated melons.

4. Discussion

A diversity of plant germplasms is valuable resources for present and future commercial producers and researchers; they can be used for breeding of new cultivars to meet the demand for food and studying the origin, evolution, and taxonomy of plant species. Melon is such a horticultural crop with abundant diversity. During the past decades, it earned worldwide attentions in scientific research (e.g., developmental biology and genetics) [32, 33] and agricultural production as its yield of fresh fruits frequently entered the top ten of the main fruits in the world (FAO Statistics from 2007-2016, http://www.fao.org/faostat/en/#data/QC).

C. melo has a large number of morphotypes, cultivated and wild [8]; cultivated melons scatter around the world and wild melons are mainly concentrated in North Africa and South Asia [34, 35]. In China, cultivated melons are commonly distinguished into TC and TN groups, which respectively correspond to the two subspecies, *melo* and *agrestis*. TN melons (subspecies *agrestis* group) are special morphotypes mainly distributed in East Asia, especially in central and eastern China, the important diversity center of subspecies *agrestis* in the world [14, 35]. This kind of melons has attracted wide attentions these years for the striking characteristics, such as adversity tolerance, early maturity, vigorous growth and good fruit set. TC melons (subspecies *melo* group), having a large number of commercial cultivars in the world, are the main cultivated forms. To utilize efficiently these melon resources in modern agriculture, we adopted 36 core microsatellite markers to examine the genetic diversity and relationships present within an Asian melon collection. Our assays detected 314 alleles (mean=8.72) with a mean PIC value of 0.67, showing a high level of variation

in the melon genome. The mean allelic number and PIC value were much higher than those of available reports with the experimental materials being a certain group or from a certain region [11–14]. This could be due to the diverse accessions used in the present study. Gao *et al*. [36] expanded the sample size to 471 melon accessions and detected a higher level of variation, a mean allelic number of 9.0 per SSR locus and PIC value of 0.68.

To date, genetic diversity and relationships in *C. melo* have been frequently reported mainly focusing on cultivated accessions. These results showed that subspecies *melo* accessions were obviously distinguished from subspecies *agrestis* accessions [4, 36–39], implying an existence of genetic divergence at the subspecies level. Also, several other reports, which involved some wild melon accessions mainly from Africa, India, and America, showed that African and Indian wild accessions were close to *conomon*, *chito*, *dudaim*, and *momordica*, but far from American and European *cantalupensis* and *inodorus* [2, 15–17]. Based on the combination of phenotypic characters and molecular marker data, American wild melons also showed genetic affinities to Asian subspecies *agrestis* [17]. These American wild accessions were assumed to be the introduction from India. As expected, the wild melon accessions in the present research, especially for Indian wild accessions, were highly diverse (Table 2) and most of them were distinctly different from the cultivated accessions (TC and Chinese TN accessions) (Figures 1 and 2; Table 4). Indian subspecies *agrestis* landraces were clustered closely with the wild accessions showing their close genetic relationship, as a probable result of the frequent gene exchanges (e.g., mutual pollination). From the records of USDA-ARS National Plant Germplasm System (Data not shown), these Indian subspecies *agrestis* accessions have similar morphological features to the wild species and are probably old indigenous landraces or semiwild forms. Since there are a variety of wild melons and semiwild forms or landraces in India, it is reasonable to assume that Indian subcontinent acts as the center of origin of this crop [7]. This view is supported by the present results as well as the

previous reports. Interestingly, the Chinese wild accessions were genetically close to the cultivated subspecies *agrestis* accessions (Chinese TN accessions) in group IV. The same finding was also described in our previous study [40]. A view proposed by Pitrat [8] is that the "wild melon" from the New World is not true wild form but a return to a wild status from cultivated melons. Chinese wild melons are likely to be such a case. There are numerous wild melons scattering across Central China, particularly in Henan province, which are customarily called "Mapao", a common weedy fruit in crop field. It bears a large number of small-size fruits ($\sim$ 20 g) but with domesticated characters, e.g., yellow skin, sweet flesh, and aroma. This kind of "wild melons" could be an escape from indigenous cultivated melons, probably the subspecies *agrestis*. Therefore, East Asian subspecies *agrestis* melons probably originated in Indian subcontinent and were intensively domesticated in Central China, as the indigenous wild melons could not be their pioneers or ancestors. The similar view has been suggested in several studies that East Asian subspecies *agrestis* melons benefited from Indian introduction (perhaps via Myanmar, Laos, and eastern China; $\sim$100 BC) and were domesticated in Huang-Huai-Hai plain in China [37, 41].

The strongest differentiation occurred between TC and TN melons ($F_{ST} = 0.380$); such a high level of differentiation could be supported by a fact that TC and TN accessions belong to two different subspecies, *melo* and *agrestis* [3, 5]. Also, the subspecies divergence in *C. melo* have been reported in several studies [2, 36–38], although the accessions of the two subspecies were of multiple-origin. As worldwide-distribution cultivated forms, subspecies *melo* (TC melons) were initially considered to originate from Africa [1, 34, 35] but the later researches evidence their Asian origin [7]. According to the taxonomy of Pitrat [3], subspecies *melo* possesses 11 varietas and is richer in morphological diversity than subspecies *agrestis* (containing 5 varietas), as also evidenced from our result (Table 2). This could be due to undergo different selection patterns; subspecies *melo* melons underwent selections over the worldwide regions while selection of subspecies *agrestis* melons was mainly restricted in Central China.

In the present research, we used SSR markers separated by polyacrylamide gel. It would be interesting to compare phylogenetic relationships among studied melon accessions using another method like capillary electrophoresis or even other molecular approaches including SNP or DArT markers. Such new researches can confirm or show a conflict with current study for the genetic clustering and it will be the next step for the investigation.

5. Conclusions

Genetic diversity and relationship are crucial for plant breeding as they determine the efficient utilization of the genetic materials and selection of potential parents. Accurate measurement of genetic diversity and relationship present within an accession collection relies on the molecular markers (e.g., SSR) with stability and even distribution across the genome. The present study revealed an existence of distinct population structure in 191 melon accessions. Indian wild accessions, revealing a close relationship to the local subspecies *agrestis* landraces, had a high level of variation and were distinguished from the cultivated accessions, subspecies *melo* and subspecies *agrestis*. Chinese wild melons also showed close lineages to the local subspecies *agrestis* accessions. Genetic differentiations occurred among wild accessions, cultivated subspecies *melo,* and subspecies *agrestis* accessions; the strongest differentiation was between cultivated subspecies *melo* and *agrestis* accessions indicating a subspecies-level divergence. The information of the genetic diversity and relationships among the cultivated and wild melons will be helpful for efficient organization and utilization of these genetic materials in current breeding programs.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

Supplementary Materials

Table 1: descriptions of the 191 melon accessions used in the present study. Table 2: the information of the 37 SSR markers used in the present study. *(Supplementary Materials)*

References

[1] J. H. Kirkbride, *Biosystematic Monograph of the Genus* Cucumis *(Cucurbitaceae)*, Parkway Publishers, Boone, NC, USA, 1993.

[2] A. Stepansky, I. Kovalski, and R. Perl-Treves, "Intraspecific classification of melons (*Cucumis melo* L.) in view of their phenotypic and molecular variation," *Plant Systematics and Evolution*, vol. 217, no. 3-4, pp. 313–332, 1999.

[3] M. Pitrat, "Melon, vegetables I," in *Handbook of Plant Breeding*, J. Prohens and F. Nuez, Eds., vol. 1, pp. 283–315, Springer, New York, NY, USA, 2008.

[4] Y. Akashi, N. Fukuda, T. Wako, M. Masuda, and K. Kato, "Genetic variation and phylogenetic relationships in East and South Asian melons, *Cucumis melo* L., based on the analysis of five isozymes," *Euphytica*, vol. 125, no. 3, pp. 385–396, 2002.

[5] D. P. Lin, "Comments on intraspecific classification of melon," *Chinese Cucubit*, vol. 25, no. 5, pp. 42–46, 2012 (Chinese).

[6] J. B. Hu, S. W. Ma, Z. H. Jian et al., "Analysis of genetic diversity of Chinese melon (*Cucumis melo* L.) germplasm resources based on morphological characters," *Journal of Plant Genetic Resources*, vol. 14, no. 4, pp. 612–619, 2013 (Chinese).

[7] P. Sebastian, H. Schaefer, I. R. H. Telford, and S. S. Renner, "Cucumber (*Cucumis sativus*) and melon (C. *melo*) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 107, no. 32, pp. 14269–14273, 2010.

[8] M. Pitrat, "Phenotypic diversity in wild and cultivated melons (*Cucumis melo*)," *Plant Biotechnology Journal*, vol. 30, no. 3, pp. 273–278, 2013.

[9] A. I. López-Sesé, J. E. Staub, and M. L. Gómez-Guillamón, "Genetic analysis of Spanish melon (*Cucumis melo* L.) germplasm using a standardized molecular-marker array and geographically diverse reference accessions," *Theoretical and Applied Genetics*, vol. 108, no. 1, pp. 41–52, 2003.

[10] S. Sensoy, S. Büyükalaca, and K. Abak, "Evaluation of genetic diversity in Turkish melons (*Cucumis melo* L.) based on phenotypic characters and RAPD markers," *Genetic Resources and Crop Evolution*, vol. 54, no. 6, pp. 1351–1365, 2007.

[11] M. Raghami, A. I. López-Sesé, M. R. Hasandokht, Z. Zamani, M. R. F. Moghadam, and A. Kashi, "Genetic diversity among melon accessions from Iran and their relationships with melon germplasm of diverse origins using microsatellite markers," *Plant Systematics and Evolution*, vol. 300, no. 1, pp. 139–151, 2014.

[12] I. Solmaz, Y. A. Kacar, O. Simsek, and N. Sari, "Genetic characterization of Turkish snake melon (*Cucumis melo* L. subsp. *meloflexuosus* Group) accessions revealed by SSR markers," *Biochemical Genetics*, vol. 54, no. 4, pp. 534–543, 2016.

[13] S. Pavan, A. R. Marcotrigiano, E. Ciani et al., "Genotyping-by-sequencing of a melon (*Cucumis melo* L.) germplasm collection from a secondary center of diversity highlights patterns of genetic variation and genomic features of different gene pools," *BMC Genomics*, vol. 18, no. 1, p. 59, 2017.

[14] Y.-L. Wang, L.-Y. Gao, S.-Y. Yang et al., "Molecular diversity and population structure of oriental thin-skinned melons, *Cucumis melo* subsp. *agrestis*, revealed by a set of core SSR markers," *Scientia Horticulturae*, vol. 229, pp. 59–64, 2018.

[15] A. J. Monforte, J. Garcia-Mas, and P. Arús, "Genetic variability in melon based on microsatellite variation," *Plant Breeding*, vol. 122, no. 2, pp. 153–157, 2003.

[16] C. Esteras, G. Formisano, C. Roig et al., "SNP genotyping in melons: Genetic variation, population structure, and linkage disequilibrium," *Theoretical and Applied Genetics*, vol. 126, no. 5, pp. 1285–1303, 2013.

[17] D. S. Decker-Walters, S.-M. Chung, J. E. Staub, H. D. Quemada, and A. I. López-Sesé, "The origin and genetic affinities of wild populations of melon (*Cucumis melo*, Cucurbitaceae) in North America," *Plant Systematics and Evolution*, vol. 233, no. 3-4, pp. 183–197, 2002.

[18] J. J. Doyle and J. L. Doyle, "Isolation of plant DNA from fresh tissue," *Focus*, vol. 12, no. 1, pp. 13–15, 1990.

[19] H. Zhu, L. Guo, P. Song et al., "Development of genome-wide SSR markers in melon with their cross-species transferability analysis and utilization in genetic diversity study," *Molecular Breeding*, vol. 36, no. 11, p. 153, 2016.

[20] A. Diaz, M. Fergany, G. Formisano et al., "A consensus linkage map for molecular markers and Quantitative Trait Loci associated with economically important traits in melon (*Cucumis melo* L.)," *BMC Plant Biology*, vol. 11, no. 1, p. 111, 2011.

[21] K. Liu and S. V. Muse, "PowerMaker: an integrated analysis environment for genetic maker analysis," *Bioinformatics*, vol. 21, no. 9, pp. 2128-2129, 2005.

[22] S. Nagy, P. Poczai, I. Cernák, A. M. Gorji, G. Hegedus, and J. Taller, "PICcalc: An online program to calculate polymorphic information content for molecular genetic studies," *Biochemical Genetics*, vol. 50, no. 9-10, pp. 670–672, 2012.

[23] D. Botstein, R. L. White, M. Skolnick, and R. W. Davis, "Construction of a genetic linkage map in man using restriction fragment length polymorphisms," *American Journal of Human Genetics*, vol. 32, no. 3, pp. 314–331, 1980.

[24] K. Tamura, G. Stecher, D. Peterson, A. Filipski, and S. Kumar, "MEGA6: Molecular Evolutionary Genetics Analysis version 6.0," *Molecular Biology and Evolution*, vol. 30, no. 12, pp. 2725–2729, 2013.

[25] F. J. Rohlf, "NTSYSpc, numerical taxonomy and multivariate analysis system," in *Exeter Software*, vol. 2.2, Setauket, NY, USA, 2005.

[26] J. K. Pritchard, M. Stephens, and P. Donnelly, "Inference of population structure using multilocus genotype data," *Genetics*, vol. 155, no. 2, pp. 945–959, 2000.

[27] G. Evanno, S. Regnaut, and J. Goudet, "Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study," *Molecular Ecology*, vol. 14, no. 8, pp. 2611–2620, 2005.

[28] M. Nei, "Estimation of average heterozygosity and genetic distance from a small number of individuals," *Genetics*, vol. 89, no. 3, pp. 583–590, 1978.

[29] R. Peakall and P. E. Smouse, "GenALEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update," *Bioinformatics*, vol. 28, no. 19, Article ID bts460, pp. 2537–2539, 2012.

[30] M. Nei and W. H. Li, "Mathematical model for studying genetic variation in terms of restriction endonucleases," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 76, no. 10, pp. 5269–5273, 1979.

[31] S. Wright, *Evolution and the Genetics of Populations, Volume 3: Experimental Results and Evolutionary Deductions*, University of Chicago Press, Chicago, IL, USA, 1984.

[32] A. Gur, I. Gonda, M. Portnoy et al., "Genomic aspects of melon fruit quality," in *Genetics and Genomics of Cucurbitaceae*, R. Grumet, N. Katzir, and J. Garcia-Mas, Eds., Plant Genetics and Genomics: Crops and Models, pp. 377–408, Springer, New York, NY, USA, 2016.

[33] M. C. Kyriacou, D. I. Leskovar, G. Colla, and Y. Rouphael, "Watermelon and melon fruit quality: The genotypic and agro-environmental factors implicated," *Scientia Horticulturae*, vol. 234, pp. 393–408, 2018.

[34] T. Kerje and M. Grum, "The origin of melon, Cucumis melo: A review of the literature," *Acta Horticulturae*, vol. 510, pp. 37–44, 2000.

[35] D. P. Lin, "Origin, classification and evolution for cultivated plants of Chinese melon," *China Cucurbit Vegetable*, vol. 23, no. 4, pp. 34–36, 2010 (Chinese).

[36] P. Gao, H. Ma, F. Luan, H. Song, and J. C. Nelson, "DNA fingerprinting of chinese melon provides evidentiary support of seed quality appraisal," *PLoS ONE*, vol. 7, no. 12, Article ID e52431, 2012.

[37] F. Luan, I. Delannay, and J. E. Staub, "Chinese melon (*Cucumis melo* L.) diversity analyses provide strategies for germplasm curation, genetic improvement, and evidentiary support of domestication patterns," *Euphytica*, vol. 164, no. 2, pp. 445–461, 2008.

[38] J. Hu, P. Wang, Y. Su, R. Wang, Q. Li, and K. Sun, "Microsatellite diversity, population structure, and core collection formation in melon germplasm," *Plant Molecular Biology Reporter*, vol. 33, no. 3, pp. 439–447, 2015.

[39] J. E. Staub, Y. Danin-Poleg, G. Fazio, T. Horejsi, N. Reis, and N. Katzir, "Comparative analysis of cultivated melon groups (*Cucumis melo* L.) using random amplified polymorphic DNA and simple sequence repeat markers," *Euphytica*, vol. 115, no. 3, pp. 225–241, 2000.

[40] J. Hu, P. Wang, Q. Li, and Y. Su, "Microsatellite analysis of genetic relationships between wild and cultivated melons in Northwest and Central China," *Molecular Biology Reports*, vol. 41, no. 12, pp. 7723–7728, 2014.

[41] J. D. McCreight, J. E. Staub, A. López-Sesé, and S.-M. Chung, "Isozyme variation in Indian and Chinese melon (*Cucumis melo* L.) germplasm collections," *Journal of the American Society for Horticultural Science*, vol. 129, no. 6, pp. 811–818, 2004.

Research Article

Tunisian Table Olive Oil Traceability and Quality Using SNP Genotyping and Bioinformatics Tools

Rayda Ben Ayed (ID) **and Ahmed Rebai** (ID)

Molecular and Cellular Screening Processes Laboratory, Centre of Biotechnology of Sfax, BP 1177, 3018, University of Sfax, Tunisia

Correspondence should be addressed to Rayda Ben Ayed; raydabenayed@yahoo.fr

To enhance and highlight the authentication and traceability of table olive oil, we considered the analysis of 11 Tunisian table olive cultivars based on seven SNP molecular markers (SOD, CALC, FAD2.1, FAD2.3, PAL70, ANTHO3, and SAD.1) localized in six different genes. Accordingly, we assessed the potential genotype-phenotypes links between the seven SNPs, on the one hand, and the quantitative and qualitative parameters, on the other. The obtained genotypes were analyzed with computational biology tools based on bivariate analysis, multinomial logistic regression, and the Bayesian networks modeling. Obtained results showed that PAL70 SNP marker was negatively influenced by the phenol rate ($r = -0.886$; $p <0.001$), the oxidative stability ($r = -0.884$; $p <0.001$), traducing a direct effect of the PAL70 genotype deviations on the proportion of total phenol for each variety. Additionally, we revealed a significant association of SAD.1 marker with the content of the linolenic unsaturated fatty acids (C18:3; $p=0.046$). Moreover, SAD.1 was positively correlated with the saturated stearic acid C18:0 ($r = 0.644$; $p = 0.032$) based on multinomial logistic regression and Bayesian networks modeling, respectively. This research work provides better understanding and characterization of the quality of Tunisian table olive and supplies a significant knowledge and data information for table olive traceability and breeding.

1. Introduction

In the Mediterranean basin countries, olive is one of the most important agricultural products. It is used for olive oil extraction or processed as either table olives. These latter are chosen from cultivated olive trees (*Olea europaea* L.) with regard to their size, volume, taste, and other organoleptic properties that make them suitable for table consumption.

According to the current data provided by the International Olive Oil Council [1], the world production of table olive is evaluated at 2,953,500 tons in the 2017/2018 season showing an impressive increase of 211 % in the global production of table olives in the period between 1991 and 2018. The most dramatic raises have been noted in Egypt, Turkey, Spain, Algeria, Greece, Argentina, Iran, and Morocco.

Initially, table olive production was restricted to the producing regions, mainly European Union, Egypt, Turkey, Algeria, Morocco, Argentina, and Syria. However, nowadays,

table olive production and exports have extended to other countries like USA and Jordan with 6 and 5 tons, respectively [1]. In recent years, several countries such as Tunisia, Argentina, Jordan, and Morocco have enhanced their production of table olives compared to the previous season unlike some producer countries that remained constant or sustained a cutback, like Syria by 47 % and Peru by 1 % [1].

According to our previous results from studies performed on main world table olive varieties [2], the genetic diversity and distribution of table olive varieties are related to several qualitative and quantitative parameters. Additionally, biological and organoleptic markers together with computational biology tools could help characteristics determination of table olives and hence start resolving its authenticity. Moreover, this study highlighted that some varieties could be more suitable as olive oil cultivars than table olive consumption regarding their high yield and consistent oil fruit content (22%) [2]. For these reasons, it is crucial to develop strategies

and procedures of traceability and authentication that allows rapid and relevant identification and then valorisation of cultivars. Generally, traceability, authenticity, and detection of fraudulency in olive oil are performed by analytical techniques. However, biochemical approach and analyses are not sufficient to assess olive oil authenticity due to the influence of environmental conditions on oil components [3–5].

More recently, the use of DNA molecule-based analyses in olive oil becomes of a great interest to meet the needs of consumers and will be essential for studying the traceability of olive oil because of their several advantages, particularly, the reliability and reproducibility of results.

Seven SNPs localized in five different genes: *fatty acid desaturase, anthocyanidin synthase (ANS), calcium-binding protein, stearoyl-acyl carrier protein desaturase (SAD),* and *L-phenylalanine ammonia lyase.* The FAD2.1 and FAD2.3 SNPs are both harboured by the *FAD2* gene which is involved in the biosynthesis of highly unsaturated fatty acids (HUFA) from the precursor polyunsaturated fatty acids (PUFA) [6]. The third studied SNP, named ANTHO3, is localized in the *anthocyanidin synthase* gene, a 2-oxoglutarate iron-dependent oxygenase, and catalyzes the penultimate step in the biosynthesis of the anthocyanin class of flavonoids [7]. The CALC SNP is carried by the calcium-binding protein gene that is involved in response to abiotic constraints (salinity, cold, and drought) [8]. The fifth SNP localized in the *Stearoyl-acyl carrier protein desaturase* gene is involved in the desaturation of C18:0 to C18:1, monounsaturated oleic acid intermediates [9]. The sixth SNP, named SOD, is an insertion/deletion polymorphism type localized in Cu–Zn superoxide dismutase gene associated with the oxidative stress response [10, 11]. The last SNP is the PAL70, located in the *L-phenylalanine ammonia lyase* gene that is implicated in phenolic biosynthesis, including the formation of flavonoids, lignin, and hydroxycinnamic acids [12].

Our study aims to assess the correlations between the seven SNPs and table olive oils quality parameters and their efficiency in the authentication and traceability of Tunisian table olive oil.

2. Materials and Methods

2.1. Plant Material. Eleven Tunisian table olive cultivars were selected from north to south geographical regions of Tunisia (Chetoui, Tounsi, Meski, Oueslati, El Horr, Fakhari, Zarrazi, Chemchali, Besbessi, Fougi, and Toffehi). Two trees were sampled for each cultivar, and olive oil was extracted from each sample, followed by DNA extraction [13].

2.2. Olive Oil Extraction. Fully ripened fruits coming from different dual purpose and table Tunisian olive varieties served for olive oil extraction. Olive fruit samples were immediately after harvest carried and stored into the laboratory for further oil extraction. In order to obtain olive oil, 2.5 kg of stoned olives was grinded, and olive oil was extracted by mechanical press. Standard methods commonly used in oil factories were followed in the procedure of monovarietal oil extraction and obtention, including milling, mixing at 25°C

for 30 min, and centrifugation for 3 min at 2000g and the final step for olive oil obtention was by natural decantation. Samples were stored at 4°C into dark glass bottles until analysis.

2.3. DNA Isolation. DNA extraction from olive oil was performed by using the QIAmp DNA tool mini kit (Qiagen) according to the protocol described by Ben Ayed et al. (2012) [12] with slight modifications. DNA quantification was carried out by spectrophotometry (Tecan GENIOS Plus spectrofluorometer) and with Hoechst H33258 dye incorporation. Dilution series of Lamda DNA (D150A Promega) were used with standard calibration. Finally, genomic DNA was diluted in TE buffer (10 mM Tris–HCl pH 8.1 mM EDTA pH 8) and stored at -20°C.

2.4. SNP Genotyping. We considered seven SNPs (FAD2.1, FAD2.3, ANTHO3, CALC, ACP1, SOD, and PAL70) in our study; all SNPs were selected in the coding regions of *FAD2, ANTHO, CALC, SAD1, SOD, and PAL70* genes, all of them being involved in fruits pomology and associated with olive oil composition and therefore easily correlatable to phenotypic characters.

The SNP SOD (insertion/deletion type) was genotyped by a simple polymerase chain reaction followed by revelation through agarose gel electrophoresis, whereas the other six SNPs (FAD2.1, FAD2.3, ANTHO3, CALC, ACP1, and PAL70) were genotyped by a polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) method (Table 1). The PCR product (171 bp) of the SNP (ANTHO3) was digested by *MspI* restriction enzyme (Fermentas, LIFE SCIENCES) at 37°C overnight. This restriction enzyme recognizes the sequence AA/GG. The G-allele carrying PCR product was cleaved once by the enzyme generating two fragments (64 and 107 bp). The PCR product (476 bp) of SNP (CALC) was digested by *BstZI* restriction enzyme (Promega) at 50°C overnight. This restriction enzyme recognizes the sequence CC/GG. The C-allele carrying PCR product was cleaved once by the enzyme producing two fragments (316-160 bp). The two other SNPs (FAD2.1 and FAD2.3) were analyzed using PCR-RFLP. The PCR product 241bp of the SNP FAD2.1 and 240 bp of the SNP FAD2.3 were digested by *BamHI* restriction enzyme (Fermentas, Life Sciences) and *Alw26I*, respectively, at 37°C overnight. The sizes of the restriction fragments of PCR product were 224 and 17 bp and 130 and 110 bp for CC genotype of FAD2.1 SNP and FAD2.3 SNP, respectively. The PCR product (330 bp) of SNP (SAD.1) was digested by *TaqI* restriction enzyme (Vivantis) at 65°C for 16 hours. This restriction enzyme recognizes the sequence CC/TT. The C-allele carrying PCR product was cleaved twice by the enzyme producing four fragments (263, 158, 105, and 67 bp). The PCR product (400 bp) of the SNP (PAL70) was digested by *HinfI* restriction enzyme (Fermentas, Life Sciences). The size of the restriction fragments of PCR product was 308, 52, and 40 bp for AA genotype of PAL70 SNP.

All digestion products were separated by electrophoresis on 3% Nusieve ethidium bromide-stained agarose gels and visualized under UV light.

TABLE 1: Characteristics of SNP studied markers.

Gene name	GenBank Accession Number	SNP code	Tm[a]	H0	He	PIC[α]
Phenylalanine ammonia lyase	AY738639	PAL70 (A/G)	60	0.545	0.396	**0.496**
Calcium Binding Protein	AF078680	SNP-H(C/G)	60	0.727	0.462	0.396
Anthocyanidin synthase	AF384050	SNP-I(G/A)	57	0.636	0.433	0.463
Fatty acid desaturase	AY083163	FAD2.1(T/C)	57	**0.909**	**0.495**	0.165
		FAD2.3(C/G)		0.636	0.433	0.463
Stearoyl-ACP desaturase	U58141	SAD.1(A/G)	57	0.363	0.297	0.463
Cu-Zn-superoxide dismutase	AF426829	SOD (InDel)	57	0.545	**0.495**	**0.595**
Mean				0,545	0.492	0.488

a: annealing temperature for PCR amplification.
α: for each locus the polymorphism content information (PIC).

2.5. Statistical Analysis. The analysis of the correlation between SNP markers and the studied parameters was performed in different steps including numerous statistical methods. In the beginning, the Chi-square test was used to evaluate the differences between the classes of qualitative traits in allele and genotype frequencies. Subsequently, a student test was used for quantitative traits, to assess the significant difference between the means of genotype groups for each SNP. R software packages were used to study the association between SNP markers and quantitative and qualitative parameters. All tests were declared statistically significant when *P values* are <0.05. Thereafter, to study the relationship of the studied seven SNPs with quantitative traits, a variance multiway analysis was carried out. In addition, multinomial logistic regression was applied in order to test the associations of the seven SNPs with qualitative traits independently.

To draw the directed acyclic graph (DAG), we used the R language and the 'grow shrink' algorithm. The algorithm proficiently filters links out of a full skeletal DAG, in which all nodes are primarily connected (excluding those having no relationships with others), based on tests of conditional independence between a pair of nodes given all possible subsets of the rest. Logical rules are applied to create the direction of links (conditional dependence between variables), so that cycles are not introduced and patterns of conditional independence are found in the data match the generated DAG. We predicted association power in the final DAG by calculating approximately the beta-coefficient for a regression for each potential causal effect in which the variable at the base of the arrow ('cause') was considered a covariate, and the variable at the head of the arrow ('effect') was considered the outcome or dependent variable [14].

3. Results and Discussion

3.1. Genotyping and Characteristics of the SNP Markers. The observed heterozygosity for the studied SNP markers ranged from 0.363 (SAD.1) to 0.909 (FAD2.1) (0.545 average), whereas the expected heterozygosity ranged from 0.297 to 0.495 with an average of 0.492 indicating a high level of heterozygosity for all markers (Table 1).

Polymorphism Information Content (PIC) value was determined for the studied table olive cultivars.

Table 1 shows that SOD Indel (PIC=0.595) and PAL70 SNP (0.496) markers have higher PIC values than the other studied markers, meaning that SOD and PAL70 are the most informative markers and therefore able to distinguish between our table olive cultivars. This result, together with mean value observed across the loci and table olive cultivars in the present study, is in accordance with the results previously obtained by Ben Ayed et al. (2014) [15]. SOD gene, for superoxide dismutase, encodes for an antioxidant enzyme that plays a pivotal role in protecting cells against superoxide radicals accumulation [16]. Phenylalanine ammonia lyase PAL70 catalyzes the reaction of trans-cinnamic acid formation *via* L-phenylalanine deamination and is therefore associated with phenolic compounds content in olives [17].

The highest discriminating power (DP) value 0.528 was shown for FAD2.3 marker. The mean value is 0.464. Using SSR markers, Reale et al. (2006) [18] and Muzzalupo et al. (2009) [19] obtained similar values, respectively, with DNA samples extracted from 65 olive cultivars and 39 Italian cultivars (0.38). However, the average values are lower than those found by Cipriani et al. (2002) [20] in 12 Italian cultivars (0.44) using SSR markers (0.71). *Fatty acid desaturase 2* gene which is involved in the biosynthesis of HUFA from PUFA precursor has been demonstrated to be associated with oleic/linoleic acid ratio content of olive oils from Tunisian olive oil cultivars [21].

The allele frequencies of the seven studied SNPs revealed a dominance of all markers, except for PAL70 whose two alleles displayed similar frequencies.

3.2. Association between SNP Polymorphisms and Olive Oil Quality Parameters. In order to illustrate the association between quality of the table olive cultivars and gene information, we applied the likelihood ratio test (LRT). Thus, a genome-wide association was carried out to identify table olive fruit quality susceptibility alleles. We studied 7 SNPs located in 6 genes for 11 table olive samples. Then, we evaluated the p values of the LRT and Ki_2 tests. The results are summarized in Table 2 and demonstrated the absence of any significant associations between the seven SNPs (SOD, FAD2.1, FAD2.3, ANTHO3, CALC, SAD.1, and PAL70) genotypes and none of the qualitative traits is considered in this work.

TABLE 2: Association between ANTHO3 and PAL70 genotypes and qualitative parameters.

Polymorphisms		ANTHO3			PAL70		
Qualitative parameters		GA	AA	Chi-square	AA	AG	Chi-square
		(%)	(%)	LRT*	(%)	(%)	RLT*
Fruit Form	Ovoid	1.9	1.1	3.592	1.1	1.9	1.946
	Elongate	4.5	2.5	0,166*	2.5	4.5	0.378*
	Spherical	0.6	0.4		0.4	0.6	
Stone Form	Elongate	4.5	2.5	5.135	3.2	3.8	3.221
	elliptic	1.3	0.7	0.052*	0.9	1.1	0.2*
	ovoid	1.3	0.7		0.9	1.1	
Symmetry stone	Asymmetric	3.8	2.2	5.238	2.7	3.3	7.639
	Slightly asymmetric	1.9	1.1	0,155*	1.4	1.6	0.054*
	Symmetric	0.6	0.4		0.5	0.5	
	Slightly symmetric	0.6	0.4		0.5	0.5	

Bold values: each variable that has statistical significance for all tests was declared when P values are <0.05.

As shown in Table 3, significant associations between CALC SNP and one parameter which is palmitic acid (C16:0) were found. However, the average rate of C16:0 between the heterozygote varieties with CG-CALC and GG-CALC genotypes ($p = 0.04$) was significantly different. This positive association among CALC polymorphisms and C16:0 parameter suggests that the heterozygote varieties with CG genotypes produce, on average, higher levels of C16:0 than GG genotypes varieties. As shown in Table 5, this significant correlation is proved by multinomial logistic regression modeling ($p = 0.039$).

Regarding the FAD2.1 SNP, a highly significant association with β-sitosterol ($p=0.018$) quantitative parameter is proved. Similar result is generated by multinomial logistic regression ($p=0.022$) (Table 5).

Besides, FAD2.3 SNP was found to be highly associated with three quantitative parameters, namely, acidity ($p=0.024$), rate of carotene ($p=0.013$), and cholesterol content ($p=0.048$) (Table 3). The homozygous varieties (CC-FAD2.3) were the main genotypes concerned by these positive associations. FAD gene is known to be involved in the synthesis pathway of the unsaturated fatty acid [6], suggesting a direct effect of FAD2.3 genotypic variations on the rate of PUFA (such as C18:2 and C18:3) for each variety and hence influenced the acidity parameter. In fact, the homozygous varieties CC (Toffehi, Fakhari, and Fougi cultivars) have a higher acidity than the heterozygous varieties CG (other cultivars). Nevertheless, the cholesterol rate was significantly higher in the varieties carrying the homozygous genotype CC-FAD2.3 (particularly Toffehi with a cholesterol rate of 1.98, and Fakhari with 2.17), than the heterozygous genotypes (Table 3). Moreover, CC varieties contain more carotene pigment than the heterozygous genotypes. Moreover, the FAD2.3 SNP marker is significantly associated with the acidity and the cholesterol rate by using the analysis of variance ($p=0.006$, $p<0.001$) (Table 4) and the multinomial logistic regression modeling ($p=0.026$, $p<0.001$) (Table 5), respectively.

The study of ANTHO3 SNP led to the identification of two genotypes: AA and AG, with a level of AG-ANTHO3

heterozygosity of 63.63 %. A significant association was established with the rate of total sterols ($p = 0.042$), which has a higher average for heterozygous cultivars (AG) (representing 63.63% % of all samples) (Table 3). This significant correlation is proved by multinomial logistic regression modeling ($p = 0.04$) (Table 5).

Moreover, PAL70 SNP is clearly associated with 3 parameters, namely, chlorophyll, total phenol contents, and the oxidative stability (Table 3). However, a variability of the rate of chlorophyll pigment among the heterozygote varieties with AG-PAL70 and AA-PAL70 genotypes ($p=0.002$) was noted. A positive correlation between the total phenol content and the genotype variation for this marker ($p<0.001$) could also be observed, where the varieties with AA genotypes displayed the highest total phenolic content. Regarding the relationship with oxidative stability ($p<0.001$), the homozygous varieties AA behaved with better oil stability than the heterozygous varieties AG. Moreover, the PAL70 SNP marker is significantly associated with the chlorophyll rate ($p=0.031$) by using the analysis of variance (Table 4).

Additionally, multivariate analyses were used to study the association between olive oil parameters and the PAL70 SNP marker, showing an important significant association between this SNP marker and the acidity parameter (Tables 4 and 5). The p values of this association were $p=0.004$ and $p = 0.036$ using the analysis of variance and the multinomial logistic regression modeling, respectively.

The association between total phenol rate and the PAL70 SNP marker is biologically relevant since the PAL70 marker is located in the *L-phenylalanine ammonia lyase* gene that is involved in the biosynthesis of phenylpropanoid compounds [12].

The relationship between the PAL70 SNP and the phenol level was assessed by Bayesian networks modeling. The derived DAG (directed acyclic graph) is shown in Figure 1 where directed arrows indicate the direction of 'causal' influence between variables. Three direct influences are identified: effect of PAL70 marker on the phenol rate, oxidative stability, and chlorophyll content. In fact, Figure 1 shows that the

TABLE 3: Associations between SNP studied markers and table olive oil composition.

(a)

Parameters SNPs		Cholesterol Mean ±SD	P	β-sitosterol Mean ±SD	P	Chlorophyll Mean ±SD	P	Carotene Mean ±SD	P	FWM Mean ±SD	P
CALC	CG	1.52 ±0.34	0.405(0.224)	1387.13 ±221.97	0.315(0.47)	1.71 ±1.17	0.75(0.73)	4.64 ±2.95	0.7(0.52)	3.58 ±2.92	0.641(**0.663**)
	GG	1.34 ±0.12		1573.67 ±360		1.46 ±0.94		3.93 ±0.11		4.55±3.01	
FAD 2.1	CT	1.48 ±0.32	0.908	1383 ±198	**0.018**	1.77 ±1.04	0.244	4.49 ±2.62	0.86	3.43 ±2.6	0.13
	TT	1.44 ±0.32		1981 ±198		0.4 ±1.04		4 ±2.62		8 ±02.6	
FAD 2.3	CC	1.71 ±0.043	0.048(0.186)	1423 ±312	0.89(0.9)	1.75 ±0.78	0.82(0.79)	6.7 ±2.58	**0.013(0.064)**	2.6 ±0.62	0.29(0.18)
	CG	1.34 ±0.078		1446 ±253		1.58 ±1.26		3.16 ±1.30		4.56 ±3.39	
ANTHO3	GA	1.56 ±0.36	0.224(0.128)	1539 ±261	0.087(0.053)	1.2 ±0.806	0.064(0.118)	4.54 ±3.1	0.87(0.84)	4.85 ±3.08	0.123(0.063)
	AA	1.32 ±0.047		1260 ±152		2.42 ±1.13		4.28 ±1.17		2.08 ±1.06	
PAL70	AA	1.31 ±0.07	0.115(0.109)	1342 ±92.266	0.287(0.264)	2.56 ±0.68	**0.002(0.003)**	4.2 ±1.03	0.78(0.76)	2.1 ±0.84	0.055(0.054)
	AG	1.61 ±0.367		1518 ±334.7		0.88 ±0.63		4.65 ±3.38		5.3 ±3.1	
SAD.1	TT	1.345 ±0.07	0.055(0.198)	1424 ±281	0.828(0.825)	1.2 ±0.91	0.064(0.082)	3.06 ±1.2	**0.005(0.039)**	4.85 ±3.16	0.123(0.06)
	CT	1.7 ±0.436		1462 ±258		2.425 ±0.95		6.88 ±2.34		2.08 ±0.45	

(b)

Parameters SNPs		C16:0 Mean ±SD	P	C16:1 Mean ±SD	P	C18:0 Mean ±SD	P	C18:3 Mean ±SD	P	I/S Mean ±SD	P
CALC	CG	15.08 ±2.1	**0.04(0.041)**	1.53 ±0.54	0.057(**0.025**)	1.9 ±0.75	0.686(0.523)	14.65 ±5.1	0.76(0.626)	4.95 ±0.91	0.075(**0.027**)
	GG	11.83 ±1.58		0.8 ±0.30		2.08 ±0.17		12.53 ±4.7		6.08 ±0.45	
FAD 2.1	CT	14.33 ±2.51	0.562	1.36 ±0.61	0.694	1.93 ±0.67	0.815	0.63 ±0.15	0.83	5.21 ±0.98	0.615
	TT	12.75 ±2.51		1.1 ±0.61		2.1 ±0.67		0.6 ±0.15		5.75 ±0.98	
FAD 2.3	CC	15.06 ±0.6	0.39(0.27)	1.73 ±0.32	0.084(**0.049**)	2.25 ±0.92	0.254(0.387)	0.57 ±0.13	0.337(0.32)	4.7 ±0.43	0.206(0.127)
	CG	13.69 ±2.97		1.1 ±0.58		1.77 ±0.39		0.66 ±0.14		5.54 ±1.07	
ANTHO3	GA	14.23 ±2.62	0.95(0.94)	1.15 ±0.56	0.19(0.19)	2.11 ±0.7	0.26(0.22)	0.66 ±0.15	0.54(0.45)	5.14 ±0.96	0.61(0.63)
	AA	14.13 ±2.42		1.65 ±0.53		1.65 ±0.47		0.57 ±0.13		5.46 ±1.03	
PAL70	AA	12.85 ±2.63	0.095(0.118)	1.21 ±0.68	0.542(0.556)	1.88 ±0.42	0.775(0.763)	0.55 ±0.07	0.108(0.097)	5.85 ±0.97	0.054(0.073)
	AG	15.31 ±1.72		1.44 ±0.52		2. ±0.81		0.69 ±0.16		4.76 ±0.63	
SAD.1	TT	14.34 ±2.64	0.809(0.804)	1.25 ±0.63	0.546(0.528)	1.67 ±0.41	0.063(0.148)	0.695 ±0.14	0.046(0.019)	5.271 ±0.86	0.958(0.963)
	CT	13.94 ±2.36		1.48 ±0.53		2.41 ±0.75		0.521±0.048		5.238±1.23	

Bold values: each variable that has statistical significance for all tests was declared when P-values are <0.05.
P: P-value of student test; **SD**: standard deviation; **FWM**: Fruit Weight at Maturation.

(c)

Parameters SNPs		MG/fruit Mean ±SD	P	Polyphenol Mean ±SD	P	Total Sterols Mean ±SD	P	Acidity Mean ±SD	P
CALC	**CG**	0.74 ±0.55	0.54(0.56)	165 ±66	0.21(0.442)	1682±315	0.26(0.35)	0.29 ±0.058	0.308(0.191)
	GG	0.98 ±0.54		249 ±151		1946 ±374		0.25 ±0.03	
FAD 2.1	**CT**	0.72 ±0.49	0.125	199 ±93.23	0.233	1715 ±327	0.245	0.28 ±0.05	0.974
	TT	1.6 ±0.49		74.4 ±93.23		2142 ±327		0.28 ±0.05	
FAD 2.3	**CC**	0.55 ±0.02	0.255(0.151)	163.28 ±70.54	0.54(0.5)	1739 ±428	0.91(0.92)	0.32 ±0.06	0.024(0.1)
	CG	0.95 ±0.63		202.2 ±110.8		1763 ±309		0.25 ±0.02	
ANTHO3	**GA**	1.02 ±0.56	0.075(0.033)	169.64 ±116	0.43(0.32)	1904 ±316	0.042(0.022)	0.28 ±0.06	0.61(0.51)
	AA	0.43 ±0.13		220 ±39.08		1492 ±176		0.27 ±0.01	
PAL70	**AA**	0.5 ±0.19	0.085(0.079)	277 ±57.54	0(0.001)	1676 ±298	0.51(0.5)	0.25 ±0.02	0.123(0.112)
	AG	1.06 ±0.61		113.92 ±36.47		1819 ±378		0.3 ±0.06	
SAD.1	**TT**	0.97 ±0.51	0.18(0.097)	191.2 ±109.53	0.894(0.885)	1741.9±340	0.878(0.883)	0.26 ±10.03	0.098(0.223)
	CT	0.51 ±0.09		182.53 ±82.17		1777 ±380		0.317 ±0.07	

(d)

Parameters SNPs		Δ-5-avenasterol Mean ±SD	P	Oxidative stability Mean ±SD	P	campesterol Mean ±SD	P
CALC	**CG**	194 ±83.93	0.234(0.261)	39.63 ±17.17	0.32(0.42)	54.07 ±15.71	0.78(0.78)
	GG	122.6 ±78.9		52.67 ±21.93		51.04 ±15.4	
FAD 2.1	**CT**	188 ±74.9	0.077	44.7 ±18.69	0.41	51.77 ±14.84	0.323
	TT	31.8 ±74.9		28 ±18.69		68.04 ±14.84	
FAD 2.3	**CC**	205.75 ±88.8	0.38(0.4)	40.75 ±11.14	0.76(0.71)	58.72 ±15.57	0.38(0.4)
	CG	156.8 ±84.84		44.57 ±22.31		50.12 ±14.76	
ANTHO3	**GA**	188 ±105	0.51(0.4)	37.29 ±20.64	0.17(0.097)	58.56 ±16.65	0.12(0.06)
	AA	150 ±26		53.5 ±7.5		43.95 ±2.28	
PAL70	**AA**	162.2 ±20.62	0.68(0.66)	60.2 ±6.14	0.000	43.4 ±3.85	0.037(0.039)
	AG	184.97 ±118.17		29 ±10.88		61.45 ±15.96	
SAD.1	**TT**	156.9 ±84.78	0.393(0.41)	42.29 ±21.39	0.844(0.826)	50.1 ±14.77	0.381(0.398)
	TC	205.5 ±89.1		44.75 ±14.56		58.77 ±15.51	

TABLE 4: *P values* of Fisher tests for the association study between SNP markers and the oil quality characteristics.

Model	SOD	CALC	ANTHO3	FAD2.1	FAD2.3	PAL70	SAD.1
FWM	--	0.894	**0.022**	--	**0.008**	**0.017**	**0.012**
MG/fruit	--	0.955	**0.028**	--	**0.009**	**0.027**	**0.018**
acidity	--	0.341	0.082	--	**0.006**	**0.004**	**0.001**
C16:0	--	0.877	0.782	--	**0.011**	0.844	0.606
C16:1	--	0.088	0.783	--	0.215	0.640	0.351
C18:0	--	0.227	0.853	--	0.175	0.140	0.205
C18:1	--	0.799	0.562	--	0.059	0.364	0.294
C18:2	--	0.963	0.336	--	0.182	0.289	0.185
C18:3	--	**0.033**	0.651	--	0.687	**0.026**	0.058
C20:0	--	0.107	0.910	--	0.079	0.149	0.174
I/S	--	0.557	0.899	--	0.066	0.794	0.592
Polyphenol	--	**0.035**	0.048	--	0.18	0.895	0.487
Oxidative Stability	--	0.696	0.054	--	**0.012**	0.935	0.229
chlorophyll	--	0.625	0.832	--	0.168	**0.031**	0.9
carotene	--	0.065	0.169	--	0.188	0.065	0.195
β-sitosterol	--	0.275	0.273	--	0.503	0.632	0.880
Campesterol	--	0.458	0	--	0.702	0.56	0.715
Δ-5-avenasterol	--	0.906	0.136	--	0.844	0.991	0.838
cholesterol	--	0.125	**0.018**	--	**0**	**0**	**0**
Total Sterols	--	0.694	0.182	--	0.564	0.92	0.956
(Uvaol+erythro)/total sterol	--	0.744	0.217	--	0.172	0.292	0.095

Bold values: each variable that has statistical significance for all tests was declared when *P values* are < 0.05; *: each variable that has statistical significance for all tests was declared when *P values* are <0.05 and has biological relevance.

TABLE 5: *P* values given by the binary logistic regression analysis.

Model	SOD	CALC	ANTHO3	FAD2.1	FAD2.3	PAL70	SAD.1
FWM	--	0.599	0.102	0.107	0.246	0.134	0.102
MG/fruit	--	0.493	0.064	0.103	0.213	0.183	0.149
acidity	--	0.261	0.565	0.970	**0.026**	0.036	0.082
C16:0	--	**0.039**	0.943	0.514	0.345	0.682	0.784
C16:1	--	0.051	0.159	0.657	0.071	0.462	0.497
C18:0	--	0.647	0.226	0.791	0.212	0.231	0.055
C18:1	--	0.146	0.519	0.173	0.710	0.351	0.607
C18:2	--	0.501	0.497	0.144	0.778	0.261	0.451
C18:3	--	0.728	0.321	0.808	0.288	0.147	**0.043**
C20:0	--	0.889	0.629	0.914	0.138	0.317	0.246
I/S	--	0.065	0.574	0.570	0.170	0.870	0.952
polyphenol	--	0.176	0.378	0.194	0.498	0.795	0.880
oxidative Stability	--	0.273	0.141	0.365	0.729	0.407	0.823
chlorophyll	--	0.723	0.057	0.203	0.798	0.420	0.057
carotene	--	0.662	0.857	0.844	**0.017**	0.096	**0.010**
β-sitosterol	--	0.267	0.074	**0.022**	0.884	0.439	0.805
Campesterol	--	0.753	0.101	0.275	0.334	0.390	0.33
Δ-5-avenasterol	--	0.194	0.462	0.066	0.338	0.208	0.342
cholesterol	--	0.354	0.186	0.896	**0.044**	0.096	**0.049**
total sterols	--	0.225	**0.04**	0.204	0.905	0.054	0.861
(Uvaol+erythro)/sterol totaux	--	0.803	0.913	0.283	0.389	0.962	0.584

Bold values: each variable that has statistical significance for all tests was declared when *P values* are < 0.05; *: each variable that has statistical significance for all tests was declared when *P values* are <0.05 and has biological relevance.

TABLE 6: Pearson's correlations of SAD.1 marker with fatty acid compositions and the PAL70 marker with olive oil parameters of the studied olive oil cultivars.

Parameters		SAD.1
	r	p
Acidity	0.684	**0.02**
C16:0	-0.083	0.809
C16:1	0.205	0.546
C18:0	0.644	**0.032**
C18:1	0.155	0.649
C18:2	-0.227	0.501
C18:3	-0.610	**0.046**
C20:0	0.350	0.291
		PAL70
	r	p
Phenol	-0.886	**<0.001**
Oxidative Stability	-0.884	**<0.001**
Chlorophyll	-0.814	**0.002**
Sterol	0.223	0.510

Bold values: each variable that has statistical significance was declared when P values are <0.05.
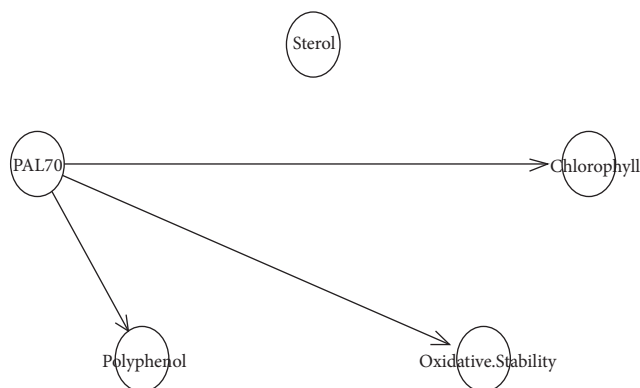P: P value. r: correlation coefficient.

FIGURE 1: Directed acyclic graph representing possible PAL70 SNP marker connections with total phenol, oxidative stability, and chlorophyll.

PAL70 SNP was negatively influenced by the phenol rate (r=-0.886; p<0.001), the oxidative stability (r=-0.884; p<0.001), and the chlorophyll (r=-0.814; p=0.002). Furthermore, PAL70 node was not influenced by the sterol level (r=0.223; p=0.510).

The oxidative stability of the olive oil samples is directly influenced by the total phenol level. Besides, total phenol amount is directly influenced by the PAL70 marker. The latter plays a key role in the total phenol level of each of the olive oil varieties. This finding could be explained by the fact that PAL70 SNP is located within a gene involved in the phenolic biosynthesis (Balsa et al. 1979), suggesting the direct effect of the PAL70 genotype variations on the percentage of total phenol for each variety.
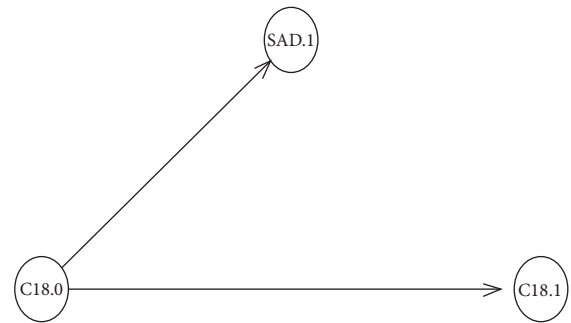

FIGURE 2: Directed acyclic graph representing possible SAD.1 SNP marker connections with stearic acid C18:0 and oleic acid C18:1.

For SAD.1 SNP study, two genotypes were identified: TT and CT. About 64 % of the varieties were homozygous TT-SAD.1 and including both two dual-use cultivars (Chemchali and Oueslati) and two table olive cultivars (Toffehi and Fakhari). Two significant associations of this marker were shown with the accumulation of the linolenic unsaturated fatty acids (C18:3; $p=0.046$) and with the rate of carotene ($p=0.005$) (Table 3).

The two significant correlations between SAD.1 SNP marker and the content of carotene pigments (p=0.01) and C18:3 (p=0.043) are confirmed by multinomial logistic regression modeling (Table 5).

The relationship between the molecular marker SAD.1 and fatty acid composition was also analyzed by Bayesian networks modeling.

Firstly, 3 nodes were considered as represented in Figure 2. Pearson correlation coefficients among fatty acid compositions in olive oil varieties are presented in Table 6. Moreover, SAD.1 was positively influenced by the saturated stearic acid C18:0 ($r = 0.644$; $p = 0.032$).

SAD gene is known to be associated with the transformation of the saturated stearic fatty acid C18:0 to the monounsaturated oleic fatty acid C18:1, therefore, suggesting the direct effect of the SAD.1 genotype variations on the fatty acid content [6].

4. Conclusions

SNP genotyping is a valuable approach for marker assisted selection in crops. For this reason, we studied in this current work the correlations between the six SNPs and table olive oils quality parameters and their usefulness in the traceability of Tunisian table olive oil. We revealed that PAL70 SNP marker was negatively influenced by the phenol rate (r=-0.886; p<0.001) and the oxidative stability (r=-0.884; p<0.001). Besides, we reported a significant association of SAD.1 marker with the accumulation of the linolenic unsaturated fatty acid (C18:3; $p=0.046$) and that SAD.1 was positively influenced by the saturated stearic acid C18:0 (r=0.644; p=0.032) based on multinomial logistic regression and Bayesian networks modeling, respectively. To the best of our knowledge, this is the first work that analyses the SNP markers of Tunisian table olive oil and the quality of the oil.

Data Availability

All data generated or analyzed during this study are included in this published article.

Ethical Approval

This paper complies with ethical standard.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this article.

Acknowledgments

References

[1] IOOC, 2018, http://www.internationaloliveoil.org/estaticos/view/132-world-table-olive-figures.

[2] R. Ben Ayed, K. Ennouri, F. Ben Amar, F. Moreau, M. A. Triki, and A. Rebai, "Bayesian and phylogenic approaches for studying relationships among table olive cultivars," *Biochemical Genetics*, vol. 55, no. 4, pp. 300–313, 2017.

[3] R. Ben Ayed, N. Kamoun, F. Moreau, and A. Rebaï, "Comparative study of microsatellite profiles of DNA from oil and leaves of two Tunisian olive cultivars," *European Food Research and Technology*, vol. 229, no. 5, pp. 757–762, 2009.

[4] R. Ben-Ayed, N. Grati, and A. Rebai, "An overview of the authentication of olive tree and oil," *Comprehensive Reviews in Food Science and Food Safety*, vol. 12, no. 2, pp. 218–227, 2013.

[5] R. Ben Ayed, H. Ben Hassen, K. Ennouri, R. Ben Marzoug, and A. Rebai, "OGDD (Olive Genetic Diversity Database): A microsatellite markers' genotypes database of worldwide olive trees for cultivar identification and virgin olive oil traceability," *Database*, vol. 2016, Article ID ebav090, 2016.

[6] V. Vance and H. Vaucheret, "RNA silencing in plants - Defense and counterdefense," *Science*, vol. 292, no. 5525, pp. 2277–2280, 2001.

[7] R. C. Wilmouth, J. J. Turnbull, R. W. D. Welford, I. J. Clifton, A. G. Prescott, and C. J. Schofield, "Structure and mechanism of anthocyanidin synthase from Arabidopsis thaliana," *Structure*, vol. 10, no. 1, pp. 93–103, 2002.

[8] K. D. Hirschi, "Expression of Arabidopsis CAX1 in tobacco: Altered calcium homeostasis and increased stress sensitivity," *The Plant Cell*, vol. 11, no. 11, pp. 2113–2122, 1999.

[9] P. M. Schlüter, S. Xu, V. Gagliardini et al., "Stearoyl-acyl carrier protein desaturases are associated with floral isolation in sexually deceptive orchids," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 108, no. 14, pp. 5696–5701, 2011.

[10] K. Apel and H. Hirt, "Reactive oxygen species: metabolism, oxidative stress, and signal transduction," *Annual Review of Plant Biology*, vol. 55, pp. 373–399, 2004.

[11] J. Flexas, J. Bota, J. Galmés, H. Medrano, and M. Ribas-Carbó, "Keeping a positive carbon balance under adverse conditions: Responses of photosynthesis and respiration to water stress," *Physiologia Plantarum*, vol. 127, no. 3, pp. 343–352, 2006.

[12] C. Balsa, G. Alibert, J. Brulfert, O. Queiroz, and A. M. Boudet, "Photoperiodic control of phenolic metabolism in Kalanchoe blossfeldiana," *Phytochemistry*, vol. 18, no. 7, pp. 1159–1163, 1979.

[13] R. Ben-Ayed, N. Grati-Kamoun, C. Sans-Grout, F. Moreau, and A. Rebai, "Characterization and authenticity of virgin olive oil (*Olea europaea* L.) cultivars by microsatellite markers," *European Food Research and Technology*, vol. 234, no. 2, pp. 263–271, 2012.

[14] R. Ben Ayed, K. Ennouri, H. B. Hlima et al., "Identification and characterization of single nucleotide polymorphism markers in *FADS2* gene associated with olive oil fatty acids composition," *Lipids in Health and Disease*, vol. 16, no. 1, 2017.

[15] R. Ben Ayed, I. Kallel, H. Ben Hassen, and A. Rebai, "SNP marker analysis for validating the authenticity of Tunisian olive oil," *Journal of Genetics*, vol. 93, no. 3, pp. e148–e154, 2014.

[16] A. Zafra, A. J. Castro, and J. d. Alché, "Identification of novel superoxide dismutase isoenzymes in the olive (*Olea europaea* L.) pollen," *BMC Plant Biology*, vol. 18, no. 1, 2018.

[17] F. Ortega-García and J. Peragón, "Phenylalanine ammonia-lyase, polyphenol oxidase, and phenol concentration in fruits of olea europaea L. cv. picual, verdial, arbequina, and frantoio during ripening," *Journal of Agricultural and Food Chemistry*, vol. 57, no. 21, pp. 10331–10340, 2009.

[18] S. Reale, S. Doveri, A. Díaz et al., "SNP-based markers for discriminating olive (*Olea europaea* L.) cultivars," *Genome*, vol. 49, no. 9, pp. 1193–1205, 2006.

[19] I. Muzzalupo, F. Stefanizzi, and E. Perri, "Evaluation of olives cultivated in southern italy by simple sequence repeat markers," *HortScience*, vol. 44, no. 3, pp. 582–588, 2009.

[20] G. Cipriani, M. T. Marrazzo, R. Marconi, A. Cimato, and R. Testolin, "Microsatellite markers isolated in olive (*Olea europaea* L.) are suitable for individual fingerprinting and reveal polymorphism within ancient cultivars," *Theoretical and Applied Genetics*, vol. 104, no. 2-3, pp. 223–228, 2002.

[21] R. Ben Ayed, K. Ennouri, S. Ercişli et al., "First study of correlation between oleic acid content and *SAD* gene polymorphism in olive oil samples through statistical and bayesian modeling analyses," *Lipids in Health and Disease*, vol. 17, no. 1, 2018.

Research Article

Morphological Seed Characterization of Common (*Phaseolus vulgaris* L.) and Runner (*Phaseolus coccineus* L.) Bean Germplasm: A Slovenian Gene Bank Example

Lovro Sinkovič ⓘ,[1] Barbara Pipan,[1] Eva Sinkovič,[2] and Vladimir Meglič[1]

[1]*Crop Science Department, Agricultural Institute of Slovenia, Hacquetova Ulica 17, 1000 Ljubljana, Slovenia*
[2]*Department of Agronomy, Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, 1000 Ljubljana, Slovenia*

Correspondence should be addressed to Lovro Sinkovič; lovro.sinkovic@kis.si

Genetic resources comprised of 953 accessions of common (*Phaseolus vulgaris* L.) and 47 accessions of runner (*Phaseolus coccineus* L.) bean from the national Slovene gene bank were characterized using fourteen morphological seed descriptors. Seeds of each accession were evaluated for six quantitative characteristics: seed length, seed thickness, seed width, seed length/width ratio, seed width/thickness ratio, and 100 or 10 seed weight. Furthermore, seeds were evaluated using eight qualitative characteristics: seed colour; number of seed colours; primary/main seed colour; predominant secondary seed colour; distribution of secondary seed colour; seed veining; seed shape; and seed colour (primary and secondary) and coat pattern. For each, common, and runner bean collection, first four components within principal component analysis explained 75.03% and 80.16% of morphological variability, respectively. Regarding Ward's method and squared Euclidian distance, three clusters with the most distinct characteristics were established for each species. The results of morphological seed characterization indicate the origin (Andean, Mesoamerican, putative hybrids between gene pools) and domestication pathways of common and runner bean. This is the first study describing morphological seed characteristics of the entire common and runner bean germplasm conserved in one of the Central European bean collections. The results obtained in this study are serving as the useful information on genetic diversity of common and runner bean accessions at the Slovene gene bank, which could be used for development of new bean varieties for studied seed characteristics.

1. Introduction

Genetic diversity or variation between different populations belonging to the same genus resulted from the evolution of crops through the history, in response to different environments and husbandry practices [1]. *Phaseolus* spp. beans are valued grain legumes or pulse crops of worldwide importance in terms of human and animal consumption [2–4]. Common bean (*Phaseolus vulgaris* L.) is the most important *Phaseolus* spp. worldwide, while the runner bean (*Phaseolus coccineus* L.) is the third, right after lima bean (*Phaseolus lunatus* L.) [5, 6]. Cultivated common and runner beans were domesticated independently within two centres of diversity, giving rise to two gene pools, i.e., Mesoamerican and Andean [4, 7]. The lack of characterization of the *Phaseolus coccineus*

L. germplasm restricts its utilization as donor species for interspecific hybridization and consequently limits its use in other *Phaseolus* spp. breeding programs, i.e., common bean [6].

Ensuring the preservation of future sources is a big challenge for plant geneticists and breeders. Seed gene banks are intended to enable the conservation of the world's crop genetic diversity against the genetic erosion of crops as an unintended consequence of the global uptake of new high-yielding green revolution agricultural varieties [8]. Seed gene banks are facilities dedicated to the medium-term storage, i.e., for a few decades in storage at 5–10°C, or long-term storage, i.e., for many decades in storage at −18 ± 3°C, of samples of seeds as a means of conserving crop or species diversity. A seed gene bank conserving crop varieties is often

called a gene bank as the seed samples, i.e., accessions, are used as a source of genes conferring desirable characteristics [9].

Genetic diversity of common bean from Central Europe was studied at the Agricultural Institute of Slovenia by AFLP and microsatellite markers [10–12]. The surveys revealed that extensive diversity resides in common bean cultivated in this area and includes variation beyond the two gene pools, i.e., Andean and Mesoamerican. As revealed by the analysis of a large set of European common bean landraces using chloroplast microsatellites (cpSSRs) and two unlinked nuclear loci (phaseolin; *PvSHP1*) a relatively high proportion of the European bean germplasm might have derived from hybridization between the two gene pools [13, 14]. Recently, 300 Croatian common bean landraces have been evaluated by phaseolin and microsatellite markers. The majority of the studied landraces belong to Andean gene pool, similarly as was found for Slovenian landraces [10]. Genetic diversity of runner bean has been less extensively investigated. The largest set of European landraces, more than 300, was evaluated by cpSSRs and a smaller set was studied also for phenotypic traits [15].

Two main gene pools, Mesoamerican and Andean associated with these two geographical areas, have been described in wild and cultivated common beans. As a result of the domestication process, a great number of varieties showing differences in morpho-agronomical quantitative traits including seed size, seed quality, and plant growing period, were obtained, and this variation has been extensively used in breeding programs or diversity studies [16]. Large or medium seed morphology characteristic was reported for Andean and mostly small seeds for Mesoamerican group genotypes [17]. Purple seed colour was found to be exclusively Andean, while pink, brown, and black predominantly Mesoamerican pool origin. Cream, yellow, and red seed colours were found in both gene pool groups. Andean beans having a tendency for higher iron seed concentration and lower seed zinc concentration than Mesoamerican and putative hybrids between gene pools [18].

In Slovenia, the majority of the bean production is based on local populations and varieties grown by small-scale farmers in low input production systems. Populations are well adapted to the specific growing conditions and microclimate agroenvironments and show a great seed morphological diversity. The aim of the present study was to evaluate the common and runner bean germplasm collections from Slovene gene bank using seed characteristics and to establish clusters of the most distinct genetic resources, based on morphological seed evaluation. Common and runner bean information based on seed descriptors would be the first step for the assessment, description and classification of large bean germplasm collections, to enable better and faster use in future research and development of new varieties.

2. Materials and Methods

2.1. Plant Material.
The plant material used in the present study comprised a total of 1000 accessions of bean germplasm from Slovene gene bank at the Agricultural Institute of Slovenia (46°03′N, 14°31′E; 297 m above sea level), being part of the National plant gene bank. Majority of the studied bean germplasm, i.e., 936 accessions, is Slovenian origin, while the minority Slovakian (60 accessions) or Hungarian origin (4 accessions). Bean germplasm, which consists of common and runner bean collections, was evaluated using morphological seed characteristics. Several *Phaseolus* spp. seed descriptors, which have been adopted by International Union for the Protection of New Varieties of Plants (UPOV), Community Plant Variety Office (CPVO), International Board for Plant Genetic Resources (IBPGR), and Improvement of sustainable Phaseolus production in Europe for human consumption (Phaselieu), were applied to characterize a total of 953 accessions of common and 47 accessions of runner bean. Both bean germplasm collections are conserved in glass jars under medium-term storage conditions at the temperature of 4°C.

Additionally, we have also observed 24 accessions with known phaseolin type (previously determined by SDS-PAGE) as standards/anchors for the Andean/Mesoamerican/mixed germplasm within whole collection. Standards for the Andean germplasm (9) are representing the accessions with T ("Tandergreen") phaseolin type and standards for Mesoamerican germplasm (7) are representing the accessions with S ("Sanilac") phaseolin type. Subset group with putative hybrids between both gene pools (8) are associated with C ("Contender") phaseolin type which originated from Andean gene pool but also represents mixed germplasm between combination of T and S phaseolin types. Table with the standard/anchor accessions are added to the supplementary data (Table S1).

2.2. Seed Evaluation Using Quantitative Characteristics.
Quantitative seed descriptors included the evaluation of the following six characteristics: seed length (L) [mm]; seed thickness (T) [mm]; seed width (W) [mm]; seed length/width ratio (L/W); seed width/thickness ratio (W/T); and 100 (for common bean) or 10 (for runner bean) seed weight [g]. For each accession the three principal axial dimensions (L, T, W) on 10-randomly selected fully developed undamaged seeds were measured using a digital Vernier calliper (Mitutoyo 500-181-30, Japan) reading to 0.1 mm. L was measured as the highest parallel to the hilum, T as lowest parallel to the hilum, and W from hilum to the opposite side. The L/W and W/T were calculated from raw seed size data. For common bean the 100 seed weight was measured in four repetitions using an electronic seed counter (Condator, Pfeuffer GmbH, Germany) and electronic balance weighing to 0.01 g. Since there was a limited quantity of seeds available in runner bean germplasm collection, the 10 seed weight within each accession was measured in one repetition only.

2.3. Seed Evaluation Using Qualitative Characteristics.
Qualitative seed descriptors included visual assessment of the following eight characteristics: seed colour (light, dark, colour mixture); number of seed colours (one, two, more than two); primary/main seed colour (white, green, grey, yellow, beige, brown, red, violet, black, and additionally beige/brown,

TABLE 1: Summary statistics for six quantitative seed characteristics in 953 accessions of common (*Phaseolus vulgaris* L.) and 47 accessions of runner bean (*Phaseolus coccineus* L.).

Characteristics	Range	Mean ± SD	CV (%)
Common bean			
L (mm)	8.06 – 19.98	13.63 ± 1.86	13.63
T (mm)	4.24 – 9.56	6.70 ± 0.94	13.97
W (mm)	5.06 – 12.63	8.22 ± 1.08	13.09
L/W	1.08 – 2.64	1.69 ± 0.28	16.57
W/T	0.97 – 2.49	1.24 ± 0.16	12.64
100 seed weight (g)	19.32 – 98.39	51.13 ± 13.84	27.06
Runner bean			
L (mm)	17.50 – 27.16	20.53 ± 2.06	10.05
T (mm)	5.16 – 10.95	8.70 ± 0.89	10.25
W (mm)	10.03 – 16.46	12.71 ± 1.17	9.20
L/W	1.42 – 1.89	1.62 ± 0.10	5.88
W/T	1.27 – 2.17	1.48 ± 0.16	10.79
10 seed weight (g)	7.56 – 26.70	13.47 ± 3.45	25.62

SD, standard deviation; CV, coefficient of variation; L, seed length; T, seed thickness; W, seed width; L/W, seed length/width ratio; W/T, seed width/thickness ratio.

beige/violet and beige/grey); predominant secondary seed colour (grey, yellow, beige, brown, red, violet, black, and additionally white, green or greenish, light/dark brown and brown/red); distribution of secondary seed colour (around hilum, on half of seed, on entire seed); seed veining (weak, medium, strong); seed shape (round/circular, oval/circular to elliptic, cuboid/elliptic, kidney shaped, truncated); seed colour (primary and secondary) and coat pattern (assessment includes a description of seed colour, shape, size and code) [19–21].

2.4. Statistical Analysis. The differences among the common and runner bean accessions were analysed through a general linear model procedure and least-squares mean tests (Statgraphics Centurion XVI 2009), at 0.05 level of significance. Statistics included mean, range (min, max), standard deviation (SD), coefficient of variation (CV), and analysis of variance (ANOVA). Correlations were obtained as Pearson's correlation coefficients. Principal component analysis (PCA) and cluster analysis were performed to reveal the most influential seed characteristics that discriminated among the accessions. Biplots for common and runner bean germplasms were constructed for two principal components, showing the accessions and the most influential seed characteristics. Dendrograms were conducted to combine individual variables into larger clusters using Ward's method and squared Euclidian distance.

3. Results and Discussion

3.1. Seed Characterization of Common Bean (Phaseolus vulgaris L.) Germplasm. A total of fourteen seed morphological characteristics, i.e., quantitative (L; T; W; L/W; W/T; and 100 seed weight) and qualitative (seed colour; number of seed colours; primary/main seed colour; predominant secondary seed colour; distribution of secondary seed colour; seed veining; seed shape; and seed colour (primary and secondary) and coat pattern), were evaluated. In the upper part of Table 1 are presented summary statistics for six quantitative common bean seed characteristics. The mean values for seed size characteristics based on all 953 accessions in common bean germplasm collection were for L 13.63 ± 1.86 mm, for T 6.70 ± 0.94 mm and for W 8.22 ± 1.08 mm (see Table 1). The minimum and maximum values ranged between 8.06 and 19.98 mm for L, 4.24 and 9.56 mm for T, and 5.06 and 12.63 mm for W. For characteristic 100 seed weight the mean value of all 953 accessions was 51.13 ± 13.84 g, while the minimum and maximum values ranged between 19.32 and 98.39 g. The highest coefficient of variation was calculated for the 100 seed weight (27.06%) and the lowest for W/T (12.64%). As discussed by Rana et al. [22] for 4274 common bean accessions conserved in Indian gene bank, L ranged 5.0–20.3 mm, W 2.0–12.0 mm and 100 seed weight 3.5–96.3 g. Similarly, Kara et al. [23] reported for 12 registered Turkish common bean genotypes L 9.1–17.8 mm, W 5.8–10.0 mm, T 4.6–6.0 mm and 100 seed weight 18.0–65.6 g. As discussed by Giurcă [24] for 9 common beans originating from northern Romania and western Ukraine, L ranged 11.8–18.0 mm, W 7.4–9.7 mm, T 4.4–6.9 mm and 100 seed weight 34.3–54.2 g. Logozzo et al. [4] evaluated 533 accessions of the European common bean germplasm and reported accessions with L 12.0–13.9 mm (35.5%), W 7.1–8.0 mm (33.0%) and T 5.0–5.9 mm (37.1%) were the most frequent.

All seed characteristics measured quantitatively or assessed qualitatively showed wide range of variation among all common bean accessions evaluated. Frequency distribution graphs of 953 common bean accessions for quantitative seed characteristics are shown in Figure 1. Figure 2 shows photo examples for common bean accessions groups distribution according to L (a, b, c) and 100 seed weight (d, e, f). Based on quantitative measurements, the common bean accessions were classified according to the
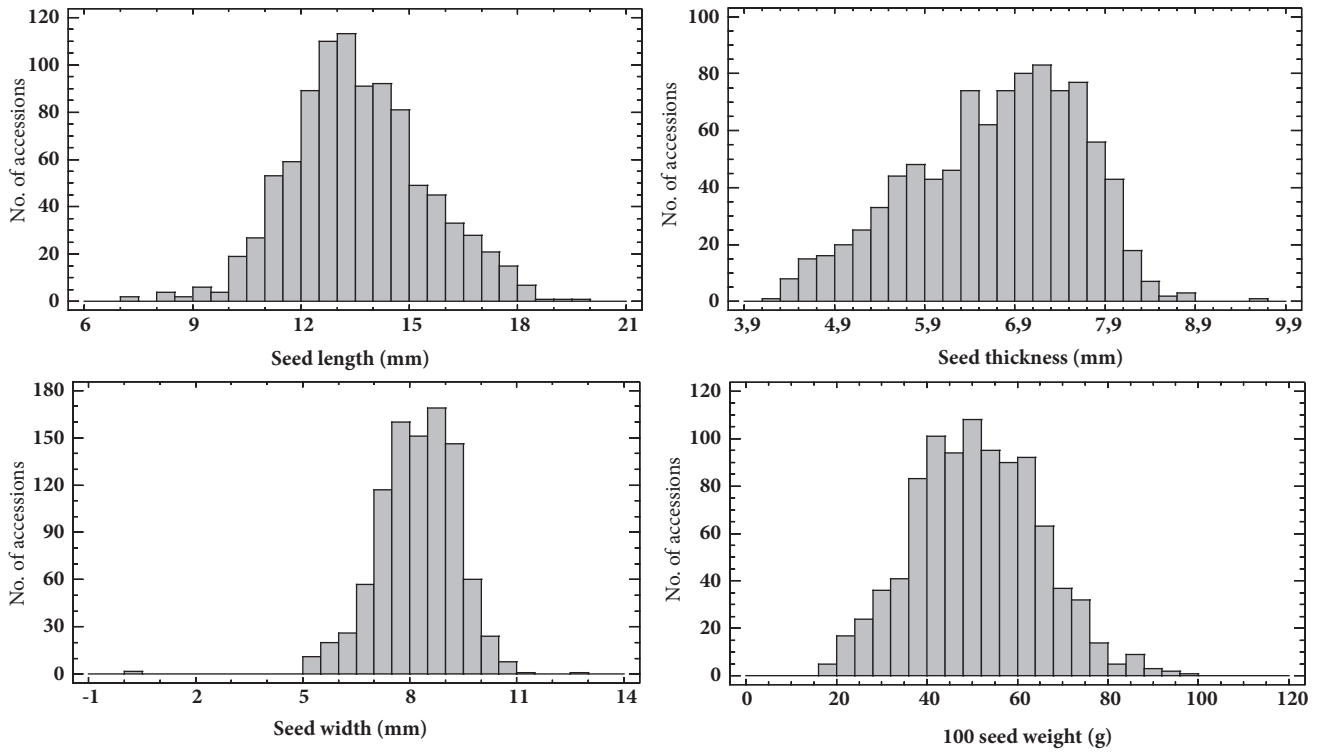
Figure 1: Frequency distribution of 953 common bean accessions (*Phaseolus vulgaris* L.) for quantitative seed characteristics.

(a)

(b)

(c)

(d)

(e)

(f)

Figure 2: Photo examples for common bean accessions (*Phaseolus vulgaris* L.) groups distribution according to seed length: (a) small, (b) medium, and (c) large; and 100 seed weight: (d) low-weight, (e) medium-weight, and (f) high-weight.

L into three groups, i.e., small, medium and large (see Figure 2). The first group included accessions with small seeds and the L < 10.0 mm (18 accessions or 2%); the second group accessions with medium seeds measuring from 10.0 to 15.0 mm (734 accessions or 77%); and the third group accessions with large seeds and L > 15.0 mm (201

accessions or 21%). Similarly, the common bean accessions were classified according to the 100 seed weight into three groups, i.e., low-weight, medium-weight and high-weight (see Figure 2). Low-weight seeds group included common bean accessions with 100 seed weight < 35.0 g (112 accessions or 12%); the medium-weight seeds group accessions with 100

seed weight measuring from 35.0 to 75.0 g (801 accessions or 84%); and the high-weight seeds group accessions with 100 seed weight > 75.0 g (40 accessions or 4%).

Frequency distribution graphs of 953 common bean accessions for qualitative seed characteristics are shown in Figure 3. For characteristic seed colour 479 common bean accessions or 50.3% belonged to the colour mixture class, 251 accessions or 26.3% had dark-coloured seeds (brown, purple, black, etc.), and 223 accessions or 23.4% light-coloured seeds (white, pale yellow, pale brown, etc.). For characteristic number of seed colours 474 common bean accessions or 49.7% had one colour, 446 accessions or 46.8% two colours, and 33 accessions or 3.5% more than two colours per seed. The most abundant primary/main seed colours among all 953 common bean accessions were brown (300 accessions), beige (222 accessions), white (151 accessions), black (79 accessions) and red (66 accessions). The most abundant predominant secondary seed colours among 479 common bean accessions with colour mixture were red (139 accessions), beige (85 accessions), brown (80 accessions) and violet (70 accessions). Distribution of secondary seed colour was for 398 accessions on the entire seed, for 53 accessions around the hilum and for 28 accessions on the half of seed. The seed shape was oval/circular to elliptic for 315 accessions, cuboid/elliptic for 296 accessions and truncated for 196 accessions. For characteristic seed veining 596 accessions had weak, 195 accessions medium and 162 accessions strong seed veining. The characteristic seed colour (primary and secondary) and coat pattern was divided into nine classes with additional subclasses. Among one-colour class the most accessions were categorized according to Phaselieu descriptor into B 13, i.e., brown round large seeds (110 accessions), followed by WH 22, i.e., white long medium seeds (59 accessions), and BL 12, i.e., black round medium (58 accessions). A total of 455 accessions were classified into bicolour class, among which 261 were Pinto type (BIP), 103 broad striped type and 91 constant mottled type. Finally, a total of 22 accessions were classified into tricolour class. Regarding results by Logozzo et al. [4] on 533 accessions of the European common bean germplasm, the accessions with cuboid seed shape (34.9%), maroon (i.e., dark brownish red; 44.3%) and white (28.1%) seed darker colour were the most frequent. Additionally, the maroon seeds were 90.3% Andean, while the white seeds were 52% phaseolin "S" types. As discussed by Piergiovanni and Lioi [25] among 168 Italian common bean landraces according to the seed coat types, white coat seed populations and Borlotto types were the most represented.

3.2. Seed Characterization of Runner Bean (Phaseolus coccineus L.) Germplasm. A total of fourteen seed characteristics, i.e., qualitative (L; T; W; L/W; W/T; and 10 seed weight) and quantitative (same as for common bean), were evaluated. In the lower part of Table 1 are presented summary statistics for six quantitative runner bean seed characteristics. The mean values for seed size characteristics based on all 47 accessions in runner bean germplasm collection were for L 20.53 ± 2.06 mm, for T 8.70 ± 0.89 mm and for W 12.71 ± 1.17 mm (see Table 1). The minimum and maximum values ranged between 17.50 and 27.16 mm for L, 5.16 and 10.95 mm

for T, and 10.03 and 16.46 mm for W. For characteristic 10 seed weight was the mean value of all 47 accessions 13.47 ± 3.45 g, while the minimum and maximum values ranged between 7.56 and 26.70 g. The highest coefficient of variation was calculated for 10 seed weight (25.62%) and the lowest for L/W (5.88%). As discussed elsewhere [6, 24] runner bean L ranged 18.6–25.0 mm, W 12.1–14.0 mm; T 7.7–12.3 mm and 10 seed weight 10.0–14.0 g.

Similarly to common bean, all seed characteristics measured quantitatively or assessed qualitatively showed wide range of variation among runner bean accessions evaluated. Frequency distribution graphs of 47 runner bean accessions for quantitative seed characteristics are shown in Figure 4. Figure 5 shows photo examples for runner bean accessions groups distribution according to L (a, b, c) and 10 seed weight (d, e, f). Based on the quantitative measurements, the accessions of runner bean were classified according to the L into three groups, i.e., small, medium and large (see Figure 5). The first group included runner bean accessions with small seeds and the L < 20.0 mm (20 accessions or 43%); the second group accessions with medium seeds measuring from 20.0 to 25.0 mm (23 accessions or 49%); and the third group accessions with large seeds and L > 25.0 mm (4 accessions or 8%). Similarly were accessions of runner bean classified according to the 10 seed weight into three groups, i.e., low-weight, medium-weight and high-weight (see Figure 5). Low-weight seeds group included common bean accessions with 10 seed weight < 10.0 g (4 accessions or 9%); the medium-weight seeds group accessions with 10 seed weight measuring from 10.0 to 20.0 g (41 accessions or 87%); and the high-weight seeds group accessions with 10 seed weight > 20.0 g (2 accessions or 4%).

Frequency distribution graphs of 47 runner bean accessions for qualitative seed characteristics are shown in Figure 6. For characteristic seed colour 36 runner bean accessions belonged to the colour mixture class, 10 accessions had light-coloured seeds and 1 accession dark-coloured seeds. For characteristic number of seed colours 11 runner bean accessions had one colour and 36 accessions two colours per seed. The most abundant primary/main seed colours among all 47 runner bean accessions were black (12 accessions), brown (11 accessions) and white (11 accessions). The most abundant predominant secondary seed colours among 36 runner bean accessions with colour mixture were brown (16 accessions) and black (13 accessions). Distribution of secondary seed colour was for 35 accessions on the entire seed and for 1 accession around the hilum. The seed shape was cuboid/elliptic for 25 runner bean accessions, oval/circular to elliptic for 14 accessions, kidney shaped for 7 accessions and truncated for 1 accession. For characteristic seed veining 38 accessions had weak, 8 accessions medium and 1 accession strong seed veining. The characteristic seed colour (primary and secondary) and coat pattern was divided into nine classes with additional subclasses. Among one-colour class the most accessions were categorized according to Phaselieu descriptor into WH 23, i.e., white long large seeds (6 accessions) and WH 13, i.e., white round large seeds (3 accessions). A total of 36 accessions were classified into bicolour class, among which 18 were Pinto type (BIP), 12 constant mottled type and
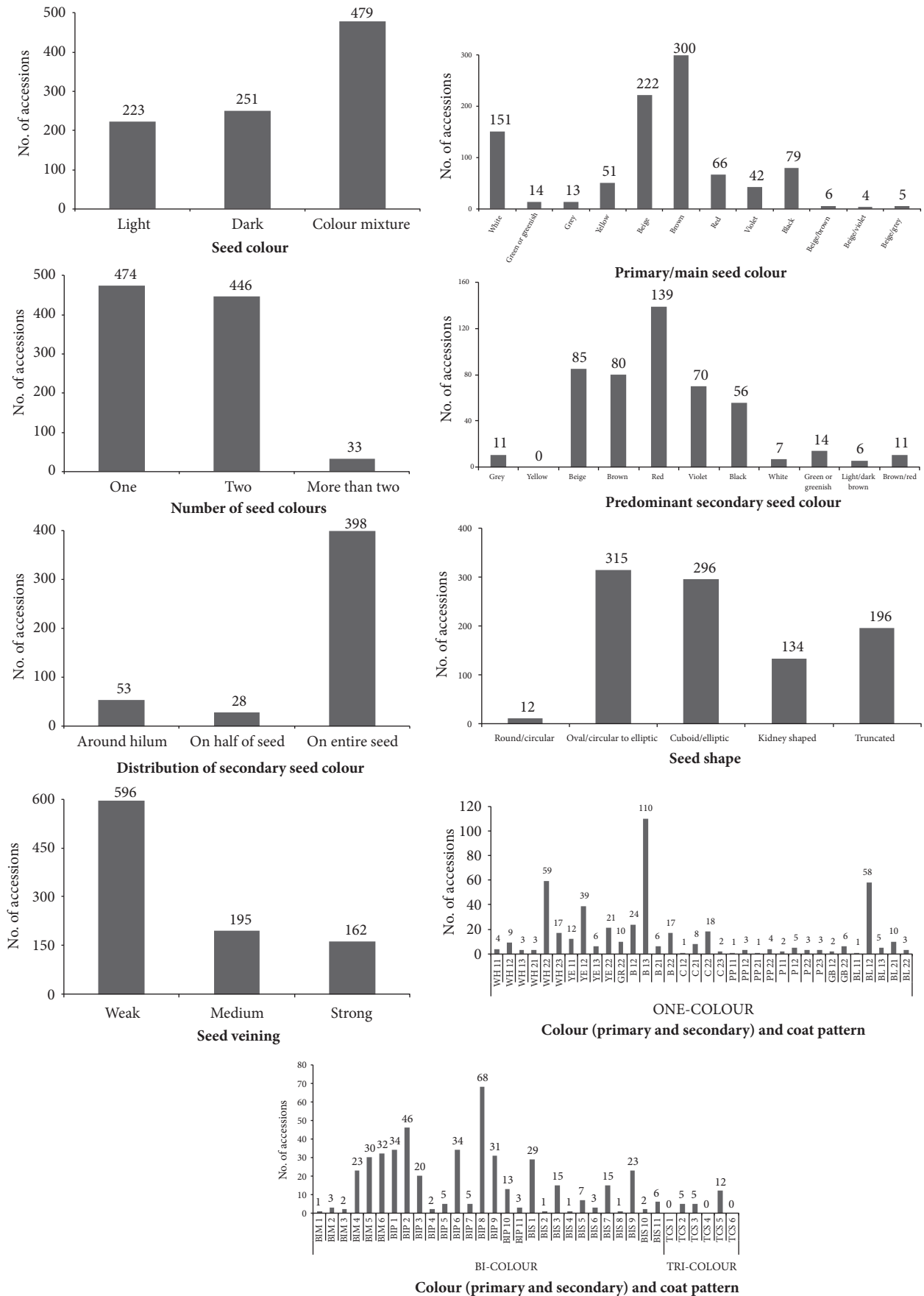
FIGURE 3: Frequency distribution of 953 common bean accessions (*Phaseolus vulgaris* L.) for qualitative seed characteristics.
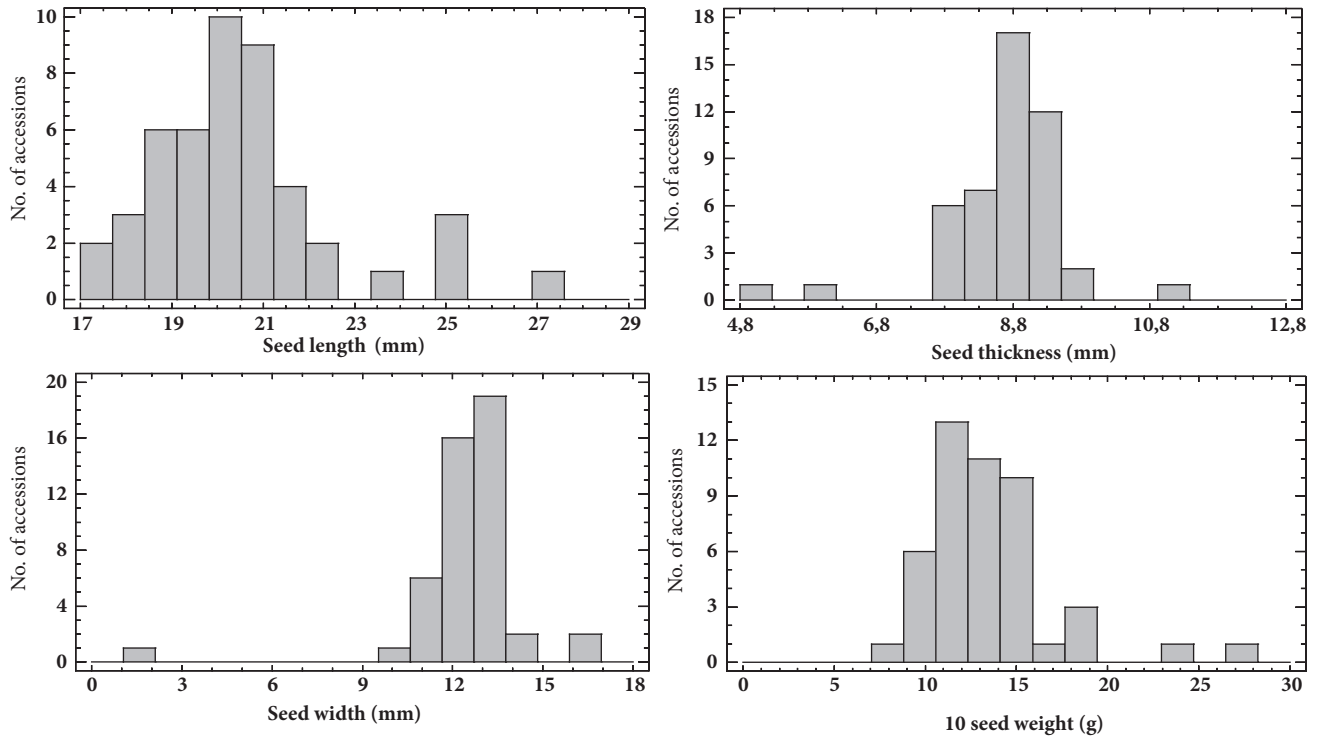
FIGURE 4: Frequency distribution of 47 runner bean accessions (*Phaseolus coccineus* L.) for quantitative seed characteristics.

FIGURE 5: Photo examples for runner bean accessions (*Phaseolus coccineus* L.) groups distribution according to seed length: (a) small, (b) medium, and (c) large; and 10 seed weight: (d) low-weight, (e) medium-weight, and (f) high-weight.

6 broad striped type. As described by Rodriguez et al. [15] within runner bean (*Phaseolus coccineus* L.) three botanical varieties exist, i.e., var. *albiflorus* with white seed colour, var. *bicolour* with beige seed and brown pattern colour, and var. *coccineus* with purple seed and black pattern colour.

3.3. Multivariate Analyses. The relationship between all fourteen seed characteristics, i.e., quantitative and qualitative,

was evaluated for common and runner bean germplasm with the Fisher's least significant difference (LSD) method. The test was performed in order to determine which seed characteristics are significantly different from others. The results showed that characteristic seed colour (primary and secondary) and coat pattern was in statistically significant correlation with all the seed characteristics, in both bean germplasm collections.
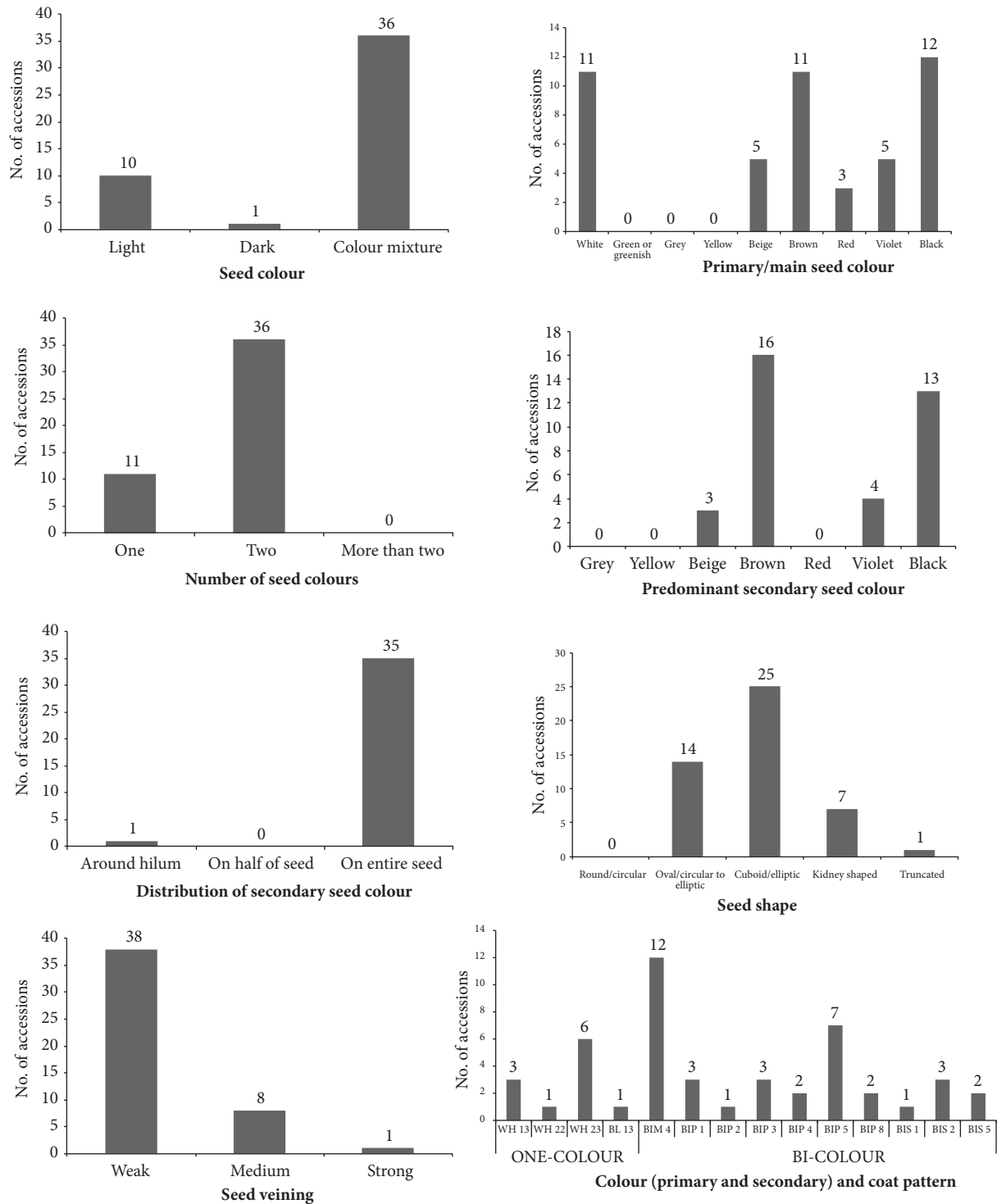
FIGURE 6: Frequency distribution of 47 runner bean accessions (*Phaseolus coccineus* L.) for qualitative seed characteristics.

The purpose of PCA analysis was to obtain a small number of linear combinations for fourteen variables which represent the majority of variability in the data within each bean germplasm collection. Based on an eigenvalue of greater than or equal to 1.0, 4 components were obtained for common bean germplasm (953 accessions), which together comprise 75.03% of the variability of the original data. PCA biplot in Figure 7 (top) defined Components 1 and 2, which together explained 52.09% of the total variance for fourteen variables in common bean germplasm. The Component 1 represented 32.90% and Component 2 19.19% of the total variance. The qualitative seed characteristics distribution of
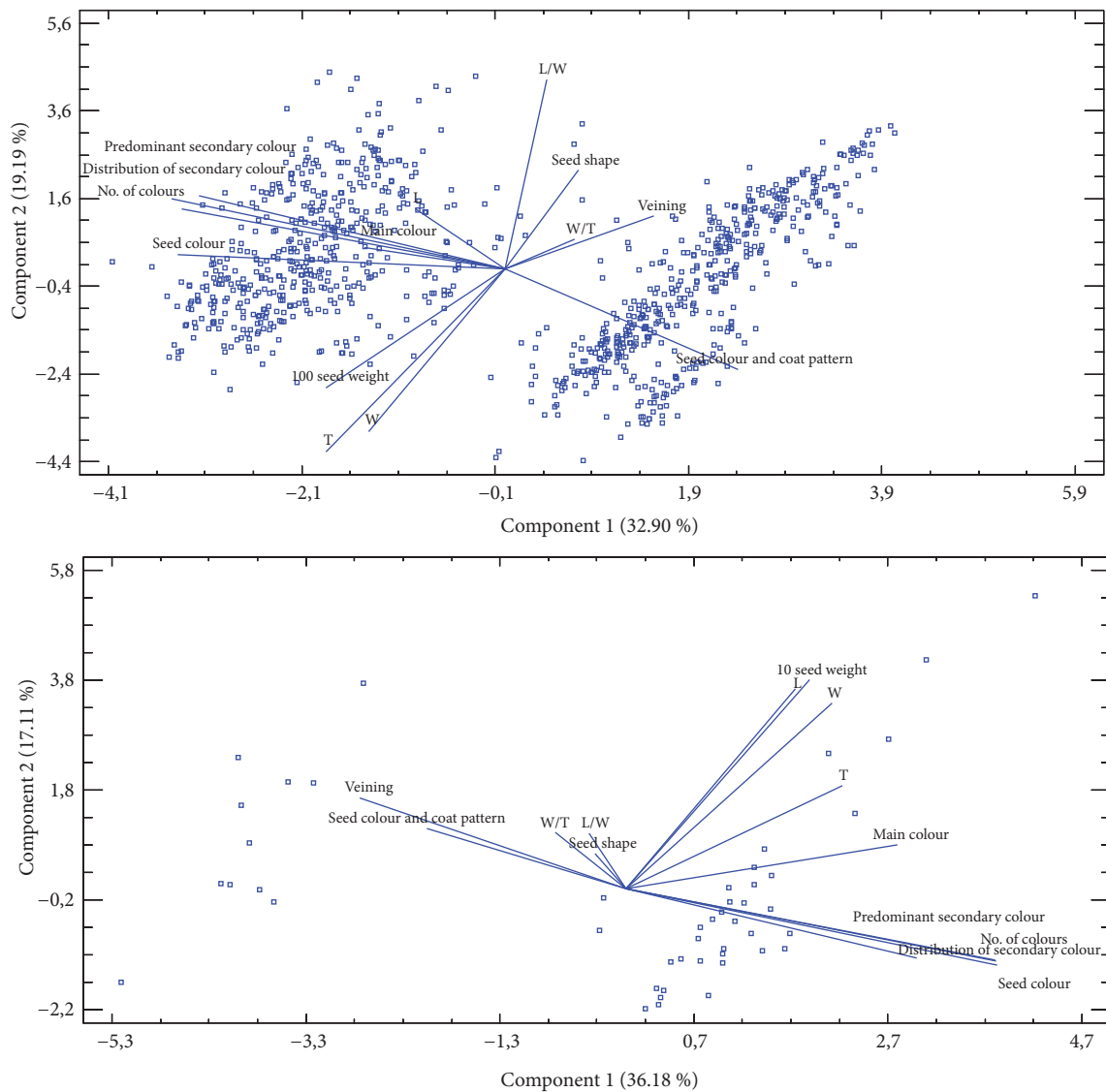
FIGURE 7: Biplot of total of fourteen seed characteristics for 953 accessions of common bean (top) and 47 accessions of runner bean germplasm (bottom). L, seed length; T, seed thickness; W, seed width; L/W, seed length/width ratio; W/T, seed width/thickness ratio.

secondary seed colour, seed colour, number of seed colours and the secondary seed colour were the major contributors to Component 1. The quantitative seed characteristics L/W, T, and W were the major contributors to Component 2. The accessions of common bean germplasm were well separated into 3 major groups (see Figure 7).

Based on an eigenvalue of greater than or equal to 1.0, 4 components were obtained for runner bean germplasm (47 accessions), which together comprise 80.16% of the variability of the original data. PCA biplot in Figure 7 (bottom) defined Components 1 and 2, which together represent 53.29% of the total variance for fourteen variables in runner bean germplasm. The Component 1 represented 36.18% and Component 2 17.11% of the total variance. The qualitative seed characteristics number of seed colours, distribution of secondary seed colour, seed colour, secondary seed colour

and seed veining were the major contributors to Component 1. The quantitative seed characteristics 10 seed weight, L, and T were the major contributors to Component 2.

Based on the cluster analyses performed on fourteen seed characteristics, all 953 common bean accessions were grouped into 3 clusters and each cluster was found to have varied number of accessions (see Figure 8 and Table 2) indicating the gene pool of origin (Mesoamerican gene pool origin, Andean gene pool origin and one subset group with putative hybrids between both gene pools). Dendrogram for accessions of common bean germplasm is presented in Figure 8 (top), where the combination of 14 individual variables established three main clusters of genotypes with standard accessions for the phaseolin type included. For this reason, it was not necessary to include pure American genotypes while we had internal standards for phaseolin type already within
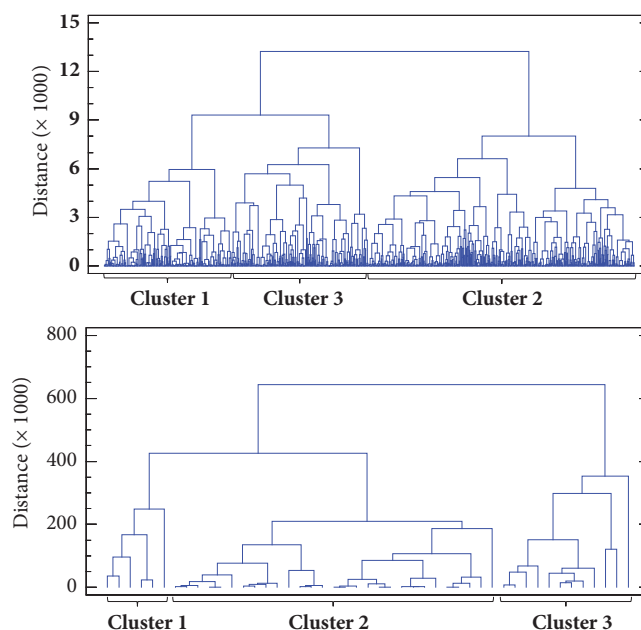
FIGURE 8: Dendrogram for 953 accessions of common bean (top) and 47 accessions of runner bean (bottom) germplasm according to fourteen seed characteristics.

TABLE 2: Means or descriptive characters of different variables in 3 clusters for 953 accessions of common bean (*Phaseolus vulgaris* L.).

Cluster variables	Cluster 1	Cluster 2	Cluster 3
Quantitative seed characteristic			
L (mm)	12.84	13.95	13.65
T (mm)	7.36	6.91	5.68
W (mm)	8.73	8.37	7.39
L/W	1.48	1.69	1.85
W/T	1.19	1.22	1.33
100 seed weight (g)	55.15	54.41	41.06
Qualitative seed characteristic			
Seed colour	Dark colour (83.70*)	Colour mixture (99.58*)	Light colour (75.10*)
Number of seed colours	One (100.00*)	Two (93.05*)	One (99.59*)
Primary/main seed colour	Brown (51.98*)	Brown (28.63*)	White (36.33*)
Predominant secondary seed colour	None (99.56*)	Red (29.05*)	None 99.59*)
Distribution of secondary seed colour	Without secondary colour (100.00*)	On entire seed (82.95*)	Without secondary colour (99.59*)
Seed veining	Weak (75.33*)	Weak (73.68*)	Strong (40.82*)
Seed shape	Oval/circular to elliptic (55.07*)	Oval/circular to elliptic (36.63*)	Cuboid/elliptic (39.59*)
Seed colour and coat pattern	One-colour: Brown (48.90*)	Bi-colour: Pinto T type (54.53*)	One-colour: White (35.92*)
Number of accessions	227	475	245

∗, % of main characteristic within each cluster; L, seed length; T, seed thickness; W, seed width; L/W, length/width ratio; W/T, width/thickness ratio.

the collection. The highest number of accessions were ranked in cluster 2 (475), followed by cluster 3 (245) and cluster 1 (227). The quantitative seed characteristics mean values of accessions grouped into each cluster showed that accessions in cluster 1 had the lowest L (12.84 mm), L/W (1.48) and W/T (1.19), and the highest T (7.36 mm), W (8.73 mm) and 100 seed weight (55.15 g). Furthermore, accessions in cluster 2 had the highest L (13.95 mm), while accessions in cluster 3 had

the lowest T (5.68 mm), W (7.39 mm) and 100 seed weight (41.06 g), and the highest L/W (1.85) and W/T (1.33). Among qualitative seed characteristics all accessions in cluster 1 had only one seed colour, which was for most accessions dark (83.70%), while most of accessions in cluster 2 had seeds with colour mixture (99.58%). The most of accessions in cluster 3 had one seed colour (99.59%), which was for most accessions (75.10%) light (Table 2). Comparing to the Indian common

TABLE 3: Means or descriptive characters of different variables in 3 clusters for 47 accessions of runner bean (*Phaseolus coccineus* L.).

Cluster variables	Cluster 1	Cluster 2	Cluster 3
Quantitative seed characteristic			
L (mm)	24.38	19.88	20.17
T (mm)	9.71	8.78	8.01
W (mm)	14.66	12.49	12.26
L/W	1.69	1.60	1.66
W/T	1.51	1.39	1.57
10 seed weight (g)	20.08	12.50	12.54
Qualitative seed characteristic			
Seed colour	Colour mixture (100.00*)	Colour mixture (100.00*)	Light colour (83.33*)
Number of seed colours	Two (100.00*)	Two (100.00*)	One (91.67*)
Primary/main seed colour	Beige (66.67*)	Black or brown (37.93*)	White (91.67*)
Predominant secondary seed colour	Brown (66.67*)	Black or brown (37.93*)	None (91.67*)
Distribution of secondary seed colour	On entire seed (100.00*)	On entire seed (96.55*)	None (91.67*)
Seed veining	Weak (83.33*)	Weak (100.00*)	Medium (58.33*)
Seed shape	Cuboid/elliptic (83.33*)	Cuboid/elliptic (62.07*)	Oval/circular to elliptic (41.67*)
Seed colour and coat pattern	Bi-colour: Pinto T type (50.00*)	Bi-colour: Pinto T type (48.28*)	One-colour: White (83.33*)
Number of accessions	6	29	12

∗, % of main characteristic within each cluster; L, seed length; T, seed thickness; W, seed width; L/W, length/width ratio; W/T, width/thickness ratio.

bean collection, 10 genetically diverse clusters were obtained regarding the phenological, morphological, and agricultural traits of common bean [21].

Similarly to the common bean, based on the cluster analyses performed on fourteen seed characteristics, all 47 runner bean accessions were grouped into 3 clusters with varied number of accessions (see Figure 8 and Table 3). Dendrogram for accessions of runner bean germplasm is presented in the Figure 8 (bottom). The highest number of accessions were ranked in cluster 2 (29), followed by cluster 3 (12) and cluster 1 (6). The quantitative seed characteristics mean values of accessions grouped into each cluster showed that accessions in cluster 1 had the highest L (24.38 mm), T (9.71 mm), W (14.66 mm), L/W (1.69) and 10 seed weight (20.08 g). Furthermore, accessions in cluster 2 had the lowest L (19.88 mm), L/W (1.60), W/T (1.39) and 10 seed weight (12.50 g), while accessions in cluster 3 had the lowest T (8.01 mm) and W (12.26 mm), and the highest W/T (1.57). Among qualitative seed characteristics all accessions in cluster 1 had seeds with colour mixture (100.00%), mainly with beige as primary/main seed colour and brown as secondary seed colour (66.67%). Accessions in cluster 2 had seeds with colour mixture (100.00%), mainly with black or brown as primary/main and secondary seed colour. The most of accessions in cluster 3 had one seed colour (91.67%) which was for most accessions (83.33%) light (Table 3). As discussed by Schwember et al. [6] extensive characterizations of runner bean germplasm are great challenges and opportunities in the future that would increase its cultivation on a broader scale worldwide.

4. Conclusion

This manuscript describes for the first time large-scale morphological seed characterization of the common and runner bean collection germplasm conserved in the Slovenian gene bank. The germplasm evaluated has a wide range of morphological variability based on fourteen seed characteristics. The study encompassed 953 accessions of common and 47 accessions of runner bean. Seeds of each accession were evaluated for quantitative characteristics: L (range 7.3–27.2 mm); T (range 4.2–11.0 mm); W (range 0.3–16.5 mm); L/W (range 0.4–2.6 mm); W/T (range 0.6–2.2 mm); and 100 for common (range 19.3–98.4 mm) or 10 for runner (range 7.6–26.7 mm) bean seed weight. Furthermore, seeds were evaluated using qualitative characteristics: seed colour (50.3% of common beans and 76.6% of runner beans were colour mixture), number of seed colours (49.7% of common beans had one colour and 76.6% of runner beans two colours per seeds), primary/main seed colour (31.5% of common beans were brown and 25.5% of runner beans black), predominant secondary seed colour (14.6% of common beans had red colour and 34.0% of runner beans brown colour), distribution of secondary seed colour (41.8% of common beans and 74.5% of runner beans had secondary colour distributed on the entire seed), seed veining (62.5% of common beans and 80.9% of runner beans had weak seed veining), seed shape (33.1% of common beans had oval/circular to elliptic and 53.2% of runner beans cuboid/elliptic seed shape), and seed colour and coat pattern (11.5% of common beans had large and one-colour brown seeds; 25.5% of runner beans had bicolour constant mottled seeds). The results of seed characterization indicate the origin (Andean, Mesoamerican, putative hybrids between gene pools) and domestication pathways of common and runner beans. To date, this is the first study reporting the morphological characteristics and comparisons of whole common and runner bean germplasm conserved in one of the Central European collections. The results obtained in this study are serving as the useful

information on genetic diversity of common and runner bean accessions at the Slovene gene bank, which could be used for development of new bean varieties for studied seed characteristics. Nondestructive screening test based on the seed characterization of large bean germplasm is shown to be an informative, noninvasive, and suitable tool for distinction of bean accessions according to the gene pool origin.

Abbreviations

AFLP: Amplified fragment length polymorphism
cpSSRs: Chloroplast microsatellites or simple sequence repeats
L: Seed length
T: Seed thickness
W: Seed width
L/W: Length/width ratio
W/T: Width/thickness ratio
PCA: Principal component analysis.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

Supplementary Materials

Table S1: List of standard common bean (*Phaseolus vulgaris* L.) accessions with known phaseolin type. *(Supplementary Materials)*

References

[1] C. Fowler, "Crop diversity: Neolithic foundations for agriculture's future adaptation to climate change," *AMBIO: A Journal of the Human Environment*, vol. 37, no. 14, pp. 498–501, 2008.

[2] E. N. Aquino-Bolaños, Y. D. García-Díaz, J. L. Chavez-Servia, J. C. Carrillo-Rodríguez, A. M. Vera-Guzmán, and E. Heredia-García, "Anthocyanin, polyphenol, and flavonoid contents and antioxidant activity in Mexican common bean (Phaseolus vulgaris L.) landraces," *Emirates Journal of Food and Agriculture*, vol. 28, no. 8, pp. 581–588, 2016.

[3] P. H. Graham and P. Ranalli, "Common bean (Phaseolus vulgaris L.)," *Field Crops Research*, vol. 53, no. 1-3, pp. 131–146, 1997.

[4] G. Logozzo, R. Donnoli, L. Macaluso, R. Papa, H. Knüpffer, and P. S. Zeuli, "Analysis of the contribution of Mesoamerican and Andean gene pools to European common bean (Phaseolus vulgaris L.) germplasm and strategies to establish a core collection," *Genetic Resources and Crop Evolution*, vol. 54, no. 8, pp. 1763–1779, 2007.

[5] M. Santalla, A. B. Monteagudo, A. M. González, and A. M. De Ron, "Agronomical and quality traits of runner bean germplasm and implications for breeding," *Euphytica*, vol. 135, no. 2, pp. 205–215, 2004.

[6] A. R. Schwember, B. Carrasco, and P. Gepts, "Unraveling agronomic and genetic aspects of runner bean (Phaseolus coccineus L.)," *Field Crops Research*, vol. 206, pp. 86–94, 2017.

[7] A. M. González, A. P. Rodiño, M. Santalla, and A. M. De Ron, "Genetics of intra-gene pool and inter-gene pool hybridization for seed traits in common bean (Phaseolus vulgaris L.) germplasm from Europe," *Field Crops Research*, vol. 112, no. 1, pp. 66–76, 2009.

[8] S. Peres, "Saving the gene pool for the future: Seed banks as archives," *Studies in History and Philosophy of Science Part C Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 55, pp. 96–104, 2016.

[9] F. R. Hay, "Seed Banks," in *Encyclopedia of Applied Plant Sciences*, B. Thomas, B. G. Murray, and D. J. Murphy, Eds., pp. 327–333, Academic Press, Oxford, UK, 2nd edition, 2017.

[10] J. Šuštar-Vozlič, M. Maras, B. Javornik, and V. Meglič, "Genetic diversity and origin of slovene common bean (Phaseolus vulgaris L.) germplasm as revealed by AFLP markers and phaseolin analysis," *Journal of the American Society for Horticultural Science*, vol. 131, no. 2, pp. 242–249, 2006.

[11] M. Maras, J. Šuštar-Vozlič, W. Kainz, and V. Meglič, "Genetic diversity and dissemination pathways of common bean in central Europe," *Journal of the American Society for Horticultural Science*, vol. 138, no. 4, pp. 297–305, 2013.

[12] M. Maras, B. Pipan, J. Šuštar-Vozlič et al., "Examination of genetic diversity of common bean from the western Balkans," *Journal of the American Society for Horticultural Science*, vol. 140, no. 4, pp. 308–316, 2015.

[13] S. A. Angioi, D. Rau, G. Attene et al., "Beans in Europe: Origin and structure of the European landraces of Phaseolus vulgaris L," *Theoretical and Applied Genetics*, vol. 121, no. 5, pp. 829–843, 2010.

[14] T. Gioia, G. Logozzo, G. Attene et al., "Evidence for introduction bottleneck and extensive inter-gene pool (Mesoamerica× Andes) hybridization in the European common bean (Phaseolus vulgaris L.) germplasm," *PLoS ONE*, vol. 8, no. 10, Article ID 0075974, 2013.

[15] M. Rodriguez, D. Rau, S. A. Angioi et al., "European Phaseolus coccineus L. landraces: population structure and adaptation, as revealed by cpSSRs and phenotypic analyses," *PLoS ONE*, vol. 8, no. 2, Article ID 0057337, 2013.

[16] E. Pérez-Vega, A. Pañeda, C. Rodríguez-Suárez, A. Campa, R. Giraldez, and J. J. Ferreira, "Mapping of QTLs for morpho-agronomic and seed quality traits in a RIL population of common bean (Phaseolus vulgaris L.)," *Theoretical and Applied Genetics*, vol. 120, no. 7, pp. 1367–1380, 2010.

[17] S. P. Singh, P. Gepts, and D. G. Debouck, "Races of common bean (Phaseolus vulgaris, Fabaceae)," *Economic Botany*, vol. 45, no. 3, pp. 379–396, 1991.

[18] M. W. Blair, L. F. González, P. M. Kimani, and L. Butare, "Genetic diversity, inter-gene pool introgression and nutritional quality of common beans (Phaseolus vulgaris L.) from Central Africa," *Theoretical and Applied Genetics*, vol. 121, no. 2, pp. 237–248, 2010.

[19] "Descriptor list for Phaseolus vulgaris," International Board for Plant Genetic Resources, 1982, http://www.bioversityinterna-tional.org/uploads/tx_news/Phaseolus_vulgaris_descriptors_160.pdf.

[20] UPOV, "International Union for the protection of new varieties of plants," 2005, http://www.upov.int/edocs/tgdocs/en/tg012.pdf.

[21] "CPVO Phaseolus vulgaris L," Protocol for Tests on Distinctness, Uniformity and Stability-French Bean, 2013, http://cpvo.europa.eu/sites/default/files/documents/phaseolus_vulgaris_4.pdf.

[22] J. C. Rana, T. R. Sharma, R. K. Tyagi et al., "Characterisation of 4274 accessions of common bean (Phaseolus vulgaris L.) germplasm conserved in the Indian gene bank for phenological, morphological and agricultural traits," *Euphytica*, vol. 205, no. 2, pp. 441–457, 2015.

[23] M. Kara, B. Sayinci, E. Elkoca, I. Öztürk, and T. B. Özmen, "Seed size and shape analysis of registered common bean (*Phaseolus vulgaris* L.) cultivars in Turkey using digital photography," *Tarim Bilimleri Dergisi*, vol. 19, no. 3, pp. 219–234, 2013.

[24] D. M. Giurca, "Morphological and phenological differences between the two species of the Phaseolus genus (Phaseolus vulgaris and Phaseolus coccineus)," *Cercetari Agronomice in Moldova*, vol. 42, no. 2, pp. 39–45, 2009.

[25] A. R. Piergiovanni and L. Lioi, "Italian common bean landraces: History, genetic diversity and seed quality," *Diversity*, vol. 2, no. 6, pp. 837–862, 2010.

Research Article

In Silico Genome-Wide Analysis of the ATP-Binding Cassette Transporter Gene Family in Soybean (*Glycine max* L.) and Their Expression Profiling

Awdhesh Kumar Mishra ⓘ**, Jinhee Choi,**
Muhammad Fazle Rabbee, and Kwang-Hyun Baek ⓘ

Department of Biotechnology, Yeungnam University, Gyeongsan, Gyeongbuk 38541, Republic of Korea

Correspondence should be addressed to Kwang-Hyun Baek; khbaek@ynu.ac.kr

ATP-binding cassette (ABC) transporters constitute one of the largest gene families in all living organisms, most of which mediate transport across biological membranes by hydrolyzing ATP. However, detailed studies of ABC transporter genes in the important oil crop, soybean, are still lacking. In the present study, we carried out genome-wide identification and phylogenetic and transcriptional analyses of the ABC gene family in *G. max*. A total of 261 *G. max* ABC (GmABCs) genes were identified and unevenly localized onto 20 chromosomes. Referring to protein-domain orientation and phylogeny, the GmABC family could be classified into eight (ABCA-ABCG and ABCI) subfamilies and ABCG were the most abundantly present. Further, investigation of whole genome duplication (WGD) signifies the role of segmental duplication in the expansion of the ABC transporter gene family in soybean. The Ka/Ks ratio indicates that several duplicated genes are governed by intense purifying selection during evolution. In addition, *in silico* expression analysis based on RNA-sequence using publicly available database revealed that ABC transporters are differentially expressed in tissues and developmental stages and in dehydration. Overall, we provide an extensive overview of the GmABC transporter gene family and it promises the primary basis for the study in development and response to dehydration tolerance.

1. Introduction

Active transport mechanisms require the intake of energy and can be operated by either of two mechanisms: (1) free energy change associated with ATP hydrolysis (primary transport), or (2) assisted by the potential energy of the chemical gradient of another molecule (secondary transport). Depending on structural homology and transport mechanism, primary transporters are subdivided into three classes — rotary motor ATPases (F-, V-, and A-ATPases), P-type ATPases, and a large family of integral membrane proteins named ATP-binding cassette (ABC) transporters [1]. The ABC transporters represent one of the largest families of membrane proteins, ubiquitously found within the kingdoms of eukarya, eubacteria, and archaea [1–3]. These proteins employ ATP hydrolysis to transport various substrates (e.g., heavy metals, endogenous metabolites, inorganic anions, drugs, lipids, sterols, hormones, amino acids, peptides, vitamins, and sugars) in and out of cells across biological

membranes [2, 4]. Primarily, plant ABCs transporters are recognized only in detoxification processes [5]. Afterwards, an abundant number of plant ABCs have been characterized in physiological and developmental processes. So far, more than 100 ABC transporters encoding genes have been characterized across plant genome. The numbers of plant ABC transporters are more than those of most other organisms, suggestive of their involvement in a broad range of biological functions [6, 7]. The Arabidopsis genome contains 131 ABC genes, whereas 121 ABC genes were reported in the rice genome [8, 9]. For instance, white brown complex (WBC) homolog, pleiotropic drug resistance (PDR), multidrug resistance (MDR), and MDR-associated protein (MRP) are best-exemplified subfamilies among plant ABC transporters. However, genome-wide surveys and expression analysis of this gene family have not been performed in *G. max*.

ABC transporters are a family of membrane-bound proteins that mediate transport across biological membranes

by hydrolyzing ATP. They contain conserved cytoplasmic domains termed as nucleotide-binding domains (NBD) (also referred to as ATP-binding cassettes). These NBD are ~200 amino acids coupled with ATPase and often serve as energy suppliers for substrate translocation as well as nontransport processes [10]. Apart from this domain, ABC proteins also consist of one or two hydrophobic transmembrane domains (TMDs). Based on the transporter class, each TMD comprises 6-10 transmembrane α-helices (frequently 6 helices present among most of the exporters). TMDs determine substrate specificity and are responsible for the translocation of a substrate across the lipid bilayer [8, 11]. The NBDs hydrolyze ATP and drive conformational changes in the attached TMDs, thus allowing substrates to cross the lipid bilayer of the membrane and be either imported into or exported out of the cytoplasm. Further, members of the ABC superfamily are classified into exporters, importers, and nontransporting proteins. Exporter and importer types of ABC proteins comprise at least two TMDs and two NBDs. Exporters are present in all kingdoms and importers solely in bacteria and plants [12]. Hence, coming to sequence comparison, there are 3 classes of ABC systems present among the last common ancestors (archaea, bacteria, and eukarya). Class 1 comprises ABC proteins in which TMDs and NBDs are fused to form a single polypeptide (exporters) (found in eukaryotes), class 2 includes nontransport ABCs lacking TMDs (nontransporters; found in both eukaryotes and prokaryotes), and class 3 includes mainly transporters with NBDs and TMDs formed by separate polypeptide chains (canonical importers; exclusively found in prokaryotes) and some bacterial exporters [13, 14].

According to sequence similarity, ABC transporter proteins are classified according to structure (full, half, or quarter molecule) and orientation (forward or reverse). Three structural types are defined as follows. A full transporter comprises two TMDs along with two NBDs and can be either in forward (TMD1-NBD1-TMD2-NBD2) or in reverse (NBD1-TMD1-NBD2-TMD2) orientation. Half transporters are composed of one TMD and one NBD and can be in forward (TMD1-NBD1) or reverse orientation (NBD1-TMD1). They usually constitute a feasible unit by forming a virtual full transporter as homo- or heterodimers [15, 16]. The third type of transporter has no TMDs, but two NBDs [17]. Owing to the lack of TMD, they are not usually involved in the transmembrane transport mechanism. Few transporters consist of single NBDs and are similar to prokaryotic ABC proteins, referred to as quarter transporters. The protein is either a half-size or full-size transporter, but the majority of subfamilies comprise both types of transporters [17, 18].

NBD is present among all three structural types and includes several highly conserved motifs within the domain: Walker A (also called a phosphate-binding loop), Walker B, Q-loop, D-loop, switch H-loop, and a signature motif or C motif (LSGGQ) [19]. The major function of D-loop is holding dimers together such as the switch H-loop interacting with the transmembrane domain. Both Walker motifs A and B form the P-loop, which binds to ATP, and, at last, the Q-loop and H-loop include special residues crucial for interacting with the γ-phosphate of the ATP [13]. Furthermore, signature

motifs exclusively found in ABC proteins make them discernible from another type of ATPase [20, 21].

Human Genome Organization (HUGO) scheme has been presently accepted and well documented for categorization of human as well as plant ABC proteins [22]. In accordance with eukaryotes, eight basic subfamilies (A–H) identified based on the domain organization and homologous relationship. However, ABCH subfamily has been reported in arthropod genomes and absent in fungi, mammals, and plants [23, 24], apart from another prokaryotic-like ABC protein with a single NBD domain called subfamily I. They are highly soluble and found only in plants but are absent in most animal genomes [17]. Hence, in total, nine subfamilies (ABCA-ABCI) have been identified, and eight of them (ABCA-ABCG and ABCI) are present in plant genomes [6]. Furthermore, proteins which belong to ABCA-ABCD subfamilies have a forward direction for domain organization (TMD-NBD) whereas ABCG and ABCH subfamilies have reverse domain organization (NBD-TMD). Proteins of ABCE and ABCF include only two NBDs and are designated as soluble proteins. They do not participate in any transport-related processes (but their NBDs can cluster with those of other ABC proteins). ABCI proteins possess only one single domain, predominantly NBD, rendering them difficult to identify and categorize. This genetic fact points out that some of ABCI proteins can bind together into systemic multisubunit ABC transporters and this subfamily is also designated as soluble proteins. Considering similar domain organization, ABCI subfamily transporters are evolved due to the movement of genes from mitochondria and plastids to the nucleus [7, 17, 18].

Proteins of the ABCA subfamily possess several half-size (ABC2 homolog, named ATH) variants and, so far, one full size (ABC1 homolog, named AOH) has been reported. Full-size variants are likely associated with the transportation of sterols, whereas the function of half-size ABC2 homologs remains unclear. ABCB subfamily possesses both full size (MDR; also known as p-glycoprotein) and half size such as transporter associated with antigen processing (TAP) and ABC transporter of the mitochondria (ATM). Full-size variants of ABCB proteins are known for detoxification and transport of auxins. Half-size ABCB is involved in heavy metal tolerance. Typically ABCC subfamily proteins are full-size ABC transporters and called MDR-associated protein. They are thought to be associated with the transportation of organic anions and xenobiotic anions into the vacuole. Few half-size ABCC are also reported in Arabidopsis and rice. ABCD subfamily proteins are generally half-size (peroxisomal membrane protein) transporters. Some exceptionally full-size ABCD variants are found in Arabidopsis and *Vitis vinifera*. They are implicated in the transport of fatty acids into the peroxisome. ABCG subfamily has both full- and half-size variants. Full-length members, also referred to as PDR transporters, have been implicated in heavy metal resistance and auxin transport [8, 25, 26]. WBC transporters are half-length members of the ABCG subfamily and requirements for the export of cuticular lipids and alkanes [27, 28]. Both ABCE and ABCF proteins are extremely conserved among all living organisms during evolution. ABCE are also referred

to as RNase L inhibitors (RLI), associated with ribosome biogenesis. In most eukaryotes including human, single ABCE have been identified [29]. ABCF subfamily is conventionally known as the general control nonrepressible (GCN) subfamily and is found to be involved in stress tolerance in plants [18]. ABCI subfamily transporters have single NBD domain. A few well-studied examples of ABCI subfamily are closely associated with iron-sulfur center biogenesis complex and cytochrome C maturation complex [7, 17, 18]. Moreover, previous studies revealed that the ABCB, ABCC, and ABCG subfamilies possess a relatively higher number of genes in the plant, in comparison to other eukaryotes such as human and yeast [17].

Soybean [*Glycine max* (L.) Merr.] is one of the most frequently cultivated crops in the world and belongs to the Leguminosae subfamily. Seeds are an important source of human food, cooking oil, and animal feed, because of their abundant protein and oil content. Soybean plants also play a significant role in soil fertility, as they fix atmospheric nitrogen during symbiosis. In addition, they are the predominant source of isoflavonoids and play a significant role in human diet and health [30]. The first draft genome of soybean (*Glycine. max* var. Williams 82) was reported in 2010 [31], and the latest version of the genome assembly, version 2.0 (*G. max* Wm82.a2.v1), contains 56,044 protein-coding genes; yet their functional contribution to crop traits remains mostly unidentified. A large number of soybean coding genes are presently annotated for Gene Ontology biological process (GOBP) terms with experimental evidence. In this study, we screened the new soybean genome database (*G. Max* Wm82.a2.v1; https://phytozome.jgi.doe.gov/pz/portal.html) for ABC transporter genes. We performed a detailed analysis of their classification, chromosome distribution, physicochemical properties, phylogenetic analysis, and duplication. Finally, we verified the differential expression profiles of all GmABC genes in nine different *G. max* tissues, seven developmental stages, and various stress conditions. This study provides important information about the origin and evolution of the GmABC family in *G. max* and lays the foundation for further studies of the functions of this gene family.

2. Materials and Methods

2.1. Identification of ABC Transporter Genes in Soybean. There are different types of putative ABC proteins such as ABC transporter domain (PF00005), the ABC-2 transporter domain (PF01061), the ABC transporter transmembrane region domain (PF00664), the CYT domain (PF01458), or the mce-related protein domain (PF02470). The Hidden Markov Model (HMM) profiles of all the abovementioned proteins were downloaded from the Pfam 27.0 database (http://pfam.xfam.org/; [32]). Proteomes of soybean were searched against all the abovementioned HMM profiles to detect all ABC transporter domains with an E-value cut-off $< 1E^{-4}$ in the new soybean genome database (*Glycine max* Wm82.a2.v1) derived from phytozomev12 (https://phytozome.jgi.doe.gov/pz/portal.html; [33]), with default settings. In addition, corresponding genomic and coding sequences along with their chromosomal positions were

also retrieved from Phytozome. Based on Verrier et al. [17], identification and classification of these ABC proteins were performed.

The protein sequences retrieved from the above approaches were collected together. Further redundancy of sequences was eliminated in order to generate a unique set of putative ABC transporter proteins. Additionally, the presence of ABC transporter domains in all these proteins was confirmed by domain search via SMART (http://smart.embl-heidelberg.de/; [34]) and PROSITE (https://prosite.expasy.org/scanprosite/; [35]). Physicochemical properties such as molecular weight (Mw) and isoelectric point (pI) for each GmABC protein were calculated using the ExPASy proteomics server (https://web.expasy.org/compute_pi/; [36].

2.2. Chromosomal Distribution, Gene Structure Prediction, and Gene Duplications in GmABCs. The chromosomal position of all GmABC genes was retrieved from *G. Max* Wm82.a2.v1 genome database in Phytozome using BLASTN searches. MapChart 2.3 (Wageningen UR, Wageningen, the Netherlands; [37]) was used to map the ABC transporter genes onto chromosomes from short-arm to long-arm telomeres, in ascending order of physical position (bps). Tandem duplications of ABC transporter genes in *G. max* were identified in Plant Tandem Duplicated Genes Database (PTGBase) (http://ocri-genomics.org/PTGBase/; [38]). Further, GmABCs in duplicated genomic regions were obtained for syntenic mapping from the batch download option of Plant Genome Duplication Database (PGDD; http://chibba.agtec.uga.edu/duplication/; [39]). Synonymous substitution (Ks) and nonsynonymous substitution (Ka) rates for each segmentally duplicated GmABCs were retrieved. Subsequently, the period of duplication incidents in soybean was predicted according to the equation $T = Ks/2\lambda$; here the mean synonymous substitution rate (λ) was taken as $6.1 * 10^{-9}$ [40, 41]. The numbers and positions of exons and introns within GmABC gene were detected by comparing the full-length cDNA sequences with the corresponding genomic DNA sequences using an online tool: gene structure display server (http://gsds.cbi.pku.edu.cn/index.php; [42]).

2.3. Phylogenetic Tree Analysis and Functional Classification of ABC Proteins. Multiple sequence alignments were executed with the deduced amino acid sequences of GmABCs by using ClustalW [43]. Afterwards, a phylogenetic tree was constructed using the Maximum-Likelihood (ML) [44] method provided in the Molecular Evolutionary Genetics Analysis (MEGA) 6.0 software tools [45] based on the Jones, Taylor, and Thornton (JTT) matrix-based model. The accuracy of an inferred tree was checked with bootstrap analysis (with 1,000 replications).

2.4. Gene Ontology Annotation and In Silico Expression Analysis through RNA-Sequencing Data. The Gene Ontology (GO) analysis of soybean ABC transporters was performed using the Blast2GO program (https://www.blast2go.com/; [46]. The full-length amino acid sequences of GmABC proteins were uploaded to the program for the blast, followed by mapping and annotation. The analysis was carried out in

three categories: biological process, molecular function, and cellular component.

We downloaded the RNA-seq data of nine different soybean tissues and analyzed the expression levels of all 261 GmABC genes based on their fragments per kilobase of transcript per million mapped reads (FPKM) values. FPKM values for each GmABC were computed to assess the transcript level of GmABCs across nine tissue genes and the data were normalized across tissues. The heatmap for GmABC genes was generated using TIGR Multi-Experiment Viewer (MeV4) software package [47]. To obtain expression patterns of GmABCs in different development stages and perturbation, we have used the publicly available mRNA sequence data from Genevestigator V3 database (https://genevestigator.com/gv/doc/intro_plant.jsp; [48]. We downloaded the program from the Genevestigator homepage and used academic license for analysis in the local computer. Using plant biology tool, we first identified 238 probes for development and perturbation conditional search for all 261 GmABC genes. Seven different developmental stages of soybean are germination, main shoot growth, inflorescence formation, flowering, fruiting, bean development, and final ripening. In the case of perturbation, light intensity, germination, dehydration, and ozone irradiation are included for transcript data analysis. All data represent the log2 ratio of relative expression pattern.

2.5. Homology Modeling for Three-Dimensional Structure. All the GmABC proteins were investigated against the Protein Data Bank (PDB) [49] by BLASTP (with the default parameters) to recognize the best template having a similar sequence and known three-dimensional structure. Further, three-dimensional (3D) structures of GmABCs were generated by intensive protein modeling using Phyre2 server (Protein Homology/AnalogY Recognition Engine; http://www.sbg.bio.ic.ac.uk/phyre2; [50]). The intensive mode uses the multitemplate alignment of experimentally solved protein structures. In addition, it incorporates ab initio folding simulation termed as Poing [51] to model regions of proteins with no apparent homology to known structures.

2.6. Cis-Elements Analysis. The upstream sequences of the GmABCs, which were 2 kb upstream from the translation start site, were retrieved from Phytozome. These sequences were analyzed for the identification of regulatory *cis*-elements important for gene expression under abiotic stress, development, and hormone signaling using PlantPAN 2.0 (http://plantpan2.itps.ncku.edu.tw/; [52].

3. Results

3.1. Identification of GmABC Transporter Genes and Their Physicochemical Properties. We identified 261 proteins with ABC transporter domains in the *G. max* genome. Different transcripts of the same gene were not considered. Among 261 GmABCs, there were 114 full transporters, 116 half transporters, and 31 soluble putative ABC proteins. ABCA, ABCB, ABCC, ABCD, and ABCG subfamilies contain full

transporters, as well as half transporters. As in most eukaryotes, soybean has only one ABCE gene [53]. The details of all 261 soybean ABC transporter proteins, including gene identifier, genomic location, domain organization, description, protein length, molecular weight, isoelectric point (pI), and their subfamily, are listed in Table S1. The lengths of ABC transporter proteins range from 67 (Glyma.16G125700) to 1826 (Glyma.03G136000) amino acids, specifying that there are huge variations within the soybean ABC transporter gene family. The predicted molecular weights of GmABC proteins ranged from 7.72 to 203.97 kDa and the pI values were between 4.41 and 10.5.

3.2. Chromosomal Distribution, Gene Structure, and Gene Duplication of GmABCs. As was previously mentioned, there are more than double ABC transporters present in soybean as compared to Arabidopsis or rice and that was likely due to recent WGD in soybean. Concerning gene duplication, it includes segmental/WGD and tandem duplication (TD) events [54]. To evaluate the expansion of the GmABC transporter gene family, a chromosome map was constructed based on the locations provided by Phytozome 12 (*G. max* Wm82.a2.v1). The 261 GmABC transporter genes are unevenly distributed throughout all 20 soybean chromosomes (2n=40, Figure 1). The largest number of GmABCs was found on chromosome 13 (34 genes), followed by chromosome 08 (26 genes) and chromosome 03 (19 genes). Both chromosomes 10 and 19 comprise 16 GmABCs each, whereas one gene could not be anchored onto any specific chromosome. It signifies that each subfamily of GmABCs was also unevenly distributed. Gene structures of the 261 soybean ABC transporters reveal high complexity with intron number ranging from 1 to 38 (Figure S1). The detailed chromosomal locations of GmABC genes and the number of introns in the gene are shown in Table S1.

Generally, tandem duplication is defined as two paralogs separated by less than five genes in the same chromosome. For detection of tandem duplication, we investigated tandem arrays of 261 GmABCs using PTGBase. Tandem arrays are described as multiple members of genes appearing inside the same or neighboring intergenic regions. We detected 28 GmABC clusters containing 69 tandemly duplicated genes (genes are depicted in a rectangle box in Figure 1 and arrays are shown in Table S2), which were identified on 15 chromosomes. Gene duplications are considered a prime factor in the evolution of genomes and expansion of gene families [55, 56]. Recent studies indicated that soybean has undergone two WGD events (approximately 59 and 13 million years ago), which resulted in ~75 % paralogous gene pairs [31, 57, 58]. Polyploidy introduces immense duplications, ultimately leading to the vital source of genetic innovation [59, 60]. Here, we identified a total of 190 duplicated gene pairs within 261 GmABC genes by using the Plant Genome Duplication Database (Table S3). The nonsynonymous/synonymous substitution ratio (Ka/Ks) explains the selective evolutionary pressure acting on a gene. The Ka/Ks ratio for majority (94.74%) of the gene pairs was found < 1, recommending that most of GmABCs are evolved under purifying selection with limited functional divergence after duplication (as shown in
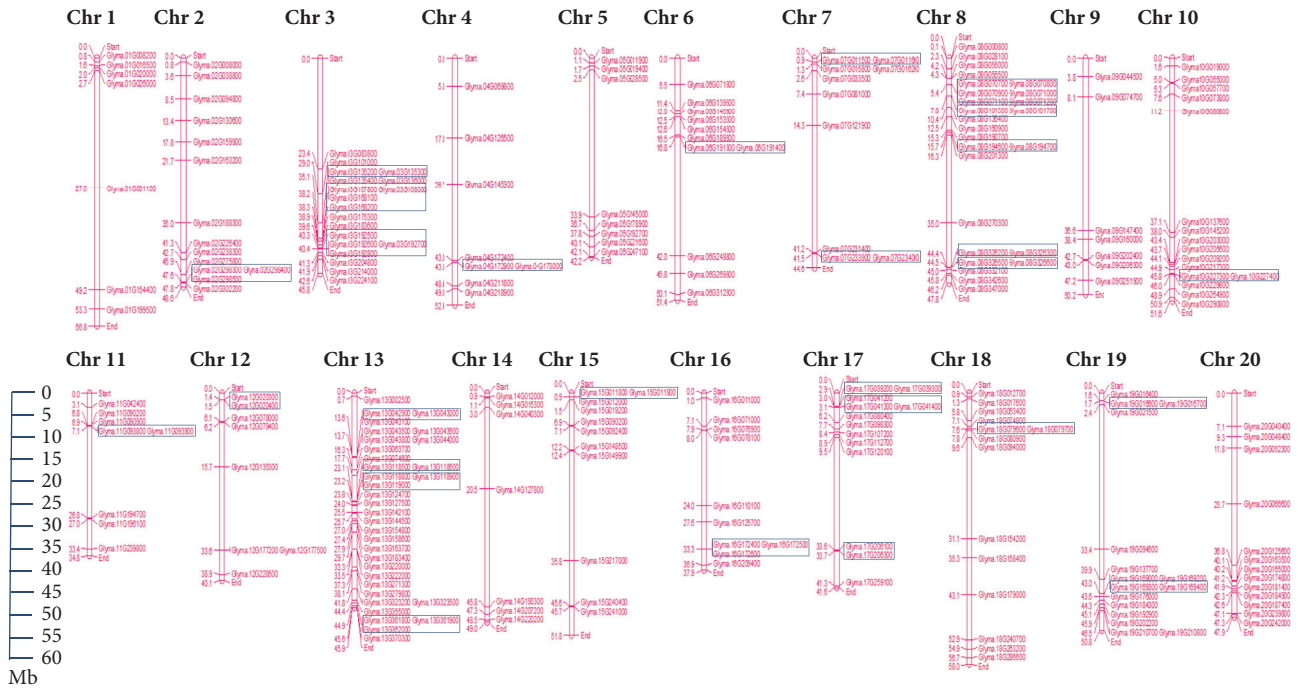
FIGURE 1: **Graphical representations of locations for putative ABC transporter genes on each soybean chromosome.** Genes within the rectangle box are putatively tandem duplicated. The chromosome size is indicated by its relative length using the information from Phytozome. The scale (in megabase) on the left depicts the relative lengths of the different chromosome in *G. max*. The gene names are depicted on the right side of each chromosome corresponding to the position of each gene. The figure was produced using the MapChart program. One GmABC could not be anchored onto any specific chromosome.

Table S3, the Ka/Ks value of most of the duplicated gene pairs ranged from 0.021 to 0.682). The Ka/Ks value of ten pairs of duplicated genes was more than 1, which signifies positive selection pressure. Thus, it concluded that segmental along with tandem duplication was involved in the expansion of the ABC transporter gene family in the soybean genome [61–63].

3.3. Phylogenetic Analysis of Soybean ABC Transporter Genes.
To elucidate the evolutionary relationships of soybean ABC transporter genes, an unrooted phylogenetic analysis was constructed from all 261 GmABC proteins using ML methods with 1,000 bootstraps. Genes from each subfamily (ABCA-ABCG and ABCI) are depicted with different bullet point colors (Figure 2). Referring to the ML phylogenetic tree analysis, the ABC family is grouped into eight subfamilies. Among ABCG, ABCB, and ABCC were the largest subfamilies containing 117, 52, and 48 members, respectively, accounting for a total of 83% of all GmABC members. ABCI, ABCF, ABCA, and ABCD included 20, 10, 8, and 5 members, respectively. There is only one member in ABCE subfamily identified (Table S1). All subfamilies were clustered together in the phylogenetic tree, except ABCI. In fact, ABCI genes were scattered throughout the tree. Our analysis corroborates previous studies, which indicated that the ABCG subfamily has the highest number of members and is accountable for the expansion of this gene family in comparison to other eukaryotes [7, 64].

3.4. In Silico Expression Analysis of GmABCs

3.4.1. Expression Analysis of GmABC Genes Exhibits a Differential Pattern in Tissues. RNA-sequencing is a robust tool for exploring the transcription patterns of certain genes using high-throughput sequencing approach. Expression levels of all 261 GmABC genes based on their FPKM values analyzed in nine different tissues and heatmap of all genes displayed differential tissue-specific expression patterns (Figure 3). Noticeably, thirty-one genes showed higher-level of transcript accumulation in all the tissues under this study suggesting that these have a unique role. They belong to all class of ABC subfamily, but, among these genes, most belong to soluble type ABC transporters (7 ABCF, 7 ABCI, and 1 ABCE) followed by 4 ABCD. Above all, 46 GmABC genes in pods, 40 genes in flower, 37 genes in leaf, 23 genes in the stem, 27 genes in shoot apical meristem, 41 genes in the root, and 17 in nodule accumulated the highest level of transcripts. Expressions of 20 genes were found to be lowest in all tissues.

3.4.2. GmABC Genes Showed Diverse Levels of Expression at Various Developmental Stages. Plant development is a very complex phenomenon that altered the expression of a number of genes to meet physiological and metabolic demand. To investigate the developmental regulation of GmABC transcripts, expression of genes was checked at seven different developmental stages, namely, germination, main shoot
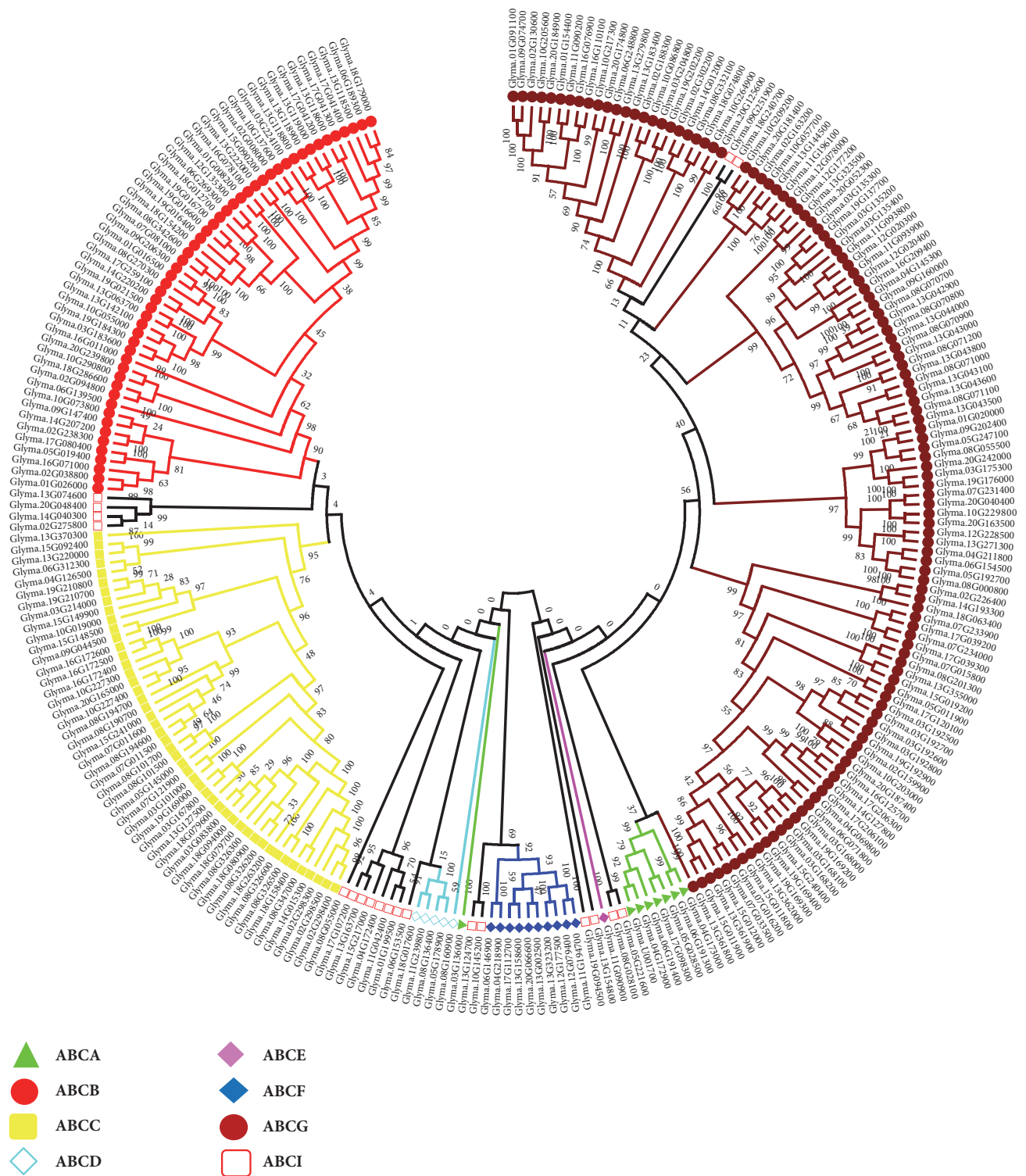
FIGURE 2: **Phylogenetic relationship of ABC transporter gene family proteins in soybean.** The phylogenetic tree was built using the ML method implemented in MEGA6.0 based on full-length amino acid sequences. The evolutionary history was inferred based on the JTT matrix-based model. The numbers at the nodes represent bootstrap percentage values based on 1000 replications. Genes from each subfamily are indicated with different bullet point colors.
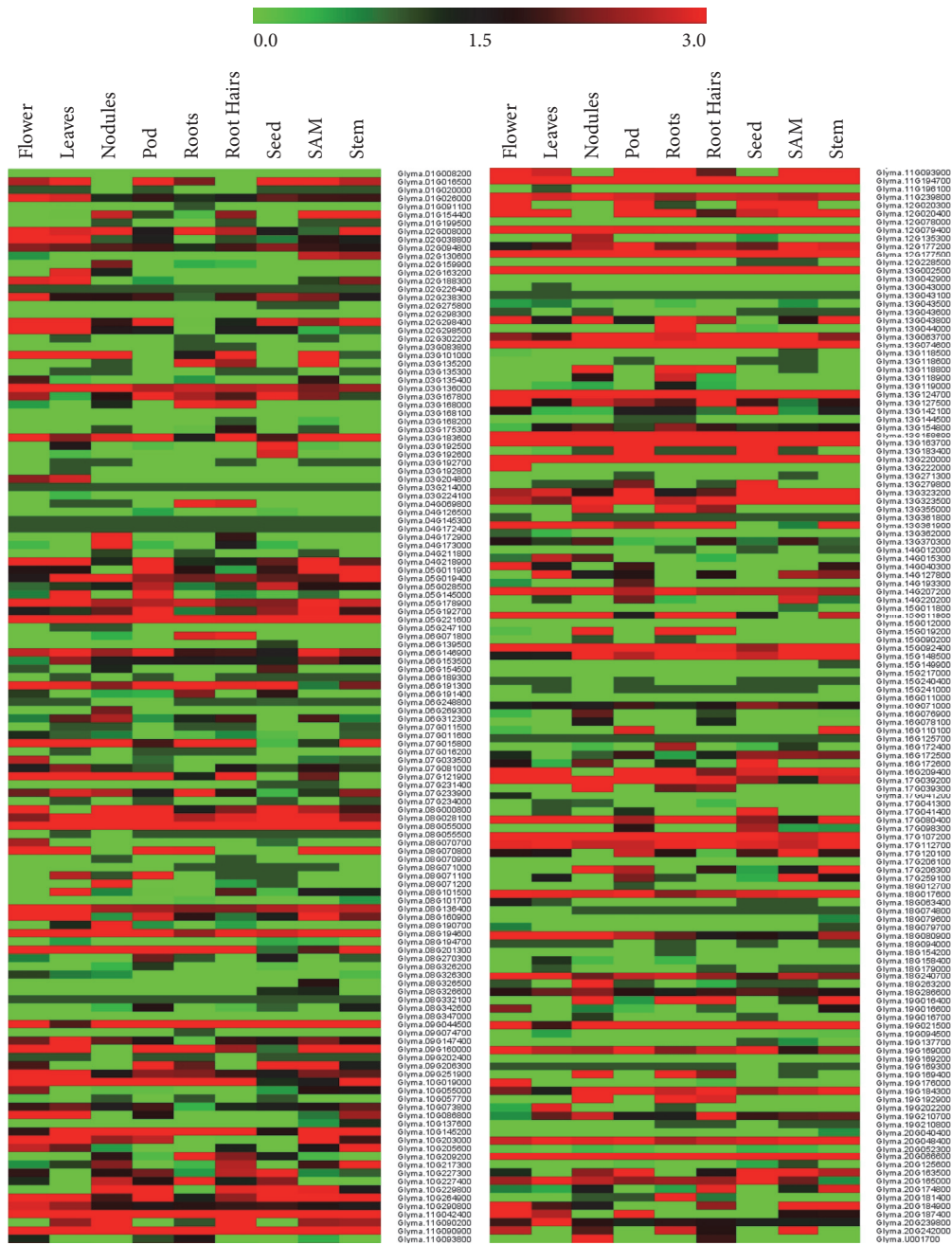
FIGURE 3: **Expression profiles of 261 soybean ABC transporter genes in nine different tissues derived from RNA-seq data of Phytozome database.** FPKM values of GmABC transporters genes were transformed by log2, and the heatmap was constructed with MeV 4. Expression levels are illustrated by graded color scale, representing relative expression pattern ranging from 0 (downregulated) to 3 (upregulated). Red, green, and black represent positive, negative, and zero, respectively. Gene IDs are given on the top. Genes with similar profiles across arrays are grouped on top by hierarchical clustering method.

growth, inflorescence formation, flowering, fruit formation, bean development, and final ripening. Expression of all these GmABC genes was investigated using Genevestigator based mRNA-seq data. Among 261 GmABC genes, 43 GmABC showed the highest level of expression at all the developmental stages of soybean (Figure S2A). About 37 genes were not detected in any of the seven development stages.

3.4.3. Expression Patterns of the GmABC Genes in Response to Abiotic Stress and Light and during Germination. The Genevestigator analysis showed expression of all 261 GmABC genes based on mRNA-sequence data. Based on the analysis, it was observed that twenty genes were highly expressed in all three dehydration study conditions (6 hr., 12 hr., and 24 hr.) in two soybean cultivar Benning and PI-416937. Thus,

they might have a role in dehydration tolerance in plants. More than 40 genes were dominantly expressed during germination in soybean cultivar Williams (Figure S2B). Remarkably, almost all genes were found to be unexpressed in both light intensity condition (mature and immature leaf) as well as in ozone study. Overall, the analyses revealed that GmABCs highest level of expression in dehydration suggests dehydration regulatory function in soybean.

3.5. Gene Ontology Annotations. The GO slim analysis conducted using Blast2GO and result showed the putative participation of GmABCs proteins in diverse biological processes (Figure 4). The majority of proteins were involved in transport, cellular component assembly, and metabolic process. Cellular localization prediction showed that GmABC proteins are equally localized in the cell membrane and in the nucleus.

3.6. Homology Modeling. Sixteen ABCs proteins having higher homology were selected and, consequently, Phyre2 was used to predict the homology modeling (Figure 5). Remarkably, these 16 proteins belong to diverse subfamily. The protein structure of all the GmABCs is modelled at >90% confidence (% residue varied from 71 to 99; Table S4). All the predicted protein structures are considered highly reliable and offer a preliminary base towards understanding the molecular function of GmABC proteins.

3.7. Cis-Regulatory Elements in the Promoters of Soybean ABC Genes. Expression of the gene can be regulated at various levels, such as transcription and posttranslational modification level. *Cis*-regulatory elements are present in the upstream region and specifically control the gene expression in higher plants [65]. Two kb promoter regions for each of the 261 GmABCs were retrieved. Then, we applied the PlantPan2.0 website to analyze the *cis*-elements and total of 298 different *cis*-elements identified in all GmABCs. The analysis identified the presence of several stress-related, development and hormone inducible motifs, such as Homeo Box/ leucine Zipper, MBS (MYB Binding Site), WRKY-domain, NAC-domain, GT1consensus, dehydration responsive element binding factors (DREB), ABRE (Abscisic Acid-Responsive Element), GATA-box, AT-hook containing transcription factors, DNA binding with one finger (DOF), GCN-4, TC-rich, and Skn-1. Identified *cis*-regulatory elements are known to regulate various stress responses, developmental processes, and hormonal signaling. Further, they are found to be distributed randomly in both positive and negative strands of the promoters of these genes. A total of 13 common *cis*-regulatory elements were observed in all GmABCs and highly conserved among all genes (Table S5).

4. Discussion

In the present study, 261 ABC transporter genes were identified in the genome of soybean; among them, 31 members are recognized as soluble proteins without TMDs. Immense variation in size (from 298 to 49491 bps) and gene structure (variable introns number from 1 to 38) indicates divergence within the soybean ABC transporters gene family. The intron/exon organization also signifies their evolutionary history across gene families. The ABC transporter family has been extensively characterized in other plants, with 131 ABCs genes in Arabidopsis [17], 121 in rice [21, 66], 130 in *Zea mays* [67], 179 *Brassica rapa* [68], 135 members in *Vitis vinifera* [69], and 154 members in tomato [70]. As compared with these plants, *G. max* contains more ABC transporter genes, which is not surprising since *G. max* is a paleopolyploid and has undergone two WGD events, as there is a significant number of segmental and tandem duplicated genes and they also contributed to the expansion of the gene family. Further, phylogenetic analysis revealed that 8 subfamilies, among all subfamilies, were clustered together, except ABCI. ABCI were clustered into different clades, suggesting more diverged gene members. Among 261 GmABC proteins, members of ABCG were most abundant followed by ABCB and ABCC. Our analysis corroborates with previous studies, which indicated that ABCG subfamily is the largest and is accountable for the expansion of this gene family in comparison to other eukaryotes [7, 64]. It has been reported that ABCG transporters are also concerned with the adaptation to the land environment [71].

To elucidate the significance of GmABCs gene, we have performed the *in silico* expression analysis in three aspects based on their transcript data: (1) tissue-specific, (2) development stage, (3) germination and stress condition. Tissue-specific expression patterns have been determined via FKPM data, while development and stress conditions are represented by m-RNA sequence data. There are 31 host genes (Table 1 and Figure 3), which were highly expressed in all nine tissues; among them, 15 belong to soluble ABC proteins.

These aforementioned genes may play a key role in tissue development. Furthermore, seven genes are exclusively expressed in nodules. Similarly, three genes are exclusively detected in flowers and have some important role in flower development. Correspondingly, Glyma.02G163200 and Glyma.19G202200 are only expressed in leaves and Glyma.13G044000 and Glyma.16G172400 are only expressed in roots. Glyma.06G154500, Glyma.13G142100, and Glyma.17G041400 are exclusively expressed in seeds, suggesting that they may play some roles in seed development. In addition, 20 genes were found to have no detectable level of expression values possible due to the very low abundance of the RNAs of these host genes.

Approximately 43 show expressions in all seven development stages and few have a high level of expression (shown in Table 2 and Figure S2A). Overall, the analysis revealed that the highest number of genes (~40 genes) was expressed during inflorescence formation stage, suggesting their role in inflorescence development. It is noteworthy that the expression of 37 host genes was not observed in any developmental stage; this may be due to inducible features of these host genes.
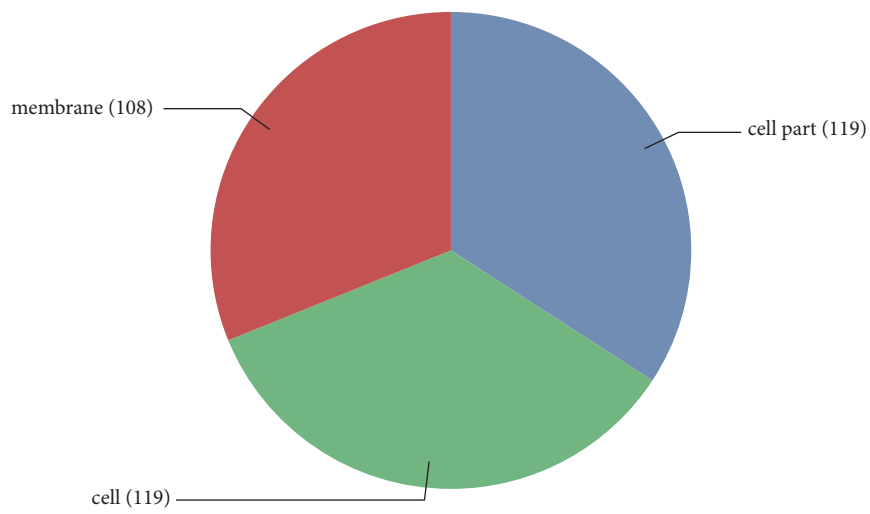
Some genes are expressed in the dehydration stage in cultivar Benning and PI-416937 (shown in Table 2 and Figure S2B). Hence, these ABC genes can play a potential role in dehydration tolerance mechanism. Most of these genes also possess stress-responsive *cis*-acting element in their promoter region. Analogously, at germination stage of some genes in
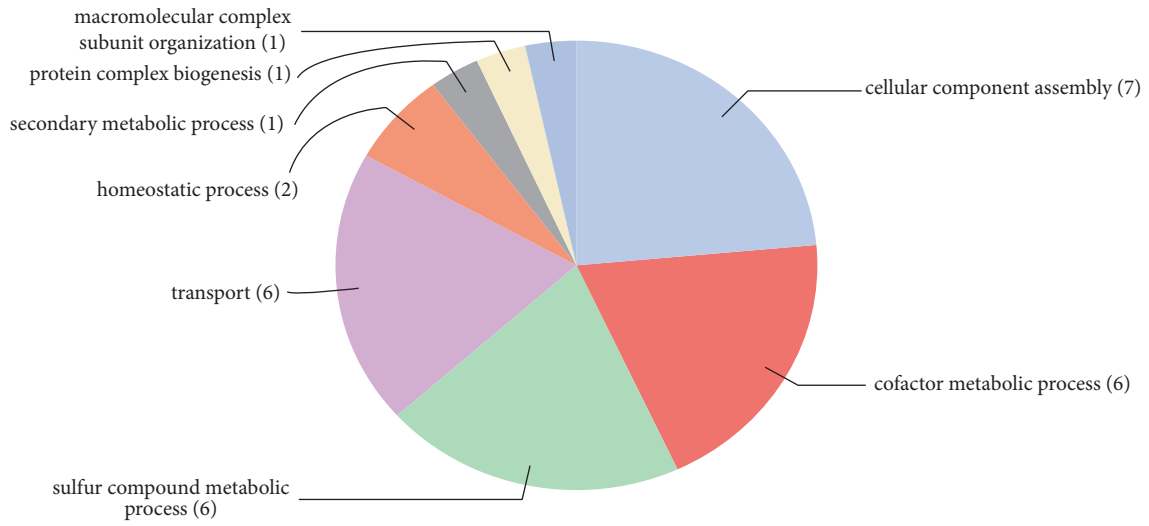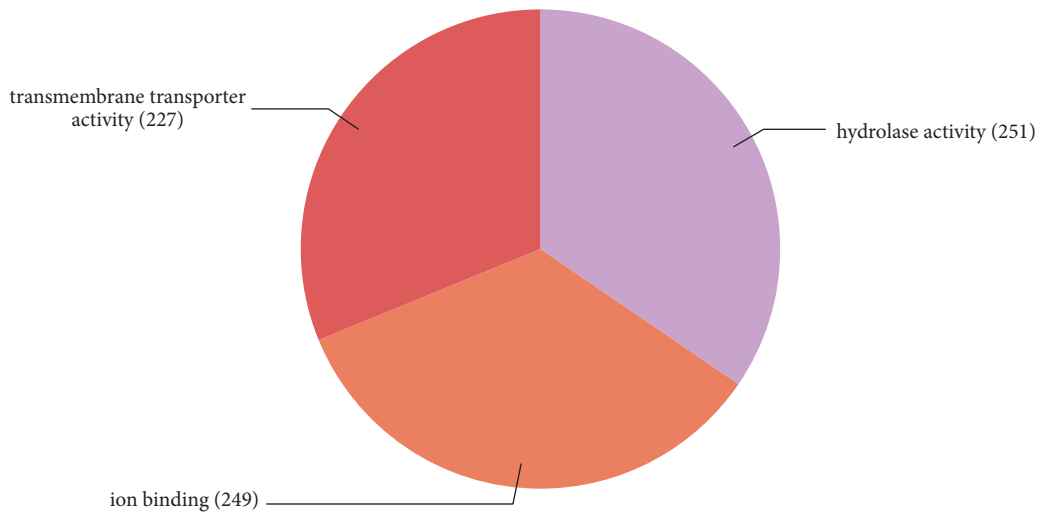
FIGURE 4: **GO annotation of GmABC transporter proteins.** The annotation was performed on three categories: (a) biological process, (b) molecular function, and (c) cellular component.

TABLE 1: List of highly expressed ABC transporter genes in different tissues.

Expressed tissue	Genes
All nine tissues	Glyma.03G136000, Glyma.05G178900, Glyma.05G221600, Glyma.08G028100, Glyma.08G055000, Glyma.08G136400, Glyma.08G194600, Glyma.09G044500, Glyma.11G042400, Glyma.11G090900, Glyma.11G194700, Glyma.11G239800, Glyma.12G079400, Glyma.13G002500, Glyma.13G063700, Glyma.13G074600, Glyma.13G124700, Glyma.13G158600, Glyma.13G163700, Glyma.13G220000, Glyma.13G323500, Glyma.14G207200, Glyma.15G092400, Glyma.17G080400, Glyma.17G107200, Glyma.17G112700, Glyma.18G017600, Glyma.19G021500, Glyma.19G169000, Glyma.20G048400, Glyma.20G066600
Only in Nodules	Glyma.02G159900, Glyma.04G172900, Glyma.04G173000, Glyma.06G269300, Glyma.08G071200, Glyma.12G135300, Glyma.U001700
Only in flowers	Glyma.07G033500, Glyma.08G070700, Glyma.13G222000
Only in leaves	Glyma.02G163200 Glyma.19G202200
Only in roots	Glyma.13G044000, Glyma.16G172400
Only in seeds	Glyma.06G154500, Glyma.13G142100, Glyma.17G041400

TABLE 2: List of GmABC transporter genes expressed during development stage, dehydration, and germination.

Expression condition	Genes
Development stage	Glyma.01G016500, Glyma.03G136000, Glyma.03G167800, Glyma.03G183600, Glyma.04G218900, Glyma.05G178900, Glyma.06G146900, Glyma.07G121900, Glyma.07G233900, Glyma.08G136400, Glyma.08G550000, Glyma.08G194600, Glyma.09G044500, Glyma.09G160000, Glyma.10G019000, Glyma.10G145200, Glyma.11G093900, Glyma.11G239800, Glyma.12G020400,Glyma.12G079400, Glyma.13G163700, Glyma.13G323200, Glyma.15G092400,Glyma.16G209400, Glyma.18G017600, Glyma.18G080900, Glyma.19G021500, Glyma.20G066600
Dehydration	Glyma.02G008000, Glyma.3G101000, Glyma.07G033500, Glyma.07G233900, Glyma.08G136400, Glyma.10G019000, Glyma.10G217300, Glyma.11G194700, Glyma.12G020400, Glyma.12G079400, Glyma.12G177500, Glyma.13G127500, Glyma.15G012000, Glyma.12G0192000,Glyma.16G076900, Glyma.16G110100, Glyma.17G039200, Glyma.17G041200, Glyma.17G112700, Glyma.20G174800
Germination stage	Glyma.01G016500, Glyma.01G026000, Glyma.02G038800, Glyma.02G0298400, Glyma.03G183600, Glyma.04G128900, Glyma.05G011900, Glyma.05G019400, Glyma.05G221600, Glyma.08G000800, Glyma.08G028100, Glyma.08G070700, Glyma.08G160900, Glyma.08G194700, Glyma.08G201300, Glyma.09G206300, Glyma.10G086800, Glyma.10G145200, Glyma.10G203000, Glyma.12G177200, Glyma.13G063700, Glyma.13G142100, Glyma.13G323200, Glyma.17G080400, Glyma.17G120100, Glyma.18G240700, Glyma.19G021500, Glyma.19G184300, Glyma.20G048400, Glyma.20G125600

Glyma.01G154400   Glyma.02G238300   Glyma.05G221600   Glyma.06G191400   Glyma.08G028100   Glyma.10G137600

Glyma.11G090900   Glyma.12G079400   Glyma.12G135300   Glyma.12G177500   Glyma.13G222000   Glyma.15G090200

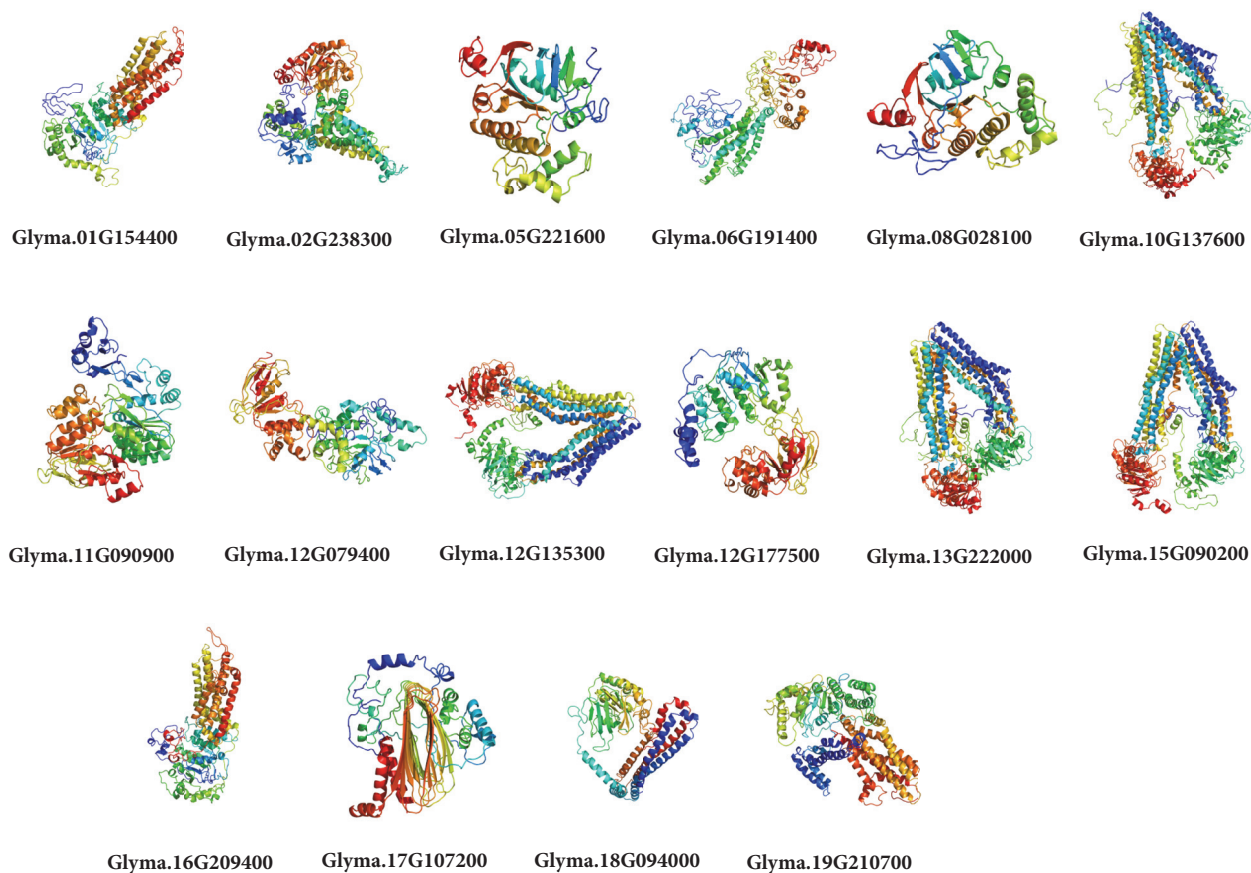Glyma.16G209400   Glyma.17G107200   Glyma.18G094000   Glyma.19G210700

FIGURE 5: Predicted structures of 16 GmABC proteins.

cultivar Williams, few host genes showed higher level of expression (shown in Table 2 and Figure S2B). Still, more experimental evidence is required to find out the specific functions of the ABC transporter gene.

5. Conclusions

The ABC transporter gene family is one of the largest known proteins, ubiquitously present in all living organisms and playing a diverse function such as regulation of growth and resistance processes across the plant kingdom. In the present study, we conducted *in silico* genome-wide annotation of soybean ABC transporter gene family. In total, 261 putative ABC transporter genes were identified and these genes were located unevenly in 20 chromosomes. The number of genes was approximately double that in Arabidopsis and rice. Since segmental duplication, tandem duplication along with WGD might have contributed to the expansion of this gene family. Further, these proteins were classified into 8 key subfamilies and results of our phylogenetic analysis corroborated those of previous studies in another plant. Data files of the physicochemical properties of ABC transporter gene display great variation in gene structure, their length, and domain orientation. The GO analysis revealed multiple functions for the GmABC transporter proteins, including their role in transport and metabolic process. The RNA-seq

and Genevestigator data illustrated that GmABC genes were expressed in various tissues, development stages, and the germination stage, as well as in response to dehydration stresses on the transcriptional level, signifying their diverse function. Besides, *cis*-regulatory element analysis of GmABCs also suggests their role in the development and various abiotic stresses. The results provide comprehensive information for the further functional study of the GmABC gene family.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

Awdhesh Kumar Mishra and Jinhee Choi contributed equally to this work.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Awdhesh Kumar Mishra and Kwang-Hyun Baek conceived and designed the experiments. Awdhesh Kumar Mishra,

Supplementary Materials

Supplementary 1. FIGURE S1: Gene structure of soybean ABC transporter genes. Gene structures were made using online tool GSDS (Hu et al., 2015; http://gsds.cbi.pku.edu.cn/). As shown in the legend, blue bars stand for untranslated regions (UTR). One yellow bar represents an exon and one black string represents an intron. The scale on the bottom is in kilobase (kb).

Supplementary 2. FIGURE S2A: Hierarchical cluster of expression profiles of 261 soybean ABC transporter genes in seven different development stages derived from Genevestigator software (https://www.genevestigator.com/). Expression levels are illustrated by graded color scale, representing the log2 ratio of relative expression pattern ranging from 0 to 100%. For the maroon/white scale, all gene-level profiles were normalized for coloring such that, for each gene, the highest signal intensity obtains the value 100% (maroon), and absence of signal obtains the value 0% (white).

Supplementary 3. FIGURE S2B: Heatmap representation of 261 differentially expressed ABC transporter genes during perturbation derived from Genevestigator V3 (https://www.genevestigator.com/). Expression levels are illustrated by graded color scale, representing the log2 ratio of relative expression pattern ranging from -2.5 (downregulated) to 2.5 (upregulated). Red, green, and black represent positive, negative, and zero, respectively. Gene IDs are given on the top.

Supplementary 4. TABLE S1: List of GmABC transporter genes identified and their basic physical properties.

Supplementary 5. TABLE S2: The ratio of Ka/Ks and distribution of duplicated soybean ABC transporter genes in respective blocks as obtained from Plant Genome Duplication Database (PGDD).

Supplementary 6. TABLE S3: Tandemly duplicated soybean ABC transporter genes obtained from PTGBase database.

Supplementary 7. TABLE S4: Characteristics of 16 GmABC proteins selected for homology modeling.

Supplementary 8. TABLE S5: Common putative *cis*-elements identified in the promoter sequences of GmABCs.

References

[1] E. Dassa and P. Bouige, "The ABC of ABCs: a phylogenetic and functional classification of ABC systems in living organisms," *Research in Microbiology*, vol. 152, no. 3-4, pp. 211–229, 2001.

[2] P. M. Jones and A. M. George, "The ABC transporter structure and mechanism: perspectives on recent research," *Cellular and Molecular Life Sciences*, vol. 61, no. 6, pp. 682–699, 2004.

[3] D. C. Rees, E. Johnson, and O. Lewinson, "ABC transporters: The power to change," *Nature Reviews Molecular Cell Biology*, vol. 10, no. 3, pp. 218–227, 2009.

[4] C. F. Higgins, "ABC Transporters: from microorganisms to man," *Annual Review of Cell and Developmental Biology*, vol. 8, pp. 67–113, 1992.

[5] E. Martinoia, E. Grill, R. Tommasini, K. Kreuz, and N. Amrhein, "ATP-dependent glutathione S-conjugate 'export' pump in the vacuolar membrane of plants," *Nature*, vol. 364, no. 6434, pp. 247–249, 1993.

[6] J. Kang, J. Park, H. Choi et al., "Plant ABC Transporters," *The Arabidopsis Book*, vol. 9, Article ID e0153, 2011.

[7] G. Andolfo, M. Ruocco, A. Di Donato et al., "Genetic variability and evolutionary diversification of membrane ABC transporters in plants," *BMC Plant Biology*, vol. 15, no. 1, p. 51, 2015.

[8] R. Sánchez-Fernández, T. G. E. Davies, J. O. D. Coleman, and P. A. Rea, "The Arabidopsis thaliana ABC Protein Superfamily, a Complete Inventory," *The Journal of Biological Chemistry*, vol. 276, no. 32, pp. 30231–30244, 2001.

[9] O. Garcia, P. Bouige, C. Forestier, and E. Dassa, "Inventory and comparative analysis of rice and Arabidopsis ATP-binding cassette (ABC) systems," *Journal of Molecular Biology*, vol. 343, no. 1, pp. 249–265, 2004.

[10] K. Hollenstein, D. C. Frei, and K. P. Locher, "Structure of an ABC transporter in complex with its binding protein," *Nature*, vol. 446, no. 7132, pp. 213–216, 2007.

[11] E. Schneider and S. Hunke, "ATP-binding-cassette (ABC) transport systems: Functional and structural aspects of the ATP-hydrolyzing subunits/domains," *FEMS Microbiology Reviews*, vol. 22, no. 1, pp. 1–20, 1998.

[12] O. Lewinson and N. Livnat-Levanon, "Mechanism of Action of ABC Importers: Conservation, Divergence, and Physiological Adaptations," *Journal of Molecular Biology*, vol. 429, no. 5, pp. 606–619, 2017.

[13] A. L. Davidson, E. Dassa, C. Orelle, and J. Chen, "Structure, function, and evolution of bacterial ATP-binding cassette systems," *Microbiology and Molecular Biology Reviews*, vol. 72, no. 2, pp. 317–364, 2008.

[14] A. Licht and E. Schneider, "ATP binding cassette systems: structures, mechanisms, and functions," *Open Life Sciences*, vol. 6, p. 785, 2011.

[15] M. Michieli, D. Damiani, A. Geromin et al., "Overexpression of multidrug resistance-associated p170-glycoprotein in acute non-lymphocytic leukemia," *European Journal of Haematology*, vol. 48, no. 2, pp. 87–92, 1992.

[16] A. M. George and P. M. Jones, "Perspectives on the structure-function of ABC transporters: The Switch and Constant Contact Models," *Progress in Biophysics and Molecular Biology*, vol. 109, no. 3, pp. 95–107, 2012.

[17] P. J. Verrier, D. Bird, B. Burla et al., "Plant ABC proteins - a unified nomenclature and updated inventory," *Trends in Plant Science*, vol. 13, no. 4, pp. 151–159, 2008.

[18] T. S. Lane, C. S. Rempe, J. Davitt et al., "Diversity of ABC transporter genes across the plant kingdom and their potential utility in biotechnology," *BMC Biotechnology*, vol. 16, no. 1, p. 47, 2016.

[19] C. F. Higgins and K. J. Linton, "The ATP switch model for ABC transporters," *Nature Structural & Molecular Biology*, vol. 11, no. 10, pp. 918–926, 2004.

[20] T. G. E. Davies and J. O. D. Coleman, "The Arabidopsis thaliana ATP-binding cassette proteins: An emerging superfamily," *Plant, Cell & Environment*, vol. 23, no. 5, pp. 431–443, 2000.

[21] P. A. Rea, "Plant ATP-binding cassette transporters," *Annual Review of Plant Biology*, vol. 58, pp. 347–375, 2007.

[22] M. Dean, A. Rzhetsky, and R. Allikmets, "The human ATP-binding cassette (ABC) transporter superfamily," *Genome Research*, vol. 11, no. 7, pp. 1156–1166, 2001.

[23] M. Dean and T. Annilo, "Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates," *Annual Review of Genomics and Human Genetics*, vol. 6, pp. 123–142, 2005.

[24] W. Dermauw and T. Van Leeuwen, "The ABC gene family in arthropods: comparative genomics and role insecticide transport and resistance," *Insect Biochemistry and Molecular Biology*, vol. 45, no. 1, pp. 89–110, 2014.

[25] M. Geisler and A. S. Murphy, "The ABC of auxin transport: The role of p-glycoproteins in plant development," *FEBS Letters*, vol. 580, no. 4, pp. 1094–1102, 2006.

[26] D.-Y. Kim, L. Bovet, M. Maeshima, E. Martinoia, and Y. Lee, "The ABC transporter AtPDR8 is a cadmium extrusion pump conferring heavy metal resistance," *The Plant Journal*, vol. 50, no. 2, pp. 207–218, 2007.

[27] J. A. Pighin, H. Zheng, L. J. Balakshin et al., "Plant cuticular lipid export requires an ABC transporter," *Science*, vol. 306, no. 5696, pp. 702–704, 2004.

[28] D. Panikashvili, S. Savaldi-Goldstein, T. Mandel et al., "The arabidopsis DESPERADO/AtWBC11 transporter is required for cutin and wax secretion," *Plant Physiology*, vol. 145, no. 4, pp. 1345–1360, 2007.

[29] A. S. K. Braz, J. Finnegan, P. Waterhouse, and R. Margis, "A plant orthologue of RNase L inhibitor (RLI) is induced in plants showing RNA interference," *Journal of Molecular Evolution*, vol. 59, no. 1, pp. 20–30, 2004.

[30] M. Messina, "Insights Gained from 20 Years of Soy Research," *Journal of Nutrition*, vol. 140, no. 12, pp. 2289S–2295S, 2010.

[31] J. Schmutz, S. B. Cannon, J. Schlueter et al., "Genome sequence of the palaeopolyploid soybean," *Nature*, vol. 463, no. 7278, pp. 178–183, 2010.

[32] R. D. Finn, P. Coggill, R. Y. Eberhardt et al., "The Pfam protein families database: towards a more sustainable future," *Nucleic Acids Research*, vol. 44, no. 1, pp. D279–D285, 2016.

[33] D. M. Goodstein, S. Shu, R. Howson et al., "Phytozome: a comparative platform for green plant genomics," *Nucleic Acids Research*, vol. 40, no. 1, pp. D1178–D1186, 2012.

[34] I. Letunic, T. Doerks, and P. Bork, "SMART 7: recent updates to the protein domain annotation resource," *Nucleic Acids Research*, vol. 40, no. 1, pp. D302–D305, 2012.

[35] E. de Castro, C. J. A. Sigrist, A. Gattiker et al., "ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins," *Nucleic Acids Research*, vol. 34, pp. W362–W365, 2006.

[36] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, and A. Bairoch, "ExPASy: the proteomics server for in-depth protein knowledge and analysis," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3784–3788, 2003.

[37] R. E. Voorrips, "Mapchart: software for the graphical presentation of linkage maps and QTLs," *Journal of Heredity*, vol. 93, no. 1, pp. 77-78, 2002.

[38] J. Yu, T. Ke, S. Tehrim, F. Sun, B. Liao, and W. Hua, "PTGBase: An integrated database to study tandem duplicated genes in plants," *Database*, vol. 2015, p. bav017, 2015.

[39] T.-H. Lee, H. Tang, X. Wang, and A. H. Paterson, "PGDD: a database of gene and genome duplication in plants," *Nucleic Acids Research*, vol. 41, no. 1, pp. D1152–D1158, 2013.

[40] M. Lynch and J. S. Conery, "The evolutionary fate and consequences of duplicate genes," *Science*, vol. 290, no. 5494, pp. 1151–1155, 2000.

[41] M. Lynch and J. S. Conery, "The evolutionary demography of duplicate genes," *Journal of Structural and Functional Genomics*, vol. 3, pp. 35–44, 2003.

[42] B. Hu, J. Jin, A.-Y. Guo, H. Zhang, J. Luo, and G. Gao, "GSDS 2.0: an upgraded gene feature visualization server," *Bioinformatics*, vol. 31, no. 8, pp. 1296-1297, 2015.

[43] J. D. Thompson, T. Gibson, and D. G. Higgins, "Multiple sequence alignment using ClustalW and ClustalX," *Current Protocols in Bioinformatics*, pp. 2.3.1–2.3.22, 2002.

[44] S. Guindon and O. Gascuel, "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood," *Systematic Biology*, vol. 52, no. 5, pp. 696–704, 2003.

[45] K. Tamura, G. Stecher, D. Peterson, A. Filipski, and S. Kumar, "MEGA6: Molecular Evolutionary Genetics Analysis version 6.0," *Molecular Biology and Evolution*, vol. 30, no. 12, pp. 2725–2729, 2013.

[46] A. Conesa and S. Götz, "Blast2GO: a comprehensive suite for functional analysis in plant genomics," *International Journal of Plant Genomics*, vol. 2008, Article ID 619832, 12 pages, 2008.

[47] A. I. Saeed, V. Sharov, J. White et al., "TM4: a free, open-source system for microarray data management and analysis," *BioTechniques*, vol. 34, no. 2, pp. 374–378, 2003.

[48] T. Hruz, O. Laule, G. Szabo et al., "Genevestigator V3: A Reference Expression Database for the Meta-Analysis of Transcriptomes," *Advances in Bioinformatics*, vol. 2008, Article ID 420747, 5 pages, 2008.

[49] H. M. Berman, J. Westbrook, Z. Feng et al., "The protein data bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000.

[50] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg, "The Phyre2 web portal for protein modeling, prediction and analysis," *Nature Protocols*, vol. 10, no. 6, pp. 845–858, 2015.

[51] B. R. Jefferys, L. A. Kelley, and M. J. E. Sternberg, "Protein folding requires crowd control in a simulated cell," *Journal of Molecular Biology*, vol. 397, no. 5, pp. 1329–1338, 2010.

[52] C.-N. Chow, H.-Q. Zheng, N.-Y. Wu et al., "PlantPAN 2.0: An update of Plant Promoter Analysis Navigator for reconstructing transcriptional regulatory networks in plants," *Nucleic Acids Research*, vol. 44, no. 1, pp. D1154–D1164, 2016.

[53] A. Zhou, B. A. Hassel, and R. H. Silverman, "Expression cloning of 2-5A-dependent RNAase: a uniquely regulated mediator of interferon action," *Cell*, vol. 72, no. 5, pp. 753–765, 1993.

[54] S. B. Cannon, A. Mitra, A. Baumgarten, N. D. Young, and G. May, "The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana," *BMC Plant Biology*, vol. 4, no. 1, p. 10, 2004.

[55] S. Ohno, *Evolution by Gene Duplication*, Springer-Verlag, New York, NY, USA, 1970.

[56] T. J. Vision, D. G. Brown, and S. D. Tanksley, "The origins of genomic duplications in Arabidopsis," *Science*, vol. 290, no. 5499, pp. 2114–2117, 2000.

[57] R. C. Shoemaker, K. Polzin, J. Labate et al., "Genome duplication in soybean (Glycine subgenus soja)," *Genetics*, vol. 144, no. 1, pp. 329–338, 1996.

[58] J. L. Shultz, D. Kurunam, K. Shopinski et al., "The Soybean Genome Database (SoyGD): a browser for display of duplicated, polyploid, regions and sequence tagged sites on the integrated physical and genetic maps of Glycine max," *Nucleic Acids Research*, vol. 34, pp. D758–D765, 2006.

[59] J. F. Wendel, "Genome evolution in polyploids," *Plant Molecular Biology*, vol. 42, no. 1, pp. 225–249, 2000.

[60] A. H. Paterson, J. E. Bowers, and B. A. Chapman, "Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 101, no. 26, pp. 9903–9908, 2004.

[61] H. Du, S.-S. Yang, Z. Liang et al., "Genome-wide analysis of the MYB transcription factor superfamily in soybean," *BMC Plant Biology*, vol. 12, article no. 106, 2012.

[62] S. Bian, X. Li, H. Mainali, L. Chen, and S. Dhaubhadel, "Genome-wide analysis of DWD proteins in soybean (Glycine max): Significance of Gm08DWD and GmMYB176 interaction in isoflavonoid biosynthesis," *PLoS ONE*, vol. 12, no. 6, Article ID e0178947, 2017.

[63] Q. Jia, Z.-X. Xiao, F.-L. Wong, S. Sun, K.-J. Liang, and H.-M. Lam, "Genome-wide analyses of the soybean F-box gene family in response to salt stress," *International Journal of Molecular Sciences*, vol. 18, no. 4, p. 818, 2017.

[64] P. Chen, Y. Li, L. Zhao et al., "Genome-wide identification and expression profiling of atp-binding cassette (ABC) transporter gene family in pineapple (ananas comosus (L.) merr.) reveal the role of AcABCG38 in pollen development," *Frontiers in Plant Science*, vol. 8, p. 2150, 2017.

[65] V. Vedel and I. Scotti, "Promoting the promoter," *Journal of Plant Sciences*, vol. 180, no. 2, pp. 182–189, 2011.

[66] V. N. T. Nguyen, S. Moon, and K.-H. Jung, "Genome-wide expression analysis of rice ABC transporter family across spatio-temporal samples and in response to abiotic stresses," *Journal of Plant Physiology*, vol. 171, no. 14, pp. 1276–1288, 2014.

[67] K. Pang, Y. Li, M. Liu, Z. Meng, and Y. Yu, "Inventory and general analysis of the ATP-binding cassette (ABC) gene superfamily in maize (Zea mays L.)," *Gene*, vol. 526, no. 2, pp. 411–428, 2013.

[68] C. Yan, W. Duan, S. Lyu, Y. Li, and X. Hou, "Genome-Wide Identification, Evolution, and Expression Analysis of the ATP-Binding Cassette Transporter Gene Family in Brassica rapa," *Frontiers in Plant Science*, vol. 8, p. 349, 2017.

[69] B. Çakir and O. Kiliçkaya, "Whole-genome survey of the putative ATP-binding cassette transporter family genes in Vitis vinifera," *PLoS ONE*, vol. 8, no. 11, Article ID e78860, 2013.

[70] P. A. Ofori, A. Mizuno, M. Suzuki et al., "Genome-wide analysis of ATP binding cassette (ABC) transporters in tomato," *PLoS ONE*, vol. 13, no. 7, pp. 1–26, 2018.

[71] J.-U. Hwang, W.-Y. Song, D. Hong et al., "Plant ABC Transporters Enable Many Unique Aspects of a Terrestrial Plant's Lifestyle," *Molecular Plant*, vol. 9, no. 3, pp. 338–355, 2016.

Research Article

Development of Clustered Resistance Gene Analogs-Based Markers of Resistance to *Phytophthora capsici* in Chili Pepper

Nayoung Kim [ID],[1] Won-Hee Kang,[2] Jundae Lee [ID],[3] and Seon-In Yeom [ID][1,2]

[1]*Department of Agricultural Plant Science, Division of Applied Life Science (BK21 Plus Program),*
 Gyeongsang National University, Jinju 52828, Republic of Korea
[2]*Institute of Agriculture & Life Science, Gyeongsang National University, Jinju 52828, Republic of Korea*
[3]*Department of Horticulture, Chonbuk National University, Jeonju 54896, Republic of Korea*

Correspondence should be addressed to Jundae Lee; ajfall@jbnu.ac.kr and Seon-In Yeom; sunin78@gnu.ac.kr

The soil-borne pathogen *Phytophthora capsici* causes severe destruction of *Capsicum* spp. Resistance in *Capsicum* against *P. capsici* is controlled by numerous minor quantitative trait loci (QTLs) and a consistent major QTL on chromosome 5. Molecular markers on *Capsicum* chromosome 5 have been developed to identify the predominant genetic contributor to resistance but have achieved little success. In this study, previously reported molecular markers were used to reanalyze the major QTL region on chromosome 5 (6.2 Mbp to 139.2 Mbp). Candidate resistance gene analogs (RGAs) were identified in the extended major QTL region including 14 nucleotide binding site leucine-rich repeats, 3 receptor-like kinases, and 1 receptor-like protein. Sequence comparison of the candidate RGAs was performed between two *Capsicum* germplasms that are resistant and susceptible, respectively, to *P. capsici*. 11 novel RGA-based markers were developed through high-resolution melting analysis which were closely linked to the major QTL for *P. capsici* resistance. Among the markers, CaNB-5480 showed the highest cosegregation rate at 86.9% and can be applied to genotyping of the germplasms that were not amenable by previous markers. With combination of three markers such as CaNB-5480, CaRP-5130 and CaNB-5330 increased genotyping accuracy for 61 *Capsicum* accessions. These could be useful to facilitate high-throughput germplasm screening and further characterize resistance genes against *P. capsici* in pepper.

1. Introduction

Hot pepper (*Capsicum* spp.) is an economically important crop that belongs to the *Solanaceae* family along with tobacco, potato, and tomato. Hot pepper provides many essential vitamins, and capsaicin is used as a major spicy flavoring in most global cuisines [1]. In 2016, the main pepper-producing countries grew 38 million tons on about 3.7 Mha [2]. The world production and trade value of hot pepper consistently increased during the last decade.

Pathogenic fungi, oomycetes, bacteria, and viruses cause economic damage and crop loss. Pepper production in negatively impacted by approximately 87 different pathogens and diseases [3]. Among them, *Phytophthora capsici* is a highly destructive, broad-host-range oomycetes that was first described in pepper in New Mexico in 1922 [4]. The host range of *P. capsici* includes Solanaceae, Cucurbitaceae,

lima beans, and other plants. *P. capsici* is found in North America, South America, Asia, Africa, and Europe. *P. capsici* causes severe disease symptoms such as foliar blight, stem blight, and root, stem, fruit, and foliar rot. The economic impact of *P. capsici* on worldwide vegetable production has been valued at over one billion dollars per year [5].

P. capsici forms oospores that survive for years in the soil and can migrate through water and wind during warm (25–28°C) and wet conditions such as those during rainy seasons [6]. After *P. capsici* becomes established at a location, it can be very difficult to control. There is no effective chemical and agricultural strategy. Resistance against *P. capsici* has been reported in cultivated peppers including cultivars 'AC2258,' 'PI201232,' 'PI201234,' and 'CM334' [7, 8]. Among the resistant cultivars, *Capsicum annuum* 'CM334' shows a very high degree of resistance to multiple races of *P. capsici*

and reported resistance to root rot, crown rot, fruit rot, and foliar blight [9–11].

Previous studies indicated that resistance of pepper to *P. capsici* is under polygenic control governed by complex quantitative trait loci (QTLs) [12–15]. Identification and QTL mapping of genes conferring partial resistance against *P. capsici* in pepper is essential for breeding cultivars with *P. capsici* resistance. The majority of QTLs for *P. capsici* resistance are conserved on *Capsicum* chromosome 5. Numerous molecular markers have been developed within the major QTL region, but little progress has been made in developing functional gene-based markers of *P. capsici* resistance, and no *P. capsici*-resistance gene (*R* gene) has yet been characterized. Recently, high-quality *de novo* sequenced *Capsicum* genomes, including the 'CM334' genome, were reported [16, 17]. This genomic information from the hot pepper could be applied to plant breeding and the selection of crops with specific traits such as disease resistance or large fruit size [16, 18, 19]. The development of functional gene-based markers and the further characterization of genes that confer resistance against *P. capsici* is key factor for efforts to breed disease-resistant pepper cultivars.

Plants have multiple layers of defense against pathogen attacks, including preformed barriers for continuous defense and programmed immune responses based on pathogen recognition [20–22]. Resistance gene analogs (RGAs) are proteins involved in plant immune responses. RGAs have highly conserved structures, which include nucleotide binding site leucine-rich repeat (NBS-LRR) proteins, receptor-like proteins (RLPs), and receptor-like kinases (RLKs) [21, 23, 24]. RGAs have been cloned as resistance genes in various plants such as wheat (*Lr10 to pathogen*), barley (*Mla6*), rice (*Xa1*), maize (*Rp1-D*), Arabidopsis (*RPM1*), lettuce (*Rgc2*), tobacco (*N*), and tomato (*Prf*) [25–32]. RGAs can be used to identify and characterize *R* genes.

In this study, the sequences of previously developed markers were screened to extend the major QTL region on *Capsicum* chromosome 5 and the RGAs were reanalyzed within that region to identify candidate resistance genes. Marker development on candidate genes was performed by comparing single nucleotide polymorphisms (SNPs) between the *P. capsici*-resistant cultivar *C. annuum* 'CM334' and the *P. capsici*-susceptible cultivar *C. annuum* 'Daepoongcho'. A total of 61 *Capsicum* accessions were used to validate the newly developed RGA-based markers through high-resolution melting (HRM) analysis. These markers could be useful to validate the genotyping of germplasms and to further characterize resistance genes against *P. capsici*.

2. Materials and Methods

2.1. Plant Materials. 61 *Capsicum* accessions (44 resistant, 3 moderately resistant, and 14 susceptible cultivars; Table S1) [33–38] were used to validate cosegregation of molecular markers. All plants were grown in 32-cell trays filled with soil. The plants were kept in a growth chamber at 25°C with a 16 h/8 h (light/darkness) photoperiod.

2.2. Pathogen Preparation and Plant Inoculation with P. capsici. The preparation of the *P. capsici* inoculum was described in previous study [11]. *P. capsici* was grown for 7 days in potato dextrose agar medium at 27°C and mycelial plugs (6 mm in diameter) were transferred on V8 agar medium for zoospore production. After 5 days, the mycelia grown on V8 agar medium were damaged using spreader and incubated under fluorescent lights for 2 days. The zoosporangia were shocked by incubating at 4°C with sterile water for 1 h 30 min to initiate release zoospores, followed by 30 min at 28°C for equilibration. The zoospores were collected and the concentration was adjusted to 1×10^5 zoospores · mL^{-1} counted by hemocytometer, and 2 mL of suspension was drenched at the root of each six-true-leaf stage of pepper plant. The inoculated pepper plants were kept at 25°C with a 16-h light photoperiod condition. Evaluation of disease symptoms was assessed using disease index of 0 to 3 scale (0 = no symptom; 1 = leaf wilting but no necrosis or less than 30% of the leaf wilted; 2 = leaf wilting and slightly necrosis stem or less than 60% of the leaf wilted; 3 = plant dead) [37]. The classification of the pepper lines, as resistant, moderately resistant or susceptible, was made based on the average disease index of each line, where if the scale was < 1, the pepper line was considered as resistant (R), $1 \leq$ disease index < 2 confer as moderately resistant (MR), and if the scale was < 3 as susceptible (S). The results of the phenotype correspond to those as shown in Kim *et al.*, 2017 [38].

2.3. Genomic DNA Extraction. Genomic DNA was extracted from the young leaves of plant samples using a slightly modified cetyltrimethylammonium bromide (CTAB) method [39]. First, the leaves were grinded using a pestle and mixed them with CTAB buffer, polyvinylpyrrolidone, and β-mercaptoethanol. The samples were incubated at 65°C for 1 h, added chloroform with isoamyl alcohol (24:1), and centrifuged them at 4°C 15,814 *g* for 15 min. Supernatant was transferred to a new 1.5 mL tube and incubated it for 30 min at −20°C with isopropyl alcohol. The precipitated genomic DNA was washed using 70% ethanol and centrifuged it at 4°C 15,814*g* for 10 min. DNA pellet was dissolved in deionized water and treated it with ribonuclease A. Concentration of genomic DNA was measured by a NanoDrop™ Spectrophotometer (ND-2000, ThermoFisher Scientific, Waltham, MA, USA) and then diluted the sample to a final DNA concentration of 20 ng·μL^{-1}.

2.4. Major QTL-Related Marker Analysis. The DNA sequences of 5 simple sequence repeat (SSR), 8 cleaved amplified polymorphic sequence (CAPS), and 18 SNP markers from the published maps of pepper chromosome 5 were used and compared with the *C. annuum* 'CM334' genome version 1.55 (http://genome.pepper.snu.ac.kr/) using BLASTn. The major QTL region on chromosome 5 was extended including all of the previously identified molecular markers and then candidate resistance genes from the extended major QTL region were selected.

2.5. Candidate Gene Selection and High-Resolution Melting Analysis. RGAs were identified according to the domain

structure of the genes in the extended major QTL region. Domain structure analysis was conducted using SMART (http://smart.embl-heidelberg.de/) and Pfam (http://pfam.xfam.org/). The RGA sequences were expanded including an additional 1 kb from both ends of the sequences using an in-house pipeline. A multiple sequence alignment tool MUSCLE (http://www.ebi.ac.uk/) was used to identify SNPs between *C. annuum* 'CM334' and *C. annuum* 'Daepoongcho' for developing RGA-based markers. The candidate genes were selected based on the locations of the SNPs within the expanded QTL region.

Peptide and DNA sequences were downloaded from the Pepper Genome Platform (http://genome.pepper.snu.ac.kr/) and designed primers using primer3plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi; Table 1) for HRM analysis of molecular markers. HRM analysis was performed to validate the cosegregation of the newly developed RGA-based markers among 61 *Capsicum* accessions using LightCycler® Real-Time PCR (Roche, Basel, Switzerland). The reaction solution had a total volume of 20 μL and contained 40 ng genomic DNA, 1 μL each of two primers at 10 pmol·μL^{-1}, 0.1 μL EasyTaq® DNA polymerase (Transgene Biotech, Beijing, China), 2 μL 10× EasyTaq® buffer (Transgene Biotech, Beijing, China), 1 μL 2.5 mM dNTPs (Transgene Biotech, Beijing, China), 1 μL SYTO®9 green fluorescent nucleic acid stain (Life Technologies, Carlsbad, CA, USA), and sterilized water. The PCR conditions were 5 min denaturation at 95°C followed by 10 s at 95°C and annealing with extension 20 s at 60°C for 40 cycles. The melting curve stage was a progression of 95°C for 1 min, 40°C for 1 min, and 65°C for 1 s with fluorescence estimated at 0.2°C intervals. Genotyping of the molecular markers was analyzed through High-Resolution Melt software version 1.1 (Roche, Basal, Switzerland).

3. Results and Discussion

3.1. Candidate RGA Markers with Integration of Genetic and Genomic Data on Pepper Chromosome 5.

Genetic and genomic data on pepper chromosome 5 were integrated to determine the extended region of a major resistance QTL. The following DNA sequences and QTL information were used in this study: CAPS markers mapped to the *Pc.5.2, Pc.5.3* region of 'H3' × 'Vania' (HV), and 'Perenial' × 'Yolo Wonder' (PY) [40]; SNP markers within the *Pc5.1* region of 'Early Jalapeno' × 'CM334' (EC), the *Phyto5* region of 'YCM334' × 'Tean,' and the *Pc5.2* region of 'CM334' × 'Chilsungcho' (CC) and 'NB1' × 'Bhut Jolokia' [41–44]; and SSR markers mapped to the *Pc5.1* region of 'Manganji' × 'CM334' [45]. The genomic locations of the previously developed markers were identified using BLASTn with the *C. annuum* 'CM334' genome. The primer sequences and the genomic location of several representative markers are shown in Table 1.

A physical map of chromosome 5 was drawn including the locations of the markers and obtained the extended major QTL region for the selection of candidate resistance genes. The extended QTL region spanned from 6.2 Mbp (U196349) to 139.2 Mbp (P5-SNAP-CM) and contained 10 QTLs and 845 genes from 56 scaffolds (Table 1 and Figure 1). The

genes included NB-ARC domain containing protein, GRAS family transcription factor, cytochrome P450, and proteins of unknown function.

Using the domain analysis tool, 18 RGAs in the extended QTL region including 14 NBS-LRR proteins, 3 RLKs, and 1 RLP were identified (Table S2). Previously, Rehrig *et al.*, 2014, reported several RGAs nearby QTL peak on chromosome 5 using 'CM334' genome (ver.1.5), but they did not developed RGA-based markers to evaluate *P. capsici*-resistance resources. In this study, additional five RGAs were identified from the updated 'CM334' genome (ver. 1.55) as shown in Table S2. The 14 NBS-LRR proteins consisted of seven partial type NBS-LRRs and seven full type NBS-LRRs. All three RLKs had a kinase-TM-kinase domain. The RLP had a signal peptide-LRR domain. The majority of candidate genes that have been cloned as *R* genes in plant species are RGAs. To date, more than 314 functional *R* genes have been identified in plants [46]. Among them, 80% encode NBS-LRR proteins (191/314) or RLPs/RLKs (60/314). In *Solanaceae*, all of the cloned *R* genes against *Phytophthora infestans* were identified RGAs such as *Rpi-blb1, Rpi-blb2, R2, R3a,* and *ELR* [47–51]. Therefore, the identification of RGAs and development of RGA-based markers could be useful in the further characterization of *P. capsici*-resistance genes.

3.2. Development of SNP Markers.

Eighteen RGAs were reanalyzed to identify SNPs between resistant and susceptible *P. capsici* germplasms by multiple sequence alignment. Among the 18 RGAs, 11 had SNPs and were selected as candidate genes for the development of molecular markers (Table 2 and Table S2). The RGA-based markers had four SNP types (A/C, A/G, T/C, and T/G) and had from one to three SNPs between the resistant and susceptible germplasms (Table 2). Using HRM analysis, nine NBS-LRR-based markers (CaNB-5390, CaNB-5410, CaNB-5440, CaNB-5480, CaNB-5550, CaNB-5720, CaNB-5330, CaNB-5530, and CaNB-5170), one RLK-based marker (CaRK-5470), and 1 RLP-based marker (CaRP-5130) were developed. The HRM curves clearly distinguished among three genotypes for each of the markers: resistant homozygous, resistant heterozygous, and susceptible homozygous (Figure 2).

The RGA-based SNP markers were applied to genotyping of 61 *Capsicum* accessions with previously developed markers such as 142964 [42] and Phyto5SAR [43]. In a previous report, the region spanning from 20.2 Mbp to 29.29 Mbp on chromosome 5 was reported as the major core QTL region [43]. Phyto5SAR was closely linked to the major core QTL, and was used to assess the genotyping accuracy of the new developed markers. The genotyping results of the newly developed markers are shown on Table 3 and Table S3, which were divided into four groups as genotyping accuracy on cultivars. The phenotypes in cultivars belong to Group 1 and Group 2 were matched with genotyping results by six markers (Table 3). The cosegregation rates varied from 37.7% to 86.9% (Figure 3(a)). Several markers adjacent to Phyto5SAR had a tendency to show a high cosegregation rate, which were CaRP-5130 (82%), CaNB-5330 (83.6%), CaRK-5470 (72.1%), and CaNB-5530 (82%). Although far from Phyto5SAR on the physical map, CaNB-5480 showed the highest cosegregation

TABLE 1: Information about the molecular markers within the extended major QTL region on chromosome 5.

Marker	QTL	Genome location (Mbp)	Forward primer (5′-3′)	Reverse primer (5′-3′)	References
UI96349	HV *Pc5.3*	6.17	CCTGGGAGAGGAGTCCTACA	GCAAGAAACAGCGCGCTTTAG	[40]
UI98I14	HV *Pc5.3*	11.35	TTGGGCTCAATTAAACCATACA	GCACCCCTTGATTGAGAGAA	[40]
CA_036I00[†]	EC *Pc5.1*	14.79	GTATATATTTATCAAAATTAATAATCTAGAGAGTATTGAGAGTGCTTAGTGAGAGATTAGC-AAATTAGCAACAATGGGTAAATTGATTTTTGTAGCAAT		[44]
UI97890	PY *Pc5.3*	14.93	CCATATGGTCCTCCTCCAGA	CTTTCAACCACTTCGGCAAT	[40]
CONTIGI896	*Phyto5*	19.99	CATAACACAACCCAATTGCAGAACC	GTCCTACACTTGTTACAGCTGCC	[43]
CA_028982[†]	EC *Pc5.1*	20.02	GTGGGAGGACGAATGTGGTATTTCTCAACCCCAACAAAGAGAAAAAAACAGTAAAAATAATT-TTGAGGAACTAAGGATAAGAAATGCAACAATGGAAGGC		[44]
CaNB-5390		20.2	AAGCCGGGTCTTATAAACTTCCAT	CTTGTGAACATCCAAAGTTAGGGG	In this study
CaNB-5410		20.25	TAGAATCCATCAACCGTACGTAGA	ATTGAGTTAGGTGGCATCTTGACT	In this study
CaNB-5440		20.3	GTTCGACGAGGCTCTACTTTGA	TCAGTCTGCCAGGATTAACAATGA	In this study
CaNB-5480		20.51	TCGAAATCAATACTCCTCCTTCCC	CTCACGCGTTGTTCTTAAAAGGTT	In this study
CAMS319	MC *Pc5.1*	20.57	CATAACACAACCCAATTGCAGAACC	CATAACACAACCCAATTGCAGAACC	[45]
CaNB-5500		20.79	TCATTTATGCGTGCAGTATCGTAC	AAATATACCCCTCAATGACACACA	In this study
CaNB-5720		22.5	GGAAGTCCATTTTATCTATCCTCAGA	ATGATTTTGGGGGATTTCAATAGT	In this study
907783	*Phyto5*	23.78	GACTCCATTTTCATGTTAGCGTGC	CTTGCACCTGGTAAACAACGC	[43]
CaRP-5130		26.43	GACACATTTTGCAGATTCATCAAC	CACCCAAAAAGGTAAAAAGAAACA	In this study
CONTIGI820	*Phyto5*	26.44	GCAAGGAGAAATCAACAGAGCC	AATACAAAGAAACGCAGTAGGAGGG	[43]
CaNB-5330		27.63	AAGAAAGCCGTCACCTTCATAGAT	GCTAATTTGCAAGGATTACGCTCA	In this study
CaRK-5470		28.5	TTTCAATATCCAAAGAACACAACA	GTGAACCCAATGTGATAGTGAAAG	In this study
Phyto5SAR	*Phyto5*	29.29	GGCACAACAATAGTCACAACGG	GAGACTAAGAAAGTTGGACGCC	[43]
CaNB-5530		30.19	GCTCCGTGTTCATCTATGTTAAACA	AACAGAAAGGATCCATCAGCATCAT	In this study
CAMS420	MC *Pc5.1*	32.04	CATAACACAACCCAATTGCAGAACC	CATAACACAACCCAATTGCAGAACC	[45]
I42964		32.29	AAAGCTTTTTACATCTCATTCAACG	AGTTTGAGCAACATAGATGTGGAG	[42]
CaNB-5170		34.56	TATCTCCTACATCCCACTGACAAC	CTTGTCATGGGTGGTGCATTAATG	In this study
Sn-2	HV *Pc5.2*	47.72	TTCGATCCACACCATCATCT	TCCTTCAATGGCTTTCCATC	[40]
P5-SNAP-CM	CC *Pc5.2*	139.2	TCATGAGGTTGCTATTAAGATTGGTCCTGTTAT-ATA	CATAGAAAGGGATATCATCTGGTACAATGCA-GAAA	[41]

Table 1 contains information about newly developed markers and several representative markers among the 31 previously developed markers. [†] Primers for CA_036I00 and CA_028982 are flanking sequences of CM334.
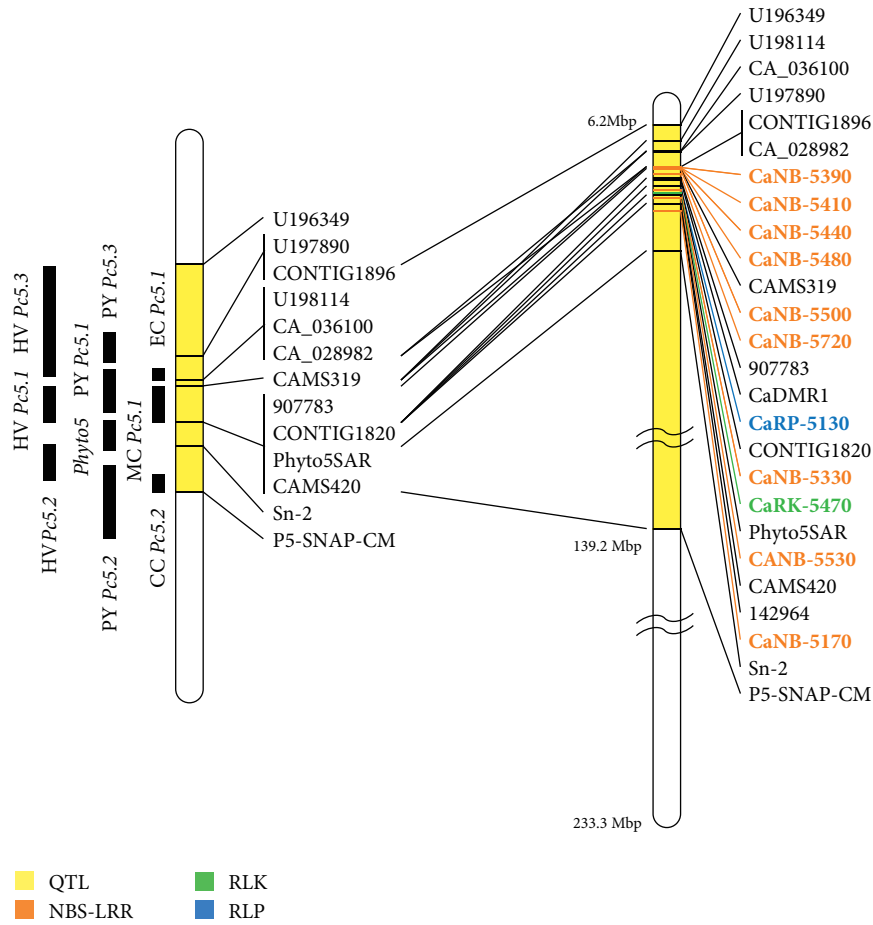
FIGURE 1: Integration of genetic and genomic data on chromosome 5. The genetic map of chromosome 5 is on the left, and the physical map is on the right. Quantitative trait loci (QTL) of the previously developed markers are shown on left side of the genetic map. The physical map is drawn to show the actual chromosome size, which is 233.3 Mbp in 'CM334.' The extended QTL region spans from 6.2 Mbp to 139.2 Mbp and in shown in yellow color. The other colors indicate information about the RGA-based markers within the QTL region, which are shown on the right side of the physical map. Previously developed markers are integrated with the genetic map and physical map with their actual locations on chromosome 5. Orange: NBS-LRR-based marker, green: RLK-based marker, and blue: RLP-based marker.
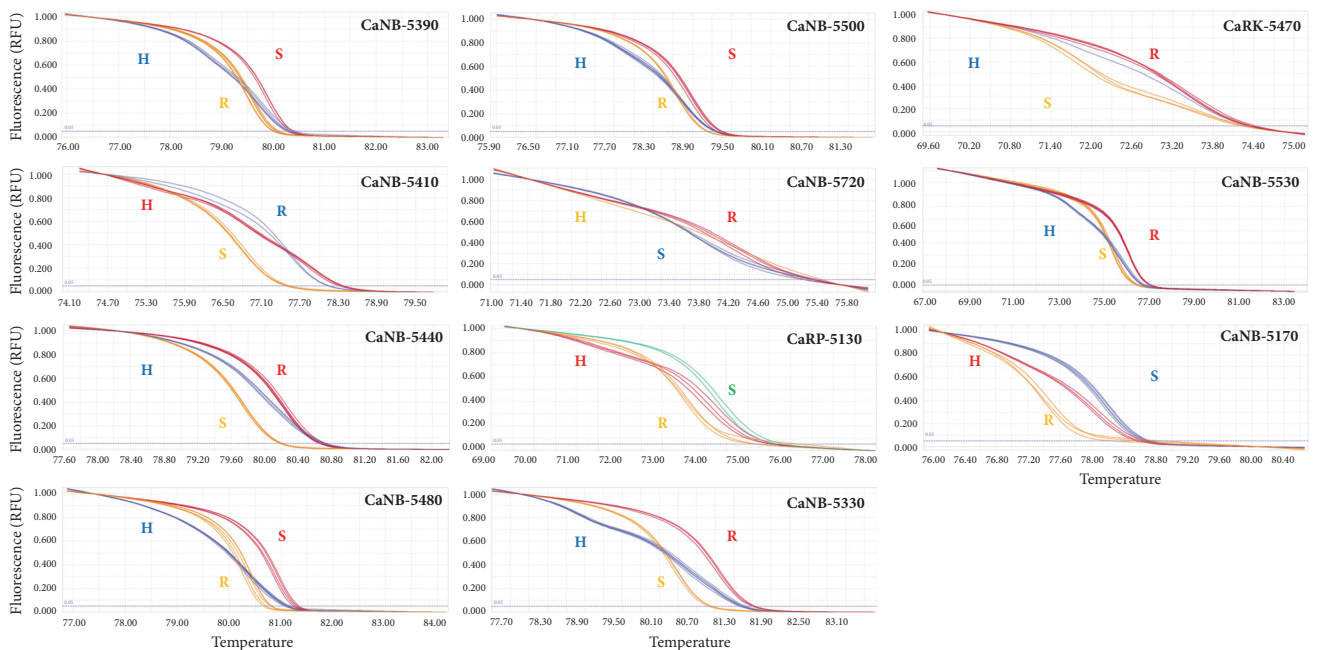

FIGURE 2: High-resolution melting (HRM) curves of the newly developed markers. Sixty-one *Capsicum* accessions were used to perform HRM analysis, and curves represent several germplasms as genotypes. The x-axis indicates temperature, and the y-axis shows fluorescence. R: resistant homozygote, H: resistant heterozygote, and S: susceptible homozygote.

TABLE 2: Comparison of SNPs between *P. capsici*-resistant and *P. capsici*-susceptible germplasms.

Marker	'CM334' allele[‡]	'Daepoongcho' allele[§]	Flanking sequence
CaNB-5390	T	C	AAGCCGGGTCTTATAAACTTCCATAGTGGTTTCCAACATGG-CATAATACATAAAAATGCCCTT(T/C)[¶]AACTCGTCCTCAA-ATCACACCTACGACCCCTAACTTTGGATGTTCACAAG
CaNB-5410	G, G	T, A	TAGAATCCATCAACCGTACGTAGAT(G/T)CTAAGTTTTT(G/A)T-TTTTTTATGAAGAAGAAAAAAGTCACAGTTCTAGTCAAG-ATGCCACCTAACTCAAT
CaNB-5440	G, C	A, T	GTTCGACGAGGCTCTACTTTGAGAAGTTCAGCTA(G/A)CAATC-TTTCTACAA(C/T)ATCCAGTAAGCTGGGAGGTAGTGCCTCCT-CATTGTTAATCCTGGCAGACTGA
CaNB-5480	T, T, G	C, C, A	TCGAAATCAATACTCCTCCTTCCCTAGGGGGTCAGTTGACAGGTAA-TTAAATTCCCAACCTCCAA(T/C)ACATTTTTCAGT-CAAAGGGTAAAGGGTGGGACTTTTTTTAC(T/C)GCGT-AAGGTAAATTAGTCTCCCCGCCTTAC(G/A)CTTTATCATTA-GCTAATGTCATTAACCTTTTAAGAACAACGGCGTGAG
CaNB-5500	T	G	TCATTTATGCGTGCAGTATCGTACTCTCATGCAAGACTTATATC-AGTGAAC(T/G)TATGCATTGAAGCATAACTTGTACCACTTAT-AAACTTATGTCCTGGAGCAGAGAATGTGTCATTGAGGGGTATATTT
CaNB-5720	A	G	GGAAGTCCATTTTATCTATCCTCAGAA(A/G)TGAGTACACTATTGAA-ATCCCCAAAATCAT
CaRP-5130	A	G	GACACATTTTGCAGATTCATCAACTAA(A/G)TTGCTGTTTTCTTTTT-ACCTTTTTGGGTG
CaNB-5330	C, G	T, A	AAGAAAGCCGTCACCTTCATAGATGATGGGTTGTCGATGGG-C(C/T)CCCTCACTACATAAGTCCACATTGGTAC(G/A)TTAAT-GGGGATTTGAGCGTAATCCTTGCAAATTAGC
CaRK-5470	G	A	TTTTCAATATCCAAAGAACACAACA(G/A)TTAATATTCTTTTCTTTTCTTT-CACTATCACATTGGGTTCA
CaNB-5530	C	T	GCTCCGTGTTCATCTATGTTAAACAAGATTTCTTGCAA(C/T)ATGTCATGA-TGCTGATGGATCCTTCTGTT
CaNB-5170	A	C	TATCTCCTACATCCCACTGACAACAGTTAGCTTTATATTGTCAACTTTA-TCTATC(A/C)AGACTCTGTTTTTCAGTCACTTTATT-GTCCATTAATGCACCACCCATGACAAG

[‡]*P. capsici*-resistant germplasm *C. annuum* 'CM334'; [§]*P. capsici*-susceptible germplasm *C. annuum* 'Daepoongcho'; [¶](CM334 allele/Daepoongcho allele).

TABLE 3: The genotypes of the major QTL for *P. capsici* resistance linked SNP markers.

Group	Code[†]	Marker					
		CaNB-5480	CaNB-5500	CaRP-5130	CaNB-5330	Phyto5SAR	CaNB-5530
	R1	R*	R*	R*	R*	R*	R*
	R2	H*	H*	H*	H*	H*	H*
	R3	R*	R*	R*	H*	H*	R*
	R4	H*	H*	H*	H*	H*	H*
	R5	H*	H*	H*	H*	H*	H*
	R6	H*	H*	H*	H*	H*	H*
	R7	H*	H*	H*	H*	H*	H*
	R8	H*	H*	H*	H*	H*	H*
	R9	H*	H*	H*	H*	H*	H*
	R10	H*	H*	H*	H*	H*	H*
	R11	R*	R*	H*	H*	H*	H*
	R12	R*	R*	H*	H*	H*	H*
	R13	R*	R*	H*	H*	H*	H*
	R14	R*	R*	R*	H*	H*	H*
	R15	R*	R*	R*	H*	H*	R*
	R16	R*	R*	R*	H*	H*	R*
Group 1	R17	R*	R*	R*	H*	R*	H*
	R18	R*	R*	H*	H*	H*	H*
	R19	R*	R*	R*	H*	H*	H*
	R20	R*	R*	R*	R*	R*	R*
	R21	H*	H*	H*	H*	H*	H*
	R22	H*	H*	R*	H*	H*	H*
	R23	R*	R*	R*	H*	H*	H*
	R24	R*	R*	R*	H*	R*	R*
	R25	H*	H*	R*	H*	H*	H*
	R26	R*	R*	R*	R*	R*	R*
	R27	R*	R*	H*	H*	H*	H*
	R28	H*	H*	H*	H*	H*	H*
	R29	H*	H*	H*	H*	H*	H*
	R30	R*	R*	R*	H*	H*	H*
	MR1	H*	H*	H*	H*	H*	H*
	MR2	R*	R*	H*	H*	H*	H*
	MR3	R*	R*	H	H	H	H

Table 3: Continued.

Group	Code[†]	Marker					
		CaNB-5480	CaNB-5500	CaRP-5130	CaNB-5330	Phyto5SAR	CaNB-5530
Group 2	S1	S*	S*	S*	S*	S*	S*
	S2	S*	S*	S*	S*	S*	S*
	S3	S*	S*	S*	S*	S*	S*
	S4	S*	S*	S*	S*	S*	S*
	R31	H*	H*	H*	H*	-**	H*
	R32	H*	H*	H*	R*	S**	S**
	R33	R*	R*	H*	R*	S**	S**
	R34	H*	H*	H*	S**	H*	H*
	R35	H*	H*	-**	R*	S**	H*
	R36	H*	H*	S**	H*	S**	S**
	R37	H*	H*	S**	R*	S**	S**
Group 3	R38	R*	R*	R*	R*	R*	R*
	R39	S**	S**	H*	S**	H*	S**
	R40	S**	S**	H*	H*	H*	H*
	R41	S**	S**	R*	S**	R*	R*
	R42	S**	S**	R*	S**	R*	R*
	R43	S**	S**	H*	S**	S**	S**
	R44	S**	S**	S**	R*	R*	R*
Group 4	S5	S*	S*	S*	R**	S*	S*
	S6	S*	S*	R**	S*	S*	R**
	S7	S*	S*	R**	S*	S*	R**
	S8	S*	S*	R**	S*	S*	R**
	S9	S*	S*	R**	H**	S*	S*
	S10	S*	S*	R**	H**	S*	H**
	S11	S*	H**	S*	H**	S*	S*
	S12	S*	H**	R**	H**	S*	R**
	S13	R**	R**	S*	S*	S*	S*
	S14	H**	H**	-**	S*	S*	S*

Cosegregation genotypes are indicated by *; ** indicates recombinant genotypes. R: resistant homozygote, H: resistant heterozygote, S: susceptible homozygote, -: not detected. Group 1: genotypes matched with R phenotypes in resistant and moderately resistant cultivars, Group 2: genotypes matched with S phenotypes in susceptible cultivars, Group 3: a group of resistant cultivars with low genotyping accuracy, and Group 4: a group of susceptible cultivars with low genotyping accuracy. [†]Accession name is shown on Table S1.
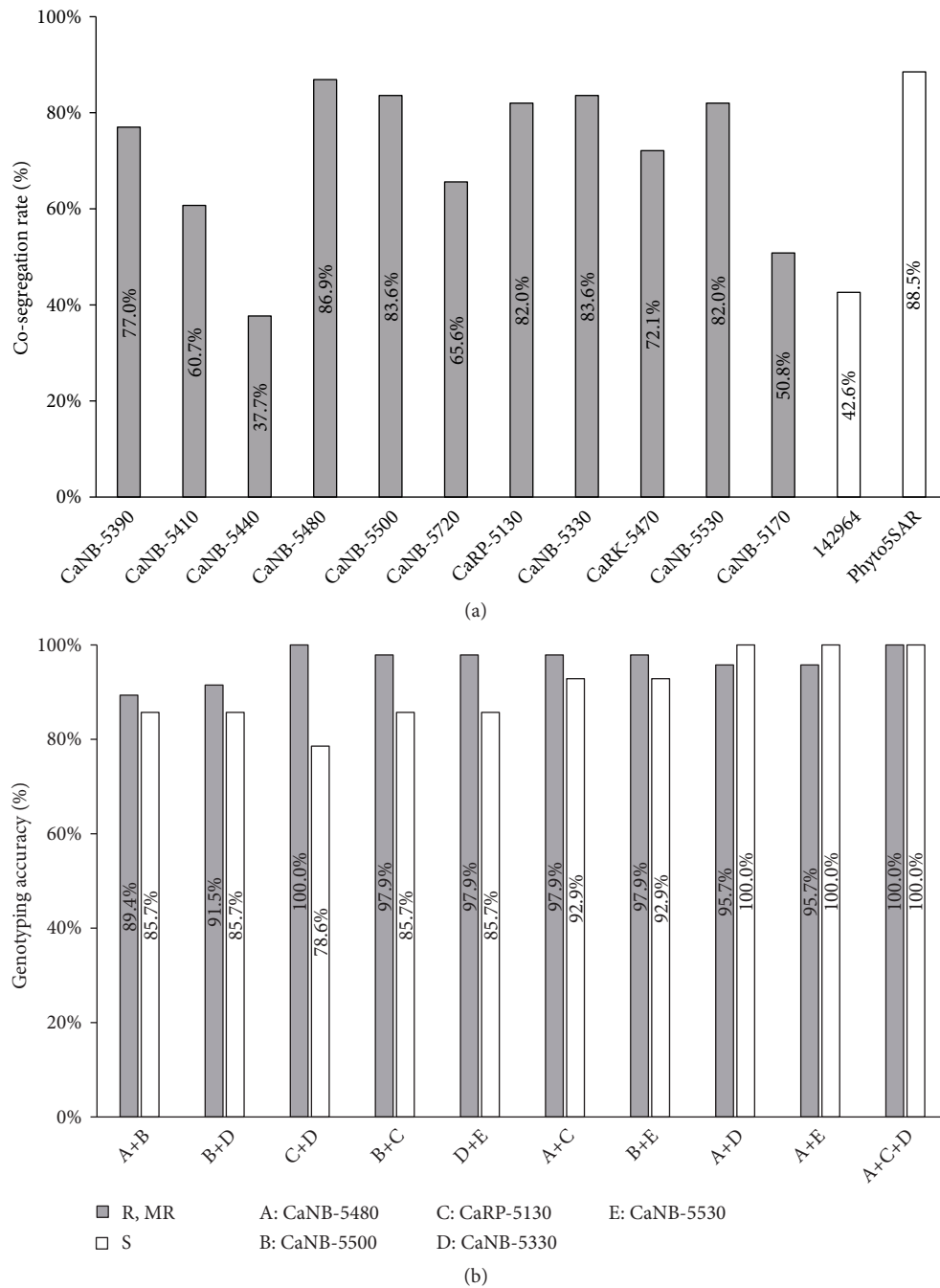
(a)

(b)

FIGURE 3: Cosegregation rate obtained by genotyping RGA-based markers in 61 cultivars. (a) The cosegregation rate of the each marker. The x-axis shows the names of the molecular markers. The y-axis shows the cosegregation rate. Newly developed markers are indicated by grey bars. White bars show previously developed markers 142964 and Phyto5SAR. (b) The genotyping accuracy when applied with combination of RGA markers at 61 cultivars. The x-axis shows combination of the markers and y-axis shows genotyping accuracy. Grey bars show genotyping accuracy of resistant and moderately resistant cultivars. Genotyping accuracy of susceptible cultivars is indicated by white bars.

rate (86.9%) except for eight cultivars (13.1%). CaNB-5480 could be applicable to genotyping of germplasms that are not amenable to genotyping (R31-33; R35-37) using Phyto5SAR. In a previous study, the combination of molecular markers increased genotyping accuracy of *Cf-9* locus in tomato cultivars [52]. In our study, by combining two or three markers, the genotyping accuracy was increased compared

to that in single RGA maker (Figure 3). Combination of three markers such as CaNB-5480, CaRP-5130, and CaNB-5330 cosegregated with the resistance and susceptible phenotypes in pepper accessions used in this study (Figure 3(b)). Genotyping results of CaNB-5480 and Phyto5SAR showed 98.4% genotyping accuracy except for one cultivar (R43) and CaNB-5130 can apply to the genotyping of the R43 cultivar.

The genotyping accuracy of combination of two or three markers ranged from 78.6% to 100% (Figure 3(b)). Taken together, combination with highly linked markers could be used efficiently for genotyping of pepper varieties for further pepper breeding.

The variation of the cosegregation rate (37.7% to 86.9%) suggested the conversion of the major QTL region between the genetic and physical maps. Kim and associates [17] constructed pseudomolecules using a high-density map with 6,281 markers derived from *C. annuum* 'Perennial' and *C. annuum* 'Dempsey'. Scaffold anchoring was conducted with genetic maps of a cross between *C. annuum* 'NuMexRNAKY' and *C. frutescens* acc. 2814-6 [53]. Neither previous study used 'CM334'-related populations. Genomic variation between 'CM334' and other pepper germplasms could represent differences in cosegregation [54]. Further reanalyses of the pepper genomic structure could reveal reanchoring or rearrangement within or among chromosomes.

To date, several markers linked to the major resistance QTL in pepper have been developed including restriction fragment length polymorphisms (RFLPs), randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLPs), and SNPs, which could lead to marker-assisted selection for breeding *P. capsici*-resistant lines [55, 56]. Wang and associates [57] also developed SSR markers that tightly linked to the resistance gene in pepper line 'PI201234'. However, the previously developed markers from the different genetic maps could be insufficient to determine whether the QTL region includes a QTL that is conserved among different progenies. *P. capsici* isolates with different virulence factors and/or inoculation concentrations also have variable disease phenotypes, which cause variation in cosegregation rates of molecular markers [43, 57]. Rehrig and associates [44] reported cosegregated *CaDMR1* with QTL *Pc5.1* as a candidate for resistance to *P. capsici* in pepper, but its function was not determined yet. Here, newly closely linked RGA-based markers of *P. capsici* resistance were developed, which can be used to genotype breeding sources and to further characterize *R* genes. Our markers may be used in combination with other markers such as CaNB-5480, CaRP-5130, and CaNB-5330 to efficiently determine the phenotypes of pepper germplasms. Such data would also be sufficient to determine the resistance gene spectrum in the QTL region on chromosome 5.

4. Conclusion

In this study, 11 novel RGA-based markers were developed that are linked to major QTL for *P. capsici* resistance. Among the markers, CaNB-5480 showed the most closely linked marker to major QTL. With combination of CaNB-5480, CaRP-5130 and CaNB-5330 provide the most accurate assessment of genotyping among 61 *Capsicum* accessions. Together, as combination with other markers, it could be more efficiently phenotyping of pepper germplasms and to further characterize resistance genes against *P. capsici*.

Supplementary Materials

Supp. Table S1: plant materials used to validate the molecular markers in the study. Supp. Table S2: the information of the reanalyzed RGAs in this study. Supp. Table S3: sequence variation (SNPs) between resistant and susceptible plants in pepper varieties. *(Supplementary Materials)*

References

[1] L. R. Howard and R. E. Wildman, "Antioxidant Vitamin and Phytochemical Content of Fresh and Processed Pepper Fruit (*Capsicum annuum*)," in *Handbook of Nutraceuticals and Functional Foods*, pp. 173–199, CRC Press, 2016.

[2] FAOSTAT, http://www.fao.org/faostat/en/#data/QC.

[3] Common Names of Plant Diseases, http://www.apsnet.org/publications/commonnames/Pages/Pepper.aspx.

[4] L. H. Leonian, "Stem and fruit blight of peppers caused by *Phytophthora capsici* sp. nov," *Phytopathology*, vol. 12, no. 9, pp. 401–408, 1922.

[5] M. K. Hausbeck and K. H. Lamour, "*Phytophthora capsici* on vegetable crops: Research progress and management challenges," *Plant Disease*, vol. 88, no. 12, pp. 1292–1303, 2004.

[6] K. H. Lamour, R. Stam, J. Jupe, and E. Huitema, "The oomycete broad-host-range pathogen *Phytophthora capsici*," *Molecular Plant Pathology*, vol. 13, no. 4, pp. 329–337, 2012.

[7] R. Bartual, E. A. Carbonell, J. I. Marsal, J. C. Tello, and T. Campos, "Gene action in the resistance of peppers (*Capsicum annuum*) to Phytophthora stem blight (*Phytophthora capsici* L.)," *Euphytica*, vol. 54, no. 2, pp. 195–200, 1991.

[8] L. M. Oelke, P. W. Bosland, and R. Steiner, "Differentiation of race specific resistance to phytophthora root rot and foliar blight in *Capsicum annuum*," *Journal of the American Society for Horticultural Science*, vol. 128, no. 2, pp. 213–218, 2003.

[9] M. K. Hausbeck and J. M. Foster, "Resistance of pepper to phytophthora crown, root, and fruit rot is affected by isolate virulence," *Plant Disease*, vol. 94, no. 1, pp. 24–30, 2010.

[10] S. J. Walker and P. W. Bosland, "Inheritance of phytophthora root rot and foliar blight resistance in pepper," *Journal of the*

American Society for Horticultural Science, vol. 124, no. 1, pp. 14–18, 1999.

[11] S.-I. Yeom, H.-K. Baek, S.-K. Oh et al., "Use of a secretion trap screen in pepper following *Phytophthora capsici* infection reveals novel functions of secreted plant proteins in modulating cell death," *Molecular Plant-Microbe Interactions*, vol. 24, no. 6, pp. 671–684, 2011.

[12] V. Lefebvre and A. Palloix, "Both epistatic and additive effects of QTLs are involved in polygenic induced resistance to disease: a case study , the interaction pepper - *Phytophthora capsici* Leonian," *Theoretical and Applied Genetics*, vol. 93, no. 4, pp. 503–511, 1996.

[13] E. A. Ogundiwin, T. F. Berke, M. Massoudi et al., "Construction of 2 intraspecific linkage maps and identification of resistance QTLs for *Phytophthora capsici* root-rot and foliar-blight diseases of pepper (*Capsicum annuum* L.)," *Genome*, vol. 48, no. 4, pp. 698–711, 2005.

[14] O. Sy, P. W. Bosland, and R. Steiner, "Inheritance of phytophthora stem blight resistance as compared to phytophthora root rot and phytophthora foliar blight resistance in *Capsicum annuum* L," *Journal of the American Society for Horticultural Science*, vol. 130, no. 1, pp. 75–78, 2005.

[15] A. Thabuis, A. Palloix, S. Pflieger, A.-M. Daubèze, C. Caranta, and V. Lefebvre, "Comparative mapping of *Phytophthora* resistance loci in pepper germplasm: evidence for conserved resistance loci across Solanaceae and for a large genetic diversity," *Theoretical and Applied Genetics*, vol. 106, no. 8, pp. 1473–1485, 2003.

[16] S. Kim, J. Park, S. Yeom et al., "New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication," *Genome Biology*, vol. 18, no. 1, Article ID 210, 2017.

[17] S. Kim, M. Park, and S.-I. Yeom, "Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species," *Nature Genetics*, vol. 46, no. 3, pp. 270–278, 2014.

[18] J. Chunthawodtiporn, T. Hill, K. Stoffel, and A. van Deynze, "Quantitative trait loci controlling fruit size and other horticultural traits in bell pepper (*Capsicum annuum*)," *The Plant Genome*, vol. 11, no. 1, Article ID 160125, 2018.

[19] S.-B. Kim, W.-H. Kang, H. N. Huy et al., "Divergent evolution of multiple virus-resistance genes from a progenitor in *Capsicum* spp," *New Phytologist*, vol. 213, no. 2, pp. 886–899, 2017.

[20] H.-A. Lee and S.-I. Yeom, "Plant NB-LRR proteins: tightly regulated sensors in a complex manner," *Briefings in Functional Genomics*, vol. 14, no. 4, pp. 233–242, 2015.

[21] M. K. Sekhwal, P. Li, I. Lam, X. Wang, S. Cloutier, and F. M. You, "Disease resistance gene analogs (RGAs) in plants," *International Journal of Molecular Sciences*, vol. 16, no. 8, pp. 19248–19290, 2015.

[22] S.-I. Yeom, E. Seo, S.-K. Oh, K. W. Kim, and D. Choi, "A common plant cell-wall protein HyPRP1 has dual roles as a positive regulator of cell death and a negative regulator of basal defense against pathogens," *The Plant Journal*, vol. 69, no. 5, pp. 755–768, 2012.

[23] E. Seo, S. Kim, S.-I. Yeom, and D. Choi, "Genome-wide comparative analyses reveal the dynamic evolution of nucleotide-binding leucine-rich repeat gene family among solanaceae plants," *Frontiers in Plant Science*, vol. 7, Article ID 1205, 2016.

[24] W.-H. Kang and S.-I. Yeom, "Genome-wide Identification, Classification, and Expression Analysis of the Receptor-Like Protein Family in Tomato," *The Plant Pathology Journal*, vol. 34, no. 5, pp. 435–444, 2018.

[25] "Center SGE: Resistance Gene *N*: Similarity to Toll and the Interleukin-1 Receptor," *Cell*, vol. 78, no. 102, pp. 1101–1115, 1994.

[26] C. Feuillet, S. Travella, N. Stein, L. Albar, A. Nublat, and B. Keller, "Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 100, no. 25, pp. 15253–15258, 2003.

[27] M. R. Grant, L. Godiard, E. Straube et al., "Structure of the Arabidopsis *RPM1* gene enabling dual specificity disease resistance," *Science*, vol. 269, no. 5225, pp. 843–846, 1995.

[28] D. Halterman, F. Zhou, F. Wei, R. P. Wise, and P. Schulze-Lefert, "The MLA6 coiled-coil, NBS-LRR protein confers *AvrMla6*-dependent resistance specificity to *Blumeria graminis* f. sp. hordei in barley and wheat," *The Plant Journal*, vol. 25, no. 3, pp. 335–348, 2001.

[29] B. C. Meyers, K. A. Shen, P. Rohani, B. S. Gaut, and R. W. Michelmore, "Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection," *The Plant Cell*, vol. 10, no. 11, pp. 1833–1846, 1998.

[30] J. M. Salmeron, G. E. D. Oldroyd, C. M. T. Rommens et al., "Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster," *Cell*, vol. 86, no. 1, pp. 123–133, 1996.

[31] S. Yoshimura, U. Yamanouchi, Y. Katayose et al., "Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inpculation," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 95, no. 4, pp. 1663–1668, 1998.

[32] N. Collins, J. Drake, M. Ayliffe et al., "Molecular characterization of the maize *Rp1-D* rust resistance haplotype and its mutants," *The Plant Cell*, vol. 11, no. 7, pp. 1365–1376, 1999.

[33] J. Hu, Z. Pang, Y. Bi et al., "Genetically diverse long-lived clonal lineages of *Phytophthora capsici* from pepper in Gansu, China," *Journal of Phytopathology*, vol. 103, no. 9, pp. 920–926, 2013.

[34] W. Lee, J. Lee, J. Han, B. Kang, and J. Yoon, "Validity Test for Molecular Markers Associated with Resistance to *Phytophthora* Root Rot in Chili Pepper (*Capsicum annuum* L.)," *Korean Journal of Horticultural Science & Technology*, vol. 30, no. 1, pp. 64–72, 2012.

[35] B. L. Candole, P. J. Conner, and P. Ji, "Evaluation of phytophthora root rot-resistant *Capsicum annuum* accessions for resistance to phytophthora foliar blight and phytophthora stem blight," *Agricultural Sciences*, vol. 3, no. 5, pp. 732–737, 2012.

[36] P. W. Bosland, "A seedling screen for *Phytophthora* root rot of pepper, *Capsicum annuum*," *Plant Disease*, vol. 75, no. 10, pp. 1048–1050, 1991.

[37] S.-J. Jo, S.-A. Shim, K. S. Jang, Y. H. Choi, J.-C. Kim, and G. J. Choi, "Resistance of chili pepper cultivars to isolates of *Phytophthora capsici*," *Korean Journal of Horticultural Science & Technology*, vol. 32, no. 1, pp. 66–76, 2014.

[38] H. Kim, J. B. Yoon, and J. Lee, "Development of Fluidigm SNP Type Genotyping Assays for Marker-assisted Breeding of Chili Pepper (*Capsicum annuum* L.)," *Korean Journal of Horticultural Science & Technology*, vol. 35, no. 4, pp. 465–479, 2017.

[39] M. G. Murray and W. F. Thompson, "Rapid isolation of high molecular weight plant DNA," *Nucleic Acids Research*, vol. 8, no. 19, pp. 4321–4326, 1980.

[40] S. Mallard, M. Cantet, A. Massire, A. Bachellez, S. Ewert, and V. Lefebvre, "A key QTL cluster is conserved among accessions

and exhibits broad-spectrum resistance to *Phytophthora capsici*: a valuable locus for pepper breeding," *Molecular Breeding*, vol. 32, no. 2, pp. 349–364, 2013.

[41] H.-J. Kim, S.-H. Nahm, H.-R. Lee et al., "BAC-derived markers converted from RFLP linked to *Phytophthora capsici* resistance in pepper (*Capsicum annuum* L.)," *Theoretical and Applied Genetics*, vol. 118, no. 1, pp. 15–27, 2008.

[42] J. Lee, S. J. Park, S. C. Hong, J.-H. Han, D. Choi, and J. B. Yoon, "QTL mapping for capsaicin and dihydrocapsaicin content in a population of *Capsicum annuum* 'NB1' × *Capsicum chinense* 'Bhut Jolokia'," *Plant Breeding*, vol. 135, no. 3, pp. 376–383, 2016.

[43] W.-Y. Liu, J.-H. Kang, H.-S. Jeong et al., "Combined use of bulked segregant analysis and microarrays reveals SNP markers pinpointing a major QTL for resistance to *Phytophthora capsici* in pepper," *Theoretical and Applied Genetics*, vol. 127, no. 11, pp. 2503–2513, 2014.

[44] W. Z. Rehrig, H. Ashrafi, T. Hill, J. Prince, and A. Van Deynze, "*CaDMR1* cosegregates with QTL Pc5.1 for resistance to *phytophthora capsici* in pepper (*Capsicum annuum*)," *The Plant Genome*, vol. 7, no. 2, 2014.

[45] Y. Minamiyama, M. Tsuro, T. Kubo, and M. Hirai, "QTL analysis for resistance to *Phytophthora capsici* in pepper using a high density SSR-based map," *Breeding Science*, vol. 57, no. 2, pp. 129–134, 2007.

[46] J. Kourelis and R. A. L. Van Der Hoorn, "Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function," *The Plant Cell*, vol. 30, no. 2, pp. 285–299, 2018.

[47] J. Du, E. Verzaux, A. Chaparro-Garcia et al., "Elicitin recognition confers enhanced resistance to *Phytophthora infestans* in potato," *Nature Plants*, vol. 1, Article ID 15034, 2015.

[48] S. Huang, E. A. G. Van Der Vossen, H. Kuang et al., "Comparative genomics enabled the isolation of the *R3a* late blight resistance gene in potato," *The Plant Journal*, vol. 42, no. 2, pp. 251–261, 2005.

[49] A. A. Lokossou, T.-H. Park, G. van Arkel et al., "Exploiting knowledge of *R/Avr* genes to rapidly clone a new LZ-NBS-LRR family of late blight resistance genes from potato linkage group IV," *Molecular plant-microbe interactions*, vol. 22, no. 6, pp. 630–641, 2009.

[50] J. Song, J. M. Bradeen, S. K. Naess et al., "Gene *RB* cloned from *Solanum bulbocastanum* confers broad spectrum resistance to potato late blight," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 100, no. 16, pp. 9128–9133, 2003.

[51] E. A. G. Van Der Vossen, J. Gros, A. Sikkema et al., "The *Rpi-blb2* gene from *Solanum bulbocastanum* is an *Mi-1* gene homolog conferring broad-spectrum late blight resistance in potato," *The Plant Journal*, vol. 44, no. 2, pp. 208–222, 2005.

[52] B. Kim, I. S. Hwang, H.-J. Lee, and C.-S. Oh, "Combination of newly developed SNP and InDel markers for genotyping the *Cf-9* locus conferring disease resistance to leaf mold disease in the tomato," *Molecular Breeding*, vol. 37, no. 5, Article ID 59, 2017.

[53] S. C. Yarnes, H. Ashrafi, S. Reyes-Chin-Wo, T. A. Hill, K. M. Stoffel, and A. Van Deynze, "Identification of QTLs for capsaicinoids, fruit quality, and plant architecture-related traits in an interspecific *Capsicum* RIL population," *Genome*, vol. 56, no. 1, pp. 61–74, 2013.

[54] B. R. Glosier, E. A. Ogundiwin, G. S. Sidhu, D. R. Sischo, and J. P. Prince, "A differential series of pepper (*Capsicum annuum*) lines delineates fourteen physiological races of *Phytophthora capsici*: Physiological races of *P. capsici* in pepper," *Euphytica*, vol. 162, no. 1, pp. 23–30, 2008.

[55] A. Thabuis, V. Lefebvre, G. Bernard et al., "Phenotypic and molecular evaluation of a recurrent selection program for a polygenic resistance to *Phytophthora capsici* in pepper," *Theoretical and Applied Genetics*, vol. 109, no. 2, pp. 342–351, 2004.

[56] A. Thabuis, A. Palloix, B. Servin et al., "Marker-assisted introgression of 4 *Phytophthora capsici* resistance QTL alleles into a bell pepper line: validation of additive and epistatic effects," *Molecular Breeding*, vol. 14, no. 1, pp. 9–20, 2004.

[57] P. Wang, L. Wang, J. Guo, W. Yang, and H. Shen, "Molecular mapping of a gene conferring resistance to *Phytophthora capsici* Leonian race 2 in pepper line PI201234 (*Capsicum annuum* L.)," *Molecular Breeding*, vol. 36, no. 6, Article ID 66, 2016.

Research Article

Molecular Cytogenetics of *Pisum sativum* L. Grown under Spaceflight-Related Stress

Olga Yu. Yurkevich (iD),[1] **Tatiana E. Samatadze,**[1] **Margarita A. Levinskikh,**[2] **Svyatoslav A. Zoshchuk,**[1] **Olga B. Signalova,**[2] **Sergei A. Surzhikov,**[1] **Vladimir N. Sychev,**[2] **Alexandra V. Amosova** (iD),[1] **and Olga V. Muravenko** (iD)[1]

[1]*Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 119991 Moscow, Russia*
[2]*Institute of Biomedical Problems, Russian Academy of Sciences, 123007 Moscow, Russia*

Correspondence should be addressed to Olga V. Muravenko; olgmur1@yandex.ru

The ontogenesis and reproduction of plants cultivated aboard a spacecraft occur inside the unique closed ecological system wherein plants are subjected to serious abiotic stresses. For the first time, a comparative molecular cytogenetic analysis of *Pisum sativum* L. (*Fabaceae*) grown on board the RS ISS during the Expedition-14 and Expedition-16 and also plants of their succeeding (F1 and F2) generations cultivated on Earth was performed in order to reveal possible structural chromosome changes in the pea genome. The karyotypes of these plants were studied by multicolour fluorescence *in situ* hybridization (FISH) with five different repeated DNA sequences (45S rDNA, 5S rDNA, PisTR-B/1, microsatellite motifs $(AG)_{12}$, and $(GAA)_9$) as probes. A chromosome aberration was revealed in one F1 plant. Significant changes in distribution of the examined repeated DNAs in karyotypes of the "space grown" pea plants as well as in F1 and F2 plants cultivated on Earth were not observed if compared with control plants. Additional oligo-$(GAA)_9$ sites were detected on chromosomes 6 and 7 in karyotypes of F1 and F2 plants. The detected changes might be related to intraspecific genomic polymorphism or plant cell adaptive responses to spaceflight-related stress factors. Our findings suggest that, despite gradual total trace contamination of the atmosphere on board the ISS associated with the extension of the space station operating life, exposure to the space environment did not induce serious chromosome reorganizations in genomes of the "space grown" pea plants and generations of these plants cultivated on Earth.

1. Introduction

The presence of growing plants aboard a spacecraft is important for creating and supporting a sustainable living environment during long-term space missions, and in the near future plant systems will become important components of any long-duration exploration scenario. At the same time, the ontogenesis and reproduction of plants occur inside the unique closed ecological system wherein plants undergo serious abiotic stress which can be induced by a number of factors including changes in gravity, radiations, vibration, aboard air composition with limited exchange of gases, humidity, nutrients, temperature, and light. They are often associated with reprogramming of gene expression and can influence plant growth, development, and yield [1–3]. Specifically, seed size reduction was observed in *Arabidopsis thaliana* (L.) Heynh, *Brassica rapa* L., and *Triticum aestivum* L. grown for full life cycles under microgravity conditions aboard the International Space Station (ISS) [4–7]. Changes in the cell wall metabolism were revealed in *A. thaliana* cultivated under microgravity conditions in space [8]. In *Hordeum vulgare* L., *A. thaliana*, and *B. rapa*, overexpression of some genes associated with stress response proteins (heat shock proteins (HSP), pathogenesis-related proteins, and antioxidant proteins) was detected under space environment [9–12]. Besides, differential organ-specific proteome responses to spaceflight environment were revealed in leaves and roots from *A. thaliana* [13]. Also, chemical contamination of the artificial atmosphere aboard a spacecraft can influence the growth and development of plants cultivated there [1, 5, 14].

The "Lada" space greenhouse installed inside the Russian Segment of the International Space Station (RS ISS) provides optimal conditions for plant growth and development [15, 16]. It was shown that plants of *Pisum sativum* L. (*Fabaceae*) cultivated in the "Lada" greenhouse during four successive expeditions (Expeditions-7–Expeditions-10) maintained their reproductive functions and viable seeds for four full life cycles [17]. Analysis of genetic polymorphism by random amplified polymorphic DNA (RAPD) did not reveal changes in "space" plants compared to the ground control [17, 18]. In karyotypes of these "space" samples, significant chromosomal rearrangements were also not detected though polymorphic organization of constitutive heterochromatin (C-band polymorphism) was observed [19]. However, the results of the 15-year monitoring of volatile organic compounds in the air onboard the ISS demonstrated that, with the extension of the space station operating life, the total chemical contamination gradually built up and diversity and toxicity of the compounds increased [20]. Over time, this process can influence the ontogenesis of the plants cultivated there and/or induce changes (e.g., chromosome aberrations) in their genomes [1, 5, 14].

In eukaryotes, heterochromatin plays a key role in epigenetic regulation of gene expression. It was shown that heterochromatin gives rise to small interfering RNAs (siRNAs) derived from transposable elements or DNA repeats [21, 22]. In plants, heterochromatic siRNAs are the most abundant class of small RNAs which play important roles in gene regulation by means of RNA-directed DNA methylation [22–25]. Environmental stress factors can induce structural changes in heterochromatin [22, 26]. C-heterochromatin comprises different repeated DNA sequences including highly repeated (satellite) DNA, transposable elements and also microsatellites, or simple sequence repeats (SSRs) which consist of tandem duplications of 1–6 bp motifs [27, 28]. Microsatellites play diverse functional roles in eukaryotic genome (e.g., modulation of gene expression, regulation of chromatin organization, DNA metabolic processes, and RNA structure) [29]. Besides, SSRs are effective genetic markers due to their common length polymorphism, and they are widely used in genetic studies [30]. Microsatellites in coding sequences can be directly linked to gene function, and mutations in microsatellites may induce the functional genomic changes providing a basis for quick adaptations to environmental changes [31–33].

The rRNA genes can serve as excellent markers in phylogenetic and cytogenetic studies of plants. These genes are organized into two distinct families (i.e., 45S and 5S rDNA) that occur as tandem repeat arrays at specific chromosomal regions. Due to high copy number, detection of rDNAs is highly reproducible and provides valuable information concerning chromosomal evolution [34]. In plant genomes, the copy numbers and chromosomal distribution of rDNAs can vary rapidly even within intraspecific taxa and can therefore provide chromosomal landmarks for genome plasticity [34–36].

Thus, DNA repeated sequences play a significant role in plant genome adaptation to stress factors, and therefore it is important to investigate the karyotype polymorphism and chromosomal distribution of these DNA fractions in plants cultivated aboard a spacecraft. In the present study, molecular cytogenetic characterization of *P. sativum* plants grown on board the RS ISS during Expedition-14 and Expedition-16 (ISS-14 and ISS-16) and also plants of the succeeding generations cultivated on Earth was performed in order to reveal possible chromosome changes in the *P. sativum* genome. The karyotypes of these plants were studied by multicolour fluorescence *in situ* hybridization (FISH) with five different repeated DNA sequences (45S rDNA, 5S rDNA, PisTR-B/1, microsatellite motifs $(AG)_{12}$, and $(GAA)_9$) as probes.

2. Materials and Methods

2.1. Plant Material. Seeds of line 131 of *P. sativum* (2n=14) (generation F0) were obtained from the collection of Department of Biology of M.V. Lomonosov Moscow State University [18].

During the ISS Expedition-14 and Expedition-16, these seeds (generations F0-14 and F0-16, correspondingly) were grown in the Lada space greenhouse installed inside the RS ISS [17]. The seeds were germinated after irrigation with water, and the seedlings were cultivated in Lada at 23 ± 1°C and 40 to 50% humidity under 24 h lighting. Fluorescent bulbs of the lighting unit generate photosynthetically active radiation with the flux at the root module surface of $250\,\mu\mathrm{mol\,m^{-2}\,s^{-1}}$ minimum at a distance of 150 mm away from the light emitting surface. The moisture supply to the substrate was controlled at a constant, optimal level by the Lada system using *in situ* moisture sensors. The substrate consisted of the porous ceramic soil conditioner Turface (Profile Products, Buffalo Grove, IL) fertilized with Osmocote (14N-14P-14K) (Scotts Professional, Geldermalsen, Netherlands).

During the ISS-14, the experiment was being carried out for 78 days (11/01/2007–13/04/2007) by a flight engineer M. Tyurin. During the ISS-16, the experiment was being carried out for 79 days (10/01/2008–13/04/2008) by a flight engineer Yu. Malenchenko. After completion of each experiment, the seedpods with the "space grown" seeds (F1 generations (F1-14 and F1-16)) were harvested and transported back to Earth for further analyses.

The "space grown" seeds were stored until use under aseptic conditions at 3-4°C and a relative humidity of 13-14 % which enabled successful long-term seed storage [37, 38].

Then "space grown" seeds were germinated and used for chromosome spread preparation and also further postflight planting (F2 generations (F2-14 and F2-16)). The seeds of line 131 cultivated only on Earth were used as a control. Ground control and postflight cultivation were carried out using the same substrate and similar greenhouse and at the same conditions.

2.2. Chromosome Slide Preparation. For FISH, the modified technique of chromosome spread preparation from pea root tips was applied [39]. The seeds were germinated in Petri

dishes on the moist filter paper at room temperature. Root tips (of 0.5 cm) were excised and treated overnight (16-20 h) in ice-cold water. After the pretreatment, the root tips were fixed in ethanol:acetic acid (3:1) for 3–24 h at room temperature. Before squashing, the roots were transferred into 1% acetocarmine solution in 45% acetic acid for 15 min. The cover slips were removed after freezing in liquid nitrogen. The slides were dehydrated in 96% ethanol and then air-dried.

2.3. DNA Probe Preparation. The following probes were used for FISH:

(1) pTa71: a 9-kb-long sequence of common wheat encoding 18S, 5.8S, and 26S rRNA genes including spacers [40]. This DNA probe was labelled directly with SpectrumAqua and SpectrumRed fluorochromes (Abbott Molecular, Wiesbaden, Germany) by nick translation according to manufacturers' protocols.

(2) pTa794: a 420-bp-long sequence of wheat containing the 5S rRNA gene and intergenic spacer [41]. This DNA probe was labelled directly with SpectrumRed fluorochrome (Abbott Molecular, Wiesbaden, Germany) by nick translation according to manufacturers' protocols.

(3) The oligo-$(GAA)_9$ probe labelled with fluorescein-12-dUTP (Roche Diagnostics, Mannheim, Germany) and oligo-$(AG)_{12}$ probe labelled with Cy3-dUTP (DNA Synthesis, Moscow, RF). These probes were synthesized using a synthesizer ABI 394 (Applied BioSystems, Redwood City, USA) in the laboratory of biological microchips of Engelhardt Institute of Molecular Biology of RAS, Moscow, RF.

(4) The PisTR-B/1 repeat sequence (GenBank number AF300830.1) (Invitrogen, California, USA) used for identification of *P. sativum* chromosomes [42]. This DNA probe was labelled directly with fluorochromes Platinum*Bright* 415 and Platinum*Bright* 495 by Nucleic Acids Labeling Kits (Kreatech Diagnostics, Amsterdam, Netherlands) according to manufacturers' protocols.

2.4. FISH Procedure. Before FISH procedure, chromosome slides were pretreated with 1 mg/ml of RNase A (Roche Diagnostics, Mannheim, Germany) in 2xSSC at 37°C for 1 h and then washed three times for 10 min in 2xSSC. The slides were dehydrated in a series of 70%, 85%, and 96% ethanol solutions and then air-dried. The hybridisation mixture (22 μl) containing 40 ng of each labelled probe was added to each slide. Coverslips were placed on the slides and sealed with rubber cement. Slides with DNA probes were codenatured at 74°C for 5 min, placed in a moisture chamber, and hybridised overnight at 37°C. After removing the coverslips, the slides were washed with 0.1xSSC at 44°C for 8 min and with 2xSSC at 44°C for 8 min with the final 5 min wash in 2xSSC at room temperature. For oligo-$(GAA)_9$ probe labelled with fluorescein-12-dUTP, fluorescent signal amplification using FITC-Alexa 488 antibodies (VectorLabs, Youngstown, USA) was performed. After incubation for 60 min at 37°C with the detection mixture, the slides were washed two times with 2xSSC for 5 min and once in 1xPBS for 5 min each at room temperature. The slides were dehydrated and air-dried in

the dark. After FISH procedure, the slides were stained with 0.125 μg/ml DAPI (Serva, Heidelberg, Germany) dissolved in Citifluor anti-fade solution (UKC Chem. Lab., Canterbury, UK).

2.5. Chromosome Analysis. For identification of pea chromosomes, we used the molecular cytogenetic marker PisTR-B/1, developed earlier for cytogenetic classification of *P. sativum* in which the chromosome numbering was brought in line with the genetic (linkage group) identification [43, 44]. For investigation of possible chromosomal polymorphism, the oligonucleotide $(AG)_{12}$ and $(GAA)_9$ probes were used for F1 and F2 plants (ISS-14 and 16). At least 8-10 individual plants from each specimen (control, F1, and F2 plants (ISS-14 and ISS-16)) were used for the karyotype analyses. At least 15 metaphase plates from each individual were analyzed. The slides were examined using an Olympus BX-61 epifluorescence microscope (Olympus, Tokyo, Japan). Images were captured with monochrome charge-coupled device camera (Cool Snap, Roper Scientific, Inc., Sarasota, FL, USA). Then they were processed with Adobe Photoshop 10.0 software (Adobe, Birmingham, USA).

3. Results

3.1. Plant Vegetation Period and External Features of the Plants. The experiment carried out during the ISS-14 showed that the plant vegetation period was about 80 days which was 10-12 days longer than it had been observed in the control ground experiments. This period of the seed-to-seed cycle was found to increase due to slower initial vegetation stages. The experiment carried out during ISS-16 showed that the plant vegetation period was about 65-70 days which was roughly comparable with the results of the control ground experiments. External features of the plants did not differ essentially from the control specimens (Figure 1).

3.2. Chromosomal Markers Revealed by FISH. For the first time, a comparative molecular cytogenetic analysis of pea plants grown on board the RS ISS during the Expedition-14 and Expedition-16 and also their succeeding (F1 and F2) generations cultivated on Earth was performed using multicolour FISH with five different repetitive DNAs (45S rDNA, 5S rDNA, the PisTR-B/1 repeat, microsatellite motifs $(AG)_{12}$, and $(GAA)_9$).

In karyotypes of all studied *P. sativum* specimens, FISH procedure revealed 45S rDNA sites in the distal regions of the long arms of the satellite chromosome pairs 4 and 7. These chromosomes pairs can easily be distinguished as a larger satellite and a larger site of $(GAA)_9$ was revealed on chromosome 7 than on chromosome 4 (Figures 2 and 3).

Hybridization sites of 5S rDNA were detected in three chromosome pairs: in the subtelomere region of the short arm of chromosome 1, in the subtelomere region of the short arm of chromosome 3 (colocalized with a large site of the oligo-$(AG)_{12}$ probe), and also in the median region of the short arm of chromosome 5 (Figures 2 and 3).
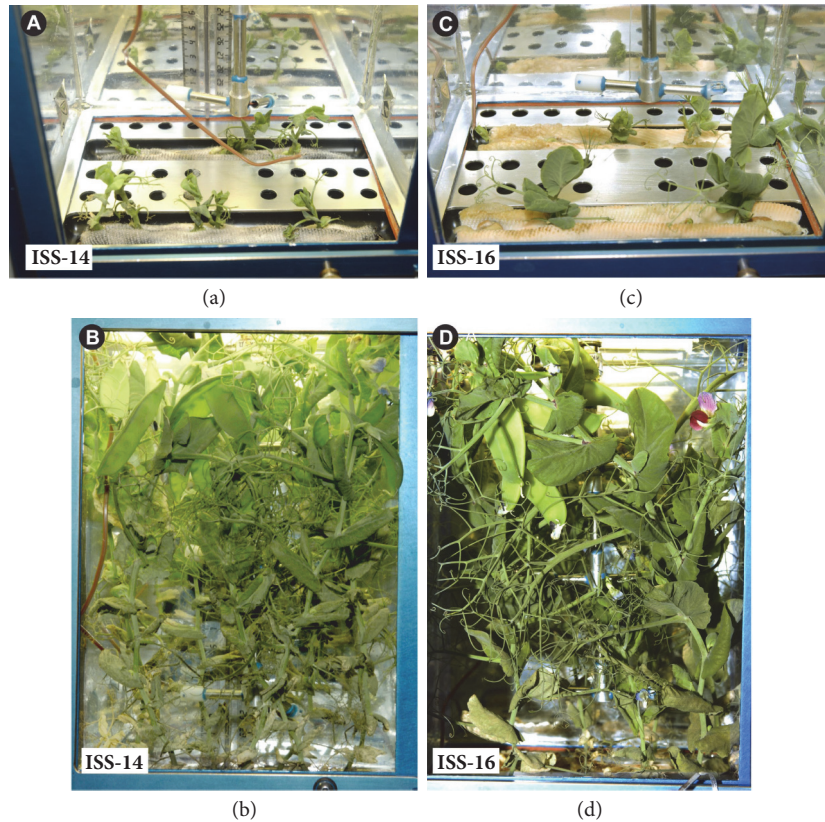
(a)

(c)

(b)

(d)

FIGURE 1: *P. sativum* **plants in the Lada greenhouse aboard the RS ISS.** Initial vegetation stages of pea plants were longer during the ISS-14 (a) compared to the ISS-16 (c). Adult plants came into flower and formed seedpods during both the ISS-14 (b) and ISS-16 (d).

The PisTR-B/1 repeat was mostly localized in subtelomeric regions of chromosomes (except for chromosome 5) and also in pericentromeric regions of chromosomes 3, 4, and 5.

In karyotypes of all studied *P. sativum* specimens, hybridization sites of $(AG)_{12}$ were clustered in the intercalary regions of both arms of all chromosome pairs. Besides, a polymorphic $(AG)_{12}$ site was revealed in the distal region of the long arm of chromosome 3 (Figure 3).

In nearly 50% of the studied F1-14 plants, multiple hybridization signals of the oligo-$(GAA)_9$ probe were detected in the short arm of satellite chromosome 7 while only one $(GAA)_9$ site was observed in karyotypes of F2-14, F1-16, and F2-16 plants and also control specimens (Figure 4). In some F1 and F2 plants (ISS-14 and 16), $(GAA)_9$ signals were also detected on chromosome 6 (Figures 3 and 4) which were absent in the control specimens.

Based on distribution patterns of the examined molecular cytogenetic markers, all chromosomes in the studied *P. sativum* plants were precisely identified (Figure 2). The karyograms of the control and experimental plant specimens (F1 and F2 generations) as well as the generalized idiogram showing the chromosomal distribution of the examined markers were constructed (Figure 3).

In the studied *P. sativum* plants, significant chromosomal aberrations were not revealed with the exception of one F1-14 plant bearing a reciprocal translocation between chromosomes 2 and 3 (Figure 2).

4. Discussion

P. sativum is a valuable nutritive crop having considerable natural genetic diversity though the pea evolution from its wild ancestor to the cultivated species involved selection for thousands of years [45]. *P. sativum* is a genetically well-characterized species which has long been used as a classic model in genetic studies [45].

The pea karyotype has a relatively low chromosome number (2n=14) and the chromosome lengths range from of 3 to 6 μm [46, 47]. The genome of *P. sativum* is rather large (1C = 4.45 Gbp) [47], and repetitive sequences make up 50-60 % of the total DNA [45, 48]. Repetitive DNA is an important component of the plant genome that is involved in genome reorganization during evolution and adaptation [34, 49]. It was shown that environmental stress factors could be responsible for quantitative changes in plant DNA (particularly, changes in satellite DNAs) [35]. Dynamic changes in the heterochromatin structure which occur during plant adaptation to a new environment can result in amplification of transposable elements as well as recombination events between repeats within satellite DNAs [50]. For instance, in *Arabidopsis* plants subjected to long-term heat stress, the increased level of transcription of satellite DNAs was observed [51]. These processes can induce structural genomic changes and formation of new structural domains and regulatory units [50]. Spaceflight-related stress
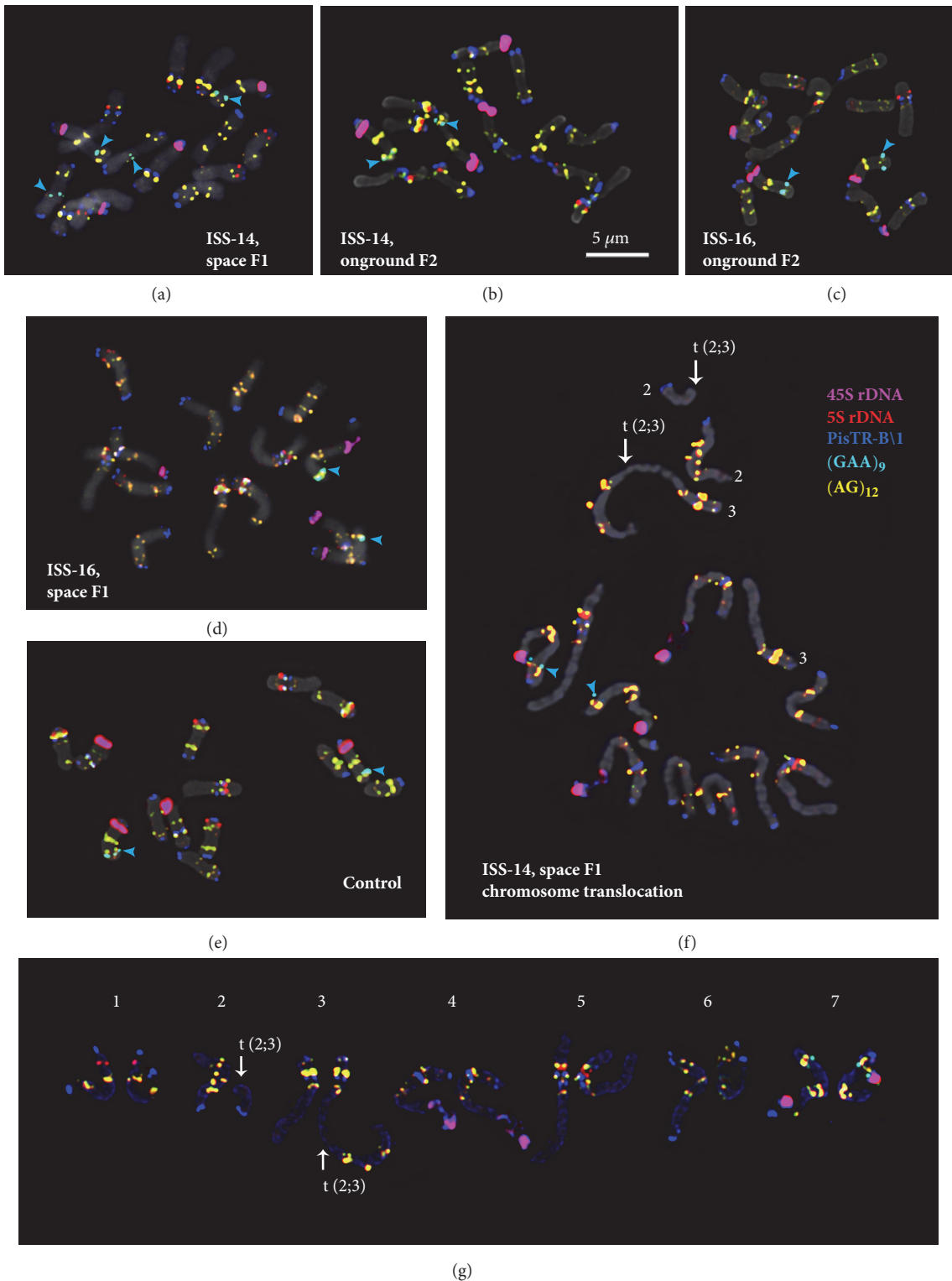
FIGURE 2: **FISH-mapping of five different DNA repeated sequences on chromosomes of experimental and control *P. sativum* samples.** Metaphase spreads of the F1-14 plant (a), F2-14 plant (b), F2-16 plant (c), F1-16 plant (d), control plant (e), F1-14 plant bearing translocation t(2;3) (f), and its karyogram (g) after FISH with 45S rDNA, 5S rDNA, PisTR-B/1, $(AG)_{12}$, and $(GAA)_9$. Arrows point to the chromosomal translocation. Heads of arrows point to $(GAA)_9$ sites on chromosomes 6 and 7. The correspondent probes and their pseudocolours are specified on the right. Bar: 5 μm.
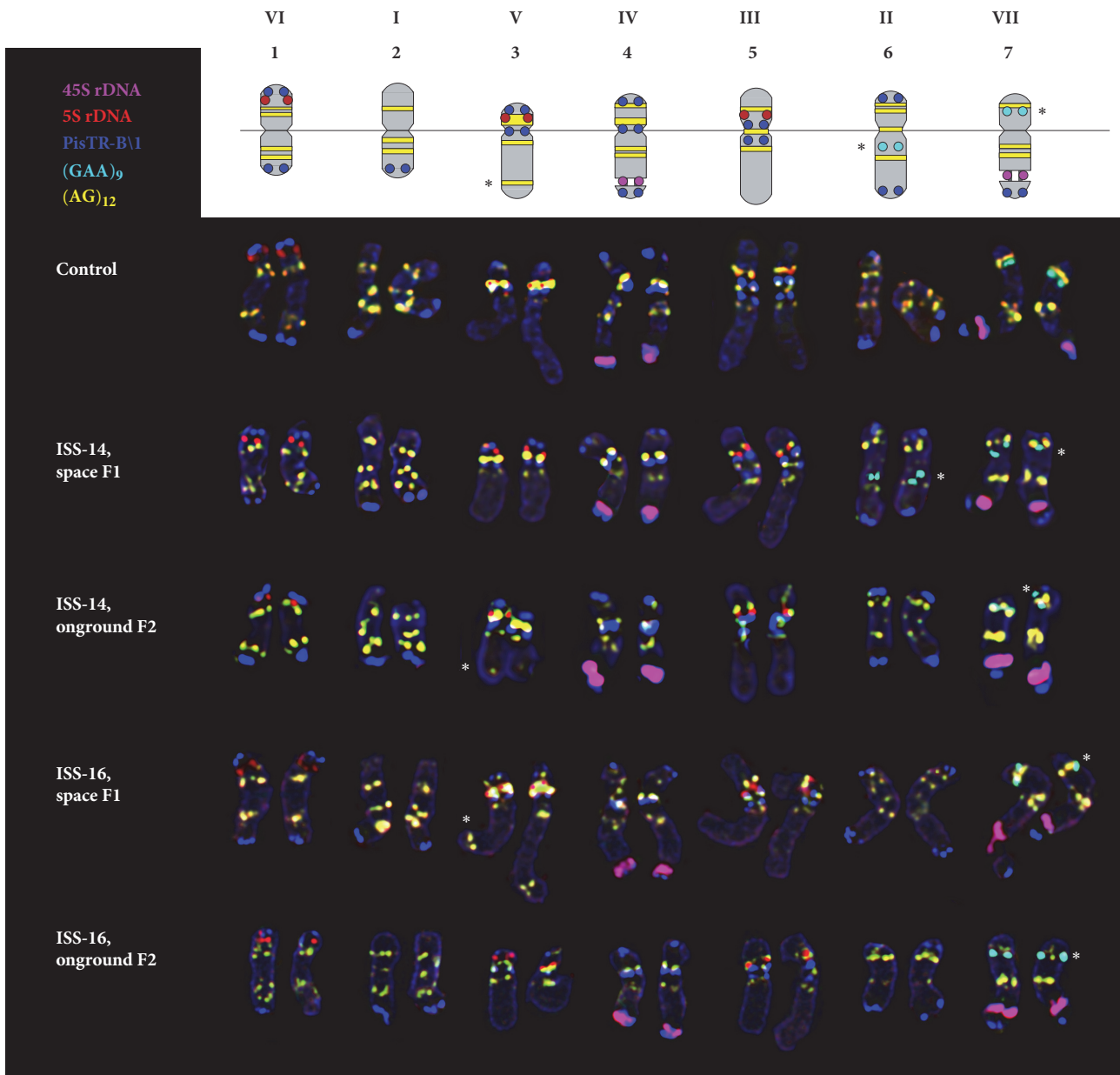
FIGURE 3: **Generalized idiogram of *P. sativum* chromosomes showing the chromosomal distribution of the examined markers and karyograms of the control and experimental samples**. Chromosome numbers and linkage groups are denoted with Arabic and Roman numerals, correspondingly. Asterisks indicate polymorphic $(GAA)_9$ and $(AG)_{12}$ sites. The correspondent probes and their pseudocolours are specified on the left.

factors can induce different genetic disorders in plants. For example, it was shown that space radiation resulted in various chromosomal aberrations and mitotic abnormalities in plants [52, 53]. In this study, the presence of a chromosomal aberration in one F1-14 plant can be related to peculiarities of seed formation occurring on the mother plant under space environment which were described earlier [4, 6, 7]. Alternatively, it might be a result of long-term seed storage that could induce chromosomal aberrations [37, 54].

It was reported earlier that number of mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide repeat motifs in genomes of different plant species could be species-specific [55]. A number of microsatellites (AAT)n, (AT)n, (GAA)n, and (AG)n were found to be frequent in the pea genome [48]. In the present study, the use of a combination of two microsatellites $((GAA)_9$ and $(AG)_{12})$ with a pea specific PisTR-B as well as 5S and 45S rDNA probes allowed us to identify all the pea chromosomes as well as perform the chromosome comparison of the plants grown from the "space grown" seeds with the plants of ground control. In the *P. sativum* plants examined here, chromosomal distribution of 45S rDNA, 5S rDNA, and PisTR-B/1 mainly corresponded to the patterns

(a)

(b)

(c)

(d)
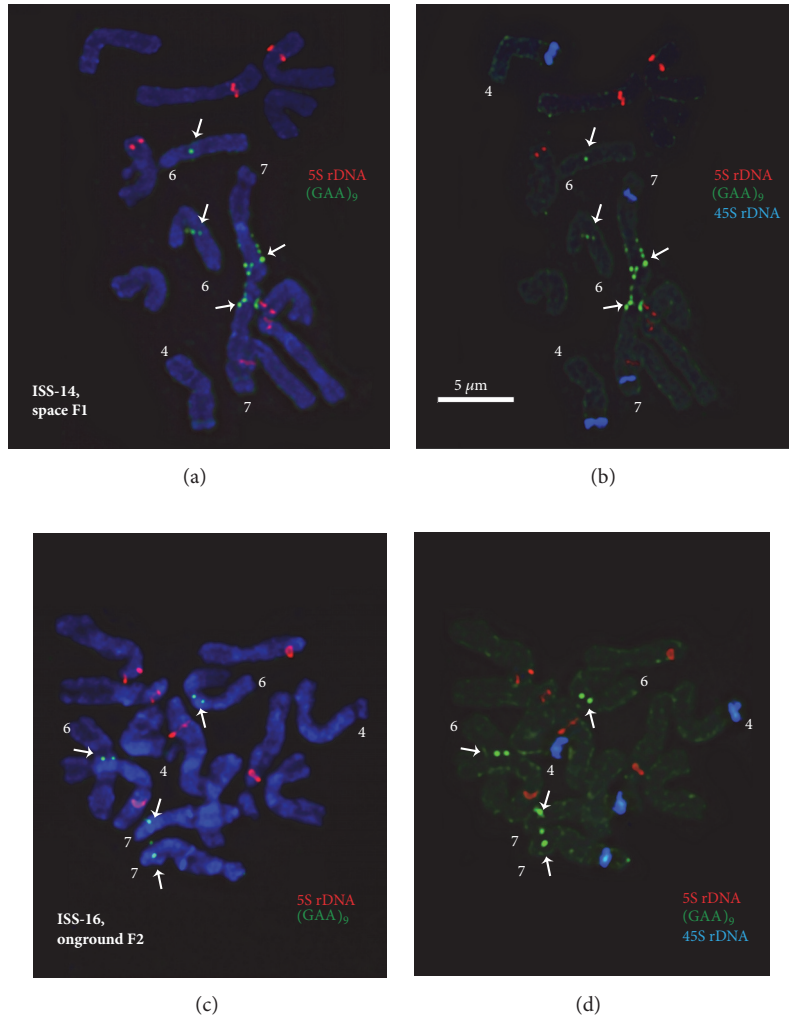
FIGURE 4: **FISH-based localization of 45S rDNA, 5S rDNA, and (GAA)$_9$ on the metaphase spreads of F1 (ISS-14) and F2 (ISS-16)** *P. sativum* **plants**. Metaphase spread of a F1 pea plant (ISS-14) after FISH with 5S rDNA, (GAA)$_9$ (a) and 45S rDNA, 5S rDNA, (GAA)$_9$ (b). Metaphase spread of a F2 pea plant (ISS-16) after FISH with 5S rDNA, (GAA)$_9$ (c) and 45S rDNA, 5S rDNA, (GAA)$_9$ (d). Arrows point to the polymorphic (GAA)$_9$ sites on chromosomes 6 and 7. The correspondent probes and their pseudocolours are specified on the right. Bar: 5 μm.

reported earlier [42, 45, 56] with the exception of the only report of four chromosome pairs bearing 5S rDNA sites in the karyotype of one *P. sativum* variety [57] which is probably due to the intraspecific variability and/or particularity of that variety. Precise chromosome identification allowed us to specify the localization of the (GAA)$_9$ and (AG)$_{12}$ sites on *P. sativum* chromosomes.

In all pea specimens studied here (including the ground control), we observed a low level of polymorphism in chromosome distribution of (AG)$_{12}$ and PisTR-B/1. However, we revealed more sites of the oligo-(AG)$_{12}$ probe in karyotypes of the studied of *P. sativum* than it was reported earlier [57]. This might indicate the presence of the intraspecific variability in chromosome distribution of the (AG)$_n$ motif in *P. sativum*. Moreover, in karyotypes of F1-14 plants, we observed multiple hybridization signals of (GAA)$_9$ on satellite chromosome 7 which were absent in

the control plants. Besides, in karyotypes of some F1 and F2 plants (ISS-14 and ISS-16) additional (GAA)$_9$ sites were also detected on chromosome 6. Considering the fact that intraspecific polymorphism in chromosomal distribution of AG and GAA motifs was described earlier for other plants (e.g., *Hordeum marinum* and *Brassica rapa* ssp. *chinensis*) [58, 59], the detected changes might also be an example of intraspecific genomic polymorphism. Alternatively, they can be related to plant cell adaptive responses to spaceflight-related stress factors. Microsatellites are shown to be involved in the processes of fast adaptation of plant populations to environmental changes and also phenotypic plasticity within and between generations, and gene-associated tandem repeats provide diverse variations promoting rapid development of new forms [31]. Some microsatellites act as cis-regulatory elements which can be recognized by transcription factors. For instance, it was shown that (GA)n and (GAA)n

repeats were included in promoters of several plant genes and participated in regulation of processes of transcription and translation [60–63].

Plant adaptation to spaceflight conditions is known to correlate with activation of genetic and epigenetic antistress mechanisms. Large-scale analyses of transcriptome in several plant species (*B. rapa*, *H. vulgare*) showed that spaceflight-related stress factors led to oxidative stress in plants and caused an increase in expression of the genes related to abiotic and biotic stresses and also in production and activities of antioxidant enzymes [9, 10, 12]. Operating experience of hermetically sealed objects showed that artificial atmosphere is known to be a multicomponent mixture involving harmful micro impurities related to various structural classes [20, 64], and an increased level of ethylene is considered to be a major cause of oxidative stress in plants [1, 5, 14]. For example, phytotoxic influence of ethylene (1.1-2.0 mg/m^3), which was early detected in the air aboard the Space Station Mir, resulted in changes in the productivity and morphometric characteristics of dwarf wheat cultivated there [5]. Spaceflight-related stress factors might induce plant cell adaptive responses manifested at the genomic level as changes in a copy number of microsatellite motifs. In support of this view, changes in plant vegetation period were observed during the ISS-14. Unfortunately, the other karyotypic studies of plants grown from the "space" seeds are next to none. This issue needs further investigations as the "Lada" space greenhouse is too small in size to produce a great number of seeds.

5. Conclusions

Significant changes in distribution of the examined repeated DNAs in karyotypes of the "space grown" *P. sativum* plants as well as in F1 and F2 pea plants cultivated on Earth were not observed if compared with control plants. The revealed changes in chromosomal distribution of the oligo-(GAA)$_9$ probe in karyotypes of F1 and F2 plants might be an example of intraspecific genomic polymorphism or related to plant cell adaptive responses to spaceflight-related stress factors.

Our findings suggest that, despite gradual total trace contamination of the atmosphere onboard the ISS associated with the extension of the space station operating life, exposure to the long-term spaceflight-related stress factors did not induce serious chromosome reorganizations in genomes of the "space grown" pea plants and generations of these plants cultivated on Earth.

Data Availability

All data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Authors' Contributions

The present study was conceived and designed by Olga Yu. Yurkevich, Olga V. Muravenko, Tatiana E. Samatadze, Alexandra V. Amosova, Margarita A. Levinskikh, and Vladimir N. Sychev. Margarita A. Levinskikh provided plant materials. Oligonucleotide probes were synthesized by Sergei A. Surzhikov. Olga Yu. Yurkevich, Tatiana E. Samatadze, and Svyatoslav A. Zoshchuk carried out the FISH experiments. Olga Yu. Yurkevich, Olga V. Muravenko, Tatiana E. Samatadze, Alexandra V. Amosova, Margarita A. Levinskikh, and Olga B. Signalova analyzed the data. Olga Yu. Yurkevich, Olga V. Muravenko, Tatiana E. Samatadze, Margarita A. Levinskikh, Vladimir N. Sychev, Alexandra V. Amosova, and Sergei A. Surzhikov participated in preparing and writing the manuscript. Olga Yu. Yurkevich, Olga V. Muravenko, Alexandra V. Amosova, Margarita A. Levinskikh, Vladimir N. Sychev, Svyatoslav A. Zoshchuk, Tatiana E. Samatadze, and Olga B. Signalova performed the analysis with constructive discussions. All authors contributed to revising the manuscript. All authors have read and approved the final manuscript. Olga Yu. Yurkevich and Tatiana E. Samatadze contributed equally to this work.

Acknowledgments

References

[1] M. E. Musgrave and A. Kuang, "Plant reproductive development during spaceflight," *Advances in Space Biology and Medicine*, vol. 9, pp. 1–23, 2003.

[2] V. De Micco, S. De Pascale, R. Paradiso, and G. Aronne, "Microgravity effects on different stages of higher plant life cycle and completion of the seed-to-seed cycle," *The Journal of Plant Biology*, vol. 16, no. 1, pp. 31–38, 2014.

[3] H. Q. Zheng, F. Han, and J. Le, "Higher plants in space: microgravity perception, response, and adaptation," *Microgravity Science and Technology*, vol. 27, no. 6, pp. 377–386, 2015.

[4] M. E. Musgrave, A. Kuang, Y. Xiao et al., "Gravity independence of seed-to-seed cycling in *Brassica rapa*," *Planta*, vol. 210, no. 3, pp. 400–406, 2000.

[5] M. A. Levinskikh, V. N. Sychev, T. A. Derendyaeva et al., "Analysis of the spaceflight effects on growth and development of Super Dwarf wheat grown on the Space Station Mir," *Journal of Plant Physiology*, vol. 156, no. 4, pp. 522–529, 2000.

[6] M. A. Levinskikh, V. N. Sychev, O. B. Signalova et al., "Some characteristics of plant seeds formed in microgravity," *Aerospace and Environmental Medicine*, vol. 36, pp. 32–35, 2002.

[7] B. M. Link, J. S. Busse, and B. Stankovic, "Seed-to-seed-to-seed growth and development of Arabidopsis in microgravity," *Astrobiology*, vol. 14, no. 10, pp. 866–875, 2014.

[8] T. Hoson, K. Soga, K. Wakabayashi et al., "Growth stimulation in inflorescences of an Arabidopsis tubulin mutant under microgravity conditions in space," *The Journal of Plant Biology*, vol. 16, no. 1, pp. 91–96, 2014.

[9] E. I. Shagimardanova, O. A. Gusev, V. N. Sychev et al., "Expression of stress response genes in barley Hordeum vulgare in a spaceflight environment," *Journal of Molecular Biology*, vol. 44, no. 5, pp. 734–740, 2010.

[10] A. L. Paul, A. K. Zupanska, D. T. Ostrow et al., "Spaceflight transcriptomes: Unique responses to a novel environment," *Astrobiology*, vol. 12, no. 1, pp. 40–56, 2012.

[11] A. K. Zupanska, F. C. Denison, R. J. Ferl, and A.-L. Paul, "Spaceflight engages heat shock protein and other molecular chaperone genes in tissue culture cells of *Arabidopsis Thaliana*," *American Journal of Botany*, vol. 100, no. 1, pp. 235–248, 2013.

[12] M. Sugimoto, Y. Oono, O. Gusev et al., "Genome-wide expression analysis of reactive oxygen species gene network in Mizuna plants grown in long-term spaceflight," *BMC Plant Biology*, vol. 14, no. 1, article 4, 2014.

[13] R. J. Ferl, J. Koh, F. Denison, and A.-L. Paul, "Spaceflight induces specific alterations in the proteomes of arabidopsis," *Astrobiology*, vol. 15, no. 1, pp. 32–56, 2015.

[14] J. G. Carman, P. Hole, F. B. Salisbury, and G. E. Bingham, "Developmental, nutritional and hormonal anomalies of weightlessness-grown wheat," *Life Sciences in Space Research*, vol. 6, pp. 59–68, 2015.

[15] V. N. Sychev, M. A. Levinskikh, and I. G. Podolsky, "Biological component of life support systems for a crew in long-duration space expeditions," *Acta Astronautica*, vol. 63, no. 7-10, pp. 1119–1125, 2008.

[16] M. A. Levinskikh, V. N. Sychev, I. G. Podolsky et al., "Modern problems of creation of the autotrophic link of perspective life-support systems of space crews," *Aviakosmicheskaia i Ekologicheskaia Meditsina*, vol. 47, pp. 91-92, 2013.

[17] V. N. Sychev, M. A. Levinskikh, S. A. Gostimsky, G. E. Bingham, and I. G. Podolsky, "Spaceflight effects on consecutive generations of peas grown onboard the Russian segment of the International Space Station," *Acta Astronautica*, vol. 60, no. 4-7, pp. 426–432, 2007.

[18] S. A. Gostimsky, M. A. Levinskikh, V. N. Sychev et al., "The study of the genetic effects in generation of pea plants cultivated during the whole cycle of ontogenesis on the board of RS ISS," *Russian Journal of Genetics*, vol. 43, no. 8, pp. 869–874, 2007.

[19] A. V. Zelenin, T. E. Samatadze, M. A. Levinskikh et al., "The cultivation of peas on the board the International Space Station did not induce changes in their karyotypes," *Pisum Genetics*, vol. 41, pp. 46-47, 2009.

[20] A. A. Pakhomova, D. S. Ozerov, D. S. Tsarkov et al., "Description of the chemical make-up of air on the International space station," *Aviakosmicheskaia i Ekologicheskaia Meditsina*, vol. 51, no. 1, pp. 58–64, 2017.

[21] H. Zhang, Z. Tao, H. Hong et al., "Transposon-derived small RNA is responsible for modified function of WRKY45 locus," *Nature Plants*, vol. 2, pp. 16016–16023, 2016.

[22] J. Wang, S. T. Jia, and S. Jia, "New insights into the regulation of heterochromatin," *Trends in Genetics*, vol. 32, no. 5, pp. 284–294, 2016.

[23] P. Neumann, H. Yan, and J. Jiang, "The centromeric retrotransposons of rice are transcribed and differentially processed by RNA interference," *Genetics*, vol. 176, no. 2, pp. 749–761, 2007.

[24] I. Djupedal and K. Ekwall, "Epigenetics: heterochromatin meets RNAi," *Cell Research*, vol. 19, no. 3, pp. 282–295, 2009.

[25] Z. Lippman and R. Martienssen, "The role of RNA interference in heterochromatic silencing," *Nature*, vol. 431, no. 7006, pp. 364–370, 2004.

[26] A. V. Probst and O. M. Scheid, "Stress-induced structural changes in plant chromatin," *Current Opinion in Plant Biology*, vol. 27, pp. 8–16, 2015.

[27] J. S. P. Heslop-Harrison and T. Schwarzacher, "Organisation of the plant genome in chromosomes," *The Plant Journal*, vol. 66, no. 1, pp. 18–33, 2011.

[28] B. J. Merritt, T. M. Culley, A. Avanesyan, R. Stokes, and J. Brzyski, "An empirical review: characteristics of plant microsatellite markers that confer higher levels of genetic variation," *Applications in Plant Sciences*, vol. 3, no. 8, p. 1500025, 2015.

[29] A. T. M. Bagshaw, "Functional mechanisms of microsatellite DNA in eukaryotic genomes," *Genome Biology and Evolution*, vol. 9, no. 9, pp. 2428–2443, 2017.

[30] R. G. Hodel, M. C. Segovia-Salcedo, J. B. Landis et al., "The report of my death was an exaggeration: a review for researchers using microsatellites in the 21st century," *Applications in Plant Sciences*, vol. 4, no. 6, p. 1600025, 2016.

[31] Y. C. Li, A. B. Korol, T. Fahima, and E. Nevo, "Microsatellites within genes: structure, function, and evolution," *Molecular Biology and Evolution*, vol. 21, no. 6, pp. 991–1007, 2004.

[32] G.-F. Richard, A. Kerrest, and B. Dujon, "Comparative genomics and molecular dynamics of DNA repeats in eukaryotes," *Microbiology and Molecular Biology Reviews*, vol. 72, no. 4, pp. 686–727, 2008.

[33] J. Shi, S. Huang, D. Fu et al., "Evolutionary dynamics of microsatellite distribution in plants: insight from the comparison of sequenced brassica, arabidopsis and other angiosperm species," *PLoS ONE*, vol. 8, no. 3, p. e59988, 2013.

[34] S. Siljak-Yakovlev, "Evolutionary implications of heterochromatin and rDNA in chromosome number and genome size changes during dysploidy: a case study in Reichardia genus," *PLoS ONE*, vol. 12, no. 8, Article ID e0182318, 2017.

[35] C. A. Cullis, "The Environment as an active generator of adaptive genomic variation," in *In Plant adaptations to stress environments*, H. R. Lerner, Ed., pp. 149–160, Marcel Dekker, New York, NY, USA, 1999.

[36] A. V. Amosova, L. V. Zemtsova, Z. E. Grushetskaya et al., "Intraspecific chromosomal and genetic polymorphism in Brassica napus L. detected by cytogenetic and molecular markers," *Journal of Genetics*, vol. 93, no. 1, pp. 133–143, 2014.

[37] M. Murata, E. E. Roos, and T. Tsuchiya, "Mitotic delay in root tips of peas induced by artificial seed aging," *Botanical Gazette*, vol. 141, no. 1, pp. 19–23, 1980.

[38] J. A. Ozga, D. M. Reinecke, N. R. Knowles, and P. Blenis, "Characterization of the loss of seedling vigor in pea (Pisum sativum L.)," *Canadian Journal of Plant Science*, vol. 84, no. 2, pp. 443–451, 2004.

[39] O. Y. Yurkevich, I. V. Kirov, N. L. Bolsheva et al., "Integration of physical, genetic, and cytogenetic mapping data for cellulose synthase (CesA) genes in flax (Linum usitatissimum L.)," *Frontiers in Plant Science*, vol. 8, article 1467, 2017.

[40] W. L. Gerlach and J. R. Bedbrook, "Cloning and characterization of ribosomal RNA genes from wheat and barley," *Nucleic Acids Research*, vol. 7, no. 7, pp. 1869–1885, 1979.

[41] W. L. Gerlach and T. A. Dyer, "Sequence organization of the repeating units in the nucleus of wheat which contain 5S rRNA genes," *Nucleic Acids Research*, vol. 8, no. 21, pp. 4851–4865, 1980.

[42] P. Neumann, M. Nouzová, and J. Macas, "Molecular and cytogenetic analysis of repetitive DNA in pea (Pisum sativum L.)," *Genome*, vol. 44, no. 4, pp. 716–728, 2001.

[43] T. H. N. Ellis and S. J. Poyser, "An integrated and comparative view of pea genetic and cytogenetic maps," *New Phytologist*, vol. 153, no. 1, pp. 17–25, 2002.

[44] P. Neumann, D. Pozárková, J. Vrána, J. Doležel, and J. Macas, "Chromosome sorting and PCR-based physical mapping in pea (Pisum sativum L.)," *Chromosome Research*, vol. 10, no. 1, pp. 63–71, 2002.

[45] P. Smýkal, G. Aubert, J. Burstin et al., "Pea (*Pisum sativum* L.) in the genomic era," *Agronomy*, vol. 2, no. 4, pp. 74–115, 2012.

[46] J. Greilhuber and I. Ebert, "Genome size variation in Pisum sativum," *Genome*, vol. 37, no. 4, pp. 646–655, 1994.

[47] J. Doležel and J. Greilhuber, "Nuclear genome size: Are we getting closer?" *Cytometry Part A*, vol. 77, no. 7, pp. 635–642, 2010.

[48] J. Macas, P. Neumann, and A. Navrátilová, "Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*," *BMC Genomics*, vol. 8, article 427, 2007.

[49] S. I. S. Grewal and S. Jia, "Heterochromatin revisited," *Nature Reviews Genetics*, vol. 8, no. 1, pp. 35–46, 2007.

[50] Ž. Pezer, J. Brajković, I. Feliciello, and D. Ugarković, "Satellite DNA-mediated effects on genome regulation," *Genome Dynamics*, vol. 7, pp. 153–169, 2012.

[51] M. Tittel-Elmer, E. Bucher, L. Broger et al., "Stress-Induced Activation of Heterochromatic Transcription," *PLoS Genetics*, vol. 6, no. 10, p. e1001175, 2010.

[52] A. D. Krikorian, "Plants and somatic embryos in space: what have we learned?" *Gravitational and Space Biology Bulletin*, vol. 11, no. 2, pp. 5–14, 1998.

[53] N. G. Platova and V. M. Petrov, "Biological estimation of space radiation effect by way of plant bio-object example," *Aviakosmicheskaia i Ekologicheskaia Meditsina*, vol. 47, pp. 117-118, 2013.

[54] W. M. Waterworth, C. M. Bray, and C. E. West, "The importance of safeguarding genome integrity in germination and seed longevity," *Journal of Experimental Botany*, vol. 66, no. 12, pp. 3549–3558, 2015.

[55] H. Sonah, R. K. Deshmukh, A. Sharma et al., "Genome-wide distribution and organization of microsatellites in plants: an insight into marker development in Brachypodium," *PLoS ONE*, vol. 6, no. 6, Article ID e21298, 2011.

[56] T. E. Samatadze, O. V. Muravenko, N. L. Bolsheva, A. V. Amosova, S. A. Gostimsky, and A. V. Zelenin, "Investigation of chromosomes in varieties and translocation lines of pea Pisum sativum L. by FISH, Ag-NOR, and differential DAPI staining," *Genetika*, vol. 41, no. 12, pp. 1665–1673, 2005.

[57] J. Fuchs, M. Kühne, and I. Schubert, "Assignment of linkage groups to pea chromosomes after karyotyping and gene mapping by fluorescent in situ hybridization," *Chromosoma*, vol. 107, no. 4, pp. 272–276, 1998.

[58] A. Carmona, E. Friero, A. D. Bustos, N. Jouve, and A. Cuadrado, "The evolutionary history of sea barley (Hordeum marinum) revealed by comparative physical mapping of repetitive DNA," *Annals of Botany*, vol. 112, no. 9, pp. 1845–1855, 2013.

[59] J. S. Zheng, C. Z. Sun, S. N. Zhang et al., "Cytogenetic diversity of simple sequences repeats in morphotypes of Brassica rapa ssp. chinensis," *Frontiers in Plant Science*, vol. 7, article 1049, 2016.

[60] M. L. C. Vieira, L. Santini, A. L. Diniz, and C. D. F. Munhoz, "Microsatellite markers: What they mean and why they are so useful," *Genetics and Molecular Biology*, vol. 39, no. 3, pp. 312–328, 2016.

[61] L. Zhang, K. Zuo, F. Zhang et al., "Conservation of noncoding microsatellites in plants: Implication for gene regulation," *BMC Genomics*, vol. 7, article no. 323, 2006.

[62] J. Reneker, E. Lyons, G. C. Conant et al., "Long identical multi-species elements in plant and animal genomes," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 109, no. 19, pp. E1183–E1191, 2012.

[63] S. Kumar and S. Bhatia, "A polymorphic (GA/CT)n- SSR influences promoter activity of Tryptophan decarboxylase gene in Catharanthus roseus L. Don," *Scientific Reports*, vol. 6, no. 1, Article ID 33280, 2016.

[64] L. N. Mukhamedieva and V. V. Bogomolov, "Characteristic of toxic risks of air pollution by chemical admixtures aboard the piloted orbital stations," *Aviakosmicheskaia i Ekologicheskaia Meditsina*, vol. 43, pp. 17–23, 2009.

Research Article

Dynamics of the Centromeric Histone CENH3 Structure in Rye-Wheat Amphidiploids (Secalotriticum)

Yulia A. Lipikhina,[1] **Elena V. Evtushenko,**[1] **Oleg M. Lyusikov,**[2] **Igor S. Gordei,**[2] **Ivan A. Gordei,**[2] **and Alexander V. Vershinin** [ID][1]

[1]*Institute of Molecular and Cellular Biology SB RAS, Novosibirsk 630090, Russia*
[2]*Institute of Genetics and Cytology, NAS of Belarus, Minsk 220072, Belarus*

Correspondence should be addressed to Alexander V. Vershinin; avershin@mcb.nsc.ru

The centromeres perform integral control of the cell division process and proper distribution of chromosomes into daughter cells. The correct course of this process is often disrupted in case of remote hybridization, which is a stress factor. The combination of parental genomes of different species in a hybrid cell leads to a "genomic shock" followed by loss of genes, changes in gene expression, deletions, inversions, and translocations of chromosome regions. The created rye-wheat allopolyploid hybrids, which were collectively called secalotriticum, represent a new interesting model for studying the effect of remote hybridization on the centromere and its components. The main feature of an active centromere is the presence of a specific histone H3 modification in the centromeric nucleosomes, which is referred to as CENH3 in plants. In this paper the results of cytogenetic analysis of the secalotriticum hybrid karyotypes and the comparison of the CENH3 N-terminal domain structure of parent and hybrid forms are presented. It is shown that the karyotypes of the created secalotriticum forms are stable balanced hexaploids not containing minichromosomes with deleted arms, in full or in part. A high level of homology between rye and wheat enables to express both parental forms of *CENH3* gene in the hybrid genomes of secalotriticum cultivars. The CENH3 structure in hybrids in each crossing combination has some specific features. The percentage of polymorphisms at several amino acid positions is much higher in one of the secalotriticum hybrids, STr VD, than in parental forms, whereas the other hybrid, STr VM, inherits a high level of amino acid substitutions at the position 25 from the maternal parent.

1. Introduction

Many species of the Triticeae tribe are natural allopolyploids and attractive objects for obtaining synthetic hybrids, which are promising material for practical breeding. The most common and studied examples of such hybrids are allopolyploid triticale. The synthetic allohexaploid triticale has a genome structure similar to hexaploid bread wheat except that it has rye as one of its progenitor instead of the D genome donor *Aegilops tauschii*. Triticale is an important model for studying the rapid changes that occur subsequent to polyploidization involving genomic remodeling and changes in gene expression [1]. However, selection and genetic analysis of the triticale gene pool showed that the genetic potential of rye adaptability is insufficiently realized. At the molecular

level, it was found that 9% of genes in octoploid triticale and up to 30% of genes in hexaploid triticale become silent [2], which may be one of the reasons for the incomplete realization of hybrid genetic potential.

The synthesis of rye-wheat amphidiploids, which have wheat (*Triticum* L.) as the pollinator and rye (*Secale* L.) as the maternal form, can offer better setting for enhancing rye gene expression and for the creation of hybrids valuable for breeding [3]. Crosses like these are normally difficult to achieve due to incompatibility, and so rye-wheat amphiploids are not yet studied. Overcoming this barrier is usually associated with severe rearrangements in the genomes of parental forms, which is the most vivid manifestation of the "genomic shock" that occurs when parents' genomes are combined in

a hybrid cell [4] and is accompanied by various chromosomal abnormalities, including those affecting the centromere structure. It has been previously determined that using an intermediary species triticale as a source of wheat genomes in crosses with rye proved to be effective for overcoming the barrier of unidirectional progamic incompatibility of parent species [3]. As a result, these rye-wheat amphidiploid hybrids were created by crossing tetraploid rye ($^{S/}$RRRR, 2n = 4x = 28) with hexaploid triticale ($^{T/}$RRAABB, 2n = 6x = 42) and were named secalotriticum (×Secalotriticum, syn. ×Secalotriticum = *Secale* L. × *Triticum* L., $^{S/}$RRAABB, 2n = 6x = 42). The presence of rye maternal cytoplasm in hybrids was proved by analyzing restriction fragments of species-specific mitochondrial DNA18S/5S-repeated sequences and the *ndhH*-region of chloroplast DNA, which showed a pattern characteristic of chloroplast DNA and rye mitochondrial DNA [3]. The created secalotriticum hybrids are characterized by a wider variability range of morphological and economically valuable traits compared to the original triticale forms [3]. Secalotriticum creates more advantageous conditions for enhancing the expression of rye genetic systems and the manifestation of its valuable adaptive traits.

The process of the hybrid genome formation and its subsequent stabilization are directly related to the normalization of meiosis and proper segregation of chromosomes. Incompatibility of centromeres of different species seems to be the main reason of the chromosome elimination of one of the parental genomes in hybrids [5, 6]. The pivotal role in the proper chromosome segregation during meiosis and mitosis lies with centromeres, their identity being defined by the presence of the centromere-specific variant of histone H3 known in plants as centromere-specific histone H3 variant CENH3 [7, 8]. This is due to the fact that at the molecular level the most specialized and universal characteristic of the active centromere is the presence of CENH3 instead of the canonical histone H3 in the nucleosomes of centromeric chromatin. As it has been shown in some mammalian species and *Drosophila*, in case of centromeric histone loss, there is no kinetochore formation and no correct chromosome segregation during cell division for some reason [9]. Unlike canonical histone H3, which has a conserved structure, CENH3 normally shows considerable variability across species [5, 10, 11]. Different domains of this molecule are diverging differently. An extended N-terminal tail (NTT) domain and loop 1 of the histone fold domain (HFD) putatively interact with centromeric DNA [12] and show signatures of positive selection in some animal and plant species [13], while the part of the HFD that lies outside loop 1 is generally conserved [14, 15].

Because of a special role CENH3 has in the formation and function of centromeres, we explore effects of remote hybridization using an unusual combination of parental forms on the structure of this protein and perform a cytogenetic analysis of the karyotypes.

2. Materials and Methods

Two secalotriticum combinations (×*Secalotriticum* Rozenst., $^{S/}$RRAABB, 2n = 6x = 42) Verasen' × Mikhas' (STr VM) and

Verasen' × Dubrava (STr VD) were developed by hybridizing the tetraploid rye cultivar Verasen' ($^{S/}$RRRR, 2n=4x=28) with the hexaploid triticale cultivars Mikhas' and Dubrava ($^{T/}$AABBRR, 2n=6x=42). Triticale acted as an intermediary species, a source of wheat genomes and an inhibitor of rye S-RNases, which allowed overcoming the rye progamic incompatibility with wheat. Stabilization of karyotypes was facilitated by a single backcross of rye-triticale pentaploid F1 hybrids (S/RRABR, 5x=35) on triticale followed by self-pollination within 15 generations and constant selection for cytological stability and phenotypic homogeneity.

Karyotyping of secalotriticum, identification of the original species chromosomes was carried out by means of root apical meristem squashed preparations and differential staining using a Giemsa technique (C-banding) [16].

Molecular analysis was conducted on plants of parental forms and hybrids grown in a greenhouse. The total RNA was isolated from leaves of 10-12-day-old seedlings from 4-5 plants of parental forms. RNA of hybrid forms was isolated from each plant separately. Isolation of RNA was performed using the TRI Reagent (MRC, Ink., USA) [17]. To avoid contamination with genomic DNA, total RNA was treated with RQ-RNase Free DNase (Promega, Madison, USA). To synthesize cDNA, a RevertAidTM H Minus First Strand cDNA Synthesis Kit (Thermo Scientific, USA) was applied. The resulting cDNA was used as a template in a series of PCR reactions with primers synthesized specifically to amplify the CENH3 N-terminal tail (NTT). The primer sequences are $5'$ – ATGGCCCGCACCAAGC (F) and $5'$ – GAAACTCGACCGACTTCTG (R). The product size was 268 bp. PCR products were cloned into the pTZ57R/T plasmid (InsTAclone PCR cloning kit, Thermo Scientific, США) and analyzed by Sanger sequencing using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA). The reaction products were separated on the 3500 Genetic Analyzer (Applied Biosystems, USA). The sequences of individual clones obtained for each sample were analyzed using the UniPro Ugene software (http://ugene.net) and FASTA software package [18]. The search for the identity of the nucleotide sequences was carried out using the BLAST algorithm in the NCBI database (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Amino acids alignments were performed online using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo) and used for downstream analysis and visualization (http://www.jalview.org). Graphic images were prepared using the Jalview program (http://www.jalview.org/).

3. Results and Discussion

The STr VM (Verasen' × Mikhas') and STr VD (Verasen' × Dubrava) (S/RRAABB, 2n=6x=42) secalotriticum karyotypes were analyzed by differential staining (C-banding) [16]. This method reveals the pattern of heterochromatic regions localization, which is specific for each chromosome of rye and wheat, thereby allowing identification of each chromosome in a hybrid karyotype. Figures 1(a) and 1(b) show that the karyotypes of the STr VM and STr VD hybrids are stable balanced hexaploids not containing minichromosomes
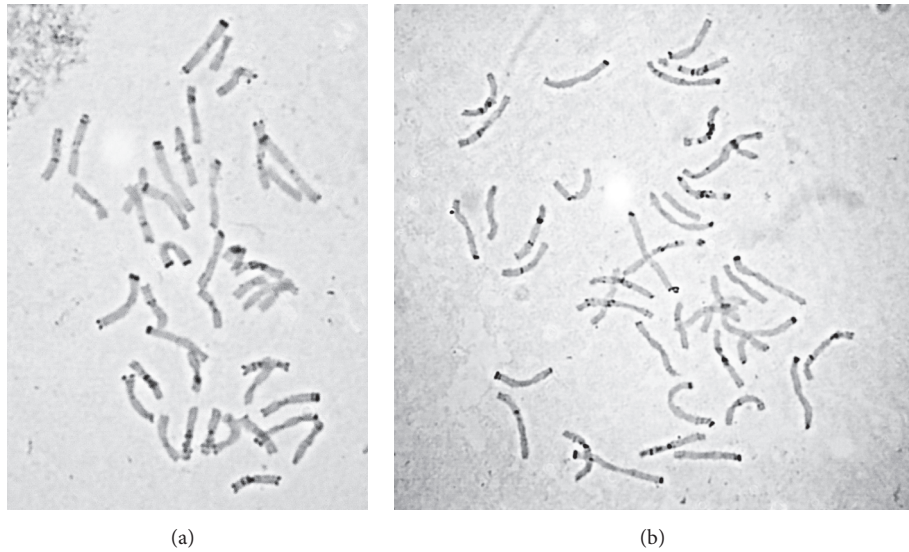
(a)

(b)

FIGURE 1: The STr VM (a) and STr VD (b) karyotypes of hybrid secalotriticum following C-banding appear with the whole sets of **R**, **A**, and **B** genomes.

with deleted arms, in full or in part. The hybrids and the original rye and wheat forms have an identical C-banding pattern.

One of the genomic shock manifestations, which arise from the combination of two parental genomes in a hybrid cell in case of remote hybridization, is various chromosomal aberrations. Deletions and translocations of individual chromosomal regions and chromosome arms are among the most common changes and have been found in the cytogenetic analysis of both wheat-rye substitution and addition lines [19, 20] and in triticale and offspring from crosses of triticale × wheat [21, 22]. The disorders herewith have been documented in the chromosomes of both parental forms. Stabilization of secalotriticum karyotypes was assisted by a single backcross of rye-triticale pentaploid F1 hybrids (S/RRABR, 5x=35) on triticale with subsequent self-pollination during 15 generations and constant selection for cytological stability and phenotypic homogeneity. Diploid RR-rye genome in rye-triticale F1 hybrids is a factor of meiosis stabilization and provides functionality of gametes with different chromosome composition. The formation of the secalotriticum genome takes place in F1BC1 on the basis of partially unreduced 21-chromosome gametes of F1 pentaploids with balanced chromosome sets of the original species haplogenomes (7(R), 7(A), 7(B)). Equation division (division into chromatids) of asyanptic univalents in anaphase I (AI) is the main source of chromosomal abnormalities during the second meiosis division (MII): individual chromatids are not included in the metaphase plate in MII; in anaphase II they often lag behind other chromosomes in division, segregate randomly, undergo fragmentation, and form micronuclei in the microspore tetrads, which leads to the elimination of genetic material and aneuploidy and the genomic instability [3, 23, 24]. Stability of the secalotriticum genome and its genetic diversity are determined by the cytoplasm type and some factors, apparently related to the structural and functional state of centromeres in its genome. Analyzing meiosis in hybrids, we observed the formation of (pseudo)univalents mainly as a result of desynapsis [23, 24]. Desynapsis here means the early decomposition of bivalents into univalents of desynaptic origin (pseudounivalents) that occur in the late prophase (prometaphase I) of meiosis. Unlike the asynaptic univalents characteristic of triticale, the pseudounivalents maintain their unipolar centromere orientation in the reductional (I) division of meiosis. The equational (II) division of meiosis was characterized by regular polar segregation and a low level of chromosome elimination [23, 24].

The secalotriticum forms created underwent rapid stabilization in meiosis during selection for productivity. Stabilization was due to the narrowing of the spectrum and a rapid decrease in the frequency of the chromosome abnormalities in the second meiotic division (by 15-50%) in F1-F3, including the level of chromosome elimination (from 30% to 6% tetrads with micronuclei). In F3-5 and subsequent secalotriticum generations, the normalization of the first division of meiosis was more pronounced: the average level of anomalies in MI metaphase decreased from 57.3% in F3 to 16.9% in F5 and also from 25-30% up to 8-10% in AI-AII anaphases and reached less than 5% at the tetrad stage [3, 23, 24]. The fact that the chromosomal changes mentioned above tend to occur quickly and most intensively in the first generations, up to the fifth, inclusive, and especially in F1 before the chromosome doubling, is probably a general trend in remote hybridization of cereals [3, 21, 23, 24]. Thus, continuous cycles of crosses in the secalotriticum production played a significant role in stabilizing the hybrid karyotype and replacing possible lost parts of the parental genomes in the first generations after hybridization.

Differences in the CENH3 structure between the parents allow estimating the expression level of the parental forms

```
                                          *
Verasen'             MARTKHPAVRKTKVPPKKKLGTRPPGGTQRRQDTDGAGTSATPRRAGRAAAPGAAEGATGQPKQRKPHRFR
Mikhas'              MARTKHPAVRKTKVPPKKKLGTRPSGGTQRRQDTDGAGTSATPRRAGRAAAPGAAEGATGQPKQRKPHRFR
Verasen'  x Mikhas'  MARTKHPAVRKTKVPPKKKLGTRPPGGTQRRQDTDGAGTSATPRRAGRAAAPGAAEGATGQPKQRKPHRFR
                                                       *              *  *           *
Verasen'             MARTKHPAVRKTKVPPKKKLGTRPPGGTQRRQDTDGAGTSATPRRAGRAAAPGAAEGATGQPKQRKPHRFR
Dubrava             MARTKHPAVRKTKVPPKKKLGTRPSGGTQRRQDTDGAGTSATPRRAGRAAAPGAAEGATGQPKQRKPHRFR
Verasen'  x Dubrava  MARTKHPAVRKTKVPPKKKLGTRPSGGTQRRQDTDGAGTSATPRRAGRAAAPGAAEGATGQPKQRKPHRFR
```

Figure 2: The amino acid consensus sequences of αCENH3 NTT domain derived from parental forms of rye (Verasen'), triticale (Mikhas', Dubrava), and secalotriticum amphidiploids (STr VM and STr VD). The asterisks indicate the positions of nonsynonymous amino acid substitutions. Amino acid polymorphic positions are color-coded according to their frequency: white, if below 50%; light-gray, if above 50%. In the positions 41, 54, 56, and 66, the percentage of substitutions does not exceed 50%, although it has high values for Dubrava and the hybrid; therefore the amino acids with the highest percentage are shown in these positions. The exact quantification of amino acid substitutions in the polymorphic positions is provided in Figure 3.
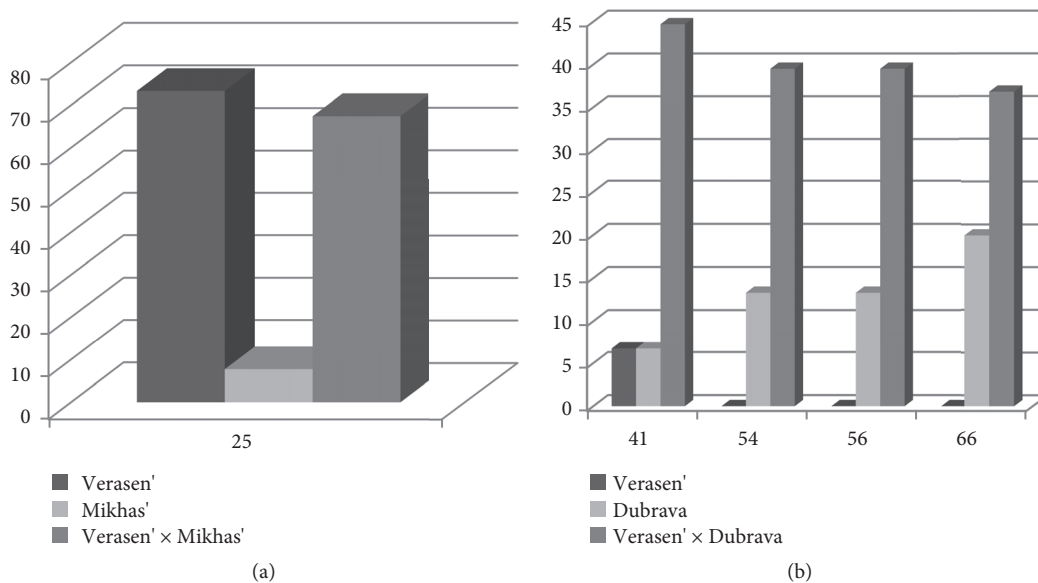

(a)

(b)

Figure 3: Quantitative determination of amino acid polymorphism in the NTT domain of αCENH3, leading to nonsynonymous amino acid substitutions. On the x-axis, the position in the amino acid sequence; on the y-axis, the percentage of clones having substitutions in polymorphic positions to the total number of clones sequenced. (a) Parental forms: Verasen' and Mikhas'; hybrids: Verasen' × Mikhas'. (b) Parental forms: Verasen' and Dubrava; hybrids: Verasen' × Dubrava. **Note:** fifteen cDNA clones were sequenced for each parental plant and the deduced amino acid sequences were obtained. About 40-45 clones were sequenced for each hybrid plant.

of the protein in a new genomic environment that arises in a hybrid cell under remote hybridization. The first work on studying possible connection between differences in the CENH3 structure in parental forms and the processes of parental genomes chromosome segregation during hybrid cell division was carried out on hybrids obtained from the crossing of cultured barley *H. vulgare* and its closest wild relative *H. bulbosum* [5]. The CENH3 molecules were not incorporated in the centromeres of *H. bulbosum* chromosomes, which were herewith inactivated and eliminated from hybrid embryos. Perhaps the absence of CENH3 in the centromeres of *H. bulbosum* chromosomes is caused by perceptible differences in the structure of this protein between barley species, especially in the structure of the N-terminal tail (NTT) [5]. Unlike barley species, DNA sequences of α and β CENH3 forms of various rye and wheat species have a very high identity (95-99%) [25], which significantly complicates

the search for interspecific differences and, accordingly, the identification of the nature of their inheritance in hybrid genomes.

Consensus amino acid sequences derived from sequenced cDNA clones obtained from the αCENH3 NTT of Verasen' and Mikhas' parental forms differ only in one nonsynonymous substitution of amino acid serine to proline at position 25 (Figure 2). It was found in 73.3% of the analyzed clones in the Verasen' variety (Figure 3(a)), but only in 7.7% of the paternal variety Mikhas' clones. The secalotriticum cultivar Verasen' × Mikhas' inherits a high content of proline from the maternal form Verasen' (67.4%, Figure 3(a)). However, a different cross does not support this tendency: Verasen' × Dubrava has as low percentage of substitutions (2.6%) as the donor triticale cultivar Dubrava (6.7%). On the other hand, Dubrava typically has more polymorphism than the donor Verasen' at several nucleotide

positions, leading to the following amino acid substitutions: alanine to valine, alanine to glycine, glutamine acid to glutamine, and lysine to asparagine (Figures 2 and 3(b)). Interestingly, although the percentage of substitutions does not exceed 50% (of the total number of clones sequenced), this value is much higher in the secalotriticum cultivar Verasen' × Dubrava than in Dubrava, something like a case of heterosis (Figure 3(b)). Thus, the results obtained indicate that both parental genomes participate in the production of the CENH3 protein in the secalotriticum hybrids. At the same time, comparison of the nucleotide and amino acid sequences of the most variable N-terminal tail of the CENH3 molecule in parental forms and hybrids does not reveal a definite tendency in inheritance.

Published data on the regulation of the expression of different genes and other classes of DNA sequences in hybrid genomes obtained by crossing different species or contrast populations of one species indicate a complex and heterogeneous inheritance pattern. The phenomenon of nucleolus dominance, characteristic for the expression of ribosomal RNA genes in triticale, as an example of genomic dominance, consists in the repression of rye rRNA genes by means of cytosine methylation [26]. However, this type of inheritance, namely, the genomic dominance of one of the parents, is not typical of most other examples. The results of gene expression study in hybrids between different ecotypes of Arabidopsis are contradictory. In one case, a general maternal dominance has been identified [27], while in the other it has been found that the maternal and paternal genomes are transcriptionally equivalent [28]. Also, it has been demonstrated that various hybrid combinations show significant variations in activation of parental alleles [29]. In some hybrids between Arabidopsis ecotypes heterosis for biomass and seed yield has been documented [30]. In these hybrids, 95% of active genes show an intermediate expression level. Among most other nonadditively expressed genes, the expression shift was towards the maternal parent [30]. When studying transcription and epigenetic adaptation of maize chromosomes in supplemented oat-maize lines containing the complete oat genome and individual maize chromosomes, most maize genes showed transcription specific for maize, but repression of gene activity was a more common trend than activation [31].

The level of parental gene expression in allopolyploid hybrids obtained by crossing wheat and rye species is of significant interest for us. We do not know such studies conducted on secalotriticum; however, an extensive analysis of rye gene expression using cDNA isolated from various tissues of hybrid plants has been carried out on allohexaploid triticale obtained from the *T. turgidum* × *S. cereale* cross [1]. The classes of absent or silent rye genes have been identified. A comparison between diploid rye and hexaploid triticale revealed 112 rye cDNA contigs (∼ 0.5% of the total amount), which were not determined by expression analysis in any of the triticale tissues, though their expression was relatively high in rye tissues. The average DNA sequence identity between rye genes not detected in triticale and their most similar contigs in the A and B genomes in *T. aestivum* was only 81%. This degree of identity was significantly lower than the global average of 91% identity between the set of 200 randomly selected rye genes and their best matches in the A and B genomes of *T. aestivum*. Thus, rye genes with a low similarity to their homeologs in *Triticum* genomes have a higher probability of being repressed or absent due to deletions in the genomes of allopolyploids. This conclusion is in good agreement with our results. High identity of rye and wheat CENH3 sequences (99%) enables a high level of expression of both parental forms in secalotriticum hybrid genomes.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] H. B. Khalil, M.-R. Ehdaeivand, Y. Xu, A. Laroche, and P. J. Gulick, "Identification and characterization of rye genes not expressed in allohexaploid triticale," *BMC Genomics*, vol. 16, no. 1, 2015.

[2] X.-F. Ma, P. Fang, and J. P. Gustafson, "Polyploidization-induced genome variation in triticale," *Genome*, vol. 47, no. 5, pp. 839–848, 2004.

[3] I. A. Gordey, N. B. Belko, and O. M. Lyusikov, *Sekalotritikum (×Secalotriticum): the genetic basis for the creation and formation of the genome*, vol. 214, Minsk: Belaruskaya Navuka Publ., 2011.

[4] B. McClintock, "The significance of responses of the genome to challenge," *Science*, vol. 226, no. 4676, pp. 792–801, 1984.

[5] M. Sanei, R. Pickering, K. Kumke, S. Nasuda, and A. Houben, "Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 108, no. 33, pp. E498–E505, 2011.

[6] T. Ishii, R. Karimi-Ashtiyani, and A. Houben, "Haploidization via Chromosome Elimination: Means and Mechanisms," *Annual Review of Plant Biology*, vol. 67, pp. 421–438, 2016.

[7] D. J. Amor, P. Kalitsis, H. Sumer, and K. H. A. Choo, "Building the centromere: From foundation proteins to 3D organization," *Trends in Cell Biology*, vol. 14, no. 7, pp. 359–368, 2004.

[8] L. Comai, S. Maheshwari, and P. A. Marimuthu, "Plant centromeres," *Current Opinion in Plant Biology*, vol. 36, pp. 156–167, 2017.

[9] P. B. Talbert, T. D. Bryson, and S. Henikoff, "Adaptive evolution of centromere proteins in plants and animals," *Journal of Biology*, vol. 3, no. 4, article no. 18, 2004.

[10] S. Maheshwari, E. H. Tan, A. West, F. C. H. Franklin, L. Comai, and S. W. L. Chan, "Naturally Occurring Differences in CENH3 Affect Chromosome Segregation in Zygotic Mitosis of Hybrids," *PLoS Genetics*, vol. 11, no. 1, p. e1004970, 2015.

[11] P. Neumann, Z. Pavlikova, A. Koblizkova et al., "Centromeres off the hook: massive changes in centromere size and structure following duplication of CenH3 gene in Fabeae species," *Molecular Biology and Evolution*, vol. 32, pp. 1862–1879, 2015.

[12] D. Vermaak, H. S. Hayden, and S. Henikoff, "Centromere targeting element within the histone fold domain of Cid," *Molecular and Cellular Biology*, vol. 22, no. 21, pp. 7553–7561, 2002.

[13] H. S. Malik and S. Henikoff, "Adaptive evolution of Cid, a centromere-specific histone in Drosophila," *Genetics*, vol. 157, no. 3, pp. 1293–1298, 2001.

[14] P. B. Talbert, R. Masuelli, A. P. Tyagi, L. Comai, and S. Henikoff, "Centromeric localization and adaptive evolution of an Arabidopsis histone H3 variant," *The Plant Cell*, vol. 14, no. 5, pp. 1053–1066, 2002.

[15] C. D. Hirsch, Y. Wu, H. Yan, and J. Jiang, "Lineage-specific adaptive evolution of the centromeric protein cenh3 in diploid and allotetraploid oryza species," *Molecular Biology and Evolution*, vol. 26, no. 12, pp. 2877–2885, 2009.

[16] E. D. Badaeva, L. F. Sozinova, N. S. Badaev, O. V. Muravenko, and A. V. Zelenin, ""Chromosomal passport" of Triticum aestivum L. em Thell. cv. Chinese Spring and standardization of chromosomal analysis of cereals," *Cereal Research Communications*, vol. 18, pp. 273–281, 1990.

[17] P. Chomczynski and N. Sacchi, "Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction," *Analytical Biochemistry*, vol. 162, no. 1, pp. 156–159, 1987.

[18] W. R. Pearson and D. J. Lipman, "Improved tools for biological sequence comparison," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 85, pp. 2444–2448, 1988.

[19] A. G. Alkhimova, J. S. Heslop-Harrison, A. I. Shchapova, and A. V. Vershinin, "Rye chromosome variability in wheat-rye addition and substitution lines," *Chromosome Research*, vol. 7, no. 3, pp. 205–212, 1999.

[20] S. Fu, Z. Lv, X. Guo, X. Zhang, and F. Han, "Alteration of Terminal Heterochromatin and Chromosome Rearrangements in Derivatives of Wheat-Rye Hybrids," *Journal of Genetics and Genomics*, vol. 40, no. 8, pp. 413–420, 2013.

[21] R. Appels, J. P. Gustafson, and C. E. May, "Structural variation in the heterochromatin of rye chromosomes in triticales," *Theoretical and Applied Genetics*, vol. 63, no. 3, pp. 235–244, 1982.

[22] X.-F. Ma and J. P. Gustafson, "Timing and rate of genome variation in triticale following allopolyploidization," *Genome*, vol. 49, no. 8, pp. 950–958, 2006.

[23] O. M. Lyusikov and I. A. Gordei, "Analysis of meiotic genome stabilization in the rye-wheat amphidiploid secalotriticum (×Secalotriticum, S/RRAABB, 2n = 42)," *Russian Journal of Genetics*, vol. 50, no. 7, pp. 693–700, 2014.

[24] O. M. Lyusikov and I. A. Gordei, "Cytogenetic stabilization of secalotriticum (×Triticosecale derzhavinii secalotriticum Rozenst., et Mittelst., $^{S/}$RRAABB, 2n = 42) genome," *Cytology and Genetics*, vol. 49, no. 2, pp. 118–124, 2015.

[25] E. V. Evtushenko, E. A. Elisafenko, S. S. Gatzkaya, Y. A. Lipikhina, A. Houben, and A. V. Vershinin, "Conserved molecular structure of the centromeric histone CENH3 in Secale and its phylogenetic relationships," *Scientific Reports*, vol. 7, no. 1, 2017.

[26] J. Lima-Brito, H. Guedes-Pinto, and J. S. Heslop-Harrison, "The activity of nucleolar organizing chromosomes in multigeneric F1 hybrids involving wheat, triticale, and tritordeum," *Genome*, vol. 41, no. 6, pp. 763–768, 1998.

[27] D. Autran, C. Baroux, M. T. Raissig et al., "Maternal epigenetic pathways control parental contributions to arabidopsis early embryogenesis," *Cell*, vol. 145, no. 5, pp. 707–719, 2011.

[28] M. D. Nodine and D. P. Bartel, "Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos," *Nature*, vol. 482, no. 7383, pp. 94–97, 2012.

[29] G. Del Toro-De León, M. García-Aguilar, and C. S. Gillmor, "Non-equivalent contributions of maternal and paternal genomes to early plant embryogenesis," *Nature*, vol. 514, no. 7524, pp. 624–627, 2014.

[30] M. M. Alonso-Peral, M. Trigueros, B. Sherman et al., "Patterns of gene expression in developing embryos of Arabidopsis hybrids," *The Plant Journal*, vol. 89, no. 5, pp. 927–939, 2017.

[31] Z. Dong, J. Yu, H. Li et al., "Transcriptional and epigenetic adaptation of maize chromosomes in Oat-Maize addition lines," *Nucleic Acids Research*, vol. 46, no. 10, pp. 5012–5028, 2018.

Research Article

Stay-Green QTLs Response in Adaptation to Post-Flowering Drought Depends on the Drought Severity

Nasrein Mohamed Kamal [iD],[1,2] **Yasir Serag Alnor Gorafi,**[2,3] **Hisashi Tsujimoto,**[2] **and Abdelbagi Mukhtar Ali Ghanim** [iD][1,4]

[1]*Biotechnology and Biosafety Research Center, Agricultural Research Corporation, P.O. Box 30, Shambat, Khartoum North, Sudan*
[2]*Arid Land Research Center, Tottori University, 1390 Hamasaka, Tottori 680-0001, Japan*
[3]*Agricultural Research Corporation, P.O. Box 126, Wad Medani, Sudan*
[4]*Plant Breeding and Genetics Laboratory, FAO/IAEA Joint Division of Nuclear Techniques in Food and Agriculture,*
 International Atomic Energy Agency (IAEA), Seibersdorf, Austria

Correspondence should be addressed to Nasrein Mohamed Kamal; renokamal@gmail.com
and Abdelbagi Mukhtar Ali Ghanim; a.mukhtar-ali-ghanim@iaea.org

Stay-green trait enhances sorghum adaptation to post-flowering drought. Six stay-green backcross introgression lines (BILs) carrying one or more stay-green QTLs (Stg1-4) and their parents were characterized under non-stress (W_{100}: 100% of soil field capacity (FC)) and two levels of post-flowering drought (W_{75}: 75% FC; W_{50}: 50% FC) in a controlled condition. We aimed to study the response and identify the drought threshold of these QTLs under different levels of post-flowering drought and find traits closely contributing to grain yield (GY) under different drought severity. W_{50} caused the highest reduction in BILs performance. From W_{100} to W_{50}, the GY of the recurrent parent reduced by 70%, whereas that of the BILs reduced by only 36%. W_{75} and W_{50} induce different behavior/response compared to W_{100}. Harvest index contributed to the GY under the three water regimes. For high GY under drought transpiration rate at the beginning of drought and mid-grain filling was important at W_{75}, whereas it was important at mid-grain filling and late-grain filling at W_{50}. Stay-green trait can be scored simply with the relative number of green leaves/plants under both irrigated and stress environments. QTL pyramiding might not always be necessary to stabilize or increase the GY under post-flowering drought. The stay-green QTLs increase GY under drought by manipulating water utilization depending on drought severity.

1. Introduction

The global population will increase to 9 billion by 2050 and most of the increase will occur in sub-Saharan Africa [1], increasing the risk of food insecurity in this region [2]. Therefore, making plants resilient to the challenges of a water-scarce planet where climate change and global warming threaten food supplies is the major challenge facing the humanity [3]. Drought is perhaps the most important abiotic stress limiting crop productivity in the rain-fed agriculture around the world [4]. Sorghum (*Sorghum bicolor* (L.) Moench) is a grain crop that is well adapted to hot and dry climates. The productivity of sorghum, the major

cereal crop grown for food, feed, and fuel, is usually under threat of terminal drought, which is likely to occur in rain-fed environments during grain filling [5]. It is reported that stay-green genotypes that exhibited the ability to retain green leaf area during grain filling under terminal drought produce higher grain yield than the non-stay-green genotypes [6–10]. Over the last 30 years, the stay-green trait has been used in breeding programs for improvement of sorghum terminal drought tolerance [11, 12]. Several sources for stay-green have been identified, including B35, SC56, and E-36 [13–15]. The quantitative trait loci (QTLs) that contribute to the stay-green trait have been mapped in a range of populations, mostly derived from crosses with B35, a derivative of an Ethiopian

dura landrace [16–23]. Xu and Sanchez [23, 24] identified four major QTLs in B35, including Stg1 (Stay-green QTL 1) and Stg2 both located on the linkage group SBI-03, Stg3 on SBI-02 and Stg4 on SBI-05. These QTLs explain 20, 30, 16, and 10%, respectively, of the phenotypic variation of the stay-green under post-flowering drought stress. Reddy [25] validated those QTLs reported in earlier studies and indicated that Stg2 and Stg3 were prominent in their expression.

The exact physiological mechanism of stay-green and the role of each individual QTL on the final phenotype of the stay-green genotypes under post-flowering drought stress are still unclear. However, a number of reports increased our understanding of this complicated drought adaptation mechanism. Borrell and Hammer [26, 27] explained the delayed onset and reduced rate of leaf senescence in stay-green genotypes by the high specific leaf nitrogen and nitrogen uptake during grain filling. Harris [28] reported that, under post-flowering water deficit, Stg2, Stg3, and Stg4 near isogenic lines (NILs) exhibited delayed onset of leaf senescence compared with the non-stay-green genotype, RTx7000, while significantly lower rates of leaf senescence in relation to RTx7000 were displayed by all of the stay-green NILs to varying degrees, but particularly by the Stg2 NIL. The Stg1 and Stg4 NILs exhibited greener leaves at flowering relative to RTx7000, indicated by higher SPAD values. Borrell [29] reported that hybrids containing B35, the source of stay-green, have higher transpiration efficiency (TE) than other eight hybrids examined. They suggested that the higher TE was due to increased photosynthetic capacity associated with higher specific leaf nitrogen, rather than reduced stomatal conductance. Vadez [9] studied the effect of different stay-green QTLs on modification of tillering and leaf area at flowering, transpiration efficiency, water extraction, harvest index, and grain yield under both terminal drought and fully irrigated conditions in 29 introgression lines with different stay-green QTLs in two backgrounds. They concluded that StgB and Stg1 modify the TE and water extraction depending on the background. Stay-green QTLs decrease the canopy size before flowering to conserve soil water for use during grain filling; the increased water uptake during grain filling in stay-green NILs relative to the non-stay-green parent RTx7000 resulted in higher biomass production, grain number, and yield [5, 30]. Moreover, Jaegglia [31] explained that tiller leaf area rather than transpiration efficiency, or transpiration per leaf area, was the main driver of weekly transpiration and the reduced pre-flowering water use in stay-green lines. According to Vadez and Borrell [5, 9, 30], the differences in TE are still unexplained and work is ongoing to investigate traits that might be related to leaf conductance aspects. In soils with good water-holding capacity, any water savings during the pre-flowering period increases water availability during the post-flowering period, therefore allowing plants to retain the photosynthetic capacity for longer by 'staying green' during grain filling [31].

The effect of the stay-green QTLs on modification of plant parameters under different levels of post-flowering drought is not extensively studied. Effect comparison between single QTL and QTL pyramiding under a specific level of drought severity has not been investigated adequately. Previously, we

introgressed the stay-green trait into the Sudanese sorghum cultivar 'Tabat' [32, 33] and evaluated the response of the BC_2F_4 stay-green lines under irrigated and drought conditions; we concluded that identification of the drought threshold is needed for better understanding the physiological reactions under specific drought severity. In this study, we analyzed the effect of single or more stay-green QTLs on modification of plant performance under controlled, non-stress condition and two levels of post-flowering drought stress by examining the plant physiological status through direct measurement of the leaves activity under drought. Also, we aimed to identify the level of drought to which those stay-green QTLs can confer drought tolerance. Moreover, we studied traits that mostly contribute to grain yield under different drought severity.

2. Materials and Methods

2.1. Plant Materials. Six BC_3F_3 stay-green backcross introgression lines (BILs) derived from a cross of drought sensitive cultivar 'Tabat' (abbrev. TAB, the recurrent parent) and stay-green donor B35 were used in this study. The backcross lines were produced at Agricultural Research Corporation, Sudan [32, 33]. Out of the six BILs, four carry a single stay-green QTL (Stg1, Stg2, Stg3, and Stg4, respectively), one BIL carries two QTLs (Stg1+4), and the other BIL carries three QTLs (Stg1+2+4). TAB is an improved Sudanese variety released for irrigated areas [34, 35] and B35 is a partially converted selection of the durra sorghum IS12555 from Ethiopia [13]. There was no big variation in the flowering of the genotypes and all flowered in the range from 70 to 76 days after sowing.

2.2. Pot Experiment. A pot experiment was conducted in a glasshouse in the Arid Land Research Center (ALRC), Tottori University (Tottori, Japan; $35°32'N, 134°13'E$) from June to November. The pots were filled with 15 kg of Tottori sandy soil. The chemical properties were reported by Fujiyama and Nagai [36]. As reported by Sohail et al. [37], we used three inorganic fertilizers: compound macronutrients N : P : K (16 : 16 : 16) in a rate of 0.4g/Kg (Central Glass Co., Ltd, Tokyo, Japan) and Ca : Mg (21 : 0.6) (Hitachi Chemical Co., Ltd, Tokyo, Japan) at a rate of 0.7g/Kg and micronutrients Mg : Mn : B (8.4 : 0.3 : 0.3) (MC Ferticom Co., Ltd, Tokyo, Japan) at a rate of 0.3g/Kg.

Five seeds were sown per pot, and two weeks after sowing the seedlings were thinned and only one plant was allowed to grow beyond until maturity stage. After flowering, we applied three different drought levels (soil water regimes): W_{100} (100% of soil field capacity (FC); water content was 120 ml/kg soil); W_{75} (75% of soil FC; water content was 90 ml/kg soil) representing moderate drought; and W_{50} (50% of soil FC representing severe drought; water content was 60 ml/kg soil). The experiment was arranged in a completely randomized design with three replications. The position of each pot was randomized and changed weekly in the glasshouse to ensure uniform environmental conditions. Usually, the pots were weighted every day and irrigated with tap water to keep the specific FC; usually, pots in W_{100} were

irrigated before reaching the FC of the pots in W_{75}, and pots in W_{75} were irrigated before reaching the FC of the pots in W_{50}, whereas pots in W_{50} were irrigated before reaching FC of 35%. During the experiment, the average maximum temperature in the glasshouse ranged from 34 to 23°C and the minimum ranged from 19 to 10°C.

2.3. Morphological Traits. Days to heading (DTH) were calculated as the number of days between the sowing date and the date when 50% of all the shoots in a pot had fully emerged spikes. At physiological maturity, plant height (PH) was measured in centimeters (cm) from the ground to the tip of the spike in each pot before harvesting. Days to maturity (DTM) were calculated from sowing date to 50% senescence of the spikes. Finally, grain yield per pot (GY) was determined as the weight (grams) of the grain from each pot; Biomass (BM) was determined as the weight (grams) of the aboveground fresh biomass, and harvest index (HI) was calculated using the formula

$$HI = \frac{Grain\ yield}{above\ ground\ biomass} * 100. \quad (1)$$

Yield susceptibility index (YSI) was calculated according to Fischer and Maurer [37]:

$$YSI = \frac{(1 - Y/Yp)}{D} \quad (2)$$

where Y is the GY of the genotype at drought, Yp is the mean GY of the genotypes at control, and D (stress intensity) = 1 – X/Xp, where X is the mean Y of all genotypes and Xp is the mean Yp of all genotypes. Genotypes were classified as highly tolerant (YSI ≤ 0.50), moderately tolerant (0.50 < YSI ≤ 1.00), or sensitive (YSI > 1.00) to drought [38, 39].

2.4. Leaf Measurement for Chlorophyll Content and Relative Number of Green Leaves/Plants. Detailed leaf observations were made on three replicates in each treatment. Fully expanded and senesced leaf number was recorded as described by Hammer [40] at one week before drought (WD), mid-grain filling (GF), and maturity (M).

A leaf was considered fully expanded when its ligule became visible above the enclosing sheath of the previous leaf. A leaf was considered senesced when more than 50% of its area turned yellow. Relative number of green leaves/plants (GN) was calculated as

$$\frac{a\ number\ of\ green\ leaves\ per\ plant}{total\ number\ of\ leaves\ per\ plant} * 100. \quad (3)$$

Leaf area of each individual fully expanded leaf was estimated nondestructively from the product of its length, greatest width, and a shape factor of 0.57, which was established by regressing the product of width and length of a leaf against its actual leaf area measured destructively at the end of the experiment. These estimates of individual leaf sizes, combined with observations of fully expanded and senesced leaves, allowed the estimation of green leaf area [41].

For the chlorophyll content (SPAD) data was represented as relative chlorophyll content (RCC) to ease the explanation and understanding of the degradation of leaf chlorophyll (senescence). Chlorophyll content was measured by the chlorophyll meter SPAD-502 (Konica Minolta). The arbitrary SPAD values can be translated into the actual value of total chlorophyll/unit area (mg cm^{-2}) using the equation: Chlorophyll content= SPAD values x 0.003 — 0.048, as described by Xu [23].

2.5. Leaf Gas Exchange Measurement. The fully expanded second leaf from the apex position and flag leaf were used for measurements. The photosynthesis rate (PR) and transpiration rate (TR) were measured using LI-6400 portable photosynthesis system (LI-COR Bioscience, Lincoln, NE, USA), at three growth stages: WD, GF, and late-grain filling (LGF) during sunny days. During measurement the chamber temperature was 25°C, the reference CO_2 concentration was 400 μmol mol^{-1}, the relative humidity was approximately 25%, and the irradiance was 1200 μmol m^{-2} s^{-1}.

2.6. Statistical Analysis. Two-way ANOVA was performed to assess the effect of genotype (G), drought treatment (W), and genotype-by-drought treatment (G×W) interaction for the different traits measured using GenStat version 17. The ANOVA was followed by Fisher's protected least significant difference (PLSD) test at $P < 0.05$. PCA analysis was performed using STAR software (STAR, version 1.4., International Rice Research Institute, Los Baños, Philippines; http://bbi.irri.org/products). Simple linear regression was performed using a linear regression model.

3. Results

3.1. The Drought Threshold of the Stay-Green BILs. The drought treatment effect was significant for all traits studied under W_{100}, W_{75}, and W_{50} ($P<0.0001$). In all traits except BM, PH, PR, and TR at WD, W_{100} had the highest mean values compared to W_{75} and W_{50} with the latter being the lowest (Tables 1 and 2). In the case of BM, W_{100} and W_{75} did not differ significantly, whereas, in the case of PH, W_{75} and W_{50} did not differ significantly. These findings indicated clearly that W_{50} as drought treatment was more severe than W_{75}. From W_{100} to W_{50}, the GY of the recurrent parent reduced by 70% whereas that of the BILs reduced by only 36%. The reduction from W_{100} to W_{50} did not exceed 50% in all traits except GLA and PR at LGF. Nevertheless, the BILs showed clear stay-green expression under W_{75} (moderate drought) (Tables 1 and 2).

3.2. The Effect of the Introgressed QTLs under No Drought at W_{100}. Generally, all the BILs with the different QTLs were comparable to or even better than TAB in their performance at W_{100} (Table 1 and Figures 1–5). However, BIL Stg1+4 had lower GY and HI, and BIL Stg3 had lower BM than TAB (Table 1). BILs Stg1, Stg2, and Stg1+4 had higher GLA than TAB at all stages (Figure 1(a)). BILs Stg3 and Stg1+2+4 had lower TR than TAB at WD and LGF, whereas at GF all BILs

TABLE 1: Grain yield, percentage (%) over TAB, biomass, harvest index, and plant height of the six stay-green introgression lines and their parents evaluated under control conditions (W_{100}, 100% soil field capacity) and two post-flowering drought conditions (W_{75}, 75%; W_{50}, 50% soil field capacity).

Genotypes	Grain yield (g)			% Over TAB		Biomass (g)			Harvest index (%)			Plant height (cm)		
	W_{100}	W_{75}	W_{50}	W_{75}	W_{50}	W_{100}	W_{75}	W_{50}	W_{100}	W_{75}	W_{50}	W_{100}	W_{75}	W_{50}
TAB	48.4	28.0	14.4	—	—	300.0	255.0	210.0	16.1	12.1	6.4	167.3	145.0	142.3
B35	15.3	16.6	14.2	—	—	280.0	256.0	236.0	6.1	8.9	6.1	115.7	116.7	117.0
Stg1	46.5	35.5	27.0	26.9	87.5	338.3	320.0	260.0	12.1	10.4	7.1	166.3	136.0	134.0
Stg2	44.4	13.5	3.2	-51.8	-77.9	320.0	350.0	333.3	14.0	3.8	1.2	161.7	150.3	142.3
Stg3	42.2	43.3	41.2	54.6	186.1	250.0	246.7	206.7	16.5	17.1	19.1	172.3	162.3	166.3
Stg4	41.2	33.4	29.1	19.2	102.3	275.0	290.0	270.0	15.5	10.5	7.8	160.3	158.7	146.3
Stg1+4	39.0	35.0	31.0	25.1	115.3	333.3	310.0	315.0	10.7	8.3	4.0	160.0	156.3	149.0
Stg1+2+4	43.6	42.1	32.1	50.2	122.9	280.0	276.7	213.3	14.3	14.3	14.9	162.0	163.3	160.3
Mean	40.1	30.9	24.0	—	—	297.1	288.0	255.5	13.2	10.7	8.3	158.2	148.6	144.7
SEM±	2.195	2.210	2.560	—	—	7.510	7.860	9.740	0.740	0.860	1.140	4.290	3.330	3.180
[a]LSD ([b]G)	8.880	6.180	8.180	—	—	38.060	34.840	23.690	3.2	3.1	1.2	24.9	10.2	12.0
LSD (G x [c]T)	3.980	3.980	3.980	—	—	18.650	18.650	18.650	1.630	1.630	1.630	9.41	9.41	9.41
P value (G)	<0.0001	<0.0001	<0.0001	—	—	0.002	0.000	<0.0001	<0.0001	<0.0001	<0.0001	0.007	<0.0001	<0.0001
P value (G x T)	<0.0001	<0.0001	<0.0001	—	—	0.004	0.004	0.004	<0.0001	<0.0001	<0.0001	0.15ns[d]	0.15ns	0.15ns
CV (%)	26	35	41	—	—	12	13	18	27	39	45	13	11	10

[a]LSD: least significant difference: $P<0.05$; [b]G: genotypes; [c]T: drought treatment; [d]ns: not significant.

TABLE 2: Green leaf area, relative chlorophyll content, relative number of green leaves/plants, photosynthesis rate, and transpiration rate of the six stay-green introgression lines and their parents evaluated under control condition (W_{100}, 100% soil field capacity) and two post-flowering drought conditions (W_{75}, 75% soil field capacity; W_{50}, 50% soil field capacity).

	Green leaf area (cm^2)			Relative chlorophyll content			Relative number of green leaves/plants			Photosynthesis rate (μmol CO_2 m^{-2} s^{-1})			Transpiration rate (mmol H_2O m^{-2} s^{-1})		
	WD	GF	M	WD	GF	M	WD	GF	M	WD	GF	LGF	WD	GF	LGF
W_{100}	679[a]	585[a]	262[a]	—	—	—	50.8[a]	43.8[a]	21.0[a]	27.5[a]	27.2[a]	16.9[a]	4.0[a]	3.4[a]	2.7[a]
W_{75}	619[a]	539[b]	251[a]	95.2[a]	95.2[a]	53.8[a]	45.8[b]	40.4[b]	18.5[b]	23.8[b]	21.2[b]	10.2[b]	4.1[ab]	1.4[b]	0.6[b]
W_{50}	604[a]	449[c]	193[b]	92.8[b]	77.9[b]	46.4[b]	43.5[b]	33.1[c]	13.9[c]	24.9[ab]	17.6[c]	8.6[c]	4.2[b]	2.2[c]	1.8[c]

WD: one week before drought; GF: mid-grain filling; LGF: late-grain filling; M: maturity. Different letters are significantly different (Fischer PLSD, $P<0.05$).
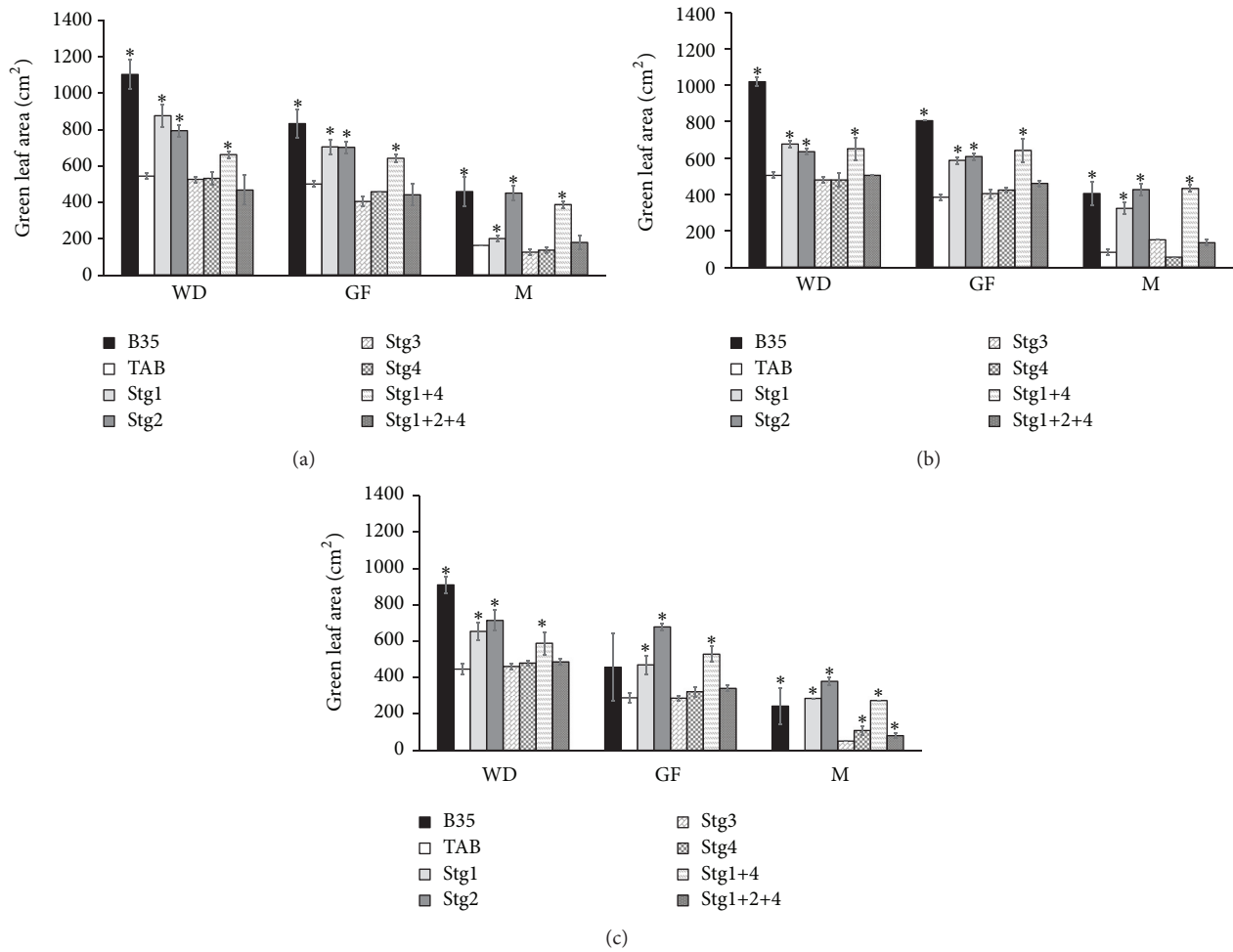
(a)

(b)

(c)

Figure 1: Green leaf area (GLA) at one week before drought (WD), mid-grain filling (GF), and maturity (M) under control W_{100} (a), 75% field capacity W_{75} (b), and 50% filed capacity W_{50} (c) of the six stay-green sorghum introgression lines evaluated with their parents under W_{100}, W_{75}, and W_{50} of soil field capacity. Asterisks indicate significant difference from Tabat (TAB) ($P < 0.05$, Fisher's PLSD test).
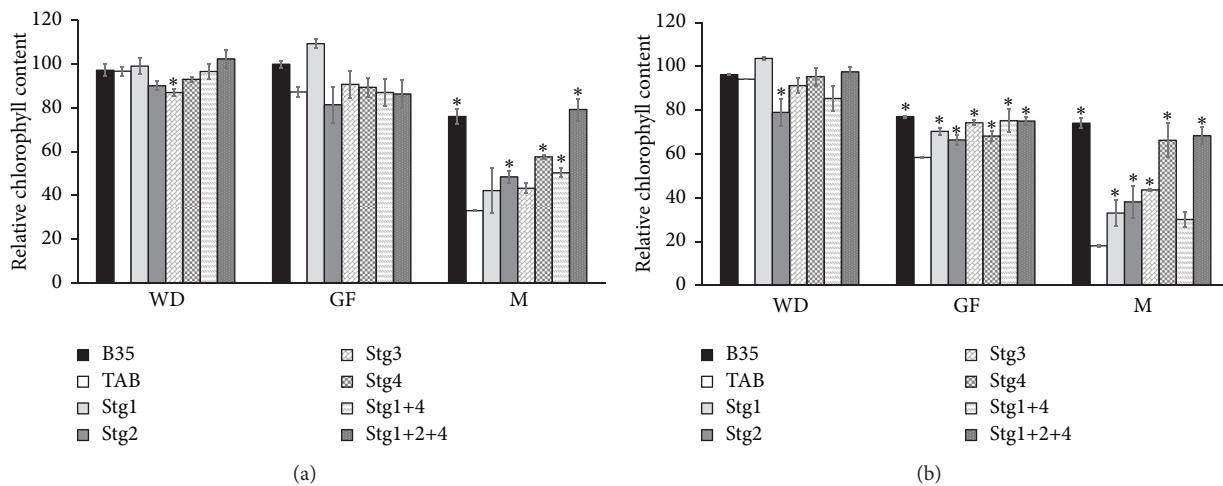

(a)

(b)

Figure 2: Relative chlorophyll content (RCC) at one week before drought (WD), mid-grain filling (GF), and maturity (M) under control W_{100} (a), 75% field capacity W_{75} (b), and 50% filed capacity W_{50} (b) of the six stay-green sorghum introgression lines evaluated with their parents under W_{100}, W_{75}, and W_{50} of soil field capacity. Asterisks indicate significant difference from Tabat (TAB) ($P < 0.05$, Fisher's PLSD test).
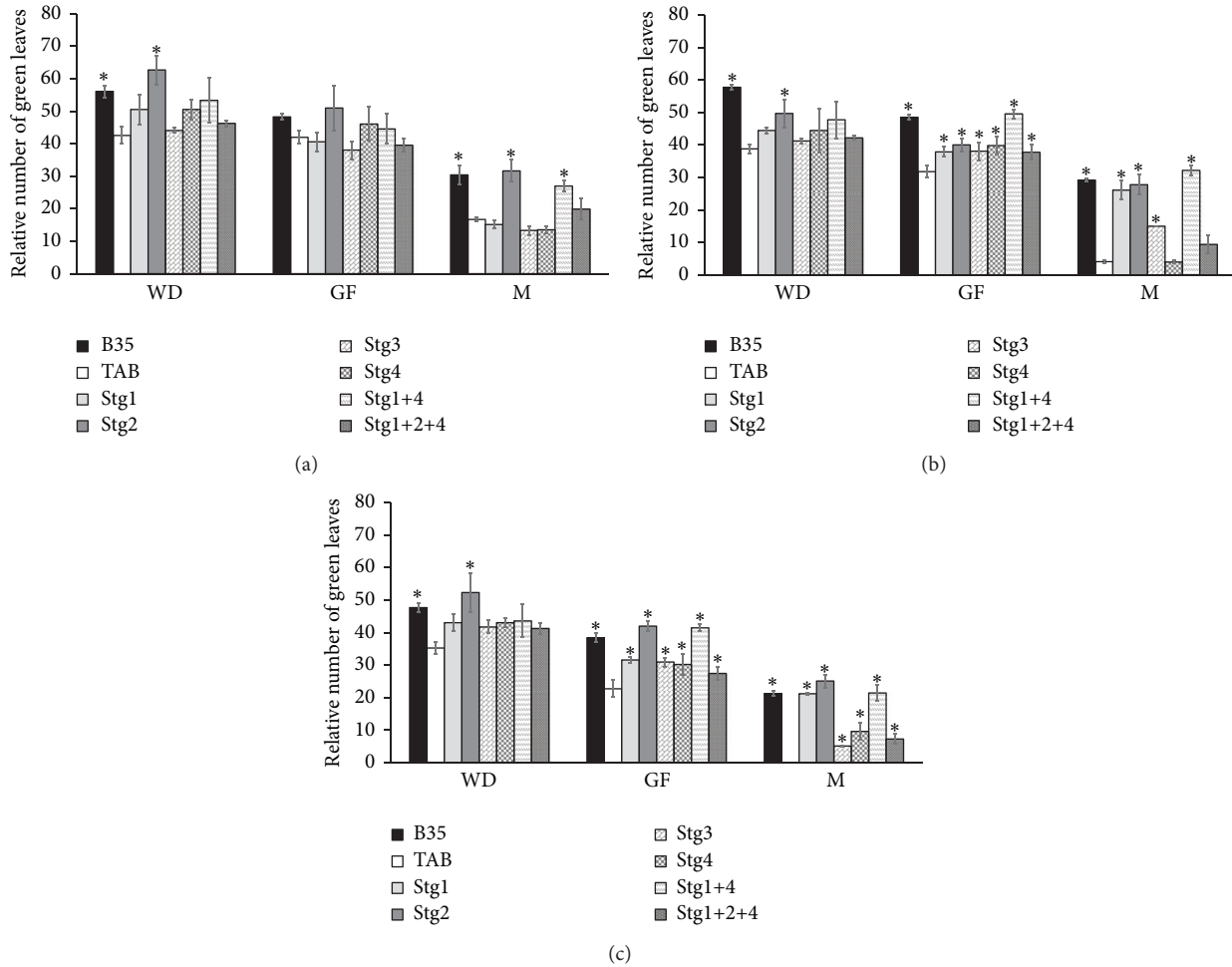
(a)

(b)

(c)

FIGURE 3: Relative number of green leaves/plants (GN) at one week before drought (WD), mid-grain filling (GF), and maturity (M) under control W_{100} (a), 75% field capacity W_{75} (b), and 50% filed capacity W_{50} (c) of the six stay-green sorghum introgression lines evaluated with their parents under W_{100}, W_{75}, and W_{50} of soil field capacity. Asterisks indicate significant difference from Tabat (TAB) ($P < 0.05$, Fisher's PLSD test).

were comparable to TAB, but BIL Stg2 showed a substantial increase in TR (Figure 5(a)).

3.3. The Effect of the Introgressed QTLs under Stress Conditions at W_{75} and W_{50}. Overall, the performance of the stay-green BILs was better than that of the recurrent parent TAB under both W_{75} and W_{50} (Table 1, Figures 1–5).

All the BILs had higher GY than TAB under both W_{75} and W_{50} except BILs Stg2 and Stg4 (Table 1). BIL Stg2 had lower GY than TAB under both W_{75} and W_{50}, whereas BIL Stg4 had comparable GY to that of TAB under W_{75}. TAB and BIL Stg2 showed the highest reduction in GY from W_{100} to W_{75} and W_{50}, whereas B35 and Stg3 showed the lowest reduction. These lines showed rather substantially higher GY under W_{75} than that under W_{100}. This may explain the ability of the stay-green genotypes to increase translocation efficiency under a specific level of drought.

The other Stg lines were intermediate between their parents in their reduction (Table 1). The stay-green QTLs

improved the GY of TAB under drought by different magnitudes, the GY of TAB improved by 25.1% with Stg1+4 to 54.6% with Stg3 under W_{75}, and by 87.5% with Stg1 to 186.1% with Stg3 under W_{50} (Table 1).

To show the degree of drought tolerance conferred by each of the stay-green QTLs, we calculated the YSI at both W_{75} and W_{50}. TAB showed 1.3 and 1.6 YSI at W_{75} and W_{50}, respectively, and BIL Stg2 showed 2.9 and 2.3 at W_{75} and W_{50}. These were classified as sensitive (YSI > 1.00). BIL Stg3 showed -0.36 and -0.08, BIL Stg1+4 showed 0.55 and 0.56, and BIL Stg1+2+4 showed 0.22 and 0.50, under W_{75} and W_{50}, respectively. These were classified as highly tolerant (YSI ≤ 0.50). BIL Stg4 showed 0.73 and 0.67 and was regarded as moderately tolerant (0.50 < YSI ≤ 1.00). BIL Stg1 was tolerant at W_{75} (0.49) but moderately tolerant at W_{50} (0.80).

In BM, BILs Stg1, Stg2, Stg4, and Stg1+4 had higher BM than TAB under both W_{75} and W_{50}. BIL Stg3 which had the highest GY under W_{50} had lower BM than TAB and was the least in the ranking of the genotypes at W_{50} (Table 1). Except for BIL Stg2, all of the BILs and their parents showed slight
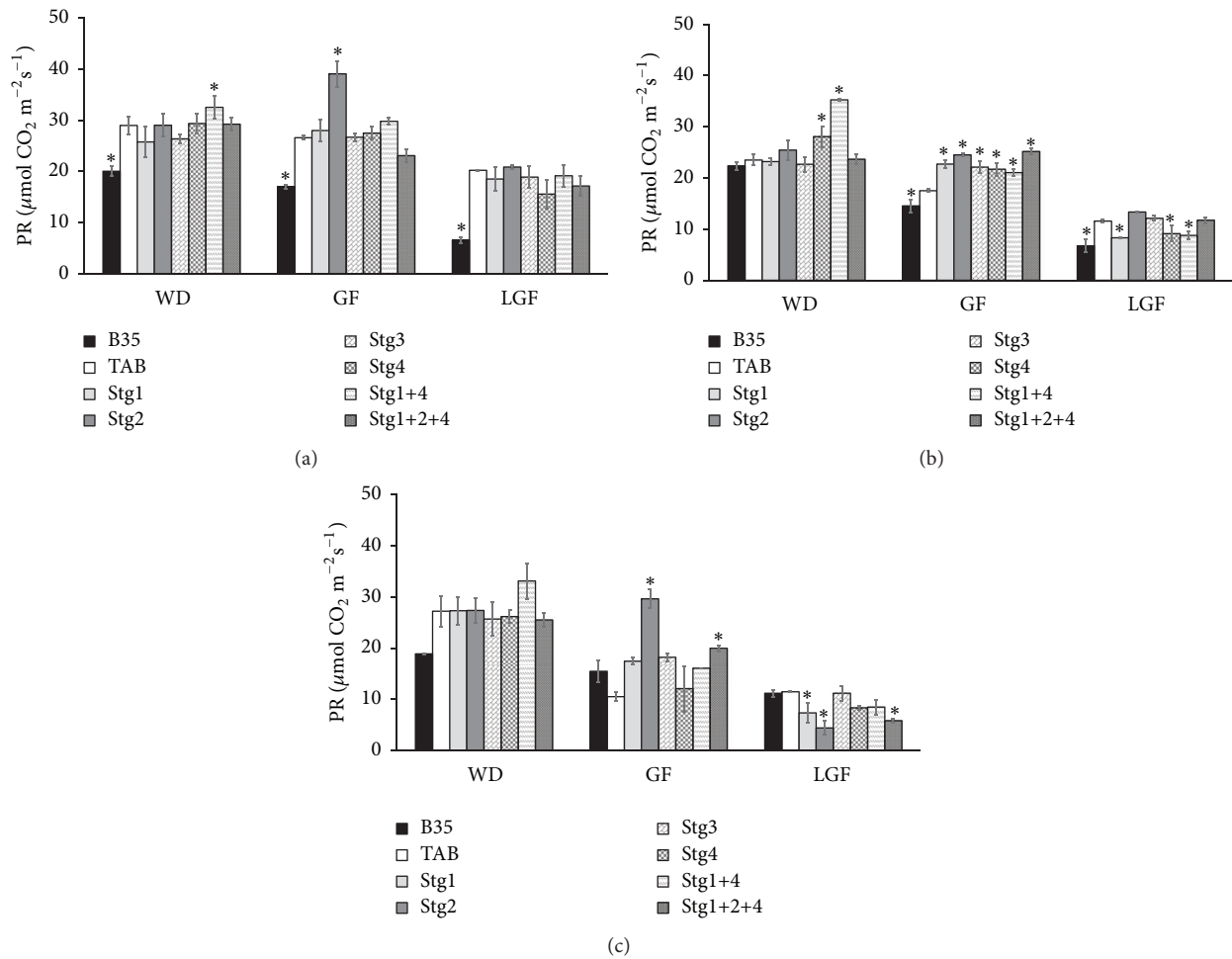
(a)

(b)

(c)

FIGURE 4: Photosynthesis rate (PR) at one week before drought (WD), mid-grain filling (GF), and late-grain filling (LGF) under control W_{100} (a), 75% field capacity W_{75} (b), and 50% filed capacity W_{50} (c) of the six stay-green sorghum introgression lines evaluated with their parents under W_{100}, W_{75}, and W_{50} of soil field capacity. Asterisks indicate significant difference from Tabat (TAB) ($P < 0.05$, Fisher's PLSD test).

or no reduction in BM from W_{100} to W_{75} and W_{50}. BIL Stg2 did not exhibit any reduction and had the highest BM at both W_{75} and W_{50} and was the top in the ranking of the genotypes (Table 1).

In HI, BILs Stg3 and Stg1+2+4 had higher HI than TAB under both W_{75} and W_{50}, whereas BILs Stg2 and Stg1+4 had lower HI than TAB under both treatments (Table 1). Among the BILs, Stg3 was the highest with no reduction from W_{100} to W_{75} and W_{50}, whereas Stg2 was the lowest with the highest reduction from W_{100} to W_{75} and W_{50} (Table 1).

In PH, BILs Stg3 and Stg1+2+4 had higher PH than TAB at both W_{75} and W_{50} (Table 1). At W_{75}, also Stg4 and Stg1+4 had higher PH than TAB. The highest reduction in PH was exhibited by Stg1, whereas Stg1+2+4 did not exhibit any reduction compared to TAB. Stg2 and Stg3 showed a slight reduction (Table 1).

In GLA, under both W_{75} and W_{50} at all stages, Stg1, Stg2, and Stg1+4 had higher GLA than TAB (Figures 1(b) and 1(c)). B35 did not show any reduction, whereas the reduction in TAB, Stg1 and Stg2, was higher than that in the other lines. From W_{100} to W_{50} the reduction in TAB and

B35 was higher than that in the other lines. All the lines showed a reduction of GLA from WD to GF with different magnitudes (Figure 1(b)). Line Stg1+4 showed the lowest reduction from WD to GF. From GF to M the reduction was higher than that from WD to GF (Figures 1(b) and 1(c)).

Generally, RCC of TAB reduced with the progress of the drought and was low at W_{50}, whereas BILs showed improved RCC than TAB (Figure 2). At WD, all the BILs were comparable to TAB, except BIL Stg3 under W_{75} and BIL Stg2 under W_{50} (Figures 2(a) and 2(b)). At GF, all the BILs were comparable to TAB under W_{75}, whereas they had higher RCC than TAB under W_{50} (Figures 2(a) and 2(b)). At M, under W_{75} all BILs except Stg1 and Stg3 had higher RCC than TAB, whereas under W_{50}, except Stg1+4, all the BILs had higher RCC than TAB. B35 had the highest RCC and maintained more than 75% of its RCC from WD to M at both W_{75} and W_{50}, whereas TAB maintained only 20% of its RCC and showed the highest rate of reduction (Figure 2). On the other hand, BILs showed different reduction magnitudes and maintained higher RCC than TAB under the W_{50} and
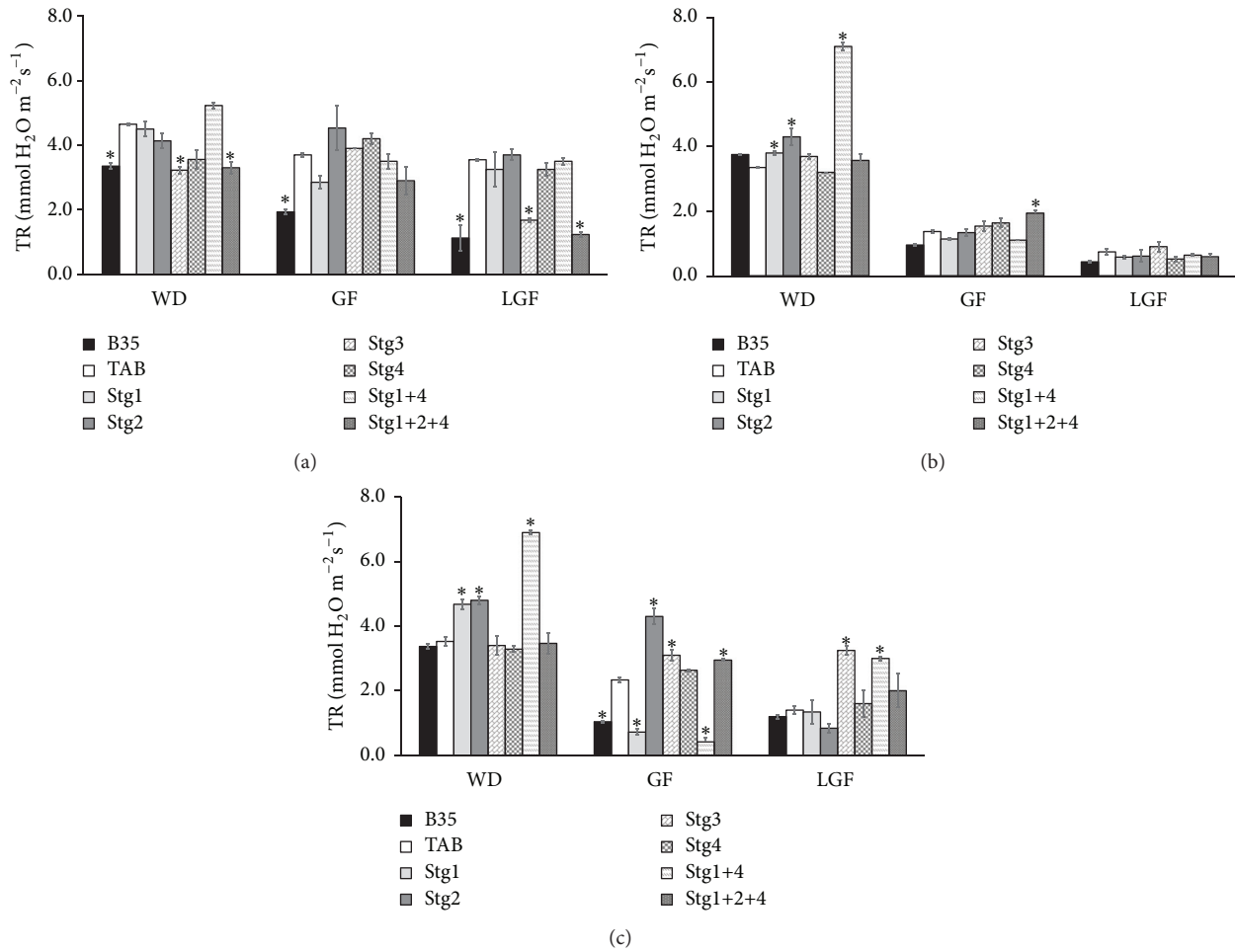
(a)

(b)

(c)

FIGURE 5: Transpiration rate (TR) at one week before drought (WD), mid-grain filling (GF), and late-grain filling (LGF) under control W_{100} (a), 75% field capacity W_{75} (b), and 50% filed capacity W_{50} (c) of the six stay-green sorghum introgression lines evaluated with their parents under W_{100}, W_{75}, and W_{50} of soil field capacity. Asterisks indicate significant difference from Tabat (TAB) ($P < 0.05$, Fisher's PLSD test).

W_{75} treatments, which indicate their ability to retain more chlorophyll content than TAB.

In GN, all the BILs performed better than TAB (Figure 3). At WD, only BIL Stg2 was higher than TAB under both W_{75} and W_{50}, whereas at GF and M, all the BILs were higher than TAB, except Stg4 and Stg1+2+4 at M under W_{75} (Figures 3(b) and 3(c)). Similar to the RCC, all BILs showed less reduction rates compared to TAB, and they were able to partially maintain their GN at M under W_{50} when TAB was completely dry (Figure 3(c)).

In PR, all the BILs and their parent PR decreased with the progress of the drought from W_{75} to W_{50} and from WD to GF and LGF (Figure 4). At WD, the BILs showed comparable PR to that of TAB under both W_{75} and W_{50}, except that Stg4 and Stg1+4 under W_{75} had higher PR than TAB (Figure 4(b)). At GF, under W_{75}, all the BILs had higher PR than TAB, whereas under W_{50}, Stg2 and Stg1+2+4 showed higher PR than TAB (Figures 4(b) and 4(c)). At LGF, some BILs had comparable PR to that of TAB and others showed less PR than that of TAB. At GF the lowest reduction from W_{100} to W_{75} and from W_{100} to W_{50} was observed in line Stg1+2+4

indicating that QTL pyramiding is essential to maintain stable photosynthesis under drought conditions (Figure 4).

As in the PR, the TR decreased with the progress of the drought from W_{75} to W_{50} and from WD to GF and LGF (Figure 5), but interestingly BILs Stg3 and Stg1+2+4 showed the ability to maintain stable TR with the progress of the drought. However, their TR was lower than TAB under W_{100} without drought (Figure 5). At WD, BILs Stg1, Stg2, and Stg1+4 showed higher TR than TAB under both W_{75} and W_{50}. At GF, BIL Stg1+2+4 showed higher TR than TAB under both W_{75} and W_{50}, and BILs Stg2 and Stg3 were higher than TAB under W_{50}. At LGF, under W_{75} Stg3 had substantially higher TR than TAB, whereas, under W_{50}, Stg3 and Stg1+4 had higher TR than TAB (Figure 5).

3.4. Principal Component Analysis. Principal component analysis (PCA) (Figure 6) illustrates the differences between the three water regimes in traits association. Sum of PC1 and PC2 explained 79.8, 63.4, and 62.4% of the total variation in W_{100}, W_{75}, and W_{50}, respectively (Figures 6(a), 6(b), and 6(c)). Under all treatments, PC1 showed high coordination
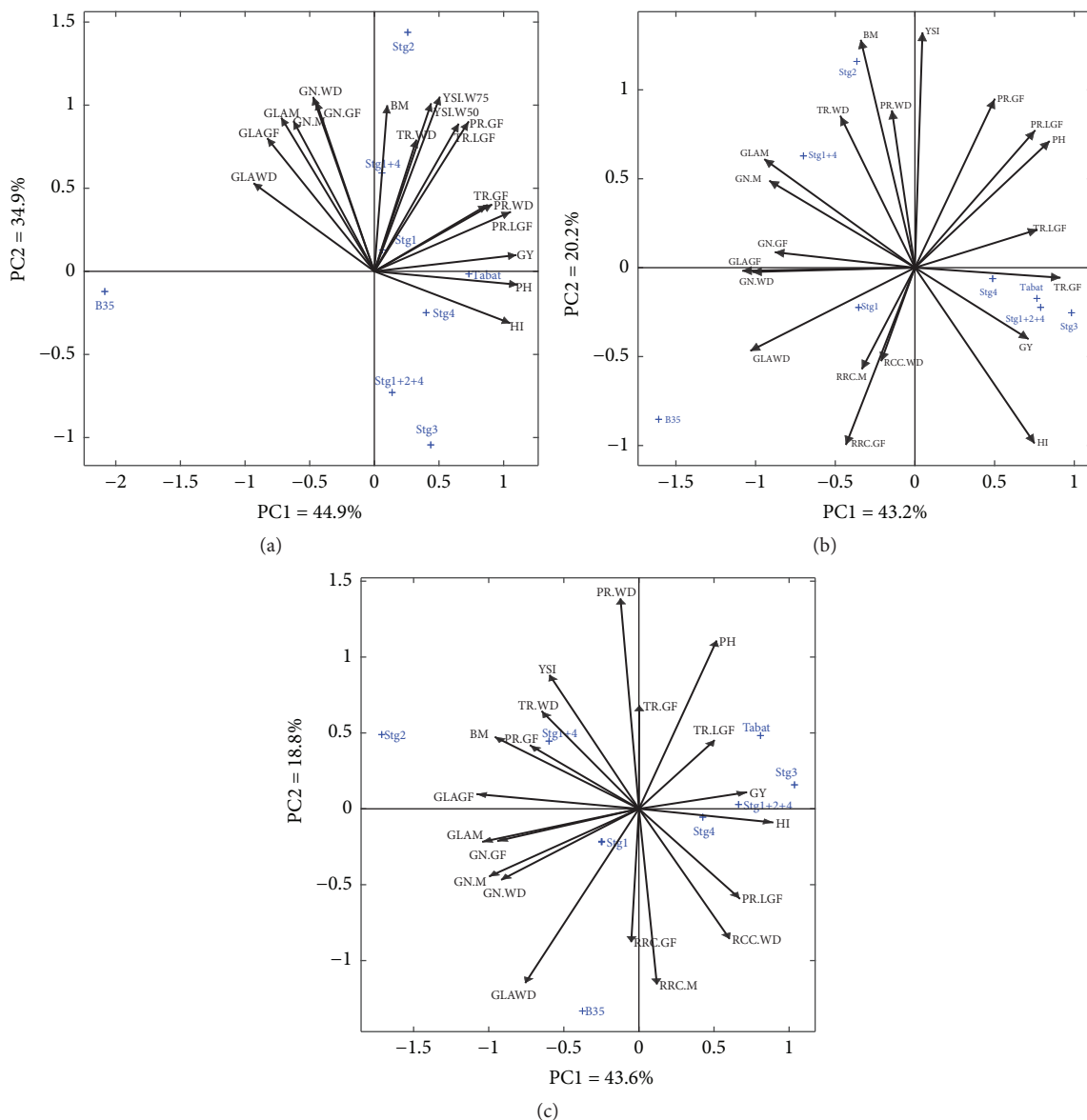
(a)

(b)

(c)

FIGURE 6: PCA analysis of stay-green and yield traits measured in six stay-green sorghum introgression lines and their parents under W_{100} (a), W_{75} (b), and W_{50} (c). Traits included in the PCA are grain yield (GY), biomass (BM), harvest index (HI), plant height (PH), yield susceptibility index (YSI), relative chlorophyll content in SPAD units (RCC), relative number of green leaves/plants (GN), green leaf area (GLA), transpiration rate (TR), and photosynthesis rate (PR). PR and TR were measured at one week before drought (WD), mid-grain filling (GF), and late-grain filling (LGF) whereas RCC, GN, and GLA were measured at WD, GF, and maturity (M).

with GY and HI, whereas PC2 showed high coordination with BM and YSI. PC2 had a negative correlation with GY. Thus, it was called stress susceptibility component. This component separated genotypes with high and low GY in different environments.

Under the W_{100} condition with no drought, GY correlated with the HI, PH, PR, and TR at WD, GF, and LGF. The stay-green traits showed negative and no association with the GY (Figure 6(a)), although YSI of W_{75} and W_{50} were strongly correlated. Under W_{75}, the GY was associated with HI, PH, and PR at WD and GF, and TR at GF and LGF. There was an association between the GY and RCC at GF. GY correlated

negatively with BM and YSI and the stay-green traits (GN and GLA) at all stages (Figure 6(b)). Under W_{50}, the GY was associated with HI, PH, and PR at WD and LGF, TR at LGF, and the RCC at WD and GF (Figure 6(c)). On the other hand, GY was negatively associated with the BM and YSI. There was a close association between BM and stay-green trait (GN and GLA) under all drought treatments.

Selection of genotypes with high PC1 and low PC2 indicates the suitable genotypes for both stress and non-stress environments [42, 43]. Based on the PCA we classified the genotypes into four groups according to their GY under non-stress and stress conditions: genotypes with high GY under

both stress and non-stress conditions (Group A), genotypes with high GY only under non-stress conditions (Group B), genotypes with high GY only under stress conditions (Group C), and at last genotypes with low GY under both conditions (Group D). Thus, Stg1+2+4 and Stg3 with rather higher PC1 and lower PC2 are superior genotypes under both stressed and non-stressed conditions (Figures 6(a), 6(b), and 6(c)). These genotypes had stable performance in the circumstances of low sensitivity to drought stress. Therefore, they belong to Group A. Stg2 could be known as Group D. This genotype is drought sensitive and had low GY and HI under drought conditions.

3.5. Regression Analysis. We applied regression analysis to provide more information on the significance of the important associations identified by the PCA analysis. The results indicated that HI contribution to the GY in the stay-green BILs was significant under the three water regimes (W_{100}, W_{75}, and W_{50}) (Figure S1). The PH had a significant association with the GY only under W_{100} without drought (Figure S2). Under W_{75}, PR was the most contributing factor ($R^2=0.85^{**}$), whereas under W_{50}, TR at GF was the most important factor contributing to GY ($R^2 = 0.69^*$) (Figure S3). GLA at M was strongly correlated with GN ($R^2 = 79^{**}$, $R^2 = 0.87^{**}$ and $R^2 = 0.87^{**}$) under W_{100}, W_{75}, and W_{50}, respectively (Figure S4).

4. Discussion

Stay-green is positively correlated with sorghum yield under post-flowering drought stress [10]. Although a positive correlation between stay-green and GY has been demonstrated in earlier studies, the physiological and molecular basis of the stay-green trait remains unclear [30]. Our results showed that stay-green QTLs affected a number of traits under terminal drought conditions in sorghum, and the significance and magnitude of the effects depended critically on the drought severity.

4.1. The Drought Threshold for the Stay-Green Lines. We performed this study at three levels of soil FC, W_{100}, W_{75}, and W_{50}. Our results indicated that the performance of the stay-green lines reduced with the reduction in soil moisture content in all traits. W_{50} was the most effective treatment in classifying and reducing the performance of the stay-green lines, and the most affected traits were the PR at LGF and GLA (Tables 1 and 2). From W_{100} to W_{50}, the GY of TAB reduced by 70%, whereas that of the BILs reduced by only 36%. This finding shows clearly the effectiveness of the stay-green QTLs in enhancing the adaptation of TAB to the post-flowering drought. Interestingly, we observed clear expression and response of the stay-green trait under the moderate drought (W_{75}) in terms of GLA and GN at all stages and RCC at maturity. An earlier report of Mahalakshmi and Bidinger [44] indicated that moderate but prolonged terminal drought stress during GF is the ideal environment for evaluating the stay-green trait. However, in this study we could identify that drought as 75% of the soil FC could

induce the expression of the stay-green and, hence, it could be enough to evaluate genotypes for stay-green traits.

4.2. The Impact of the Stay-Green QTLs under Non-Stress Conditions. Understanding the effect of the stay-green QTLs in the modification of plant performance under adequate soil moisture is essential as drought is fluctuating from year to year and place to place, and thus elasticity in crop performance is essential to have good yield with less or adequate water. Our results indicated that the single stay-green QTLs have no negative impact on the GY of TAB. However, the combination of Stg1+4 showed a decrease in TAB GY under W_{100} conditions (Table 1). Other stay-green QTLs showed different impacts in TAB background; Stg1, Stg2, and Stg1+4 increased the GLA (Figure 1) and Stg3 and Stg1+2+4 decreased the TR at WD and LGF but not at the GF (Figure 5). The PCA analysis indicated that HI contributes to the GY (Figure 6(a)). Thus, we attribute the lower GY of Stg1+4 to its low HI. Borrell [30] reported that stay-green QTLs have no consistent yield penalty under irrigated conditions without drought. We attribute this contradiction of findings to the following: (1) Borrell [30] studied the effect of single QTLs only and did not study the effect of multiple QTLs, (2) there is a difference between the genetic backgrounds used in this study and that employed by Borrell [30], and (3) there is linkage drag as these lines still at BC_3. Vadez [9] showed that the stay-green QTLs, Stg1 and Stg3, decreased the tillering and leaf area in S35 background, whereas there was no such effect in R16 background. Thus, they concluded that the impact of the stay-green QTLs depends on their interaction with the genetic background. Therefore, it is important for the breeding programs to consider this interaction to assure good performance and yield in the wet periods.

4.3. Association between the Studied Traits and GY under Normal and Drought Conditions. Understanding the association between the traits and the GY under different soil moisture contents is a prerequisite to decide which traits should be focused on in the breeding programs to increase the GY and also to understand how the stay-green contributes to increasing or stabilization of GY under drought conditions. Our results indicated that HI, PH, TR, and PR are the major traits contributing to the GY at all soil moisture levels (Table 1, Figure 6). Interestingly, the TR contribution to the GY differs among the treatments; it was important at the W_{100} condition and W_{75} at GF and LGF, whereas it was important at LGF at W_{50}. The contribution of the PR varied with the variation in the soil moisture content. The contribution was high at the W_{100} and W_{75}, whereas it was less at W_{50} (Figures 6(a), 6(b), and 6(c)). At the W_{75} and W_{50}, PR at LGF was more important than at WD and GF for higher GY. These findings indicated clearly that the response of the stay-green lines to the post-flowering drought depends on the drought severity level. These findings are consistent with the findings of Vadez [9]. They concluded that variation in GY of stay-green QTLs in different genetic backgrounds was due to HI and transpiration efficiency. The difference between our study and Vadez [9] is that they estimated

the transpiration efficiency by measuring the water supply and consumption, and we estimated the TR of the plants leaves by direct measurement of the leaf activity under two levels of drought at three different developmental stages. Thus, we were able to understand the change in the plant behavior with the progress of drought and development of the plant. Our findings and that of Vadez [9] indicate that the stay-green genotypes stabilize their GY under drought by manipulation of their behavior of water uptake and utilization, photosynthesis performance, and increasing the mobilization of the photo-assimilate to the grains (high HI). This manipulation of physiological performance or behavior depends on the degree of drought severity and the genetic background.

In this study, PH was associated with GY at the three water regimes (Figures 6(a), 6(b), and 6(c)). The association was high and significant under W_{100} compared to that at W_{75} and W_{50} (Figure S2). This could be explained by the findings of Sabadin [45] that PH QTLs were colocalized with the GY and stay-green QTLs. On the other hand, taking into consideration that also HI association with GY was higher under W_{100} compared to that under W_{75} and W_{50} (Figure S1), we can suggest that stay-green BILs are less reliant on the stem reserve under drought and operate another mechanism to maintain or stabilize their GY under drought. In addition, high GY could be reasonably predicted from PH ($R^2 = 0.92^{***}$) under non-stress environment.

HI was correlated with RCC at WD in W_{100} and W_{75} but not in W_{50}, where it was correlated with PR and RCC at GF and M (Figure 6). These results implied that remobilized reserves explained at least more than 60% of the variation observed in the GY of stay-green BILs tested under normal and moderate or severe drought condition. Ongom [46] suggested that plants with high remobilization could perform well under post-flowering drought. Under severe drought, PR and RCC at GF and M could be good indicators for high GY. Furthermore, PR had a positive association with GLA only under W_{50}, and this may explain its role under severe drought and could be explained by the findings of Swain [47]; that is, variation in photosynthesis is associated with leaf protein content.

The GLA and RCC were correlated at WD, GF, and M under W_{100}, W_{75}, and W_{50} and were not correlated with the GY (Figure 6). These findings indicate that higher leaf senescence is due to higher translocation of food reserve from leaves to grains for better grain filling and increased GY as reported by Reddy [25]. This finding also explains the importance of the high HI trait in the stabilization of the GY under the drought conditions. This is typically the case of the BIL Stg3 as it had more HI and GY than the other BILs, but it had low GLA and RCC compared to the other BILs. Furthermore, GLA at M and GF stages were significant and positively correlated with GN under all soil moisture levels (W_{100}, W_{75}, and W_{50}) indicating that GN can be used as easy/fast indicator for stay-green trait instead of GLA.

4.4. Stay-Green QTLs Contribution to the GY Stabilization under Drought.

As a result of their low rate of leaf senesce,

all the stay-green single QTLs improved or increased the GN under both the moderate and the severe drought (Figure 3), especially at M. Under W_{75}, Stg3 and Stg4 did not affect the GLA (Figure 1), whereas at W_{50} all the Stg QTLs increased/maintained high GLA. These results indicate that the impact of the Stg QTLs depends on the degree of the drought severity. Our results indicated that Stg2 and Stg1+4 had the lowest reduction in GN and GLA, and similar results were reported by Jordan [48].

The stay-green trait is positively correlated with GY in field conditions under terminal drought [6, 8, 10, 47]. The QTLs contribution in the GY varied with a variation in the soil moisture content or, in other words, with the severity of the drought. At the W_{100} condition, none of the QTLs increased the GY in TAB, and Stg1+4 decreased the GY, whereas at W_{75} and W_{50} all the stay-green QTLs increased the GY. Stg3 and the combination of Stg1+2+4 were the most efficient QTLs in terms of GY performance. These lines decreased the TR under W_{100} condition, whereas under drought both possessed higher TR. Interestingly the high TR in these two lines was related to specific stages depending on the drought severity. At W_{75} Stg3 increased the TR at WD and LGF, whereas at W_{50} the TR was high at the GF and LGF. Stg1+2+4 increased the TR at GF under both W_{75} and W_{50} (Figure 5). In addition, these two lines possessed high HI and increased PR at GF and LGF (Table 1, Figure 5). Based on these results we attribute the high GY of Stg3 and Stg1+2+4 to their high HI, PR, and TR. Stg3 is found to be positively important for improving GY under post-flowering drought stress (Table 1), and similar results were reported by Reddy [25]. Also, Sabadin [45] pointed out the colocalization of Stg3 and GY QTL and suggested the potential of indirect selection based on stay-green to improve sorghum GY under drought.

Stg2 showed very low HI and had a great reduction in GY under W_{75} and W_{50} compared to the other Stgs (Table 1), although it has a lower reduction in GN (Table 1). This result explains that variation in GY reduction depends on the stress severity since the variation in GY/panicle was found to be a function of terminal drought [39]. Moreover, introducing Stg QTL into highly senescent background could affect the sink source relationship as reported by earlier studies of Kassahun [49]. Our results confirmed previous reports that Stg2 is an important QTL for maintaining higher GLA contributing to slow senescence (Figure 4). This QTL was also reported to contribute to higher GLA at WD and M [50] and to %GLA at 45 days after flowering [15] in different genetic backgrounds. Reddy [25] indicated that the expression of Stg2 QTL was consistent and formed an important QTL for marker-assisted improvement of post-rainy sorghum lines for terminal drought tolerance. We attributed the low GY observed for Stg2 in our study to the presence of linkage drag and the difference of the genetic background and the environments.

4.5. Impact of the QTL Pyramiding.

In this study, we compared the effect of the single QTLs and two combinations of double and triple QTLs on the adaptation to post-flowering drought stress. The performance in terms of GY of the

single QTL Stg3 under both drought treatments W_{75} and W_{50} was similar or better than that of the double or triple QTLs (Table 1). However, the effect of the QTL pyramiding was evident in the tolerance of the BILs; using the YSI we classified Stg1 and Stg4 as moderately tolerant but when these QTLs were combined in Stg1+4 the tolerance increased, and Stg1+4 was classified as tolerant. Interestingly, when the sensitive QTL Stg2 was coupled with Stg1+4 the combination Stg1+2+4 was classified as tolerant. On the other hand, Stg2 had lower GY and HI under drought, but, when combined with Stg1+4 (the combination Stg1+2+4), it increased the GY and HI under drought (Table 1). These findings indicate that QTL pyramiding can enhance the adaptation to post-flowering drought. However, this effect needs to be investigated across different backgrounds and environments, as the stay-green QTLs effect on improving adaptation to post-flowering drought was found to be dependent on the genetic background and the environment [9, 51]. In addition, in this study, the single QTL Stg3 had comparable GY to that of the combination Stg1+2+4 which suggests that QTL pyramiding might not always be necessary depending on the environment and genetic background.

4.6. The Putative Model of Stay-Green Adaptation to Post-Flowering Drought. Using stay-green introgression lines in the background of RTX7000, Borrell [5, 30] concluded that stay-green genotypes adapt to the post-flowering drought through decreased tillering and the size of upper leaves, which reduced canopy size at flowering. This reduction in transpirational leaf area reduced pre-flowering water demand, thereby increasing water availability during GF and, ultimately, GY. Recently, Borrell [30] using the same introgression lines reported that tiller leaf area rather than transpiration efficiency, or transpiration per leaf area, was the main driver of weekly transpiration and the reduced pre-flowering water use in stay-green lines. In this study, we did not observe any reduction in the GLA of the stay-green lines, and stay-green QTLs Stg1, Sg2, and Stg1+4 increased the GLA of the senescent parent TAB before flowering under both control and drought conditions. We attribute this contradiction in findings to the difference in the genetic backgrounds used, especially that Vadez [9] showed that the stay-green QTLs, Stg1, and Stg3 decreased the tillering and leaf area in S35 background, whereas there was no such effect in R16. This contradiction demonstrates that (1) the stay-green plants adapt to the post-flowering drought through other different mechanisms and not only GLA reduction and water saving before flowering and (2) the stay-green effect depends largely on the genetic background.

In these earlier reports, plants water utilization behavior was evaluated by measuring the plant water consumption and transpiration efficiency. In this study, we measured the actual leaf TR at three different stages under two different levels of post-flowering drought severities. This enabled us to examine in more detail the behavior of the Stg QTLs in the modification of the plant behavior under drought. Thus, based on our results and the other reports we can suggest that stay-green genotypes adapt to post-flowering drought by reducing the transpirational leaf area and the TR per leaf that reduce pre-flowering water demand, thereby increasing water availability during grain filling and utilizing the conserved water depending on the drought severity and the genetic background.

5. Conclusion

In conclusion, our results clearly showed that the stay-green QTLs enhance post-flowering drought response to a level up to 50% of the soil FC and the stay-green trait is expressed under the moderate drought (W_{75}). The stay-green QTLs help to increase or stabilize the GY under drought through efficient water utilization (TR) depending on the drought severity coupled with the high rate of photo-assimilates translocation (high HI). QTL pyramiding could increase the drought tolerance but might not always be necessary to stabilize and increase the GY under post-flowering drought. The understanding of the physiological mechanisms associated with drought severity, senescence, and photosynthetic efficiency and the connection between QTL expression/interaction with genetic background and physiological response to drought could be the key to remove the plateau of productivity associated with sorghum adaptation to unfavorable environmental conditions.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Acknowledgments

Supplementary Materials

Figure S1. Relationship between grain yield (GY) and harvest index (HI) for six stay-green introgression lines and their parents evaluated under W_{100} (a), W_{75} (b) and W_{50} (c) of soil field capacity. ∗ denote significant differences at $P \leq 0.05$.
Figure S2. Relationship between grain yield (GY) and plant height (PH) for six stay-green introgression lines and their parents evaluated under W_{100} (a), W_{75} (b) and W_{50} (c) of soil field capacity. ∗ ∗ ∗ denote significant differences at $P \leq 0.001$, ns denote not significant. **Figure S3.** Relationship between photosynthesis rate (PR) at grain filling and grain yield (GY) under W_{100} (a), transpiration rate (TR) at grain filling and grain yield (GY) under W_{50} (b) for six stay-green introgression lines and their parents. ∗, ∗∗ denote significant differences at $P \leq 0.05$ and $P \leq 0.01$, respectively. **Figure S4.** Relationship between green leaf area at grain filling and maturity (GLA.GF and GLA.M, respectively) and relative number of green leaves/plant under W_{100} (a), W_{75} (b) and W_{50} (c) for six stay-green introgression lines and their parents. ∗, ∗∗, ∗∗∗ denote significant differences at $P \leq 0.05$, $P \leq 0.01$, $P \leq 0.001$, respectively. *(Supplementary Materials)*

References

[1] C. Haub, "World population data sheet," 2013, http://www.prb.org/pdf13/2013-population-data-sheet_eng.pdf.

[2] United Nations Development Programme, *Africa Human Development Report 2012, Towards A Food Secure Future*, UNDP, New York, USA, 2012.

[3] J. Vidal, "Climate change: how a warming world is a threat to our food supplies," *The Observer*, 2013.

[4] C. Nguyen, M. Mannino, K. Gardiner, and K. J. Cio, "ClusFCM: An algorithm for predicting protein function using homologies and protein interactions," *Journal of Bioinformatics and Computational Biology*, vol. 6, no. 1, pp. 203–222, 2008.

[5] A. K. Borrell, J. E. Mullet, B. George-Jaeggli et al., "Drought adaptation of stay-green sorghum is associated with canopy development, leaf anatomy, root growth, and water uptake," *Journal of Experimental Botany*, vol. 65, no. 21, pp. 6251–6263, 2014.

[6] A. K. Borrell, F. R. Bidinger, and K. Sunitha, "Stay-green associated with yield in recombinant inbred sorghum lines varying in rate of leaf senescence," *Intl Sorghum and Millets Newsl*, vol. 40, pp. 31–33, 1999.

[7] A. K. Borrell, G. L. Hammer, and A. C. L. Douglas, "Does maintaining green leaf area in sorghum improve yield under drought? I. Leaf growth and senescence," *Crop Science*, vol. 40, no. 4, pp. 1026–1037, 2000.

[8] A. K. Borrell, G. L. Hammer, and R. G. Henzell, "Does maintaining green leaf area in sorghum improve yield under drought? II. Dry matter production and yield," *Crop Science*, vol. 40, no. 4, pp. 1037–1048, 2000.

[9] V. Vadez, S. P. Deshpande, J. Kholova et al., "Stay-green quantitative trait loci's effects on water extraction, transpiration efficiency and seed yield depend on recipient parent background," *Functional Plant Biology*, vol. 38, no. 7, pp. 553–566, 2011.

[10] D. R. Jordan, C. H. Hunt, A. W. Cruickshank, A. K. Borrell, and R. G. Henzell, "The relationship between the stay-green trait and grain yield in elite sorghum hybrids grown in a range of environments," *Crop Science*, vol. 52, no. 3, pp. 1153–1161, 2012.

[11] D. T. Rosenow, "Breeding for lodging resistance in sorghum," in *Proceedings of the 32nd Annual Corn and Sorghum Research Conference*, H. D. Loden and D. Wilkinson, Eds., pp. 171–185, American Seed Trade Association, Washington, DC, USA, 1977.

[12] R. G. Henzell, R. Brengman, D. Fletcher, and T. McCosker, "Relationships between yield and non-senescence ('stay-green') in some grain sorghum hybrids grown under terminal drought stress," in *Proceedings of the Second Australian Sorghum Conference*, M. A. Foale, R. G. Henzell, and P. N. Vance, Eds., pp. 355–359, Australian Institute of Agricultural Science, Melbourne, Australia, 1992.

[13] D. T. Rosenow, J. E. Quisenberry, C. W. Wendt, and L. E. Clark, "Drought tolerant sorghum and cotton germplasm," *Agricultural Water Management*, vol. 7, no. 1-3, pp. 207–222, 1983.

[14] H. Kebede, P. K. Subudhi, D. T. Rosenow, and H. T. Nguyen, "Quantitative trait loci influencing drought tolerance in grain sorghum (Sorghum bicolor L. Moench)," *Theoretical and Applied Genetics*, vol. 103, no. 2-3, pp. 266–276, 2001.

[15] B. I. G. Haussmann, V. Mahalakshmi, B. V. S. Reddy, N. Seetharama, C. T. Hash, and H. H. Geiger, "QTL mapping of stay-green in two sorghum recombinant inbred populations," *Theoretical and Applied Genetics*, vol. 106, no. 1, pp. 143–148, 2002.

[16] M. R. Tuinstra, E. M. Grote, P. B. Goldsbrough, and G. Ejeta, "Identification of quantitative trait loci associated with pre-flowering drought tolerance in sorghum," *Crop Science*, vol. 36, no. 5, pp. 1337–1344, 1996.

[17] M. R. Tuinstra, E. M. Grote, P. B. Goldsbrough, and G. Ejeta, "Genetic analysis of post-flowering drought tolerance and components of grain development in Sorghum bicolor (L.) Moench," *Molecular Breeding*, vol. 3, no. 6, pp. 439–448, 1997.

[18] M. R. Tuinstra, G. Ejeta, and P. Goldsbrough, "Evaluation of near-isogenic sorghum lines contrasting for QTL markers associated with drought tolerance," *Crop Science*, vol. 38, no. 3, pp. 835–842, 1998.

[19] O. R. Crasta, W. Xu, D. T. Rosenow, J. Mulletand, and H. T. Nguyen, "Mapping of post-flowering drought resistance traits in grain sorghum: association between QTLs influencing premature senescence and maturity," *Molecular Genetics and Genomics*, vol. 262, pp. 579–588, 1999.

[20] J.-M. Boffa, S. J.-B. Taonda, J. B. Dickey, and D. M. Knudson, "Field-scale influence of karite (Vitellaria paradoxa) on sorghum production in the Sudan zone of Burkina Faso," *Agroforestry Systems*, vol. 49, no. 2, pp. 153–175, 2000.

[21] P. K. Subudhi and H. T. Nguyen, "New horizons in biotechnology," in *Sorghum: Origin, History, Technology, and Production*, Smith et al., Ed., pp. 349–397, John Wiley & Sons, New York, USA, 2000.

[22] Y. Z. Tao, R. G. Henzell, D. R. Jordan, D. G. Butler, A. M. Kelly, and C. L. McIntyre, "Identification of genomic regions associated with stay green in sorghum by testing RILs in multiple environments," *Theoratical and Appllied Genetics*, vol. 100, pp. 1225–1232, 2000.

[23] W. Xu, K. Subudhi, O. R. Crasta, D. T. Rosenow, J. E. Mullet, and H. T. Nguyen, "Molecular mapping of QTLs conferring stay-green in grain sorghum (*Sorghum bicolor* L. Moench)," *Genome*, vol. 43, pp. 461–469, 2000.

[24] A. C. Sanchez, P. K. Subudhi, D. T. Rosenow, and H. T. Nguyen, "Mapping QTLs associated with drought resistance in sorghum

(*Sorghum bicolor* L. Moench),” *Plant Molecular Biology*, vol. 84, pp. 713–726, 2002.

[25] N. R. R. Rama Reddy, M. Ragimasalawada, M. M. Sabbavarapu, S. Nadoor, and J. V. Patil, “Detection and validation of stay-green QTL in post-rainy sorghum involving widely adapted cultivar, M35-1 and a popular stay-green genotype B35,” *BMC Genomics*, vol. 15, no. 1, p. 909, 2014.

[26] A. K. Borrell and G. L. Hammer, “Nitrogen dynamics and the physiological basis of stay-green in Sorghum,” *Crop Science*, vol. 40, no. 5, pp. 1295–1307, 2000.

[27] A. K. Borrell, G. L. Hammer, and E. Van Oosterom, “Stay-green: A consequence of the balance between supply and demand for nitrogen during grain filling?” *Annals of Applied Biology*, vol. 138, no. 1, pp. 91–95, 2001.

[28] K. Harris, P. K. Subudhi, A. Borrell et al., “Sorghum stay-green QTL individually reduce post-flowering drought-induced leaf senescence,” *Journal of Experimental Botany*, vol. 58, no. 2, pp. 327–338, 2007.

[29] A. K. Borrell, E. Van Oosterom, G. Hammer, D. Jordan, and A. Douglas, “The physiology of stay-green in sorghum,” in *Proceedings of the 11th Australian Agronomy Conference*, M. Unkovich, G. O’Leary, and M. sorghum. Unkovich, Eds., Solutions for a better environment, Geelong, Victoria, 2003.

[30] A. K. Borrell, E. J. van Oosterom, J. E. Mullet et al., “Stay-green alleles individually enhance grain yield in sorghum under drought by modifying canopy development and water uptake patterns,” *New Phytologist*, vol. 203, no. 3, pp. 817–830, 2014.

[31] B. G. Jaegglia, M. Y. Mortlockb, and A. Borrell, “Bigger is not always better: Reducing leaf area helps stay-green sorghum use soil water more slowly,” *Environmental and Experimental Botany*, vol. 138, pp. 119–129, 2017.

[32] A. M. Ali, N. M. Kamal, I. Noureldin, Y. Hiraoaka, Y. Yamauchi, and Y. Sugimoto, “Marker assisted breeding of the stay-green trait of sorghum to enhance terminal drought tolerance for Sudan: Candidate donor and recipient genotypes,” *Sudan Journal of Agricultural Research*, vol. 10, pp. 133–141, 2007.

[33] N. M. Kamal, Y. S. A. Gorafi, and A. M. A. Ghanim, “Performance of Sorghum Stay-green Introgression Lines Under Post-Flowering Drought,” *International Journal of Plant Research*, vol. 7, pp. 65–74, 2017.

[34] A. A. Salih, I. A. Ali, A. Lux et al., “Rooting, water uptake, and xylem structure adaptation to drought of two sorghum cultivars,” *Crop Science*, vol. 39, no. 1, pp. 168–173, 1999.

[35] W. Tsuji, M. E. K. Ali, S. Inanaga, and Y. Sugimoto, “Growth and gas change of three sorghum cultivars under drought stress,” *Biology Plant*, vol. 46, pp. 583–587, 2003.

[36] H. Fujiyama and T. Nagai, “Studies on improvement of nutrient and water supply in crop cultivation on sand dune soil, I. comparison of irrigation methods,” *Soil Science & Plant Nutrition*, vol. 32, no. 4, pp. 511–521, 1986.

[37] Q. Sohail, T. Inoue, H. Tanaka, A. E. Eltayeb, Y. Matsuoka, and H. Tsujimoto, “Applicability of Aegilops tauschii drought tolerance traits to breeding of hexaploid wheat,” *Breeding Science*, vol. 61, no. 4, pp. 347–357, 2011.

[38] R. A. Fischer and R. Maurer, “Drought resistance in spring wheat cultivars. I. Grain yield responses,” *Australian Journal of Agricultural Research*, vol. 29, no. 5, pp. 897–912, 1978.

[39] R. Khanna-Chopra, P. S. S. Rao, M. Maheswari, L. Xiaobing, and K. S. Shivshankar, “Effect of Water Deficit on Accumulation of Dry Matter, Carbon and Nitrogen in the Kernel of Wheat Genotypes Differing in Yield Stability,” *Annals of Botany*, vol. 74, no. 5, pp. 503–511, 1994.

[40] G. L. Hammer, P. S. Carberry, and R. C. Muchow, “Modelling genotypic and environmental control of leaf area dynamics in grain sorghum. I. Whole plant level,” *Field Crops Research*, vol. 33, no. 3, pp. 293–310, 1993.

[41] R. C. Muchow and P. S. Carberry, “Phenology and leaf-area development in a tropical grain sorghum,” *Field Crops Research*, vol. 23, no. 3-4, pp. 221–237, 1990.

[42] M. Golabadi, A. Arzani, Mirmohammadi, and S. A. M. Meibody, “Assessment of drought tolerance in segregating populations in durum wheat,” *African Journal of Agricultural Research*, vol. 5, pp. 162–171, 2006.

[43] R. Shahryari and V. Mollasadeghi, “Harvest index and its Associated characters in Winter wheat genotypes against terminal drought at presence of a peat derived humic fertilizer,” *Advances in Environmental Biology*, vol. 5, no. 1, pp. 162–165, 2011.

[44] V. Mahalakshmi and F. R. Bidinger, “Evaluation of stay-green sorghum germplasm lines at ICRISAT,” *Crop Science*, vol. 42, no. 3, pp. 965–974, 2002.

[45] P. K. Sabadin, M. Malosetti, M. P. Boer et al., “Studying the genetic basis of drought tolerance in sorghum by managed stress trials and adjustments for phenological and plant height differences,” *Theoretical and Applied Genetics*, vol. 124, no. 8, pp. 1389–1402, 2012.

[46] P. O. Ongom, J. J. Volenec, and G. Ejeta, “Selection for drought tolerance in sorghum using desiccants to simulate post-anthesis drought stress,” *Field Crops Research*, vol. 198, pp. 312–321, 2016.

[47] M. A. Swain, R. W. Peet, and D. A. Galloway, “Characterization of the gene encoding herpes simplex virus type 2 glycoprotein C and comparison with the type 1 counterpart,” *Journal of Virology*, vol. 53, no. 2, pp. 561–569, 1985.

[48] D. R. Jordan, Y. Tao, I. D. Godwin, R. G. Henzell, M. Cooper, and C. L. McIntyre, “Prediction of hybrid performance in grain sorghum using RFLP markers,” *Theoratical and Appllied Genetics*, vol. 106, pp. 559–567, 2003.

[49] B. Kassahun, F. Bidinger, C. Hash, and M. Kuruvinashetti, “Stay-green expression in early generation sorghum [*Sorghum bicolor* (L.) Moench] QTL introgression lines,” *Euphytica*, vol. 172, no. 3, pp. 351–362, 2010.

[50] G. Srinivas, K. Satish, R. Madhusudhana, R. Nagaraja Reddy, S. Murali Mohan, and N. Seetharama, “Identification of quantitative trait loci for agronomically important traits and their association with genic-microsatellite markers in sorghum,” *Theoretical and Applied Genetics*, vol. 118, no. 8, pp. 1439–1454, 2009.

[51] B. V. S. Reddy, P. S. Reddy, A. R. Sadananda et al., “Postrainy season sorghum: Constraints and breeding approaches,” *Journal of SAT Agricultural Research*, vol. 10, pp. 1–12, 2012.